Hardware Sizing Guide

About Genesys

Genesys is the world's leading provider of customer service and contact center software—with more than 4,000 customers in 80 countries. Drawing on its more than 20 years of customer service innovation and experience, Genesys is uniquely positioned to help companies bring their people, insights and customer channels together to effectively drive today's customer conversation. Genesys software directs more than 100 million interactions every day, maximizing the value of customer engagement and differentiating the experience by driving personalization and multi-channel customer service—and extending customer service across the enterprise to optimize processes and the performance of customer-facing employees. Go to www.genesys.com for more information.

Each product has its own documentation for online viewing at the Genesys Documentation website or on the Documentation Library DVD, which is available from Genesys upon request. For more information, contact your sales representative.

Notice

Although reasonable effort is made to ensure that the information in this document is complete and accurate at the time of release, Genesys Telecommunications Laboratories, Inc., cannot assume responsibility for any existing errors. Changes and/or corrections to the information contained in this document may be incorporated in future versions.

Your Responsibility for Your System's Security

You are responsible for the security of your system. Product administration to prevent unauthorized use is your responsibility. Your system administrator should read all documents provided with this product to fully understand the features available that reduce your risk of incurring charges for unlicensed use of Genesys products.

Trademarks

Technical Support from VARs

If you have purchased support from a value-added reseller (VAR), please contact the VAR for technical support.

Customer Care from Genesys

If you have purchased support directly from Genesys, please contact Genesys Customer Care. Before contacting Customer Care, please refer to the ***Genesys Care Support Guide for On-Premises*** for complete contact information and procedures.

Ordering and Licensing Information

Complete information on ordering and licensing Genesys products can be found in the ***Genesys Licensing Guide.***

Released by

Genesys Telecommunications Laboratories, Inc. www.genesys.com

Document Version: 80g_sizing_11-2016_v8.0.001.56

Table of Contents

Preface **11**

Intended Audience 11
Scope 12
Recommendations 12
Making Comments on This Document 18
Contacting Genesys Customer Care 18

Chapter 1 **Small Contact Centers** **19**

Recommendations 20
Call Progress Detection Server 20
eServices (Multi-Channel Routing and Multimedia) 21
Genesys Agent Scripting 21
Blue Pumpkin Integration 21
Genesys Call Director Route 22
Interaction Workspace 22
User Interface Hardware 22

Chapter 2 **Medium Contact Centers** **23**

Framework, Reporting, Routing, Outbound, Voice Callback 24
Call Progress Detection Server 25
eServices (Multi-Channel Routing and Multimedia) 25
Component Distribution 26
Sample Architecture 26
Minimum Hardware Sizes 27
Interaction Processing Loads 28
Sample Log File Sizes in a Chat Solution 29
Genesys Agent Desktop 30
Genesys Desktop .NET Server 30
Genesys Agent Scripting 31
Blue Pumpkin Integration 31
Genesys Call Director Route 31
Interaction Workspace 31
User Interface Hardware 32

Chapter 3 **Hardware Architecture Diagrams.. 33**

Small Configurations 33
Microsoft Windows Platform 34
Sun Solaris Platform 35
IBM/AIX Platform 36
HP-UX Platform 37
Medium Configurations... 37
Microsoft Windows Platform 38
Sun Solaris Platform 39
IBM/AIX Platform 40
Sample Medium Configuration on an IBM/AIX Platform......................... 41
HP-UX Platform 41

Chapter 4 **Management Framework 43**

Client Connections 43
Configuration Server Memory Utilization 44
Test Results 44
Recommendations... 45
Configuration Performance Guidelines... 46
Solution Control Server Monitoring Limits 46

Chapter 5 **License Reporting Manager 8.x Solution 49**

About License Reporting Manager 8.x 49
Architecture in Release 8.x... 51
Hardware Sizing and Performance Information...................................... 51
Performance Testing Scenarios 53
Performance Test Results 54

Chapter 6 **Genesys Interaction Concentrator.. 57**

Hardware Architectures 58
Deployment Scenarios.. 58
Hardware Guidelines 62
Interaction Concentrator Performance 64
Interaction Concentrator 8.0.0 Performance 64
Hardware and Software 65
Voice Test Environment 66
Voice Tests and Results... 67
Multimedia Test Environment.. 70
Multimedia Tests and Results 71
ICON 8.0.0 Performance Conclusions.. 72
Interaction Concentrator 7.6.1 Performance 73

Voice Test Environment 73
Open Media Test Environment 77
ICON 7.6.1 Performance Conclusions 86
Interaction Concentrator 7.6.0 Performance 86
Voice-Specific Environment 86
Open Media Environment 94
ICON 7.6.0 Performance Conclusions 97
Interaction Concentrator 7.5 Performance 97
ICON 7.5 Performance Conclusions 104

Chapter 7 **Genesys Info Mart 8.x Solution 107**

About Genesys Info Mart 8.x 108
Hardware Architectures in Release 8.x 108
Genesys Info Mart 8.1.2 Performance 109
About the Genesys Info Mart 8.1.2 Performance Testing 109
Release 8.1.2 Performance Test Setup 111
Release 8.1.2 Endurance Test 128
Endurance Test Results 128
Release 8.1.2 Recovery Test—Two-Hour Outage 145
Two-Hour Recovery Test Results 146
Release 8.1.2 Recovery Test—Ten-Hour Outage 148
Ten-Hour Recovery Test Results 148
Genesys Info Mart 8.0 Performance 153
About Genesys Info Mart 8.0 Performance Testing 153
Release 8.0 Performance Testing Configuration 154
Release 8.0 Baseline Performance Test 157
Baseline Performance Test Setup 157
Baseline Performance Test Results 163
Release 8.0 Benchmark Performance 171
The Benchmark Performance Tests 172
Benchmark Performance Test Setup 172
Benchmark Performance Test Results 178
Genesys Info Mart 8.x Performance Tuning 184
Factors Affecting Release 8.x ETL Performance 184
Performance Tuning Guidelines 185
Genesys Info Mart 8.x Database Size Estimation 186

Chapter 8 **Reporting and Analytics Aggregates 189**

RAA Release 8.1 Performance 189
Release 8.1.1 Test Environment 190
Release 8.1.1 Endurance Test 192
RAA Results for the Endurance Test 193

Release 8.1.1 Interaction Lifespan Limit 199
Maximum Interaction Lifespan Test 199

Chapter 9 **Genesys Interactive Insights.. 201**

GI2 Release 8.0 Performance 201
Environment.. 202
Release 8.0 Report Performance—Baseline.. 203
General Results 203
Factors Affecting Report Performance.. 206
Release 8.0 Report Performance—Benchmark 211
General Results 212
Factors Affecting Report Performance.. 212
Release 8.0 Conclusions.. 215
GI2 Release 7.6 Performance 215
Environment.. 216
GI2 7.6 Report Performance for a Microsoft SQL Server Info Mart....... 219
Report Performance for a Microsoft SQL Server Info Mart............... 221
GI2 7.6 Report Performance for an Oracle Info Mart 223
Results for a Small-Sized Oracle Info Mart...................................... 224
Results for a Medium-Sized Oracle Info Mart 226
Release 7.6 Conclusions.. 229

Chapter 10 **Genesys Workforce Management.. 231**

Software Co-location Recommendations 231
Small Contact Centers.. 232
Medium Contact Centers.. 234
Dedicated Report Server 235
Client Workstation Requirements 235

Chapter 11 **Genesys Voice Platform 8.1 .... 237**

Capacity and Traffic Terminology 238
Capacity Metrics and Formulas 238
Performance Metrics and Formulas.. 239
Recommended Hardware and Operating Systems 242
Reporting Server.. 244
Traffic and Capacity Testing 245
VoIP Capacity Test Summaries.. 245
Component Capacity Test Cases.. 257
Performance Planning and Scalability... 293
Performance Test Cases 293
Application Test Cases 293

Component Test Cases 319
Single Server Test Cases 334
Multiple MCP Instances and Virtual Machines Test Cases 336
Performance and Scalability Comparisons 346
Hardware and Bandwidth Usage 349
Reporting Server Hardware Usage 349
Bandwidth Usage 358

Chapter 12 **Genesys SIP Server Solution 363**

SIP Server 8.1.1 and SIP Proxy 8.1.1: Sizing and Performance 363
SIP Server 8.1.0: Sizing and Performance 365
Solution Capacities 367
Reference Architecture 367
Scenarios and Benchmarks 370
CPU Usage Estimation 371
Memory Usage 378
Network Traffic Estimation 378
Example 379
SIP Server 8.1.0: Reference Call Flows 382
SIP Inbound Call with Treatment 382
SIP Inbound Call with Single-Step Transfer 383
SIP Inbound Call with Two-Step Transfer 384
SIP Inbound Call with Recording 385
SIP Inbound Call with Single-Step Conference 386
SIP Inbound Call with Two-Step Conference 387
SIP Inbound Call and Consultation Call 388
Internal Call 389
SIP Server 8.0: Sizing and Performance 389
Solution Capacities 391
Reference Architecture 391
Scenarios and Benchmarks 392
CPU Usage Estimation 393
Memory Usage 397
Network Traffic Estimation 397
Examples 398
SIP Server 8.0: Reference Call Flows 402
SIP Inbound Call 402
SIP Inbound Call with Single-Step Transfer 403
SIP Inbound Call with Recording 404
SIP Inbound Call with Single-Step Conference 405
SIP Inbound Call and Consultation Call 406
SIP Inbound Call with Treatment 407
Internal Call 408

Genesys Media Server Sizing with SIP Server 408
Genesys SIP Voicemail 8.1 416

Chapter 13 **Performance Management Advisors 417**

Overview 417
Performance Considerations... 418
General Guidelines for Contact Center Sizing Categories 420
  Example Configurations for Contact Centers based on Size 421
Stat Server Sizing.. 424
Advisors Genesys Adapter Performance Information 424
  Determining Message Rates 426
  Estimating the Number of Requested Statistics for Frontline Advisor 427
Apache Tuning Tips... 429
  XML Generator Compression Options for Apache 429
Desktop 430
Capacity, Measurement, and Sample Architecture 430
  Performance Measurement Environment....................................... 431
  Load-Carrying Capacity 431
  Improving Supervisor Desktop Service Performance...................... 457

Chapter 14 **Workspace Desktop Edition.. 459**

Agent Workstation Requirements... 459
ClickOnce Requirements.. 460
(Virtualized) Server Requirements 460

Chapter 15 **Unified Communications (UC) Connector 461**

Server Requirements 461
Desktop Requirements... 462

Chapter 16 **Genesys Administrator... 463**

Genesys Administrator... 463
Genesys Administrator Extension 463
Network Connections 464
  Genesys Administrator 464
  Genesys Administrator Extension 465
Hardware Recommendations... 466
  Genesys Administrator 466
  Genesys Administrator Extension 466

Chapter 17 **Genesys Orchestration Platform 467**

Genesys Orchestration Platform Overview 467
Genesys Orchestration Platform Reference Architecture...................... 468
Software Versions Utilized in Reference Architecture....................... 469
Call Model Scenarios and Benchmarks.. 470
Voice Only Call Model Scenario 470
Two ORS Applications in a Cluster on Windows 474
Five ORS Applications in a Cluster on Windows 475
Two ORS Applications in a Cluster on Linux 477
Five ORS Applications in a Cluster on Linux 479
Five ORS Nodes Cluster: Single WEB Server on Windows 481
Performance Test Result Conclusions.. 483
Hardware Specifications 485

Chapter 18 **intelligent Workload Distribution 487**

Methodology 487
Sizing Calculations 489
Estimating Database File Size 490
Calculation Formulae 490
Estimating the Effective Task Rate 491
Database Server Hardware Layout Selection 492
Estimation of CPU Utilization 493
Memory Usage 494
System Sizing Example 495

Chapter 19 **Genesys Co-browse 499**

Hardware 499
Node Calculation 499
Disk Space Usage Calculation 500
Sizing Calculation 501

Chapter 20 **Generated System Traffic 503**

Traffic Generated by Components and Solutions 503
Traffic Generated by Framework 7.x Components 503
Traffic Generated by Inbound Solutions 506

Chapter 21 **Database Sizing Estimation.. 511**

Database Sizing Estimator Spreadsheets 512
Reference Information 513

Chapter 22 **Genesys WebRTC Service** 515

Chapter 23 **Genesys Web Engagement** 517

Appendix A **Performance Improvements** 519

Configuration Server Performance Improvements 519
Logging Improvements 519
Log Test Description 520
Solution Performance Test 520
Stat Server Performance Improvements 7.0 vs. 6.5 522
Skill-Based Routing Performance Enhancement 7.0 523

Appendix B **Log Database 7.0 Sizing Examples** 525

Examples 525
Case 1 526
Case 2 526
Case 3 527
Log Length Dependence for 7.0 528

Supplements **Related Documentation Resources** 531

Document Conventions 533

Index 535

Preface

Welcome to the *Genesys Hardware Sizing Guide*. This document provides you with system-level information about Genesys hardware sizing guidelines for

This Preface contains the following sections:

- Intended Audience, page 11
- Scope, page 12
- Recommendations, page 12
- Making Comments on This Document, page 18
- Contacting Genesys Customer Care, page 18

For information about related resources and about the conventions that are used in this document, see the supplementary material starting on page 531.

Note: This guide reflects the products currently available for purchase from Genesys. See Table 1 on page 13.

Intended Audience

The *Genesys Hardware Sizing Guide* is primarily intended for those who make hardware and network LAN/WAN bandwidth purchasing recommendations. It assumes that you have a basic understanding of:

- Computer-telephony integration (CTI) concepts, processes, terminology, and applications.
- Network design and operation.
- Your own network configurations.

You should also be familiar with Genesys Framework architecture and functions.

Note: The Database Administrator must participate in all database sizing and configuration decisions.

Scope

This system-level guide has been prepared by Genesys. It is intended as a pre-sales tool for estimating the hardware required for installations.

The suggestions are based on field experience and are conservative estimates: a slight over-estimation of hardware requirements does not generate the production issues represented by under-estimating these requirements.

While the information in this guide can assist you in choosing hardware appropriate for your implementation, it is not intended to provide detailed information for every possible solution. These guidelines assume an “average” customer configuration of Genesys software deployed at a single site.

Unusual circumstances—such as a configuration with very complex routing strategies, multiple database lookups, or one distributed across multiple geographic locations—may require additional hardware or processors.

Be sure to consider the worst-case or peak scenarios when determining the appropriate configuration for your situation.

Once you have determined the hardware configuration that will accommodate your business needs, we recommend that you verify your proposed hardware solutions with Genesys Professional Services.

Recommendations

This guide shows the average complexity and call-flow scenario for small and medium contact centers, group-based routing, and queue routing. The recommendations are organized by contact center size, operating system, and Genesys solution.

Since hardware models change frequently, you will need to verify that the recommended models coincide with the current information from the hardware vendor.

This document covers four of the more common platforms and operating systems: IBM/AIX, Sun/Solaris, Microsoft Windows, and HP-UX. Other platforms may also be supported. Refer to the *Genesys Supported Operating Environment Reference Guide* at http://docs.genesys.com/.

The solution types, such as Framework and Enterprise Routing, have been factored out as much as possible so that you can assemble the elements for your specific needs.

Installation

It is necessary to use a **DVD ROM** when installing Genesys software.

Sizing Parameters

The parameters that have to be taken into consideration for hardware sizing are the following:

- Number of Agents
- Calls Per Hour
- Service Level
- Average Waiting Time
- Service Time

Note: Parameters cannot be changed independently.
For specific solutions and releases, particular parameters may have a greater impact on sizing. For example, Reporting statistics filters in 6.x have a high performance impact but not in 7.0.

Guide to Recommendations

Table 1 is a guide to recommendations and contains the following information:

- Contact center sizes.
- Corresponding number of agents.
- Maximum number of interactions the contact center should be receiving.
- Operating systems referred to in this document.

Note: Some Genesys solutions use different numbers to determine contact center size categories. In these cases, the different sizing criteria are provided in the applicable chapters in this document.

Table 1: Guide to Recommendations

Contact Center Size	Number of Agents	Average Interactions Per Second
Small Contact Center	<150	1.5
Medium Contact Center	150 to 1000	1.5-10
Large Contact Center	>1000, multi-site, distributed environment	

Table 1: Guide to Recommendations (Continued)

Contact Center Size	Number of Agents	Average Interactions Per Second
Operating Systems		
Sun Solaris		
Microsoft Windows		
IBM/AIX		
HP-UX		

Notes:

- The **Number of Agents** varies when using multi-channel routing. See MCR section of this document for specific MCR recommendations.
- The **Average Interactions/Second** numbers represent the average number of interactions per second that can be processed using reasonably complex routing strategies. With simpler routing strategies, the number of interactions per second that can be processed may be higher.
- In terms of **Contact Center Size**, this version of the guide is only intended to cover recommendations for small and medium installations. Large configurations may require special architectural design and should be done only in consultation with Genesys Professional Services.

This guide reflects products currently available for purchase from Genesys 6.5, 7.x, and 8.x products, with the exception of Genesys Express, which is 4.x.

Recommended Platform Configurations

This guide introduces standard hardware configurations to assign to the majority of Genesys products. However, sometimes you might need to modify various parameters, such as RAM, HDD, or network cards. These modifications are specified in parenthesis in the tables in this guide, to substitute for the original values in the standard configurations.

Recommended Server Platform Configurations

Genesys recommends using the following hardware configurations for server platforms.

Table 2: Server Platform Configurations

HP-UX Operating System: GEN_HP_SERVER	
OS	HP-UX 11i
Processor Type, Quantity, Speed	PA type processor: HP9000 rp 4410, for example
Memory Size	4 GB RAM
Hard Disk Space	40 GB per server
Ports	Networking Ports: Two TP Ethernet 1000/100/10 BASE-T cards that work in full duplex mode
IBM AIX Operating System: GEN_IBM_SERVER	
OS	IBM AIX 5.3
Processor Type, Quantity, Speed	4 CPU P5 510 type server or similar
Memory Size	4 GB RAM
Hard Disk Space	40 GB per server
Ports	Networking Ports: 2 TP Ethernet 1000/100/10 BASE-T cards that work in full duplex mode

Table 2: Server Platform Configurations (Continued)

Linux Operating System: GEN_LINUX_SERVER	
OS	Red Hat Linux Enterprise Server 5, 6, or 7
Processor Type, Quantity, Speed	4 Core CPU 2.6 GHz Intel Xeon or similar
Memory Size	4 GB RAM
Hard Disk Space	40 GB per server
Ports	Networking Ports: 2 TP Ethernet 1000/100/10 BASE-T cards that work in full duplex mode
Microsoft Windows Operating System: GEN_WIN_SERVER	
OS	Microsoft Windows 2008 or better
Processor Type, Quantity, Speed	4 Core CPU 2.6 GHz Intel Xeon or similar; in some cases an Intel processor with a total of two cores can be used in relatively small contact centers. Genesys recommends using server-class machines for all server applications.
Memory Size	4 GB RAM
Hard Disk Space	80 GB per server
Ports	Networking Ports: 2 Ethernet 1000/100/10 BASE-T ports
Sun Solaris Operating System: GEN_SUN_SERVER	
OS	Solaris 2.6-2.10, 32- or 64-bit
Processor Type, Quantity, Speed	4 CPU 1.5 GHz UltraSPARC IIIi or similar with total 4 CPU cores
Memory Size	8 GB RAM
Hard Disk Space	146 GB per server
Ports	Networking Ports: 2 Ethernet 1000/100/10 BASE-T ports

Recommended Desktop Platform Configurations

Genesys recommends using the following hardware configurations for desktop platforms.

Table 3: Desktop Platform Configurations

Microsoft Windows Operating System: GEN_WIN_DESKTOP	
OS	Microsoft Windows XP/Vista
Processor Type, Quantity, Speed	Intel Core 2 Duo CPU 2.6 GHz or similar with at least 2 cores
Memory Size	2 GB
Hard Disk Space	40 GB per server
Ports	Networking Ports: TP Ethernet 1000/100/10 BASE-T card that works in full duplex mode
Mac Operating System: GEN_MAC_DESKTOP	
OS	Mac OS X
Processor Type, Quantity, Speed	Intel Core 2 Duo CPU 2.6 GHz or similar with at least 2 cores
Memory Size	2 GB
Hard Disk Space	40 GB per server
Ports	Networking Ports - TP Ethernet 1000/100/10 BASE-T card that works in full duplex mode

Multi-core/Processors

Genesys software applications are tested against Operating Systems and Databases only, and not specific hardware. Genesys does, however, indicate minimum sizing requirements, such as processor speed and RAM with regards to the various Operating Systems platforms it works on.

The majority of Genesys software has been designed as single-threaded applications and, as such, will run on a variety of hardware, provided the appropriate Operating System is installed and the minimum sizing requirements are met. Genesys is aware that there are new multi-core/processor platform architectures currently being designed and sold that are specifically created to handle new advanced multi-threaded software, such as Java web applications.

Although Genesys products will operate as designed on these new hardware platforms, it is important that you contact your hardware vendor and discuss the levels of performance that will be expected from single-threaded applications on their products. This question not only needs to be asked for

Genesys single-threaded applications, but also for any other single-threaded software product you may have in your enterprise environment that is being considered to reside/run on these new architectures.

Genesys is aware of the new trends in hardware and has been monitoring the marketplace closely. Genesys has begun the process, as a result of this monitoring, of planning the required steps necessary to accommodate these changes in the marketplace.

Most hardware vendors who have these new multi-core/processor platform architectures do provide accurate information on their products about performance of various applications.

Making Comments on This Document

If you especially like or dislike anything about this document, feel free to e-mail your comments to Techpubs.webadmin@genesys.com.

You can comment on what you regard as specific errors or omissions, and on the accuracy, organization, subject matter, or completeness of this document. Please limit your comments to the scope of this document only and to the way in which the information is presented. Contact your Genesys Account Representative or Genesys Customer Care if you have suggestions about the product itself.

When you send us comments, you grant Genesys a nonexclusive right to use or distribute your comments in any way it believes appropriate, without incurring any obligation to you.

Contacting Genesys Customer Care

If you have purchased support directly from Genesys, please contact Genesys Customer Care.

Before contacting Customer Care, please refer to the *Genesys Care Support Guide for On-Premises* for complete contact information and procedures.

Chapter

1 Small Contact Centers

This chapter presents hardware guidelines and recommendations for small contact centers. Small contact centers generally have less than 150 agents and receive a maximum of 1.5 interactions per second (IPS).

The information in this chapter contains the following topics:

- Recommendations, page 20
- Call Progress Detection Server, page 20
- eServices (Multi-Channel Routing and Multimedia), page 21
- Genesys Agent Scripting, page 21
- Genesys Agent Scripting, page 21
- Blue Pumpkin Integration, page 21
- Genesys Call Director Route, page 22
- Interaction Workspace, page 22
- User Interface Hardware, page 22

Note: Hardware architecture diagrams of sample configurations are in Chapter 3, "Hardware Architecture Diagrams" on page 33

Recommendations

This section describes recommendations for small contact centers using Framework, Reporting, Routing, Outbound, Voice Callback, and others.

Table 4 on page 20 shows the recommendations for small contact centers using Framework Management and Configuration Layers, T-Server, Historical Reporting, Enterprise Routing and/or Outbound Contact and Voice Callback on HP-UX, IBM AIX, Linux, MS Windows, and /or Sun Solaris operating systems. All the solutions, including Stat Server and Reporting, may be located on one box or distributed as described in the following table.

Note: In a single Solution Control Server (SCS) environment, you might need to limit the number of hosts controlled by that single Solution Control Server. Refer to the “Solution Control Server Monitoring Limits” on page 46 for more information.

Table 4: Framework, Reporting, Routing, Outbound, and Voice Callback

Operating System	Recommendations
HP-UX	GEN_HP_SERVER
IBM AIX	GEN_IBM_SERVER
Linux	GEN_LINUX_SERVER
MS Windows	GEN_WIN_SERVER
Sun Solaris	GEN_SUN_SERVER

Call Progress Detection Server

The recommendation for Call Progress Detection Server is GEN_WIN_SERVER.

For the CPD Board, see the Supported Dialogic Boards Table in the *Genesys Supported Media Interfaces Reference Manual*.

In most cases, 2-core CPU and one Ethernet card are sufficient for small contact centers.

Note: Hardware configuration for systems where Call Progress Detection Server is used with Intel HMP, software should comply with the requirements specified in Intel HMP documentation.

eServices (Multi-Channel Routing and Multimedia)

The recommendation for eServices (Multi-Channel Routing/Multimedia) is GEN_WIN_SERVER.

Hard disk consumption depends on the size of the database (the amount of stored contact and interaction history), and also on the accumulated size of the log files, which you can regulate using options in Configuration Manager.

Notes:

- The name of Multimedia is changed to eServices beginning with release 8.0.1.
- For purposes of eServices (Multi-Channel Routing/Multimedia) only, a small contact center is defined as one with 75 or fewer agents.

Genesys Agent Scripting

The recommendation for Genesys Agent Scripting is GEN_WIN_SERVER. In most cases, two-core CPU, RAM 2 GB, and one Ethernet card are sufficient. Recommended for the Server: IIS 5.0, IIS 6.0, IIS 7, IIS 7.5 Apache/Tomcat 4.3.1, Apache/Tomcat 5.0, or Apache/Tomcat 6.0 Web Server running on Microsoft Windows 2003 server or Microsoft Windows 2008 server.

Notes:

- Genesys Agent Scripting is installed in standalone mode (one Tomcat and Apache or IIS web server).
- The sizing information is for running scripts without integration with other products. If scripts are used or integrated with other products, then the sizing requirements for those products must be considered, in addition to what is defined above.

Blue Pumpkin Integration

The recommendation for Blue Pumpkin Integration is GEN_WIN_SERVER with two-core 1 GHz CPU; RAM 2 GB.

Client Workstation Requirements

Computers running the Statistics Configuration Utility, at minimum, should meet these specifications:

- Pentium III, 1 GHz or greater, with 1 GB RAM

Genesys Call Director Route

The recommendation for Blue Pumpkin Integration is GEN_WIN_SERVER with two-core 1 GHz CPU; RAM 2 GB; one Ethernet card; and Tomcat 4.3.1 or 5.0 Web Server.

Interaction Workspace

For information about Interaction Workspace, refer to Chapter 16, "Workspace Desktop Edition," on page 459.

User Interface Hardware

User interfaces such as Configuration Management Environment (CME), Solution Control Interface (SCI), and CCPulse+, should use GEN_WIN_DESKTOP computer.

Note: RAM 4GB is preferable when non-Genesys applications are being run, or if it is necessary to reduce a screen-pop delay.

Genesys Agent and Genesys Supervisor Desktops also support Mac OS X workstations GEN_MAC_DESKTOP.

Chapter

2 Medium Contact Centers

This chapter presents hardware guidelines and recommendations for medium contact centers. Medium contact centers are generally those that have from 150 to 1000 agents and receive a maximum of 12 interactions per second (IPS).

The information in this chapter contains the following topics:

- Framework, Reporting, Routing, Outbound, Voice Callback, page 24
- Call Progress Detection Server, page 25
- eServices (Multi-Channel Routing and Multimedia), page 25
- Genesys Agent Desktop, page 30
- Genesys Desktop .NET Server, page 30
- Genesys Agent Scripting, page 31
- Blue Pumpkin Integration, page 31
- Genesys Call Director Route, page 31
- Interaction Workspace, page 31
- User Interface Hardware, page 32

Hardware architecture diagrams of sample configurations are in Chapter 3, "Hardware Architecture Diagrams" on page 33.

Note: You have to run License Server on one of the servers. This server requires a good network connection to other servers on which Genesys software is installed.

Framework, Reporting, Routing, Outbound, Voice Callback

Table 5 shows the relationships between small and medium contact center hardware recommendations. Hardware for medium contact centers is specified as an upgrade of the small contact center hardware.

Table 5: Framework, Reporting, Routing, Outbound, and Voice Callback Hardware Upgrade

Operating System	Small Contact Center	Medium Contact Center
HP UX	GEN_HP_SERVER	GEN_HP_SERVER2 = GEN_HP_SERVER (4 CPU PA8700+ or similar, HDD 80 GB)
IBM AIX	GEN_IBM_SERVER	GEN_IBM_SERVER2 = GEN_IBM_SERVER (4 CPU 1 GHz POWER 4, HDD 80 GB)
Linux	GEN_LINUX_SERVER	GEN_LINUX_SERVER2 = GEN_LINUX_SERVER (HDD 80 GB)
MS Windows	GEN_WIN_SERVER	GEN_WIN_SERVER2 = GEN_WIN_SERVER (HDD 80 GB)
Sun Solaris	GEN_SUN_SERVER	GEN_SUN_SERVER2 = GEN_SUN_SERVER)

- A single Solution Control Server should be assigned to a maximum of 15 hosts, to ensure that Solution Control Server can react within 20 seconds to changes in the status of any number of hosts. If more than 15 hosts are used in the configuration, Genesys recommends that you add additional Solution Control Servers in a Distributed Solution Control Server configuration to limit the load on each server. If a Solution Control Server is set up to control more than 15 hosts, the time to respond to changes in host status may increase and exceed 20 seconds, especially in those cases where multiple hosts fail simultaneously.
- Stat Server may need to be located on a separate box depending on interaction volume. It is recommended that a separate Stat Server should be installed if Genesys Routing is deployed.
- The Database Server may be installed on the same box as the database; however accurate database sizing information is essential. The number of HDDs and/or size of HDDs may need to be increased depending on data storage requirements and workflow.

- You should have the DVD-ROM for eServices/MultiMedia (T-Server's) installation.

Note: Important: The Database Administrator needs to participate in all database sizing decisions.

Call Progress Detection Server

Recommendation for Call Progress Detection Server is the following: GEN_WIN_SERVER. For the CPD Board, see the Supported Dialogic Boards Table in the *Genesys Supported Media Interfaces Reference Manual*; 2-core CPU; and Networking card: 10/100 TX PCI UTP Microsoft Windows .

Note: Hardware configuration for systems where Call Progress Detection Server is used with Intel HMP software should comply with the requirements specified in Intel HMP documentation.

eServices (Multi-Channel Routing and Multimedia)

The recommendations for eServices (Multi-Channel Routing/Multimedia) are:

- GEN_WIN_SERVER (6 core CPU; HDD 160 GB).
- Database: GEN_WIN_SERVER.
- Hard Disk Drive: Depends upon the contact center’s policy regarding the preservation of log files.

Notes:

- The name of Multi-Channel Routing is changed to Multimedia beginning with release 7.2.
- The name of Multimedia is changed to eServices beginning with release 8.0.1.
- For purposes of eServices (MCR and Multimedia) only, a medium contact center is defined as one with no more than 300 agents doing simultaneous chat processing and no more than 500 agents for e-mail processing, with an e-mail volume of no more than five interactions per second.

Component Distribution

Genesys recommends that you distribute eServices (MCR/Multimedia) and associated components among several host machines, as follows:

- **Server**: eServices (Multimedia/MCR) core components: Interaction Server (with separate installation of DB Server), Classification Server, Universal Contact Server (UCS), E-mail Server Java*, Chat Server*
- **Server**: eServices (Multimedia) web components
- **Database**: Universal Contact Server database and Interaction Server cache
- **Server**: Kana Response Live Server (part of Genesys web collaboration, requires a separate application container from Web API Server)
- **Desktop**: Interaction Routing Designer (IRD), Knowledge Manager, UCS Manager

* Based on the load and nature of contact center media (email, chat or blended) you might need to deploy these components on separate machines.

Note: The corporate mail server should also be deployed on a separate computer.

Sample Architecture

The sample architecture in Figure 1 shows all the components required for a eServices (Multimedia) solution. Some key points about these requirements that you should be aware of:

- The number of servers and distribution of components complies with the recommendations in "Component Distribution".
- Adjust the number of servers and component distribution as necessary, according to the sizing information in the tables later in this section.
- If necessary, substitute any of the Microsoft Windows Server 2003 hosts in this architecture for a UNIX host (AIX, Solaris, Linux), thus allowing for a mixed approach.

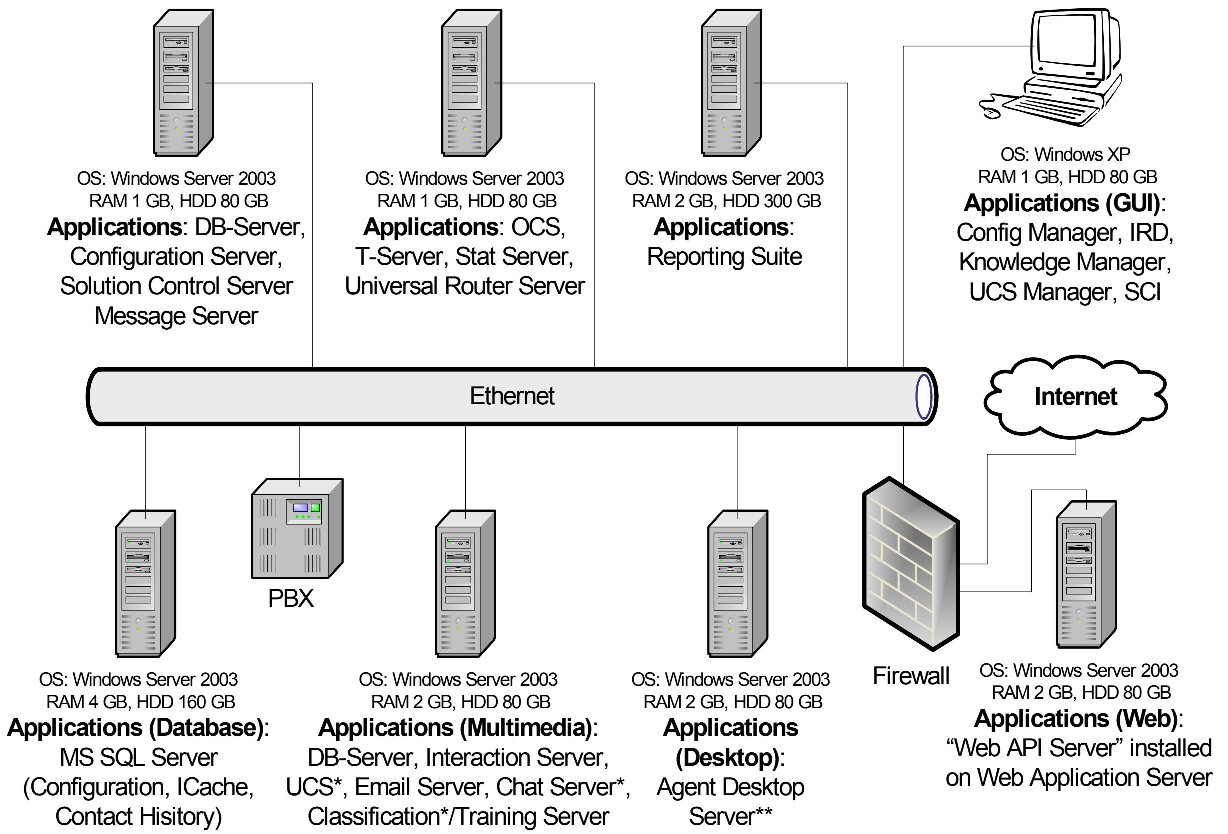


Figure 1: Sample eServices (Multimedia) Architecture

* Depending on your solution, some of these servers might need to be installed on a separate machine. For example, in an E-mail solution, both Classification Server and UCS must be installed on separate machines.

** A computer with the listed specifications can support approximately 150-200 agents. For more agents, you must distribute the load over additional machines. For further explanation, see "Genesys Agent Desktop" on page 30.

Minimum Hardware Sizes

Recommendation for eServices (Multi-Channel Routing/Multimedia:

- GEN_WIN_SERVER(6-core CPU, HDD 160 GB). A possible CPU configuration could be: distribution between three 2xCPU boxes, each with 1.5 GB RAM.
- Database: GEN_WIN_SERVER.
- Hard Disk Drive: Depends upon the contact center's policy regarding the preservation of log files.

Interaction Processing Loads

See the following tables for information about:

- "Processing Loads for E-mail Interactions"
- "Processing Loads for Chat Interactions"

The numbers in these tables are derived from hardware configurations tested in Genesys laboratories—different configurations yield different results, so consider these values as guidelines to help plan the basic layout of your deployment. For example, in a blended solution (E-mail and Chat), depending on the capability of the various systems on your network and the volume of interactions you anticipate, you might consider setting up E-mail and Chat functionality on separate machines.

Processing Loads for E-mail Interactions

Table 6 shows both CPU and memory consumption in a contact center servicing 340 agents logged into a single instance of Genesys Agent Desktop. Components were deployed on computers with at least two 2.33 GHz Xeon-powered processors.

Table 6: Processing Loads for E-mail Interactions

eServices (MCR/Multimedia) Components	Average CPU usage	Maximum CPU usage	Memory footprint
E-mail Server Java	4%	8%	less than 100 MB
Interaction Server	18%	57%	less than 24 MB
Universal Contact Server	22%	74%	less than 100 MB
Classification Server	22%	74%	less than 100 MB
Databases	**Average CPU usage**	**Maximum CPU usage**	**Memory footprint**
Interaction Server Database (ICache) on MS SQL Server	8%	11%	1.0 GB
Universal Contact Server Database on MS SQL Server	15%	50%	980 MB

Processing Loads for Chat Interactions

Table 7 shows both CPU and memory consumption in a contact center with 300 logged in agents, handling simultaneous chat sessions.

Table 7: Processing Loads for Chat Interactions

eServices (MCR/Multimedia) Components	Average CPU usage	Memory footprint
Chat Server	44%	Total less than 50 MB
Web API Server	70% Apache usage: 6%	400 MB (Maximum 640 MB) Apache usage: 47 MB
Interaction Server (and DB Server)	Total 2%	Total less than 100 MB
Universal Contact Server	1%	65 MB
Databases	**Average CPU usage**	**Memory footprint**
Interaction Server Database (ICache) on MS SQL Server	1%	1.7 GB
Universal Contact Server Database on MS SQL Server	2%	

Sample Log File Sizes in a Chat Solution

The values in Table 8 give you a sample of how much hard disk space you will need for the log files that the various eServices (Multimedia) components generate when running a Chat solution at a medium output level for 30 days. Use these numbers as guidelines only—actual size requirements vary, depending on the number and length of the chat sessions that your solution serves.

Note: Lab testing of Genesys components used a total of 300 simultaneous chat sessions, each with a duration of 5 minutes, creating approximately two and half million interactions over a one-month period.

Table 8: Log File Sizes in a Chat Solution

eServices/Multimedia Component	Average consumption per chat session	Total monthly HDD space consumption
Chat Server	30 K [a]	About 100 GB
Interaction Server	100 - 150 K [b]	260 - 390 GB

Table 8: Log File Sizes in a Chat Solution (Continued)

eServices/Multimedia Component	Average consumption per chat session	Total monthly HDD space consumption
Universal Contact Server	25 K	100 GB
Classification Server	5K	100 GB
Training Server	Negligible consumption [c]	
Knowledge Manager	This GUI application has no impact on log size.	

a. Based on a 5-minute session containing 10 messages of 1.1 K each.

b. Very much dependent on the complexity of the interaction workflow (ie. strategies)

c. Since the server is typically run only once a week or month.

Genesys Agent Desktop

The recommendation for Genesys Agent Desktop is GEN_WIN_SERVER, with HDD 80 GB; and Apache 2.2.4 Web Server. In most cases, one Ethernet card is sufficient.

- Genesys Agent Desktop Server is installed in either standalone or in load balancing mode (several Tomcats and Apache in front), as follows:
  a. Standalone mode: Genesys Agent Desktop Server can support a maximum of 400 agents with Apache 2.2.4 with one Tomcat 5.5 (Catalina) with a maximum of 1.8 EMS or 8 CPS.

  Or:

  b. Load balancing mode: Use network load balancing hardware or software (for example: Microsoft Network Load Balancing for W2K Advanced Server) for configuration of more than 400 agents. Desirable configuration is one network node (2x2.4 MHz 2 GB memory box) with one Web Server and one Genesys Agent Desktop Server per each 400 agents.

  c. Java Server Pages should already be compiled using Java Development Kit 1.4.2_xx or 1.5.0_yy from Sun.

Genesys Desktop .NET Server

Genesys Desktop .NET Server is no longer available. From release 7.2, all .NET Server capabilities have been delivered in Genesys Integration Server (GIS). See Chapter 17, “Genesys Integration Server,” page 481 for details.

Genesys Agent Scripting

The recommendation for Genesys Agent Scripting is GEN_WIN_SERVER with two-core CPU; RAM 2 GB; and one Ethernet card. For the Server: IIS 5.0, IIS 6.0, Apache/Tomcat 4.3.1, or Apache/Tomcat 5.0.Web Server running on Microsoft Windows 2003 server.

Notes:

- Genesys Agent Scripting is installed in standalone mode (one Tomcat and Apache or IIS web server.
- The sizing information is for running scripts without integration with other products. If scripts are used or integrated with other products, then the sizing requirements for those products must be considered in addition to what is defined above.

Blue Pumpkin Integration

The recommendation for Blue Pumpkin Integration is GEN_WIN_SERVER with two-core 1 GHz CPU; RAM 2 GB.

Contact Genesys Customer Care website at http://genesys.com/support for specific requirements, because HDD 40 GB is given as a baseline, only.

Client Workstation Requirements

Computers running the Statistics Configuration Utility, at minimum, should meet these specifications:

- Pentium III, 1 GHz or greater, with 1 GB RAM

Genesys Call Director Route

The recommendation for Call Director Route: GEN_WIN_SERVER with two-core 1 GHz CPU; RAM 2 GB; and one Ethernet card. For the Server: Tomcat 4.3.1 or 5.0 Web Server.’

Interaction Workspace

For information about Interaction Workspace, refer to Chapter 16, “Workspace Desktop Edition,” on page 459.

User Interface Hardware

User interfaces such as Configuration Management Environment (CME), Solution Control Interface (SCI), and CCPulse+, should use GEN_WIN_DESKTOP computer.

Note: RAM 4GB is preferable when non-Genesys applications are being run, or if it is necessary to reduce a screen-pop delay.

Genesys Agent and Genesys Supervisor Desktops also support GEN_MAC_DESKTOP computer.

Chapter

3 Hardware Architecture Diagrams

This chapter presents hardware architecture diagrams prepared by Genesys. It includes generic sizing information and diagrams for small and medium contact centers running Microsoft Windows, Sun Solaris, IBM/AIX, and HP-UX operating systems.

The information in this chapter contains the following topics:

- Small Configurations, page 33
- Medium Configurations, page 37
- Sample Medium Configuration on an IBM/AIX Platform, page 41

Note: Genesys did not size box for SQL Server engine.

Small Configurations

The diagrams in this section are for small configurations on Microsoft Microsoft Windows, Sun Solaris, and IBM/AIX platforms.

Microsoft Windows Platform

Figure 2 shows a diagram of a sample small configuration on a Microsoft Windows Platform.

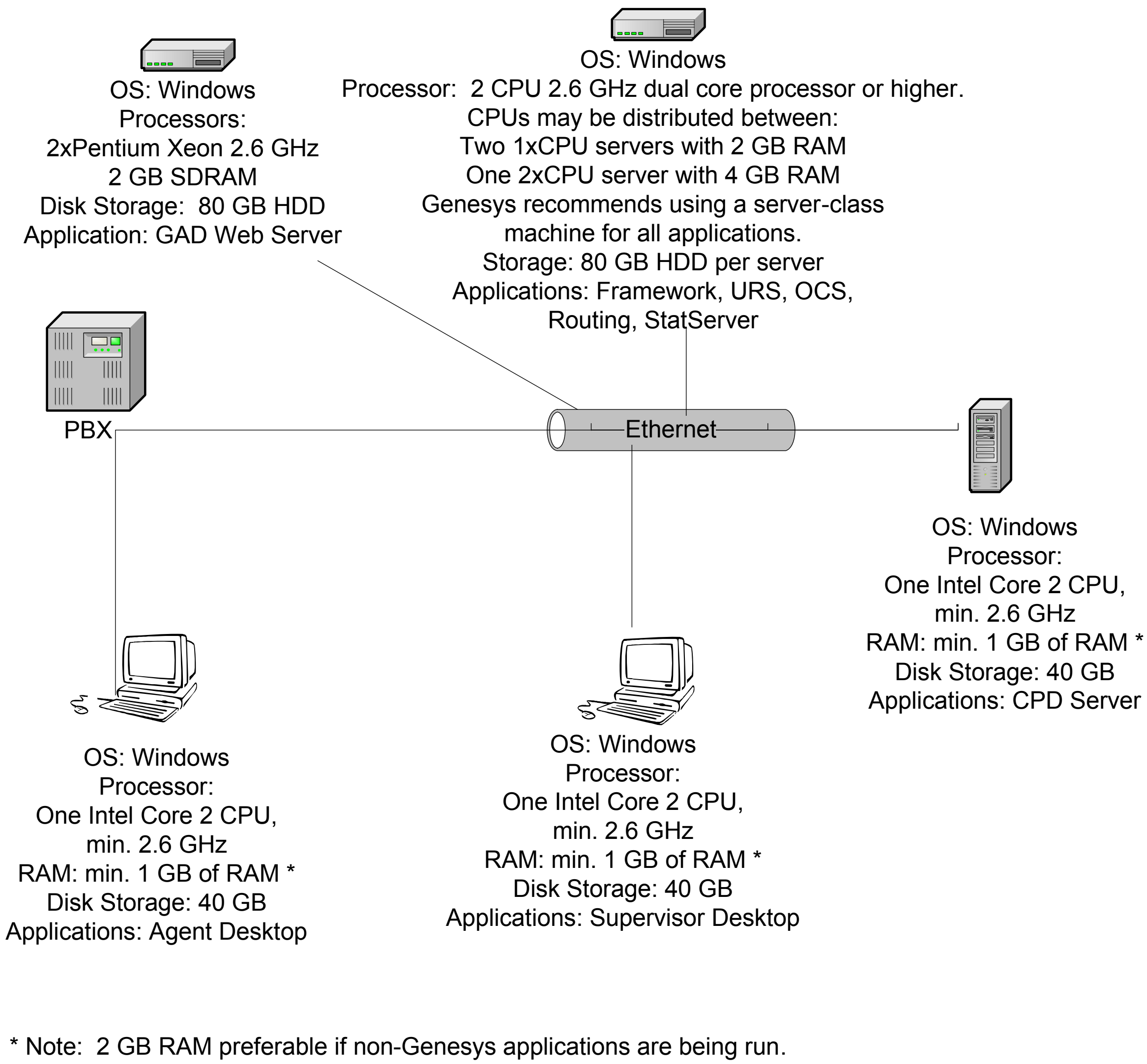


Figure 2: Sample Small Configuration on a Microsoft Windows Platform

Sun Solaris Platform

Figure 3 shows a diagram of a sample small configuration on a Sun Solaris platform.

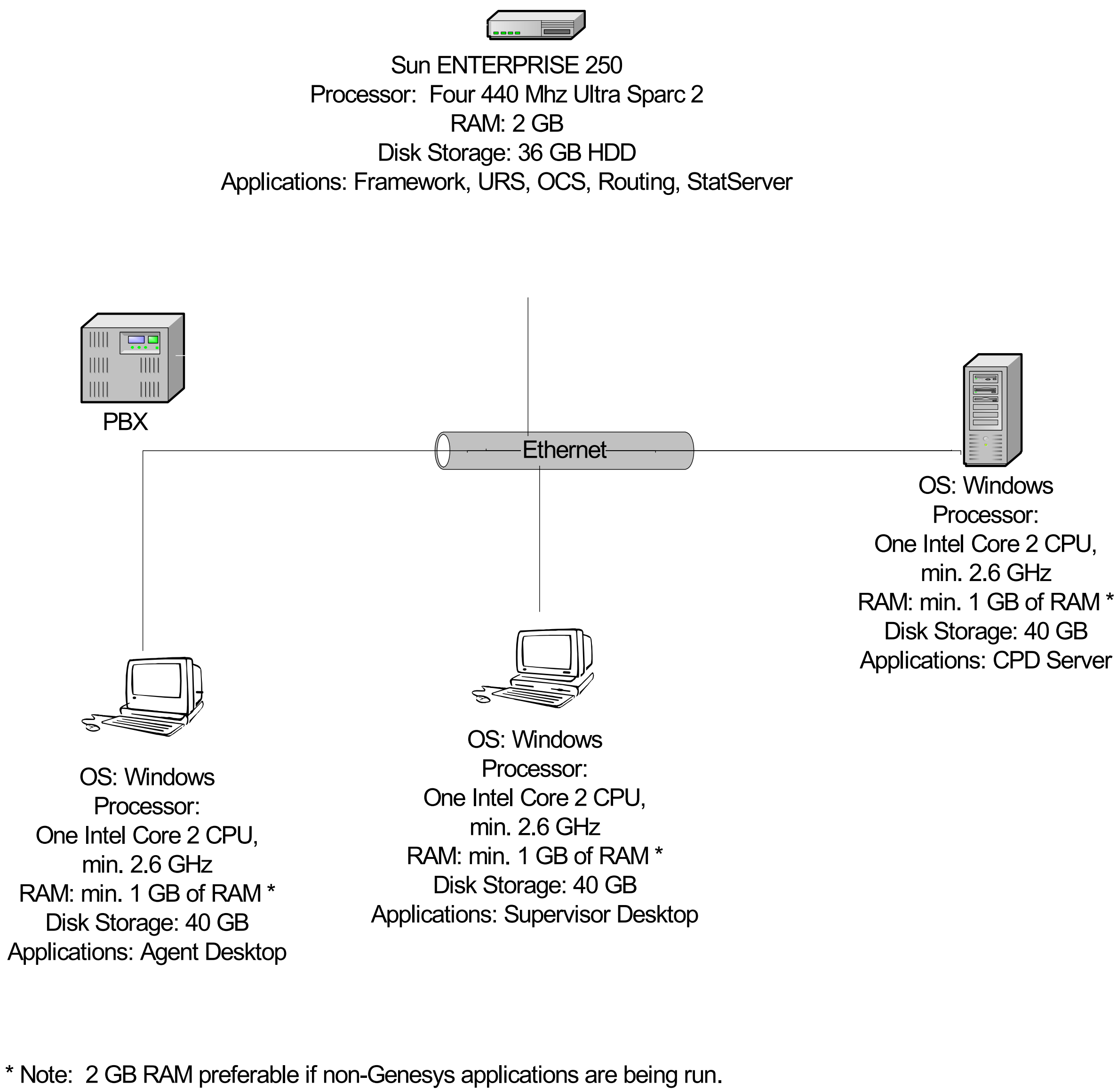


Figure 3: Sample Small Configuration on a Sun Solaris Platform

IBM/AIX Platform

Figure 4 shows a diagram of a sample small configuration on an IBM/AIX platform.

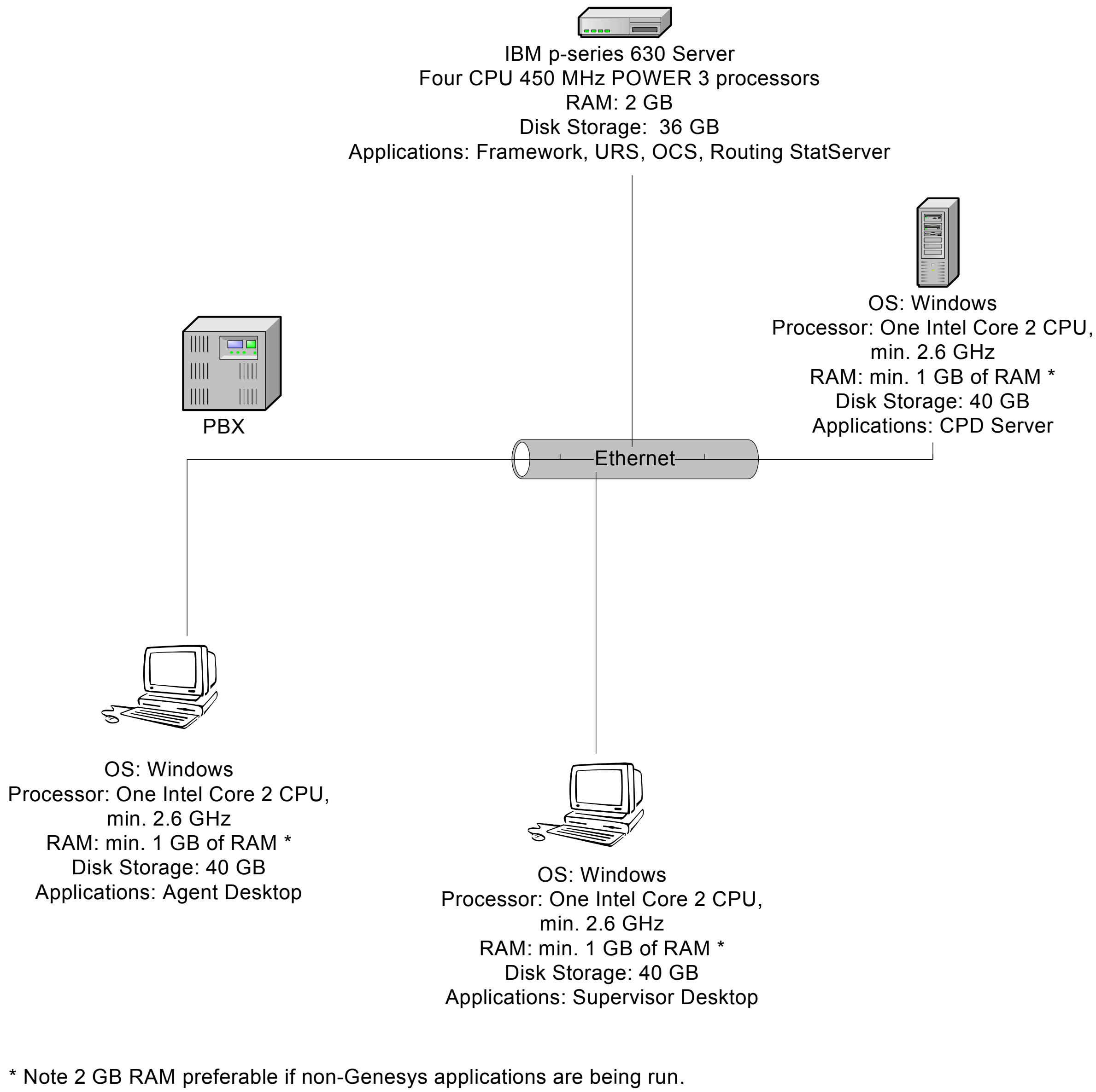


Figure 4: Sample Small Configuration on an IBM/AIX Platform

HP-UX Platform

Figure 5 shows a diagram of a sample small configuration on an HP-UX platform.

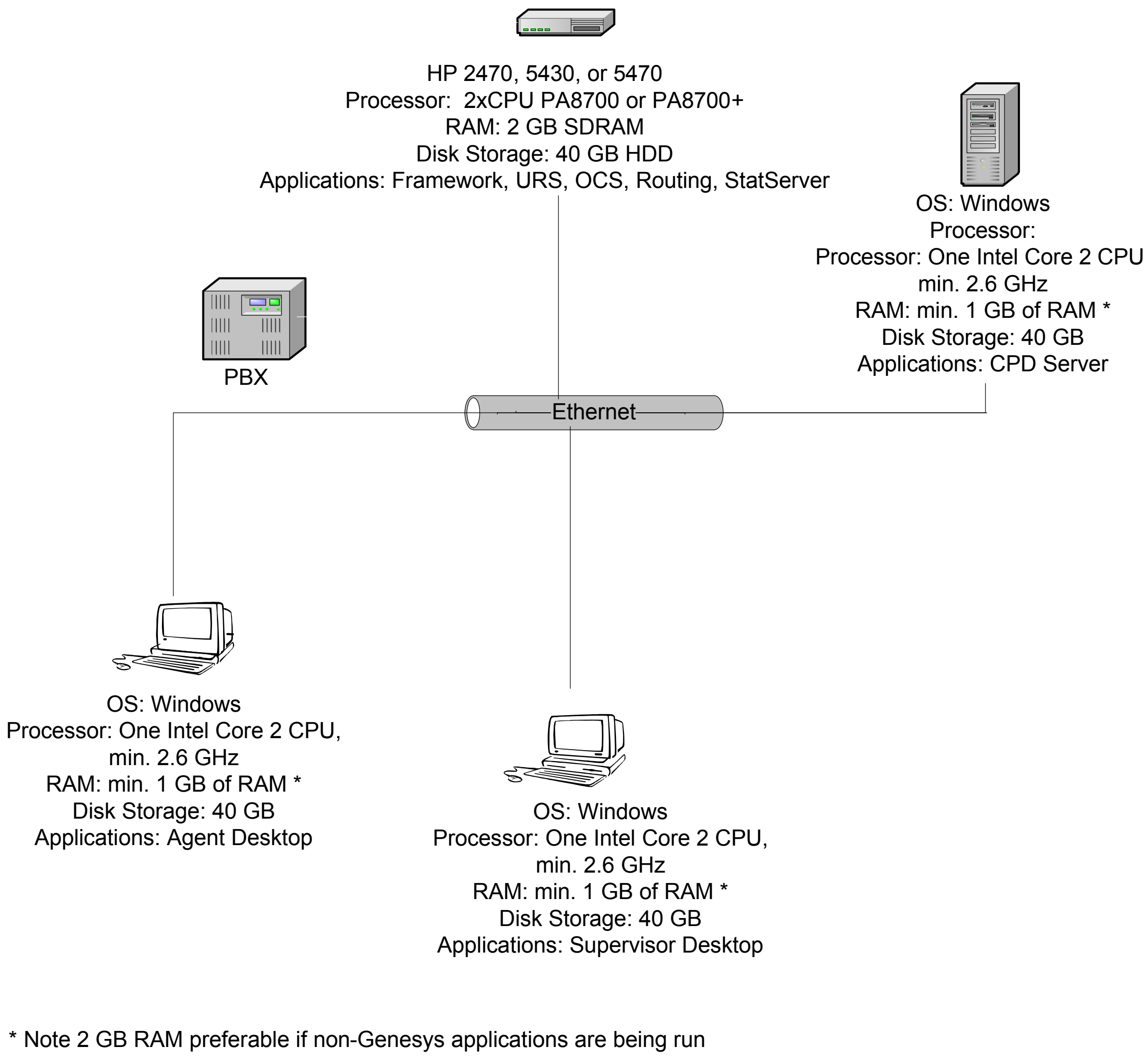


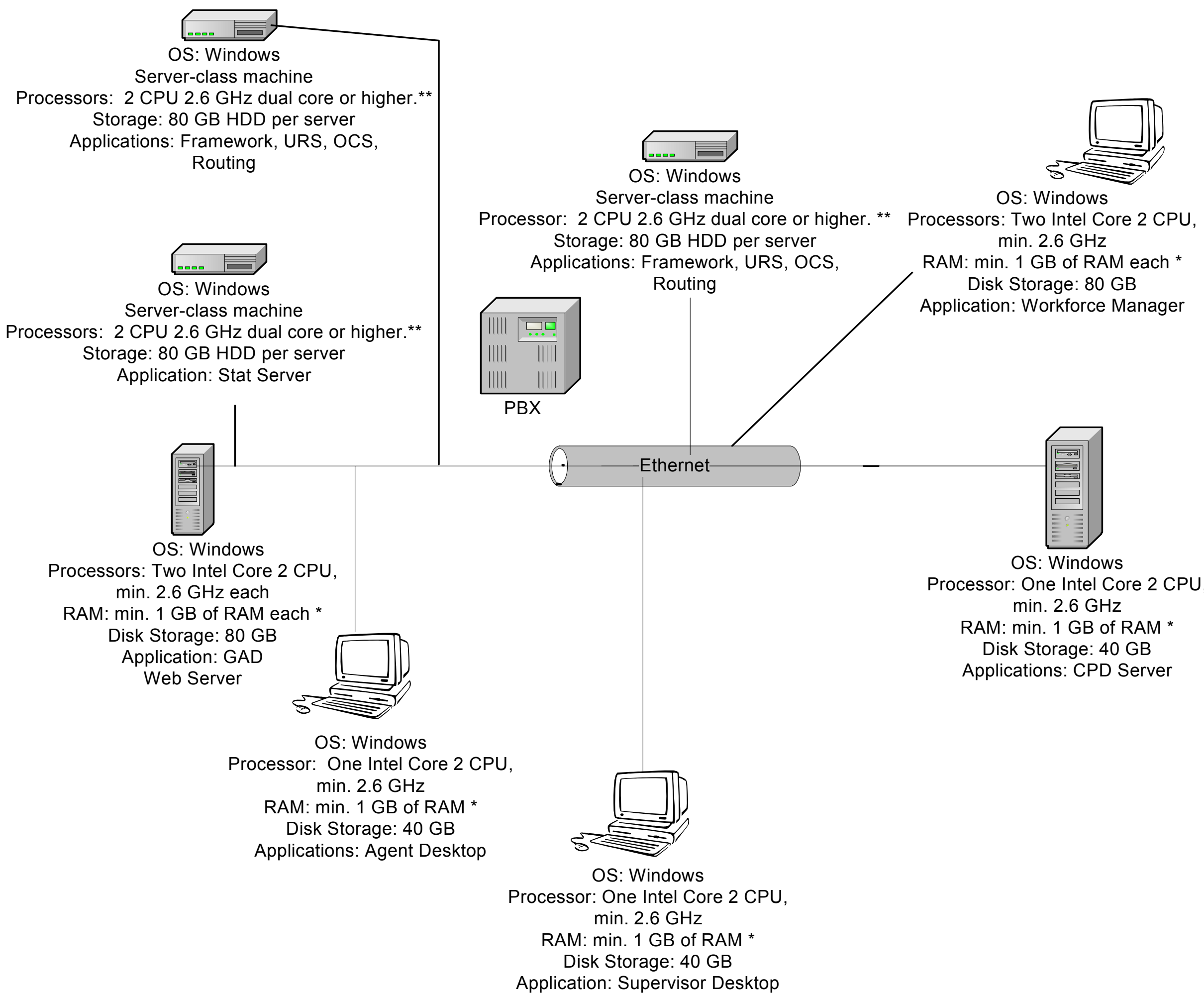
Figure 5: Sample Small Configuration on an HP-UX Platform

Medium Configurations

The diagrams in this section are for medium configurations on Microsoft Windows, Sun Solaris, and IBM/AIX platforms.

Microsoft Windows Platform

Figure 6 shows a diagram of a sample medium configuration on a Microsoft Windows platform.



* 2 GB RAM preferable if non-Genesys applications are being run.
** Possible choices for CPU configuration could be: Two 1xCPU servers with 2 GB RAM, or, One 2xCPU server with 4 GB RAM. Genesys recommends using a server-class machine for all applications.

Figure 6: Sample Medium Configuration on a Microsoft Windows Platform

Sun Solaris Platform

Figure 7 shows a diagram of a sample medium configuration on a Sun Solaris platform.

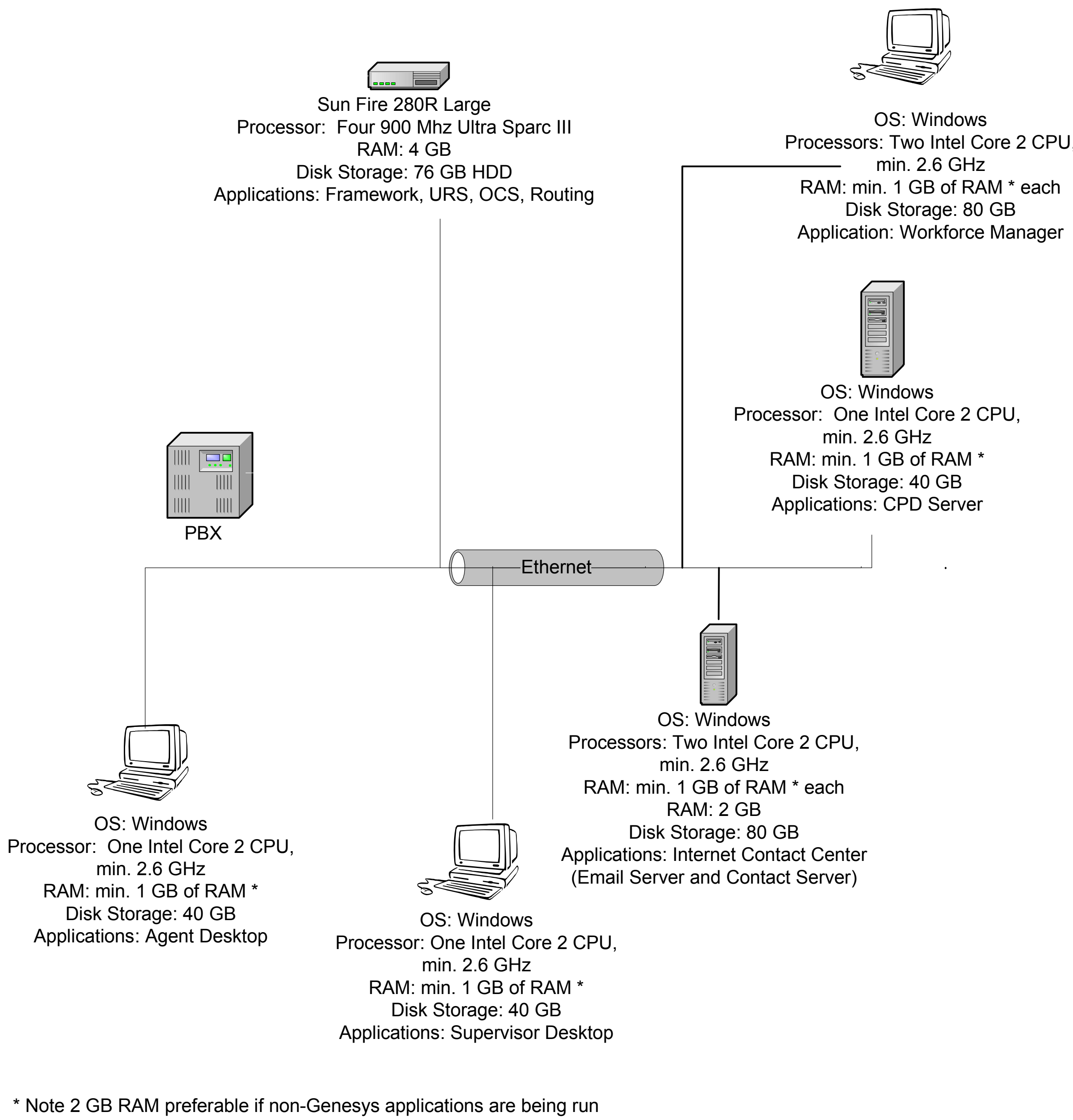


Figure 7: Sample Medium Configuration on a Sun Solaris Platform

Note: Box for DB Engine and corporate Mail Server was not counted.

IBM/AIX Platform

Figure 8 shows a diagram of a sample medium configuration on an IBM/AIX platform.

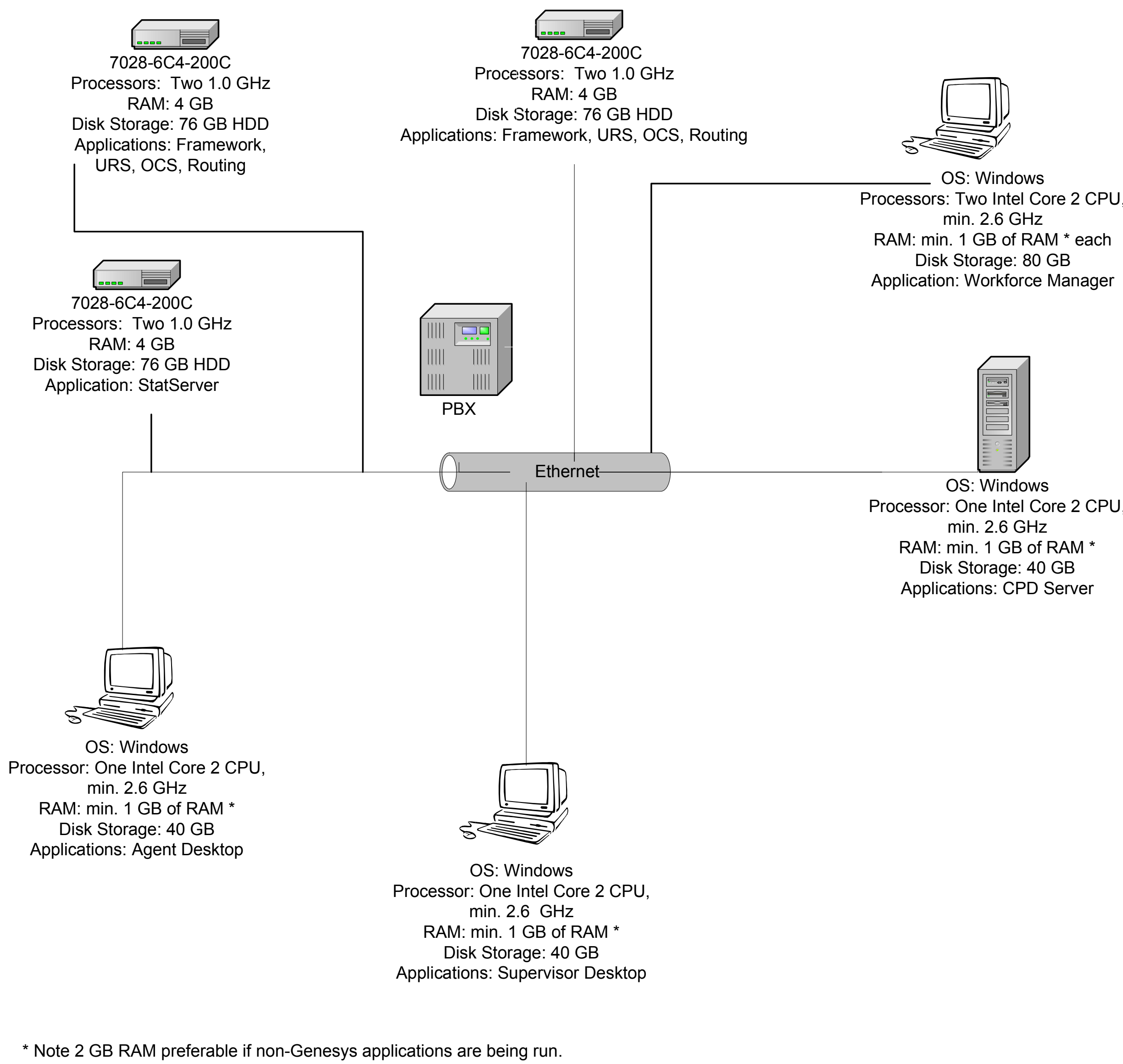


Figure 8: Sample Medium Configuration on a IBM/AIX Platform

Note: Box for DB Engine and corporate Mail Server was not counted.

Sample Medium Configuration on an IBM/AIX Platform

HP-UX Platform

Figure 9 shows a diagram of a sample medium configuration on an HP-UX platform.

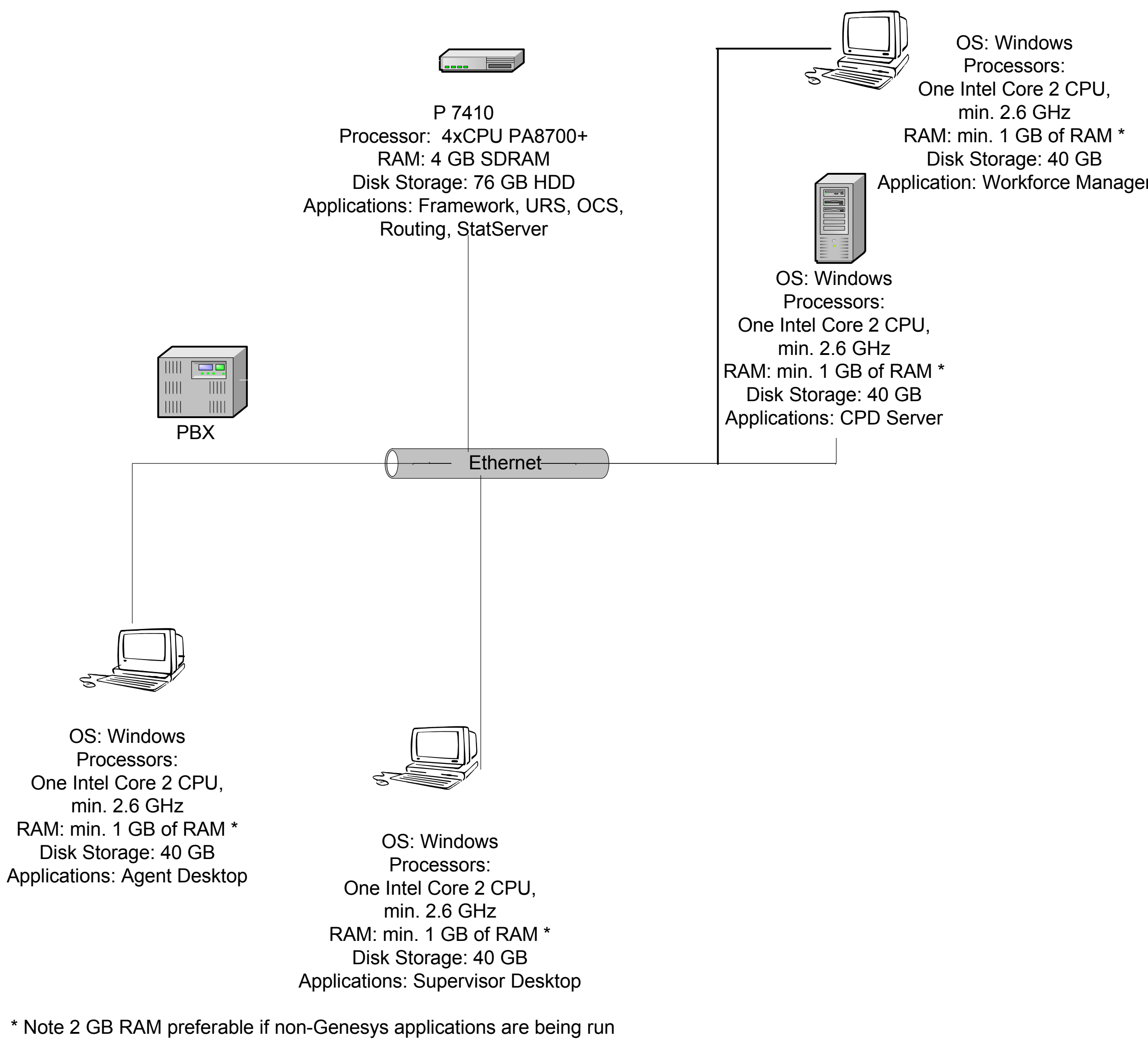


Figure 9: Sample Medium Configuration on an HP-UX Platform

Note: Box for DB Engine and corporate Mail Server was not counted.

Chapter

4 Management Framework

This chapter provides sizing information and recommendations for Management Framework components.

This chapter contains the following topics:

- Client Connections, page 43
- Configuration Server Memory Utilization, page 44
- Configuration Performance Guidelines, page 46
- Solution Control Server Monitoring Limits, page 46

Client Connections

A single instance of Configuration Server is designed to serve up to 500 simultaneous connections. If you expect or require more than 500 client applications connecting to Configuration Server, consider installing additional instances of Configuration Server Proxy to distribute the load.

Using Configuration Server Proxy with Workspace Desktop Edition

If you are using Configuration Server Proxy dedicated to Workspace Desktop Edition (WDE) client application version 8.5.113.11 (or newer), the number of connections can be increased to a maximum of 1,500 for an environment with no more than 25,000 configured agents. The capacity of a single Configuration Server Proxy increases when the Team Communicator feature of all WDE instances is configured to use the brief information format, and caching the agent directory instead of reloading it on each reconnect.

Configure the brief information format as follows:

1. In the Configuration Server Proxy application, in the [security] section, set the objbrief-api-permission-check option to true.

2. In each WDE client application, set the following configuration options as indicated in Table 9:

Table 9: WDE Client Application Configuration Options

Option	Value
general.configuration-agent-collection-loading-method	brief-info
general.configuration-object-collection-cache-timeout	A value > 72 hours
general.configuration-update-notification	ThisAgent, ThisApplication
options.record-option-locally-only	True OR False, if options.record-location set to a valid shared directory

For more information about how these configuration options affect performance, refer to Effects of Configuration Options and Privileges on Performance in the Planning Your Deployment section of the WDE Deployment Guide.

Note: If your environment has more than 25,000 configured agents, additional load tests might be required to confirm the target number of connections.

Configuration Server Memory Utilization

The information in this section was derived from tests run in the following environment:

- Configuration Server version 8.1.300.09, 64-bit executable
- Host computer with the following specifications:
  - Processor—2 x Intel Clovertown Quad Xeon X5365 3.0 GHz/L2=2x4 MB
  - RAM—20 GB
  - Drives—Disk2 x SCSI HDD x 146 GB X 15K RPM
  - Network peripheral—Dual Gigabyte Ethernet
- Configuration Database prepopulated with 700,000 objects

Test Results

Table 10 on page 45 provides the results obtained using the configuration previously described. Figure 10 depicts the relationship between memory used and the number of clients working with a specified number of configuration objects.

Table 10: Configuration Server Memory Usage

# of Clients	# of Requested Objects	Maximum Configuration Server Memory Usage (GB)
1	0	3.162
100	2,000,000	5.677
200	4,000,000	8.077
300	6,000,000	10.550
400	8,000,000	12.880
500	10,000,000	15.519

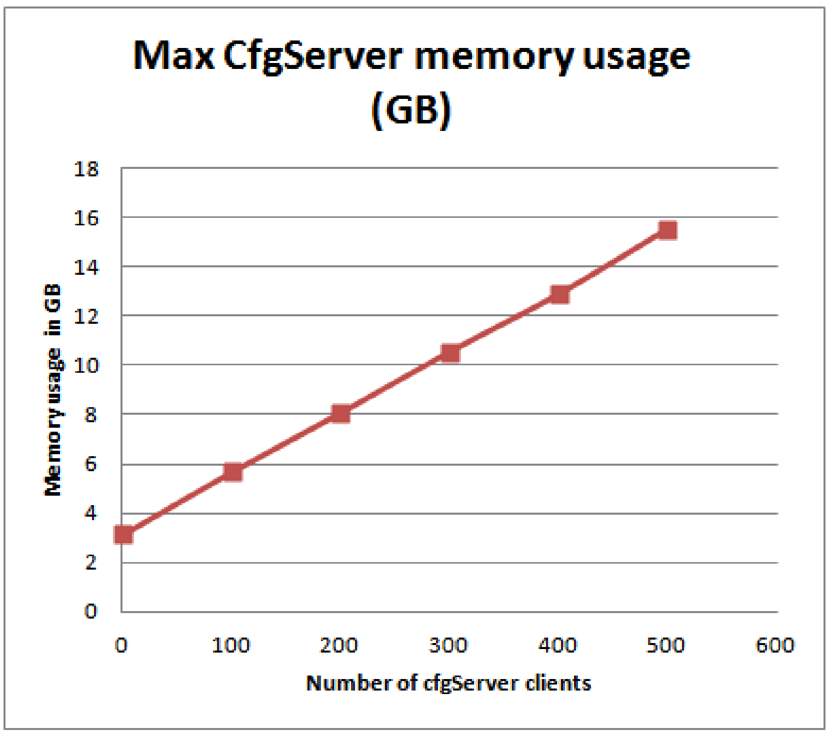


Figure 10: Maximum Configuration Server Memory Usage (GB)

Recommendations

Genesys recommends that the amount of physical RAM dedicated to Configuration Server processes be not less than the maximum expected memory usage for the selected environment (as provided in Table 10) plus 25%.

For 32-bit environments, the total amount of memory that Configuration Server can consume is 2GB; or 3GB on Windows 2003 with enabled address extensions.

Configuration Performance Guidelines

In addition to the guidelines and recommendations given previously in this chapter, Genesys strongly recommends that you consider these guidelines when operating your configuration environment:

- Consider using Folder objects when creating a large number of configuration objects. The recommended number of configuration objects per folder is up to 4,000. Anything larger significantly increases Genesys Administrator time for loading configuration objects.
- When creating configuration objects of the Script type (for example, routing strategies), keep in mind that both the number of Script objects and the script size significantly affect the time it takes Genesys Administrator to load the Script configuration objects. If you create large scripts, reduce the number of Script objects in a subfolder to achieve an acceptable loading speed. For instance, for script-type configuration objects approximately 150 KB in size, limiting the number of script-type objects to 30 per subfolder guarantees an acceptable loading speed.
- When creating a large number of configuration objects of the Agent Login type, assign them to User configuration objects as you create them. When the Configuration Database contains too many unassigned Agent Logins, Genesys Administrator takes a long time to open the Agent Login browse dialog box from the Configuration tab or the Person Properties dialog box. To guarantee an acceptable loading speed, keep the number of unassigned Agent Login objects below 1000 per Tenant object.
- For all configuration objects, do not store large amounts of data as text properties in an object's Annex, unless it is explicitly required by Genesys applications.
- Use Genesys Administrator and other Configuration Server clients with special care, to prevent loading problems. For example, create user accounts with different configuration access capabilities, so that contact center staff can log in to Genesys Administrator and perform only those 48 Genesys tasks they are required to perform over the configuration objects for which they have permissions. This saves Genesys Administrator from loading all the objects from the Configuration Database.

Solution Control Server Monitoring Limits

In a single Solution Control Server (SCS) environment, Genesys recommends that you limit the number of hosts controlled by that single SCS to 250. This will ensure that SCS is able to react to host failures within 20 seconds if the overall number of hosts having failed is less than 50. You can assign fewer hosts to SCS if you want to decrease reaction time, or if you expect that a large number of hosts might fail at any one time.

If you want to control more hosts than the single SCS can support, either 250 or fewer depending on expected failure rates and required reaction time, consider acquiring additional Solution Control Servers and deploying them in a Distributed SCS configuration. Refer to the *Framework Deployment Guide* for more information about this configuration, and to the *Genesys Licensing Guide* for additional licensing requirements.

Genesys

Chapter

5 License Reporting Manager 8.x Solution

The information in this chapter applies to the 8.x releases of License Reporting Manager (LRM).

This chapter describes the factors that affect LRM 8.1 performance, and lists sample performance measurements for reference platforms in large-scale, single- and multi-tenant deployments.

Note: Before proceeding, review the "Architecture" section in the *License Reporting Manager Deployment Guide* for your release of LRM to familiarize yourself with the product architecture.

This chapter contains the following sections:

- About License Reporting Manager 8.x, page 49
- Architecture in Release 8.x, page 51
- Hardware Sizing and Performance Information, page 51
- Performance Test Results, page 54

About License Reporting Manager 8.x

License Reporting Manager (LRM) measures and stores usage data for licensed Genesys products and user-defined bundles, providing Genesys users with license management reports and Hosted Service Providers with billing data. The existing Genesys reporting components ICON and GVP Reporting Server perform the first level of event analysis and data storage. LRM then performs data analysis and aggregation from these reporting components into usage data for the various sellable items.The usage data is collected in LRM and may be accessed in one of two ways:

- Using a custom billing adapter that is designed to extract data in the form that the hosted service provider's back-office billing systems need.
- Using on-demand reports that can be run from the Genesys Administrator Extension (GAX) user interface. See the GAX help for details.

LRM is a server application that uses the data from Interaction Concentrator Database (IDB), the GVP Reporting Servers, and Configuration Manager to execute data analysis and summarization. The LRM Server should always be running, so it is available to respond to HTTP requests and generate reports from the LRM Database (LRM DB).

The LRM has a nightly statistics job which you can schedule to run at the same time each day when there is a low volume of interaction in the contact center (for example, at night). During this nightly statistics job, LRM performs the following tasks:

- Reads the configuration data for the various Genesys components to calculate the concurrent peak usage for certain sellable items.
- Generates concurrent peak usage data for various sellable items from ICON:
  - Reads the login session data from all the ICON instances connected to the LRM and temporarily stores the results in the LRM DB.
  - Calculates all sellable items in minute intervals at the tenant level.
  - Calculates and stores the daily value for all sellable items at the tenant and system levels.
  - Calculates and stores the daily value for all user-defined bundles at the tenant and system levels. See Predefined and User-Defined Bundles for details.

Note: For LRM to retrieve data from ICON, the ICON DB must be running and available, although ICON itself is not required.

- Generates concurrent peak usage data for some sellable items for GVP, for each GVP Reporting Server, and GVP-related sellable item:
  - Creates HTTP requests for each of these sellable items, all the tenants, and the system.
  - Stores the data in the LRM DB.
- Generates enabled seat count data for sellable items from Configuration Server:
  - Takes a snapshot of the Places and DN objects in configuration.
  - Stores the data in the LRM DB.
- Generates Agent Groups and Place Groups usage data if the configuration options are set to enable this functionality.

LRM also accepts HTTP requests for reports and generates HTTP responses containing these reports. The GAX plug-in for LRM uses this HTTP service to generate reports.

You can also use other services to generate their own reports by using the LRM web services API.

Architecture in Release 8.x

The LRM architecture is flexible and scalable. However, there are certain rules and requirements that must be observed. For detailed information about the architecture that LRM 8.x supports, see the "About LRM" chapter in the *License Reporting Manager Deployment Guide*.

Hardware Sizing and Performance Information

The Standard Server configurations (GEN_LINUX_SERVER and GEN_WIN_SERVER) that are listed in "Recommended Platform Configurations" in the Preface of the *Genesys Hardware Sizing Guide* should be sufficient for LRM to process concurrent usage and enabled-seats data in a medium-to-large contact center. Genesys recommends using a separate server or Virtual Machine (VM) with equivalent dedicated resources for the LRM server in a medium to large contact center. For sizing information on other Genesys products that are used with LRM, see the *Genesys Hardware Sizing Guide*.

For a small to medium contact center, you can host LRM on the same server or VM as other Genesys applications, if the other applications are not very active when LRM runs the scheduled nightly statistics job to gather and calculate statistics from ICON, GVP and Configuration Server.

Genesys conducted both single-tenant and multi-tenant measurements with 100,000 Agents, 1500 Agent Groups and 100 Place Groups configured in the Configuration Servers. Table 11 provides details about the number of logged-in simulated agents. LRM retrieves data from Configuration Server every 10 minutes and the enabled and concurrent seats reports show the distribution between different agent groups and place groups.

In the large contact center, Genesys configured 16 tenants in the multi-tenant system. A total of 21,000 enabled simulated agents distributed among the 16 tenants logged in daily (7000 were simultaneously logged in during 3 shifts as shown in Table 11).

LRM reported the usage for the following sellable items: CIM MS, SIP Server, and Agent Connector seats.

Table 11: Summary of Recommendations

	Medium Contact Centre	Large Contact Centre
Total agents	1000	21,000
Total shifts	2	3
Agents/shift	500	7000
Logins/day	1000	21,000
Products	CIM	CIM
Minimum recommended free disk space for LRM DB	10 GB	20 GB
Disk space recommended for LRM DB to keep >13 months of reporting data	Approximately 20 GB, depending upon number of tenants, agents, places, tenants, sellable items, bundles	Approximately 40 GB, depending upon number of tenants, agents, places, tenants, sellable items, bundles
Minimum recommended CPU	1 Intel CPU 2.6 GHz core 2 duo	2 Intel CPU 2.6 GHz core 2 duo
Minimum recommended Memory	4 GB	8 GB *
Dedicated server requirement	No	Yes
VM info/recommendations	VMs are acceptable if they have similar CPU, memory, and disk resources	VMs are acceptable if they have similar CPU, memory, and disk resource

For a large contact center, if the system has more than 50,000 agents, LRM requires additional Java Heap space to load the agent information from the Configuration Server. See the *License Reporting Manager 8.1 Deployment Guide* for details.

In addition if the number of agents exceeds 100,000 agents, Genesys recommends that LRM is deployed on a host or Virtual Machine with a minimum of 8 GB RAM.

Performance Testing Scenarios

Genesys performed LRM load testing for the following hardware and software platforms, and databases to create the sizing guidelines for LRM 8.1.2.

LRM Server Platform Configurations

Virtual machines (VM) hosted LRM and other Genesys components.

Base ESX server configuration:

- CPU Core = 16 (Intel(R) Xeon® CPU E5-26650 @ 2.40GHZ)
- RAM = 128 GB
- Hard Disk Space = 4 TB
- NICs = 16

Table 12: Performance Testing Configuration

	Linux Virtual Machine: GEN_LINUX_SERVER	Microsoft Windows Virtual Machine: GEN_WIN_SERVER	Microsoft Windows Physical host: R2_WIN_SERVER
OS	Red Hat Linux Enterprise Server v6	Windows Server 2008 R2 Enterprise	Windows Server 2008 R2 Enterprise
Processor Type	2 cores of Intel® Xeon® CPU E5-26650	2 cores of Intel® Xeon® CPU E5-26650	2 processors (16 cores) of Intel® Xeon® CPU E5630
Speed	2.40 GHZ	2.40 GHZ	2.53 GHZ
Memory Size (RAM)	4 GB	4 GB	8 GB
Hard Disk Space	40 GB	40 GB	80 GB
Ports	2 TP Ethernet 1000/100 BASE-T cards in full duplex mode	2 Ethernet 1000/100 BASE-T ports in full duplex mode	2 Ethernet 1000/100 BASE-T ports in full duplex mode

For each Virtual Machine (VM) configuration, Genesys hosted LRM on one VM, and hosted Management Framework, ICON, Database Server, SIP Server and other Genesys components on another virtual machine of type GEN_WIN_SERVER (VM).

Databases:

- MS SQL Server 2008
- Oracle 11g
- PostgreSQL 9.0

LRM Operating Systems:

- Windows Server 2008 R2 Enterprise
- Red Hat Linux v6 64-bit

Performance Test Results

For a large multi-tenant contact center with 100,000 configured agents distributed on 16 tenants, 1500 agent groups, and 100 place groups, License Reporting Manager release 8.1.2 running on a GEN_LINUX_SERVER VM, GEN_WINDOWS_SERVER VM, or Physical R2_WIN_SERVER successfully loaded and processed the data from the ICON DB in less than 5 minutes for all 3 database systems (MS SQL 2008, PostgreSQL 9, and Oracle 11g). Although only 21,000 logins were simulated and processed each day, the existence of 100,000 agents, 16 tenants, 1500 agent groups and 500 place groups in the configuration server is significant, because LRM not only has to read that information from the configuration server every 10 minutes but also process and report the enabled seats and concurrent peaks for the sellable item distributed among all the tenants, agent groups and place groups. The simulated agents logged into a Genesys_CIM_Platform_MS environment. The places/DNs used for agent logins were configured for a SIP Switch. Therefore, LRM computed concurrent peaks for Genesys_CIM_Platform_MS, Agent Connector, and SIP Server sellable items.

CPU Usage

The maximum levels of CPU usage occurred for less than five minutes when the nightly job processed the data for the entire day. There was also a spike in the CPU utilization for a few seconds every 10 minutes when LRM collects data from the Configuration Server.

Table 13: CPU Usage

Environment	Machine	RAM	Maximum CPU usage (%)
Single-tenant	Physical	8 GB	40% (averaged over all the cores)
Multi-tenant	Physical	8 GB	80%
Multi-tenant	Virtual	4 GB	95%

Note: No measurements were conducted for a single-tenant environment on a 4 GB Virtual Machine, but its CPU utilization should be less than that of the multi-tenanted environment.

Memory Usage

Genesys recommends that LRM is deployed on a host or Virtual Machine with a minimum of 8 GB RAM for very large configurations with 100,000 agents or more.

Table 14 shows the actual memory (virtual and physical) used by the system on the host or VM running the LRM service.

Table 14: Memory Usage

Platform	Virtual Memory	Physical Memory
Windows	2.5 GB	2.5 GB
Linux	2.5 GB	3.4 GB

Disk Usage on Database Server host

The database size grows overall by an average of less than 10 MB each day. The database can grow temporarily by as much as 300 MB to store temporary data for the group usage details.

Small and Medium Contact Center

For a medium contact center with 2 tenants, each having 500 simulated agents, LRM running on a Linux or Windows VM as described above, processed the data for the entire day within two minutes on all 3 database systems. The size of the database grew by less than 5MB each day.

Note: As the performance measurements were conducted with simulated logins in a SIP Server-based MS CIM Platform only, it is possible that in a more complex customer environment where Genesys Interaction Workspace, Genesys Agent Desktop with GAD Social Media Plug-ins, SIP Voicemail server ports, GVP ASR, TTS and Media ports, and additional sellable items are used, the processing time for LRM could increase by a factor of 5-10. However, because LRM requires so little CPU and memory resources to process 21,000 simple agent logins for a day suggests that even with processing times increasing by a factor of 5-10, LRM would still require less than an hour to finish processing a day's worth of data.

Chapter

6 Genesys Interaction Concentrator

Interaction Concentrator collects and stores detailed data about the interactions and resources in customer interaction networks that use Genesys Framework (contact center, enterprise-wide, or multi-enterprise telephony and computer networks). Interaction Concentrator consists of two components:

- Interaction Concentrator (ICON) server
- Interaction Database (IDB)

This chapter describes the hardware architecture for the ICON server and IDB components, providing examples of architectures for single-site and multi-site deployments. It also describes the factors that affect Interaction Concentrator performance, and lists sample performance measurements for reference platforms for both Windows and UNIX in single-site and multi-site deployments. This chapter also describes testing that was conducted in a large-scale environment to examine the performance and scalability of ICON 8.0.0 and 7.6.x. For testing results for ICON 8.1.5, see the ICON 8.1.5 *Sizing Guide Whitepaper*.

Note: Before proceeding, review the product overview chapter in the *Interaction Concentrator Deployment Guide,* to familiarize yourself with the product architecture.

This chapter contains the following sections:

- Hardware Architectures, page 58
- Interaction Concentrator Performance, page 64
- Interaction Concentrator 8.0.0 Performance, page 64
- Interaction Concentrator 7.6.1 Performance, page 73
- Interaction Concentrator 7.6.0 Performance, page 86
- Interaction Concentrator 7.5 Performance, page 97

Note: Because of their variability and complexity, large contact centers require special planning. If you want to plan an Interaction Concentrator deployment for a large contact center, contact Genesys Professional Services.

Hardware Architectures

This section provides examples of architectures for single-site and multi-site Interaction Concentrator deployments, and also provides hardware guidelines for the ICON server and IDB components.

The Interaction Concentrator architecture is flexible enough to store reporting data for a contact center environment of practically any size. The hardware architecture that you select for your Interaction Concentrator deployment depends primarily on:

- The size of your contact center, in terms of the number of daily interactions and the number of agents. This determines the requirements for your Genesys Framework Configuration Layer and Media Layer components, which are the major sources of data for Interaction Concentrator.
- The type of data that you need to collect. In a multi-site environment with multiple IDBs, considerations include whether and how your downstream reporting application will merge inter-site interactions.

  You can balance the database-writing load by distributing some ICON roles among ICON instances. For more information, see the section about ICON roles in the product overview chapter in the *Interaction Concentrator Deployment Guide.*

For information about the hardware requirements for the Framework Configuration Layer, Management Layer, and Media Layer components, see the chapters about Small Contact Centers and Medium Contact Centers.

For information about the hardware requirements for the ICON server, see “ICON Server” on page 62.

For information about sizing IDB in order to calculate the hardware requirements for the database, see “Interaction Database” on page 63.

Deployment Scenarios

This section provides examples of the following basic deployment types:

- A single-site deployment, with a single ICON instance writing to a single IDB instance (see page 59)
- A multi-site deployment, with a single ICON instance writing to a single IDB instance for the entire contact center (see page 59)

- A multi-site deployment, with multiple ICON instances writing to a single IDB instance (see page 60)
- A multi-site deployment, with multiple ICON instances writing to multiple IDB instances (see page 61)

Note: For network deployments, the ICON application connects to both the Network T-Server and the premise T-Servers.

For more information about the supported Interaction Concentrator deployments, see the section about deployment scenarios in the product overview chapter in the *Interaction Concentrator Deployment Guide*.

Diagram Conventions

To maintain focus, the diagrams in this section omit the Genesys application servers that are required for additional, optional Interaction Concentrator functionality—for example, Interaction Server, SIP Server, Network T-Server, or Outbound Contact Server (OCS).

Single-Site Deployment—One ICON, One IDB

Figure 11 depicts a sample hardware architecture for a small-size contact center in which a single ICON instance writes to a single IDB instance. The ICON instance is connected to a single T-Server.

This architecture is also suitable for eServices/Multimedia solutions—the Interaction Server occupies the same position in the architecture as T-Server.

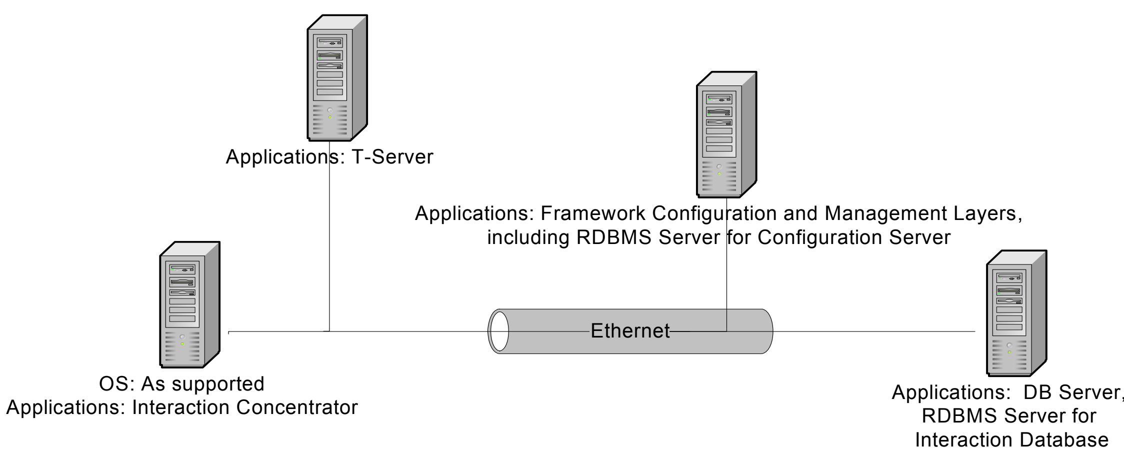


Figure 11: Single-Site Deployment with One ICON and One IDB

Multi-Site Deployment—One ICON, One IDB

Figure 12 depicts a sample hardware architecture for a medium-size contact center in which a single ICON instance, writing to a single IDB instance, serves all sites in the contact center. The ICON instance is connected to multiple T-Servers.

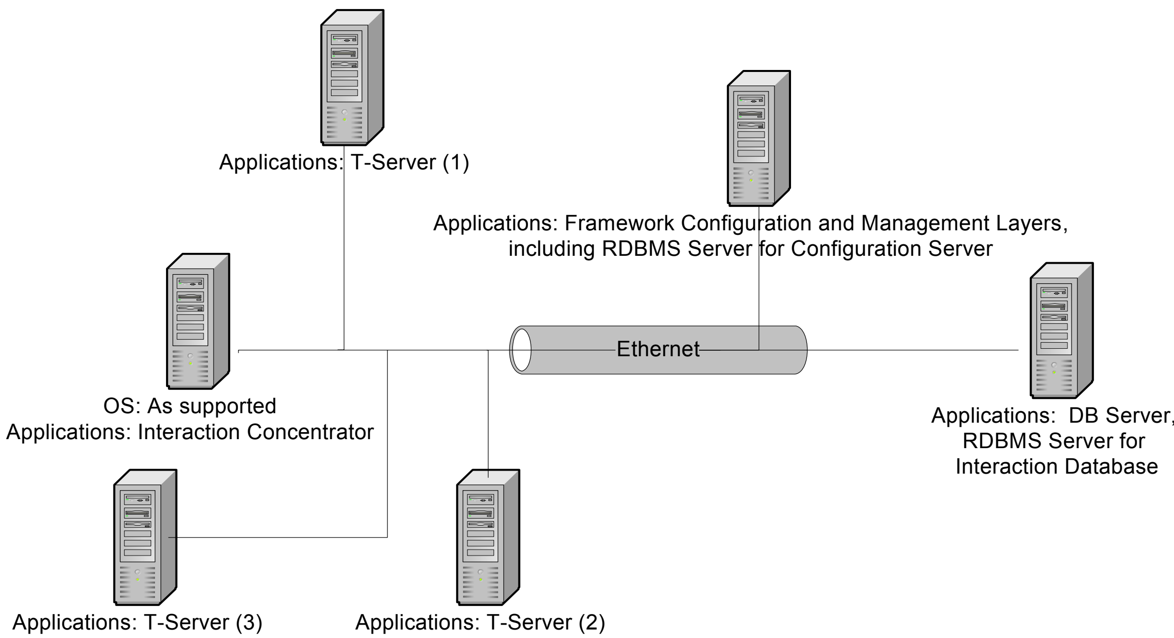


Figure 12: Multi-Site Deployment with One ICON and One IDB

Multi-Site Deployment—Multiple ICONs, One IDB

Figure 13 depicts a sample hardware architecture for a medium-size contact center in which there are multiple ICON instances, all of which write to the same IDB instance. Each ICON instance is connected to a single T-Server.

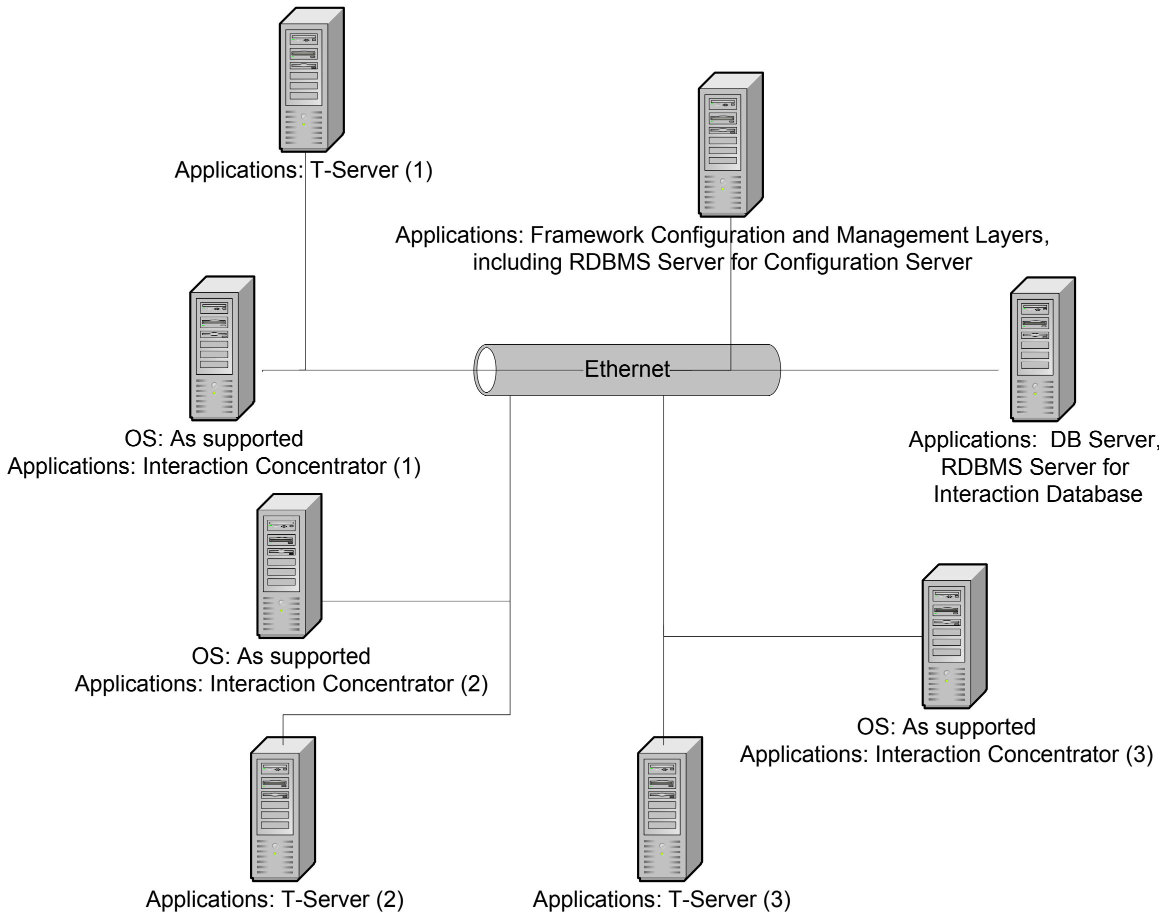


Figure 13: Multi-Site Deployment with Multiple ICONs and One IDB

Note: Genesys recommends using the multiple ICON, single IDB deployment only when there is a compelling reason to do so (for example, to use multiple network T-Servers in load-balancing mode). Genesys test results show that the performance of IDB is the limiting factor for overall Interaction Concentrator performance. Compared with the multiple ICON, multiple IDB deployment (see Figure 14), the multiple ICON, single IDB deployment requires more RDBMS resources and is more likely to encounter database conflicts that adversely impact IDB performance.

Multi-Site Deployment—Multiple ICONs, Multiple IDBs

Figure 14 depicts a sample hardware architecture for a medium-size or large-size contact center in which there are multiple ICON instances, each of which writes to its own IDB instance.

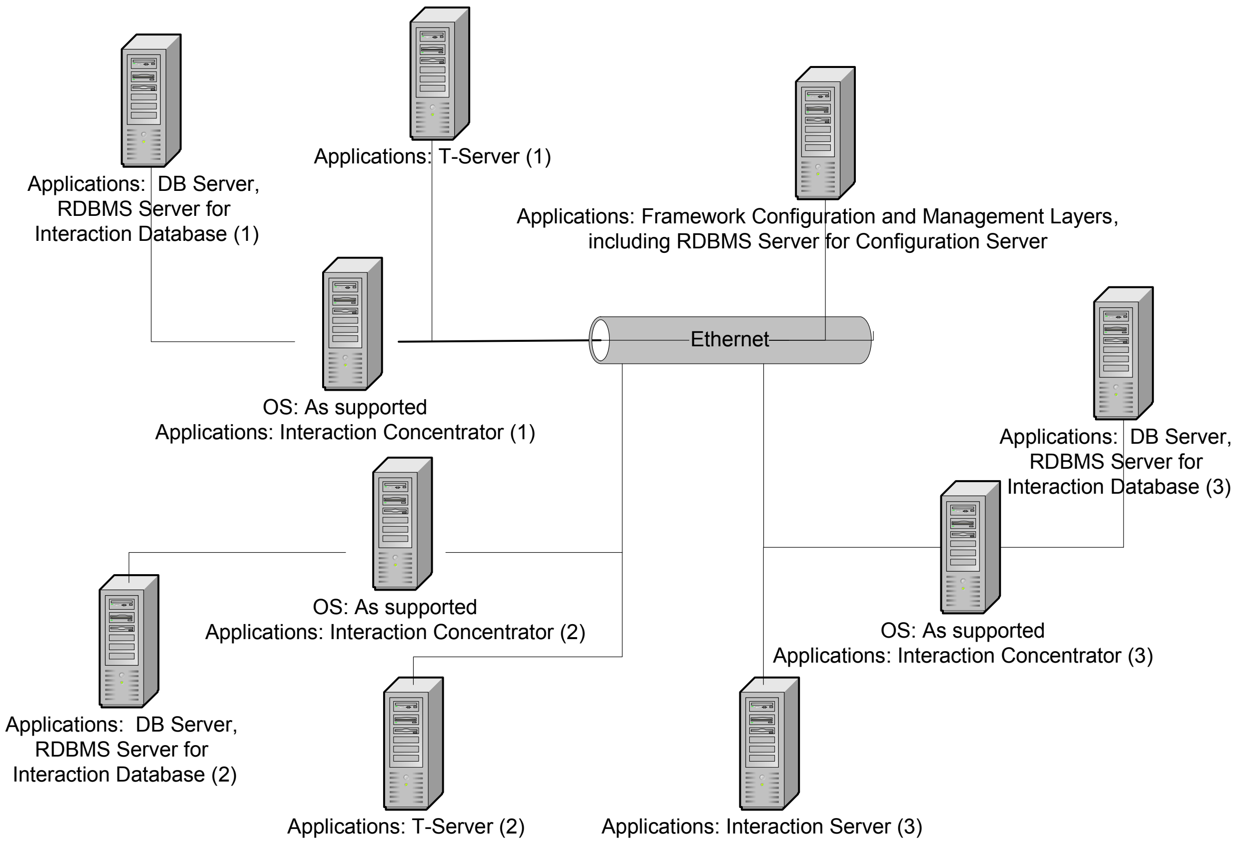


Figure 14: Multi-Site Deployment with Multiple ICONs and Multiple IDBs

Hardware Guidelines

This section describes hardware guidelines for the ICON server and IDB components. For information about the hardware requirements for the Genesys Framework components, see the chapters about Small Contact Centers and Medium Contact Centers.

ICON Server

The memory requirements for the ICON server application, including the persistent queue, depend on the specific deployment and contact center characteristics.

The following factors affect the amount of memory and processing capacity that ICON requires:

- The number of interactions, which is a function of both the overall contact center size and the complexity of interaction flows

- The ICON configuration—for example, the roles that the ICON application has been configured to perform, or configuration settings that affect persistent queue, operational memory, and database-writing operations
- The requirements for merging interaction data

Because requirements are so deployment-specific, it is not possible to provide precise hardware specifications for the ICON server. However, Genesys has conducted performance tests for Interaction Concentrator on reference platforms for both Windows and UNIX. The test hardware, in combination with the reported usage results, can serve as a baseline that you can use when you calculate your own requirements.

Interaction Database

The hardware requirements for IDB depend on the estimated size of your IDB instance or instances.

Genesys provides an interactive tool to help you estimate the required size of your IDB. This tool, the *Interaction Concentrator Database Size Estimator*, is a Microsoft Office Excel spreadsheet that is available from the Genesys Customer Care website.

The spreadsheet uses relational database management system (RDBMS)–specific information and user input about general flow characteristics to provide an estimate of table, index, and total database size by day, month, and year, assuming that IDB is not periodically purged. Tooltips in the spreadsheet itself explain the information that you are required to enter.

Note: The *Interaction Concentrator Database Size Estimator* provides a reasonably generous estimate, based on average projected activity. The estimator does not factor in growth associated with increased business.

Database Configuration and Optimization

Tablespace configuration and database tuning can significantly affect the performance and stability of Interaction Concentrator and your downstream reporting application.

Reducing I/O Contention

Any system configuration or database strategies to optimize database input and output (I/O) will significantly improve Interaction Concentrator performance. In particular, storing RDBMS logs, indexes, and table data on different disk drives reduces I/O contention. Genesys strongly recommends that you equip your RDBMS host with a fiber array or with a disk subsystem that contains multiple SCSI disk drives.

Recommendations for High Call Volumes

For environments with high call volumes, Genesys strongly recommends a multi-spindle disk subsystem, preferably with an advanced controller with write-back cache. Genesys also recommends locating the database log and temporary tablespace on disks that are separate from the disks where tables and indexes are stored.

In addition, for high call volumes, Genesys recommends that you configure IDB with a large buffer cache (hundreds of megabytes, if not gigabytes).

Database Settings For information about the database settings that were used for the Genesys 7.5 performance tests, see "Database Settings" on page 99.

Interaction Concentrator Performance

The primary factors that affect Interaction Concentrator performance are the following:

- The size of IDB
- The type of RDBMS
- RDBMS settings
- For multi-site deployments that write to a centralized IDB, the frequency with which the merge procedure is run
- Tuning of IDB
- The speed of the network connections between components
- The amount of business data attached to interactions

The following sections illustrate the significance of each of these factors in the context of specific ICON releases.

Note: Unless specified otherwise, call rates that are cited in the descriptions of the test conditions are for the contact center as a whole, not for each T-Server or Interaction Server.

Interaction Concentrator 8.0.0 Performance

Genesys performance testing of ICON 8.0.0 focused on validating the following performance-based requirements of ICON 8.0.0 in a large-scale testing environment:

- Ability to handle a large number of active interactions
- Ability to catch up on a reasonable backlog without loss of data
- Benchmark performance and reliability in a Voice environment
- Endurance performance in a Multimedia environment

Identical tests were conducted on the following platforms:

- Microsoft Windows Server 2003 (32-bit), with Microsoft SQL Server 2005

- Solaris 10 SPARC (64-bit), with Oracle Database 10g Release 2 (version 10.2.0.4 for IDB; version 10.2.0.3 for Configuration Database)

Two separate test environments were utilized—a voice-specific deployment and a multimedia (e-mail) deployment.

For comparison against 7.6.1 performance, the same baseline tests were run with a preproduction version of ICON 8.0.0 and with ICON 7.6.1 on both Windows and Solaris platforms. Aside from the ICON release and associated IDB schema version, the comparison tests utilized the same test environments.

The following sections describe the sample ICON 8.0.0 and 7.6.1 environments, call flows, tests, results, and recommendations.

Hardware and Software

Table 15 describes the hardware that hosted Interaction Concentrator and the software versions for Interaction Concentrator and other components in the ICON 8.0.0 and 7.6.1 environments.

The hardware setup was similar to the architecture depicted in Figure 12 on page 60, except that, for the tests conducted on a Solaris platform with Oracle, the ICON application and IDB were hosted on the same machine.

Table 15: Hardware and Software in the ICON 8.0.0 and ICON 7.6.1 Environments

Application	Application Software Version	
	On Windows Server 2003[a]	On Solaris 10[b]
ICON (Windows)[c] **Host:** E5450 dual quad core Xeon 3.00 GHz **Memory:** 8 GB RAM	8.0.000.21 7.6.100.26	n/a
ICON (Solaris)[c] **Host:** Quad SPARC64 VI 2.15 GHz[d] **Memory:** 32 GB RAM	n/a	8.0.000.21 7.6.100.26
RDBMS for IDB (Windows) **Host:** E5450 dual quad core Xeon 3.00 GHz **Memory:** 8 GB RAM	Microsoft SQL Server 2005	n/a
RDBMS for IDB (Solaris)[e] **Host:** Quad SPARC64 VI 2.15 GHz[d] **Memory:** 32 GB RAM	n/a	Oracle 10.2.0.4

Table 15: Hardware and Software in the ICON 8.0.0 and ICON 7.6.1 Environments (Continued)

Application	Application Software Version	
	On Windows Server 2003[a]	On Solaris 10[b]
T-Server	Avaya 7.0.006.04	n/a
Interaction Server	7.6.100.21	n/a
Interaction Proxies	n/a	7.6.100.04
Configuration Server	7.6.000.06	n/a
RDBMS for Configuration Database	n/a	Oracle 10.2.0.3
DB Server (ICON)	7.6.000.08	n/a
DB Server (Interaction Server)	7.6.000.08	n/a
Stat Server	7.6.100.12	n/a
Universal Routing Server (URS)	7.6.100.04	n/a

a. Windows Server 2003 R2 Enterprise Edition SP2 (32-bit)

b. Solaris 10 (Sun OS 5.10 Generic) 64-bit 10/09

c. Although ICON can be deployed on a multithreaded machine, it is not a multi-threaded application.

d. The ICON application, RDBMS server, and IDB were hosted on the same machine.

e. Multi-threaded application.

Voice Test Environment

The voice test environment was organized as an enterprise with two tenants, each containing 20 T-Servers. On each switch, 640 agents and 200 behind-the-switch IVR ports were configured. Call configuration included 54 key-value pairs (KVPs) of attached data. A total of 20,000 agents were logged in. All the T-Servers were served by a single ICON and IDB. The deployment included one URS application. During the tests, the logging level was set to Standard for all processes.

The topology was similar to the environment that was originally used to test the ICON 7.6.1 release (see Figure 15 on page 74).

Call Flow and Configuration

Two Network T-Servers received generated calls and presented them to the premise IVRs. Twenty percent of the calls were transferred to agents for handling, and 10 percent of these were subsequently transferred.

The call flow is as follows:

- Inbound calls arrive at a route point on each Network T-Server.
- A strategy on the route point attaches 100 bytes of attached data to the call.
- A tenant is selected, and a DN group that contains all of the IVR queues for that tenant is targeted.
- The call is delivered to the IVR queue on a premise switch and is then diverted to an IVR port.
- Eighty percent of the calls are released after 1 minute of talk time on the IVR. The remaining 20 percent are transferred to a local agent queue.
- Ninety percent of the transferred calls are handled by the targeted agent. Ten percent of the transferred calls are further transferred by single-step or two-step transfer, conference, or circular transfer (transferred off-premise and then transferred back). Total handle time is 120 seconds.

A single ICON monitored all call activity and stored data in a single IDB. ICON was configured to store 54 attached data keys, and the merge procedure was configured to run at 5-minute intervals. ICON options to filter the data that is stored in IDB were turned off. Each test started with an empty IDB, and IDB was not purged during the test.

Voice Tests and Results

The following tests were run in the voice-specific environment:

- ICON 8.0.0 and 7.6.1 Voice baseline comparison test
- ICON 8.0.0 and 7.6.1 Voice baseline comparison test, with backlog (page 68)
- ICON 8.0.0 Voice benchmark test (page 68)
- ICON 8.0.0 Voice benchmark test with network load balancing (page 69)

ICON 8.0.0 and 7.6.1 Voice Baseline Comparison Test

Genesys monitored ICON 8.0.0 and ICON 7.6.1 over a period of 24 hours to compare the performance of each version. The tests were conducted under the following conditions:

- 2-hour run with a call rate of 16 calls/second.
- 3-hour run with a call rate of 34 calls/second.
- 19-hour run with a call rate of 16 calls/second.

- The contact center environment and ICON application were configured as described in "Call Flow and Configuration" on page 67. The Network T-Servers did not function in load-balancing mode.

Test Results

ICON performed without failure or performance degradation during the test. Compared with ICON 7.6.1, memory utilization in ICON 8.0.0 shows an improvement of approximately 30 percent in both Windows and Solaris. However, CPU utilization shows an increase of approximately 70 percent in Windows and 26 percent in Solaris.

For details about CPU and memory utilization during the test, see Table 16 on page 69.

ICON 8.0.0 and 7.6.1 Voice Baseline Comparison Test, with Backlog

Genesys monitored ICON 8.0.0 and ICON 7.6.1 over a period of 24 hours to compare the performance of each version when there is a backlog of data in the persistent queue (for example, when an unavailable IDB or DB Server failure prevents ICON from storing data in the database). The tests were conducted under the same conditions as the voice baseline comparison test (see page 67), with the following additional condition:

- A backlog of three hours' worth of data was deliberately caused during the peak period of call volume (34 calls/second)—the data was available in the ICON persistent queue, and ICON continued to collect data during and after the peak period backlog, but ICON was prevented from writing the data to IDB until the next off-peak period (16 calls/second).

Test Results

ICON performed without failure during the test, and no data was lost. ICON recovered within a timeframe that was significantly shorter than the backlog, so that ICON did not continue to fall behind.

For details about CPU and memory utilization during the test, see Table 16 on page 69.

ICON 8.0.0 Voice Benchmark Test

Genesys monitored ICON 8.0.0 over a period of 24 hours to verify performance when all the capabilities that are implemented in ICON 8.0.0 are enabled in a network environment. The test was conducted under the same conditions as the voice baseline comparison test (see page 67).

Test Results

ICON performed without failure during the test.

For details about ICON CPU and memory utilization during the test, see Table 16 on page 69.

ICON 8.0.0 Voice Benchmark Test with Network Load Balancing

Genesys monitored ICON 8.0.0 over a period of 24 hours to identify if there is any impact on performance when network load balancing is implemented. The test was conducted under the same conditions as the voice benchmark test (see page 67), except for the following condition:

- The network switch and ICON were configured to identify that the Network T-Servers were working with the network switch in load-balancing mode.

Test Results

There was no significant impact on ICON performance in a network deployment in which the Network T-Servers operated in load-balancing mode.

For details about CPU and memory utilization during the test, see Table 16 on page 69.

Summary of Voice Test Results

Table 16 summarizes the results of the voice performance tests, in terms of CPU and memory utilization by the ICON application.

Table 16: ICON 8.0.0 Voice Performance Test Results

Performance Test	ICON 8.0.0 Test Results			
	Windows		Solaris	
	Average CPU (%)[a]	Maximum RAM (MB)[b]	Average CPU (%)[a]	Maximum RAM (MB)[b]
Baseline: ICON 8.0.0 (Baseline: ICON 7.6.1)	13.8 (8.1)	728.4 (1045.5)	15.6 (12.4)	570.3 (811.6)
Baseline, with backlog (Baseline, with backlog: ICON 7.6.1)	13.8[c] (8.1)	728.4 (1045.5)	15.6[d] (12.4)	570.3 (811.6)

Table 16: ICON 8.0.0 Voice Performance Test Results (Continued)

Performance Test	ICON 8.0.0 Test Results			
	Windows		Solaris	
	Average CPU (%)[a]	Maximum RAM (MB)[b]	Average CPU (%)[a]	Maximum RAM (MB)[b]
Voice benchmark	13.5	724.5	15.5	572
Voice benchmark, with load balancing	12.7	749.0	9.76	566.5

a. Average utilization, in percent, of one core per thread.

b. Maximum memory utilization, in MB of RAM.

c. CPU utilization went up to 27.6 percent during the catchup phase. ICON recovered from the backlog within 1 hour 20 minutes.

d. CPU utilization went up to 37.5 percent during the catchup phase. ICON recovered from the backlog within 1 hour 10 minutes.

Multimedia Test Environment

The environment for testing multimedia (e-mail) interactions was organized as a single contact center under a single tenant that used a multi-tenant Configuration Server. The contact center blended support for voice and multimedia interactions. The deployment included one Interaction Server and two Interaction Server Proxies. It also included four URS applications. A single ICON application was configured to collect multimedia interactions from Interaction Server. During the tests, the logging level was set to Standard for all processes.

A total of 32,000 agents were logged in. The average number of skills configured for each agent was 30. All agents were able to handle voice and multimedia interactions concurrently. For the performance tests, 4000 agents were logged in for the e-mail media type, and 16,000 agents were logged in for the chat media type. Routing was performed for the e-mail media type only.

Interaction Flow and Configuration

The workflow delivered interactions of the e-mail media type to agents and stopped the interaction on completion of the workflow. The workflow included:

- 5 business processes
- 31 queues
- 15 URS strategies
- 3 database dips

- 1 ESP server call
- 2 HTTP web service calls
- A queue archiving process
- A time-in-queue condition
- More than 150 attached data KVPs
- Skills-based routing
- Stat Server statistics calls

On average, each interaction used 6 strategies and 7 queues, including 1 virtual queue.

A single ICON monitored all multimedia interaction activity and stored data in a single IDB, which was located on a different host from the Interaction Server database. ICON was configured to store all the attached data KVPs that were included in the workflow. ICON options to filter the data that is stored in IDB were turned off, but the memory management options (for example, om-memory-optimization and om-max-in-memory) were set to optimize memory usage. Each test started with an empty IDB, and IDB was not purged during the test.

Multimedia Tests and Results

The following tests were run in the multimedia environment:

- ICON 8.0.0 Multimedia baseline test
- ICON 8.0.0 Multimedia endurance run (page 72)

ICON 8.0.0 Multimedia Baseline Test

Genesys monitored ICON 8.0.0 over a period of 24 hours to verify acceptable baseline performance. The test was conducted under the following conditions:

- 14 e-mails/second.
- 4000 e-mail agents.
- The contact center environment and ICON application were configured as described in “Interaction Flow and Configuration” on page 70.

Test Results

ICON performed without failure or performance degradation during the test, with acceptable memory and CPU utilization.

For details about CPU and memory utilization during the test, see Table 17 on page 72.

ICON 8.0.0 Multimedia Endurance Run

Genesys monitored ICON 8.0.0 over a period of 7 days to verify performance during an extended, continuous run with a large number of complex interaction flows. The test was conducted under the same conditions as the multimedia baseline test (see page 71).

Test Results

ICON was able to handle 14 multimedia interactions per second for 7 days without failure or performance degradation. The memory management options enabled ICON to handle large accumulations of interactions without utilizing an excessive amount of memory.

For details about CPU and memory utilization during the test, see Table 17 on page 72.

Summary of Multimedia Test Results

Table 17 summarizes the results of the multimedia performance tests, in terms of CPU and memory utilization by the ICON application and, for the endurance run, CPU utilization by the RDBMS server.

Table 17: ICON 8.0.0 Multimedia Performance Test Results

Performance Test	ICON 8.0.0 Test Results			
	Windows		Solaris	
	Average CPU (%)[a]	Maximum RAM (MB)[b]	Average CPU (%)[a]	Maximum RAM (MB)[b]
Baseline: ICON 8.0.0	8.28	Not recorded	9.44	623.0
Multimedia endurance: ICON 8.0.0 RDBMS server	6.3 5.5 (MSSQL)	1044.4 n/a	28.8 16.8 (Oracle)	619.9 n/a

a. Average utilization, in percent, of one core per thread.

b. Maximum memory utilization, in MB of RAM.

ICON 8.0.0 Performance Conclusions

Based on the results of performance testing that was conducted in a large-scale environment, Genesys has the following observations and conclusions:

- ICON 8.0.0 demonstrates improvement over ICON 7.6.1.

- ICON can recover successfully if a backlog builds up under load conditions.
- The memory management options are important in controlling ICON memory consumption, so that ICON can continue to operate under load conditions while a large number of interactions are backlogged.
- In a multimedia deployment, it is important that IDB and the Interaction Server database are not hosted on the same machine.

Interaction Concentrator 7.6.1 Performance

Genesys performance testing of ICON 7.6.1 focused on validating the following performance-based requirements of ICON 7.6.1 in a large-scale testing environment:

- Handling a large number of active interactions
- Filtering out unnecessary Multimedia interaction data
- Purging Multimedia interaction data

Two separate test environments were utilized—a voice-specific deployment and an open media (e-mail) deployment. The following sections describe the sample ICON 7.6.1 environments, call flows, tests, results, and recommendations.

Voice Test Environment

The voice test environment was organized as a single contact center under a single tenant that used a multi-tenant Configuration Server (see Figure 15). The contact center had four T-Servers that represented four separate sites. The routing targets consisted of 5,000 inbound agents per T-Server, for a total of 20,000 agents. During the tests, the logging level was set to Standard for all processes.

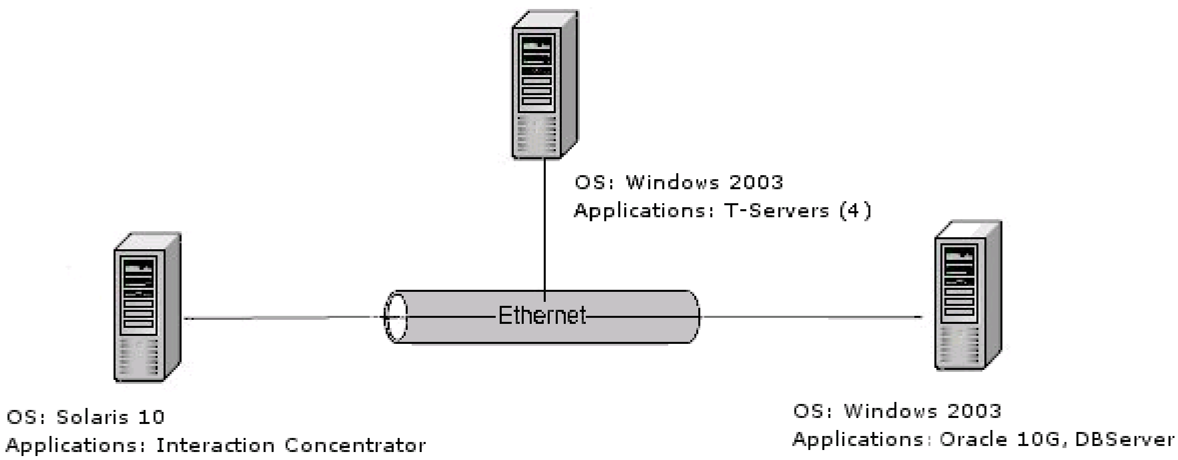


Figure 15: ICON 7.6.1 Voice Test Environment

Hardware and Software

Table 18 describes the hardware and software that hosted the ICON 7.6.1 application in the voice-specific test environment.

Table 18: Hardware and Software in the ICON 7.6.1 Voice Test Environment

Application	Processor	Memory	Application Software Version	
			On Windows 2003 Server[a]	On Solaris 10[b]
ICON	Quad dual core SPARC64 IV 2.15 GHz	32 GB RAM	n/a	See Table 19
RDBMS (Oracle 10)	E5410 dual quad core Intel Xeon 2.33 GHz	8 GB RAM	10.2.0.4	n/a
DB Server (ICON)	E5410 dual quad core Intel Xeon 2.33 GHz	8 GB RAM	7.6.000.08	n/a
Stat Server	X5355 dual quad core Intel Xeon 2.66 GHz	8 GB RAM	7.6.100.12	n/a
URS	X5355 dual quad core Intel Xeon 2.66 GHz	8 GB RAM	7.6.100.04	n/a

a. Windows Server 2003 Enterprise Edition SP2 32-bit

b. Solaris 10 (Sun OS 5.10) 64-bit 11/06

Table 19: ICON Releases Tested

Test	ICON Release
ICON 7.6.0 and 7.6.1 Comparison Test	7.6.000.16 and 7.6.100.09
Voice Purge Test	7.6.100.09

Call Flow and Configuration

Figure 16 shows the call flow and configuration that were used in the voice-specific tests.

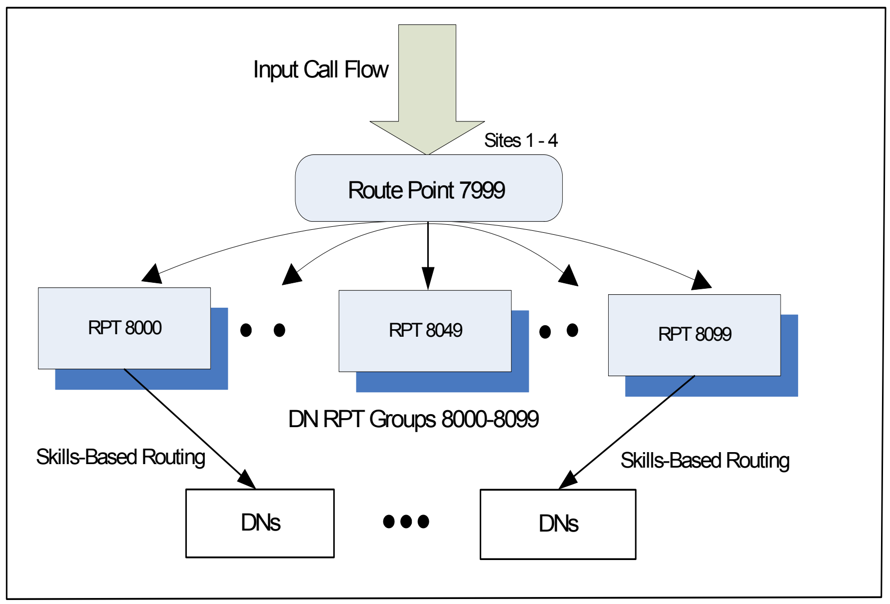


Figure 16: Voice-Specific Call Flow for ICON 7.6.1 Tests

The call flow is as follows:

- Inbound calls arrive at a route point, where they are routed in turn to a DN group of 100 route points.
- At the second route point, Configuration Server options are assigned to routing-strategy variables for reference at various points in the routing strategy.
- One KB of data is attached to each call.
- A skills-based routing strategy routes the call to the most suitable agent (across all four sites), where the call is handled for 120 seconds.

Tests and Results

The following tests were run in the voice-specific environment:

- ICON 7.6.0 and 7.6.1 Comparison Test (Voice) (see page 76)
- Voice Purge Test (see page 76)

The details of these tests and the obtained results are discussed in the following sections.

ICON 7.6.0 and 7.6.1 Comparison Test (Voice)

Genesys monitored ICON 7.6.0 and ICON 7.6.1 over a period of two hours to compare the performance of each version. The tests were conducted under the following conditions:

- Call rate of 34 calls/second
- Duration of test: 2 hours
- Twenty thousand (20,000) agents configured
- ICON application running on Solaris 10

Test Results

Performance results are consistent between the 7.6.0 and 7.6.1 releases of ICON. Table 20 provides the observed results.

Table 20: ICON 7.6.0 and 7.6.1 Comparison Test (Voice) Results

ICON Release	ICON		Oracle[a]	
	Maximum CPU (%)[b]	Maximum RAM (MB)	Maximum CPU (%)[b]	Maximum RAM (MB)
7.6.0	32	756	102	2,421
7.6.1	32	768	106	2,423

a. Multi-threaded application

b. Peak usage of one core/thread (%)

Voice Purge Test

Genesys tested the new gsysPurge76 stored procedure in the voice-specific environment over a period of three days while under a constant load. The test was conducted under the following conditions:

- Call rate of 17 calls/second
- Twenty thousand (20,000) agents configured

- Memory-management options enabled: om-memory-optimization set to true; om-max-in-memory set to 100 (default)
- Non-partitioned database

Test Results

The purge procedure was executed once every 24 hours for two days. It was able to clear approximately 204 million rows from IDB tables in 150 minutes. Although ICON did fall behind processing calls and attached-data queues while IDB tables were purged, it recovered quickly after the purge procedure completed. Table 21 displays the observed results.

Table 21: Voice Purge Test Results

Application	Maximum CPU (%)[a]	Maximum RAM (MB)
ICON	36	686
Oracle (RDBMS)[b]	175	2,440

a. Peak usage of one core/thread (%)

b. Multi-threaded application

Open Media Test Environment

The open media test environment was organized as a single contact center under a single tenant that used a multi-tenant Configuration Server. A single ICON application on a Windows platform, writing to an Oracle database on Microsoft Windows Server 2003, was configured to collect open media (e-mail) interactions from Interaction Server (see Figure 17). During the tests, the logging level was set to Standard for all processes.

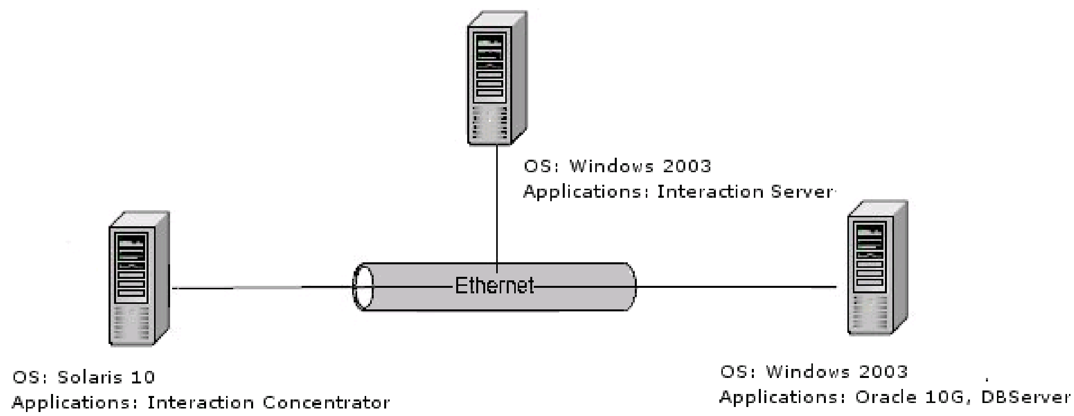


Figure 17: ICON 7.6.1 Open Media Test Environment

Hardware and Software

Table 22 describes the hardware that hosted the ICON application and other components in the 7.6.1 open media environment.

Table 22: Hardware and Software in ICON 7.6.1 Open Media Environment

Application	Processor	Memory	Application Software Version	
			On Windows 2003 Server[a]	On Solaris 10[b]
ICON	Quad dual core SPARC64 IV 2.15 GHz	32 GB RAM	n/a	See Table 23
Interaction Server	X5355 dual quad core Intel Xeon 2.66 GHz	8 GB RAM	7.6.100.24	n/a
Interaction Proxies	Dual UltraSPARC-IIIi 1.0 GHz	2 GB RAM	n/a	7.6.100.04
RDBMS (Oracle 10)	E5410 dual quad core Intel Xeon 2.33 GHz	8 GB RAM	10.2.0.4	n/a
DB Server (ICON)	E5410 dual quad core Intel Xeon 2.33 GHz	8 GB RAM	7.6.000.08	n/a
DB Server (Interaction Server)	E5410 dual quad core Intel Xeon 2.33 GHz	8 GB RAM	7.6.000.08	n/a

Table 22: Hardware and Software in ICON 7.6.1 Open Media Environment (Continued)

Application	Processor	Memory	Application Software Version	
			On Windows 2003 Server[a]	On Solaris 10[b]
Stat Server	X5355 dual quad core Intel Xeon 2.66 GHz	8 GB RAM	7.6.100.12	n/a
URS	X5355 dual quad core Intel Xeon 2.66 GHz	8 GB RAM	7.6.100.04	n/a

a. Windows Server 2003 Enterprise Edition SP2 32-bit

b. Solaris 10 (Sun OS 5.10) 64-bit 11/06

Table 23: ICON Releases Tested

Test	ICON Release
Open Media Backlog Test	7.6.100.09 (with 400,000 e-mails backlogged) 7.6.100.05 (with 0, 1 million, and 10 million e-mails backlogged)
Data Filtering Test	7.6.100.05
ICON 7.6.0 and 7.6.1 Comparison Test (Open Media)	7.6.000.16 and 7.6.100.09
Multimedia Purge Test1	7.6.100.05
Multimedia Purge Test2	7.6.100.05

Configuration and Interaction Flow

In the 7.6.1 open media model, the interaction flow is as follows:

- An interaction arrives at the entry queue and is routed to a target queue based on the business strategy.
- Five (5) KB of data, consisting of 100 key-value pairs of data, is attached.
- The interaction is directed by a skills-based routing strategy to one of the 2,000 agents who are logged in to two separate Interaction Server proxies (total of 4,000 agents).
- The agent handles the interaction for 5 minutes (300 seconds), which simulates the approximate amount of time that is required to respond to an e-mail interaction.
- The interaction is routed to an archive queue.

Figure 18 shows the call flow and configuration used in the open media tests.

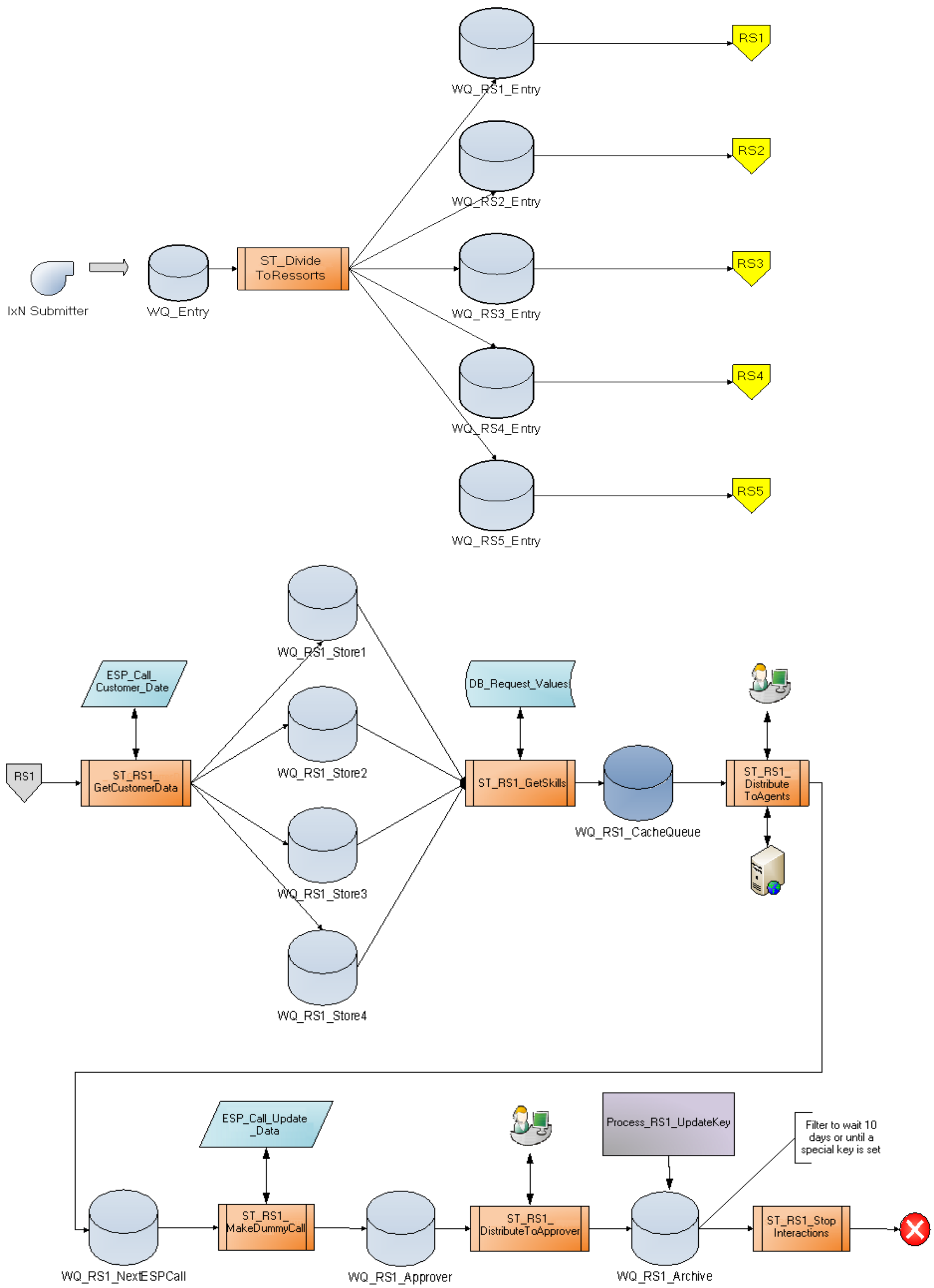


Figure 18: Open Media Call Flow for ICON 7.6.1 Tests

Tests and Results

The following tests were run in the open media environment:

- Open Media Backlog Test (see page 81)
- Data Filtering Test (see page 82)
- ICON 7.6.0 and 7.6.1 Comparison Test (Open Media) (see page 83)
- Multimedia Purge Test 1 (see page 84)
- Multimedia Purge Test 2 (see page 85)

The details of these tests and the obtained results are discussed in the following sections.

Open Media Backlog Test

In this test, Genesys varied the number of e-mail interactions that were backlogged to determine the effect on memory consumption. It was conducted under the following conditions:

- Four thousand (4,000 agents) logged in to queue
- Duration of test: 2 hours of e-mail generation
- Memory-management options enabled: om-memory-optimization set to true; om-max-in-memory set to 100 (default); om-memory-clean set to 1 on the interaction queues that are receiving the interactions.
- Four-hundred thousand (400,000) processed interactions in IDB
- Number of e-mails backlogged: 0; 400,000; 1 million; and 10 million
- ICON application running on Solaris 10

Test Results

Table 24 displays the following results:

- Oracle CPU utilization gradually increased with the backlog of e-mails until it stabilized after approximately 6 hours.
- ICON CPU utilization remained constant at 48 percent, as the number of backlogged e-mails increased.

Table 24: Open Media Backlog Test Results

Application	ICON		Oracle[a]	
# E-mails Backlogged	Maximum CPU (%)[b]	Maximum RAM (MB)	Maximum CPU (%)[b]	Maximum RAM (MB)
0	48	538	113	2,302
400,000	48	559	185	2,421

Table 24: Open Media Backlog Test Results (Continued)

Application	ICON		Oracle[a]	
# E-mails Backlogged	Maximum CPU (%)[b]	Maximum RAM (MB)	Maximum CPU (%)[b]	Maximum RAM (MB)
1 million	48	563	182	2,300
10 million	48	550	152	2,307

a. Multi-threaded application

b. Peak usage of one core/thread (%)

Data-Filtering Test

In this test, Genesys monitored the performance of ICON as the data-filtering configuration option, udata-history-terminated, was set first to 0 (default) and then to 1.

Note: By setting the udata-history-terminated configuration option to 1 on the ICON Application object, ICON will not store the call-termination value of UserData keys to IDB. For more information on this feature, see the *Interaction Concentrator 7.6 User's Guide*.

The test was conducted under the following conditions:

- Four thousand (4,000) agents logged into queue
- Duration of test: 2 hours
- No backlogged interactions
- Four-hundred thousand (400,000) processed interactions in IDB
- Memory management options enabled

Test Results

Table 25 displays the following results:

- There was no impact on ICON's CPU utilization when filtering was enabled compared with when filtering was disabled.
- Fifty-five (55) percent fewer attached-data keys were recorded in the G_USERDATA_HISTORY table, when filtering was enabled.

Table 25: Data-Filtering Test Results

Application	ICON		Oracle[a]	
Filtering	Maximum CPU (%)[b]	Maximum RAM (MB)	Maximum CPU (%)[b]	Maximum RAM (MB)
OFF	48	538	113	2,302
ON	48	535	104	2,426

a. Multi-threaded application

b. Peak usage of one core/thread (%)

ICON 7.6.0 and 7.6.1 Comparison Test (Open Media)

Genesys monitored ICON 7.6.0 and 7.6.1 over a period of 2 hours to measure any changes in performance between releases. The test was conducted under the following conditions:

- Four thousand (4,000) agents configured
- Interaction rate of 14 e-mails/second
- Duration of test: 2 hours
- Two-hundred twenty thousand (220,000) previously processed interactions in IDB
- Memory-management options enabled

Test Results

Memory utilization, CPU utilization, and Oracle CPU utilization in ICON 7.6.1 were much better than in ICON 7.6.0. Table 26 displays the following results:

- ICON 7.6.0 memory utilization increased as expected throughout the test, reaching a maximum of 4.3 GB. By contrast, ICON 7.6.1 memory utilization did not increase significantly throughout the test.
- CPU utilization in ICON 7.6.0 reached 80 percent of the available CPU, compared with a maximum of 48 percent in ICON 7.6.1. In other words, CPU utilization in ICON 7.6.0 was 67 percent higher than in ICON 7.6.1.
- With CPU utilization by Oracle (which is multithreaded) expressed in terms of core usage by a single-threaded application, Oracle CPU utilization in ICON 7.6.0 reached 334 percent, compared with a maximum of 113 percent in ICON 7.6.1. In other words, Oracle CPU utilization in ICON 7.6.0 was almost 300 percent higher than in ICON 7.6.1.

Table 26: ICON 7.6.0 and 7.6.1 Comparison Test (Open Media) Results

ICON Release	ICON		Oracle[a]	
	Maximum CPU (%)[b]	Maximum RAM (MB)	Maximum CPU (%)[b]	Maximum RAM (MB)
7.6.0	80	4,315	334	2,422
7.6.1	48	538	113	2,302

a. Multi-threaded application

b. Peak usage of one core/thread (%)

Multimedia Purge Test 1

Genesys tested the new gsysPurge76 stored procedure in the open media environment over a period of three days while under a constant load. No sizing parameters were set. The test was conducted under the following conditions:

- Interaction rate of 14 e-mails/second
- Four thousand (4,000) agents configured
- Memory-management options enabled
- Test duration: 3 days under constant load
- Non-partitioned database
- ICON application running on Solaris 10

Test Results

Testing produced the following results:

- In two main executions, the purge procedure cleared approximately 361 million and 363 million rows from IDB tables in 293 minutes and 333 minutes, respectively.
- The execution of the purge procedure did not affect ICON CPU or memory consumption (see Table 27).
- Although ICON did fall behind processing calls and attached-data queues while IDB tables were purged, it recovered quickly after the purge procedure completed.

Table 27: Multimedia Purge Test 1 Results

Application	Maximum CPU (%)[a]	Maximum RAM (MB)
ICON	48	580
Oracle[b]	300	2480

a. Peak usage of one core/thread (%)

b. Multi-threaded application

Multimedia Purge Test 2

Genesys repeated the multimedia purge test—this time, setting an IDB sizing parameter (rowspertransaction) to specify the maximum size of one database transaction. This test was conducted under the following conditions:

- Interaction rate of 14 e-mails/second
- Four thousand (4,000) agents configured
- Memory-management options set to true
- rowspertransaction parameter set to 10 million
- Duration of test: 3 days under constant load
- Non-partitioned database
- ICON application running on Solaris 10

Test Results

Testing produced the following results:

- In two purge procedure executions, 360.8 million and 360.91 million rows were cleared from IDB tables in 303 minutes and 323 minutes, respectively.
- Although ICON did fall behind processing calls and attached-data queues while IDB tables were purged, it recovered quickly after the purge procedure completed.

Table 28 provides the observed results.

Table 28: Multimedia Purge Test 2 Results

Application	Maximum CPU (%)[a]	Maximum RAM (MB)
ICON	48	575
Oracle[b]	265	2464

a. Peak usage of one core/thread (%)

b. Multi-threaded application

ICON 7.6.1 Performance Conclusions

Based on the results of performance testing that was conducted in a large-scale environment, Genesys has the following observations and conclusions:

- ICON 7.6.1 demonstrates a significant improvement over ICON 7.6.0.
- The new memory-management options are effective in controlling ICON memory consumption, in that ICON can continue to operate under load conditions while having a large number of interactions backlogged.
- ICON 7.6.1 demonstrates a significant (approximately 67 percent) improvement over ICON 7.6.0 in handling open media (e-mail) interactions under load conditions. There is also a significant (295 percent) reduction to the Oracle database load.
- ICON can operate successfully under load conditions while the new gsysPurge76 procedure executes in a timely manner to clear IDB tables. Although ICON does fall behind with database insertion while IDB tables are purged, it recovers quickly after the purge procedure has completed.

Interaction Concentrator 7.6.0 Performance

Genesys has conducted a number of tests to examine the performance and scalability of ICON 7.6.0 in a large-scale environment. Two separate test environments were utilized—a voice-specific deployment and an open media (e-mail) deployment.

The following sections describe the sample 7.6.0 environments, call flows, tests, results, and recommendations.

Voice-Specific Environment

In the voice-specific test environment, two separate ICON applications were configured to collect parallel calls on two separate platforms, as shown in Figure 19.

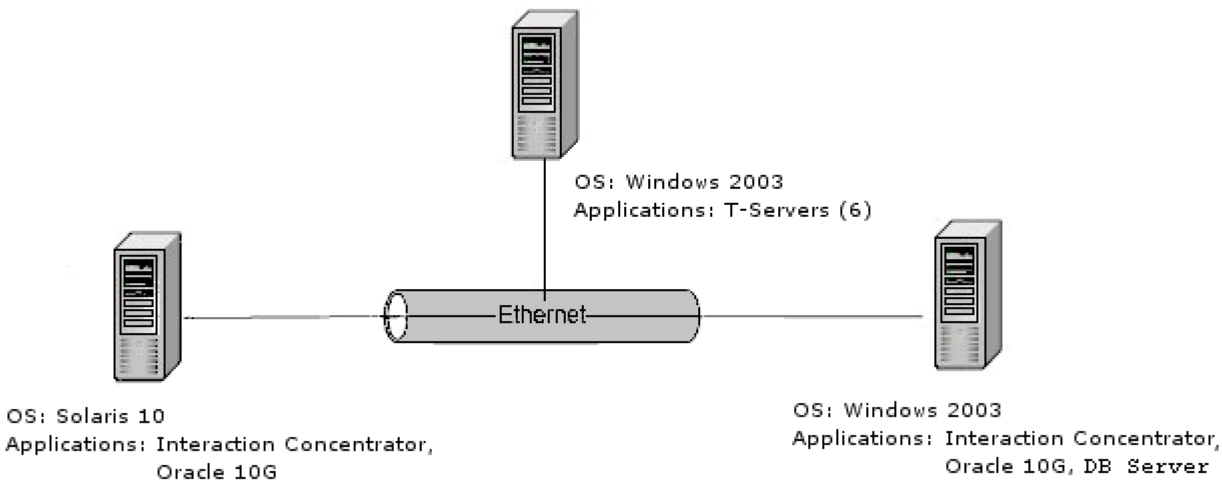


Figure 19: ICON 7.6.0 Voice-Specific Test Environment

Hardware and Software

Table 29 describes the hardware that hosted the ICON 7.6.0 application, persistent queue, and other components in the voice-specific test environment.

Table 29: Hardware and Software in Voice Test Environment

Application	Processor	Memory	Application Software Version	
			On Windows 2003 Server[a]	On Solaris 10[b]
ICON	Sunfire Dual Ultra-Sparc IIIi 1.504 GHz	2 GB RAM	See Table 30	See Table 30
RDBMS (Oracle 10)	Dual dual-core Xeon 5160 3.0 GHz	4 GB RAM	10.2.0.4.0	n/a
DB Server	Dual dual-core Xeon 5160 3.0 GHz	4 GB RAM	7.6.000.08	n/a
Stat Server	Dual Xeon 2.8 GHz	1 GB RAM	7.6.000.18	n/a
URS	Dual Xeon 2.8 GHz	1 GB RAM	7.6.001.06	n/a
T-Server (Avaya G3)	Dual Xeon 3.06 GHz	2 GB RAM	7.6.000.04	n/a

a. Windows Server 2003 Enterprise Edition SP2 32-bit

b. Solaris 10 (Sun OS 5.10) 64-bit 11/06

Table 30: ICON Releases Tested

Test	ICON Release
Comparison Test	7.5.000.22 and 7.6.000.14
Attached Data	7.6.000.16
Vary Call Rate	7.6.000.14
Data Filtering	7.6.000.14
ICON Recovery	7.6.000.16

Call Flow and Configuration

Figure 20 shows the sample call flow and configuration for the scenarios that were used in the voice-specific tests.

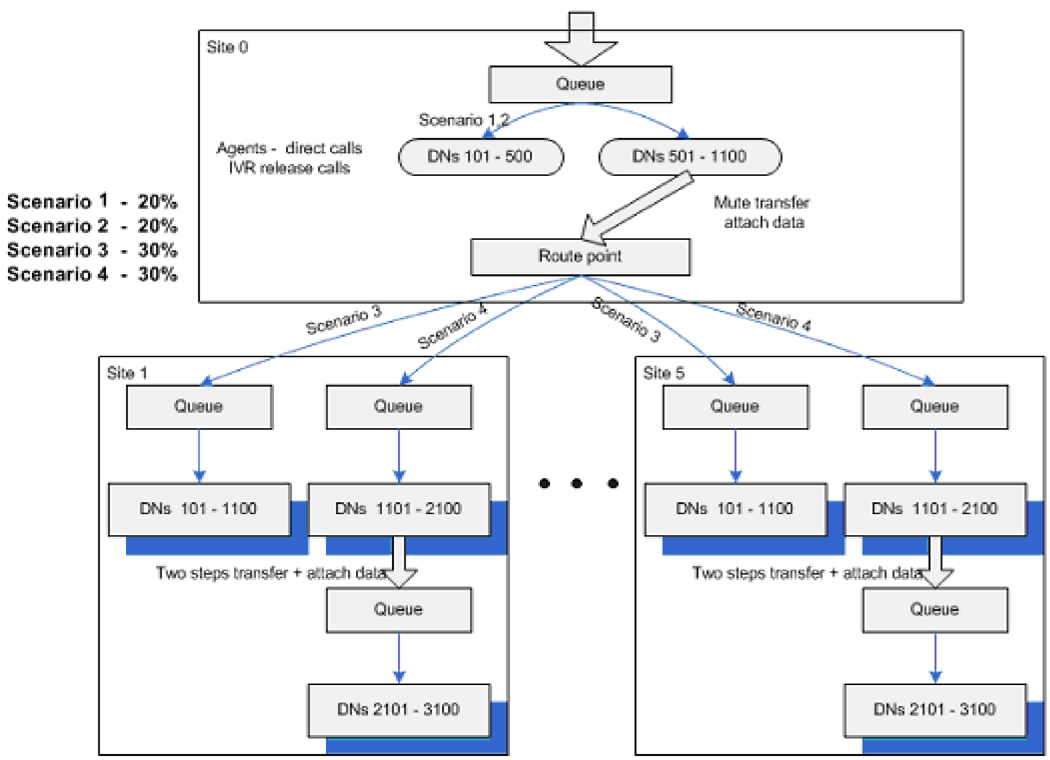


Figure 20: Sample Call Flow and Configuration (Voice Only)

The following four scenarios were tested with this configuration:

Scenario 1 Calls routed directly to an agent and released (20 percent of total calls)

Scenario 2 Calls routed to an IVR and released (20 percent of total calls)

Scenario 3 Calls routed to an IVR, transferred to a routing point, transferred to an agent, and then released (30 percent of total calls)

Scenario 4 Call routed to an IVR, transferred to a routing point, transferred to an agent, transferred and routed to a second agent, and then released (30 percent of total calls)

Tests and Results

The following tests were run in the voice-specific environment:

- ICON 7.5 and 7.6.0 Comparison Test (see page 89)
- Attached-Data Varied Test (see page 90)
- Call-Rate (Voice) Varied Test (see page 91)
- ICON Recovery (DB Server Disconnect) Test (see page 92)
- Data-Filtering Test (see page 93)

The details of these tests and the obtained results are discussed in the following sections.

ICON 7.5 and 7.6.0 Comparison Test

Genesys monitored ICON 7.5 and 7.6.0 over an extended period of time to ensure that no performance degradation occurred between releases. The test was conducted under the following conditions:

- Merge procedure run at 5-minute intervals
- Ten (10) predefined key-value pairs (KVPs) of data attached to every call.
- Five (5) days' endurance run
- Call rate of 14 calls/second (cps)

Test Results

Performance results are consistent between the 7.5 and 7.6.0 releases of ICON. Table 31 provides the observed results. Note that all values have been rounded up and are approximate.

Table 31: ICON 7.5 and 7.6.0 Comparison Test Results

Application	Average CPU (%)	Maximum CPU (%)	Maximum RAM (MB)	Average RAM (MB)
Solaris Operating System				
ICON 7.6.0	33	42	1,235	1,235
ICON 7.5	33	42	1,021	1,021

Table 31: ICON 7.5 and 7.6.0 Comparison Test Results (Continued)

Application	Average CPU (%)	Maximum CPU (%)	Maximum RAM (MB)	Average RAM (MB)
Windows Operating System				
ICON 7.6.0	22	~60-80	712	708
ICON 7.5	19	~70	600	600

Attached-Data Varied Test

In this test, Genesys varied the number of key-value pairs that were attached to each call to observe the effect on CPU and RAM utilization. The test was conducted under the following conditions:

- Number of attached KVPs varied from 10 to 50
- Average of 255 bytes of data attached to KVPs
- Merge procedure run at 5-minute intervals
- Duration of test: 2 hours
- Five (5) days of data prior to start

Test Results

Table 32 shows that increasing the number of KVPs had a linear affect on ICON storage. CPU utlization increased quickly with the number of KVPs, but had no effect on the merge time, which took an average time of 4 seconds. Note that all values have been rounded up and are approximate.

Table 32: Attached-Data Varied Test Results

# KVPs Attached/call	Average CPU (%)	Maximum CPU (%)	Maximum RAM (MB)	Average RAM (MB)	Merge Time (s)
10	33	40	887	880	3.6
30	40	48	1,017	1,002	4
50	46	58	998	984	4

Call-Rate (Voice) Varied Test

In this test, Genesys varied the call rate to observe the effect on CPU utilization and the merge time. The test was conducted under the following conditions:

- Call volume: Variable, starting at 7 calls/second, incrementing to 30 calls/second
- Merge procedure run at 5-minute intervals
- Duration of test: 2 hours
- Five (5) days of data prior to start

Test Results

Table 33 shows that the following results were produced, as the call rate increased:

- The average RAM utilization was fairly constant: approximately 890 MB for call rates of 10 and 14 calls/second, and a slightly higher 950 MB for call rates of 10, 15, 20, 25, 28, and 30 calls/second. Note that all values have been rounded up and are approximate.
- The duration of the merge procedure increased from 2 to 11 seconds (see Figure 21).

Table 33: Call Rate Varied Test Results

Call Rate (cps)	Average CPU (%)	Maximum CPU (%)	Maximum RAM (MB)	Average RAM (MB)
7	16	20	887	887
10	23	28	948	948
14	33	40	887	880
15	35	42	948	948
20	48	56	948	948
25	60	70	948	948
28	68	80	948	945
30	73	84	961	958

Figure 21 shows the linear relationship between the duration of the merge procedure and the call rate.

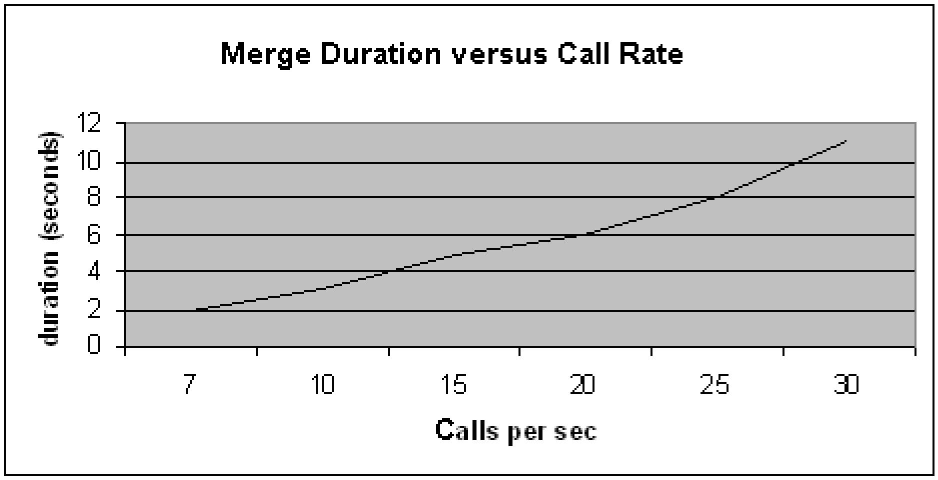


Figure 21: ICON Merge Duration Versus Call Rate

ICON Recovery (DB Server Disconnect) Test

Genesys has conducted a recovery test in which ICON was disconnected from DB Server for a period of time and then reconnected to determine how long it would take for ICON to recover (clear the persistent queue). The test was conducted under the following conditions:

- Call flow of 14 calls/second
- Merge procedure run at 5-minute intervals
- Duration of disconnection from DB Server: 0 minutes and 120 minutes

Test Results

After a disconnect from DB server of 2 hours, it took ICON 72 minutes to clear the persistent queue.

During the period of disconnect, testing produced the following results:

- The average CPU utilization increased to 50 percent.
- RAM utilization increased slightly from an average of 880 MB to 962 MB (see Table 34).
- Merge time increased to an average of 20 seconds (peak of 69 seconds). Typically, the merge time is about 4 seconds (see Figure 22).
- ICON took approximately 60 percent of the time that it was disconnected from DB Server to recover (clear the persistent queue, or PQ).

Note that all values in Table 34 have been rounded up and are approximate.

Table 34: ICON Recovery Test Results

DB Server Disconnect Time (minutes)	Time to Clear PQ (minutes)	Average CPU (%)	Maximum CPU (%)	Maximum RAM (MB)	Average RAM (MB)
0	N/A	33	40	887	880
120	72	50	64	962	962

Figure 22 shows the backlog of transactions in the persistent queue while ICON was disconnected from DB Server. After its reconnection to DB Server, ICON cleared the backlog within approximately 72 minutes.

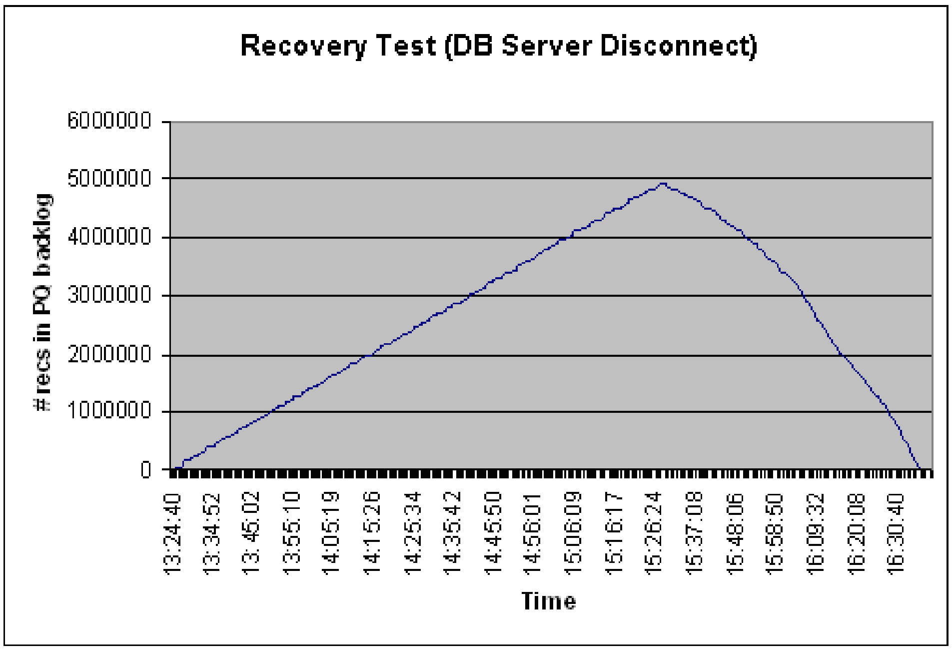


Figure 22: Records in ICON's Persistent Queue During and After DB Server Disconnect

Data-Filtering Test

In this test, the performance of ICON was measured when data filter options were set to filter out (exclude) data from IDB storage. The tests were conducted under the following conditions:

- Filters set to: none, three (call-metrics, call-history, ir-history), and all
- Merge procedure run at 5-minute intervals
- Call flow of 14 calls/second
- Duration of test: 60 minutes and 130 minutes

Note: For more information about ICON's data-filtering feature, see the *Interaction Concentrator 7.6 User's Guide.*

Test Results

There was no impact on CPU utilization when three data filters were applied versus none. When all of the data filters were applied, however, the maximum CPU utilization decreased by 25 percent, as shown in Table 35. Increasing the number of data filters also decreased the rate of data storage (see Figure 23).

Table 35: Data Filtering Test Results

Data Filters Set	Average CPU (%)	Maximum CPU (%)	Maximum RAM (MB)	Average RAM (MB)
None	33	40	887	880
Three (call-metrics, call-history, ir-history)	32	40	924	913
All	24	30	752	747

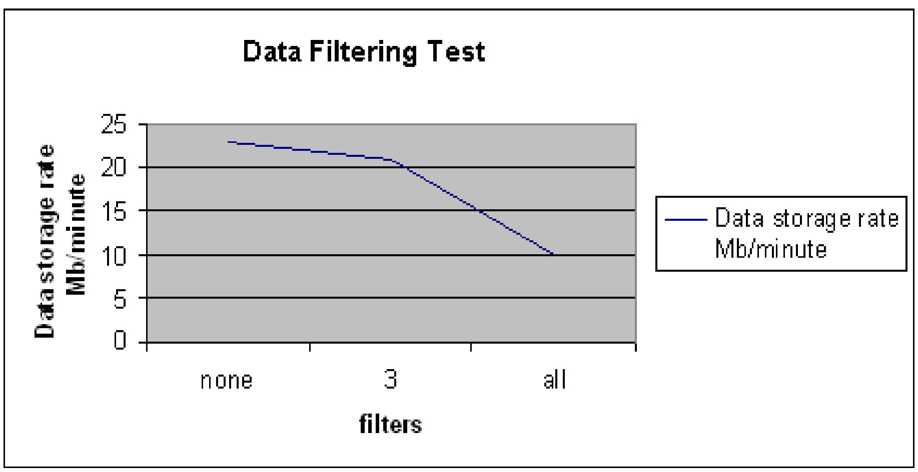


Figure 23: Rate of Data Storage Versus Number of Data Filters

Open Media Environment

In the open media test environment, a single ICON application—running on a Windows platform, and writing to an Oracle database on a Windows Server 2003—was configured to collect open media (e-mail) interactions from Interaction Server (see Figure 24).

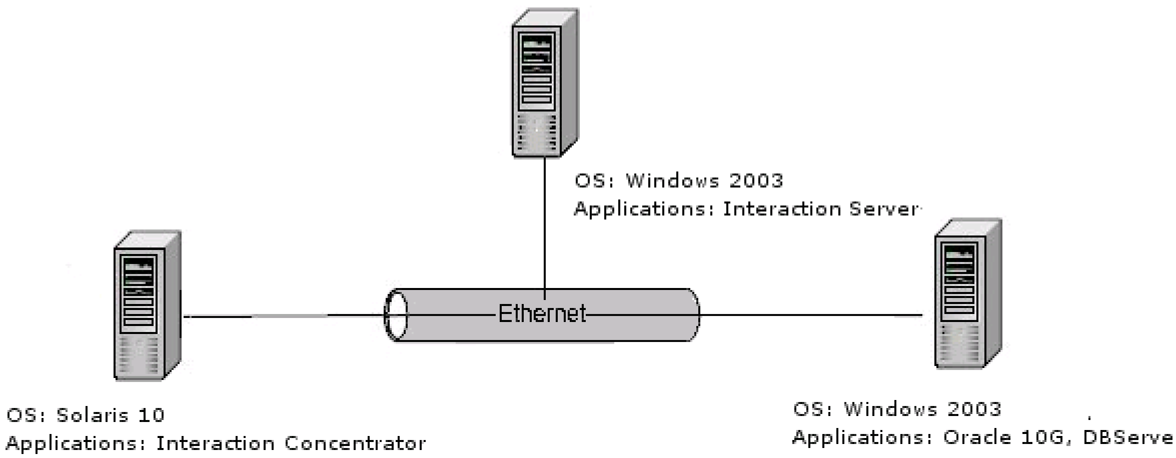


Figure 24: ICON 7.6.0 Open Media Test Environment

Hardware and Software

Table 36 describes the hardware that hosted the ICON application, persistent queue, and other components in the open media environment.

Table 36: Hardware and Software in ICON Open Media Environment

Application	Operating System	Processor	Memory	Software Version
ICON	Solaris	T5220 1.2 GHz 8 Core, 64 thread	32 GB RAM	7.6.000.16 7.6.100.00
RDBMS (Oracle 10)	Windows	Dual dual-core Xeon 5160 3.0 MHz	4 GB RAM	10.2.0.4.0
DB Server	Windows	Dual dual-core Xeon 5160 3.0 MHz	4 GB RAM	7.6.000.08
Stat Server	Windows	Dual dual-core Xeon 5160 3.0 MHz	4 GB RAM	7.6.000.18
URS	Windows	Dual dual-core Xeon 5160 3.0 MHz	4 GB RAM	7.6.001.06
Interaction Server and Proxies	Windows	Dual dual-core Xeon 5160 3.0 MHz	4 GB RAM	7.6.000.10

Call Flow and Configuration

In the open media model, the interaction flow is as follows:

3. The e-mail arrives at the first queue, where 10 KVPs of data are attached by the routing strategy.
4. The e-mail moves through the next two queues each of which has a strategy loaded onto it.
5. The e-mail arrives at an agent, where it is handled for 120 seconds.
6. The interaction moves through a final queue and routing strategy, before it is terminated.

Tests and Results

The following test was run in the open media environment:

- Rate of E-ails Varied Test (see page 96)

The following sections provide a detailed description of this test and a discussion of the obtained results.

Rate of E-Mails Varied Test

In this test, Genesys varied the number of e-mail interactions that were submitted during 2 hours of testing and with 1 day of previously processed data in IDB (approximately 1.2 million e-mails). The test was conducted under the following conditions:

- Data attached as KVPs to each e-mail at the first queue
- Forty-five hundred (4500) agents and four (4) queues configured
- One-hundred twenty (120) seconds handling time
- Rate of submitted e-mails: 7 e-mails/second and 14 e-mails/second

Test Results

Table 37 shows that, at a rate of 14 e-mails per second, CPU utilization averaged 93 percent and reached a maximum of 99 percent. As the rate of submitted e-mails increased, ICON's CPU utilization increased linearly (see Figure 25).

Table 37: Rate of E-Mails Varied Test Results

Submitted Rate (e-mails/sec)	Average CPU (%)	Maximum CPU (%)
7	50	53
14	92	99

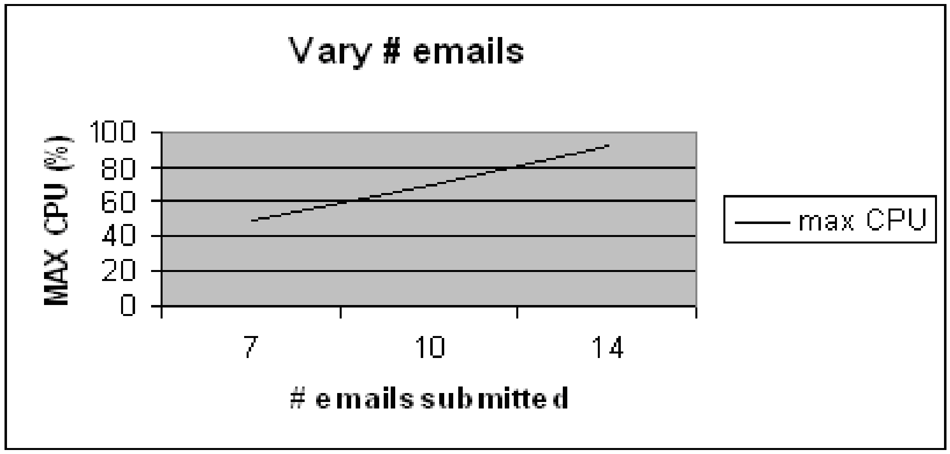


Figure 25: E-Mails Submitted Versus CPU Utilization

ICON 7.6.0 Performance Conclusions

Based on the results of performance testing that was conducted in a large-scale environment, Genesys has the following conclusions and recommendations:

- There is no degradation in performance between ICON releases 7.5 and 7.6.0 on either Solaris or Windows operating systems.
- The call rate and the number of key-value pairs (attached data) are significant factors that affect ICON’s scalability.
- ICON recovers quickly from DB Server disconnects—typically, recovering in a shorter time interval than the original outage.
- Filtering of data did not have a significant impact on RAM and CPU utilization, when a small number of filters (three) was set. The reduction was more significant— 43 percent reduction in RAM and 25 percent reduction in CPU utilization— when all filters were set.

Interaction Concentrator 7.5 Performance

Genesys has run several 7.5 performance tests for selected combinations of the following variables:

- Operating system (OS)—Windows (Microsoft Windows Server 2003 Enterprise), UNIX (Solaris 10)
- Relational database management system (RDBMS)—Microsoft SQL Server 2000, Oracle 10
- Call volume—starting at 10 calls/second, incrementing by 5 calls/second

The following sections describe the sample ICON 7.5 environment, call flows, tests, results, and recommendations.

Sample Environments

This subsection describes the Genesys sample 7.5 environments and conditions:

- Hardware specifications for the ICON server (see page 98)
- Hardware specifications for the DB Server, RDBMS server, and IDB (see page 99)
- Database settings (see page 99)
- Environment configuration and call flows (see page 100)

Hardware Specifications

Table 38 describes the hardware that hosted the ICON 7.5 application and persistent queue.

Table 38: Hardware Specifications—ICON 7.5 Application

Windows	
Operating System	Microsoft Windows Server 2003 Enterprise
Processor	2 x Intel Xeon 3.2 GHz/2 MB
Memory	4 GB RAM
Disk Storage	2 HDD x 80 GB SATA
UNIX	
Operating System	Solaris 10
Processor	2 x Sun UltraSPARC-IIIi 1.0 GHz/1 MB
Memory	2 GB RAM
Disk Storage	2 x 36GB Ultra160 SCSI drives

Table 39 describes the hardware that hosted the DB Server, RDBMS Server, and IDB.

Table 39: Hardware Specifications—DB Server, RDBMS Server, and IDB

Windows	
Operating System	Microsoft Windows Server 2003 Enterprise
RDBMS	Microsoft SQL Server 2000
Processor	2 x Intel Xeon 3.2 GHz/2 MB
Memory	4 GB RAM
Disk Storage	2 HDD x 80 GB SATA
UNIX	
Operating System	Solaris 10
RDBMS	Oracle 10
Processor	SUN SPARC Enterprise 9000, 16 GB memory, 4 x CPU quad core 2.8 GHZ [a]
Memory	8 GB Real RAM
Disk Storage	2 Gbit Sun StorEdge 3510 Fiber Channel Array

a. Each core supports four hardware threads of execution. The hardware threads are scheduled on the core's processing unit in round-robin order. A different software thread can run on each hardware thread, so 32 software threads can run in parallel on a single T1 processor.

Database Settings

Table 40 summarizes the Microsoft SQL database settings that were used in the Genesys sample 7.5 configurations.

Table 40: Microsoft SQL Database Settings

Setting	Value
Memory	2497 MB
Auto update statistics	On
Torn page detection	On

Table 40: Microsoft SQL Database Settings (Continued)

Setting	Value
Auto create statistics	On
Model	Simple

Table 41 summarizes the Oracle database settings that were used in the Genesys sample 7.5 configurations.

Table 41: Oracle Database Settings

Setting	Value
Memory	2,543 MB
Database block size (db_block_size)	8 KB
File system I/O options (filesystemio_options)*	directIO
*The Solaris file system was mounted with the option forcedirectio.	

Environment Configurations and Call Flows

The Genesys 7.5 performance results were obtained under the following conditions:

- Genesys used the following site configurations and call-flow scenarios:
  - Single site, with simple call flow (see page 101)
  - Single site, with complex call flow (see page 101)
  - Multi-site (single ICON, single IDB), with complex call flow (see page 103)

  The call flows include several Genesys-defined attached-data items, and eight user-defined attached data key-value pairs (approximately 200 bytes).
- The ICON Application used the default configuration settings.
- The logging level was set to Standard.
- All tests started with an empty database.
- In the tests that used the multi-site scenario with complex call flow, the merge procedure was run at 5-minute intervals.
- The ICON Performance Counter pages were queried frequently to obtain the reported measurements.

Single Site (Simple Call Flow)

Figure 26 shows the configuration and call flow for the simple call-flow scenario in a single-site deployment. The routing strategy has nine target queues, and each queue has 100 logged-in agents.

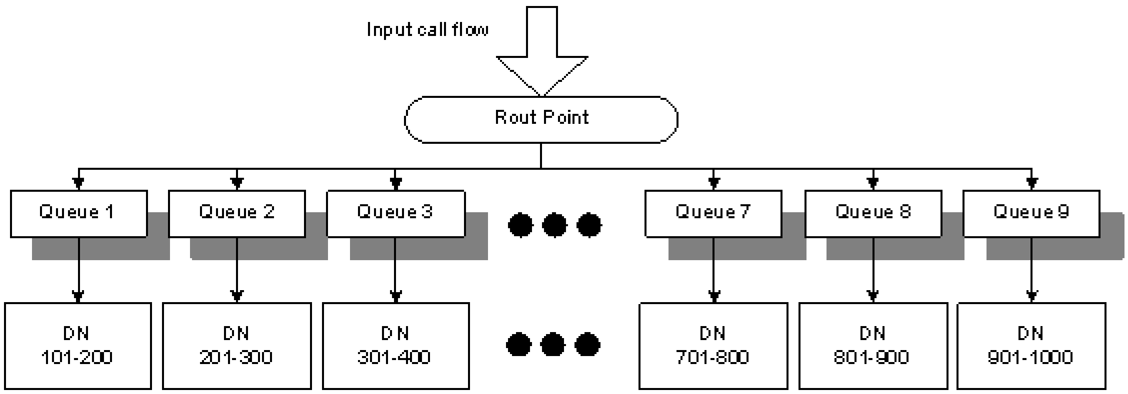


Figure 26: Single Site, Simple Call Flow

The call flow is as follows:

1. A new inbound call arrives at the Routing Point.
2. URS sends the call to one of the nine target queues.
3. The call is directed to one of the agents who are logged in to the queue.
4. The agent answers the call, handles the interaction, and then releases the call. The agent talk time is 40 seconds.

Single Site (Complex Call Flow)

Figure 27 shows the configuration and call flow for the complex call-flow scenario in a single-site deployment. There is an input queue, which has 400 logged-in DNs. The routing strategy has six target queues, and each queue has 400 logged-in DNs. There are 600 agents logged in to two additional queues.

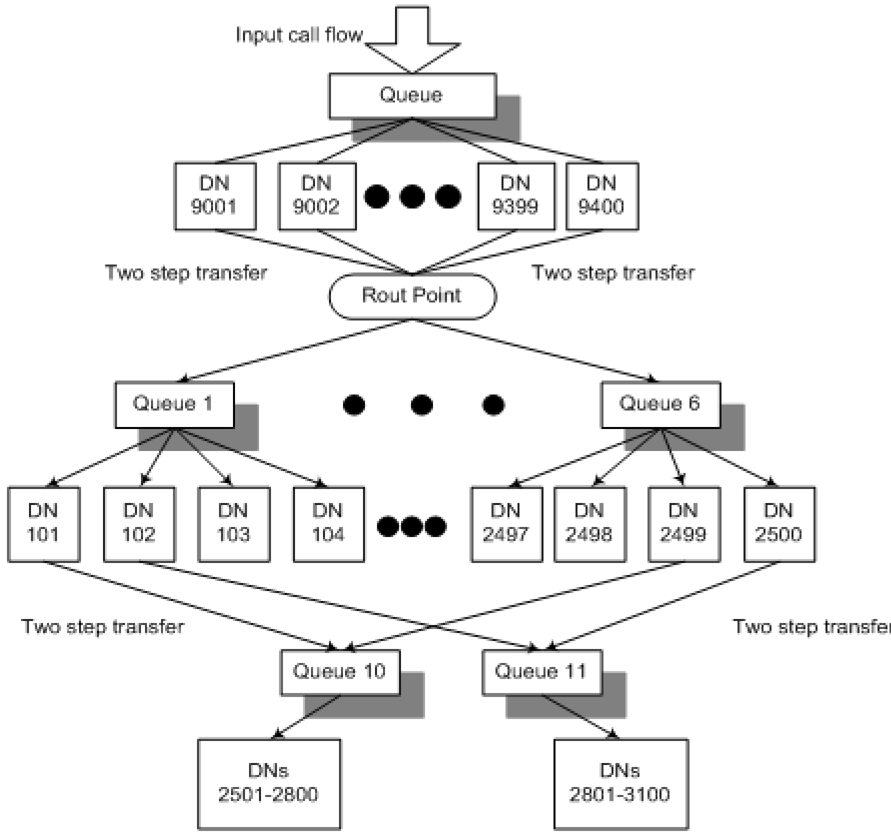


Figure 27: Single Site, Complex Call Flow

The call flow is as follows:

1. A new inbound call arrives at the input queue.
2. The call is directed to 1 of the 400 DNs that are logged in to the queue.
3. An agent simulator on these DNs attaches some business data and performs a two-step transfer to send the call to the Routing Point.
4. URS sends the call to one of the six target queues.
5. The call is directed to one of the agents who are logged in to the queue.
6. The call flow scenario splits:
   - For 90 percent of the agents logged in to the target queues, the agent answers the call, handles the interaction, and then releases the call. The agent talk time is 30 seconds.
   - For the remaining 10 percent of the agents logged in to the target queues, the agent answers the call, attaches some additional data, and then makes a two-step transfer to one of two additional queues, where there are a total of 600 logged in agents. The talk time of the agent in the last queue is 15 seconds, before the agent releases the call.

Multi-Site (Complex Call Flow)

Figure 28 shows the configuration and call flow for the complex call flow scenario in a multi-site deployment. There is an input queue, which has 900 logged-in DNs. The routing strategy has one target queue at each of five sites, and each queue has 900 logged-in DNs. Each site also has 1,100 agents logged in to each of two additional queues.

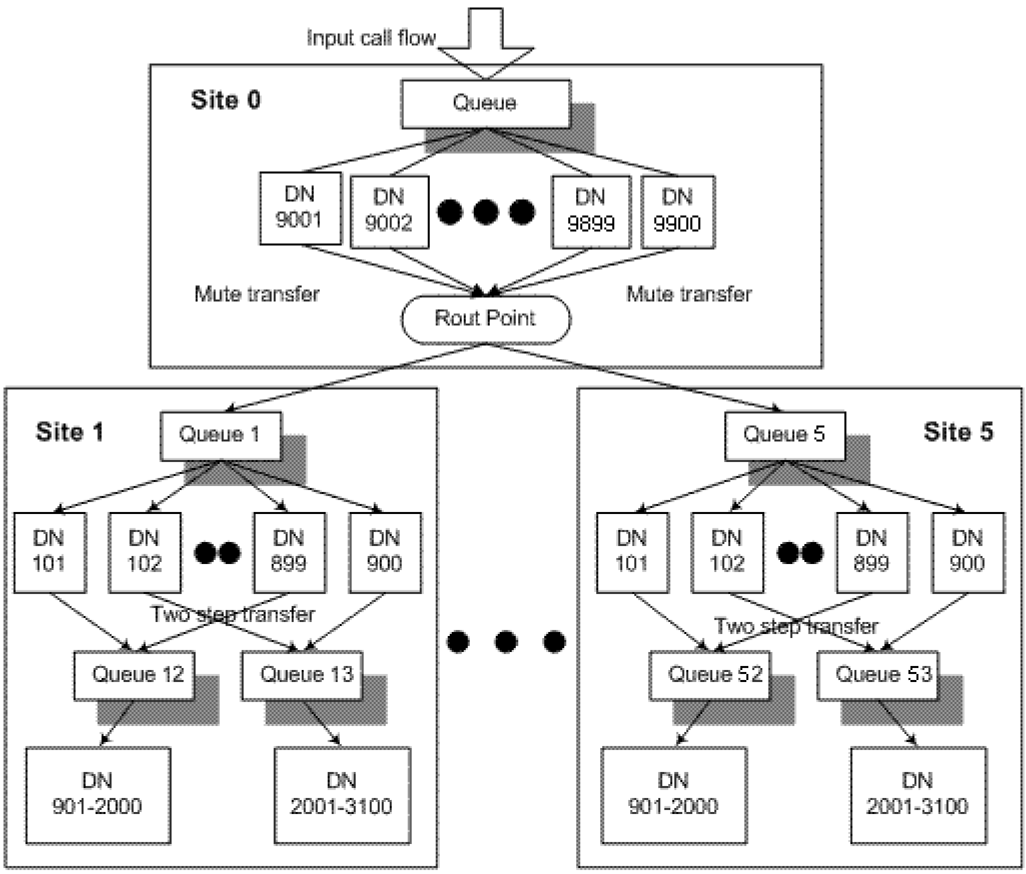


Figure 28: Multi-Site, Complex Call Flow

The call flow is as follows:

1. A new inbound call arrives at the input queue.
2. The call is directed to 1 of the 900 DNs that are logged in to the queue.
3. An agent simulator on these DNs attaches some business data and performs a mute transfer to send the call to the Routing Point.
4. URS sends the call to one of the five target queues.
5. The call is directed to one of the agents who are logged in to the queue.
6. The agent answers the call, attaches some additional data, and then makes a two-step transfer to one of two additional queues.
7. The last agent talk time is 40 seconds, before the agent releases the call.

ICON 7.5 Performance Conclusions

The Genesys performance results report Interaction Concentrator 7.5 runtime performance, related to the database-writing side of Interaction Concentrator activity. Call simulation and performance measurement started after ICON initialization was complete.

Table 42 summarizes the application processing usage and related recommendations for maximum call flows, on the basis of the performance results. The maximum call flows represent the average number of interactions per second that ICON can write to IDB without unsustainable persistent queue backlogs being generated.

For more information about the call-flow scenarios that produced these performance results, see "Environment Configurations and Call Flows" on page 100.

Table 42: Maximum Call Flows and Application Usage, by Scenario

Scenario	Maximum Call Flow (calls/sec)	Application Processing Usage		
		ICON CPU (%)	ICON Memory (MB)	RDBMS CPU (%)
OS: Windows Server 2003 Enterprise RDBMS: Microsoft SQL Server 2000				
Single-Site, Simple Call Flow	75	46	431	42.9
Single-Site, Complex Call Flow	60	57.9	421	70.5
Multi-Site, Complex Call Flow*	25	40.9	437	58.6
OS: Solaris 10 RDBMS: Oracle 10 * **				
Single-Site, Simple Call Flow	30**	38	725	4
Single-Site, Complex Call Flow	15**	40	757	4
Multi-Site, Complex Call Flow*	32**	42	854	52

*The merge procedure was run at 5-minute intervals.

**The hardware configuration of the database host had an adverse impact on ICON performance.

***Single side data was obtained using T2000 class of Servers while data for complex call flow was obtained using l M9000 class Server for DB hosting.

Endurance Test

In addition to the incremental call-flow performance tests, Genesys subjected Interaction Concentrator 7.5 to an endurance test under the following conditions:

- OS: Windows Server 2003 Enterprise
- RDBMS: Microsoft SQL Server 2000
- Call-flow scenario: Multi-Site (Complex Call Flow) (for more information, see page 103)
- Call speed: 25 calls/second
- Merge-procedure frequency: At 5-minute call intervals
- Duration of test: Approximately 72 hours

Endurance Test Results

Table 43 summarizes the endurance test results. All values have been rounded up and are approximate.

Table 43: Endurance Test Results

Measurement	Value
ICON CPU usage (%)	35
ICON memory usage (MB)	612
RDBMS CPU usage (%)	52
RDBMS memory usage (MB)	1,750
DB Server multiserver CPU usage (%)	0.5
DB Server multiserver memory usage (MB)	12
DB Server client 1 CPU usage (%)	9
DB Server client 1 memory usage (MB)	50
DB Server client 2 CPU usage (%)	6
DB Server client 2 memory usage (MB)	32
Merge time (sec)	30–40
Total number of calls	6,210,000
Total number of ICON transactions	372,887,000

Recovery Test

In addition to the incremental call-flow performance tests, Genesys subjected Interaction Concentrator 7.5 to a recovery test, following disconnection from DB Server. The test was conducted under the following conditions:

- OS: Windows Server 2003 Enterprise.
- RDBMS: Microsoft SQL Server 2000.
- Call flow scenario: Multi-Site (Complex Call Flow) (for more information, see page 103).
- Call speed: 10 calls/second. Call flow was not stopped when DB Server reconnected.
- Duration of disconnection from DB Server: 15 minutes.

Recovery Test Results

While ICON was disconnected from DB Server, transactions backlogged in the persistent queue. After reconnection to DB Server, ICON cleared the backlog within approximately 17 minutes.

On the basis of the recovery test results, Genesys suggests that the recovery time can be estimated as follows:

Recovery time = Time_without_DBServer * Input_call_flow / (Max_call_flow – Input_call_flow)

Chapter

7 Genesys Info Mart 8.x Solution

The information in this chapter applies to the 8.x releases of Genesys Info Mart.

This chapter describes the hardware architecture for Genesys Info Mart components, providing an example of the architecture for a multi-site deployment. It also describes the factors that affect Genesys Info Mart 8.0 and 8.1 performance, and lists sample performance measurements for reference platforms in large-scale, multi-site deployments that included aggregation.

Note: Before proceeding, review the "Architecture" section in the *Genesys Info Mart Deployment Guide* for your release of Genesys Info Mart to familiarize yourself with the product architecture.

This chapter contains the following sections:

- About Genesys Info Mart 8.x, page 108
- Hardware Architectures in Release 8.x, page 108
- Genesys Info Mart 8.1.2 Performance, page 109
- Release 8.1.2 Endurance Test, page 128
- Release 8.1.2 Recovery Test—Two-Hour Outage, page 145
- Release 8.1.2 Recovery Test—Ten-Hour Outage, page 148
- Genesys Info Mart 8.0 Performance, page 153
- Release 8.0 Baseline Performance Test, page 157
- Release 8.0 Benchmark Performance, page 171
- Genesys Info Mart 8.x Performance Tuning, page 184
- Genesys Info Mart 8.x Database Size Estimation, page 186

For related information about Reporting and Analytics Aggregates (RAA) 8.x reporting performance, see Chapter 8 on page 189.

For related information about Genesys Interactive Insights (GI2) 8.x summary reporting performance, see Chapter 9 on page 201.

About Genesys Info Mart 8.x

Genesys Info Mart produces a data mart containing several star schemas that you can use for contact center historical reporting.

Genesys Info Mart includes a server component, administration graphical user interface (GUI), and database. The Genesys Info Mart Server runs a set of predefined jobs to extract, transform, and load (ETL) data that has been gathered by Interaction Concentrator (ICON) from data sources such as Configuration Server, T-Server, Interaction Server, and Outbound Contact Server.

Genesys Info Mart initially stores the low-level interaction data, which is consolidated from Interaction Concentrator databases (Interaction Databases [IDBs]), in the Global Interaction Database (GIDB) tables in the Info Mart database. Genesys Info Mart then transforms the low-level interaction data and loads it into the dimensional model (or star schemas) in the Info Mart database.

In deployments that include GI2 or the separately installed Reporting and Analytics Aggregates (RAA) package, Genesys Info Mart also hosts an aggregation engine that aggregates or re-aggregates the data and populates aggregate tables in the Info Mart database.

Hardware Architectures in Release 8.x

The Genesys Info Mart architecture is flexible and scalable. However, there are certain rules and requirements that must be observed. For detailed information about the hardware architectures that Genesys Info Mart 8.x supports, see the chapter about supported topologies in the *Genesys Info Mart 8.x Deployment Guide,* which also includes hosting recommendations.

Figure 29 on page 109 depicts a generic architecture that embodies the rules and requirements for Genesys Info Mart. For detailed information about the hardware architectures that were used in the Genesys Info Mart 8.x performance tests, see “Release 8.1.2 Performance Test Setup” on page 111 and “Release 8.0 Performance Testing Configuration” on page 154.

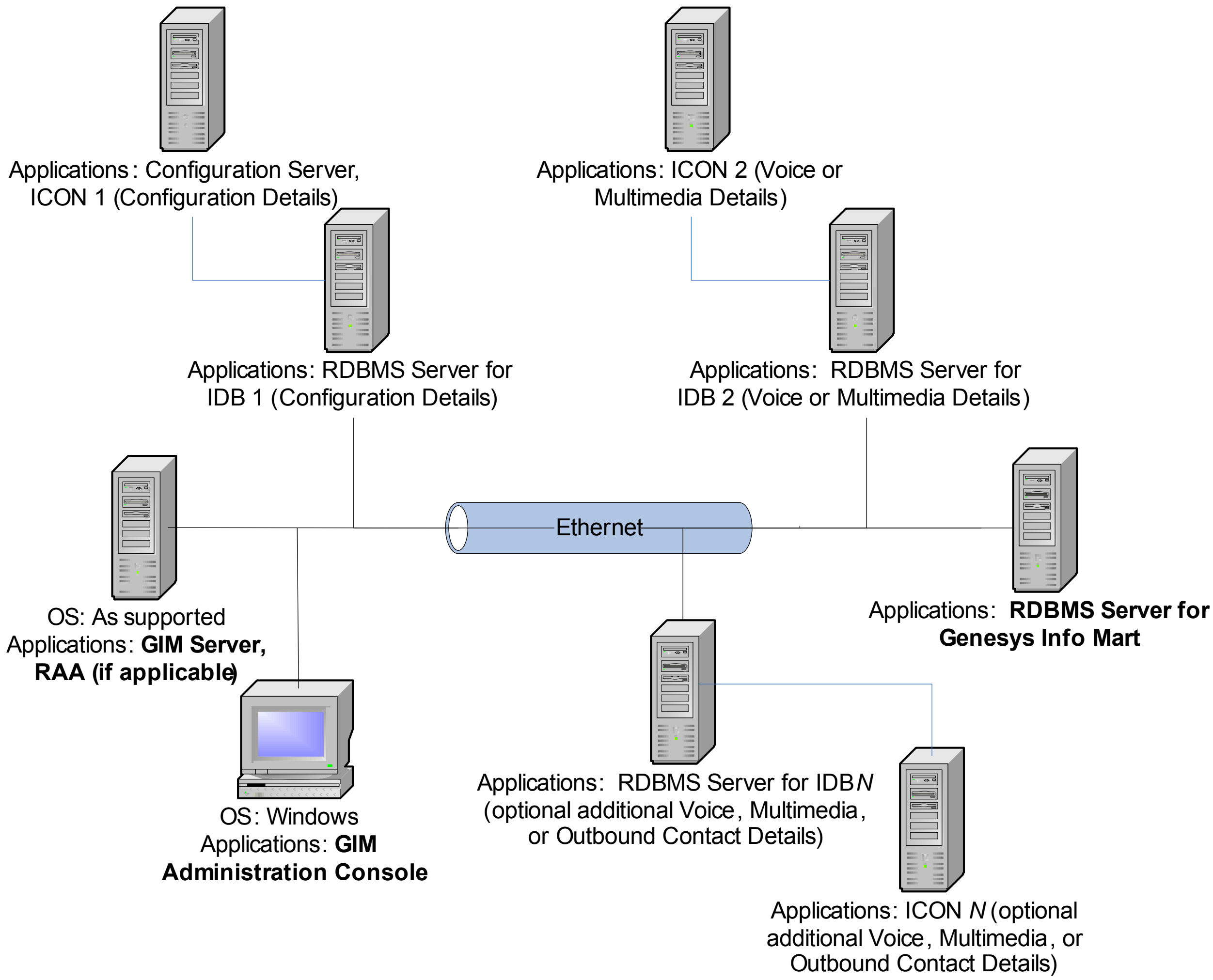


Figure 29: Generic Genesys Info Mart 8.x Hardware Architecture

Genesys Info Mart 8.1.2 Performance

This section provides general background about the Genesys Info Mart 8.1.2 performance testing and test setup.

About the Genesys Info Mart 8.1.2 Performance Testing

Objectives

The purpose of the Genesys Info Mart 8.1.2 performance testing was to validate the ability of Genesys Info Mart to support very high volumes of data in a “blended” environment, which included inbound voice, Outbound Contact voice, and multimedia (e-mail and chat) interactions. Genesys Info Mart was deployed on a Linux platform. The Info Mart database and IDBs were deployed on a four-node Oracle 11g Release 2 (R2) Real Application Clusters (RAC) database platform, also on Linux.

The ETL cycle and aggregation ran throughout the day to process actively populated source data, which was continuously generated for a high-complexity call flow. For details about the call flows, see "Release 8.1.2 Performance Test Setup" on page 111.

Tests Performed

The Genesys Info Mart 8.1.2 performance testing consisted of:

- An endurance test in which Genesys Info Mart ran for 14 consecutive days at a peak rate of 19 million interactions a day, or 220 interactions per second.

  For detailed results, see "Endurance Test Results" on page 128.
- A recovery test in which ETL processing was stopped for two hours while call generation continued at peak rates. When the ETL resumed, Genesys Info Mart was able to recover from the outage—in other words, process the backlog of data and catch up to regular, steady-state processing—in approximately one hour.

  For detailed results, see "Two-Hour Recovery Test Results" on page 146.
- A recovery test in which ETL processing was stopped for ten hours while call generation continued at peak rates. When the ETL resumed, Genesys Info Mart was able to recover from the outage in approximately five hours.

  For detailed results, see "Ten-Hour Recovery Test Results" on page 148.

Test Applications

The test environment included:

- Interaction Concentrator 8.1.1, which introduced support for partitioned IDBs and a streamlined, partition-aware purge procedure
- RAA 8.1.101, which introduced an improved aggregate dispatcher that enables notifications to be prioritized, so that continual re-aggregation of long-living multimedia interactions does not delay processing of more recent interactions
- Genesys Info Mart 8.1.2, which introduced additional performance improvements

For full details about the applications in the test environment, see “Product Versions” on page 118.

Note: Genesys Info Mart 8.1 performance testing was combined with testing of RAA release 8.1.1. Test results for Genesys Info Mart 8.1.2 report on activity that included aggregation.

For detailed information about the associated test results for RAA release 8.1.1 performance, see “RAA Release 8.1 Performance” on page 189.

Release 8.1.2 Performance Test Setup

The Genesys Info Mart 8.1 testing was designed to evaluate the scalability of the historical reporting solution.

Interaction Volumes

Voice and multimedia interactions were generated nonstop, for 24 hours a day for the entire duration of the extended test period, at a flat rate of 220 interactions per second. Table 44 shows the breakdown by media type.

Table 44: Daily Interaction Volume, by Media Type

Media Type	Number of Interactions	
	Per Day	Per Second
Voice	15,001,058	173.6
Voice (Outbound Contact)	3,116,883	36.1
E-mail	713,960	8.3
Chat	236,152	2.7
Total	**19,068,053**	**220**

Configuration History Environment

The environment for recording Configuration details consisted of the Configuration Server for the entire contact center, monitored by one Interaction Concentrator instance. (An *Interaction Concentrator instance* consists of an ICON server and the IDB it populates.) The contact-center configuration did not change during the test. The Configuration details IDB was not purged.

Inbound Voice Environment

Although many of the T-Server and other applications resided on the same hosts in the test laboratory, the setup for inbound voice calls modelled a highly distributed environment consisting of eight “sites,” with each site having one T-Server for self-service IVR calls (“IVR T-Server”) and three T-Servers for agent-assisted calls (“Agent T-Servers”). On each site, the IVR T-Server was monitored by one Interaction Concentrator instance, and the three agent T-Servers were monitored by a second Interaction Concentrator instance. Altogether, there were 32 T-Servers, monitored by 16 ICONs for Voice details (in other words, with role=gcc,gud,gls), with each ICON populating its own IDB.

To handle the calls, the environment provided a combined total of 40,000 IVR ports and agents. Agents logged in and out in eight-hour shifts.

For details about the hardware setup, see Figure 30 on page 115.

Call Topologies

All the calls were initially delivered to an IVR, following which:

- 75 percent were completed in the IVR.
- 25 percent were transferred to an agent on another switch in the same site. Of these:
  - 50 percent were transferred by the agent to an IVR on another switch in the same site, and the call was completed in that IVR.
  - 50 percent were transferred by the agent to a second agent on another switch in the same site. The second agent then transferred the call to an IVR on another switch in the same site, and the call was completed in that IVR.

  The average agent talk time was approximately 350 seconds.

User Data

User-data processing was significant.

- In 90 percent of calls, the first IVR attached 18 key-value pairs (KVPs), including 5 high-cardinality KVPs (10 million values), which were mapped to custom user-data facts in the Info Mart database, and 7 low-cardinality KVPs (ranging from 2 to 100 possible values), which were mapped to custom user-data dimensions.
- In 10 percent of calls, the first IVR attached 180 KVPs, including 144 high-cardinality KVPs (10 million values), which were mapped to custom user-data facts, and 30 low-cardinality KVPs (ranging from 2 to 100 possible values), which were mapped to custom user-data dimensions.

To improve performance, storage of KVP data in IDB was split between the G_USERDATA_HISTORY table and the G_SECURE_USERDATA_HISTORY table. Splitting data storage improves performance in two ways: for Interaction Concentrator, database operations against two separate, smaller tables consume fewer resources; for Genesys Info Mart, extraction performance is improved because the extraction job can extract separate tables in parallel.

Outbound Contact Voice Environment

The setup for Outbound Contact voice calls modelled a distributed environment consisting of three "sites," with each site having one Outbound Contact Server (OCS), which ran a single campaign to a single campaign group, and one T-Server for the interaction side of the calls. Each OCS was monitored by one Interaction Concentrator instance for Outbound Contact details (in other words, with ICON role=gos), and all three T-Servers were monitored by a single Interaction Concentrator instance for Voice details (ICON role=gcc,gud,gls).

A total of 1800 agents handled the calls.

For details about the hardware setup, see Figure 31 on page 116.

Call Topology

In all cases, the outbound call was transferred to an agent, who handled the call and then released it. There were no transfers to other agents or an IVR.

Record Field Data

In addition to the mandatory OCS record fields that are listed in the *Genesys Info Mart 8.1 Deployment Guide,* 20 custom record fields were stored in IDB and extracted by Genesys Info Mart.

Storage of record field data in IDB was split between the GO_FIELDHIST table and the GO_SEC_FIELDHIST table.

Multimedia Environment

The environment for multimedia interactions consisted of one "site," with one Interaction Server monitored by one Interaction Concentrator instance for Multimedia details (in other words, with ICON role=gcc,gud,gls).

The environment supported a combined daily total of 950,000 e-mail and chat interactions, which were handled by 7000 agents.

For details about the hardware setup, see Figure 32 on page 117.

Call Topologies

Interactions were handled by the agent to whom they were routed. There were no agent transfers or collaborations.

- 25 percent of the interactions were routed through one Interaction Queue to an agent, who completed the interaction on the same day.

- 25 percent of the interactions were routed through two Interaction Queues to an agent, who completed the interaction in two days.
- 25 percent of the interactions were routed through three Interaction Queues to an agent, who completed the interaction in three days.
- 25 percent of the interactions were routed through three Interaction Queues to an agent, who completed the interaction in four days.

User Data

User-data processing was not significant. There was no custom user data, and the volume of default attached data for multimedia interactions was small.

Hardware Architecture

Figures 30 through 32 show the architecture and specifications for the hardware that hosted Interaction Concentrator, Genesys Info Mart, and RAA software components. Because of the size and complexity of the test environment, there are separate diagrams for the different media domains.

All the IDBs and the Info Mart database were deployed on a four-node Oracle RAC platform. The Oracle Cluster nodes connected with each other and with the storage area network (SAN) devices over 4 Gbps Fiber Switch. The Genesys application servers connected with each other and with the Oracle Cluster nodes over 1 Gbps Ethernet. There were no WAN connections in the configuration.

Voice Figure 30 on page 115 shows the architecture and hardware specifications for the Inbound Voice environment described on page 112. For simplicity, the diagram does not show the Configuration details ICON, which was installed on the Voice 3 host and which populated a Configuration details IDB on Node 1 of the Oracle RAC.

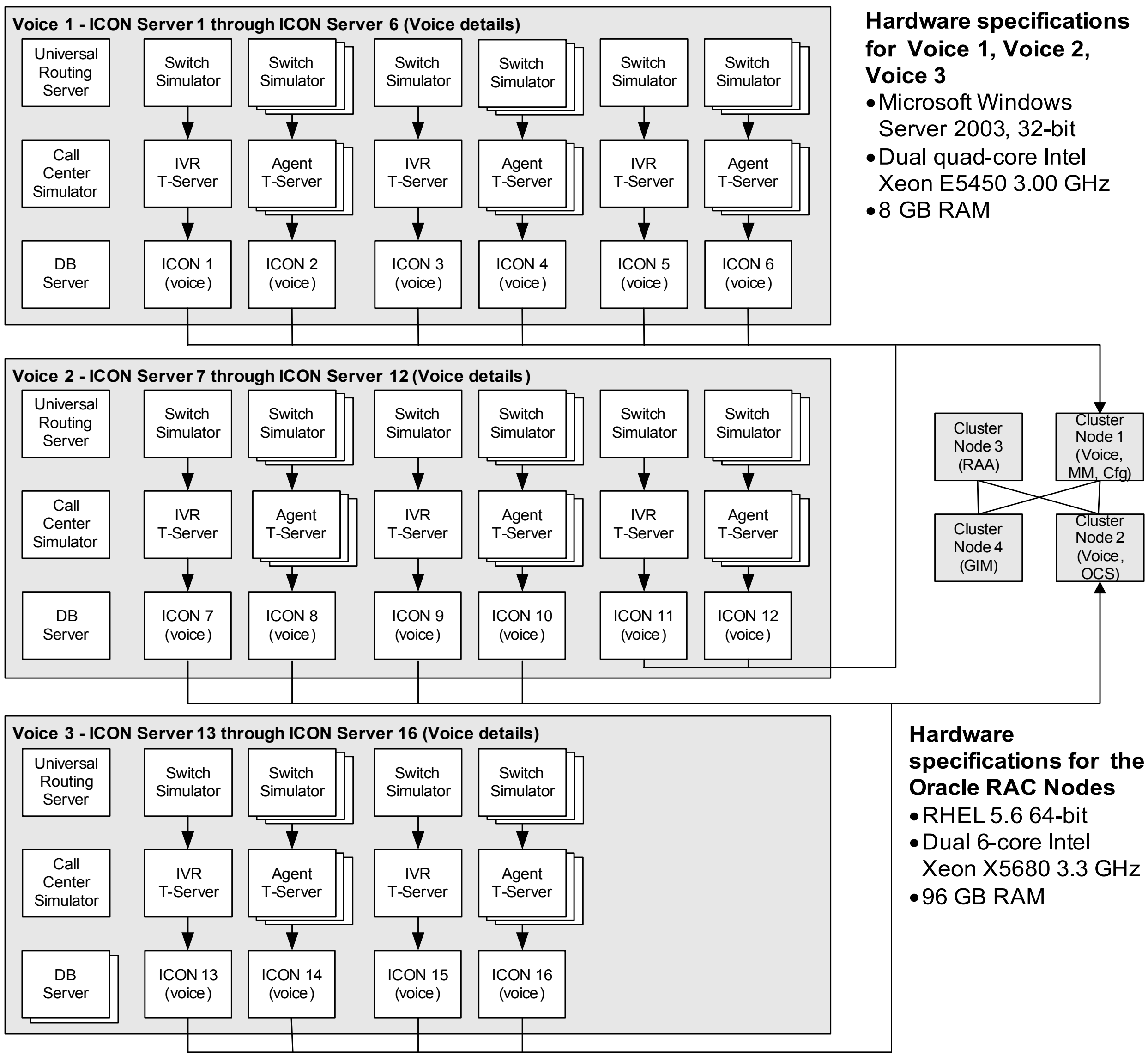
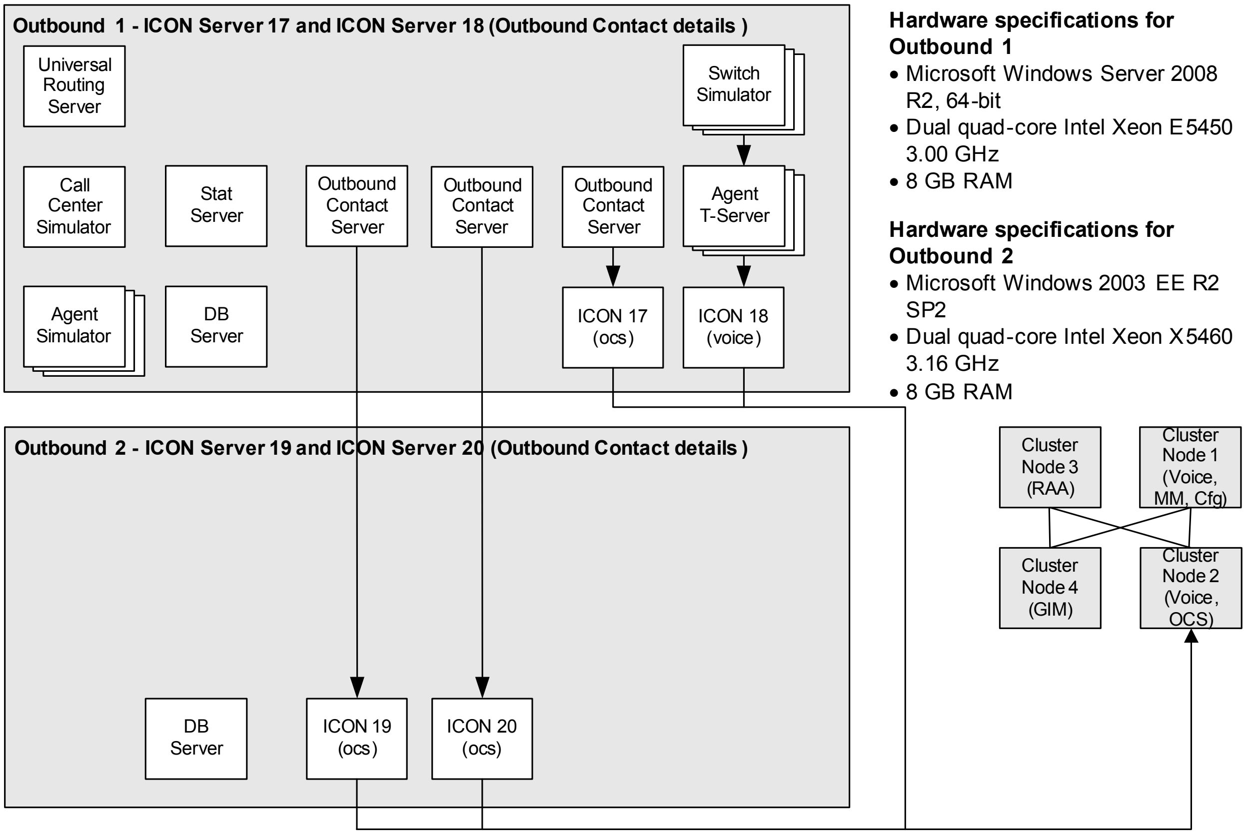


Figure 30: Hardware Architecture—Voice

Outbound Contact

Figure 31 on page 116 shows the architecture and hardware specifications for the Outbound Contact environment described on page 113. For the hardware specifications for the Oracle RAC nodes, see Figure 30.



Hardware specifications for Outbound 1
- Microsoft Windows Server 2008 R2, 64-bit
- Dual quad-core Intel Xeon E5450 3.00 GHz
- 8 GB RAM

Hardware specifications for Outbound 2
- Microsoft Windows 2003 EE R2 SP2
- Dual quad-core Intel Xeon X5460 3.16 GHz
- 8 GB RAM

Figure 31: Hardware Architecture—Outbound Contact

Multimedia

Figure 32 on page 117 shows the architecture and specifications for the Multimedia environment described on page 113. The Multimedia details ICON resided on a server that also hosted the Genesys Info Mart Server and RAA applications. For the hardware specifications for the Oracle RAC nodes, see Figure 30 on page 115.

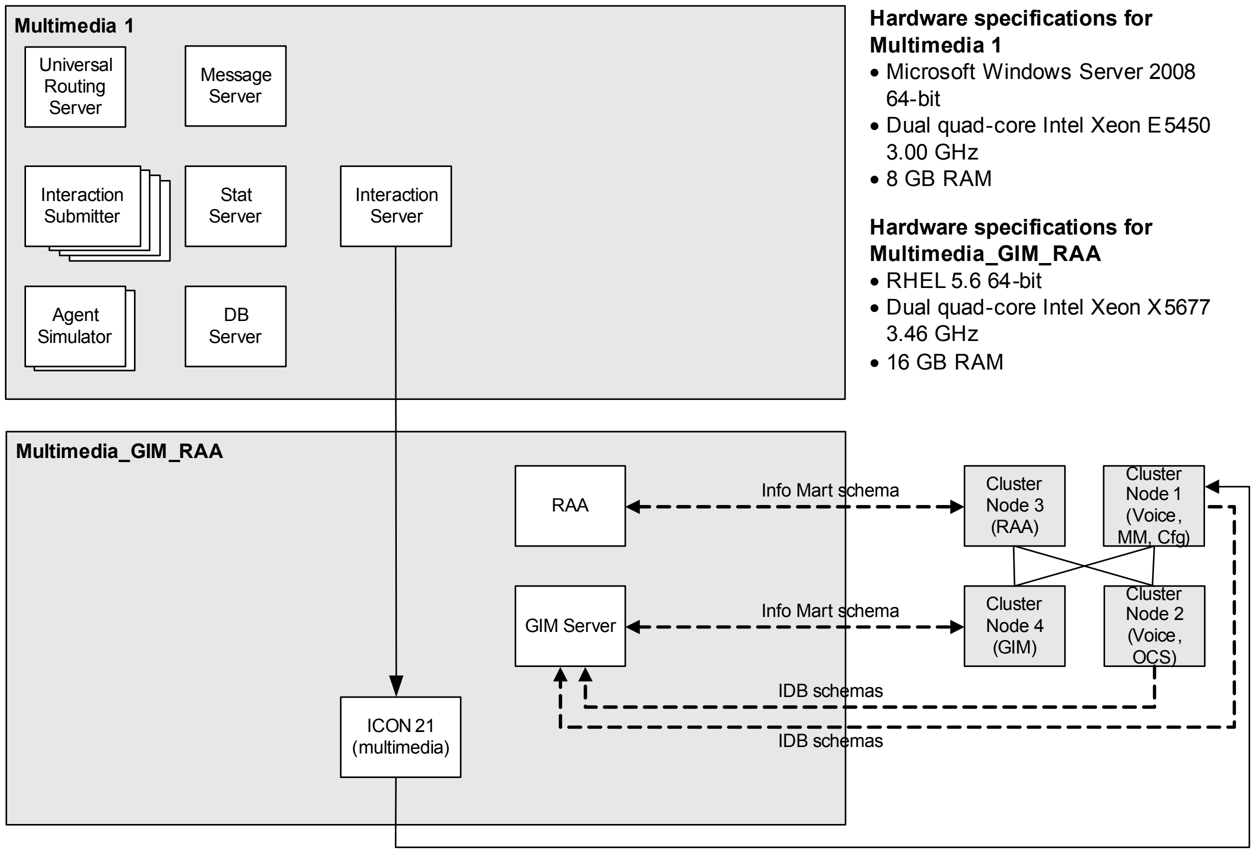


Figure 32: Hardware Architecture—Multimedia, Genesys Info Mart Server, and RAA

Database Storage

Two network-attached storage devices were integrated with the database through Oracle Automatic Storage Management (ASM):

- EMC CX4-120, with a total of 60 drives, each 300 GB:
    - 4 disk shelves
    - 15 x 300 GB 15K RPM drives per shelf

  Ten RAID 0 arrays used all 15 drives on a shelf, so each array provided 1.6 TB of storage. The disk arrays were mapped to 10 Oracle ASM disks and collected in the Oracle disk group named ORADATA, which provided a total of 17 TB of storage.

  The ORADATA disk group contained all the IDB, Info Mart, system, and SYSAUX tablespaces. All tablespace datafiles were striped across all disks in the subgroup. All application datafiles used the same DATAFILE ASM template, with no redundancy and with coarse striping.
- Sun StorEdge 6140:
    - 1 disk shelf
    - 16 x 146 GB 15K RPM drives

One RAID 0 array used all 16 drives, providing 1.7 TB of storage. The disk array was mapped to an Oracle ASM disk and collected in the Oracle disk group named ORAREDO.

The ORAREDO disk group contained the UNDO tablespace (8 x 60 GB datafiles) and Online Redo logs (8 x 50 GB files, 2 per node). All UNDO datafiles used the same DATAFILE ASM template and all the log files used the same ONLINELOG ASM template, with no redundancy and with coarse striping.

The theoretical maximum data transfer rate of each array was 400 MB per second.

The four database servers in the cluster were connected to the storage devices through a single Brocade 4 Gb fiber switch.

Product Versions

Table 45 lists the versions of Genesys Info Mart, RAA, and other supporting software products that were used in the testing.

Table 45: Product Versions Tested

Software	Version
Genesys Products	
Interaction Concentrator	8.1.100.18 **Note:** The performance testing used a prerelease version of the software.
Genesys Info Mart	8.1.200.14 **Note:** The performance testing used a prerelease version of the software. The testing did not cover certain capabilities that were included in the initial general release of Genesys Info Mart 8.1.2.
Reporting and Analytics Aggregates (RAA)	8.1.101.00 **Note:** The performance testing used a prerelease version of the software.
T-Server for Avaya Communication Manager	8.0.101.03
Interaction Server	8.0.200.07
Outbound Contact Server	8.1.000.12
Configuration Server	8.1.000.16

Table 45: Product Versions Tested (Continued)

Software	Version
Miscellaneous additional Genesys products, such as DB Server, Message Server, and Universal Routing Server	Various 8.x releases
Third-Party Products	
Java Development Kit (JDK) on the Genesys Info Mart application host	1.7 Build 10
Oracle	Oracle Database 11g Enterprise Edition Release 11.2.0.3.0 for 64-bit with Real Application Clusters (RAC), Partitioning and Automatic Storage Management (ASM) options

Oracle RAC Configuration

Based on previous experience with Oracle RAC, Genesys determined that it was important for Genesys Info Mart performance to separate processing by function onto specific nodes in the cluster, so that, as far as possible, data could be kept local to the respective processes. Accordingly:

- Separate named services were set up for each function—GIM, RAA, and two ICON services—and each service was directed to a specific node in the cluster. For more information, see "IDBs" and "DAP Objects for ICON", starting on page 123, and "Info Mart Database" and "DAP Objects for Genesys Info Mart", starting on page 126.
- The Oracle parameter PARALLEL_FORCE_LOCAL was set to TRUE. With this setting, the parallel parts of a query execute only on the node on which the query is submitted; they are not distributed across all nodes in the cluster.
- The Oracle RAC was not configured in failover or connection load-balancing modes.

Database links were not configured between the RDBMS servers that hosted the Info Mart and IDB database schemas. Early experience during the testing showed reduced performance with database links, because of extensive node synchronization.

Table 46 lists non-default Oracle settings that were used for the cluster.

Table 46: Non-Default Oracle Settings

Parameter	Value
audit_file_dest	/u01/app/oracle/admin/genrac/adump
audit_trail	NONE
cluster_database	TRUE
compatible	11.2.0.0.0
control_files	ORADATA/genrac/controlfile/current.256. 800535439
cursor_sharing	EXACT (default value)[a]
db_block_size	16384
db_create_file_dest	+ORADATA
db_domain	testlab.genesys.ca
db_name	genrac
diagnostic_dest	/u01/app/oracle
instance_number	1
open_cursors	4000
pga_aggregate_target	5000M
processes	2048
recyclebin	off
remote_listener	gen-scan.testlab.genesys.ca:1521
remote_login_passwordfile	EXCLUSIVE
session_cached_cursors	4000
sessions	3096
sga_target	75008M
spfile	+ORADATA/genrac/spfilegenrac.ora
thread	1
undo_tablespace	UNDO1
parallel_force_local	TRUE

a. Preliminary testing used cursor_sharing=SIMILAR, but ETL performance was significantly worse with this value. For all the test results reported in this document, cursor_sharing was set to EXACT.

To reduce the overhead of managing sequences, which are heavily used by ICON, the sequence cache sizes for high-volume IDB tables were increased from the default of 20 to 1000.

The Oracle-defined program to automatically optimize statistics collection, SYS.GATHER_STATS_PROG, was used to collect statistics for the IDB and Info Mart database schemas during the daily maintenance window. The default Oracle settings for Optimizer Statistics were used.

Archive logging was not used, because of limited storage.

Interaction Concentrator–Related Configuration

ICON Applications

Table 47 on page 121 lists important ICON release 8.1.1 configuration options for which non-default values were used.

Non-default settings for the acc-* options optimized transaction sizes in the persistent queue, and the [filter-data] options reduced the volume of unneeded data in IDB. For the Multimedia details ICON, non-default settings for the om-memory-* options in the ICON application and Script objects reduced the amount of ICON memory used.

For the Voice details ICONs, in addition to the options listed in Table 47, the attached data specification file (adata_spec.xml) was modified to enable ICON to store both predefined and custom attached data, as described in “User Data” on page 112.

Table 47: Non-Default ICON 8.1.1 Application Settings

Option	Value			
	Voice ICONs	Outbound Contact ICONs	Multimedia ICON	Default Value
callconcentrator Section				
acc-queue-lifespan	10	5	5	5
acc-queue-size	5000	5000	500	500
calls-in-the-past*	true	true	true	false
dest-busy-processing	true	false	false	false
dss-no-data-tout	60	60	60	300

Table 47: Non-Default ICON 8.1.1 Application Settings (Continued)

Option	Value			
	Voice ICONs	Outbound Contact ICONs	Multimedia ICON	Default Value
extended-route-result	1	0	1	0
gls-active-reason-codes*	true	true	true	false
gls-acw-first	true	false	false	false
om-force-adata*	true	true	true	false
om-memory-optimization			true	false
partition-type*	2	2	2	0
role*	gcc,gud,gls	gos	gcc,gud,gls	all
store-releasing-party	true	false	false	false
store-route-result-reliability	0	0	1	0
use-dss-monitor*	true	true	true	false
*Mandatory option—Genesys Info Mart will not function if the option value is not set as specified.				
dbw-error-reactions Section				
dbw-error1	error=Duplicate;reaction=ignore			
filter-data Section				
acd-party-metrics	1	0	1	0
call-history	0	0	1	0
gls-ivr	1	0	1	0
gls-metrics	1	0	1	0
ir-history	1	0	1	0
observer-party	1	0	1	0
udata-history-terminated	1	0	1	0
ICON-Related Options on Script (Interaction Queue) Objects—callconcentrator Section				
om-memory-clean			1	0

Field Objects—Outbound Contact Records

In the [default] section on the Annex tab of Field objects, the icon_attribute option was set to 1 or 2, so that some record field data was stored in the GO_FIELDHIST table and some in the GO_SEC_FIELDHIST table in IDB.

IDBs

Except for the Configuration details IDB, all the IDBs in the deployment were partitioned, to enable efficient, partition-aware purging in parallel with other processing.

The IDBs were deployed on two nodes of the Oracle RAC cluster. Two named services were defined to share the function of ICON operations with IDB (ICON1 and ICON2—see "tnsnames.ora File"), and each service connected to a specific node (identified as Node 1 and Node 2, respectively, in the architecture diagrams starting on page 115). For more information about the Oracle RAC setup, see "Oracle RAC Configuration" on page 119.

A DB Server was deployed on each host on which ICONs were running. The Oracle client used the DBMS Name that each ICON supplied to DB Server through the ICON database access point (DAP), to do a lookup in the tnsnames.ora file on the host to get the connection string to use to connect to the RAC cluster.

tnsnames.ora File

In the tnsnames.ora file on each DB Server host, the following two services were defined for ICON:

```
ICON1 =
(DESCRIPTION =
  (ADDRESS = (PROTOCOL = TCP)(HOST = gen-scan.testlab.genesys.ca)(PORT =
1521))
  (LOAD_BALANCE = NO)
  (CONNECT_DATA =
   (SERVER = DEDICATED)
   (SERVICE_NAME = ICON1.testlab.genesys.ca)
  )
)
```

and

```
ICON2 =
(DESCRIPTION =
  (ADDRESS = (PROTOCOL = TCP)(HOST = gen-scan.testlab.genesys.ca)(PORT =
1521))
  (LOAD_BALANCE = NO)
  (CONNECT_DATA =
   (SERVER = DEDICATED)
   (SERVICE_NAME = ICON2.testlab.genesys.ca)
  )
)
```

where (HOST = gen-scan.testlab.genesys.ca) is the name of the Listener service defined in the Oracle REMOTE_LISTENER initialization parameter.

DAP Objects for ICON

On each DAP that enabled an ICON to access its IDB (*ICON DAP*), the DBMS Name field on the DB Info tab specified the named service to be used. Half the ICONs were directed to ICON1 and half to ICON2.

Separate schemas in the database corresponded to each IDB. The User Name field on the DB Info tab specified the schema name. As shown in Figures 30 through 32 starting on page 115, the ICONs that were directed to the ICON1 service populated IDB schemas on Oracle RAC Node 1, and the ICONs that were directed to the ICON2 service populated IDB schemas on Oracle RAC Node 2.

For example, Figure 33 shows the DB Info tab of the ICON DAP for a Voice details ICON that used the ICON1 service to access the IDB schema named ICON_ORACLE_S01_IVR.

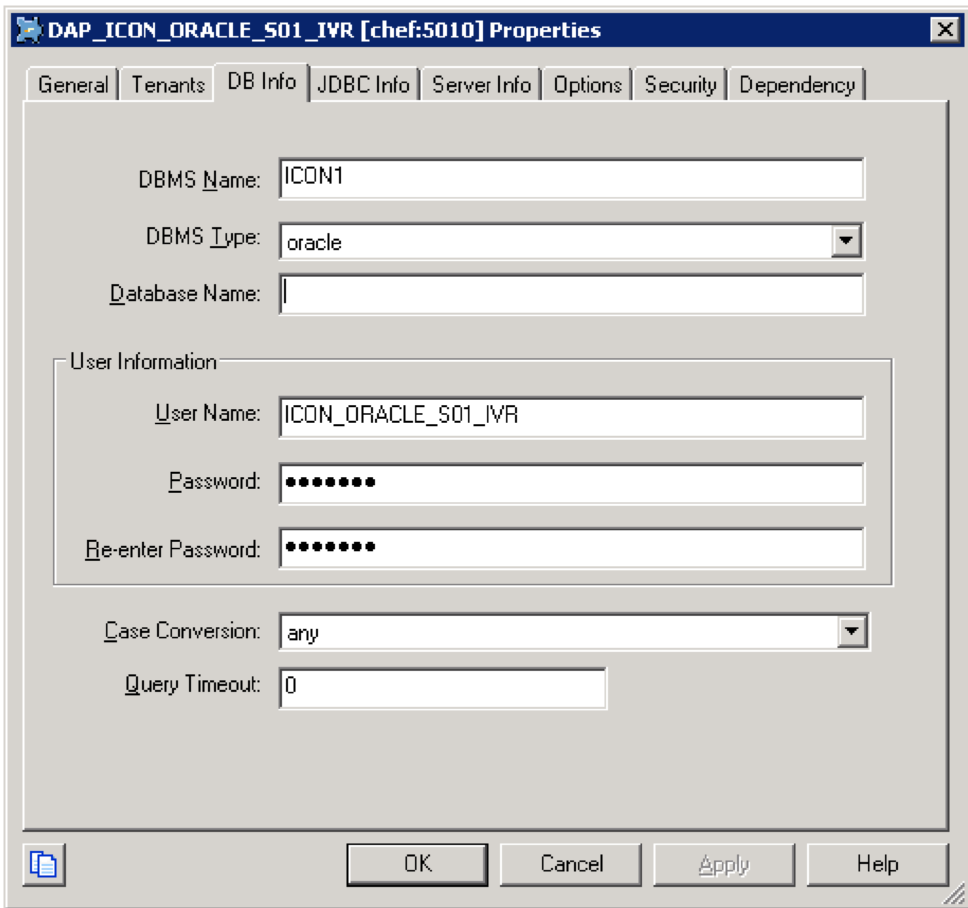


Figure 33: ICON DAP Example

The same DAP objects were used as *extraction DAPs*, to enable Genesys Info Mart to access the IDBs for the purposes of data extraction. For Genesys Info Mart–related configuration on the DAPs, see “Extraction DAPs” on page 127.

Genesys Info Mart–Related Configuration

Genesys Info Mart Application

Table 48 lists those Genesys Info Mart release 8.1.2 configuration options that were set to non-default values for the testing. The table includes only options that affected test results. Either the default values were used for all the other Genesys Info Mart application options or else, where non-default settings were used, the options relate to functionality or resource usage that would not have affected test results. For additional option settings that affected aggregation, see “Aggregation–Related Configuration” on page 191.

Table 48: Non-Default Genesys Info Mart 8.1.2 Application Settings

Option	Value Used	Default Value
gim-etl Section		
days-to-keep-active-facts	5	30
days-to-keep-gidb-facts	4	14
extract-data-chunk-size	1000	900
extract-data-max-conn	350	128
extract-data-thread-pool-size	350	32
max-thread-duration-after-inactive-in-days	10	30
merge-chunk-size	250000	200000
partitioning-ahead-range	50	14
purge-transaction-size	1000000	100000
gim-etl-populate Section		
populate-mm-ixn-queue-facts	true	false
gim-transformation Section		
default-ivr-to-self-service	true	false
irf-io-parallelism	8	4
ud-io-parallelism	8	5
log4j Section		
logging-level	DEBUG	INFO

Table 48: Non-Default Genesys Info Mart 8.1.2 Application Settings (Continued)

Option	Value Used	Default Value
max-backup-index	99	10
schedule Section		
aggregate-duration aggregate-schedule **Note:** These option settings mean that the aggregation job ran continuously from 00:10 until 23:40 every day.	23:30 10 0	05:00 0 1
etl-end-time etl-start-time **Note:** With etl-frequency at the default value of 1 (minute), these option settings mean that the ETL cycle ran almost continuously between 00:10 and midnight every day.	00:00 00:10	22:00 06:00
maintain-start-time	00:00	03:00
run-aggregates	TRUE	FALSE
run-scheduler	TRUE	FALSE
timezone	EST	GMT

Field Objects—Outbound Contact Records

In the [gim-etl-mapping] section on the Annex tab of Field objects, the table-name and column-name options were configured to define storage for the Outbound Contact record field data. High-cardinality data was stored in the CONTACT_ATTEMPT_FACT (CAF) table, and low-cardinality data was stored in the RECORD_FIELD_GROUP_1 and RECORD_FIELD_GROUP_2 dimension tables.

Info Mart Database

The Info Mart database was partitioned, using the default partition size of 24 hours for GIDB tables and dimensional model fact tables. Partitioning enabled database maintenance to be streamlined.

Given the call flows that were used (see page 113), there were no very long-living multimedia interactions. GIDB data was purged relatively aggressively (see the days-to-keep-* option values in Table 48 on page 125), but no data was purged from the dimensional model during the testing.

The Info Mart database was deployed on two nodes of the Oracle RAC cluster: the core Info Mart tables on one node (identified as Node 4 in the architecture diagrams starting on page 115) and the aggregate tables on another (Node 3). Two named services (GIM and RAA) were defined in the tnsnames.ora file to perform the functions of Genesys Info Mart and RAA operations with the Info Mart database.

For more information about the Oracle RAC setup, see "Oracle RAC Configuration" on page 119.

DAP Objects for Genesys Info Mart

Extraction DAPs

For the DAPs that enable the extraction job to access the IDBs (*extraction DAPs*), Genesys Info Mart reused the non-JDBC ICON DAPs, with the default-schema, jdbc-url, and role options configured in the [gim-etl] section on the Options tab. The SERVICE_NAME parameter in the jdbc-url option value, which specified the named ICON service, and the default-schema option value, which identified the IDB schema, matched the equivalent information on the DB Info tab. For example, for the Voice details ICON DAP shown in Figure 33 on page 124, the following options were configured in the [gim-etl] section:

- default-schema=ICON_ORACLE_S01_IVR
- jdbc-url=jdbc:oracle:thin:@(DESCRIPTION=(ADDRESS=(PROTOCOL=TCP)(HOST=gen-scan.testlab.genesys.ca)(PORT=1521))(LOAD_BALANCE=NO)(CONNECT_DATA=(SERVER=DEDICATED)(SERVICE_NAME=ICON1.testlab.genesys.ca)))
- role=ICON_CORE

Info Mart and Admin Console DAP

The deployment used a single, non-JDBC DAP object to enable Genesys Info Mart and the Genesys Info Mart Administration Console to access the Info Mart database (the *Info Mart DAP* and *Admin Console DAP,* respectively).

On the DB Info tab, the DBMS Name field specified the GIM service, and the User Name field specified the Info Mart schema (GEN_GIM_ETL_812).

In the [gim-etl] section on the Options tab, a supplementary JDBC URL option enabled aggregation to be configured to use a separate Cluster node. The following options were configured in the [gim-etl] section:

- default-schema=GEN_GIM_ETL_812
- agg-jdbc-url = jdbc:oracle:thin:@(DESCRIPTION=(LOAD_BALANCE=OFF)(ADDRESS=(PROTOCOL=TCP) (HOST=gen-scan.testlab.genesys.ca)(PORT=1521))(CONNECT_DATA=(SERVER=DEDICATED)(SERVICE_NAME=RAA.testlab.genesys.ca)))
- jdbc-url = jdbc:oracle:thin:@(DESCRIPTION = (ADDRESS = (PROTOCOL = TCP)(HOST = gen-scan.testlab.genesys.ca)(PORT = 1521)) (LOAD_BALANCE = NO) (CONNECT_DATA = (SERVER = DEDICATED) (SERVICE_NAME = GIM.testlab.genesys.ca)))
- role=ADMIN_CONSOLE,INFO_MART

JVM and System-Related Startup Parameters

In the gim_etl_server file, the Java memory setting and other startup parameters for the Genesys Info Mart Server process were set to the following:

```
{JAVACMD} -server -cp "${GIM_ETL_CP}" -Xmx4000m -Xms32m -Xss512k
-XX:GCTimeRatio=4 -DqueryParallelism=8 -DreadCfgDNInfo=false
-Doracle.jdbc.mapDateToTimestamp=true
-Doracle.jdbc.J2EE13Compliant=true
-Dcom.ibm.tools.attach.enable=no -Duser.country=US
-Duser.language=en -Djava.library.path="${GIM_LIBPATH}"
-Duser.timezone=GMT com.genesyslab.gim.etl.server.GIMServer "$@"
```

The queryParallelism startup parameter controls the degree of parallelism used in certain SQL queries. For Oracle, specifying -DqueryParallelism=8 means that PARALLEL(8) will be inserted as a query hint into certain SQL statements. Genesys tested queryParallelism values of 0, 4, 8, and 16 in the test environment, and a value of 8 yielded the best job times.

Release 8.1.2 Endurance Test

The endurance testing measured the ability of Genesys Info Mart 8.1.2 to process 19 million interactions per day for 14 consecutive days, including aggregation.

Interactions were generated at flat hourly rates, as described in "Interaction Volumes" on page 111, for a total of more than 50 days, in order to populate the IDB and Info Mart databases with significant quantities of data so that realistic, steady-state test measurements could be made for 14 days at the end of the extended test period. With the ETL frequency at the default value of 1 minute, the ETL jobs (Job_ExtractGIM and Job_TransformGIM) ran throughout the day, except for a daily 10-minute maintenance window. Aggregation ran continuously throughout the day, except for a half-hour maintenance window.

For information about the hardware, software, and call flows that were used for the endurance test, see "Release 8.1.2 Performance Test Setup" on page 111.

For the detailed endurance test results for Genesys Info Mart, see "Endurance Test Results". For the detailed performance test results for RAA, see "RAA Results for the Endurance Test" on page 193.

Endurance Test Results

The endurance performance test was successful, based on the following criteria:

- The Genesys Info Mart and RAA applications ran without interruption for 14 consecutive days and were stable for the entire duration of the test.

- Genesys Info Mart was comfortably able to sustain the test plan call volume of 19 million interactions per day and support intraday aggregation and reporting.

  Genesys Info Mart processed and loaded all the generated interactions into the Info Mart database within 30 minutes of call termination.

ETL cycle times averaged 53 seconds. Data latency, defined as the time from call termination to the time that the corresponding fact is available in the Info Mart tables, averaged 106 seconds.

Figure 34 shows the data latency for each ETL cycle over the course of a typical day late in the endurance run. The ETL cycle was suspended for 10 minutes at cycle 268 for the daily maintenance window. When the ETL cycle resumed, latency spiked to 347 seconds while Genesys Info Mart caught up after the stoppage, but latency returned to less than the average of 106 seconds within 4 ETL cycles.

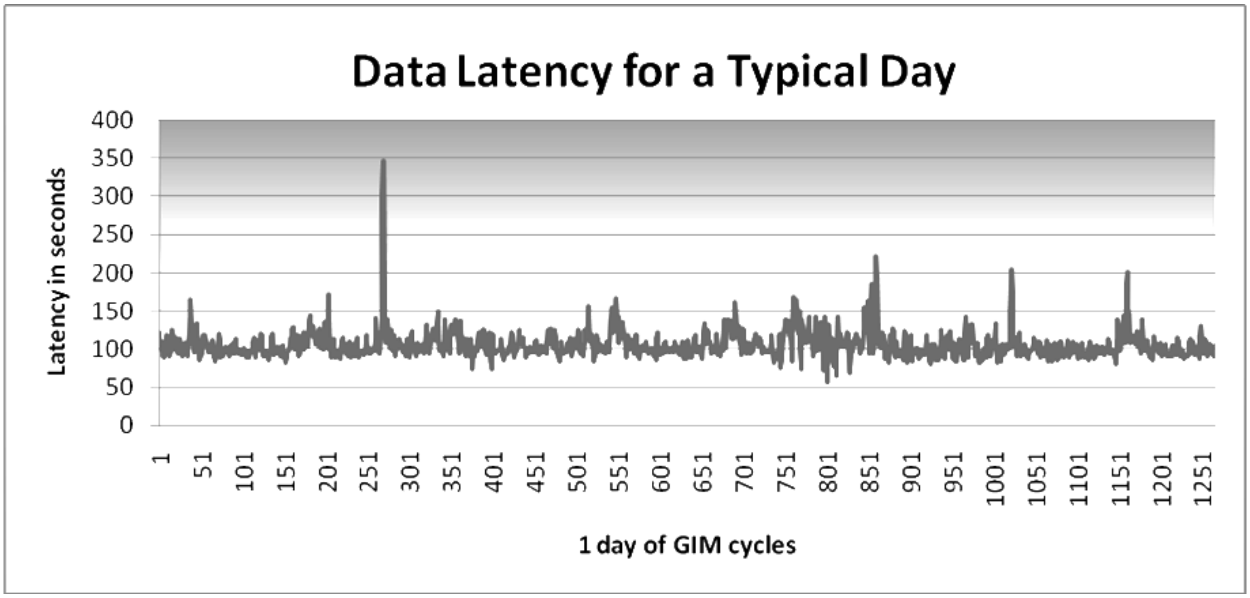


Figure 34: Data Latency—Endurance Test

Performance of Genesys Info Mart Jobs

The primary measure of Genesys Info Mart performance is the amount of time it takes for jobs to execute (the *duration* of the job or cycle). This subsection presents detailed results for the ETL and maintenance jobs.

For related information about aggregation performance, see “RAA Results for the Endurance Test” on page 193.

ETL Job Durations

ETL Cycle

Figure 35 on page 130 shows the durations of the ETL cycles during Day 50 of the extended test period. The 10 minute ETL stoppage during the maintenance window caused the spike at cycle 268. The average time for an ETL cycle was

53 seconds, with a minimum of 8 seconds and a maximum of 252 seconds. The maximum duration was for the first cycle following the maintenance stoppage.

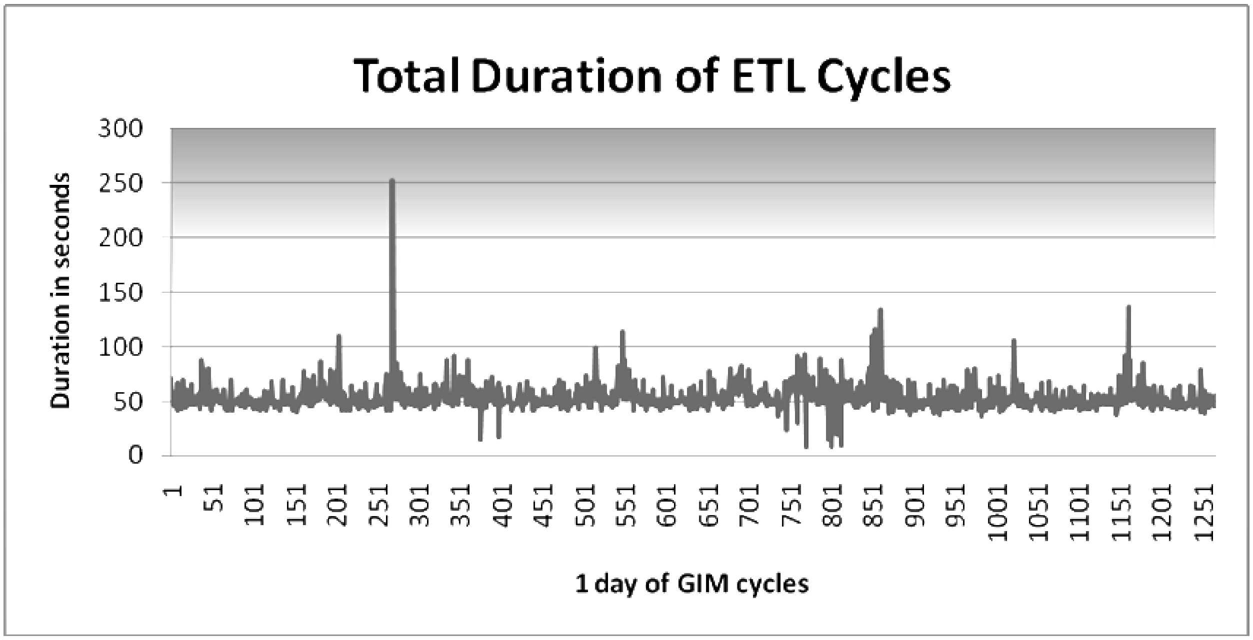


Figure 35: Durations of ETL Cycles—Endurance Test

Extraction and Transformation Jobs Separately

Figure 36 and Figure 37 on page 131 depict the durations of the extraction and transformation jobs, respectively, during Day 50 of the extended test period. The results indicate that, compared with Genesys Info Mart 8.0 (see page 163), extraction performance has improved significantly. In particular:

- On average, data extraction is quicker than transformation: The extraction job averaged 24 seconds, with a minimum of 9 seconds and a maximum of 94 seconds. The transformation job averaged 29 seconds, with a minimum of 1 second and a maximum of 158 seconds. The maximum durations were for the first jobs following the maintenance window.
- The extraction job no longer exhibits a trend for the job duration to increase gradually over time. With a partitioned IDB, improved IDB purge performance means that the size of IDB no longer continues to grow, degrading the performance of the extraction job.

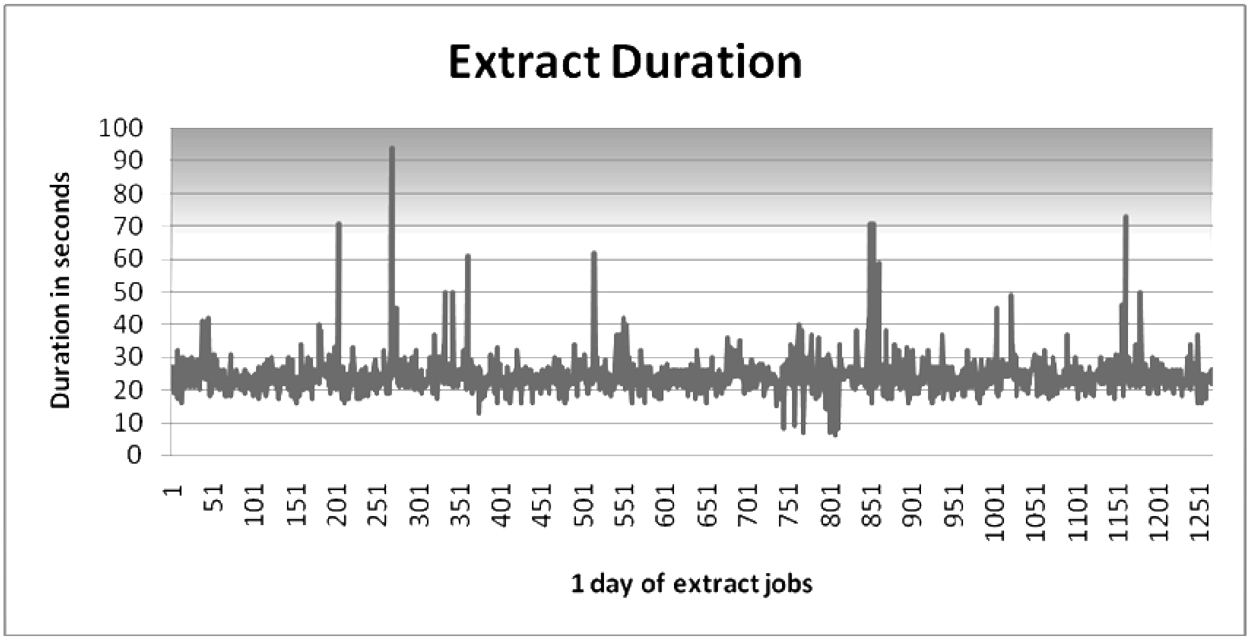


Figure 36: Durations of Extraction Job—Endurance Test

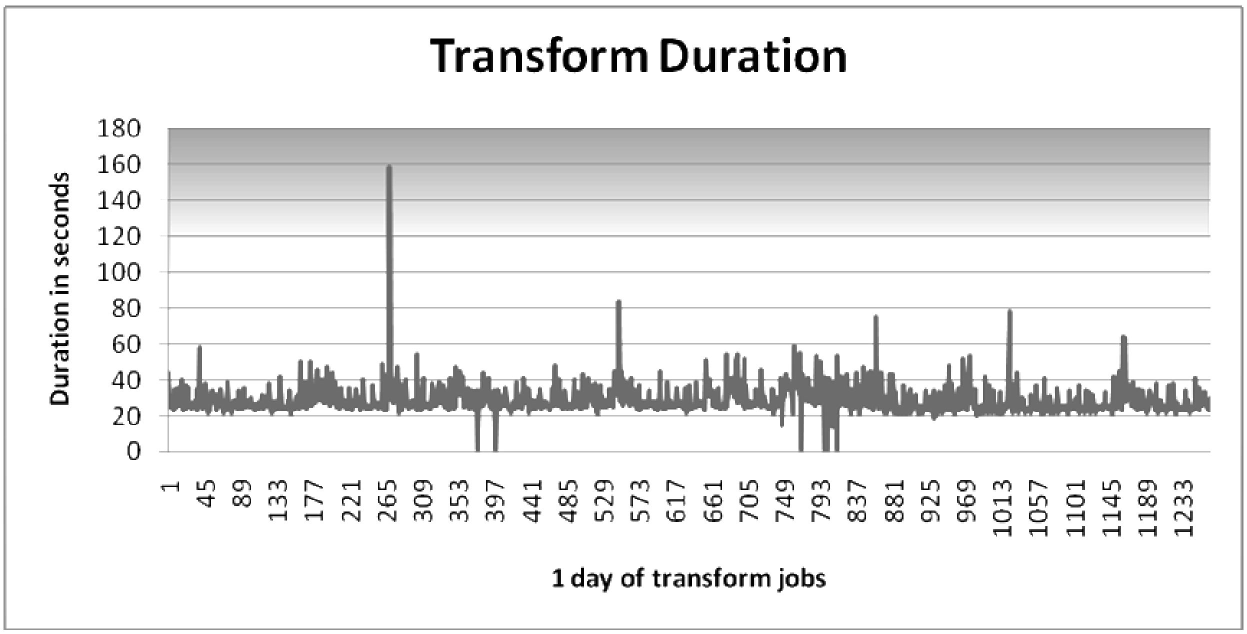


Figure 37: Durations of Transformation Job—Endurance Test

Maintenance Job Duration

Figure 38 on page 132 shows the durations of the daily maintenance job (Job_MaintainGIM) over the last 51 days of the extended test period. There was no long-term growth in job times.

To enable continuing performance to be measured against as full a database as possible, no facts were purged from the Info Mart dimensional model during the extended test period. The maintenance job purged GIDB data and operated on internal Info Mart control and staging tables. Because the Info Mart database was partitioned, with the size of partitions in GIDB set to the default 86400 seconds (24 hours) and with the partitioned tables having only local indexes, a day's worth of extracted IDB interaction data in GIDB was purged very quickly at each run of the daily maintenance job.

Note: The Genesys Info Mart maintenance job does not purge IDB. Interaction Concentrator functionality was used to purge the IDBs on a daily schedule throughout the extended test period.

For information about IDB purge during the testing, see "IDB Purge Performance" on page 143.

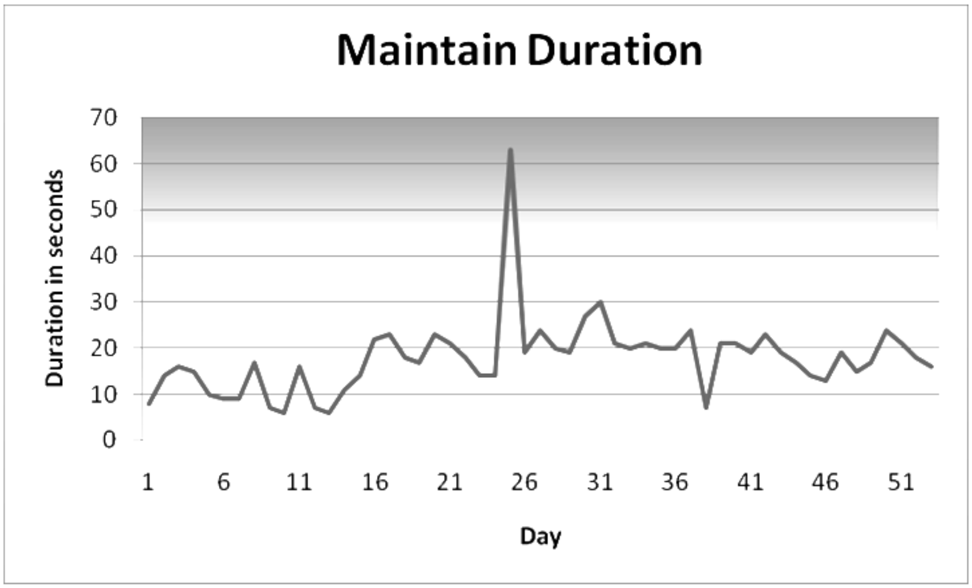


Figure 38: Durations of Maintenance Job—Endurance Test

Hardware Resource Usage—Applications

This subsection provides information about hardware resource usage for the Genesys Info Mart application and the ICON applications during one day of the endurance run.

For details about the hardware that was used in the endurance test, see "Hardware Architecture" on page 114. For details about the Java and application settings that were used, see "Genesys Info Mart–Related Configuration" on page 125.

Genesys Info Mart Server

Memory was fixed at a maximum of 4000 MB (see "JVM and System-Related Startup Parameters" on page 128), the only input and output (I/O) was for local log files, and network traffic was not a limiting factor. Therefore, Genesys Info Mart Server performance was measured solely in terms of CPU utilization.

Figure 39 shows CPU usage by the ETL process, as reported by the Linux top command over a 5-minute period, sampled at 3-second intervals. The ETL process is multithreaded and can execute on all available CPUs at the same time. Therefore, usage is expressed as a percent of the CPU available for the entire server, not a single core.

Figure 39 covers five ETL cycles. The larger spikes indicate the start of extraction, and the smaller spikes that follow indicate transformation activity. ETL usage peaked during one extraction cycle at 83 percent of the total available CPU capacity.

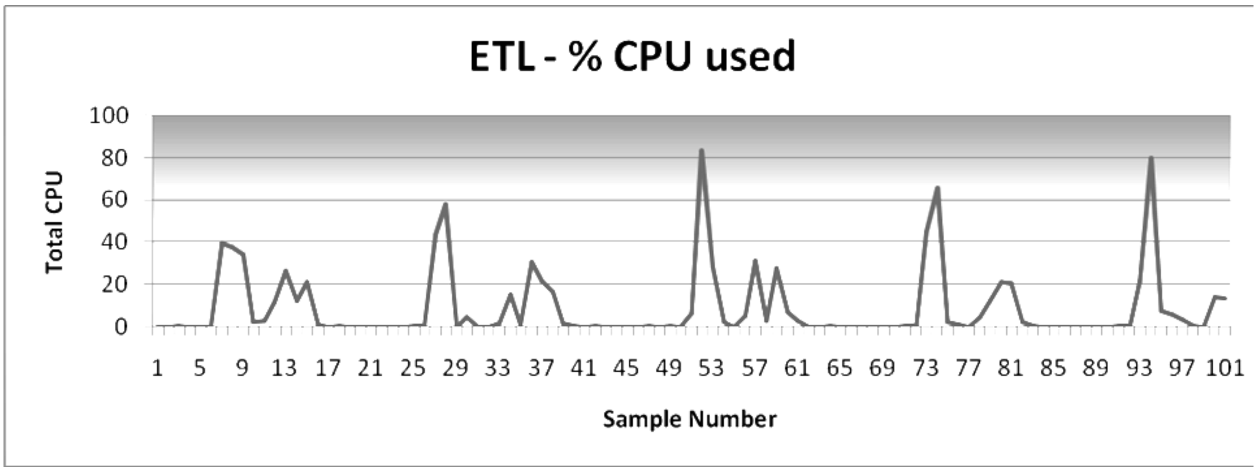


Figure 39: CPU Usage—Genesys Info Mart Server

Figure 40 on page 134 shows monitoring information for the JVM, for a period of five minutes of ETL activity during steady-state operations. The representations of CPU usage, memory usage, and thread counts show similar patterns of spikes for extraction and transformation processing during five ETL cycles.

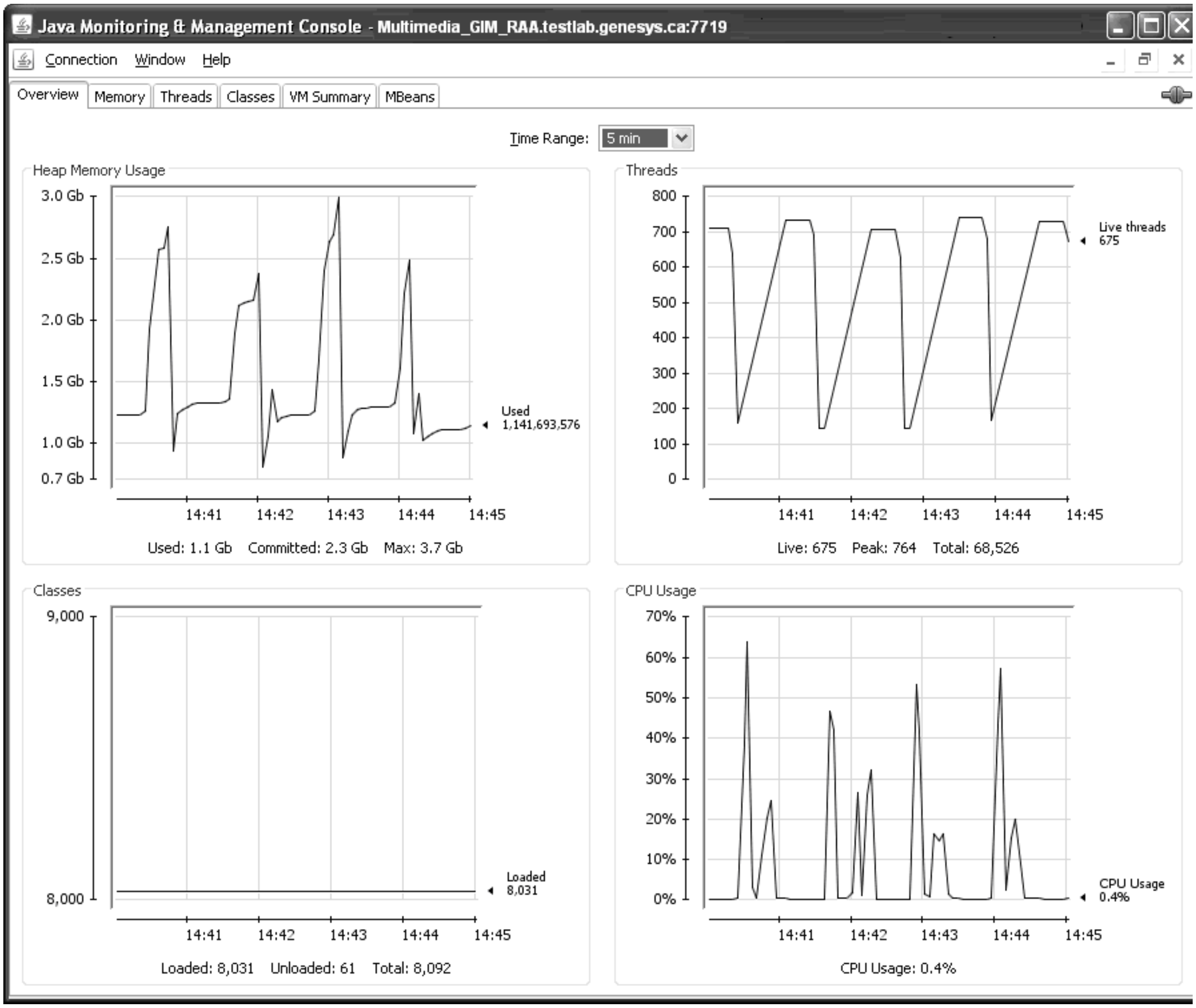


Figure 40: JVM Activity

ICON Applications

The general rule for deploying the ICON applications among the available hosts was:

- No more than one ICON process per core
- 1 GB RAM per ICON process

As described in “Hardware Architecture” on page 114, a total of 22 ICON applications were deployed—1 Configuration details, 16 Voice details for inbound calls, 1 Voice details for Outbound Contact calls, 3 Outbound Contact details, 1 Multimedia details. For details about the distribution of the ICON applications among the available hosts, see Figures 30 through 32, starting on page 115.

Table 49 reports CPU and memory usage by the ICON applications over one hour of operations at the call rates described in “Interaction Volumes” on

page 111. In all cases, CPU usage is expressed as a percent of the CPU available for a single core. The names of the hosts and ICON application processes match the labels in the hardware architecture diagrams starting on page 115.

Table 49: CPU and Memory Usage—ICON Applications

Process	Average CPU Usage (%)	CPU Usage Range (%)	Memory Usage (MB)	Approx. Call Rate (cps)
Host = Voice 1 (6 Voice details ICONs)				
ICON 1	7.2	1–13	678	15
ICON 2	7.2	1–12	687	15
ICON 3	7.2	1–15	673	15
ICON 4	8.9	1–14	580	29
ICON 5	9.2	1–25	580	29
ICON 6	9.1	1–14	583	29
Host = Voice 2 (6 Voice details ICONs)				
ICON 7	8.3	1–15	684	15
ICON 8	7.1	1–13	678	15
ICON 9	9.0	2–20	683	15
ICON 10	9.3	1–14	583	29
ICON 11	8.9	1–15	582	29
ICON 12	9.9	1–20	590	29
Host = Voice 3 (4 Voice details ICONs, with lower call rate and more attached data)				
ICON 13	3.3	1–6	700	10
ICON 14	5.7	1–12	585	10
ICON 15	4.9	1–11	587	10
ICON 16	2.3	1–6	668	10
Host = Outbound 1 (1 Outbound Contact details ICON, 1 Voice details ICON to monitor Outbound Contact calls)				
No data recorded				

Table 49: CPU and Memory Usage—ICON Applications (Continued)

Process	Average CPU Usage (%)	CPU Usage Range (%)	Memory Usage (MB)	Approx. Call Rate (cps)
Host = Outbound 2 (2 Outbound Contact details ICONs)				
ICON 19	10.0	1–41	588	12
ICON 20	10.0	1–43	588	12
Host = Multimedia_GIM_RAA (1 Multimedia details ICON)				
ICON 1	11–24	1–48	700–1700	10

Table 49 shows the following results:

- Voice details ICONs—Memory usage was steady, and CPU usage was low.
- Outbound Contact details ICONs—Memory usage was steady, and CPU usage was comparatively high, in the range of 1–40 percent. The CPU usage of 40 percent likely coincided with the loading of groups of records in OCS
- Multimedia details ICON—Memory usage was high, because of the significant percentage of multiday interactions. Memory usage started at 700 MB and grew steadily to 1700 MB, where it stayed. CPU usage grew along with the memory increase.

Operations during the measurement period included IDB purging. The ICON applications exhibited no differences in resource usage during purging.

Partitioned vs. Non-Partitioned IDBs

The IDBs that were deployed for the performance testing used partitioned schemas. For data volumes and retention periods that were used in supplementary testing, there were no significant differences in performance of the ICON processes between partitioned and nonpartitioned schemas. During normal steady-state operations to process data for peak contact-center activity data, without IDB purging in process, partitioned and nonpartitioned IDBs used equivalent database resources.

However, under the same test conditions, Genesys Info Mart extraction times are slightly faster when the IDB is partitioned than when it is not partitioned.

Hardware Resource Usage—Oracle RAC

The Info Mart and Interaction Concentrator databases were deployed on a four-node Oracle RAC. This subsection provides information about hardware resource usage for all four nodes in the Oracle cluster during one day of the endurance run.

For details about the hardware that was used in the endurance test, see “Hardware Architecture” on page 114. For details about the Oracle RAC setup, see “Oracle RAC Configuration” on page 119.

Memory was fixed for each node in the Oracle cluster, with an Oracle System Global Area (SGA) size of 72 GB. Network traffic was not a limiting factor. Therefore, Oracle RAC performance was measured solely in terms of CPU utilization and disk I/O for each node.

CPU Utilization

The graphs in this subsection show CPU usage on each node in the Oracle cluster, as reported by the Linux top command for the same 5-minute period covered by Figure 39 on page 133, sampled at 3-second intervals.

GIM Node

Figure 41 shows CPU usage on the GIM node.

ETL usage peaked at 67 percent of the total available CPU capacity. Spikes coincide with the start of extraction and with user-data processing during transformation. On average, 5 percent of the time that the CPU was active (*CPU time*) was I/O wait time.

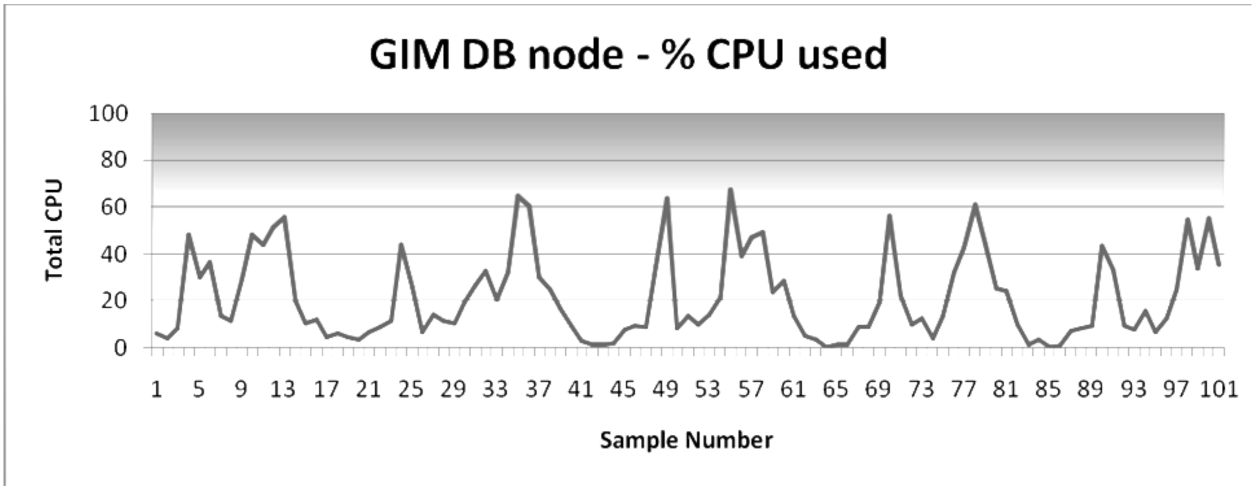


Figure 41: CPU Usage—GIM Node

RAA Node

Figure 42 on page 138 shows CPU usage on the RAA node.

RAA usage was steady and averaged 58 percent of the total available CPU capacity. On average, 25 percent of CPU time was I/O wait time. The relatively high percentage of I/O wait time reflects the fact that RAA is frequently aggregating data, which is no longer cached in the database memory and must be read from disk.

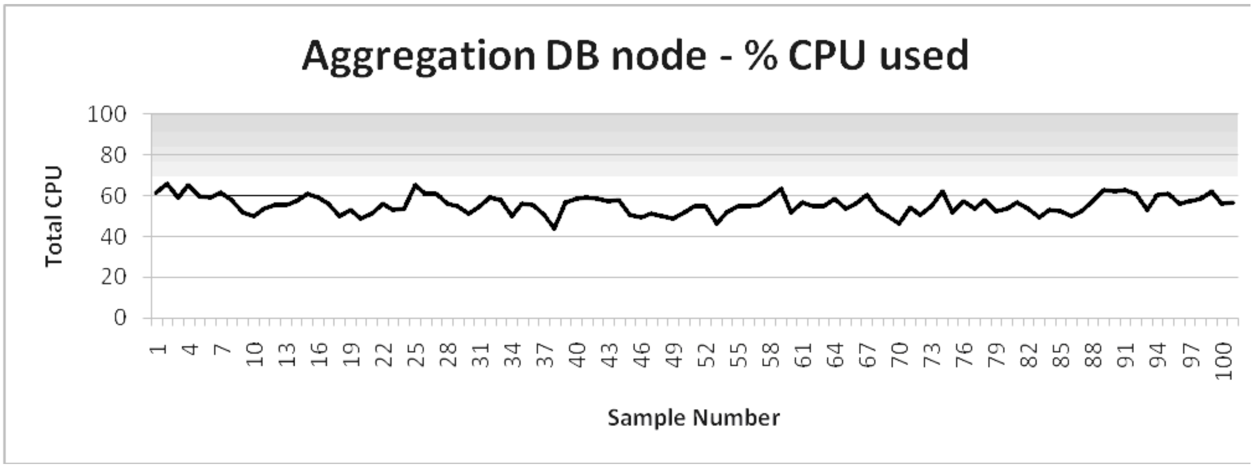


Figure 42: CPU Usage—RAA Node

ICON Nodes

Figures 43 and 44 show CPU usage on the nodes that were used by the ICON1 and ICON2 services (*ICON1 node* and *ICON2 node*), respectively.

Figure 43 shows database activity for the node that hosted the Configuration details IDB, eight Voice details IDBs, and the Multimedia details IDB.

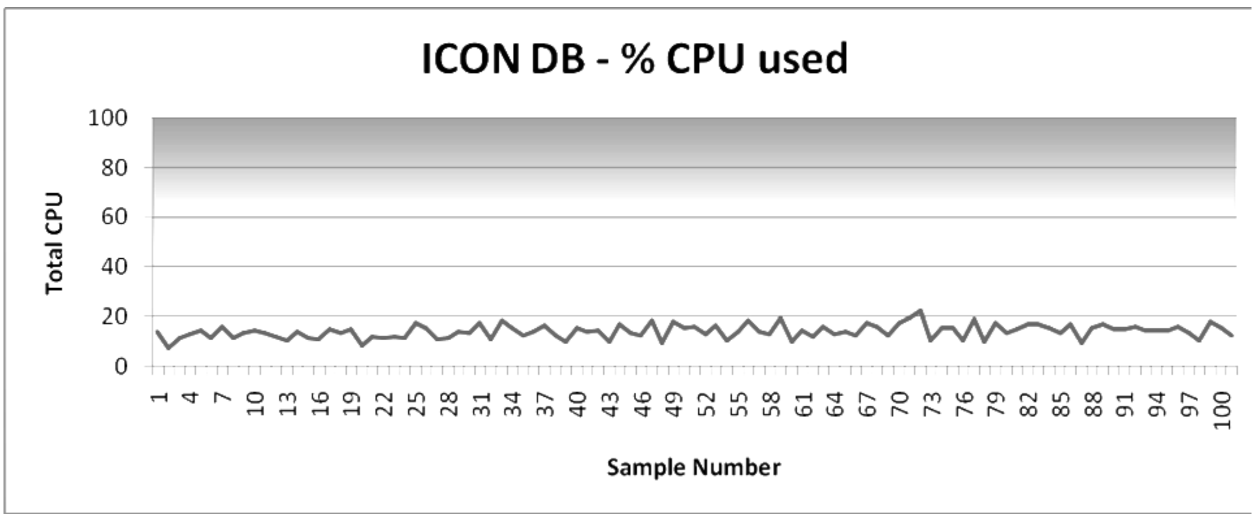


Figure 43: CPU Usage—ICON1 Node

Figure 44 on page 139 shows database activity for the node that hosted the remaining nine Voice details IDBs and the three Outbound Contact details IDBs. Levels of CPU activity on this node are higher than on the other ICON node, because of additional Outbound Contact processing.

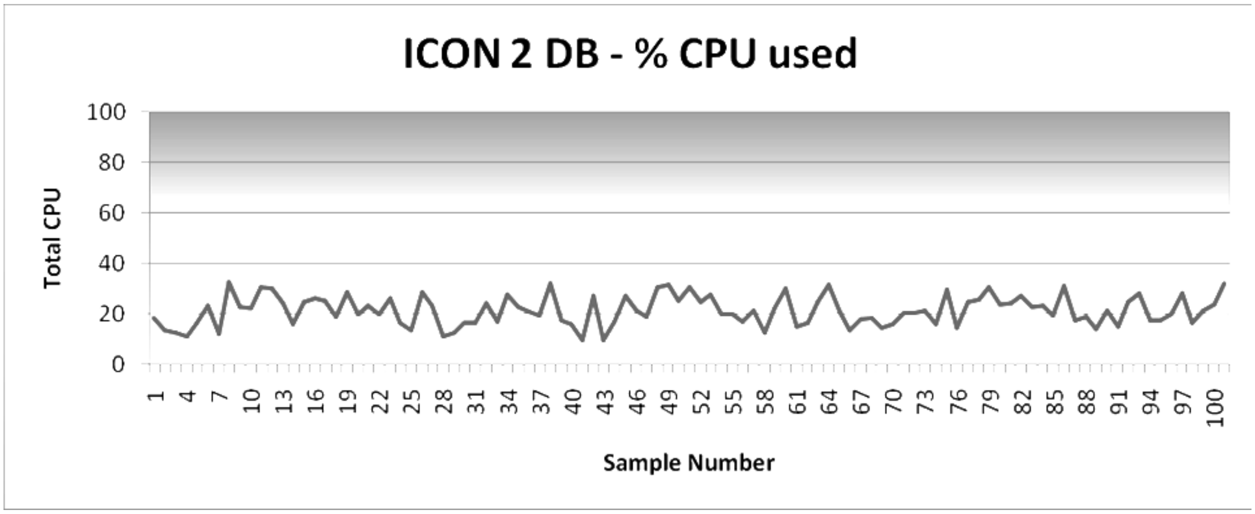


Figure 44: CPU Usage—ICON2 Node

Disk I/O

The graphs in this subsection show I/O activity in the cluster as reported by the Oracle Enterprise Manager, broken down by type of database activity, for a one-hour period during steady-state operations.

Figure 45 on page 140 shows combined I/O activity for all four nodes in the cluster. Subsequent graphs in this subsection show I/O activity separately on each node, for the same one-hour period.

Combined Cluster

As Figure 45 shows, combined cluster I/O peaked at about 250 MB per second. Traffic was divided between the two network-attached storage arrays, with approximately 150 MB per second of data traffic on one array and 100 MB per second of log traffic on the second array.

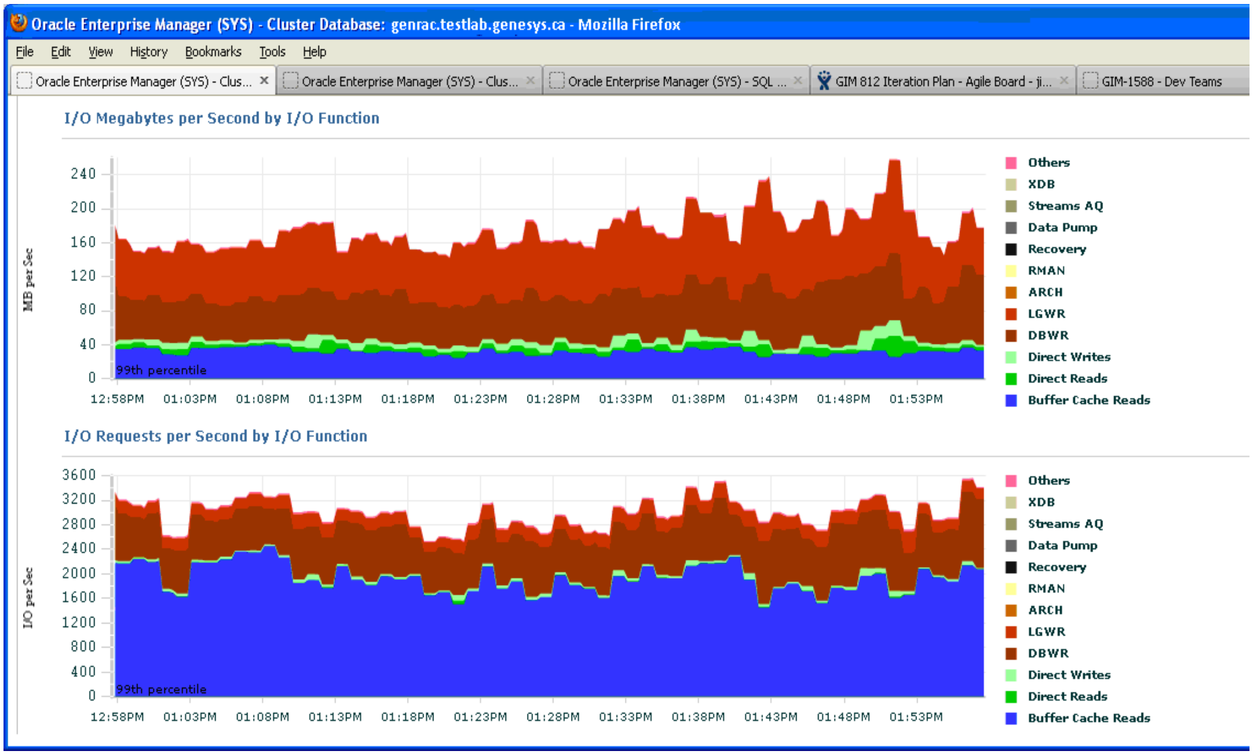


Figure 45: Total I/O Activity for the Oracle Cluster

GIM Node Figure 46 on page 141 shows I/O activity on the GIM node. I/O load was in bursts that coincided with the ETL jobs. Predictably, the majority of the activity was from the Database Writer (DBWR) process and the Log Writer (LGWR) process, which writes asynchronously to the transaction log.

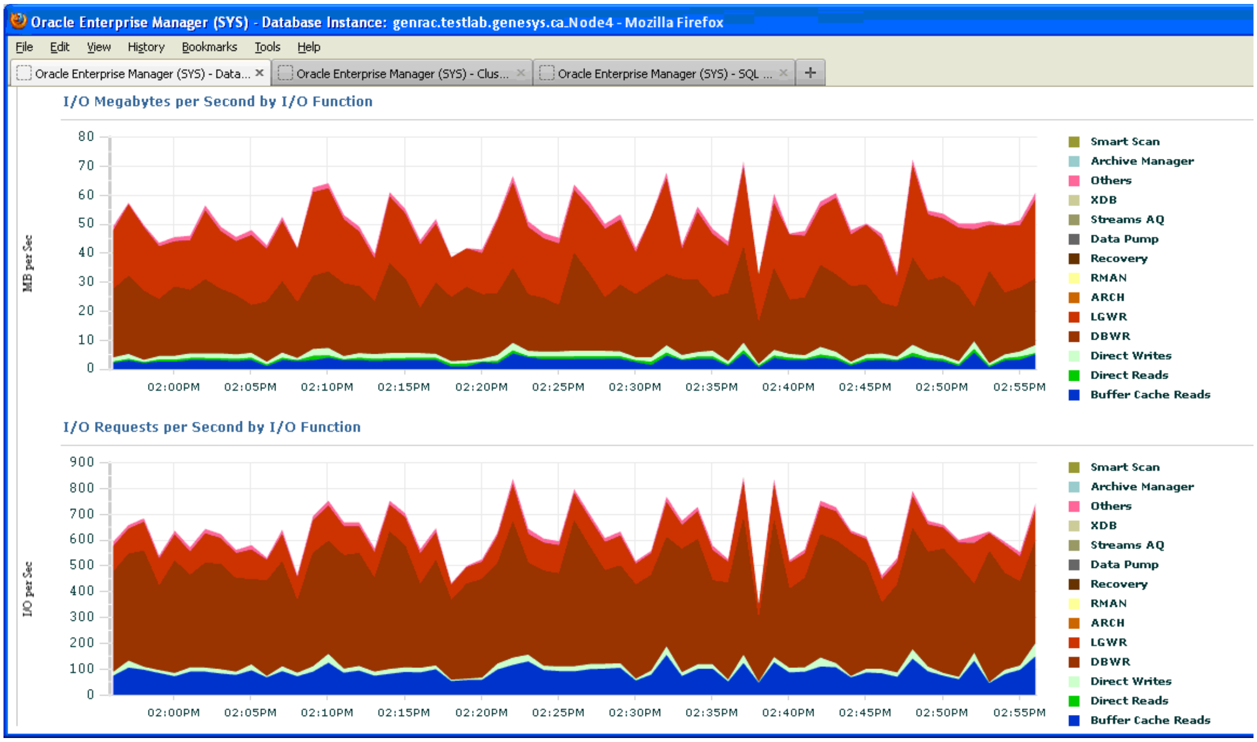


Figure 46: I/O Activity—GIM Node

RAA Node Figure 47 shows I/O activity on the RAA node. Spikes in the I/O load coincided with the execution of large aggregation queries.

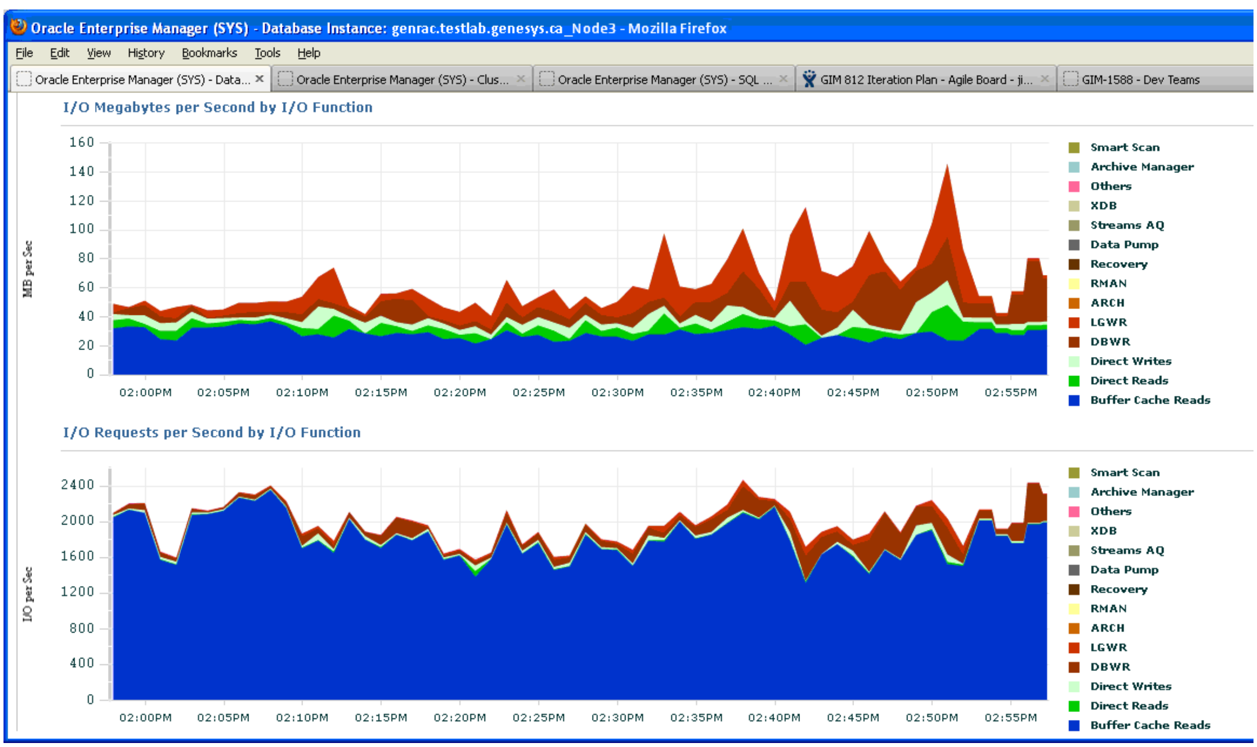


Figure 47: I/O Activity—RAA Node

ICON Nodes Figures 48 and 49 show I/O activity on the ICON1 and ICON2 nodes, respectively. I/O on both nodes was fairly constant, because the call rate was constant.

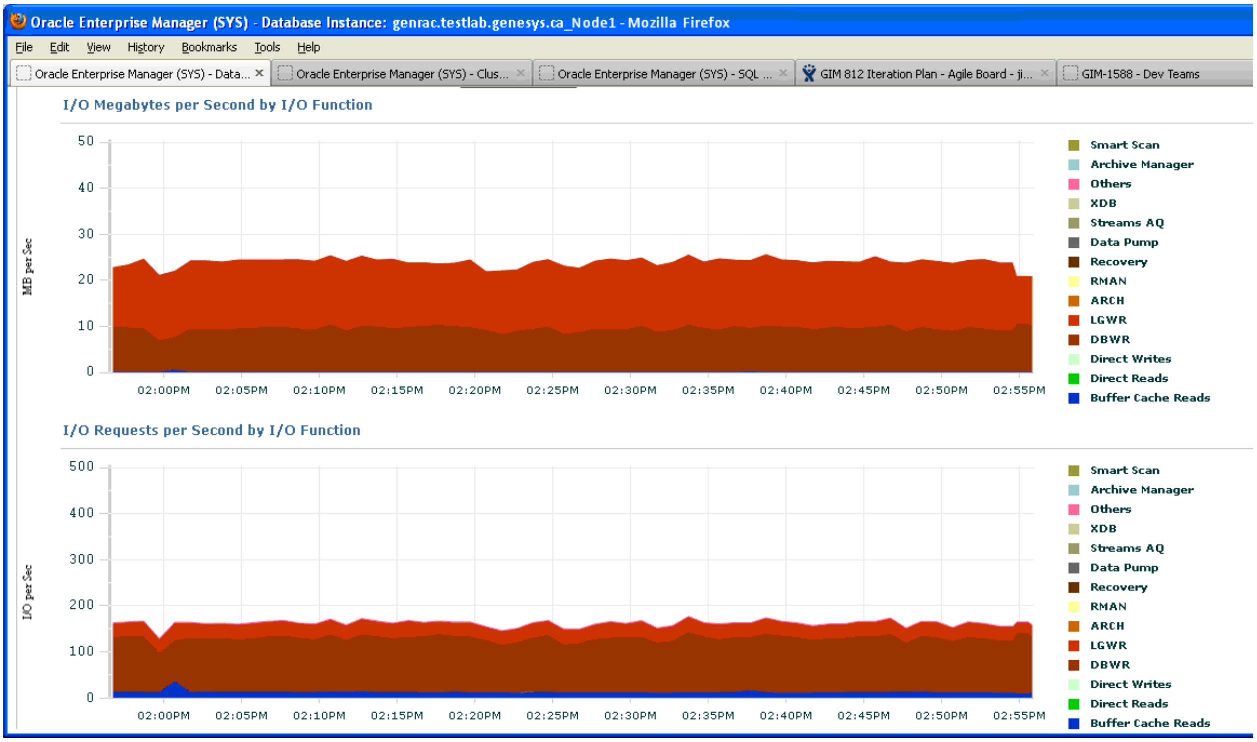


Figure 48: I/O Activity—ICON1 Node

The I/O load on the ICON2 node was approximately double the load on the ICON1 node, because of the additional load from the three Outbound Contact details ICONs.

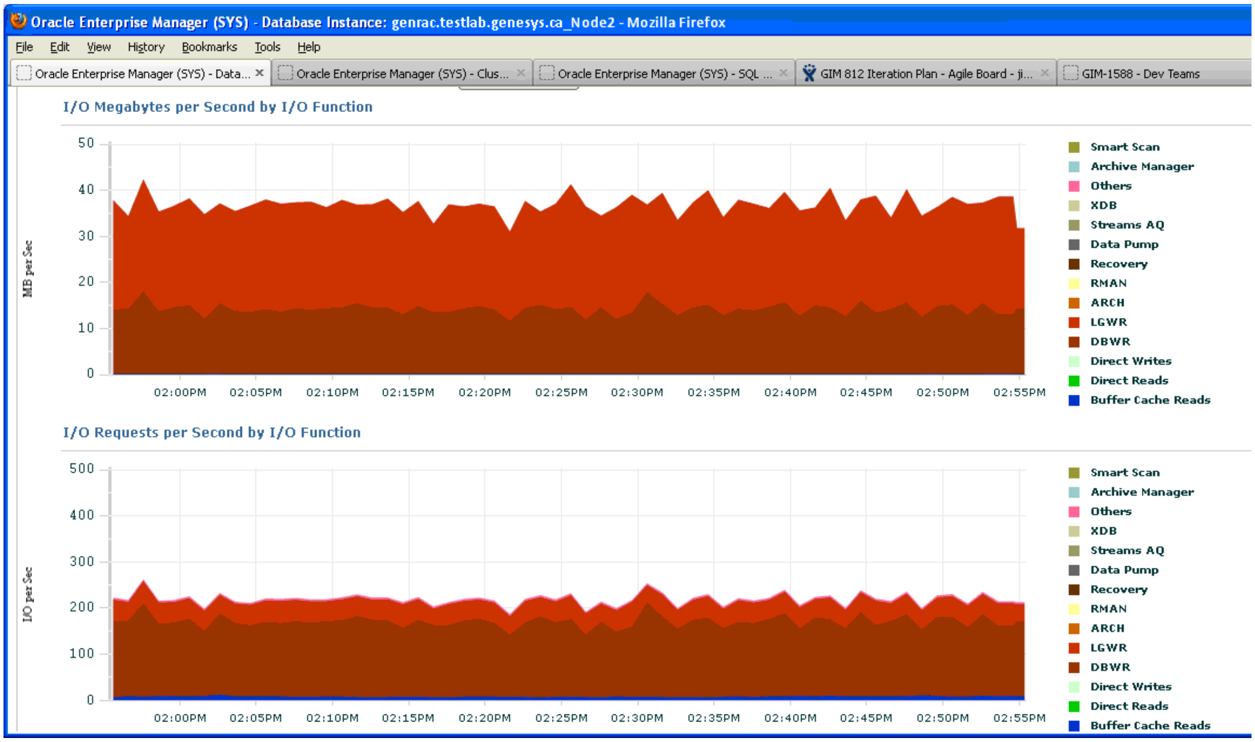


Figure 49: I/O Activity—ICON2 Node

Database Storage Usage

At the end of the extended test run:

- The Info Mart tablespace was 8.558 TB—3.6 TB for the GIDB tables and 5.0 TB for facts and aggregates.
- IDB tablespaces totalled 4.030 TB. Each tablespace held a maximum of seven days of calls; after the first seven days of the test run, the IDBs were purged daily, with a four-day retention.
- There were 988,064,100 rows in the INTERACTION_FACT table. Facts covered 59 days, with only 6 days having fewer than 10 million interactions. As described in "Interaction Volumes" on page 111, the maximum daily volume was 19.1 million interactions.

IDB Purge Performance

Previous Genesys testing has identified that the size of IDB is a primary factor in ICON and Genesys Info Mart performance. Therefore, the ability to purge IDB efficiently is a significant consideration.

As mentioned in "Test Applications" on page 110, preliminary Genesys Info Mart 8.1 testing demonstrated that IDB purge performance with Interaction Concentrator 8.1.0 and nonpartitioned IDBs was an insurmountable performance limitation. At the test volumes, nonpartitioned IDB purge, which uses row deletes to purge data, took more than six hours to purge a day's worth of data and considerably delayed both ICON and Genesys Info Mart database activity.

Accordingly, the test environment was upgraded to use Interaction Concentrator 8.1.1, and the IDBs were deployed on partitioned schemas. On partitioned IDBs, most indexes are local and a new partition-aware purge procedure, which uses TRUNCATE PARTITION to delete old data, is available. The resulting performance improvement was substantial.

Table 50 summarizes the volumes of data purged and the time required to purge one day's worth of data from the partitioned IDBs, while the environment was running at the full interaction rate. For ease of reference to other 8.1.2 performance results, Table 50 uses the names of the associated ICON applications (as shown in the hardware architecture diagrams starting on page 115) to identify the tablespaces for the respective IDBs.

Table 50 reports purging time when the IDBs were purged one at a time. However, for the majority of the extended test run, the IDBs were purged in parallel.

Table 50: Partitioned IDB Purge Performance

Tablespace, by Related ICON Application	Free (MB)		Amount Purged (MB)	Purging Time (seconds)
	Before Purge	After Purge		
ICON 1 (voice IVR)	140,107	171,257	31,150	205
ICON 2 (voice agent)	93,485	114,589	21,104	246
ICON 3 (voice IVR)	132,591	162,692	30,101	217
ICON 4 (voice agent)	93,426	114,541	21,115	231
ICON 5 (voice IVR)	135,778	165,892	30,114	221
ICON 6 (voice agent)	95,328	116,471	21,143	232
ICON 7 (voice IVR)	132,744	162,896	30,152	227
ICON 8 (voice agent)	93,407	114,588	21,181	223
ICON 9 (voice IVR)	132,921	162,881	29,960	218
ICON 10 (voice agent)	93,488	114,716	21,228	225
ICON 11 (voice IVR)	132,584	162,608	30,024	209
ICON 12 (voice agent)	93,793	115,166	21,373	238
ICON 13 (voice IVR)	224,398	273,575	49,177	105
ICON 14 (voice agent)	93,787	113,991	20,204	135

Table 50: Partitioned IDB Purge Performance (Continued)

Tablespace, by Related ICON Application	Free (MB)		Amount Purged (MB)	Purging Time (seconds)
	Before Purge	After Purge		
ICON 15 (voice IVR)	169,453	207,198	37,745	103
ICON 16 (voice agent)	41,012	50,099	9,087	126
ICON 17 (ocs)	81,336	116,016	34,680	32
ICON 18 (voice ocs)	79,877	113,900	34,023	230
ICON 19 (ocs)	60,880	95,902	35,022	27
ICON 20 (ocs)	60,766	95,628	34,862	31
ICON 21 (multimedia)	47,270	72,205	24,935	241
Overall Total			**588,380**	**3722**[a]

a. The total of approximately 60 minutes was the time required to purge the IDBs one at a time. Total time to purge all the IDBs in parallel ranged from 12 to 20 minutes.

Table 50 demonstrates the following results:

- Different functional types of IDBs purge at different rates. In particular, IDBs for Outbound Contact details purge faster than IDBs for Voice details or Multimedia details, because the Outbound Contact details IDBs have fewer data tables.
- IDBs of the same functional type might purge at different rates, depending on the mix of data between the various tables.

Release 8.1.2 Recovery Test—Two-Hour Outage

This recovery test measured the ability of Genesys Info Mart 8.1.2 to recover from a short-term outage, when ETL processing was suspended for two hours. Call generation continued at peak rates during the outage and recovery period.

The recovery test used the same test environment as the endurance test. For details about the hardware, software, and call flows, see "Release 8.1.2 Performance Test Setup" on page 111.

Two-Hour Recovery Test Results

The recovery test was successful, based on the following criterion:

- When ETL processing resumed, Genesys Info Mart was able to process the backlog of two hours of data and catch up to regular, steady-state processing within 1 hour and 8 minutes.

ETL Job Durations

ETL Cycle

Figure 50 shows the durations of the ETL cycles during the recovery period. Before the outage, the average duration of ETL cycles was 53 seconds. The outage occurred between Cycle 1 and Cycle 2. By Cycle 15, the ETL had caught up, and average ETL cycle time was back to 53 seconds. The time between Cycle 2 and Cycle 15 was 1 hour and 8 minutes.

During recovery, the ETL processed the configured maximum of 1000 seconds of source data per cycle. The average time for ETL cycles during recovery was 340 seconds.

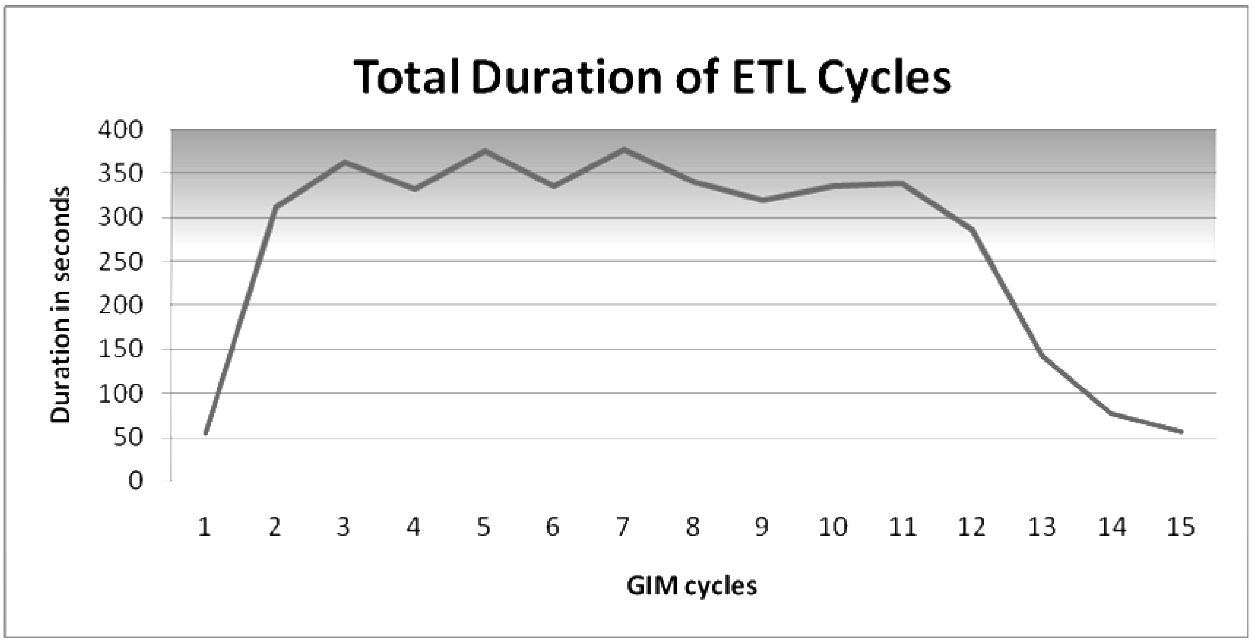


Figure 50: Durations of ETL Cycles Following Two-Hour Outage

Extraction and Transformation Jobs Separately

Figure 51 and Figure 52 on page 147 depict the durations of the extraction and transformation jobs, respectively, during the recovery period.

- The average time to extract a full 1000-second chunk of data was 162 seconds.
- The average time to transform a full 1000-second chunk of data was 179 seconds.

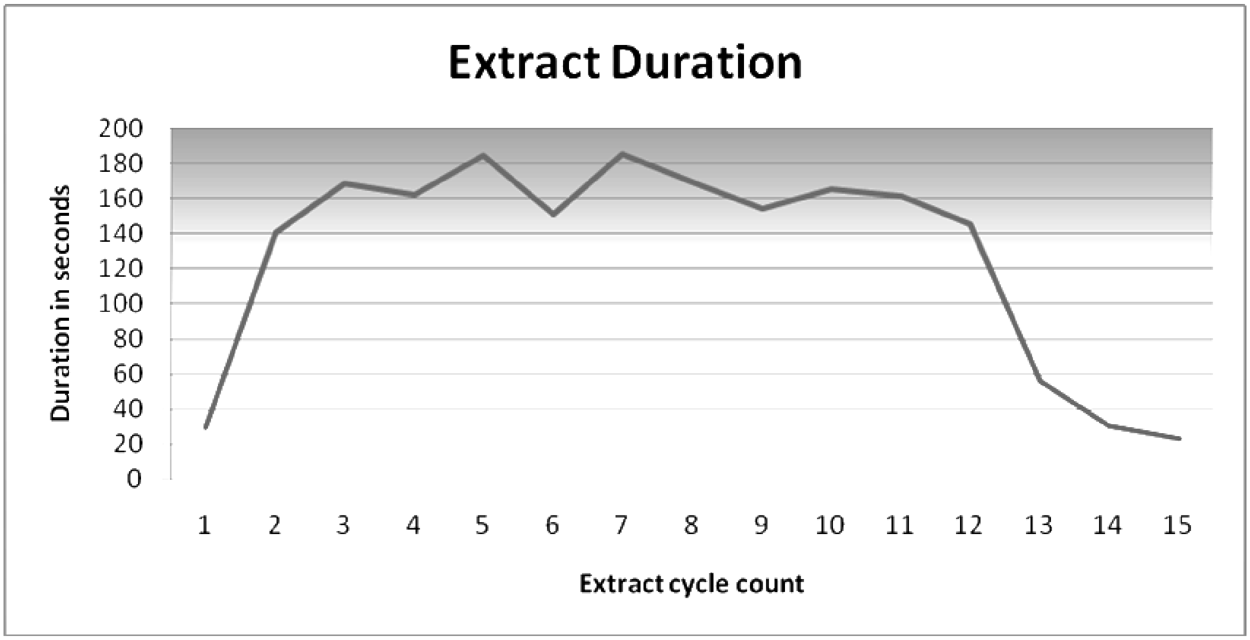


Figure 51: Durations of Extraction Job Following Two-Hour Outage

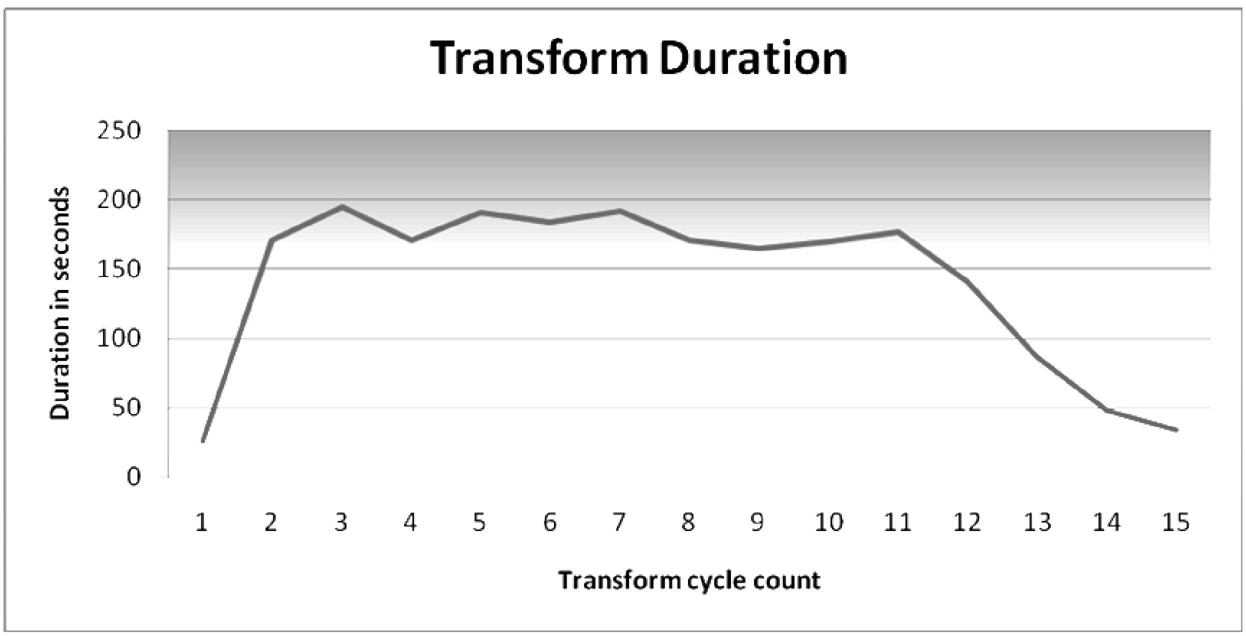


Figure 52: Durations of Transformation Job Following Two-Hour Outage

Release 8.1.2 Recovery Test—Ten-Hour Outage

This recovery test measured the ability of Genesys Info Mart 8.1.2 to recover from a reasonably long outage, when ETL processing was suspended for ten hours. Call generation continued at peak rates during the outage and recovery period.

The recovery test used the same test environment as the endurance test. For details about the hardware, software, and call flows, see "Release 8.1.2 Performance Test Setup" on page 111.

Ten-Hour Recovery Test Results

The recovery test was successful, based on the following criterion:

- When ETL processing resumed, Genesys Info Mart was able to process the backlog of ten hours of data and catch up to regular, steady-state processing within 5 hours and 12 minutes.

ETL Job Durations

ETL Cycle

Figure 53 on page 149 shows the durations of the ETL cycles during the recovery period. Before the outage, the average duration of ETL cycles was 53 seconds. The outage occurred between Cycle 1 and Cycle 2. By Cycle 58, the ETL had caught up, and average ETL cycle time was back to 53 seconds. The time between Cycle 2 and Cycle 58 was 5 hours and 12 minutes.

During recovery, the ETL processed the configured maximum of 1000 seconds of source data per cycle. The average time for ETL cycles during recovery was 340 seconds.

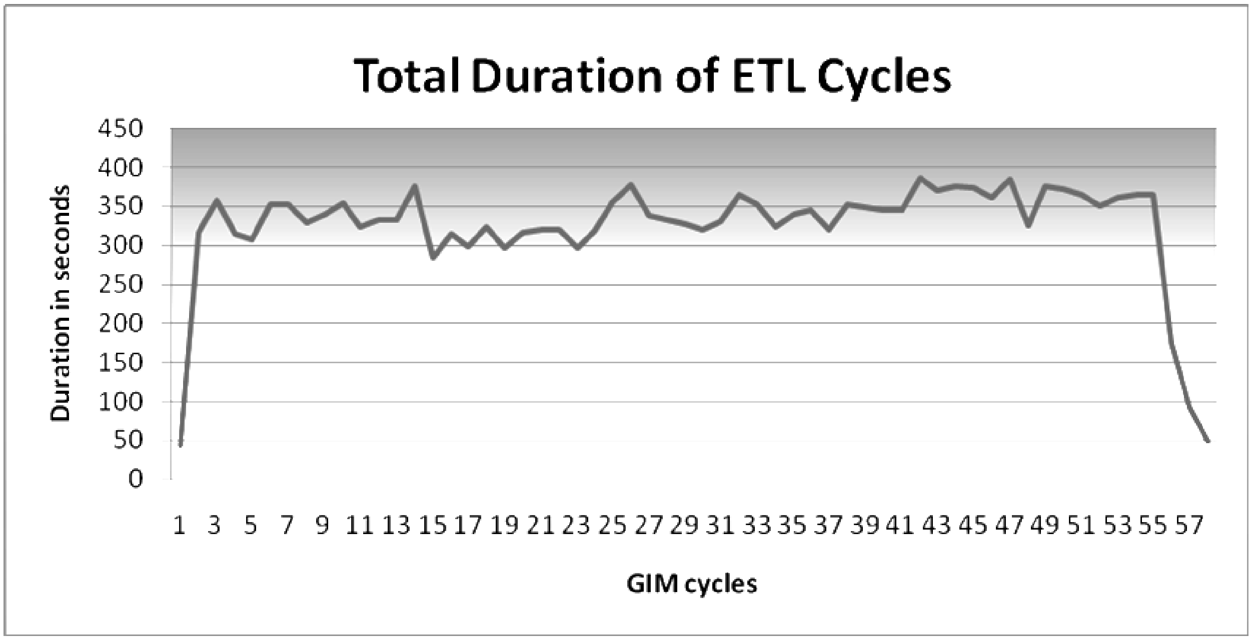


Figure 53: Durations of ETL Cycles Following Ten-Hour Outage

Extraction and Transformation Jobs Separately

Figure 54 and Figure 55 on page 150 depict the durations of the extraction and transformation jobs, respectively, during the recovery period.

- The average time to extract a full 1000-second chunk of data was 147 seconds. There was no significant difference in extraction times throughout the recovery period, even though earlier source data would have aged out of the Oracle cache and had to be read in from disk.
- The average time to transform a full 1000-second chunk of data was 195 seconds.

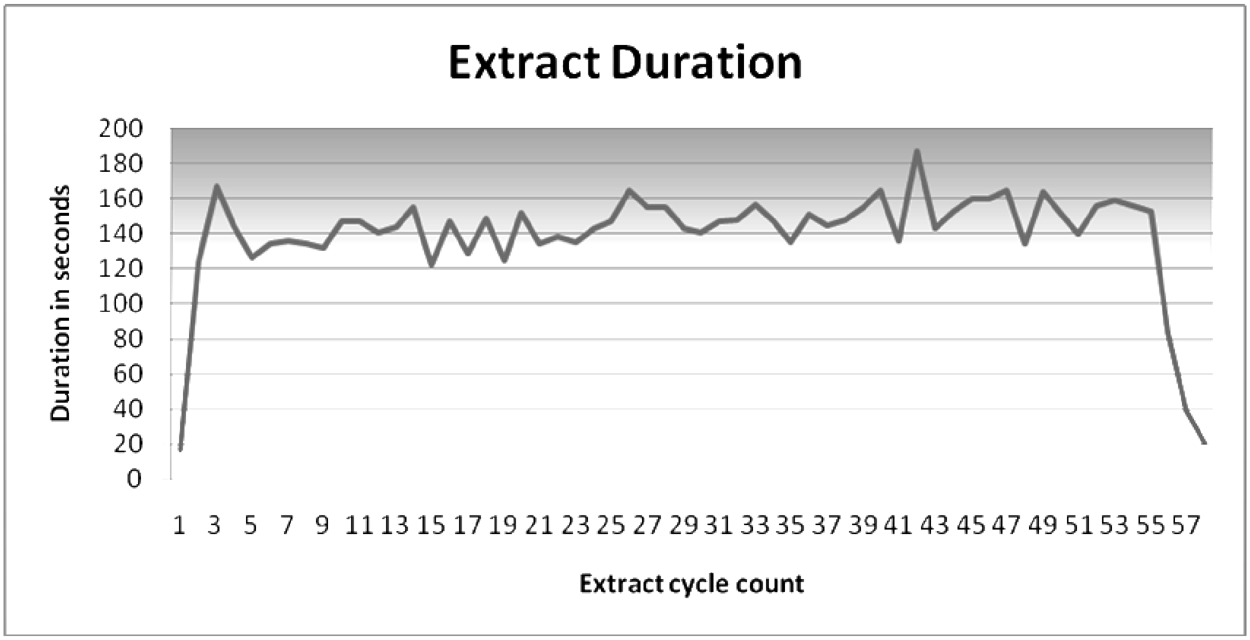


Figure 54: Durations of Extraction Job Following Ten-Hour Outage

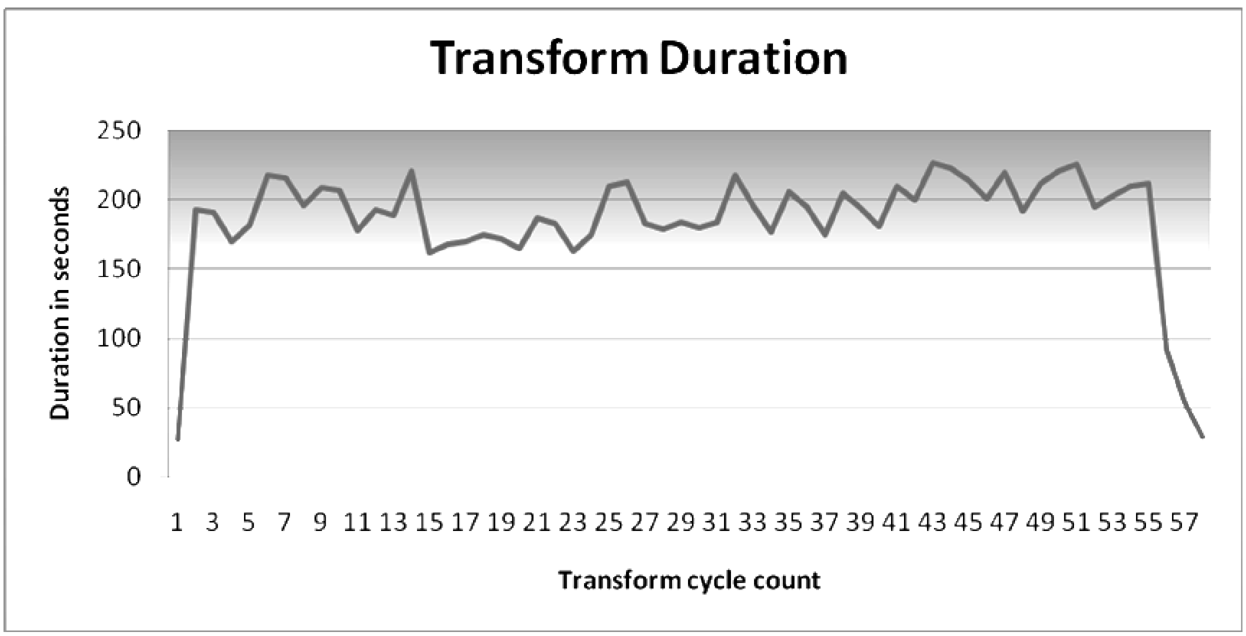


Figure 55: Durations of Transformation Job Following Ten-Hour Outage

Hardware Resource Usage

This subsection provides information about hardware resource usage for the Genesys Info Mart application and the Oracle cluster during the recovery test.

Genesys Info Mart Server

Figure 56 shows CPU usage by the ETL process during the recovery period, as reported by the Linux top command over a 15-minute period, sampled at 3-second intervals. Usage is expressed as a percent of the CPU available for the entire server, not a single core.

Figure 56 covers two full ETL cycles and the extraction part of a third cycle.

- Samples 21–140 cover one ETL cycle:
  - Samples 21–70 cover the extraction job.
  - Samples 71–140 cover the transformation job.
- Samples 141–256 cover the next ETL cycle:
  - Samples 141–190 cover the extraction job.
  - Samples 191–256 cover the transformation job.

As comparison with Figure 39 on page 133 shows, peak usage during recovery was no higher than during steady-state operations.

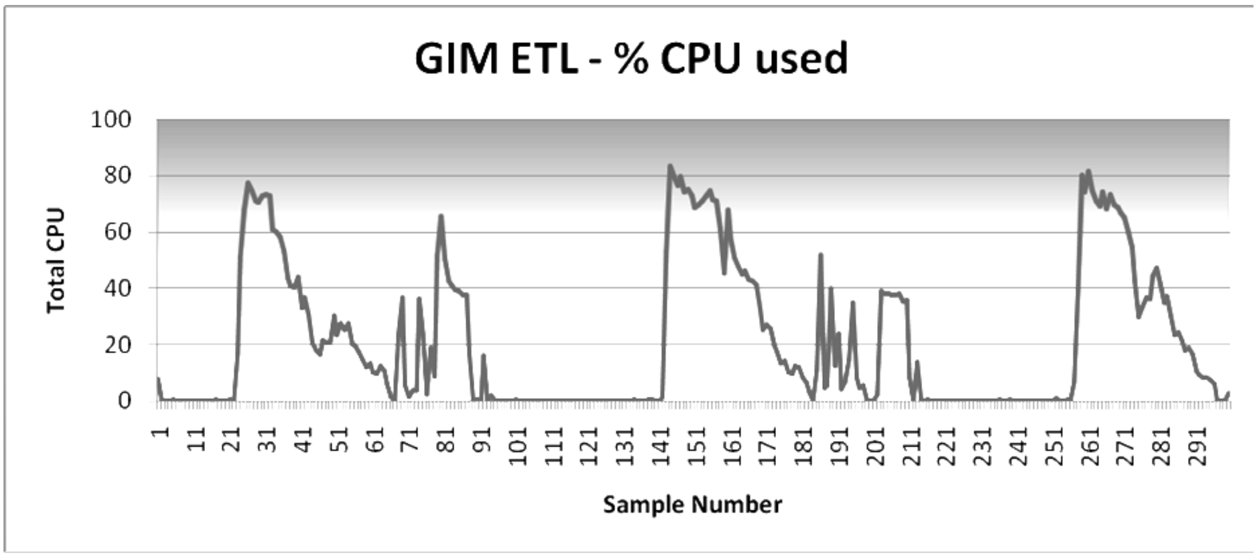


Figure 56: CPU Usage—Genesys Info Mart Server During Recovery Following Ten-Hour Outage

Oracle RAC—CPU Usage

GIM Node

Figure 57 on page 152 shows CPU usage on the GIM node of the Oracle cluster, for the same 15-minute period during recovery. As in Figure 56, Samples 21–70 and 141–190 cover the extraction job, and Samples 71–140 and 191–256 cover the transformation job (which includes loading).

Unlike during steady-state operations, CPU capacity was almost fully utilized at times during the transformation job. During the extraction job, CPU use was limited by I/O waits.

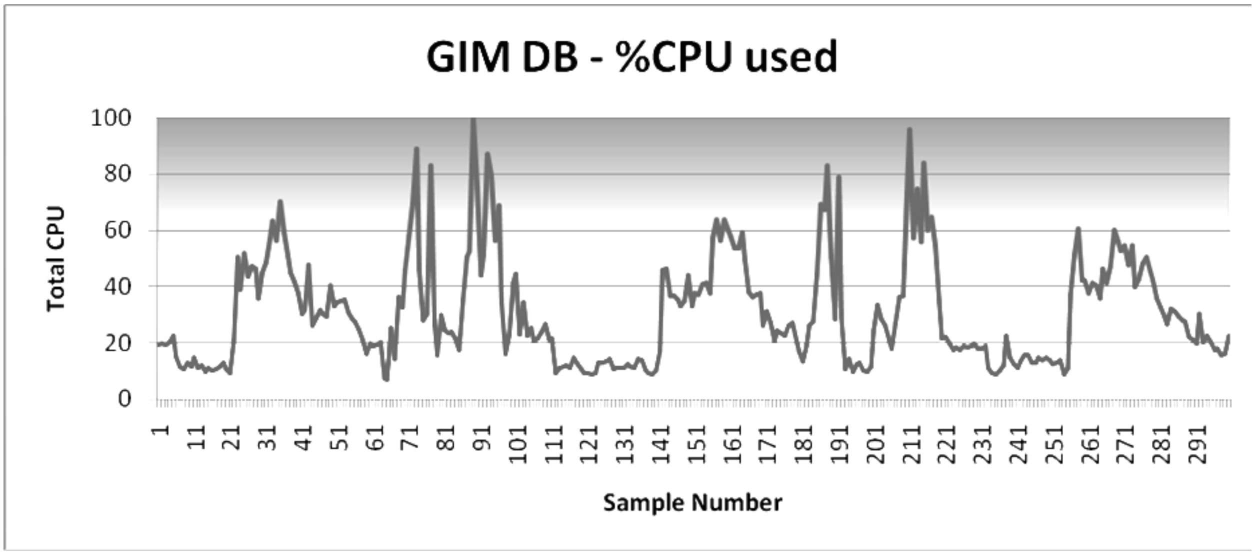


Figure 57: CPU Usage—GIM Node During Recovery Following Ten-Hour Outage

ICON Nodes There was no significant difference in CPU use between recovery and steady-state operations on the Oracle nodes for the IDB schemas.

Oracle RAC—Disk I/O

Figure 58 on page 153 shows combined I/O activity for all four nodes in the cluster as reported by the Oracle Enterprise Manager, broken down by type of database activity, for a one-hour period during recovery from the ten-hour outage.

As Figure 58 shows, I/O load was in bursts, which coincide with the ETL jobs. Predictably, the majority of the activity was from the DBWR process and the LGWR process, which writes asynchronously to the transaction log.

Combined cluster I/O peaked at about 320 MB per second. Traffic was divided between the two network-attached storage arrays, with approximately 200 MB per second of data traffic on one array and 120 MB per second of log traffic on the second array.

I/O requests during recovery were about double the number of requests during steady-state operations.

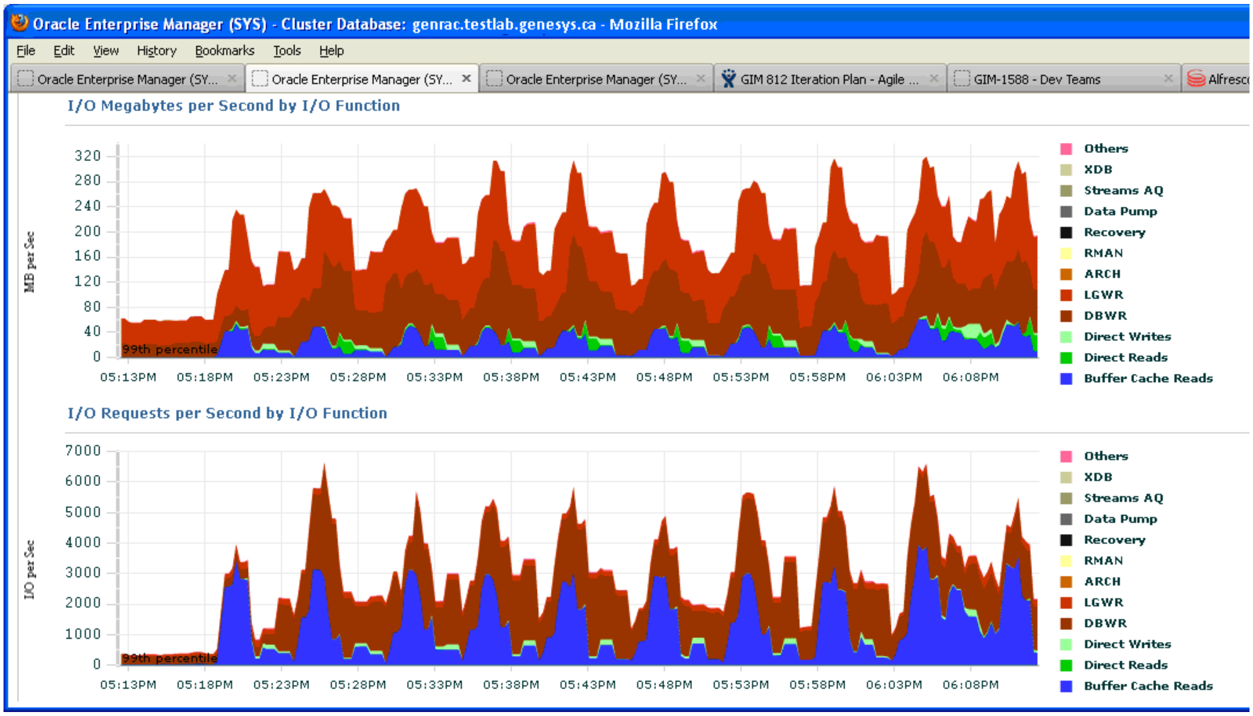


Figure 58: Total I/O Activity for the Oracle Cluster During Recovery Following Ten-Hour Outage

Genesys Info Mart 8.0 Performance

This section provides general background about the Genesys Info Mart 8.0 performance testing.

About Genesys Info Mart 8.0 Performance Testing

The purpose of the Genesys Info Mart 8.0 performance testing was twofold:

- Baseline—An endurance test to validate the ability of Genesys Info Mart to sustain a seven-day run in a voice-only environment, with performance equal to or better than Genesys Info Mart 7.6 in terms of data volumes and latency, ETL execution times, acceptable CPU utilization, and database performance. Genesys Info Mart was deployed on a Windows platform with an Oracle 11g database.
- Benchmark—A capacity test to validate the ability of Genesys Info Mart to sustain performance in an environment that included voice, e-mail, and chat (“blended environment”), and to determine the maximum interaction volume that Genesys Info Mart can handle. Genesys Info Mart was deployed on a Solaris platform with an Oracle 10g database.

The ETL cycle was run on actively populated source data, which was continuously generated for a high-complexity call flow. For details about the call flows, see “Call Flows—Baseline” on page 158 and “Interaction Flows—Benchmark” on page 173.

Genesys Info Mart ran without interruption for more than seven days. However, the detailed Genesys Info Mart 8.0 performance test results that are reported in this document relate only to the last three days of the test runs.

Note: Genesys Info Mart 8.0 performance testing was combined with testing of GI2 release 8.0. Test results for Genesys Info Mart 8.0 report on activity that included aggregation and generation of intraday GI2 reports.

For detailed information about the associated test results for GI2 release 8.0 performance, see Chapter 9 on page 201.

Release 8.0 Performance Testing Configuration

This subsection describes the ICON, user data, and Info Mart database configuration settings that were common to both the baseline and benchmark performance test environments. For additional information about the configuration of the test environments, see “Baseline Performance Test Setup” on page 157 and “Benchmark Performance Test Setup” on page 172, respectively.

Interaction Concentrator Application

Table 51 lists important ICON release 8.0 configuration options that were used for the testing.

The table does not include the options in the [filter-data] section, which enable selective population of IDB data. Genesys Info Mart supports some filtering of IDB data. However, the ICON data-filtering options were not enabled for the baseline performance tests. For the benchmark performance tests, the data-filtering options were enabled. For information about the data-filtering options that Genesys Info Mart supports, see the chapter about preparing Interaction Concentrator in the *Genesys Info Mart 8.x Deployment Guide.*

In addition to the options listed in Table 51, the ICON attached data specification file (adata_spec.xml) was modified to enable ICON to store both predefined and custom attached data. For more information, see “User Data Configuration” on page 155.

Table 51: ICON 8.0 Application Settings

Section	Option	
	Value Used	**Default Value**
callconcentrator	adata-extensions-history=none adata-reasons-history=none adata-userdata-history=none calls-in-the-past=true* dss-no-data-tout=60 gls-active-reason-codes=true* om-force-adata=true* partition-type=2* role=cfg (for the Configuration details ICON) role=gcc,gud,gls (for the Voice/Multimedia details ICONs) use-dss-monitor=true* *Mandatory option—Genesys Info Mart will not function if the option value is not set as specified.	adata-extensions-history=none adata-reasons-history=none adata-userdata-history=none calls-in-the-past=false dss-no-data-tout=300 gls-active-reason-codes=false om-force-adata=false partition-type=0 role=all use-dss-monitor=false
custom-states	EventData=char,postcallKVP1,char,postcallKVP2 store-event-data=conf	EventData= (no default value) store-event-data=none
ICON-Related Options on the Switch Object		
gts	gls-enable-acw-busy=0	gls-enable-acw-busy=1

User Data Configuration

ICON and Genesys Info Mart were configured to store a total of 14 user data key-value pairs (KVPs). In addition to the KVPs for which storage is predefined in the INTERACTION_DESCRIPTOR table, custom user data included five high-cardinality KVPs and five low-cardinality KVPs.

Approximately 100 bytes of user data was attached to all interactions early in the interaction flow. In addition, a small percentage of agents attached user data during after-call work (ACW). For more information about when user data was attached for the baseline and benchmark tests, see "Call Flows—Baseline" on page 158 and "Interaction Flows—Benchmark" on page 173, respectively.

Table 52 shows the user data that was stored in the Info Mart database.

Table 52: User Data Stored in Info Mart

Info Mart Table	Key	Cardinality
IRF_USER_DATA_CUST_1	HighData1	1000
	HighData2	2000
	HighData3	3000
	postcallKVP1	5
	postcallKVP2	5
USER_DATA_CUST_DIM_1	AgentAction	8
	AgentLogout	2
	CurrentTenant	2
	Destination	
	DistributionMethod	2
INTERACTION_DESCRIPTOR	CustomerSegment	
	ServiceType	2
	ServiceSubType	2
	Business Result	10

Info Mart Database

The Info Mart database was partitioned, using the default partition sizes of 24 hours for GIDB tables and 7 days for dimensional model fact tables. Partitioning enabled database maintenance to be streamlined.

Database links were configured between the RDBMS servers that hosted the Info Mart and IDB database schemas. Database links supported performance of the extraction job by enabling data to be copied directly from IDB into the Info Mart database, instead of through Genesys Info Mart memory.

For additional information about the database settings that were used for the Info Mart database and IDBs, see “Oracle Database Settings—Baseline” on page 162 and “Oracle Database Settings—Benchmark” on page 177.

Release 8.0 Baseline Performance Test

The baseline performance testing measured the ability of Genesys Info Mart 8.0 to process more than 2 million calls per day for a sustained period, including generation of intraday GI2 reports. With voice calls being generated in a variable daily call pattern for nine days, the ETL jobs were run on a 3-minute cycle. Aggregation ran continuously throughout the day.

For information about the hardware, software, and call flows that were used for the performance tests, see "Baseline Performance Test Setup" on page 157.

For the detailed performance test results for Genesys Info Mart, see "Baseline Performance Test Results" on page 163.

For the detailed performance test results for GI2, see "Release 8.0 Report Performance—Baseline" on page 203.

Baseline Performance Test Setup

Testing was performed on a Windows Server 2003 64-bit platform, with the Info Mart database and IDBs deployed on Oracle. To avoid known issues with Oracle 10g (for example, NOWAIT issues), Oracle 11g was used.

Configuration Settings

For information about the Interaction Concentrator application settings that were used for the release 8.0 performance tests, see "Interaction Concentrator Application" on page 154. The following subsections provide additional details about the environment that was used for the baseline tests:

- Call Volumes—Baseline, page 157
- Call Flows—Baseline, page 158
- Hardware Architecture—Baseline, page 159
- Product Versions—Baseline, page 160
- Genesys Info Mart Application—Baseline, page 161
- Oracle Database Settings—Baseline, page 162

Call Volumes—Baseline

The call pattern was based on a typical call center for a 24-hour period, with skills-based routing used to route voice interactions. The peak call rate was 40 calls per second, the average call rate was 28 calls per second, and the lowest call rate was 14 calls per second. A total of approximately 2 million inbound voice interactions were generated per day during testing. Figure 59 on page 158 shows the daily call pattern, in calls per second.

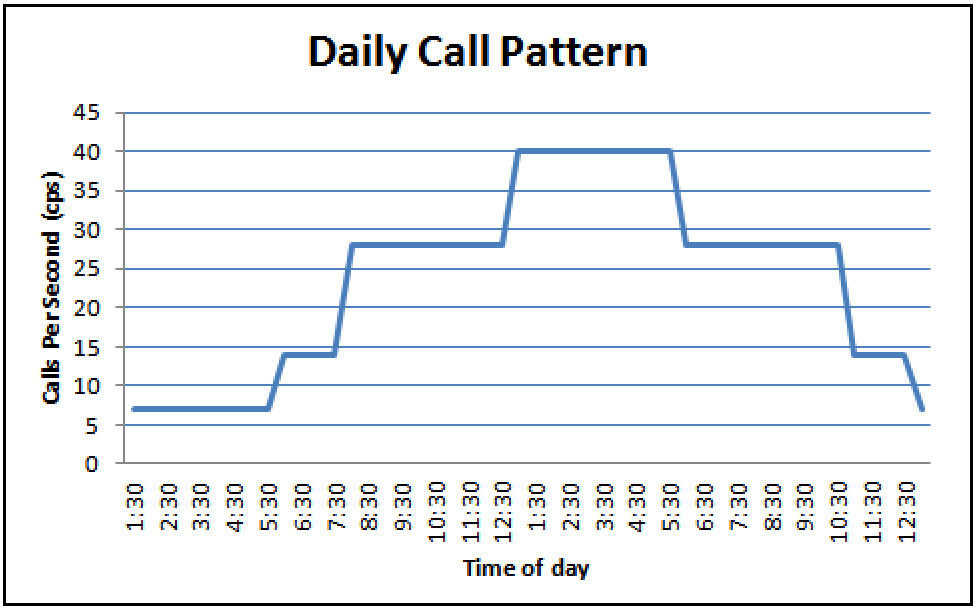


Figure 59: Daily Call Pattern—Baseline Performance Test

Call Flows—Baseline

Call Flow Summary

The fairly complex call flows provided a large number of ISCC calls and used a mixture of skills- and queue-based routing to target agents.

Seventy percent of calls were handled by only one agent and terminated. The remaining 30 percent of calls involved more than one agent, with transfer or conference scenarios.

Caller and agent behavior was randomized. Behavior included arbitrary hangups at random points in the call progress, no answer, and different forms of transfers. Agents' shifts were also factored in.

Call Flow Details

In more detail, the voice call flow was as follows:

- A T-Server simulator sent a call to a local target Routing Point (6000) on each switch.
- A strategy on the Routing Point attached user data.
- The call was routed to a Queue Group that contained queues (8001) from all the switches. In 98 percent of cases, the call becomes inter-site (using ISCC) at this leg.
- The call was diverted to an IVR port (7xxx).
  - Five percent of calls were released after one minute of talk time on the IVR.
  - The remaining 95 percent of calls were transferred to the local 6001 Routing Point for further routing.

- Of the calls that were transferred to the Routing Point:
  - Seventy percent were handled by the targeted agent. (Routing used three skill levels as well as virtual agent groups.)
  - Twenty percent were sent to a queue, and the available agent handled the call.
  - Ten percent were abandoned (sent to an "abandoned queue" and terminated).
- In the case of the calls that were handled by agents:
  - Thirty percent were transferred to another skill-based agent by single- or two-step transfers.
  - Thirty percent of the agents added user data in after-call work (ACW) mode.

Hardware Architecture—Baseline

Four high availability (HA) pairs of Interaction Concentrator instances monitored release 8.0 T-Servers to generate redundant sets of Voice details, and one HA pair of Interaction Concentrator instances monitored Configuration Server to generate a redundant set of Configuration details. (An *Interaction Concentrator instance* consists of an ICON server and the IDB it populates.)

Figure 60 shows the specifications for the servers that hosted Interaction Concentrator, Genesys Info Mart, and GI2 software components.

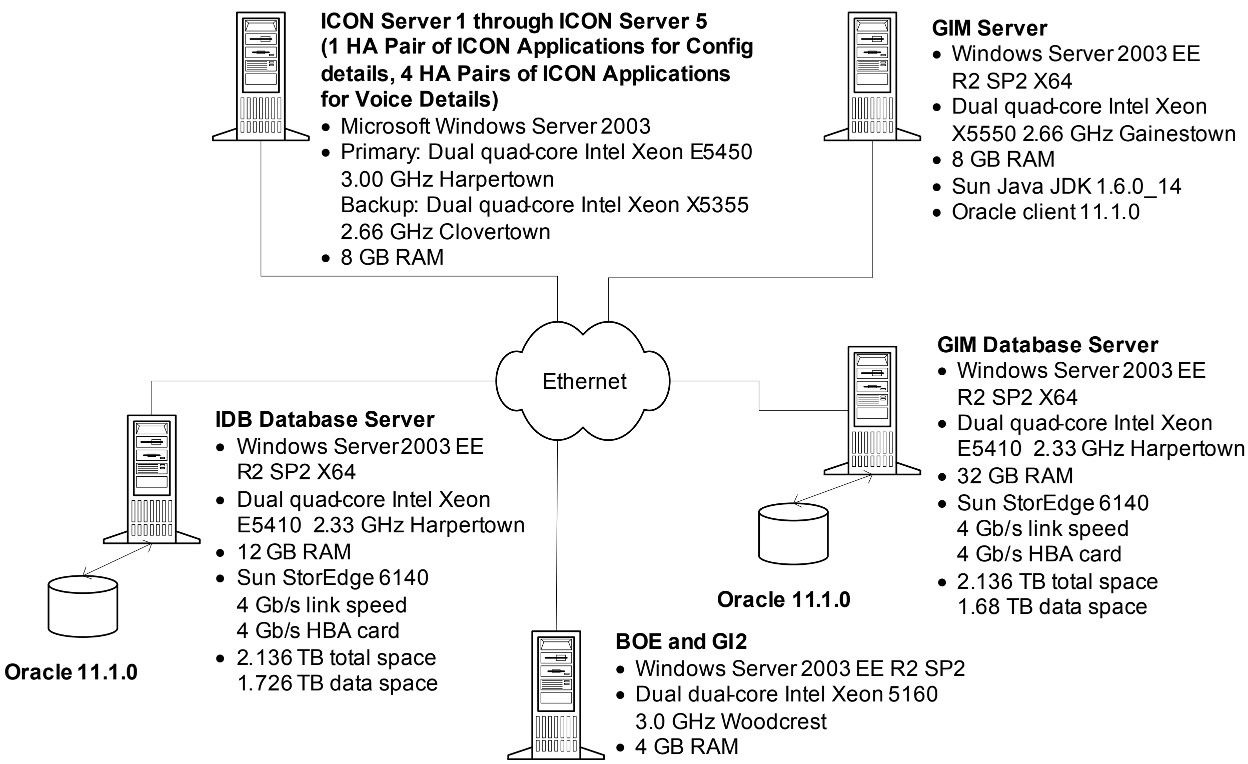


Figure 60: Hardware Architecture—Baseline Performance Test

Product Versions—Baseline

Table 53 lists the versions of Genesys Info Mart, GI2, and other supporting software products that were used in the testing.

Table 53: Product Versions Tested

Genesys Software	Version (for Windows 2003 Server)
Genesys Products	
Interaction Concentrator	8.0.000.30 **Note:** ICON was upgraded to release 8.0.000.35 during the testing.
Genesys Info Mart	8.0.000.65 **Note:** The performance testing used a prerelease version of the software. The testing did not cover certain capabilities that were included in the initial general release of Genesys Info Mart 8.0.
Genesys Interactive Insights (GI2)	8.0.000.20

Table 53: Product Versions Tested (Continued)

Genesys Software	Version (for Windows 2003 Server)
Reporting and Analytics Aggregates (RAA)	8.0.000.25 build 1
T-Server for Avaya Communication Manager	8.0.006.01
Third-Party Products	
Java Development Kit (JDK) on the Genesys Info Mart application server	1.6.0_14
Oracle client on the Genesys Info Mart application server Updated .jar files on the Oracle client	11.1.0 ojdbc18n.jar ojdbc6.jar
Business Objects	Enterprise 3.1 SP3 Fix Pack 1.8

Genesys Info Mart Application—Baseline

Table 54 lists those Genesys Info Mart release 8.0 configuration options that were set to non-default values for the testing. The table includes only options that affected test results. Either the default values were used for all the other Genesys Info Mart application options or else, where non-default settings were used, the options relate to functionality or resource usage that would not have affected test results. For options that affect aggregation, "default values" refers to the required values that are specified in the RAA and GI2 documentation.

Table 54: Genesys Info Mart 8.0 Application Settings

Section	Option	
	Value Used	Default Value
error-policy	error-policy-irf-exception=resume	error-policy-irf-exception=log_db_resume

Table 54: Genesys Info Mart 8.0 Application Settings (Continued)

Section	Option	
	Value Used	**Default Value**
gim-etl	days-to-keep-active-facts=3 days-to-keep-gidb-facts=3 days-to-keep-gim-facts=7 extract-data-cfg-facts-chunk-size=-1 extract-data-chunk-size=1800 extract-data-stuck-threshold=26000 max-call-duration=5400	days-to-keep-active-facts=600 days-to-keep-gidb-facts=14 days-to-keep-gim-facts=400 extract-data-cfg-facts-chunk-size=90000 extract-data-chunk-size=900 extract-data-stuck-threshold=28860 max-call-duration=3600
gim-etl-populate	populate-mm-ixn-queue-facts=true	populate-mm-ixn-queue-facts=false
schedule	etl-end-time=23:59 etl-frequency=3 etl-start-time=01:00 maintain-start-time=00:10 run-aggregates=TRUE run-scheduler=TRUE **Note:** In addition, options to schedule the aggregation job were set so that aggregation ran continuously from 01:00 until 23:00 daily.	etl-end-time=22:00 etl-frequency=1 etl-start-time=06:00 maintain-start-time=03:00 run-aggregates=FALSE run-scheduler=FALSE
gim-transformation	irf-io-parallelism=8	irf-io-parallelism=4

JVM Parameters—Baseline

The Java memory setting and other parameters for the Genesys Info Mart Server process in the gim_etl_server.bat file were set to the following:

-Xmx6000m -Xms32m -Xss512k -XX:GCTimeRatio=9
-Dirf-ignore-removeObserver-exception=true
-Doracle.jdbc.mapDateToTimestamp=true
-Doracle.jdbc.J2EE13Compliant=true -Duser.country=US
-Duser.language=en -Djava.library.path="%Genesys Info Mart_LIBPATH%" -
Duser.timezone=GMT com.genesyslab.gim.etl.server.Genesys Info MartServer %*

Oracle Database Settings—Baseline

Table 55 lists non-default Oracle settings that were used for the Info Mart database and IDBs in the baseline performance test.

Table 55: Non-Default Oracle Settings—Baseline Performance Test

Info Mart Database	IDBs
The Database Resource Manager was de-activated: ALTER SYSTEM SET resource_manager_plan=" SCOPE=BOTH;	
Statistics collection for the Info Mart schemas was scheduled for a non-busy time (1:30 a.m.).	Statistics were not collected for the IDB schemas. The daily maintenance window was disabled.
audit_trail = NONE cursor_sharing = exact db_writer_processes = 2 session_cached_cursors = 2000 undo_retention = 500 compatible = 11.1.0.0.0 db_block_size = 16384 db_recovery_file_dest_size = 2G open_cursors = 2000 pga_aggregate_target = 6G processes = 600 remote_login_passwordfile = EXCLUSIVE sga_target = 22G undo_tablespace = UNDOTBS1	audit_trail = DB cursor_sharing = EXACT db_writer_processes = 2 session_cached_cursors = 1200 compatible = 11.1.0.0.0 db_block_size = 16384 db_recovery_file_dest_size = 2G memory_target = 10G open_cursors = 1200 processes = 1000 remote_login_passwordfile = EXCLUSIVE undo_tablespace = UNDOTBS1

Baseline Performance Test Results

The baseline performance test was successful, based on the following criteria:

- The endurance run was conducted for nine consecutive days. The Genesys Info Mart and GI2 applications ran without interruption and were stable for the entire duration of the test.
- Genesys Info Mart was able to sustain the test plan call volume and support intraday reporting.

  Within 30 minutes of call termination, all the generated calls were loaded into the Info Mart database and aggregated, with aggregates also stored in the Info Mart database.

The following subsections provide detailed measurements of Genesys Info Mart performance during the last three days of the test run.

Performance of Genesys Info Mart Jobs

The primary measure of Genesys Info Mart performance is the amount of time it takes for jobs to execute (the *duration* of the job or cycle). This subsection presents detailed results for the ETL and maintenance jobs.

Extraction Rates

Aside from two anomalous ETL cycles (see "Effect of Execution Plans"), all the extraction cycles pulled T-Server records well within 30 minutes of the T-Server timestamps. On average, the extraction job started 2 minutes after the leading edge of the extraction window. (For example, an extraction cycle that started at 2:00 p.m. was likely extracting interaction records with a maximum timestamp of 1:58 p.m.)

ETL Job Durations

ETL Cycle

Figure 61 on page 164 depicts the durations of the ETL cycles during the last three days of the endurance test. The average time for an ETL cycle was 307 seconds. During the periods of high call volumes (see "Call Volumes—Baseline" on page 157), the time for an ETL cycle to complete typically reached 400 seconds.

Effect of Execution Plans

As Figure 61 shows, two spikes were observed (cycle 237 and cycle 1004). The anomalously high durations were the result of poor execution plans.

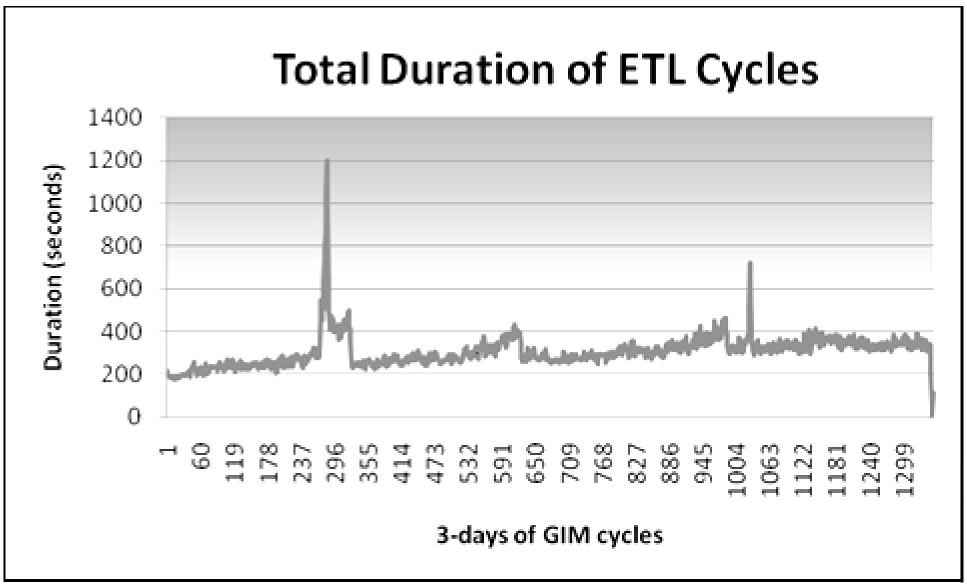


Figure 61: Duration of ETL Cycles—Baseline Performance Test

Extraction and Transformation Jobs Separately

Figure 62 and Figure 63 on page 165 depict the durations of the extraction and transformation jobs, respectively, during the last three days of the endurance test. The results indicate that:

- On average, the extraction job took twice as long as the transformation job.
- The poor execution plans that were used in cycles 237 and 1004 had proportionally more of an impact on the transformation job than on the extraction job.
- The extraction job exhibited a trend for the job duration to increase gradually over time. The performance of the extraction job degrades as the size of IDB grows.

- The duration of the transformation job remained relatively constant. Many of the Staging tables in the Info Mart database are truncated at the end of each ETL cycle, so the size of the Staging tables does not become a factor in ongoing performance.

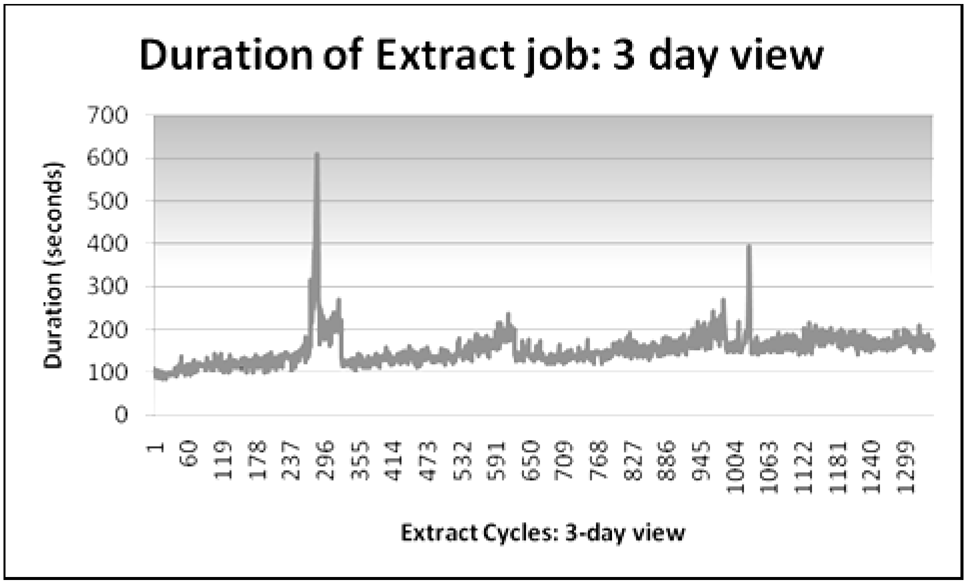


Figure 62: Duration of Extraction Job—Baseline Performance Test

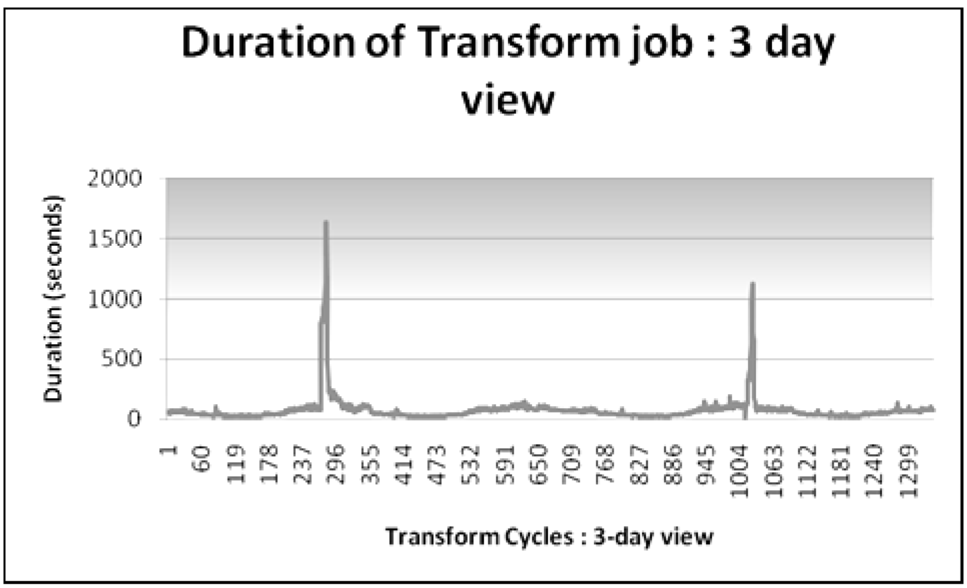


Figure 63: Duration of Transformation Job—Baseline Performance Test

Extraction and Transformation Jobs by Call Rate

Figure 64 on page 166 compares durations of the extraction and transformation jobs, based on call rates. The chart shows the average durations of the ETL cycles for the timespans in which calls were generated at 7 cps, 14 cps, 28 cps, and 40 cps, respectively. The results indicate that:

- The call rate had a linear impact on the average duration of the extraction job. The average extraction time for calls generated at 40 cps was approximately 12 percent longer than for calls generated at 28 cps.
- There was no significant difference in the average duration of the transformation job at the lower call rates. Higher call rates had a linear impact. The increase in transformation times between the 14 cps, 28 cps, and 40 cps cycles was proportional to the increase in call rates.

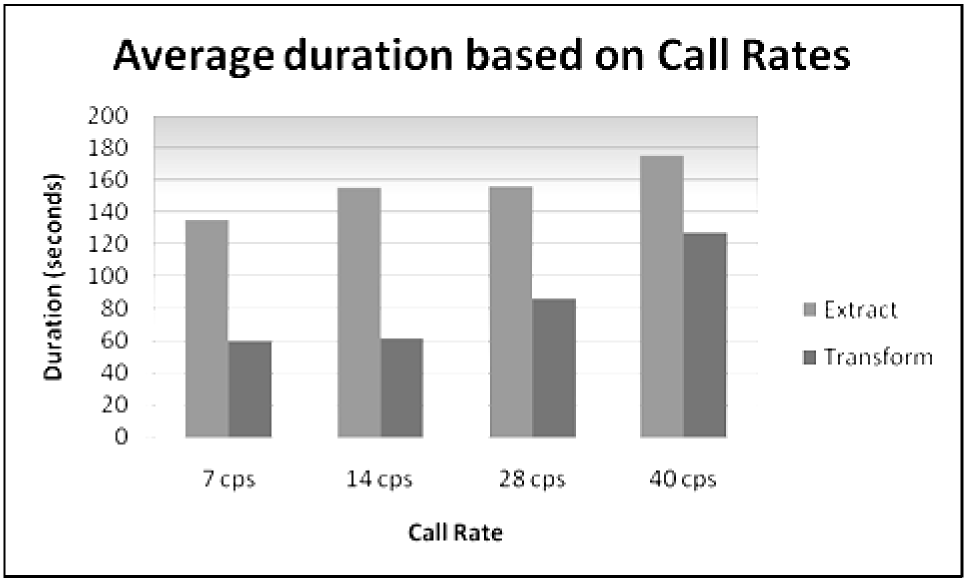


Figure 64: Duration of ETL Jobs by Call Rate—Baseline Performance Test

Maintenance Job Duration

Figure 65 depicts the durations of the daily maintenance job (Job_MaintainGIM) during the last three days of the endurance test. Because the Info Mart database was partitioned, with the size of partitions in GIDB set to the default 86400 seconds (24 hours), a day's worth of extracted ICON call data in GIDB was purged very quickly at each run of the daily maintenance job. The job duration was consistent for all three days.

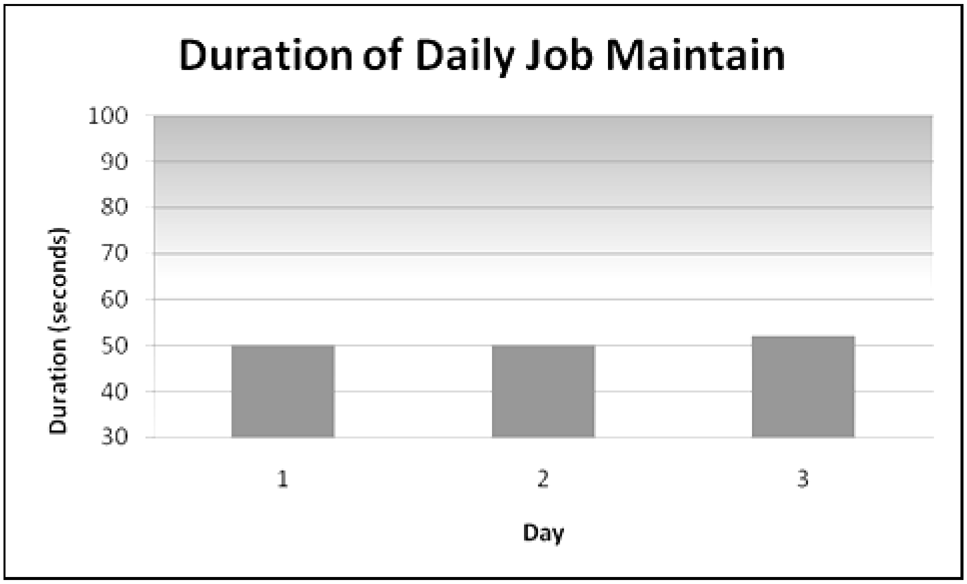


Figure 65: Duration of Maintenance Job—Baseline Performance Test

Hardware Resource Usage

This subsection provides information about hardware resource usage for the Genesys Info Mart application and the Info Mart and Interaction Concentrator databases during one day of the endurance run.

For details about the hardware that was used in the baseline performance test, see Figure 60 on page 160.

Genesys Info Mart Server

The best measure of Genesys Info Mart application performance is the amount of time that the jobs take, as described in "Performance of Genesys Info Mart Jobs" on page 163. Genesys did not extensively measure Genesys Info Mart Server performance in terms of hardware resource usage, beyond verifying acceptable CPU utilization.

CPU Utilization

During the busiest ETL cycle, the Genesys Info Mart application used less than 50 percent of the CPU on the server.

Memory Utilization

The Java engine limited memory use to a maximum of 6000 MB, in accordance with the Java memory setting in the ETL Server batch file (see "JVM Parameters—Baseline" on page 162).

Info Mart Database

CPU Utilization

Figure 66 depicts the CPU usage on the Info Mart database engine during one day of the endurance run. The Oracle server was hosted on an eight-core machine. Figure 66 shows that, during the busiest part of the day, the Info Mart database used less than one-eighth of the CPU resources on the Oracle server.

The chart shows CPU utilization by Oracle (which is multithreaded) expressed in terms of core usage by a single-threaded application.

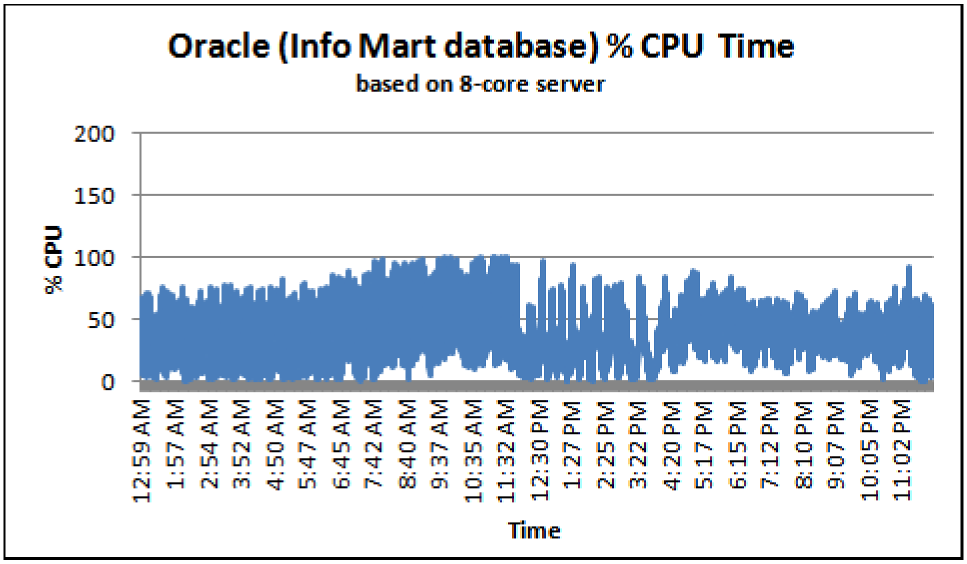


Figure 66: Oracle CPU Usage—Info Mart Database

Memory Utilization The Info Mart database used an average of 23.06 GB of RAM. The maximum amount of RAM used was 24.38 GB.

I/O Operations Table 56 reports Info Mart database performance in terms of input and output (I/O) rates—the average and maximum size and number of disk read and write operations per second.

Table 56: I/O Operations per Second—Info Mart Database

Measure	Disk Reads per Second		Disk Writes per Second	
	MB	Number	MB	Number
Average	14.68	155	14.80	211
Maximum	298.78	3610	153.79	4902

IDB

CPU Utilization Figure 67 depicts the CPU usage on the IDB database engine during one day of the endurance run.

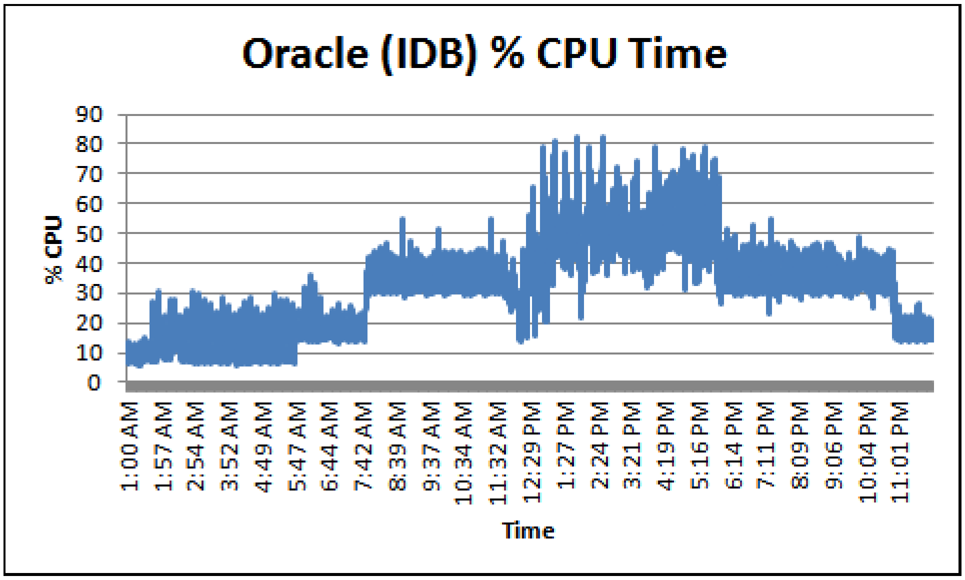


Figure 67: Oracle CPU Usage—IDB

Memory Utilization

IDB used an average of 8.14 GB of RAM. The maximum amount of RAM used was 8.29 GB.

I/O Operations

Table 57 reports IDB database performance in terms of I/O rates—the average and maximum size and number of disk read and write operations per second.

Table 57: I/O Operations per Second—IDB

Measure	Disk Reads per Second		Disk Writes per Second	
	MB	Number	MB	Number
Average	5.05	287	27.93	330
Maximum	194.87	5043	77.18	5475

ICON

Four HA pairs of Interaction Concentrator servers (ICONs) were deployed to manage the call details of the system. An additional HA pair of ICONs managed configuration details.

CPU Utilization

During the majority of testing, CPU utilization was within acceptable ranges. CPU usage spiked during periods of higher call rates. No data issues were observed—for example, there was no persistent queue (PQ file) backlog.

Figure 68 depicts the CPU usage of one of the primary ICONs during one day of the endurance run. The average CPU of one ICON was 7 percent of a single CPU core, and the maximum was 29 percent.

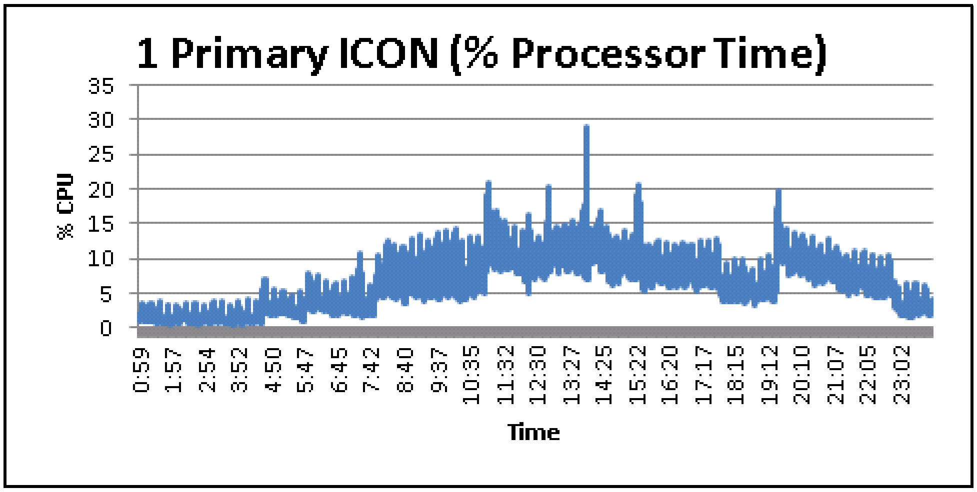


Figure 68: ICON CPU Usage

Memory Utilization

The primary ICON used an average of 0.24 GB of RAM. There was no significant difference between the average and maximum amount of RAM used.

Comparison of Call Records in IDB and Info Mart

Table 58 compares the number of call records in the Info Mart INTERACTION_FACT table that resulted from call details that were extracted from only one of the ICONs, which started collecting data a few days before the Genesys Info Mart ETL started processing.

At the end of the endurance run, there were approximately 55 million records in the Info Mart INTERACTION_FACT table, which represented the last seven days' worth of data.

Table 58: Number of Call Records in IDB and Info Mart

Day	Total Number of Records (millions)	
	IDB G_CALL Table	Info Mart INTERACTION_FACT Table
Day 1	1.51	
Day 2	1.24	
Day 3	0.71	
Day 4	0.40	
Day 5	1.46	1.72
Day 6	1.40	2.06

Table 58: Number of Call Records in IDB and Info Mart (Continued)

Day	Total Number of Records (millions)	
	IDB G_CALL Table	Info Mart INTERACTION_FACT Table
Day 7	1.59	2.03
Day 8	1.13	1.72
Day 9	1.04	1.49
Day 10	1.57	2.05
Day 11	1.58	2.05
Day 12	1.58	2.05
Day 13	1.00	1.03

Release 8.0 Benchmark Performance

The benchmark performance testing measured the ability of Genesys Info Mart 8.0 to sustain ETL and aggregation processing in a blended environment and to determine the maximum traffic that Genesys Info Mart can handle, with interaction flows that represent a typical Genesys customer base. The blended environment included voice and multimedia (e-mail and chat) interactions.

Genesys Info Mart and GI2 were able to process 8.6 million interactions per day and to support intraday reporting. With multimedia interactions being generated at a constant rate of 14 interactions per second, the call rate for voice calls was increased to a maximum rate of 85.5 calls per second, which was sustained for ten days.

The maximum traffic rate did not challenge the limits of Genesys Info Mart and GI2 processing capacity. Lack of infrastructure prevented Genesys from generating higher call rates to establish maximum capacity.

For information about the hardware, software, and call flows that were used for the performance tests, see “Benchmark Performance Test Setup” on page 172.

For the detailed performance test results for Genesys Info Mart, see “Benchmark Performance Test Results” on page 178.

For the detailed performance test results for GI2, see “Release 8.0 Report Performance—Benchmark” on page 211.

The Benchmark Performance Tests

Genesys performed the following benchmark tests:

- ETL-oriented benchmark tests: Obtain the maximum daily call volume for Genesys Info Mart and GI2 release 8.0 in a voice and multimedia environment.

  Starting from a total of 2.6 million interactions a day and with a fixed rate of 14 interactions per second for multimedia interactions, the call rate for voice calls was increased until it reached 85.5 calls per second, for a combined total of approximately 8.6 million voice and multimedia interactions a day. The constant call rate was sustained for ten days, with purging of the Info Mart database starting after seven days.
- Reporting-oriented benchmark tests: Vary the number and frequency of active GI2 reports on both an active Genesys Info Mart and a static Genesys Info Mart, to validate performance of both Genesys Info Mart and GI2 under load.

  For more information about the GI2 test scenarios and results, see "Release 8.0 Report Performance—Benchmark" on page 211.

Benchmark Performance Test Setup

Testing was performed with Genesys Info Mart deployed on a Solaris 10 64-bit platform, with the Info Mart database deployed on Oracle 10g. GI2 was deployed on Microsoft Windows Server 2003. For more details about the application and RDBMS servers that were used for Genesys Info Mart, GI2, and Interaction Concentrator, see "Hardware Architecture—Benchmark" on page 174.

Configuration Settings

For information about the Interaction Concentrator settings that were used for the release 8.0 performance tests, see "Interaction Concentrator Application" on page 154. The following subsections provide additional details about the environment that was used for the benchmark tests:

- Interaction Volumes—Benchmark, page 172
- Interaction Flows—Benchmark, page 173
- Hardware Architecture—Benchmark, page 174
- Product Versions—Benchmark, page 175
- Genesys Info Mart Application—Benchmark, page 176
- Oracle Database Settings—Benchmark, page 177

Interaction Volumes—Benchmark

The media contacts in the blended environment were voice, e-mail, and chat.

The daily pattern was flat:

- E-mail and chat interactions were generated at a constant rate of 14 interactions per second for the entire duration of the benchmark testing:
  - Approximately 7 e-mails per second
  - Approximately 7 chats per second
- The call rate for voice interactions started at 60 cps and increased to a maximum of 85.5 cps.

Interaction Flows—Benchmark

The Genesys configuration environment was organized as a single contact center under a single tenant.

A total of 32,000 agents were configured. On average, each agent had 30 skills.

All agents could handle voice calls, but only 4000 agents could also log in for the e-mail and chat media types. These agents could handle voice and multimedia concurrently. The benchmark tests were performed with only these 4000 agents logged in. Routing was performed only for e-mail and chat interactions.

Voice Call Flow

Call Flow Summary

The voice call flow encompassed nine Avaya T-Servers and Avaya switch simulators, with cross-switch routing between all of them. Each switch received external voice calls directly to a main routing point, from which the calls were evenly distributed across 100 routing points on the same switch (900 routing points in total). From these routing points, strategies used skill expressions, which could target an agent on any of the switches, to target agents.

Call Flow Detail

In more detail, the voice call flow was as follows:

- A dialer placed a call to a local target Routing Point (7999).
- A strategy on the Routing Point attached 100 bytes of attached data.
- A tenant was selected, and a DN Group that contained all of the IVR Queues (8001) for that tenant was targeted.
- The call was delivered to 8001 on a premise switch, where it was diverted to a Routing Point (7xxx).

  Eighty percent of calls were then released after one minute of talk time on the IVR.

  The remaining 20 percent were transferred to the local 8000 queue for agent handling.
- Of the calls that were transferred to the queue for agent handling:
  - Ninety percent were handled by the targeted agent.

- Ten percent were transferred by single-step, two-step, conference, or circular transfer (where *circular transfer* means that the call moves from one premise to another and then is transferred back).

Typical Call Activity

The call flow yielded the following typical call activity:

- Average talk time was 120 seconds.
- Thirty-five percent of calls ended in the IVR.
- Eighteen percent of calls were transferred at least once.
- Eight percent of calls were abandoned.

Multimedia Interaction Flows

Media routing used five business processes with many queues and strategies. The business processes used web service calls, ESP server calls, database dips, and skills-based agent targeting.

Attached Data

The first strategy in the interaction flow, which divided the interactions among the business processes, attached predefined customer and routing attached data, as well as certain custom KVPs. The next strategy, which obtained customer data, attached additional user data from an ESP server call. A subsequent strategy, which obtained agent skills, used the attached data from the ESP server call to perform a database dip to determine the appropriate segment skill, which was used by subsequent strategies to deliver the interaction to an agent; this strategy also attached additional user data.

Hardware Architecture—Benchmark

Three HA pairs of Interaction Concentrator instances monitored release 8.0 Premise T-Servers to generate redundant sets of Voice details, one HA pair of Interaction Concentrator instances monitored a release 8.0 Interaction Server to generate a redundant set of Multimedia details, and one HA pair of Interaction Concentrator instances monitored Configuration Server to generate a redundant set of Configuration details.

Figure 69 shows the specifications for the servers that hosted Interaction Concentrator, Genesys Info Mart, and GI2 software components during the benchmark performance tests.

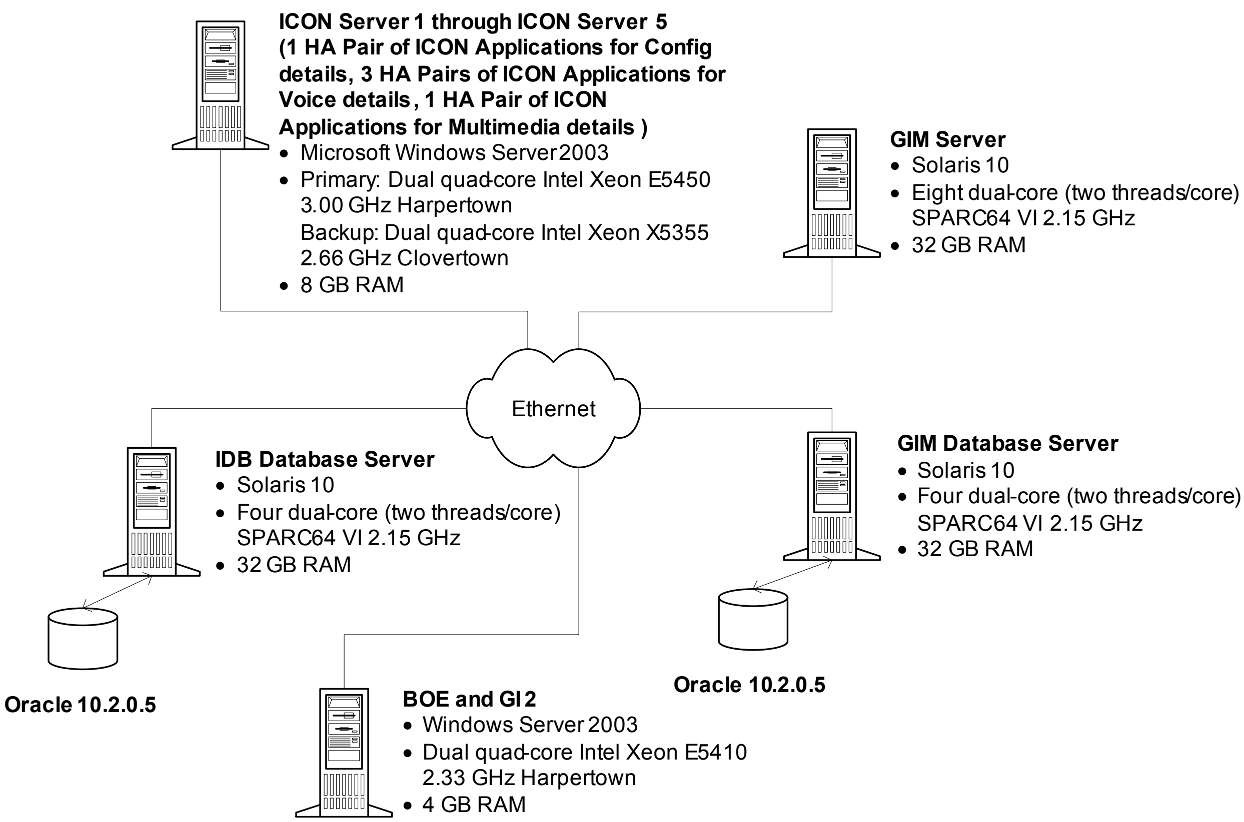


Figure 69: Hardware Architecture—Benchmark Performance Test

Product Versions—Benchmark

Table 59 lists the versions of Genesys Info Mart, GI2, and other supporting software products that were used in the testing.

Table 59: Product Versions Tested

Genesys Software	Operating System	Product Version
Genesys Products		
Interaction Concentrator	Windows Server 2003	8.0.000.35
Genesys Info Mart	Solaris 10	8.0.001.05
Genesys Interactive Insights (GI2)	Windows Server 2003	8.0
Reporting and Analytics Aggregates (RAA)	Windows Server 2003	8.0
T-Server for Avaya Communication Manager	Windows Server 2003	8.0

Table 59: Product Versions Tested (Continued)

Genesys Software	Operating System	Product Version
Interaction Server	Windows Server 2003	8.0
Universal Routing Server	Windows Server 2003	7.6
Third-Party Products		
Oracle client on the Genesys Info Mart application server	Solaris 10	10.2.0.5
Business Objects	Windows Server 2003	Enterprise 3.1 SP3 Fix Pack 1.8

Genesys Info Mart Application—Benchmark

Table 54 lists those Genesys Info Mart release 8.0 configuration options that were set to non-default values for the testing. The table includes only options that affected test results. Either the default values were used for all the other Genesys Info Mart application options or else, where non-default settings were used, the options relate to functionality or resource usage that would not have affected test results. For options that affect aggregation, "default values" refers to the required values that are specified in the RAA and GI2 documentation.

Table 60: Genesys Info Mart 8.0 Application Settings

Section	Option	
	Value Used	Default Value
error-policy	error-policy-irf-exception=resume	error-policy-irf-exception=log_db_resume
gim-etl	days-to-keep-gidb-facts=7 days-to-keep-gim-facts=7 extract-data-cfg-facts-chunk-size=-1 extract-data-stuck-threshold=26000 extract-data-thread-pool-size=64	days-to-keep-gidb-facts=14 days-to-keep-gim-facts=400 extract-data-cfg-facts-chunk-size=90000 extract-data-stuck-threshold=28860 extract-data-thread-pool-size=32
gim-etl-populate	populate-mm-ixn-queue-facts=true	populate-mm-ixn-queue-facts=false

Table 60: Genesys Info Mart 8.0 Application Settings (Continued)

Section	Option	
	Value Used	Default Value
schedule	etl-end-time=23:59 etl-frequency=3 etl-start-time=01:00 maintain-start-time=00:10 run-aggregates=TRUE run-scheduler=TRUE **Note:** In addition, options to schedule the aggregation job were set so that aggregation ran continuously from 01:00 until 23:00 daily.	etl-end-time=22:00 etl-frequency=1 etl-start-time=06:00 maintain-start-time=03:00 run-aggregates=FALSE run-scheduler=FALSE
gim-transformation	irf-io-parallelism=8	irf-io-parallelism=4

JVM Parameters—Benchmark

The Java memory setting and other parameters for the Genesys Info Mart Server process in the gim_etl_server.bat file were set to the following:

-Xmx5000m -Xms32m -Xss512k -XX:GCTimeRatio=9
-Dirf-ignore-removeObserver-exception=true
-Doracle.jdbc.mapDateToTimestamp=true
-Doracle.jdbc.J2EE13Compliant=true -Duser.country=US
-Duser.language=en -Djava.library.path="%GIM_LIBPATH%"
-Duser.timezone=GMT com.genesyslab.gim.etl.server.GIMServer %*

Oracle Database Settings—Benchmark

The following settings were specified in the Oracle parameters file for the Info Mart database:

 # Cache and I/O

db_block_size=16384

#db_file_multiblock_read_count=16

db_writer_processes=2

log_checkpoint_timeout=0

filesystemio_options=setall

fast_start_mttr_target=0

 # Job Queues

job_queue_processes=10

 # Miscellaneous

compatible=10.2.0.3.0

 # Processes and Sessions

processes=400

#sessions=225

 # Cursors and Library Cache

open_cursors=2000

session_cached_cursors=2000

cursor_sharing=force

 # Diagnostics and Statistics

background_dump_dest=/cti/oracle/admin/m4000/bdump

core_dump_dest=/cti/oracle/admin/m4000/cdump

user_dump_dest=/cti/oracle/admin/m4000/udump

 # File Configuration

control_files=("/cti/oracle/oradata/m4000/control01.ctl", "/cti/oracle/oradata/m4000/control02.ctl")

db_recovery_file_dest=/oradata/flash_recovery_area

db_recovery_file_dest_size=2147483648

 # SGA Memory

sga_target=25g

sga_max_size=25g

 # Security and Auditing

audit_file_dest=/cti/oracle/admin/m4000/adump

remote_login_passwordfile=EXCLUSIVE

 # Sort, Hash Joins, Bitmap Indexes

pga_aggregate_target=2g

 # System Managed Undo and Rollback Segments

undo_management=AUTO

undo_retention=500

undo_tablespace=UNDOTBS1

Benchmark Performance Test Results

The benchmark tests yielded the following performance results:

- At 8.6 million interactions per day in a blended environment with reasonably complex interaction flows and requirements for aggregation and intraday reporting, the limits of Genesys Info Mart and GI2 processing capacity were not reached. The maximum traffic rates achieved during testing were 85.5 cps for voice and 14 interactions per second for multimedia (e-mail and chat). The maximum rates were sustained for ten days without interruption.
- In almost all cases, the duration of the ETL cycle was less than 30 minutes. In fact, the average duration of the ETL cycle was below the configured frequency (etl-frequency=3). Therefore, the average frequency of the ETL cycle was 3 minutes. On average, the ETL jobs took 135 seconds to process 3 minutes' worth of data in each cycle.
- The times required to generate the GI2 reports, which were selected to represent the different sets of aggregate tables and which ran in various combinations and at varying frequencies, ranged from 30 seconds to 90 seconds. The GI2 report requests returned report sets that ranged from 1 page to 185 pages.

The following subsections provide detailed measurements of Genesys Info Mart performance during the last three days of the test run.

For detailed information about the GI2 reporting results, see "Release 8.0 Report Performance—Benchmark" on page 211.

Performance of Genesys Info Mart Jobs

The durations of the Genesys Info Mart jobs are the primary measure of Genesys Info Mart performance. This subsection presents detailed results for the ETL and maintenance jobs.

ETL Job Durations

Genesys measured the durations of the extraction and transformation jobs at various interaction volumes. With a fixed rate of 14 multimedia interactions per second, the volume of daily interactions was increased by increasing the rate of voice calls—60 cps, 70 cps, 80 cps, 85.5 cps. The call rates remained constant over an entire day.

At the maximum daily interaction volume (voice calls at 85.5 cps), the average duration of the ETL cycle was 135 seconds:

- The average duration of the extraction job was 50 seconds.
- The average duration of the transformation job was 85 seconds.

ETL Cycle

Figure 70 depicts the average durations of the ETL cycles with the voice call rate at 60 cps, 70 cps, 80 cps, and 85.5 cps, respectively.

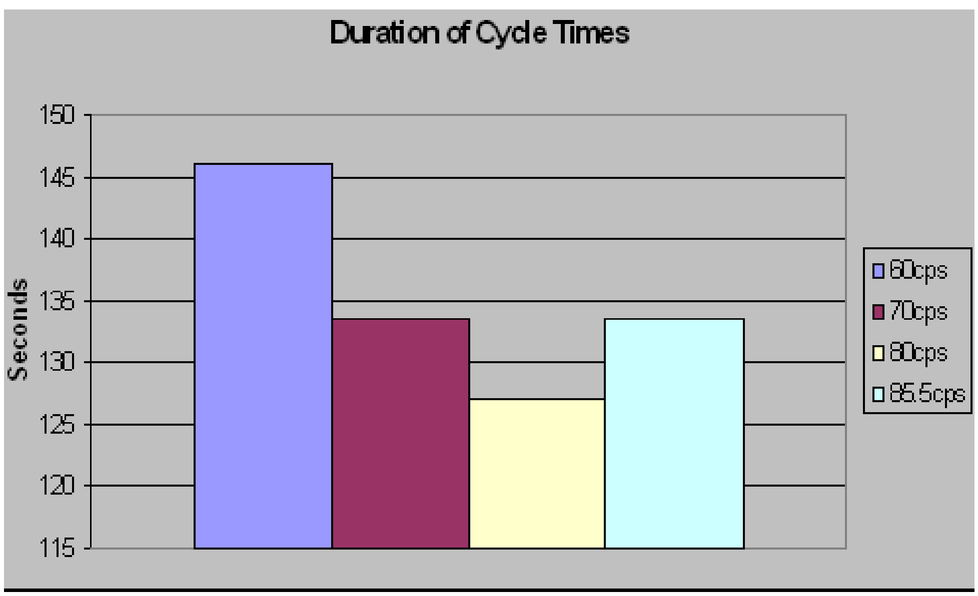


Figure 70: Average Duration of ETL Cycles at Various Call Rates

Extraction and Transformation Jobs Separately

Figures 71 through 74 depict the average and maximum durations of the extraction and transformation jobs with the voice call rate at 60 cps, 70 cps, 80 cps, and 85.5 cps, respectively.

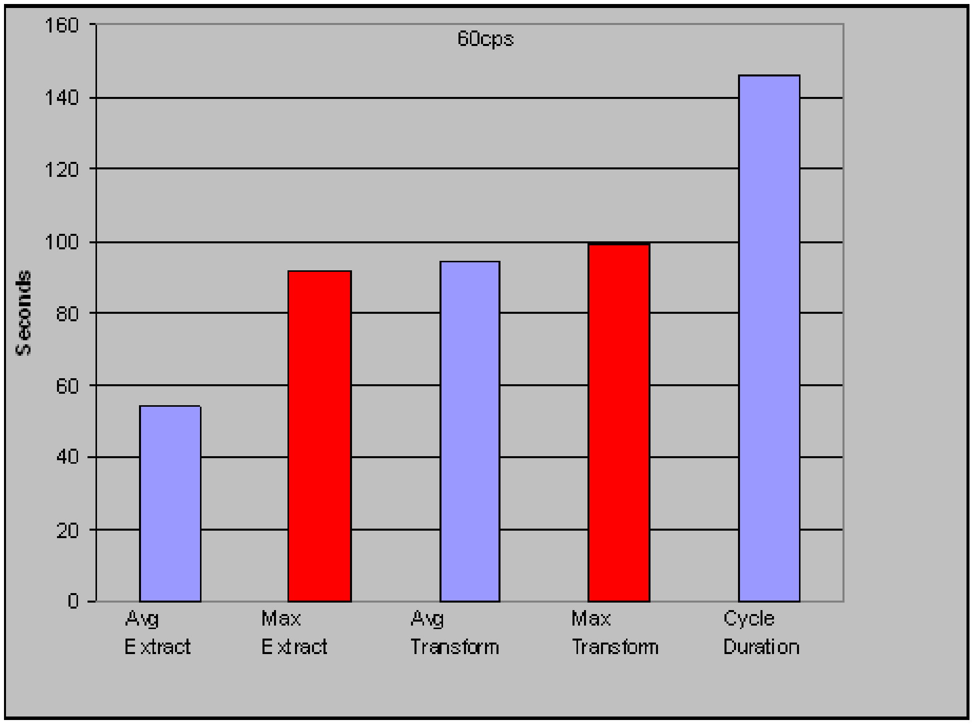


Figure 71: Extraction and Transformation Times—60 cps

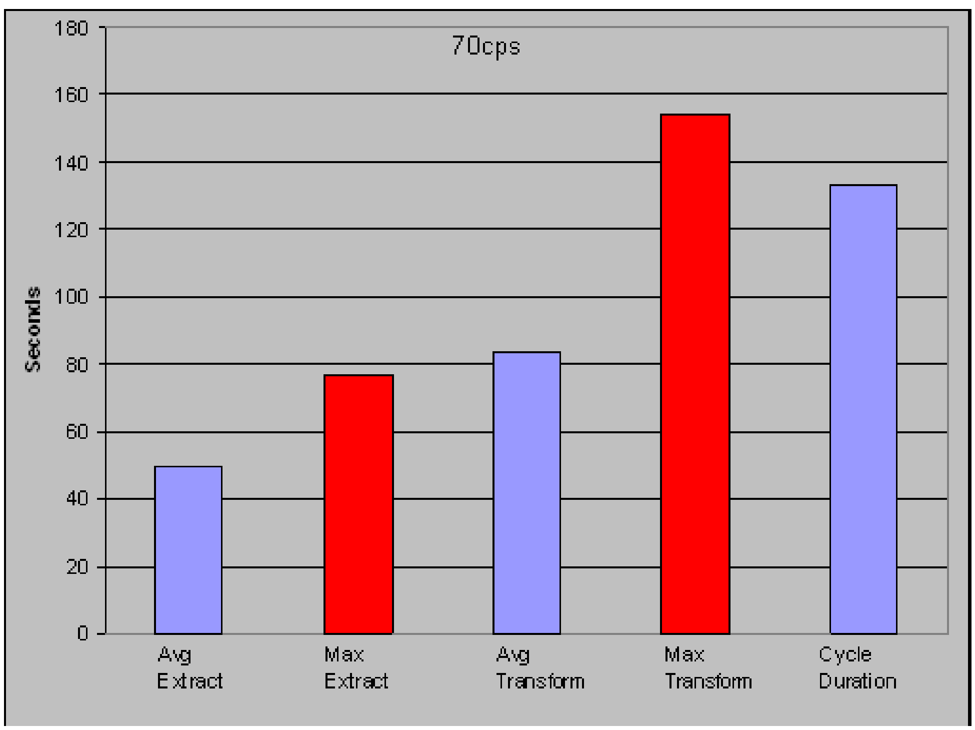


Figure 72: Extraction and Transformation Times—70 cps

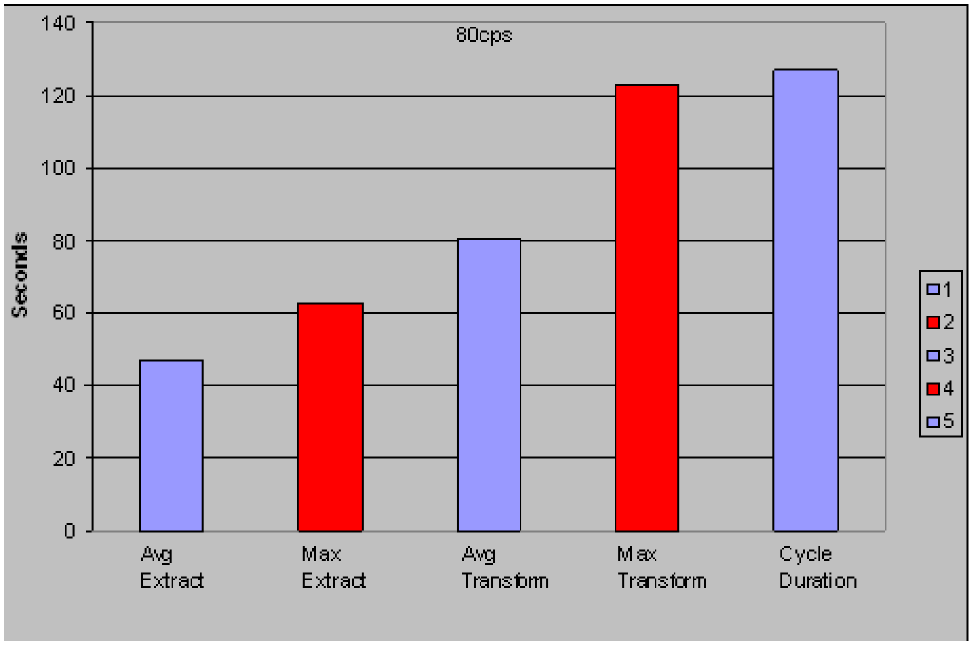


Figure 73: Extraction and Transformation Times—80 cps

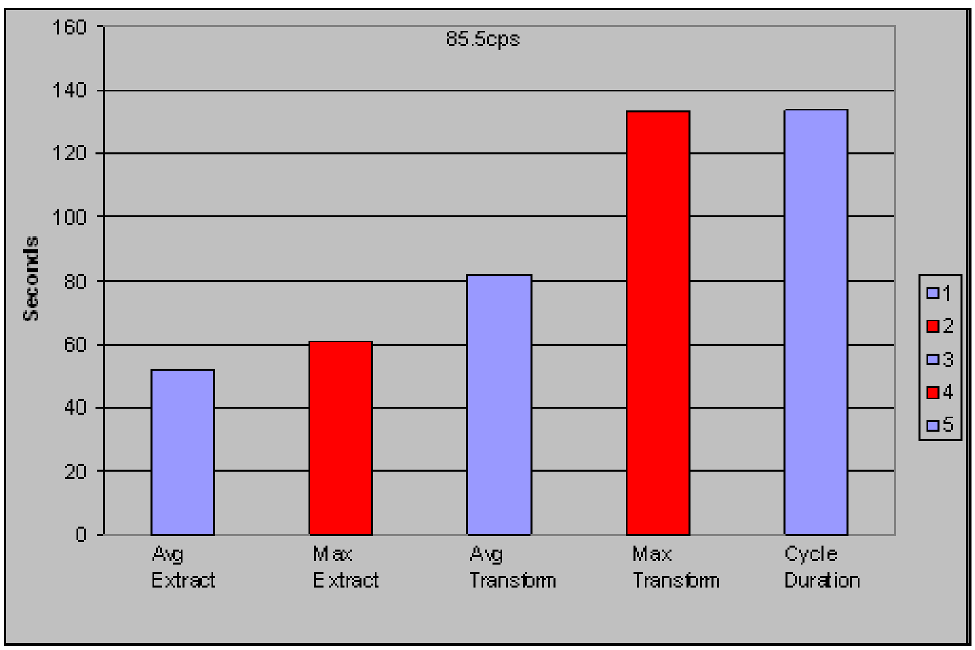


Figure 74: Extraction and Transformation Times—85.5 cps

Maintenance Job Duration

Figure 75 depicts the durations of the daily maintenance job (Job_MaintainGIM) on four days, during which the voice call rate was set at 60 cps, 70 cps, 80 cps, and 85.5 cps, respectively.

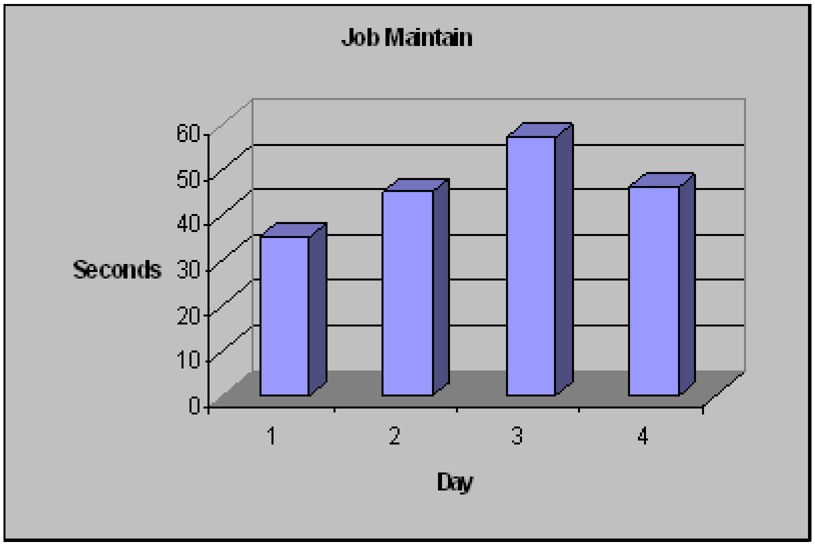


Figure 75: Duration of Maintenance Job—Benchmark Performance Test

Hardware Resource Usage

This subsection provides information about hardware resource usage for the Genesys Info Mart application and the Info Mart and Interaction Concentrator databases during one day of the benchmark run at the maximum daily interaction volume.

For details about the hardware that was used in the benchmark performance tests, see Figure 60 on page 160.

Genesys Info Mart Server

The best measure of Genesys Info Mart application performance is the amount of time that the jobs take, as described in "Performance of Genesys Info Mart Jobs" on page 179. Genesys did not extensively measure Genesys Info Mart Server performance in terms of hardware resource usage, beyond verifying acceptable CPU utilization.

CPU Utilization During the busiest ETL cycle, the Genesys Info Mart application used less than 40 percent of the CPU on the server.

Memory Utilization The Java engine limited memory use to a maximum of 5000 MB, in accordance with the Java memory setting in the ETL Server batch file (see "JVM Parameters—Benchmark" on page 177).

Info Mart Database

CPU Utilization Figure 76 depicts the CPU usage for one core on the Info Mart database engine during one day of the benchmark run at the maximum daily interaction volume. The Oracle server was hosted on an eight-core machine. The Info

Mart database engine used a maximum of 39.25 percent of total server capacity.

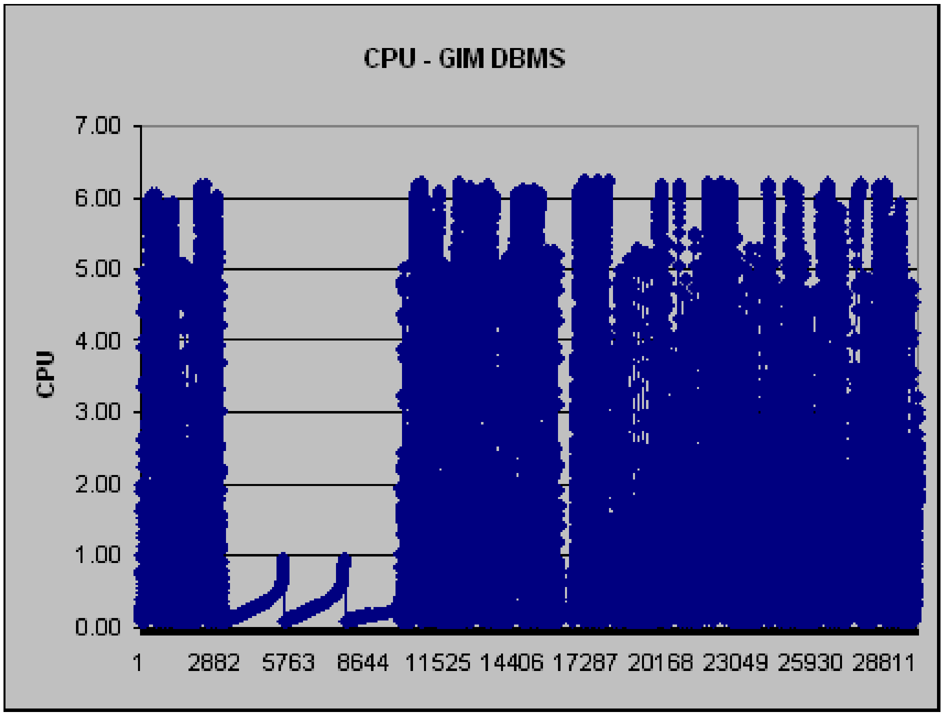


Figure 76: Oracle CPU Usage—Info Mart Database

IDB

CPU Utilization

Figure 77 depicts the CPU usage for one core on the IDB database engine during one day of the benchmark run at the maximum daily interaction volume. The IDB database engine used a maximum of 19 percent of total server capacity.

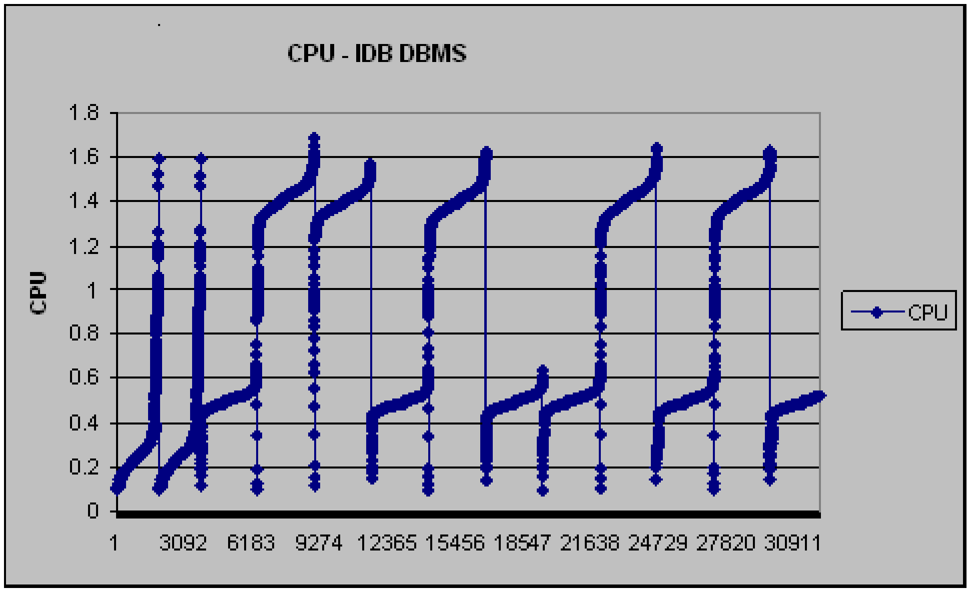


Figure 77: Oracle CPU Usage—IDB

Genesys Info Mart 8.x Performance Tuning

This section describes conclusions and recommendations that arise from the Genesys Info Mart 8.x performance testing.

Factors Affecting Release 8.x ETL Performance

The following factors affect ETL performance:

- The number of daily interactions in your contact center
- The number of agents in your contact center
- The complexity of your interaction flows
- The amount of business data attached to interactions
- The hardware on which the Genesys Info Mart Server is running, primarily the CPU speed and available memory
- The hardware on which the Info Mart RDBMS server is running, primarily the CPU speed, disk speed, and available memory
- The tuning of the Info Mart database
- The hardware on which the RDBMS server for IDB is running, primarily the CPU speed, disk speed, and available memory
- In very large-scale deployments using Oracle RAC, distribution of functions across nodes in the cluster
- The amount of historical data retained in IDB
- The speed of the network connections between components

- The degree of parallelism configured for application and database processes

Performance Tuning Guidelines

Based on the performance testing, Genesys has the following recommendations for tuning Genesys Info Mart 8.x performance, particularly in large-scale environments:

Genesys Info Mart Recommendations

- Partition the Info Mart database.

  Partitioning of the Info Mart database significantly streamlines database maintenance, because purging of potentially large amounts of GIDB and dimensional model fact data is accomplished by a simple DROP PARTITION statement. Because the maintenance job cannot run at the same time as the ETL jobs, streamlining maintenance increases the amount of time that is available for ETL processing.

  Genesys recommends using the default partition sizes of 86400 seconds (1 day) for GIDB and dimensional model fact tables. (Until release 8.1.1, the default partition size for the dimensional model was 7 days.)
- Utilize RDBMS capabilities and standard database strategies to optimize database I/O—for example, in non-cluster deployments, use database links.
- Do not use database compression. The overall storage savings are not significant, while compression imposes a significant performance penalty, because it causes large amounts of contention among parallel threads.

  Do not remove any compression that has been defined by RAA.
- In Oracle deployments, pre-allocate the Info Mart tablespace, instead of relying on AUTOEXTEND during job execution.
- Observe the following recommendations for important Oracle initialization parameters:
  - Dynamic sampling enabled
  - cursor-sharing=EXACT
  - Increased block size (db_block_size=16384 instead of 8192)
  - PARALLEL_FORCE_LOCAL=TRUE
  - (Genesys Info Mart startup parameter) -DqueryParallelism=8

  For additional Oracle settings, see Table 46 on page 120 and Table 55 on page 163.
- If you use archive logging for the Info Mart database, do not put the IDBs in the same RDBMS instance.
- Increase the values of the Genesys Info Mart configuration options that control the parallelism of ETL processing—extract-data-max-conn, irf-io-parallelism, and ud-io-parallelism. For recommended settings, see Table 48 on page 125.

Interaction Concentrator Recommendations

- To minimize the load on each IDB, distribute monitoring and storage among multiple Interaction Concentrator instances.
- To improve both Interaction Concentrator and Genesys Info Mart performance in deployments that include large quantities of user data, divide user-data storage between the G_USERDATA_HISTORY and G_SECURE_USERDATA_HISTORY tables, and use multiple custom user-data tables, as required.

  Similarly, in deployments that include Outbound Contact, divide storage of field data between the GO_FIELD_HIST and GO_SEC_FIELD_HIST tables.
- Increase the cache sizes for the most frequently used sequences in IDB. As described on page 121, Genesys used a cache size of 1000.
- Purge IDB regularly. In new database deployments with Interaction Concentrator 8.1.1 or later, partition the IDB schemas, so that you can streamline IDB purge.

  The number of interaction records in IDB has a known impact on the performance of the extraction job. To maintain ETL job performance over time, Genesys recommends that you regularly purge IDB.

  For the Interaction Concentrator release that was used during 8.0 testing, which used nonpartitioned IDB schemas, the performance of the stored procedures to purge IDB was not sufficient, and executing the purge procedures interfered with ICON's ability to write calls. As an interim solution, truncate commands were used to purge data from the G_* tables.

  For the 8.1.2 testing, use of IDB partitioning and the partition-aware purge procedures was found to be essential to meet performance requirements.
- Use high availability (HA) architecture and functionality for disaster recovery, instead of archive logs and database backups of individual IDBs.

Genesys Info Mart 8.x Database Size Estimation

Genesys Info Mart 8.x reads data from IDB(s) and writes data to various tables in the Info Mart database. The Genesys Info Mart database schema includes the following categories of tables, which the ETL jobs use for data processing and storage:

- Control tables
- Merge tables
- GIDB tables
- Staging tables
- Temporary tables
- Fact and dimension tables (collectively referred to as dimensional model)

In deployments that use GI2 or RAA, the Info Mart database also includes aggregate tables and views.

For database sizing purposes, the GIDB tables and the dimensional model are the most significant. The variables that have the greatest effect on the size of the GIDB and dimensional model tables are:

- The number of daily interactions in your contact center
- The number of agents in your contact center
- The complexity of your interaction flows
- The amount of business data attached to interactions
- In deployments that include eServices/Multimedia, the average length of time that multimedia interactions remain active
- The length of time that you want to retain data in the dimensional model and GIDB

Genesys Info Mart provides the *Genesys Info Mart 8.x Database Size Estimator,* a spreadsheet in Microsoft Excel 2007, to help you estimate the data size of the Info Mart database.

The spreadsheet requires you to select the RDBMS that you use and to fill in the basic information about the resources and daily interaction volumes for your contact center, including campaign information for deployments that include Outbound Contact. You also specify the number of days that you intend to store fact data in the GIDB and dimensional model tables, respectively. The spreadsheet then calculates the estimated average size of the Info Mart database. For deployments that include aggregation, a separate tab in the spreadsheet enables you to estimate the additional size of the aggregate tables.

The *Genesys Info Mart Database Size Estimator* spreadsheet for your release is available from the Genesys Customer Care website.

Notes:

- The estimates include only raw data; they do not include database overhead.
- The data sizes are only estimates. Be sure to factor in extra space to accommodate variations in average data lengths.
- By default, the estimate allows for storage of up to 14 days of extracted data, which is the default value of the applicable purge-related configuration option (days-to-keep-gidb-facts). If you decide to retain GIDB data for a shorter period, ensure that you provide sufficient database storage, and configure Genesys Info Mart, to retain the data long enough to enable Genesys Info Mart to recover from possible network outage, database outage, or other errors that might temporarily prevent Genesys Info Mart from transforming or loading data.

- The Info Mart size estimate includes default indexes. Ensure that you provide additional storage for any indexes that you plan to add to the Info Mart database schema to enhance query performance.
- The *Genesys Info Mart Database Size Estimator* estimates the size of the Info Mart database only and does not include Interaction Concentrator requirements. For information about estimating the size of IDB, see the Interaction Concentrator chapter in this guide.

Genesys

Chapter

8 Reporting and Analytics Aggregates

Reporting and Analytics Aggregates (RAA) is an optional aggregation process that users can add to a Genesys Info Mart environment to create and populate predefined aggregation tables and views in the Info Mart database.

To assist you in estimating how RAA might perform in your environment, this chapter provides sample measurements of RAA performance on reference platforms in large-scale, multi-site Genesys Info Mart deployments that included aggregation.

This chapter contains the following sections:

- RAA Release 8.1 Performance, page 189
- Release 8.1.1 Endurance Test, page 192
- Release 8.1.1 Interaction Lifespan Limit, page 199

RAA Release 8.1 Performance

RAA release 8.1 performance testing was conducted as part of performance testing for Genesys Info Mart release 8.1.2. The primary purpose of the release 8.1 testing was to validate the ability of Genesys Info Mart to support very high volumes of data in a deployment that included aggregation.

The primary purpose of the RAA testing was to validate that aggregation was able to keep pace with the volume of data. A secondary purpose was to determine the maximum interaction lifespan that RAA could support in the test environment.

The Genesys Info Mart 8.1.2 and RAA 8.1 performance testing consisted of:

- An endurance test in which Genesys Info Mart ran for 14 consecutive days at a peak rate of 19 million interactions a day, or 220 interactions per second.

For detailed Genesys Info Mart results, including aggregation-related database performance, see "Endurance Test Results" on page 128.

For detailed RAA results, see "RAA Results for the Endurance Test" on page 193.

- A recovery test in which extract, transform, and load (ETL) processing was stopped for ten hours while call generation continued at peak rates. When the ETL resumed, Genesys Info Mart was able to recover from the outage within approximately five hours. RAA was able to keep pace with Genesys Info Mart during recovery.

  For detailed Genesys Info Mart results, see "Ten-Hour Recovery Test Results" on page 148.

- A test to identify the longest lifespan for a multimedia interaction that RAA could support in the test environment.

  For more information, see "Release 8.1.1 Interaction Lifespan Limit" on page 199.

This chapter provides detailed results about the aspects of the Genesys Info Mart solution release 8.1 performance testing that relate only to RAA. For more information about the aspects of release 8.1 performance testing that relate to Genesys Info Mart, see "Genesys Info Mart 8.1.2 Performance" on page 109.

Release 8.1.1 Test Environment

Genesys Info Mart and RAA were deployed on a Linux platform. The Info Mart database and IDBs were deployed on a four-node Oracle 11g Release 2 (R2) Real Application Clusters (RAC) database platform.

For details about the Genesys Info Mart hardware and software environment, including the RAA node in the Oracle cluster, see "Release 8.1.2 Performance Test Setup" on page 111.

The ETL cycle and aggregation ran throughout the day to process actively populated source data, which was continuously generated for a high-complexity call flow. The ETL jobs were scheduled to run effectively continuously from 00:10 AM to midnight every day. The aggregation job was scheduled to run continuously from 00:10 AM to 23:40 every day. In other words, the ETL cycle and the aggregation job ran concurrently throughout most of the day. The maintenance job ran between midnight and 00:10 AM every day.

RAA Application Version

Preliminary testing with Genesys Info Mart 8.1.0 and RAA 8.1.0 identified an issue with the aggregate dispatcher that prevented Genesys Info Mart from meeting performance test requirements.

In release 8.1.0, the aggregate dispatcher worked from the oldest notification to the newest. With e-mail interactions lasting multiple days, the ETL produced notifications spanning multiple days nearly every ETL cycle. As a result, aggregates fell days behind the facts and were unable to catch up.

RAA release 8.1.101 introduced a new dispatcher implementation that provides more flexibility in associating aggregation resources with different windows of time. When the test environment was upgraded to use RAA 8.1.101.01, the changes in aggregate dispatcher functioning enabled aggregates to remain current with the fact data.

The test results that are reported in this document relate only to the upgraded test environment. For more information about the applications that were used for the 8.1 performance testing, see “Test Applications” on page 110 and “Product Versions” on page 118.

Aggregation–Related Configuration

With Genesys Info Mart hosting the aggregation process, options that control functioning of the aggregation engine are configured in the Genesys Info Mart Application object. In addition, certain runtime parameters affect aggregation engine operations.

Table 61 lists those aggregation-related options that were set to non-default values for the testing. The table includes only settings that affected test results. Table 61 includes both configuration option settings and runtime parameters.

Note: By default, population of all RAA aggregate hierarchies is enabled. This default setting was not changed.

Table 61: Non-Default Aggregation-Related Settings

Option or Parameter	Value Used	Default Value
Configuration Options		
agg Section		
level-of-log	.:FINEST	.:INFO
gim-etl Section		
aggregation-engine-class-name	GIMAgg.GimInterfaceImpl.AggregationImpl	

Table 61: Non-Default Aggregation-Related Settings (Continued)

Option or Parameter	Value Used	Default Value
schedule Section		
aggregate-duration aggregate-schedule **Note:** These option settings mean that the aggregation job ran continuously from 00:10 until 23:40 every day.	23:30 10 0	05:00 0 1
Runtime Parameters		
-writerSchedule -zoneOffset -realtimeOffset	default=flex(4:5) 36000 (10 hours) 3600 (1 hour)	default=flex(3:1) 115200 (32 hours) 900 (15 minutes)
Note: These parameter settings mean that four aggregate writer processes were allocated to the zone for recent notifications (Zone1), which spanned 10 hours, and five aggregate writer processes handled the backlog of re-aggregation requests in the zone for older notifications (Zone2). The leading edge of Zone1 was 1 hour behind the time the notifications were received, to avoid conflicts with active ETL processing.		

Other RAA customizations included use of:

- A suitable partition-kit.ss file to alter the RAA hierarchies to function with the partitioned Info Mart database
- A patch-agg.ss file to disable month-level aggregates for the AGENT and AGENT_QUEUE hierarchies
- A user-data-map.ss file to map a total of four custom dimensions to various hierarchies

Release 8.1.1 Endurance Test

The RAA aspect of the Genesys Info Mart 8.1.2 endurance testing measured the ability of RAA 8.1.1 to keep pace with Genesys Info Mart processing of 19 million interactions per day for 14 consecutive days, with new aggregates being produced one hour behind new facts.

For information about the hardware, software, and call flows that were used for the endurance test, see “Release 8.1.2 Performance Test Setup” on page 111.

In the test environment, the data from the respective data domains had the following distinct characteristics:

- The majority of inbound and Outbound Contact voice data notifications fell into Zone1 (the zone for recent new activity). The facts were not aggregated at all until the data was one hour old, and the majority of the voice data was effectively aggregated only once.
- Multimedia data had facts that lasted between four and five days. Notifications for multimedia facts fell across both notification zones (recent new activity and updates about older facts), and all the multimedia data was aggregated more than once, in some cases almost 40 times (see Figure 78 on page 198).

RAA Results for the Endurance Test

The improved aggregate-dispatcher functionality enabled RAA to keep producing new aggregates one hour behind the new facts. Updates about activity relating to long-lived multimedia interactions did cause a constant backlog of re-aggregation requests. However, with the aggregation resources divided between Zone1 and Zone2, RAA kept the Zone1 aggregates current with the facts.

This subsection provides results for the processing side of RAA performance during the Genesys Info Mart 8.1.2 endurance test. For information about database performance of the RAA node in the Oracle cluster, see "Hardware Resource Usage—Oracle RAC" on page 136.

Query Performance

One measure of RAA performance is the time required for aggregation queries to complete. Test measurements indicate that RAA was comfortably able to process new fact data.

Measurement Method—Query Performance

Genesys used the checkpoint messages that the Agg.Writer processes produce in the log after a set of aggregate result rows are inserted. For example:

Agg.Writer.8   Checkpointed: 37 H_CAMPAIGN-HOUR, zone1, 4 key(s) in 64,852 ms, deleted 105, inserted 128

The messages indicate how many DATE_TIME keys were used (in the above example, 4), which correlates to the timespan of the query, the dispatcher zone in which the keys fall, and the runtime of the query; the messages also indicate the number of rows produced.

Genesys used these checkpoint messages for Zone1 over a day of operations, normalizing the runtimes for one hour of data, to calculate typical runtimes for the various aggregation queries.

Query Runtimes

Table 62 reports typical runtimes for the various queries to aggregate one hour of fact data.

Table 62: Average Runtimes to Aggregate 1 Hour of Fact Data

Aggregate Name	Average Time (seconds)
I_AGENT	192
I_SESS_STATE	300[a]
I_STATE_RSN	24
CAMPAIGN	16
AGENT_CAMPAIGN	17
ID	110
AGENT	53
AGENT_QUEUE	60
QUEUE	160
QUEUE_ABN	8
QUEUE_ACC_AGENT	30

a. Except for I_SESS_STATE, the runtimes for most aggregates were fairly consistent. For I_SESS_STATE, runtimes for a single key were either around 20 seconds or around 500 seconds.

Aggregates Latency

Another measure of RAA performance is the latency between new fact data and aggregates. Test measurements indicate that RAA consistently produced new aggregates one hour behind new fact data, which corresponds to the configured real-time offset for Zone1. In other words, for new facts and updates to facts that were up to ten hours old, new and updated aggregates were produced approximately one hour after the activity occurred.

However, there was a constant backlog of re-aggregation requests in Zone2, which affected interactions that were up to five days old. The latency for re-aggregation depended on the extent of other ETL activity. The extended outage test demonstrated that, when no new notifications were coming in, RAA processed the backlog of notifications in Zone2 in about three hours (see “Aggregation During Extended Outage” on page 198). In other words, for updates to facts that were between ten hours and five days old, updated aggregates were produced no later than three hours after the activity occurred.

Measurement Method—Aggregates Latency

Genesys queried the fact tables, aggregate tables, and pending aggregation requests at a point in time, then compared timestamps to determine the relationships between new fact data and completed and pending aggregation.

Latency Results

Tables 63 and 64 summarize the results of queries of the Info Mart *_FACT and AG2_* tables and the RAA PENDING_AGR internal queue at 15:40 on January 23, 2013.

Comparison of the timestamps of leading and trailing edges, respectively, indicate that:

- Aggregates are one hour behind the facts. Facts are being produced in the hour between 15:00 and 16:00, and aggregates are being produced for the hour between 14:00 and 15:00.
- Different aggregate tables have different backlogs of re-aggregation requests.
- Unprocessed notifications cover the range from 15:00 on January 23 back to 16:00 on January 18.

Table 63 shows the latest timestamps of the facts for the various types of interactions.

Table 63: Summary of Fact Data at 15:40

Number of Interactions	Leading Edge of Facts	Media	Interaction Type
2147	2013-01-23 15:30-15:45	Chat	Inbound
6505	2013-01-23 15:30-15:45	Email	Inbound
119707	2013-01-23 15:30-15:45	Voice	Inbound
27188	2013-01-23 15:30-15:45	Voice	Outbound
6784	2013-01-23 15:30-15:45	Voice	Unknown

Table 64 shows the latest timestamps (leading edge) of the aggregates in various aggregate tables, together with the oldest timestamps (trailing edge) of facts for which there were pending re-aggregation requests. In all cases, the leading edge of aggregates in Table 64 was approximately one hour behind the leading edge of the facts in Table 63. The trailing edge of pending notifications was up to five days behind the leading edge of the facts.

Table 64: Summary of Aggregate Data and Pending Notifications at 15:40

Table Name	Leading Edge of Aggregates	Trailing Edge of Pending Notifications
AG2_I_AGENT_SUBHR	2013-01-23 14:00-14:30	2013-01-21 15:00
AG2_I_SESS_STATE_SUBHR	2013-01-23 14:00-14:30	2013-01-22 13:00
AG2_I_STATE_RSN_SUBHR	2013-01-23 14:00-14:30	2013-01-22 12:30
AG2_AGENT_HOUR	2013-01-23 14:00-15:00	2013-01-21 15:00
AG2_AGENT_GRP_HOUR	2013-01-23 14:00-15:00	2013-01-21 15:00
AG2_AGENT_QUEUE_HOUR	2013-01-23 14:00-15:00	2013-01-18 16:00
AG2_CAMPAIGN_HOUR	2013-01-23 14:00-15:00	2013-01-21 15:00
AG2_QUEUE_HOUR	2013-01-23 14:00-15:00	2013-01-18 16:00
AG2_QUEUE_GRP_HOUR	2013-01-23 14:00-15:00	2013-01-21 13:00
AG2_QUEUE_ACC_AGENT_HOUR	2013-01-23 14:00-15:00	2013-01-18 16:00
AG2_QUEUE_ABN_HOUR	2013-01-23 14:00-15:00	2013-01-18 16:00
AG2_AGENT_CAMPAIGN_HOUR	2013-01-23 14:00-15:00	2013-01-21 21:00
AG2_ID_HOUR	2013-01-23 14:00-15:00	2013-01-18 16:00

Re-Aggregation Requests

To obtain a more detailed picture of the degree of re-aggregation required and the time to process backlogged requests, Genesys performed additional measurements to ascertain the frequency and distribution of re-aggregation requests.

For Voice and Outbound Contact details, notifications about available fact data typically contained time ranges of 15 minutes or less. However, for Multimedia details, notifications about available fact data usually contained time ranges of multiple days. For example, Table 65 shows the ranges for a typical set of notifications for an ETL cycle during the test run:

Table 65: Typical Notification Ranges for Facts

Fact Table	Notification Range (seconds)	
	Inserted Facts	Updated Facts
SM_RES_STATE_FACT	632	86460
IXN_RESOURCE_STATE_FACT	60	
INTERACTION_FACT	1533	318333
MEDIATION_SEGMENT_FACT	716	318049
INTERACTION_RESOURCE_FACT	713	86071

The facts in Table 65 cover agent states that last up to one day, queue mediations that last for up to four days, and e-mail replies that span up to five days. Changed facts over these timespans lead to backlogs of pending notifications, as shown in “Aggregates Latency” on page 194.

As Table 65 shows, aggregates that are based on INTERACTION_FACT and MEDIATION_SEGMENT_FACT data have the longest range of pending re-aggregation requests, while aggregates that are based on agent states have the shortest range of pending re-aggregation requests.

The number of times that a fact is re-aggregated depends on the ranges of the notifications. For an aggregate such as AG2_ID_HOUR, which is triggered from INTERACTION_FACT, a given hour of fact data is aggregated about 40 times before the trailing edge of the notifications passes that hour in time.

Measurement Method—Re-Aggregation Frequency

Re-aggregating a time range (in other words, recalculating the aggregates for a timespan) involves deleting all rows in the aggregate tables for that timespan, then inserting the updated aggregate rows. Genesys added a trigger to the AG2_ID_HOUR aggregate table, to insert a row in an audit table every time the aggregate was deleted. Each unique delete time in the audit table corresponds to a re-aggregation request as a result of an INTERACTION_FACT update. Graphing the number of times a row was deleted over 5–6 days (the number of aggregation requests) against the age of the row in the audit table (a proxy for the age of the underlying fact) yields the re-aggregation frequency.

Re-Aggregation Frequency

Figure 78 shows the frequency of re-aggregation requests generated by fact updates during a 5-day period of steady-state operations. The number of re-aggregation requests is related to the age, in hours, of the underlying INTERACTION_FACT data. Not surprisingly, the re-aggregation frequency increases with the lifetime of the interactions.

Figure 78 illustrates the following results in the test environment:

- Facts that were 24 hours old were aggregated 10 times.
- Facts that were 48 hours old were aggregated 16 times.
- Facts that were 72 hours old were aggregated 20 times.
- For the longest-living e-mail interactions (5 days), facts were aggregated 38–40 times before the underlying fact data stopped changing.

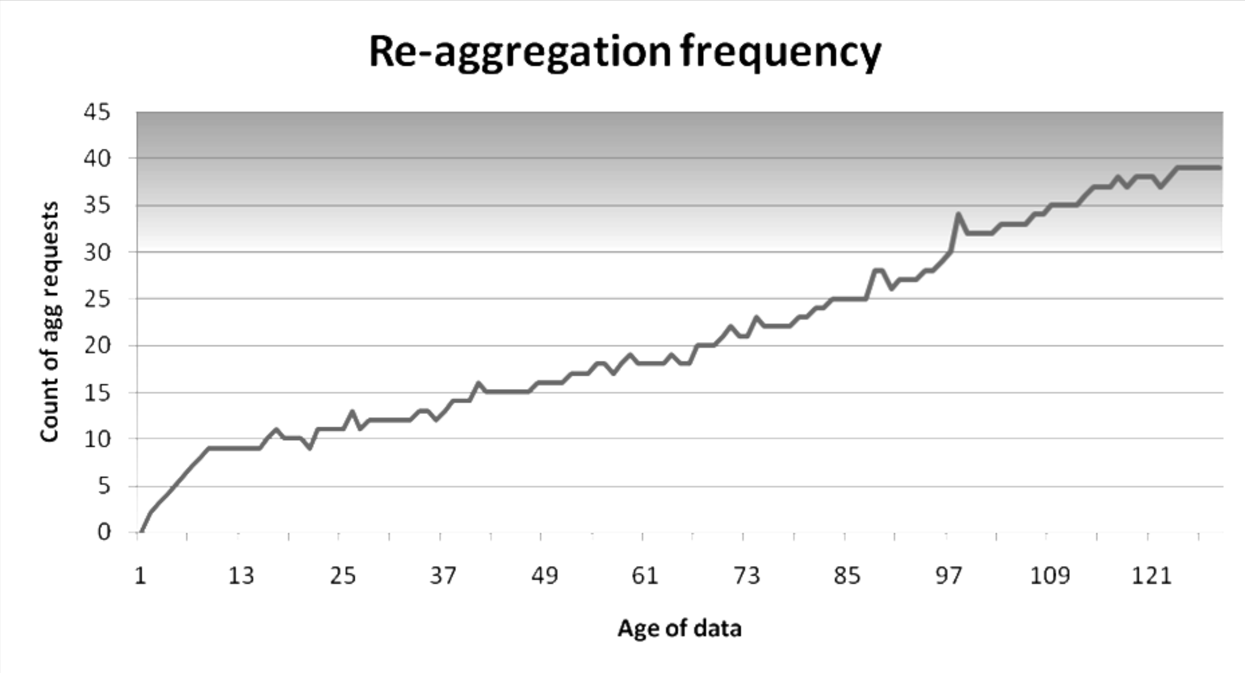


Figure 78: Re-Aggregation Frequency

Aggregation During Extended Outage

Another way to interpret the results in Figure 78 is that, on average, the AG2_ID_HOUR aggregate was calculated about eight times a day. Since the ETL constantly produced large notifications, eight times a day is the fastest that RAA can produce re-aggregates.

This is consistent with aggregation results observed during the extended outage test (see "Release 8.1.2 Recovery Test—Ten-Hour Outage" on page 148): When the ETL was stopped for an extended period, RAA took about three hours to fully process the backlog of notifications that were in Zone2 when the ETL was stopped. Within three hours, the PENDING_AGR queue was empty.

Release 8.1.1 Interaction Lifespan Limit

As a supplementary test, Genesys attempted to identify the longest-living multimedia interaction that RAA could support in the test environment, before RAA started falling behind.

With the RAA 8.1.1 aggregate dispatcher implementation, the notifications are split between a set that keeps the leading edge of the aggregates up to date with the leading edge of the facts and a set for the backlog of re-aggregation requests for older time periods. In the test environment, RAA was configured so that, regardless of how large the backlog became, RAA was always able to keep the aggregates up to date with the facts (see page 192). For the purposes of identifying the limit for the lifetime of an interaction, falling behind refers to the situation when the backlog is no longer bounded, and the trailing edge of the backlog advances slower than the leading edge of the facts.

As other RAA testing showed (see page 197), it took approximately 3 hours to process notifications for a 5-day timespan in the test environment. Extrapolating from this result, it would take 24 hours to process a 40-day timespan. For a timespan longer than 40 days, the backlog would grow faster than it can be processed.

A further consideration is how often business users require the aggregates to be updated. For example, do they require yesterday's aggregates to be updated once a day, twice a day, or more?

Formula

Genesys postulates the following formula for estimating the maximum lifespan of interactions in a deployment, based on actual time to process a particular timespan of notifications:

> max interaction span = [(*x*-day timespan of notification)/(time for RAA to process *x*-day timespan) * (max available backlog processing time)] / (minimum daily update count)

For example, in the Genesys test environment:

- For one update a day for previous days' aggregates:

  max interaction span = [(5*24 hrs)/3 hrs) * 24 hrs] / 1 = 960 hrs or 40 days

- For four updates a day for previous days' aggregates:

  max interaction span = [(5*24 hrs)/3 hrs) * 24 hrs] / 4 = 240 hrs or 10 days

Maximum Interaction Lifespan Test

To validate the calculations for estimating the maximum supported lifespan for interactions, Genesys injected an artificial notification for both the INTERACTION_FACT (IF) table and the MEDIATION_SEGMENT_FACT (MSF) table, covering 40 days of full-volume data. Genesys measured the time for RAA to process this 40-day notification while the ETL continued to process interactions at the full call rate, with e-mail interactions up to two days long.

Maximum Interaction Lifespan Test Results

Throughout the test, new aggregates stayed current with new facts, offset by the configured one-hour delay. Furthermore, the ETL job times were consistent with previous steady-state tests; in other words, Genesys Info Mart ETL processing was not affected by the additional RAA activity.

After 8 hours, the backlog had been reduced by 13 days. After 23 hours and 20 minutes, the backlog of ID and QUEUE aggregates had been processed and the remaining backlog aggregates were being processed.

The test confirms that, under the Genesys Info Mart 8.1.2 test conditions, a 40-day notification can be processed in about 24 hours.

Note: Genesys tested a single 40-day notification sent during a 24-hour processing period. Repeated notifications of this size—for example, every hour—would likely exceed the processing capability of the configured number of backlog Aggregate Writer threads. For frequent notifications, the sustainable maximum lifespan for interactions in the test environment is likely 30 days.

Chapter

9 Genesys Interactive Insights

Genesys Interactive Insights (GI2) is the presentation layer that Genesys has designed for the business-like interpretation of source data that is collected by Genesys Info Mart and stored in the Info Mart database. GI2 is powered by Business Objects Enterprise (BO) XI 3.1 software in release 8.1.1 and earlier, and by SAP BusinessObjects Business Intelligence Platform (BI) 4.1 in releases 8.1.3 and later.

To assist you in estimating how GI2 might perform in your environment, this chapter provides the results of GI2 performance tests that Genesys executed on sample reference platforms.

This chapter contains the following sections:

- GI2 Release 8.0 Performance, page 201
- Release 8.0 Report Performance—Baseline, page 203
- Release 8.0 Report Performance—Benchmark, page 211
- Release 8.0 Conclusions, page 215
- GI2 Release 7.6 Performance, page 215
- GI2 7.6 Report Performance for a Microsoft SQL Server Info Mart, page 219
- GI2 7.6 Report Performance for an Oracle Info Mart, page 223
- Release 7.6 Conclusions, page 229

GI2 Release 8.0 Performance

GI2 release 8.0 performance testing was conducted as part of performance testing for Genesys Info Mart release 8.0. The primary purpose of the release 8.0 testing was to validate the ability of the Genesys Info Mart and GI2 solution to sustain intraday reporting with high rates of traffic. The particular purpose of the GI2 testing was to evaluate GI2 report performance under load.

Genesys conducted two kinds of tests for GI2 release 8.0 performance:

- Baseline—An endurance test to measure the ability of GI2 to sustain intraday reporting when Genesys Info Mart ran for more than 7 days with 2 millions calls per day in a voice-only environment.
- Benchmark—A capacity test to validate the ability of GI2 to sustain intraday reporting in a blended environment (voice, e-mail, and chat) under the maximum load of 8.6 million interactions a day (14 multimedia interactions per second and 85.5 voice calls per second [cps]).

During the Genesys Info Mart testing and after completion, selected GI2 reports were run in various combinations and frequencies, both on dynamic data (while calls were being generated and Genesys Info Mart was actively processing them) and on static data (while Genesys Info Mart was not actively processing new interactions).

To represent the range of reports that are available with GI2, the selected reports accessed different sets of aggregate tables—agent, queue, and business attribute/tenant.

This chapter provides detailed results about the aspects of the Genesys Info Mart solution release 8.0 performance testing that relate only to GI2. For more information about the aspects of release 8.0 performance testing that relate to Genesys Info Mart, see Chapter 7 on page 107.

Environment

Genesys Info Mart was configured to run the extract, transform, and load (ETL) jobs effectively continuously from 01:00 a.m. to midnight. Similarly, the aggregation job was scheduled to run continuously from 01:00 a.m. to midnight. In other words, the ETL cycle and the aggregation job ran concurrently throughout the day, except for a one-hour maintenance interval from midnight to 01:00 a.m.

In other respects, the testing environments for the baseline and benchmark tests were different.

Baseline Test Environment

All tests were run with the following software:

- Business Objects Enterprise (BOE) XI 3.1 SP3 Fix Pack 1.8
- GI2 8.0

BOE was deployed on a dual dual-core Intel Xeon 5160 Woodcrest box with the following specifications:

- 3.0 GHz operating speed
- 4 GB RAM

The operating system was Microsoft Windows Server 2003 Enterprise Edition R2 SP2.

For details about the Genesys Info Mart hardware and software environment, including the Reporting and Analytics Aggregates (RAA) software and Genesys Info Mart application configuration, see "Baseline Performance Test Setup" on page 157.

Benchmark Test Environment

All tests were run with BO XI 3.1 and GI2 release 8.0 software.

The BOE-related software was deployed on a dual quad-core Intel Xeon E5410 box with the following specifications:

- 2.33 GHz operating speed
- 4 GB RAM

The operating system was Microsoft Windows Server 2003.

For details about the Genesys Info Mart hardware and software environment, including the RAA software and Genesys Info Mart application configuration, see "Benchmark Performance Test Setup" on page 172.

Release 8.0 Report Performance—Baseline

The GI2 aspect of the Genesys Info Mart 8.0 baseline performance testing measured GI2 performance in terms of the time required to generate the selected report (*duration* of the report).

General Results

As an indication of overall Genesys Info Mart and GI2 release 8.0 performance during the baseline tests, Table 66 on page 204 provides results for a number of GI2 reports.

The Parameters column in Table 66 specifies User Prompt values that were selected for the applicable report.

Table 66: Durations of Various Reports—Baseline Testing

Report Name	Frequency	Duration (secs)	Number of Pages	Parameters
Agent Login-Logout Details Report	Daily	20	6324	Pre-set Day Filter = None Report Date = <date> From Hour = 0 To Hour = 24 All other prompts = all (default)
Agent Not Ready Report	Daily	38	38	Pre-set Day Filter = None Report Date = <date> From Hour = 0 To Hour = 24 All other prompts = all (default)
Agent Not Ready Report	Daily	11	6	Pre-set Day Filter = None Report Date = <date> From Hour = 0 To Hour = 24 Agent Group = <group> All other prompts = all (default)
Agent State Details Report	Hourly	117	2	Pre-set Day Filter = None Report Date = <date> From Hour = <hour> To Hour = <hour + 1> Agent = <agent> Reason Code Type = SOFTWARE All other prompts = all (default)
Interaction Traffic Report	One-time	2	2	Pre-set Date Filter = None Start Date = <date> End Date = <date + 1 hour> All other prompts = all (default)
Interaction Volume Summary Report	One-time	21	2	Start Date = <date> End Date = <date + 1 hour> All other prompts = all (default)
Interaction Volume Customer Segment Report	One-time	21	2	Pre-set Date Filter = None Start Date = <date> End Date = <date + 5 minutes> Customer Segment = <customer segment> All other prompts = all (default)

Table 66: Durations of Various Reports—Baseline Testing (Continued)

Report Name	Frequency	Duration (secs)	Number of Pages	Parameters
Interaction Volume Business Result Report	One-time	20	2	Pre-set Date Filter = None Start Date = <date> End Date = <date + 1 hour> Business Result = <business result> All other prompts = all (default)
Interaction Handling Attempt Report	One-time	17	51	Pre-set Day Filter = None Start Time = <date1> End Time = <date2> Last Queue = 8001 All other prompts = all (default)
Queue Summary Report	One-time	1	2	Pre-set Date Filter = None Start Date = <date> End Date = <date + 1 day> All other prompts = all (default)
Agent Summary Activity Report	Weekly	50	42	Pre-set Date Filter = None Start Date = <date1> End Date = <date2> All other prompts = all (default)
	Weekly	16	6	Pre-set Date Filter = None Start Date = <date1> End Date = <date2> Agent Group = <group> All other prompts = all (default)
	Monthly	9	1	Pre-set Date Filter = None Start Date = <date1> End Date = <date2> Agent = <agent> All other prompts = all (default)
	Monthly	40	6	Pre-set Date Filter = None Start Date = <date1> End Date = <date2> Agent Group = <group> All other prompts = all (default)
Abandon Delay Report	Monthly	3	2	Pre-set Date Filter = None Start Date = <date1> End Date = <date2> All other prompts = all (default)

Factors Affecting Report Performance

The baseline performance testing examined the impact of the following variables on GI2 report performance:

- Concurrency of reports—Two representative reports were run on their own and in combination with other reports, to identify the impact of concurrency on report performance. To isolate the effect of concurrency, the reports were run in a static environment—in other words, when no calls were being generated.

  For the test results, see “Effect of Concurrency” on page 206.
- Amount of data to be processed—A detailed agent activity report was run at daily intervals for a week to identify if an increasing amount of data in the Info Mart database had an impact on report performance.

  For the test results, see “Effect of Data Growth” on page 207.
- Concurrency with other processing—Three reports in the Agents category were run during call generation and then in a static environment, to identify if ETL and aggregation processing of new call data had an impact on report performance.

  For the test results, see “Effect of Concurrent Processing” on page 208.
- Call rate—A detailed agent activity report was run at various call rates over the course of a day, to identify if the call rate affected report performance.

  For the test results, see “Effect of Traffic Rate” on page 210.

Effect of Concurrency

To identify if the number of concurrent report queries affects the amount of time that is required to generate a report, two reports were run in different combinations of concurrency.

Table 67 summarizes the report and test parameters for the concurrent report test.

Table 67: Concurrent Report Test Conditions

Report Name	Frequency	Parameters	Test Scenario
Interaction Volume Business Result Report	Daily	Pre-set Date Filter = None Start Date = <date 09:00:00 AM> End Date = <date 12:00:00 PM> Media Type = Voice All other prompts = all (default)	The report was executed three times with the same parameters in a static environment: • Exclusively (no other reports at the same time) • Concurrently with two other reports (three reports in total) • Concurrently with five other reports (six reports in total)
Agent Conduct Report	Daily	Pre-set Day Filter = None Report Date = <date> From Hour = 8 To Hour = 9 Agent Group = <group> All other prompts = all (default)	The report was executed twice with the same parameters in a static environment: • Exclusively (no other reports at the same time) • Concurrently with two other reports (three reports in total)

Table 68 summarizes the report performance results for the concurrent report test.

Table 68: Concurrent Report Test Results

Report Name	Concurrency	Duration (secs)	Conclusions
Interaction Volume Business Result Report	1	3	Running the report concurrently with: • Two others did not significantly impact the duration. • Five others had a significant impact on the duration.
	3	4	
	6	7	
Agent Conduct Report	1	10	Running the report concurrently with two others did not significantly impact the duration.
	3	13	

Effect of Data Growth

To identify if the growth of data in the Info Mart database affects the amount of time that is required to generate a report, the Agent Interval Based Report was run daily for eight days, at a time when calls were not being generated. The Genesys Info Mart application was configured to purge Info Mart fact data after seven days.

The same parameters were used for each run of the report: Pre-set Day Filter = Today, From Hour = 9, To Hour = 24, Media Type = Voice, all other prompts = all (default).

Table 69 summarizes the report performance results for the data size test.

Table 69: A Week's View of the Agent Interval Based Report

Day	Duration (secs)	Number of Pages	Conclusions
Day 1	9	3388	There was no significant deviation in the time it took to generate the report (duration) from the start of the week to the end of the week.
Day 2	9	3360	
Day 3	8	2892	
Day 4	9	2477	
Day 5	8	3354	
Day 6	10	3362	
Day 7	11	3501	
Day 8	10	3338	

Effect of Concurrent Processing

To identify the impact of concurrent ETL and aggregation processing on report generation time, three daily reports on agent activity were each run on static and dynamic data.

Table 70 summarizes the report and test parameters for the concurrent processing test.

Table 70: Concurrent Processing Test Conditions—Agents Category Reports

Report Name	Frequency	Parameters	Test Scenarios
Agent Wrap Report	Daily	Pre-set Day Filter = Today From Hour = 0 To Hour = 24 All other prompts = all (default)	For three days, the report was executed twice with the same parameters: • Once while no calls were being generated • Once while calls were being generated at 28 cps

Table 70: Concurrent Processing Test Conditions—Agents Category Reports (Continued)

Report Name	Frequency	Parameters	Test Scenarios
Agent Group Business Result Report	Daily	For Day 1: Pre-set Date Filter = Today Agent Group = <group1> Media Type = Voice All other prompts = all (default) For Day 2: Pre-set Date Filter = Today Agent Group = <group2> Media Type = Voice All other prompts = all (default)	For two days, the report was executed twice: • Once while no calls were being generated • Once while calls were being generated at 28 cps Each day's reports used the same parameters
Agent Interval Based Report	Daily	Pre-set Day Filter = today From Hour = 0 To Hour = 24 Media Type = Voice All other prompts = all (default)	The report was executed twice on one day with the same parameters: • Once while no calls were being generated • Once while calls were being generated at 40 cps

Table 71 summarizes the report performance results for the concurrent processing test. The table compares durations for three agent reports that were generated on dynamic data (while calls were being generated and Genesys Info Mart ETL processing and aggregation were continuing) and on static data (while calls were not being generated).

Table 71: Comparison of Report Durations for Dynamic and Static Data

Report Name	Day	Durations (secs)		Number of Pages	Conclusions
		Dynamic	Static		
Daily Agent Group Business Result Report	Day 1	6[a]	2	1	• Report duration increased when new call data was being generated. • Increasing the call rate increases the report duration further. • The number of pages in the report did not affect the duration of the report.
	Day 2	4[a]	2	1	
Agent Wrap Report	Day 2	44[a]	27	50	
	Day 3	32[a]	30	50	
	Day 4	43[a]	25	113[b]	
Daily Agent Interval Based Report	Day 1	29[c]	7	414	

a. Calls were generated at 28 cps.

b. The call generation environment had to be restarted, and this increased the length of the report.

c. Calls were generated at 40 cps.

Effect of Traffic Rate

To identify if increased call rates affect the amount of time that is required to generate a report, the Agent Conduct Report was run hourly over the course of a day, with the call rate periodically varied.

The same parameters were used for each run of the report: Pre-set Day Filter=today, From Hour=0, To Hour=24, Agent Group=<group>, all other parameters = all (default).

Table 72 summarizes the report performance results for the call rate test.

Table 72: Report Duration for Different Call Rates

Time of Day	Call Rate (cps)	Duration (secs)
1:14 a.m.	7	7
2:14 a.m.	7	9
3:14 a.m.	7	7
4:14 a.m.	7	8
5:14 a.m.	7	7
6:14 a.m.	7	8

Table 72: Report Duration for Different Call Rates (Continued)

Time of Day	Call Rate (cps)	Duration (secs)
7:14 a.m.	14	7
8:14 a.m.	14	8
9:14 a.m.	28	11
10:14 a.m.	28	9
11:14 a.m.	28	11
12:14 a.m.	28	17
1:14 p.m.	28	11
2:14 p.m.	40	13
3:14 p.m.	40	105
4:14 p.m.	40	15
5:14 p.m.	40	33
6:14 p.m.	28	63
7:14 p.m.	28	14
8:14 p.m.	28	78
9:14 p.m.	28	17
10:14 p.m.	28	35
11:14 p.m.	14	66
12:14 p.m.	14	7

Release 8.0 Report Performance—Benchmark

As part of the Genesys Info Mart benchmark testing, selected GI2 reports were run in various combinations and frequencies. The GI2 aspect of the Genesys Info Mart 8.0 benchmark performance testing measured GI2 performance in terms of the time required to generate the selected report (*duration* of the report).

General Results

As an indication of overall Genesys Info Mart and GI2 release 8.0 performance during the benchmark tests, Table 73 provides results for a number of GI2 reports.

The Parameters column in Table 73 specifies User Prompt values that were selected for the applicable report.

Table 73: Durations of Various Reports—Benchmark Testing

Report Name	Frequency	Duration (secs)	Number of Pages	Parameters
Agent Not Ready Report	Daily	55	1	Pre-set Day Filter = None Report Date = <date> From Hour = 0 To Hour = 24 All other prompts = all (default)
Queue Summary Report • Email • Chat • Voice	One-time	 83 85 116	 21 21 181	Pre-set Date Filter = None Start Date = <date> End Date = <date + 1 day> Media Type = Email/Chat/Voice All other prompts = all (default)
Agent Summary Activity Report	Daily	56	183	Pre-set Date Filter = None Start Date = <date1> End Date = <date2> All other prompts = all (default)
	Weekly	68	185	Pre-set Date Filter = None Start Date = <date1> End Date = <date2> Agent Group = <group> All other prompts = all (default)

Factors Affecting Report Performance

Similarly to the baseline testing, the benchmark performance testing examined the impact of the following variables on GI2 report performance under load:

- Concurrency of reports—Two representative reports were run on their own and in combination with other reports, to identify the impact of concurrency on report performance. To isolate the effect of concurrency, the reports were run in a static environment—in other words, when no calls were being generated.

  For the test results, see “Effect of Concurrency” on page 213.

- Concurrency with other processing—Two reports in the Agents category were run during generation of new interactions and then in a static environment, to identify if ETL and aggregation processing of new interaction data had an impact on report performance.

  For the test results, see “Effect of Concurrent Processing” on page 214.

Effect of Concurrency

To identify if the number of concurrent report queries affects the amount of time that is required to generate a report, two reports were run in different combinations of concurrency.

Table 74 summarizes the report and test parameters for the concurrent report test.

Table 74: Concurrent Report Test Conditions

Report Name	Frequency	Parameters	Test Scenario
Interaction Volume Business Result Report	Daily	Pre-set Date Filter = None Start Date = <date 09:00:00 AM> End Date = <date 12:00:00 PM> Media Type = Voice All other prompts = all (default)	The report was executed three times with the same parameters in a static environment: • Exclusively (no other reports at the same time) • Concurrently with two other reports (three reports in total) • Concurrently with five other reports (six reports in total)
Agent Conduct Report	Daily	Pre-set Day Filter = None Report Date = <date> From Hour = 8 To Hour = 9 Agent Group = <group> All other prompts = all (default)	The report was executed twice with the same parameters in a static environment: • Exclusively (no other reports at the same time) • Concurrently with two other reports (three reports in total)

Table 75 summarizes the report performance results for the concurrent report test.

Table 75: Concurrent Report Test Results

Report Name	Concurrency	Duration (secs)	Conclusions
Interaction Volume Business Result Report	1	25	Running the report concurrently with: • Two others did not significantly impact the duration. • Five others had a significant impact on the duration.
	3	26	
	6	39	
Agent Conduct Report	1	39	Running the report concurrently with two others did not significantly impact the duration.
	3	41	

Effect of Concurrent Processing

To identify the impact of concurrent ETL and aggregation processing on report generation time, two daily reports on agent activity were each run on static and dynamic data.

Table 76 summarizes the report and test parameters for the concurrent processing test.

Table 76: Concurrent Processing Test Conditions

Report Name	Frequency	Parameters	Test Scenarios
Agent Conduct Report	Daily	Pre-set Day Filter = today From Hour = 0 To Hour = 24 All other prompts = all (default)	The reports were executed twice on one day with the same parameters: • Once while no interactions were being generated • Once while calls were being generated
Agent Group Business Result Report	Daily	Pre-set Date Filter = None Start Date = <date> End Date = <date + 1 day> All other prompts = all (default)	

Table 77 summarizes the report performance results for the concurrent processing test. The table compares durations for two agent reports that were generated on dynamic data (while interactions were being generated and Genesys Info Mart ETL processing and aggregation were continuing) and on static data (while interactions were not being generated).

Table 77: Comparison of Report Durations for Dynamic and Static Data

Report Name	Durations (secs)		Number of Pages	Conclusions
	Dynamic	Static		
Agent Conduct Report	54	39	1	• Report duration increased when new call data was being generated. • The number of pages in the report did not affect the duration of the report.
Agent Group Business Result Report	40	25	1	

Release 8.0 Conclusions

The GI2 release 8.0 baseline and benchmark performance tests yield the following general conclusions:

- GI2 can sustain satisfactory intraday reporting performance at traffic rates up to 8.6 million interactions per day in a blended environment (voice and multimedia interactions).
- As expected, it generally takes longer to generate a report on an active Genesys Info Mart than on a static Genesys Info Mart.
- The volume of traffic also affects the amount of time that it takes to generate a report.
- At low levels of concurrency (for example, three reports at a time), requesting multiple reports does not significantly impact the time that it takes to generate a report. However, at higher higher levels of concurrency (for example, six reports at a time), there is an impact on report duration.
- The cardinality of attached user data can have a significant impact on aggregation and report performance. Very high cardinality results in a very large number of unique occurrences that the Aggregator must process.

GI2 Release 7.6 Performance

In the 7.6 release, GI2 introduced 17 reports that you can run to summarize the inbound, call-related data stored in your Info Mart 7.6 database. Genesys has conducted a number of laboratory tests on the GI2 reports to assess how long it takes to create or open a report instance under varying conditions such as different database size, different RDBMSs, different numbers of object selections for the hourly and daily reports, and different number of generated pages. The results of these tests can help you determine how GI2 might perform in your own environment.

This section describes the environment in which the tests were run and summarizes the results in the following subsections:

- Environment, page 216
- GI2 7.6 Report Performance for a Microsoft SQL Server Info Mart, page 219
- GI2 7.6 Report Performance for an Oracle Info Mart, page 223
- Release 7.6 Conclusions, page 229

Environment

Hardware/Software Used

All tests were run with BO XI 3.0 and GI2 7.6, operating on an Intel Xeon box with the following specifications:

- 2 x Intel Xeon 3.2 GHz/2 MB
- 4 GB RAM
- 2 HDD x 80 GB SATA
- Dual GigEthernet

The tests were also run with Genesys Info Mart 7.6 and either Oracle 10g or Microsoft SQL Server 2005, on an Intel Clovertown Quad Xeon box with the following specifications:

- 2 x Intel Clovertown Quad Xeon X5365 3.0 GHz/L2=2x4 MB
- 4 GB RAM
- 16 x SCSI HDD x 146 GB x 15K RPM
- Dual GigEthernet

The operating system running on both boxes was Microsoft Windows Server 2003 Enterprise edition. Except where noted, Genesys Info Mart ETL processes were scheduled to run in such a way that they did not compete for system resources with the running of GI2 reports.

Call Flow Model

The call flow scenario for these tests simulated 4 to 6 inbound calls per second entering queue 81001 with 40 percent of them directly routed to agent DNs or IVR ports that were configured as handling resources. The remaining 60 percent were routed through routing point 22002 to one of four ACD queues before distribution to agent DNs, as shown in Figure 79. Additional inbound call flows entered other queues at a low rate, to simulate abandoned and other interactions.

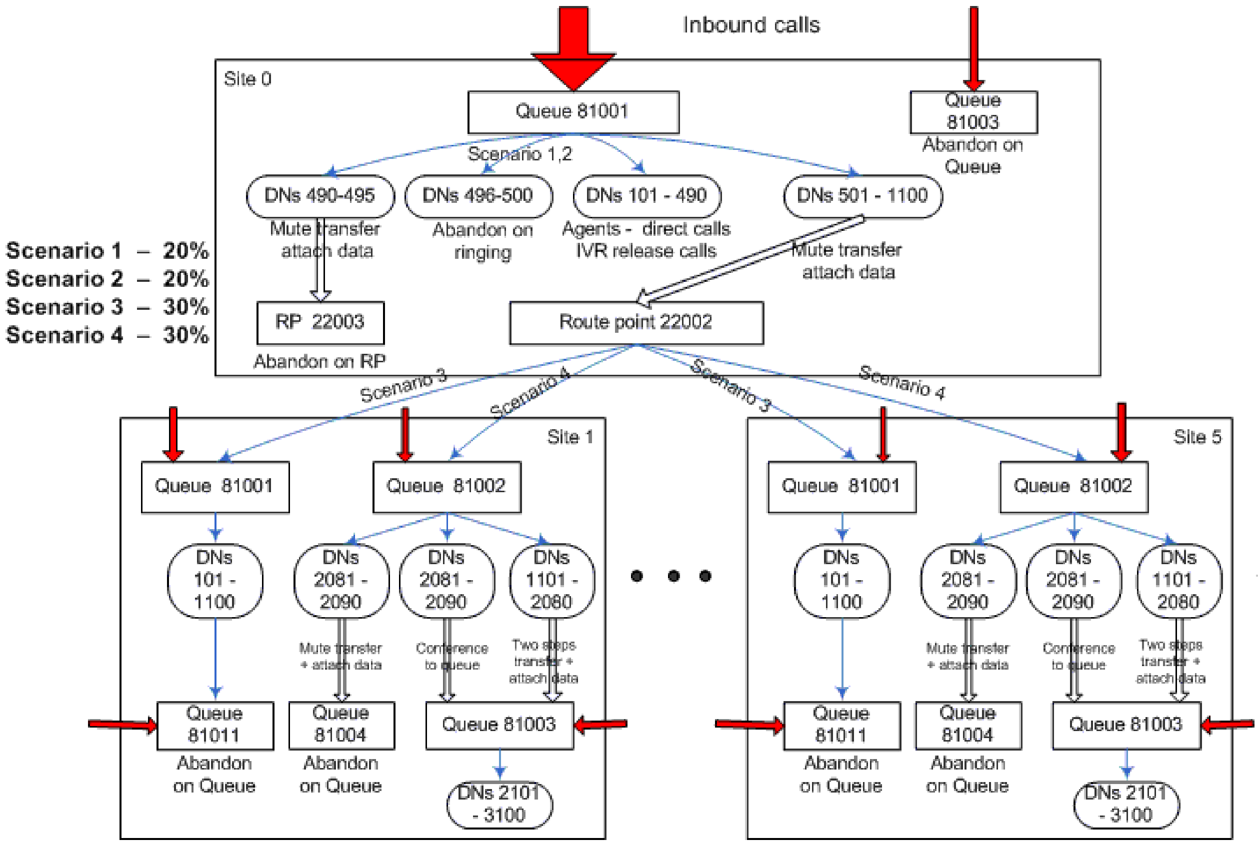


Figure 79: Simulated Call Flows

Attached Data Description

User data was affiliated with each interaction to associate a customer segment, service type and subtypes, business result, and/or service objective, and to apply one of ten string user data keys. This configuration enables data for the Call Volume reports to be populated.

How Time Was Measured

Opening a report consists of refreshing the report's data to re-query and re-retrieve data from the Info Mart, as opposed to displaying the results of an already generated report instance for which data was stored in the report's cube. (Refer to Business Objects documentation for information about the content of a report cube.) The time required to open a report begins when the Run button is clicked in the User Prompt Input area of a report, and continues until the report's results are displayed on-screen. Web Intelligence's internal timer provided the official measurement; the results of this timer are displayed the next time the query is run, in the Running Query message box shown in Figure 80. The level of precision available for this measurement is seconds.

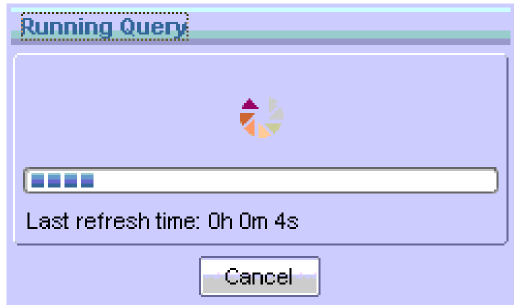


Figure 80: Web Intelligence Running Query Reflects Last Refresh Time

For scheduled reports, two time measurements were taken for the performance results gathered—that of report creation and that of report display. The time spent displaying a report was gathered in the same manner as described above. The time spent creating a report was gathered by viewing the details of the report's success history within Web Intelligence. The information about start and end times is recorded on the HistoryStatus page shown in Figure 81.

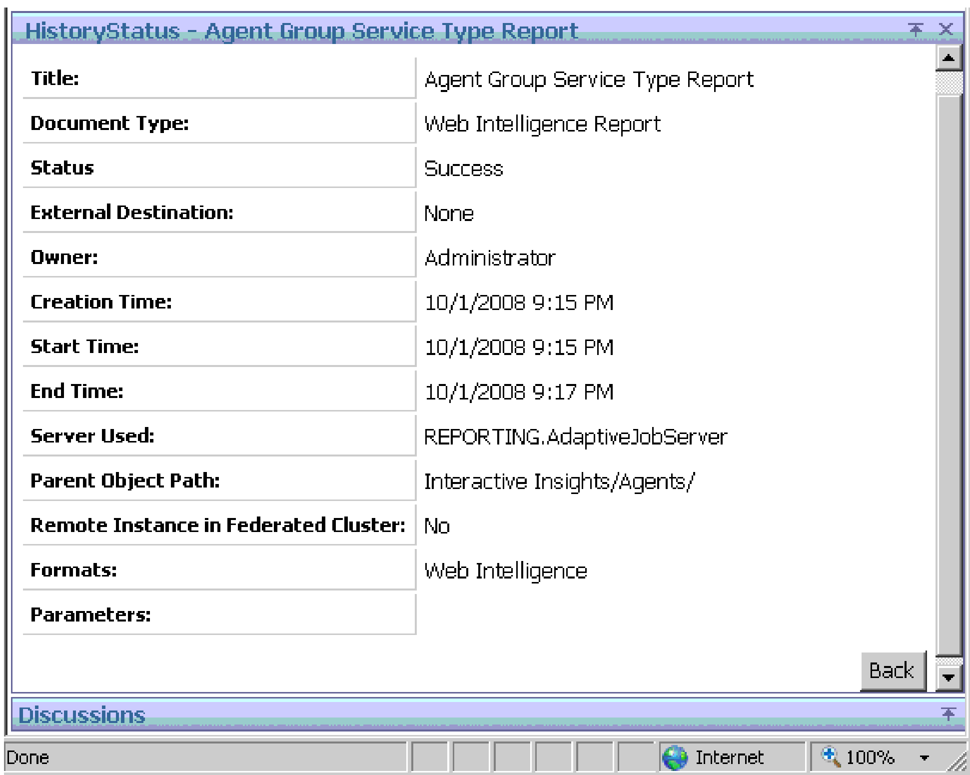


Figure 81: HistoryStatus Page Shows Minutes As Unit of Precision

The level of precision available for report creation measurements is minutes; therefore, a span of time is provided in the results. In Figure 81, for example, the performance results in this chapter would report two to three minutes for the time required to create the Agent Group Service Type Report.

Configuration Objects Used

The objects selected for the Agent reports were three different agent groups: AG1, AG2, and AG3, comprised of 300, 500, and 800 agents, respectively. The objects selected for the Call Volume reports were three service types or service subtypes: T1, T2, and T3. Finally, different report runs for the queue-type reports used selections of:

- One queue object, queue 81003—abbreviated 1Q.
- Two queue objects, queues 81003 and 81004— abbreviated 2Q.
- Three queue objects, queues 81003, 81004, and 81011—abbreviated 3Q.

GI2 7.6 Report Performance for a Microsoft SQL Server Info Mart

Tables 78 through Table 80 show the amount of time, in second(s), that was required to open each Interactive Insight report given a Genesys Info Mart 7.6 source on a Microsoft SQL Server 2005 RDBMS. Database size was 61 gigabytes, the data amassed over 18 days, and none other than the requisite processes were simultaneously competing for system resources while the reports were running. No GenesysInfo Mart 7.6 processes were simultaneously running.

Table 78 shows the time it took to run and open the GI2 reports, which provided hourly breakdowns of the results for the 18-day range over which data was collected.

Table 78: Results of Hourly Reports on Microsoft SQL Server

Agent Reports		Time to Open (s)		
		AG1	AG2	AG3
Agent ACW Report	Time	2	2	3
	Pages	37	61	98
Agent Group Inbound Call Handling Report	Time	3	3	4
	Pages	2	2	2
Agent Inbound Call Handling VQ Report	Time	5	3	5
	Pages	603	690	690
Agent Inbound Utilization Report	Time	5	3	3
	Pages	61	101	161
Agent Interval Based Report	Time	5	5	7
	Pages	101	167	267
Agent Not Ready Reason Code Report	Time	60	60	60
	Pages	43	72	115
Agent Not Ready Report	Time	5	5	8
	Pages	155	233	276

Table 78: Results of Hourly Reports on Microsoft SQL Server (Continued)

Agent Service Report	Time	6	6	6
	Pages	3	3	3
Daily Agent Login-Logout Report	Time	2	2	2
	Pages	185	305	485
Daily Agent State Detail Report	Time	13	13	13
	Pages	57	89	124
Business Result Reports		**T1**	**T1,T2**	**T1-T3**
Call Volume Service Subtype Report	Time	3	5	5
	Pages	2	3	4
Call Volume Service Report	Time	3	3	3
	Pages	1	1	1
Queue Reports		**1Q**	**2Q**	**3Q**
Abandon Delay Report	Time	4	4	5
	Pages	1	1	2
Inbound Voice Traffic Group Report	Time	3	4	4
	Pages	1	1	1
Inbound Voice Traffic Report	Time	3	3	3
	Pages	1	1	1
Queue-Virtual Queue Summary Report	Time	3	3	3
	Pages	1	1	1
Speed of Answer Report	Time	3	3	4
	Pages	1	1	1

Table 79 provides these results in daily breakdowns of the same data over the same date range.

Table 79: Results of Daily Reports on Microsoft SQL Server

Agent Reports		**Time to Open (s)**		
		AG1	**AG2**	**AG3**
Agent Group Inbound Call Handling Report	Time	60	60	65
	Pages	5	5	5
Agent Inbound Call Handling VQ Report	Time	56	56	60
	Pages	162	269	424
Agent Inbound Utilization Report	Time	120	120	125
	Pages	281	462	741
Agent Not Ready Reason Code Report	Time	62	62	65
	Pages	18	18	21
Business Result Reports		**T1**	**T1-T2**	
Call Volume Service Subtype Report	Time	352	680	
	Pages	3	5	
Call Volume Service Report	Time	356	690	
	Pages	2	3	

Table 79: Results of Daily Reports on Microsoft SQL Server (Continued)

Queue Reports		1Q	2Q	3Q
Abandon Delay Report	Time	50	236	415
	Pages	2	2	3
Inbound Voice Traffic Group Report	Time	190	320	410
	Pages	1	1	1
Inbound Voice Traffic Report	Time	315	575	
	Pages	2	2	
Queue-Virtual Queue Summary Report	Time	220	480	
	Pages	2	3	
Speed of Answer Report	Time	220	580	
	Pages	2	3	

Table 80 shows the time it took to create report instances and to open the reports for two scheduled reports.

Table 80: Results of Scheduled Hourly Reports on Microsoft SQL Server

Agent Reports for AG3	Time to Create Report	Time to Open Report
Agent Not Ready Reason Code Report	1– 2 minutes	2 seconds
Daily Agent Login-Logout Report	1– 2 minutes	3 seconds

Report Performance for a Microsoft SQL Server Info Mart

Tables 81 through 83 show the amount of time, in second(s), that was required to open each Interactive Insight report given a Genesys Info Mart 7.6 source on a Microsoft SQL Server 2005 RDBMS. Database size was 61 gigabytes, the data amassed over 18 days, and none other than the requisite processes were simultaneously competing for system resources while the reports were running. No Genesys Info Mart 7.6 processes were simultaneously running.

Table 81 shows the time it took to run and open the GI2 reports, which provided hourly breakdowns of the results for the 18-day range over which data was collected. Table 82 provides these results in daily breakdowns of the same data over the same date range. Table 83 shows the time it took to create report instances and to open the reports for two scheduled reports.

Table 81: Results of Hourly Reports on Microsoft SQL Server

Agent Reports		Time to Open (s)		
		AG1	AG2	AG3
Agent ACW Report	Time	2	2	3
	Pages	37	61	98
Agent Group Inbound Call Handling Report	Time	3	3	4
	Pages	2	2	2

Table 81: Results of Hourly Reports on Microsoft SQL Server (Continued)

Agent Inbound Call Handling VQ Report	Time	5	3	5
	Pages	603	690	690
Agent Inbound Utilization Report	Time	5	3	3
	Pages	61	101	161
Agent Interval Based Report	Time	5	5	7
	Pages	101	167	267
Agent Not Ready Reason Code Report	Time	60	60	60
	Pages	43	72	115
Agent Not Ready Report	Time	5	5	8
	Pages	155	233	276
Agent Service Report	Time	6	6	6
	Pages	3	3	3
Daily Agent Login-Logout Report	Time	2	2	2
	Pages	185	305	485
Daily Agent State Detail Report	Time	13	13	13
	Pages	57	89	124
Business Result Reports		**T1**	**T1,T2**	**T1-T3**
Call Volume Service Subtype Report	Time	3	5	5
	Pages	2	3	4
Call Volume Service Report	Time	3	3	3
	Pages	1	1	1
Queue Reports		**1Q**	**2Q**	**3Q**
Abandon Delay Report	Time	4	4	5
	Pages	1	1	2
Inbound Voice Traffic Group Report	Time	3	4	4
	Pages	1	1	1
Inbound Voice Traffic Report	Time	3	3	3
	Pages	1	1	1
Queue-Virtual Queue Summary Report	Time	3	3	3
	Pages	1	1	1
Speed of Answer Report	Time	3	3	4
	Pages	1	1	1

Table 82: Results of Daily Reports on Microsoft SQL Server

Agent Reports		**Time to Open (s)**		
		AG1	**AG2**	**AG3**
Agent Group Inbound Call Handling Report	Time	60	60	65
	Pages	5	5	5
Agent Inbound Call Handling VQ Report	Time	56	56	60
	Pages	162	269	424

Table 82: Results of Daily Reports on Microsoft SQL Server (Continued)

Agent Inbound Utilization Report	Time	120	120	125
	Pages	281	462	741
Agent Not Ready Reason Code Report	Time	62	62	65
	Pages	18	18	21
Business Result Reports		**T1**	**T1-T2**	
Call Volume Service Subtype Report	Time	352	680	
	Pages	3	5	
Call Volume Service Report	Time	356	690	
	Pages	2	3	
Queue Reports		**1Q**	**2Q**	**3Q**
Abandon Delay Report	Time	50	236	415
	Pages	2	2	3
Inbound Voice Traffic Group Report	Time	190	320	410
	Pages	1	1	1
Inbound Voice Traffic Report	Time	315	575	
	Pages	2	2	
Queue-Virtual Queue Summary Report	Time	220	480	
	Pages	2	3	
Speed of Answer Report	Time	220	580	
	Pages	2	3	

Table 83: Results of Scheduled Hourly Reports on Microsoft SQL Server

Agent Reports for AG3	Time to Create Report	Time to Open
Agent Not Ready Reason Code Report	1– 2 minutes	2 seconds
Daily Agent Login-Logout Report	1– 2 minutes	3 seconds

GI2 7.6 Report Performance for an Oracle Info Mart

Two sets of tests were run and results gathered, to measure the time required to open reports given a relatively small-sized Info Mart (30 gigabytes) and a relatively medium-sized one (125 gigabytes) on an Oracle 10g RDBMS. For the smaller Info Mart, data amassed over a 7-day period with call flow volume of 4 to 6 calls per second. For the larger Info Mart, data amassed over a simulated 365-day period with the same call flow volume. The results are provided in "Results for a Small-Sized Oracle Info Mart" and "Results for a Medium-Sized Oracle Info Mart" on page 226 sections below.

Results for a Small-Sized Oracle Info Mart

The objects selected for the reports were the same as those selected for the Microsoft SQL Server tests on page 219. Table 84 shows an hourly breakdown of the time it took to run and open reports for the 7-day range over which data was collected. Table 85 shows a daily breakdown of the same data. Table 86 shows the time it took for a scheduled process to create a report instance for the Agent Not Ready Reason Code Report and open the report. Finally, Table 87 shows results, comparable to those in Table 84, when Genesys Info Mart ETL processes were running and competing for system resources.

Table 84: Results of Hourly Reports for a Small-Sized Oracle Info Mart

Agent Reports		Time to Open (s)		
		AG1	AG2	AG3
Agent ACW Report	Time	3	3	3
	Pages	133	218	347
Agent Group Inbound Call Handling Report	Time	4	4	4
	Pages	1	1	2
Agent Inbound Call Handling VQ Report	Time	4	4	4
	Pages	222	222	610
Agent Inbound Utilization Report	Time	5	5	5
	Pages	14	54	114
Agent Interval Based Report	Time	4	4	5
	Pages	198	327	521
Agent Not Ready Reason Code Report	Time	36	36	36
	Pages	38	63	101
Agent Not Ready Report	Time	3	3	4
	Pages	130	214	341
Agent Service Type Report	Time	6	6	6
	Pages	3	3	3
Daily Agent Login-Logout Report	Time	3	3	3
	Pages	110	167	301
Daily Agent State Detail Report	Time	3	3	3
	Pages	148	249	400

Table 84: Results of Hourly Reports for a Small-Sized Oracle Info Mart (Continued)

Business Result Reports		**T1**	**T1,T2**	**T1-T3**
Call Volume Service Subtype Report	Time	3	3	3
	Pages	2	3	4
Call Volume Service Type Report	Time	3	3	3
	Pages	1	1	1
Queues Reports		**1Q**	**2Q**	**3Q**
Abandon Delay Report	Time	3	3	3
	Pages	1	1	1
Inbound Voice Traffic Group Report	Time	3	3	3
	Pages	1	1	1
Inbound Voice Traffic Report	Time	3	3	3
	Pages	1	1	1
Queue-Virtual Queue Summary Report	Time	3	3	3
	Pages	1	1	1
Speed of Answer Report	Time	3	3	3
	Pages	1	1	1

Table 85: Results of Daily Reports for a Small-Sized Oracle Info Mart

Agent Reports		Time to Open (s)		
		AG1	**AG2**	**AG3**
Agent Group Inbound Call Handling Report	Time	18	18	22
	Pages	3	3	4
Agent Inbound Call Handling VQ Report	Time	5	5	5
	Pages	102	177	279
Agent Inbound Utilization Report	Time	10	10	14
	Pages	157	277	457
Agent Service Type Report	Time	8	8	8
	Pages	12	12	14
Business Result Reports		**1T**	**2T**	**3T**
Call Volume Service Subtype Report	Time	2	2	2
	Pages	4	8	11
Call Volume Service Report	Time	2	2	2
	Pages	1	2	2

Table 85: Results of Daily Reports for a Small-Sized Oracle Info Mart (Continued)

Queue Reports		1Q	2Q	3Q
Abandon Delay Report	Time	2	2	2
	Pages	1	2	2
Inbound Voice Traffic Group Report	Time	2	2	3
	Pages	1	1	1
Inbound Voice Traffic Report	Time	3	3	3
	Pages	1	2	2
Queue-Virtual Queue Summary Report	Time	3	3	3
	Pages	1	2	2
Speed of Answer Report	Time	3	3	3
	Pages	1	1	1

Table 86: Results of Scheduled Hourly Reports on Oracle

Agent Reports for AG3	Time to Create	Time to Open
Agent Not Ready Reason Code Report	< =1 minute	2 seconds

Table 87: Results of Scheduled Hourly Reports with Genesys Info Mart Running

Agent Reports for AG1	Time to Open (s)	
Agent ACW Report	Time	8–12
	Pages	134
Daily Agent State Detail Report	Time	22–28
	Pages	161
Business Result Reports (All Service Types/Subtypes)	**Time to Open (s)**	
Call Volume Service Type Report	Time	2–4
	Pages	2
Queues Reports (All Queues)		
Inbound Voice Traffic Group Report	Time	2
	Pages	1
Queue-Virtual Queue Summary Report	Time	2–3
	Pages	2

Results for a Medium-Sized Oracle Info Mart

The objects selected for the reports were the same as those selected for the Microsoft SQL Server tests on page 219. Table 88 shows an hourly breakdown of the time it took to run and open reports for the simulated 300-day range over which data was collected with no Genesys Info Mart processes simultaneously running. Table 89 shows the same results in a daily breakdown. Table 90 shows the time to create and open one scheduled report, Agent Not Ready Reason

Code Report. Finally, Table 91 shows the results when six users concurrent running reports. This table repeats the corresponding tests and results that were run in Table 88, to enable comparison between single and concurrent report runs.

Table 88: Results of Hourly Reports for a Medium-Sized Oracle Info Mart

Agent Reports		Time to open (s)		
		AG1	AG2	AG3
Agent ACW Report	Time	146	157	166
	Pages	124	244	327
Agent Group Inbound Call Handling Report	Time	12	15	16
	Pages	2	2	2
Agent Inbound Call Handling VQ Report	Time	45	52	74
	Pages	217	217	603
Agent Inbound Utilization Report	Time	101	111	117
	Pages	61	101	161
Agent Interval Based Report	Time	290	330	330
	Pages	186	365	490
Agent Not Ready Reason Code Report	Time	690	745	770
	Pages	61	101	161
Agent Not Ready Report	Time	146	158	170
	Pages	124	243	326
Agent Service Type Report	Time	56	64	73
	Pages	13	18	20
Daily Agent State Detail Report	Time	112	115	113
	Pages	91	151	200
Business Result Reports		**T1**	**T1,T2**	**T1-T3**
Call Volume Service Subtype Report	Time	2	2	2
	Pages	2	4	4
Call Volume Service Type Report	Time	2	2	2
	Pages	1	1	2
Queue Reports		**1Q**	**2Q**	**3Q**
Abandon Delay Report	Time	2	2	2
	Pages	1	1	1
Inbound Voice Traffic Group Report	Time	2	2	2
	Pages	1	1	1
Inbound Voice Traffic Report	Time	2	2	2
	Pages	1	1	1
Queue-Virtual Queue Summary Report	Time	2	2	2
	Pages	1	1	1
Speed of Answer Report	Time	2	2	2
	Pages	1	1	1

Table 89: Results of Daily Reports for a Medium-Sized Oracle Info Mart

Agent Reports		Time to open (s)		
		AG1	AG2	AG3
Agent Group Inbound Call Handling Report	Time	33	34	35
	Pages	2	2	2
Agent Inbound Call Handling VQ Report	Time	14	14	27
	Pages	262	262	728
Agent Inbound Utilization Report	Time	112	113	114
	Pages	76	126	201
Agent Service Type Report	Time	55	63	67
	Pages	13	18	20
Business Result Reports		**T1**	**T1,T2**	**T1-T3**
Call Volume Service Subtype Report	Time	2	2	2
	Pages	2	4	6
Call Volume Service Report	Time	2	2	2
	Pages	1	1	2
Queues Reports		**1Q**	**2Q**	**3Q**
Abandon Delay Report	Time	2	2	2
	Pages	1	1	1
Inbound Voice Traffic Group Report	Time	2	2	2
	Pages	1	1	1
Inbound Voice Traffic Report	Time	2	2	2
	Pages	1	1	1
Queue-Virtual Queue Summary Report	Time	2	2	2
	Pages	1	1	1
Speed of Answer Report	Time	2	2	2
	Pages	1	1	1

Table 90: Scheduled Daily Report on Oracle

Agent Report for AG3		Time to Create	Time to Open
Agent Not Ready Reason Code Report	Time	<= 12 minutes	2 seconds

Note: The daily results of the single report runs in Table 91 were generated based on data from a different day or span of days than the results reported in Table 88.

Table 91: Comparison of Daily Results for Concurrent and Noncurrent Report Users on Oracle

	Pages	Time to Open (s)	
		Single	Concurrent
Agent Reports for AG1			
Agent ACW Report	124	157	166–173
Agent Group Inbound Call Handling Report	2	7	7–9
Agent Inbound Call Handling VQ Report	42	45	101–107
Agent Inbound Utilization Report	61	81	88–94
Agent Interval Based Report	186	278	316–340
Agent Not Ready Reason Code Report	61	660	726–752
Agent Not Ready Report	124	140	179–182
Agent Service Type Report	3	20	22–23
Daily Agent State Detail Report	91	109	116–120
Business Result Reports for T1			
Call Volume Service Subtype Report	2	2	2
Call Volume Service Type Report	1	2	2
Queue Reports 1Q			
Abandon Delay Report	1	2	2
Inbound Voice Traffic Group Report	1	2	2
Inbound Voice Traffic Report	1	2	2
Queue/Virtual Queue Summary Report	2	1	1–2
Speed of Answer Report	2	2	1–2

Release 7.6 Conclusions

As expected, the number of objects selected affects the time required to create and open a report. It generally takes longer to create a report than it does to open it. The larger the selection pool (the number of generated pages), the longer the report creation and open time. Reports take longer to open with concurrent usage than nonconcurrent (single) usage.

Chapter

10 Genesys Workforce Management

This chapter presents the following information about Genesys Workforce Management:

- Software Co-location Recommendations, page 231
- Small Contact Centers, page 232
- Medium Contact Centers, page 234
- Dedicated Report Server, page 235
- Client Workstation Requirements, page 235

Software Co-location Recommendations

Genesys makes the following co-location recommendations for WFM deployments:

- Co-locate the WFM Server and the WFM database on the same LAN segment. The network latency between these components should be no more than 1 millisecond (ms).
- Co-locate the WFM Data Aggregator (and its backup, if any) on the same LAN segment as the WFM database. The network latency between these components should be no more than 1 ms.
- Co-locate the WFM Server and WFM Data Aggregator. The network latency between these components should be less than 20 ms.
- Co-locate the WFM Data Aggregator and Genesys Stat Server. The network latency between these components should be less than 20 ms. Genesys recommends that you use a Stat Server that is dedicated for the WFM Solution, especially in environments where Stat Server is initially located on a different LAN segment than WFM Data Aggregator.

- Co-locate the WFM Configuration Utility and the WFM database. The network latency between these components should be less than 10 ms.
- Co-locate the WFM Database Utility and the WFM database. The network latency between these components should be less than 10 ms.
- In environments where Configuration Server is deployed in a WAN, Genesys recommends that you use a Configuration Server Proxy that is dedicated for WFM Solution.

Notes: Insufficient network bandwidth or excessive network latency can result in the following behaviors:

- Between WFM Server and WFM database, there could be significant delays in the WFM Web interface for most operations. Most noticeable, saving schedules and calendar operations.
- Between WFM Configuration Utility and WFM database, there can be significant delays in interface responses, when objects are viewed or saved.
- Between WFM Data Aggregator and Genesys Stat Server, there is a risk of delays in communication between these components, which might result in missing data. Also, Stat Server could close the connection to WFM Data Aggregator with the reason client too slow.
- Latency between WFM Data Aggregator and WFM database can lead to excessive memory consumption on the Data Aggregator, due to the extended amount of time the server takes to write data to the WFM database.
- Latency between WFM Data Aggregator and WFM Server can cause WFM Web to display outdated schedule state data in the real-time adherence view.

Note: If a dedicated Configuration Server Proxy is not used for the WFM Solution and timeouts or frequent disconnects from Configuration Server occur, these issues might be resolved by adding a dedicated Configuration Server Proxy.

Small Contact Centers

Workforce Management hardware recommendations for small contact centers (less than 500 agents) have two options—Options A and B.

Option A is to use several servers and is valid for 32-bit and 64-bit operating systems. See Table 92.

Note: For small contact centers, limit the schedule building process to 150 activities at the time.

Table 92: Small Contact Centers (less than 500 Agents)—Option A

Small Contact Center	
Workforce Management - Server 1 **Tomcat/WebSphere, WFM Web Server, WFM Server, WFM Data Aggregator:**	
Server 1	GEN_WIN_SERVER (40 GB HDD). **Note:** A 40 GB HDD is just a baseline For specific requirements, see the Genesys Customer Care web site at http://www.genesys.com/customer-care. In most cases, one Ethernet card is enough.
Workforce Management - Server 2 **WFM Builder, WFM Daemon:**	
Server 2	GEN_WIN_SERVER (2 GB RAM, 40 GB HDD). **Note:** 2 GB RAM and a 40 GB HDD is just a baseline. For specific requirements, see the Genesys Customer Care web site at http://www.genesys.com/customer-care. In most cases, one Ethernet card is enough.

Use Option B if you plan to co-locate the whole WFM back end on one server It requires a 64-bit operating system. See Table 93.

Table 93: Small Contact Centers (less than 500 Agents)—Option B

Small Contact Center	
Workforce Management - Server 1 **Tomcat/WebSphere, WFM Web Server, WFM Server, WFM Data Aggregator, WFM Builder, WFM Daemon:**	
Server 1	Double the GEN_WIN_SERVER requirements for the number of CPU cores and the amount of memory. 2 CPU Intel, 2.6 GHz quad core, 8 GB RAM, 150 GB HDD, 64-bit OS. In most cases, one Ethernet card will be enough.

Medium Contact Centers

Workforce Management hardware recommendations for medium contact center (500-1000 agents) have two options.

Option A is to use several servers and is valid for 32-bit and 64-bit operating systems. See Table 94.

Note: For medium sized contact centers, limit the schedule building process to 400 activities at the time.

Table 94: Medium Sized Contact Centers (500-1000 Agents)—Option A

Medium Sized Contact Center	
Workforce Management - Server 1 Tomcat/WebSphere, WFM Web Server:	
Server 1	GEN_WIN_SERVER (40 GB HDD). **Note:** A 40 GB HDD is just a baseline For specific requirements, see the Genesys Customer Care web site at http://www.genesys.com/customer-care. In most cases, one Ethernet card is enough.
Workforce Management - Server 2 WFM Server, WFM Data Aggregator, WFM Daemon:	
Server 2	GEN_WIN_SERVER (2 GB RAM, 40 GBHDD). **Note:** 2 GB RAM and a 40 GB HDD is just a baseline. For specific requirements, see the Genesys Customer Care web site at http://www.genesys.com/customer-care. In most cases, one Ethernet card is enough.
Workforce Management - Server 3 WFM Builder:	
Server 3	GEN_WIN_SERVER (2 GB RAM 40 GBHDD). **Note:** 2 GB RAM and a 40 GB HDD is just a baseline, for specific requirements, see the Genesys Customer Care web site at http://www.genesys.com/customer-care. In most cases, one Ethernet card is enough.

Use option B if you plan to co-locate the whole WFM back end on one server. It requires a 64-bit operating system. See Table 95.

Table 95: Medium Sized Contact Centers (500-1000 Agents)—Option B

Meeium Sized Contact Center	
Workforce Management - Server 1 **Tomcat/WebSphere, WFM Web Server, WFM Server, WFM Data Aggregator, WFM Builder, WFM Daemon:**	
Server 1	Triple the GEN_WIN_SERVER requirements for the number of CPU cores and the amount of memory. 2 CPU Intel, 2.6 GHz six core or 4 CPU Intel 2.6 GHz quad core, 16 GB RAM, 300 GB HDD, 64-bit OS. In most cases, one Ethernet card is enough.

Dedicated Report Server

Starting with WFM 7.2, there is no longer a separate WFM server component for building reports. Starting with WFM 7.6.1 the report-generating functionality is integrated into the WFM Web component. Some contact centers might want to maintain a dedicated server for this process, because report generation can be CPU-intensive. For the same reason, contact centers might not want the same instance of WFM Web that is used to generate reports to also serve up all agent and supervisor-facing user interfaces. Therefore, an instance of WFM Web can be deployed on an additional server and dedicated to report-building only. For instructions on how to configure this, see the current *Workforce Management Administrator's Guide.*

Note: When WFM Builder and the instance of WFM Web that is dedicated to report generation are deployed on the same machine, in some cases usage might overlap and cause delays for pending user requests. If this becomes an issue, deploy WFM Builder and the WFM Web dedicated to report generation on separate machines.

Client Workstation Requirements

Computers running the WFM client applications WFM Configuration Utility, WFM Database Utility, WFM Web for Supervisors, and WFM Web for Agents at minimum, must meet these specifications:

- WFM Configuration Utility, WFM Database Utility, or WFM Web for Supervisors: GEN_WIN_DESKTOP
- WFM Web for Agents, see the hardware requirements for the specific version of internet browser used.

Chapter

11 Genesys Voice Platform 8.1

This chapter specifies the recommended hardware and operating systems for Genesys Voice Platform (GVP) 8.1, and provides information about capacity and performance.

The performance criteria, machine setup, application profile and typical call volume information is written with the following audience in mind: Engineers, Product Managers, Program Managers, Quality Assurance testers, Technical Publications writers, Production personnel, Genesys Partners and Genesys customers.

This chapter contains the following sections:

- Capacity and Traffic Terminology, page 238
- Recommended Hardware and Operating Systems, page 242
- Traffic and Capacity Testing, page 245
- Performance Planning and Scalability, page 293
- Hardware and Bandwidth Usage, page 349

Capacity and Traffic Terminology

This chapter defines and describes the relationships between these capacity and traffic measurement metrics in a GVP system:

Capacity Metrics

- Call Arrival Per Second (CAPS)
- Port Density (PD), also known as Peak Ports

Performance Metrics

- Call Setup Latency (CSL)
- Caller Perceived Latency (CPL), also known as ResponseTime Latency
- Call Passrate (PR)
- Call Duration (CD)
- Peak Capacity (PC)

System capacity is defined as a function of the maximum number of ports or maximum call-arrival rate (CAPS) at which GVP can maximize its use of hardware resources while maintaining all of the criteria within a predefined threshold.

This section also provides the formulas used to calculate capacity and performance.

Capacity Metrics and Formulas

Two units of measure are used for capacity planning—Call Arrival Per Second and Port Density.

Call Arrivals Per Second

CAPS is used to measure traffic within the system. For example, 10 CAPS means that GVP is receiving 10 calls every second, which is considered *busy traffic*.

CAPS is similar to Busy Hour Call Attempts (BHCA) or Centum Call Seconds (CCS), which is the legacy engineering term for telephony traffic.

Use the following formula to calculate CAPS in terms of CCS:

CAPS = CCS/36

CAPS measures can be applied to components which handle messages or data associated with a call. For example, the reporting server will have a CAPS value based on the number of call records written to it, which will often relate one-to-one with a completed call.

Throughout this chapter, including in the tables, the capacity of a function and/or component is defined by its Peak CAPS (the maximum number of calls per second that the system can handle for that function without suffering from latency). See the Capacity Criteria section below for more information.

Port Density

PD is the maximum number of simultaneous calls that can be served by GVP at any given time. In the tables of this document, PD is called “Peak Ports” because it also specifies the number of ports that are required to handle the call traffic.

Use the following formula to calculate Port Density:

PD = CAPS x Avg(CD)

...where Avg = Average.

Performance Metrics and Formulas

Four units of measure are used to assess performance—Call Duration, Call Setup Latency, Caller Perceived Latency, and Call Passrate.

Call Setup Latency

CSL is defined as the delay between the initial SIP INVITE message and the first audible RTP packet sent from GVP. One example is the dialogue in a typical call flow shown in Figure 82:

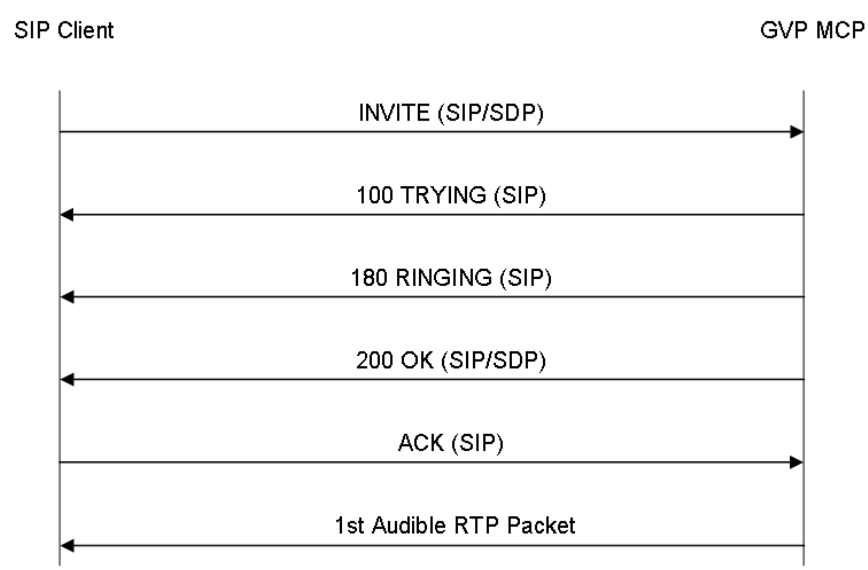


Figure 82: Typical SIP Call Flow #1

CSL consists of the following requests and responses:

- User SIP INVITE request received > SIP 200 OK response sent.
- SIP 200 OK response sent > User SIP ACK request sent.
- User SIP ACK request sent > First audible media response sent.

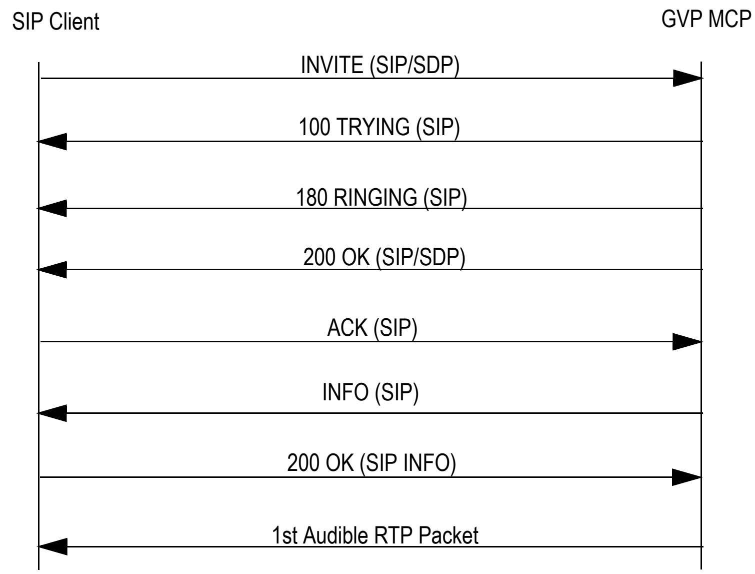


Figure 83: Typical SIP Call Flow #2

Call setup latency is the delay between when the initial SIP INVITE (top line from the typical call flow shown in Figure 83) is received to the time that the first audible packet (bottom line from the diagram above) is sent out by GVP.

Other than call setup latency, SIP INFO response latency (from SIP INFO with MSML embedded to 200 OK response from MCP for SIP INFO request) is also an important factor. It should be measured and reported separately, although this duration is already part of call setup latency.

Caller Perceived Latency

CPL (or Response Time Latency [RTL]) is defined as the time between the last user input (speech or DTMF) and the next prompt. In Figure 84, "Caller Perceived Latency," on page 241, the time between T1 and T2 is the period of CPL.

Calculate CPL by using the formula, CPL = T2 - T1.

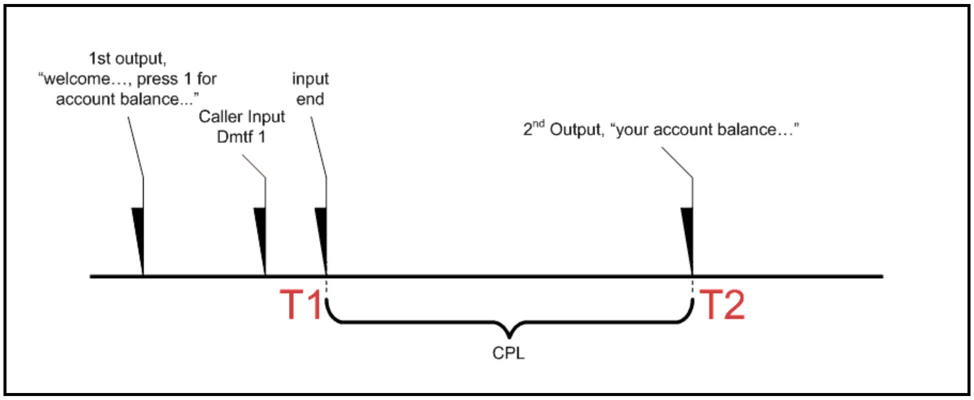


Figure 84: Caller Perceived Latency

CPL is impacted by the following factors:

- Recognition engines.
- End of speech or a DTMF time out.
- Application pages and prompts.
- Grammars caching and fetching mechanisms.
- The size of application pages.
- Call traffic, including call arrival rate and call duration.
- Speech density—during a call, the percentage of time that speech or DTMF detection is on and the caller can provide input.
- The size of speech recognition grammars and the how often they are used in an application.
- Back end operation—the length of time required to obtain information from the business layer (such as, database queries and CRM findings) and return the results to the caller.

Call Passrate

Call Passrate (CP) is defined as the number of calls that finished the predefined call flow successfully during a performance load test.

Use the following formula to calculate the CP:

Assume the number of calls is 1000 and 5 calls did not finish the predefined call flow successfully.

(1000-5)/1000 = 99.5%

The capacity measurement formulas are:

ASR/TTS-dependent application	Passrate >= 99.95%	Error rate <= 0.05%
DTMF-only application	Passrate >= 99.99%	Error rate <= 0.01%

Similar to the formulas for Call Duration, *x* is considered the peak capacity for this criteria when the threshold is reached at call volume *x*.

Call Duration and Peak Capacity

Call Duration (CD) is the length of time that a call stays in the GVP system.

Use CAPS and CD to calculate the port density required for handling such traffic.

Instead of measuring individual Caller Perceived Latencies within an application under test, you can use data collected by GVP to measure the increase in the total call duration to determine system loading.

For a single call, the assumption is that the normal call duration (where the hang-up is performed by the application) is CD1. When the load increases on the system, the call duration is expected to increase due to an increase of latencies within the application. Assume that for x simultaneous calls in the system, the measured average call duration is Avg(CDx) and the measured 95th percentile call duration is 95%tile(CDx). The capacity measurement goal is

Avg(CDx) / CD1 <= 110%
95%tile(CDx) / CD1 <= 120%

When the 110% (and 120%) threshold is reached, the call volume x is considered to be the peak capacity for this criterion.

Recommended Hardware and Operating Systems

With the exception of the operating system itself, the recommended specifications in Table 96 are the same for Windows and Linux.

Table 96: Hardware and Operating Systems

Hardware specification	Recommendation
CPU	Dual Quad Core Xeon 2.66 GHz or higher (For optimal performance, Genesys recommends Xeon with Core 2 technology.)
Memory	4 GB RAM minimum, 8GB recommended
Network	Gigabit or 100 Megabit Ethernet
Storage	RAID 1 HDD with at least 40 GB with 15 K RPM

Table 96: Hardware and Operating Systems (Continued)

Hardware specification	Recommendation
Operating systems	Microsoft Windows 2003 Standard Edition Microsoft Windows 2003 R2, Enterprise Edition, SP2 Microsoft Windows 2008, Enterprise Edition, SP2, x64 Microsoft Windows 2008, Enterprise Edition, SP2, x86 Microsoft Windows 2008 R2, Enterprise Edition, SP1, x64
	Redhat Enterprise Linux 4.0, Update 6 (or later) Redhat Enterprise Linux 5.0, Update 5 (or later), x64 Redhat Enterprise Linux 5.0, Update 5 (or later), x86 Redhat Enterprise Linux 6.0, x64

Note: The benchmark CPU is a Dual Quad Core Xeon X5355 2.66GHz, but other CPUs are also used for testing:

- Dual Quad Core Xeon E5620 2.40GHz 16GB RAM
- Dual Hexa Core Xeon X5675 3.06GHz 32GB RAM
- Single Hex Core Xeon X5675 3.06GHz 12GB RAM
- Single Hex Core Xeon X5670 2.93GHz 12GB RAM
- Single Dual Core Xeon X5160 3.0GHz 8GB RAM

Reporting Server

Recommendations

If you intend to deploy Reporting Server in partitioning mode in your high performance environment, you must ensure you are using a supported operating system version. Before you deploy the Reporting Server, consider the information in Table 97.

Table 97: Reporting Server Modes

Reporting Server mode	Supported operating systems
partitioning mode	• Oracle 10g or 11g Enterprise Edition • Microsoft SQL Server 2008 Enterprise Edition
standard mode (no partitioning)	• Oracle 10g or 11g Standard Edition • Oracle 10g or 11g Enterprise Edition • Microsoft SQL Server 2005 or 2008 Standard Edition • Microsoft SQL Server 2005 or 2008 Enterprise Edition
Notes: All versions of Microsoft SQL Server are supported on Windows only. Reporting Server can be installed on Linux, however, the database must be installed off-board on a separate Windows host.	

Reporting Server Performance in Different Modes

The Reporting Server can potentially perform at optimal levels when in partitioning mode. In standard mode, Reporting Server 8.1.3 performance is below optimal and comparable to Reporting Server 8.1.1 performance, which does not support partitioning. However, Reporting Server 8.1.2 and later releases in partitioning mode are much improved over standard mode Reporting Server 8.1.1 in terms of performance.

Traffic and Capacity Testing

Use the information in this section to determine the required capacity of your GVP servers, based on anticipated traffic characteristics or by running tests on an existing system.

When measuring peak capacity on a single GVP machine, CPU usage is usually the determining factor—memory has not been an issue in most test cases. Therefore, the sample test results in this section concentrates on CPU usage and other criteria.

In addition, the Media Resource Control Protocol (MRCP) server that supports Automatic Speech Recognition (ASR) applications, must not share a host with a GVP server. You can use multiple MRCP servers for a particular test, however, it is important that the MRCP resources do not cause a bottleneck during testing.

This section contains VoIP capacity summaries for both Windows and Linux (see Table 100 on page 250) and the following capacity test cases:

- Call Setup Latency Test Case, page 289
- Caller Perceived Latency Test Case, page 290
- Cachable VoiceXML Content Test Cases, page 291

VoIP Capacity Test Summaries

The complexity of VoiceXML and CCXML applications impacts capacity testing, therefore, the Genesys QA performance testing results in this section are derived from test cases using four different VoiceXML applications and two different CCXML applications (some more complex than others).

VoiceXML Application Profiles

VoiceXML performance testing was conducted on four major application profiles. Their characteristics are outlined in Table 98. The call flow duration for each application profile is for a single call or CD1 (see “Call Duration and Peak Capacity” on page 242).

Table 98: VoiceXML Test Application Profiles

Profile Name	Type	Details
VoiceXML_App1	A simple DTMF-only application designed to refill calling cards.	• Total number of digits (DTMF input only) = 52, including: ◦ Toll free number from the back of the card ◦ Refill card PIN number ◦ Refill dollar amount ◦ Credit card number ◦ Credit card expiration date ◦ Zip Code of caller • Number of VoiceXML pages = 18 • VoiceXML complexity = low • Number of audio prompts = 9 • Number of audio files used in prompts (no TTS) = 107 • ECMA script complexity = moderate • Number of VoiceXML pages = 6 • Number of Java script help functions in each VoiceXML page = 13 • Call flow duration: ◦ 74 seconds (pre GVP 8.1.2) ◦ 76 seconds (GVP 8.1.2 and higher)
VoiceXML_App2	A complex application designed for insurance coverage inquiries.	• Speech input, including: ◦ Type of request ◦ ID card number ◦ Confirmation ◦ Relationship with insurance plan holder ◦ Date of birth confirmation • Number of VoiceXML pages = 10 • VoiceXML complexity (~ 1 MB of content) = High • Number of audio prompts = 7 • Number of audio files used in prompts (3 with TTS) = 29 • ECMA script complexity = high • Call flow duration: ◦ 70 seconds (ASR engine) ◦ 55 seconds (ASR engine simulator)

Table 98: VoiceXML Test Application Profiles (Continued)

Profile Name	Type	Details
VoiceXML_App3	QA ASR/TTS load application.	• Speech input, including: ◆ Words ◆ Digits ◆ Hotkey (NGI) ◆ Yes or no confirmation • Number of VoiceXML pages = 1 • VoiceXML complexity = low • Number of audio prompts = 7 prompts involve 7 audio files and 7 TTS • ECMA script complexity = low • Call flow duration = 62 seconds (Next Generation Interpreter [NGI])
VoiceXML_App4	Composer-generated application designed for IVR-assisted banking.	• Input (DTMF only) ◆ Input current customer number ◆ Confirm contact ID ◆ Input debit menu option ◆ Input debit banking menu ◆ Input personal option ◆ Input 6 digit secure code For a total of 20 DTMF digits. • Number of VoiceXML pages = 20 • VoiceXML complexity = medium (~ 400 KB of content) • Number of audio prompts = 6 (no TTS, 12 audio files) • ECMA script complexity = moderate (4 general JavaScript function files) • Call duration = 85 seconds

CCXML Application Profiles

CallControlXML (CCXML) performance testing was conducted on two major application profiles. Their characteristics are outlined in Table 99. The call flow duration for each application profile is for a single call or CD1 (see "Call Duration and Peak Capacity" on page 242).

Table 99: CCXML Test Application Profiles

Profile Name	Type	Details
CCXML_App1	An outbound application that joins multiple call legs, dialogs, and conferences.	• Includes the following steps: ◦ Call customer and connect to a dialog ◦ Call agent and connect to dialog ◦ Exit agent dialog ◦ Exit customer dialog ◦ Create conference ◦ Join customer and agent to conference ◦ Disconnect agent ◦ Disconnect customer ◦ Destroy conference • Number of CCXML (JSP) pages = 2 • CCXML complexity = medium • Customer call duration = 8.7 seconds • Agent call duration = 8.6 seconds • Conference call duration = 6 seconds
CCXML_App2	Simple conference recording call.	• Includes the following steps: ◦ Create a call to agent ◦ Agent receives an invite and a dialog is created for agent to ring back ◦ Agent answers the call and a conference is created to join caller and agent ◦ Conference is established and dialog is created for recording ◦ Call is disconnected from caller after 15 seconds of recording • Number of CCXML pages = 1 • Number of VoiceXML pages = 2 • CCXML complexity = medium • Call duration = 21 seconds

VoIP Capacity Summaries

Some capacity test summaries in this section were performed on systems with hardware specifications other than those recommended in Table 96 on page 242. Major differences in test results can occur, depending on the CPU model and the number of CPUs that are used.

In Table 100, the hardware that was used to test each application type is specified. Unless otherwise stated, the results are based on the Media Control Platforms NGI configuration.

Certain tests may not be conducted with the hardware specified in "Recommended Hardware and Operating Systems" on page 242; the major difference is the CPU model and the number of CPUs being used. The Hardware column in the tables below describes the CPU setup that was used in each test and the observed capacity. The results are based on NGi configured in MCP, unless it is stated as GVPi.

Note: VoiceXML_App3 was used for both single server testing and PSTNC testing. See Table 102, "Single Server All-In-One Capacity Testing," on page 256 and PSTNC Table 108, "PSTN Connector and SSG Capacity Testing," on page 287.

How to Use These Tables

How does a reader approach the difficult task of sizing, in the face of so much raw data contained by the tables in the following section? Each table is prefaced with a description of its intent, and suggestions for interpreting and applying the data.

Click on a link in the list below, and find specific details about intent and use above each table:

- Table 100, "GVP VOIP VXML/CCXML Capacity Testing," on page 250
- Table 101, "Multiple VMs Versus Multiple MCP Capacity Testing," on page 252
- Table 102, "Single Server All-In-One Capacity Testing," on page 256
- Table 103, "Media Control Platform Capacity Testing (Physical Servers)," on page 258Table 103, "Media Control Platform Capacity Testing (Physical Servers)," on page 258
- Table 104, "Media Control Platform / Media Server Capacity (Virtual Servers)," on page 263
- Table 105, "Resource Manager and MRCP Proxy Capacity Testing," on page 273
- Table 106, "Reporting Server Capacity Testing," on page 279
- Table 107, "CTI Connector with IVRSC and CTI Connector with ICM Capacity Testing," on page 281

- Table 108, "PSTN Connector and SSG Capacity Testing," on page 287

GVP VOIP Capacity Testing Table

Table 100 shows the fundamental performance of a single physical server process in terms of peak throughput and peak port capacity, either VoiceXML applications for MCP or CCXML for CCP. You can use this table as the first basis of your assessment.

Table 100: GVP VOIP VXML/CCXML Capacity Testing

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
		Windows		
VoiceXML_App1	2x Core 2 Quad Xeon x5355, 2.66 GHz	17 (preferred)	1300 (preferred)	See note at the end of this table for a definition of *preferred* and *peak*.
VoiceXML_App1	2x Core 2 Quad Xeon x5355, 2.66 GHz	17	1300	Using TCP and TLS.
VoiceXML_App1	2x Core 2 Quad Xeon x5355, 2.66 GHz	23.6 (peak)	1800 (peak)	Ignore call setup latency threshold on Window 2003 and 2008 R2, x64.
VoiceXML_App1	1x HexCore Xeon x5770 2.66GHz	26 (peak)	2000 (peak)	Ignore call setup latency threshold, Windows 2008 R2 x64 SP1
VoiceXML_App1	2x Core 2 Quad Xeon x5355, 2.66 GHz	10 (preferred)	800 (preferred)	Using GVPi.
VoiceXML_App2	1x HexCore Xeon X5670 2.93GHz	7.2	400	MCP on a physical server. Tested with offboard NSS engine MRCP v1.
VoiceXML_App2	1x HexCore Xeon X5670 2.93GHz	7.2	400	MCP on a physical server. Tested with offboard NSS engine MRCP v2 (NSS 6.2.x + NR 10.2.x + NV 5.7.x) with session XML enabled. GVP 8.1.7 or later.
VoiceXML_App2	1x HexCore Xeon X5670 2.93GHz	8	450	MCP on a physical server. Tested with offboard NSS engine MRCP v2 (NSS 6.2.x + NR 10.2.x + NV 5.7.x) with session XML disabled. GVP 8.1.7 or later.

Table 100: GVP VOIP VXML/CCXML Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
VoiceXML_App2	2x Core 2 Dual Xeon x5160, 3.00 GHz	4.5	250	MCP on a physical server. Tested with simulated speech server.
VoiceXML_App2	2x Core 2 Quad Xeon x5355, 2.66 GHz	1	60 (GVPi)	MCP on a physical server. Tested with Nuance Speech Server.
VoiceXML_App4	2x Core 2 Quad Xeon x5355, 2.66 GHz	9.4	800	
CCXML_App1	2x Core 2 Quad Xeon x5355, 2.66 GHz	30	N/A	
CCXML_App2	2x Core 2 Quad Xeon x5355, 2.66 GHz	20	420	
Linux				
VoiceXML_App1	2x Core 2 Quad Xeon x5355, 2.66 GHz	17 (preferred)	1300 (preferred)	
VoiceXML_App1	2x Core 2 Quad Xeon x5355, 2.66 GHz	23.6 (peak)	1800 (peak)	Peak, ignoring call setup and tear-down latency threshold.
VoiceXML_App1	2x Core 2 Quad Xeon x5355, 2.66 GHz	23.6	1800	Using TCP and TLS.
VoiceXML_App1	2x Core 2 Quad Xeon x5355, 2.66 GHz	14.5	1100	Inband DTMF.
VoiceXML_App2	2x Core 2 Quad Xeon x5355, 2.66 GHz	7.2	400	MCP on a physical server. Tested with simulated speech server.
Note: *Preferred* means the highest capacity that the system can sustain while maintaining optimal user experience. *Peak* means the highest capacity that the system can sustain regardless of the user experience.				

Multiple VMs Versus Multiple MCP Capacity Testing Table

Table 101 provides a comparison of capacity testing results when multiple virtual machines (VMs) are used versus multiple Media Control Platform instances.

This table shows the effect of stacking server processes on the same hardware server where there is one MCP associated with a VM instance on the same hardware server. The effect is the increased total port capacity that you can achieve using stacked processes.

Table 101: Multiple VMs Versus Multiple MCP Capacity Testing

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
Using VMWare				
VoiceXML_App1 1 VM	2x Core 2 Quad Xeon x5355, 2.66 GHz, 12 GB RAM	17	1300	VM images (using VMWare ESXi 5.0) are configured and enabled with 1 Media Control Platform instance only installed in each image. Guest OS is Windows 2008 Server SP2 x86.
VoiceXML_App1 2 VMs	2x Core 2 Quad Xeon x5355, 2.66 GHz, 12 GB RAM	21	1600	
VoiceXML_App1 2 VMs	2x Core 2 Quad Xeon x5355, 2.66 GHz, 12 GB RAM	21	1600	VM images (using VMWare ESXi 5.0) are configured and enabled with 4 Media Control Platform instances—2 installed in each image. Guest OS is Windows 2008 Server SP2 x86.
VoiceXML_App1 4 VMs	2x Core 2 Quad Xeon x5355, 2.66 GHz, 12 GB RAM	29	2200	VM images (using VMWare ESXi 5.0) are configured and enabled with 1 Media Control Platform instance only installed in each image. Guest OS is Windows 2008 Server SP2 x86.

Table 101: Multiple VMs Versus Multiple MCP Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
VoiceXML_App1 4 VMs	2x Core 2 Quad Xeon x5355, 2.66 GHz, 12 GB RAM	26	2000	VM images (using VMWare ESXi 5.0) are configured and enabled with 8 Media Control Platform instances—2 installed in each image. Guest OS is Windows 2008 Server SP2 x86.
VoiceXML_App1 8 VMs	2x Core 2 Quad Xeon x5355, 2.66 GHz, 12 GB RAM	34	2600	VM images (using VMWare ESXi 5.0) are configured and enabled with 1 Media Control Platform instance only installed in each image. Guest OS is Windows 2008 Server SP2 x86.
VoiceXML_App1 1 VM	2x Core 2 Quad Xeon x5355, 2.66 GHz, 4 GB RAM	8 (tested)	600 (tested)	VM image (using VMWare ESXi) is configured and enabled with all GVP components (except Reporting Server) together with SIP server. Guest OS is Windows 2003 Server.
VoiceXML_App1 (4 VMs, 4 MCPs, 1 MCP per VM)	2x Quad-Core Xeon E5620 2.40GHz 16GB RAM	39	3000	4 VMs under EXSi 5.0 are configured and enabled with only one MCP installed in each VM. Guest OS on each VM is Windows 2008 Server R2 x64 SP1

Table 101: Multiple VMs Versus Multiple MCP Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
VoiceXML_App2 (4 VMs, 4 MCPs, 1 MCP per VM)	2x Quad-Core Xeon E5620 2.40GHz 16GB RAM	8.6	600	4 VMs under EXSi 5.0 are configured and enabled with only one MCP installed in each VM. Guest OS on each VM is Windows 2008 Server R2 x64 SP1. Tested with Nuance Speech Servers which run on another 4VMs of same hardware spec as MCP.
VoiceXML_App4 (4 VMs, 4 MCPs, 1 MCP per VM)	2x Quad Core Xeon E5620 2.40GHz 16GB RAM	21	1800	4 VMs under EXSi 5.0 are configured and enabled with only one MCP installed in each VM. Guest OS on each VM is Windows 2008 Server R2 x64 SP1
VXML_App1 (6 VMs, 6 MCPs, 1 MCP per VM)	2x Hex-Core Xeon X5675 3.06GHz 32GB RAM	52	4000	6 VMs under EXSi 5.0 are configured and enabled with only one MCP installed in each VM. Guest OS on each VM is RHEL 5.8 x64
VXML_App1 (6 VMs, 6 MCPs, 1 MCP per VM)	2x Hex-Core Xeon X5675 3.06GHz 32GB RAM	3.9	300	6 VMs under EXSi 5.0 are configured and enabled with only one MCP installed in each VM. Guest OS on each VM is RHEL 6.4 x64. GVP 8.1.7 or later.
VXML_App2 (6 VMs, 6 MCPs, 1 MCP per VM)	2x Hex-Core Xeon X5675 3.06GHz 32GB RAM	8.6	600	6 VMs under EXSi 5.0 are configured and enabled with only one MCP installed in each VM. Guest OS on each VM is RHEL 5.8 x64. Tested with Nuance Speech Servers which run on another 4VMs of a host of 2x Quad Core Xeon E5620.

Table 101: Multiple VMs Versus Multiple MCP Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
No VMWare				
VoiceXML_App1 1 Media Control Platform instance	2x Core 2 Quad Xeon x5355, 2.66 GHz, 12 GB RAM	17	1300	All Media Control Platform instances are configured on one server. Windows 2008 Server, SP2, x86.
VoiceXML_App1 2 Media Control Platform instances	2x Core 2 Quad Xeon x5355, 2.66 GHz, 12 GB RAM	27.5	2100	
VoiceXML_App1 4 Media Control Platform instances	2x Core 2 Quad Xeon x5355, 2.66 GHz, 12 GB RAM	30	2300	
VoiceXML_App1 8 Media Control Platform instances	2x Core 2 Quad Xeon x5355, 2.66 GHz, 12 GB RAM	27.5	2100	
VoiceXML_App1 2 Media Control Platform instances	2x Core 2 Quad Xeon x5355, 2.66 GHz, 4 GB RAM	39.4 (peak)	3000 (peak)	Media Control Platform instances are configured on one server. Squid is bypassed and call setup latency threshold is ignored Window 2003 Server only

Single Server All-In-One Capacity Testing Table

Table 102 describes the capacity testing for single server with multiple components installed (see comments column). Tests were performed by using a single instance of the Media Control Platform on Windows and Linux systems with 1 Core 2 Dual Xeon x5160, 3.0 GHz CPUs with 8 GB RAM.

This table shows the effect of having many GVP processes, including Nuance speech components, on just one physical server, which Genesys calls “the single server solution.”

Table 102: Single Server All-In-One Capacity Testing

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
Windows 2008, SP2, x86 and Windows 2008 R2				
VoiceXML_App1	1x Core 2 Dual Xeon x5160, 3.0 GHz, 8 GB RAM	7.9	600	A single server hosting Management Framework, Media Control Platform, Resource Manager, Reporting Server, Web Application Server (WAS), and SIP Server.
VoiceXML_App2 MRCP v1	1x Core 2 Dual Xeon x5160, 3.0 GHz, 8 GB RAM	1.2	100	
VoiceXML_App3 MRCP v1	1x Core 2 Dual Xeon x5160, 3.0 GHz, 8 GB RAM	2.5	160	
VoiceXML_App3 MRCP v2	1x Core 2 Dual Xeon x5160, 3.0 GHz, 8 GB RAM	1.9	120	
Red Hat Enterprise Linux 4				
VoiceXML_App1	1x Core 2 Dual Xeon x5160, 3.0 GHz, 8 GB RAM	4 (maximum CAPS tested)	300 (ports tested)	A single server hosting an Oracle DB Server, Management Framework, Reporting Server, Media Control Platform, Resource Manager, SIP Server, Web Application Server, and Linux.

Component Capacity Test Cases

This section describes capacity test case results for the GVP components and includes the following tables:

- Media Control Platform / Media Server Capacity (Physical Servers)—Table 103 on page 258
- Media Control Platform / Media Server Capacity (Virtual Servers)—Table 104 on page 263
- Resource Manager—Table 105 on page 273
- MRCP Proxy—Table 105 on page 273
- Reporting Server—Table 106 on page 279
- CTI Connector—Table 107 on page 281
- CTI Connector/ICM—Table 107 on page 281
- PSTN Connector—Table 108 on page 287
- Supplementary Services Gateway—Table 108 on page 287

For additional sizing information for Genesys Media Server with SIP Server, see See Chapter 18, "Genesys Administrator," page 463 of this guide.

The capacity testing results for the Media Control Platform are described in Table 103 on page 258. Tests were performed by using a single instance of the Media Control Platform on Windows and Linux systems with 2x Core 2 Quad, Xeon x5355, 2.66 GHz CPUs.

Media Control Platform Capacity Testing Table

Table 103 does not focus on GVP as a whole, but rather shows the impact of media services (announcements, call parking, bridging, conferencing, transcoding and video) on the performance of the Media Control Platform (MCP).

Note: Media Services Only: If your deployment is limited to Media Services, then see critical information for sizing the MCP in Table 103, "Media Control Platform Capacity Testing (Physical Servers)," on page 258 and the section "Genesys Media Server Sizing with SIP Server" in the SIP Server Sizing Guide.

Media Services plus VoiceXML Applications: If you have both types of services on the same GVP system, media and VoiceXML, then the actual performance will be a roughly proportional combination of media service performance and VoiceXML performance. This will be rather difficult to determine. We recommend that you default to the media performance metrics if transcoding is prevalent or media services are significant.

Table 103: Media Control Platform Capacity Testing (Physical Servers)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
Windows				
Audio bridge transfer G.711u <-> G.711u (baseline ~117 second duration)	2x Core2Quad Xeon x5355 2.66GHz	6.8	800	Bi-directional audio streams. Tested on Windows 2003.
Audio transcoding G.711u <-> AMR (~117 second duration)	2x Core2Quad Xeon x5355 2.66GHz	2.6	300	
Audio transcoding G.711u <-> AMR-WB (~117 second duration)	2x Core2Quad Xeon x5355 2.66GHz	2.0	230	
Audio transcoding G.711u <-> G.722 (~117 second duration)	2x Core2Quad Xeon x5355 2.66GHz	3.0	350	
Audio transcoding G.711u <-> G.726 (~117 second duration)	2x Core2Quad Xeon x5355 2.66GHz	2.6	300	
Audio transcoding G.711u <-> G.729 (~117 second duration)	2x Core2Quad Xeon x5355 2.66GHz	3.0	350	
SRTP with bridge transfer G.711u (~67 second duration)	2x Core2Quad Xeon x5355 2.66GHz	10	1200	Uni-directional audio streams. The capacity is the same for RTP and SRTP when SRTP is in default mode (both encryption and authentication are enabled), encrypted mode, or decrypted mode. Tested on Windows 2003.

Table 103: Media Control Platform Capacity Testing (Physical Servers) (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
MSML CPD + VoiceXML dialog (helloworld) codec N/A (~8 seconds duration, including 2.5 seconds of CPD)	2x Core2Quad Xeon x5355 2.66GHz	30	N/A	CPD is enabled within MSML which also invokes a VoiceXML dialog using the default helloworld page after the CPD result is returned. The VoiceXML dialog starts after the CPD returned successfully, the result of human.
Netann announcement - 3 seconds audio	2x Core2Quad Xeon x5355 2.66GHz	120 (preferred) 200 (peak)	500 (preferred) 1100 (peak)	Preferred - with call setup + call tear down latency < 1sec (500ms each) Peak - ignore call setup/tear down delay
Netann Play Treatment - G.711u (~60 seconds duration)	2x Core2Quad Xeon x5355 2.66GHz	30	1800	No transcoding.
Netann Play Treatment - video h263(+) (~120 seconds duration)	2x Core2Quad Xeon x5355 2.66GHz	10	1200	No transcoding.
Netann Play Treatment - video 3gp/avi (h263) (~120 seconds duration)	2x Core2Quad Xeon x5355 2.66GHz	8.3	1000	No transcoding.
Netann Recording Single Call - G.711u (raw, au & wav), G.722, G.726 (~120 seconds duration)	2x Core2Quad Xeon x5355 2.66GHz	8.3	1000	The capacity is the same for G.711u, G.722 & G.726.
Netann Recording Single Call - G.729, AMR (~120 seconds duration)	2x Core2Quad Xeon x5355 2.66GHz	5.8	700	The capacity is the same for G.729 & AMR.
Netann Recording Single Call - AMR-WB (~120 seconds duration)	2x Core2Quad Xeon x5355 2.66GHz	6.6	800	—

Table 103: Media Control Platform Capacity Testing (Physical Servers) (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
Netann Recording Single Call - video raw h263(+) (~120 seconds duration)	2x Core2Quad Xeon x5355 2.66GHz	4.2	500	—
Netann Recording Single Call - video avi (h263+G.711u) (~120 seconds duration)	2x Core2Quad Xeon x5355 2.66GHz	4	480	—
Netann Recording Single Call - video 3gp (h263+amr) (~120 seconds duration)	2x Core2Quad Xeon x5355 2.66GHz	2	240	—
Netann Recording Single Call - video raw h264 (~120 seconds duration)	2x Core2Quad Xeon x5355 2.66GHz	2	250	—
Netann 2 party Call Recording - G.711u (~60 seconds duration)	2x Core2Quad Xeon x5355 2.66GHz	11	660 call legs (330 recording sessions)	—
MSML Play announcement - one prompt (SIP INFO), one audio file - 3 seconds	2x Core2Quad Xeon x5355 2.66GHz	80	260	Call duration 3.13 seconds and gvp precheck is on.
MSML Play announcement - one prompt (SIP INFO), one audio file - 10 seconds	2x Core2Quad Xeon x5355 2.66GHz	200	2000	Call duration 10.34s and gvp precheck is off.
MSML Play announcement - one prompt (SIP INFO), two audio files - 4 + 6seconds	2x Core2Quad Xeon x5355 2.66GHz	200	2000	Call duration 10.34s and gvp precheck is off.

Table 103: Media Control Platform Capacity Testing (Physical Servers) (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
MSML Play announcement - two prompts (SIP INFO), two audio file - 4 + 6 seconds, one file per prompt.	2x Core2Quad Xeon x5355 2.66GHz	130	1400	Call duration 10.46s and gvp precheck is off.
MSML Play announcement - one prompt (SIP INFO), one audio file - 20 seconds	2x Core2Quad Xeon x5355 2.66GHz	150	3000	Call duration 20.34s and gvp precheck is off.
MSML Play announcement - one prompt (SIP INFO), three audio files - 4+6+10 seconds	2x Core2Quad Xeon x5355 2.66GHz	130	2600	Call duration 20.35s and gvp precheck is off.
MSML Play announcement - three prompts (SIP INFO), three audio filed - 4+6+10 seconds, one file per prompt	2x Core2Quad Xeon x5355 2.66GHz	100	2000	Call duration 20.60s and gvp precheck is off.
MSML Conference (all participants using the same codec - G711u) 3-party; ~60 seconds duration)	2x Core2Quad Xeon x5355 2.66GHz	6	360 participants (120 conference sessions)	The capacity is the same for G.711u, G.729, or GSM.
Linux				
Audio bridge transfer G.711u <-> G.711u (baseline ~117 second duration)	2x Core2Quad Xeon x5355 2.66GHz	9.4	1100	Bi-directional audio streams.

Table 103: Media Control Platform Capacity Testing (Physical Servers) (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
Transcoding with bridge transfer G.711u <-> or G.722 (~117 seconds duration)	2x Core2Quad Xeon x5355 2.66GHz	3	350	Bi-directional transcoding.
Transcoding with bridge transfer G.711u <-> or G.726 (~117 seconds duration)	2x Core2Quad Xeon x5355 2.66GHz	2	240	
Transcoding with bridge transfer G.711u <-> or G.729 (~117 seconds duration)	2x Core2Quad Xeon x5355 2.66GHz	2	240	
Transcoding with bridge transfer G.711u <-> or AMR-WB (~117 seconds duration)	2x Core2Quad Xeon x5355 2.66GHz	2	240	Bi-directional transcoding.
Transcoding with bridge transfer G.711u <-> or AMR (~120 seconds duration)	2x Core2Quad Xeon x5355 2.66GHz	1.7	200	
SRTP with bridge transfer G.711u (~67 second duration)	2x Core2Quad Xeon x5355 2.66GHz	12.8	1500	Uni-directional audio streams. The capacity is the same for RTP and SRTP when SRTP is in default mode (both encryption and authentication are enabled), encrypted mode, or decrypted mode. Tested on Red Hat Enterprise Linux 5.

Table 103: Media Control Platform Capacity Testing (Physical Servers) (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
SRTP with bridge transfer AMR audio + 3 gp of H.264 video (352 x 288) (~125 second duration)	2x Core2Quad Xeon x5355 2.66GHz	3.2	400	Uni-directional RTP streams. The capacity is the same for RTP and SRTP when SRTP is in default mode (both encryption and authentication are enabled), encrypted mode, or decrypted mode. Tested on Red Hat Enterprise Linux 5, x64.
MSML CPD + VoiceXML dialog (helloworld) codec: N/A (~8 seconds duration, including 2.5 seconds of CPD)	2x Core2Quad Xeon x5355 2.66GHz	40	N/A	CPD is enabled within MSML which also invokes a VoiceXML dialog using the default helloworld page after the CPD result is returned. The VoiceXML dialog starts after the CPD returned successfully, the result of human.
Note: *_preferred_ means the highest capacity that the system can sustain while maintaining optimal user experience. _Peak_ means the highest capacity that the system can sustain regardless of the user experience.				

Table 104: Media Control Platform / Media Server Capacity (Virtual Servers)

Application Type (Windows)	Hardware	Peak CAPS	Peak Ports	Comment
Video bridge transfer (H264 + AMR, 720P, 30fps, 1Mbps, level 3.1, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	6	400 calls	Uni-directional rtp (video + audio) stream. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video transcoding with bridge transfer (H264 + AMR, 30fps, 1Mbps, level 3.1, 720P -> CIF, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	0.11	8 calls	Uni-directional down scale transcoding. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.

Table 104: Media Control Platform / Media Server Capacity (Virtual Servers)

Application Type (Windows)	Hardware	Peak CAPS	Peak Ports	Comment
Video transcoding with bridge transfer (H264 + AMR, 30fps, 1Mbps, level 3.1, 720P -> QCIF, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	0.21	16 calls	Uni-directional down scale transcoding. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video bridge transfer (H264 + AMR, VGA, 30fps, 1Mbps, level 3.0, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	7.2	500 calls	Uni-directional rtp (video + audio) stream. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video transcoding with bridge transfer (H264 + AMR, 30fps, 1Mbps, level 3.0, VGA -> CIF, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	0.43	30 calls	Uni-directional down scale transcoding. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video transcoding with bridge transfer (H264 + AMR, 30fps, 1Mbps, level 3.0, VGA -> QCIF, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	0.72	50 calls	Uni-directional down scale transcoding. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video bridge transfer (H264 + AMR, CIF, 30fps, 256Kbps, level 2.0, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	11.3	800 calls	Uni-directional rtp (video + audio) stream. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video transcoding with bridge transfer (H264 + AMR, 30fps, 256Kbps, level 2.0, CIF -> QCIF, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	1.43	100 calls	Uni-directional down scale transcoding. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video bridge transfer (H264 + AMR, VGA, 60fps, 1Mbps, level 3.0, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	4.3	300 calls	Uni-directional rtp (video + audio) stream. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.

Table 104: Media Control Platform / Media Server Capacity (Virtual Servers)

Application Type (Windows)	Hardware	Peak CAPS	Peak Ports	Comment
Video bridge transfer (H264 + AMR, VGA, 60fps, 1Mbps, level 3.1, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	6.43	450 calls	Uni-directional rtp (video + audio) stream. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video transcoding with bridge transfer (H264 + AMR, VGA 1Mbps, level 3.1, 60fps -> 30fps, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	0.43	30 calls	Uni-directional down scale transcoding. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video transcoding with bridge transfer (H264 + AMR, VGA, 1Mbps, level 3.1, 60fps -> 15fps, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	0.13	9 calls	Uni-directional down scale transcoding. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video bridge transfer (H264 + AMR, VGA, 30fps, 1Mbps, level 3.0, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	6	420 calls	Uni-directional rtp (video + audio) stream. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video transcoding with bridge transfer (H264 + AMR, VGA 1Mbps, level 3.0, 30fps -> 15fps, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	0.34	24 calls	Uni-directional down scale transcoding. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video transcoding with bridge transfer (H264 + AMR, CIF, 30fps, 1.5Mbps -> 500Kbps, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	0.5	35 calls	Uni-directional down scale transcoding. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video transcoding with bridge transfer (H264 + AMR, CIF, 30fps, 1Mbps -> 192Kbps, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	1.12	80 calls	Uni-directional down scale transcoding. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.

Table 104: Media Control Platform / Media Server Capacity (Virtual Servers)

Application Type (Windows)	Hardware	Peak CAPS	Peak Ports	Comment
Video transcoding with bridge transfer (H264 + AMR, CIF, 30fps, 500Kbps -> 192Kbps, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	2	140 calls	Uni-directional down scale transcoding. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML CPA Answer Machine (~12.8 seconds duration)	2x Six-Core Xeon X5675 3.06GHz	150	n/a	MSML CPA only. Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML CPA Busy Machine (~7.7 seconds duration)	2x Six-Core Xeon X5675 3.06GHz	120	n/a	MSML CPA only. Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML CPA Fax Machine (6.3 seconds duration)	2x Six-Core Xeon X5675 3.06GHz	140	n/a	MSML CPA only. Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML CPA Human Machine (7.8 seconds duration)	2x Six-Core Xeon X5675 3.06GHz	170	n/a	MSML CPA only. Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML CPA SIT VC Machine (2.3 seconds duration)	2x Six-Core Xeon X5675 3.06GHz	130	n/a	MSML CPA only. Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec G711, 120 seconds duration)	2x Quad-Core Xeon E5620 2.40GHz	50	6000 calls	Tested on 4 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.

Table 104: Media Control Platform / Media Server Capacity (Virtual Servers)

Application Type (Windows)	Hardware	Peak CAPS	Peak Ports	Comment
MSML Play Announcement, (MP3, Any KHz, Any Kb, Cache enabled, Negotiated codec G711, 120 seconds duration)	2x Hex-Core Xeon X5675 3.06GHz	60	7200 calls	Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement, (MP3, 320Kbit, 44.1KHz, Cache disabled, Negotiated codec: G.711, 120 seconds duration)	2x Hex-Core Xeon X5675 3.06GHz	18	2160 calls	Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (MP3, 92Kbit, 32KHz, Cache disabled, Negotiated codec: G.711, 120 seconds duration)	2x Hex-Core Xeon X5675 3.06GHz	25	3000 calls	Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (MP3, 92Kbit, 32KHz Negotiated codec: G.711, 120 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	13	1560 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (MP3, 320Kbit, 44.1KHz, Negotiated codec: G.711, 120 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	12	1440 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H263 and AMR, CIF, 128Kbps 10fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	25	1500 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec AMR and H263 CIF, 512Kbps 30fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	8.5	500 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.

Table 104: Media Control Platform / Media Server Capacity (Virtual Servers)

Application Type (Windows)	Hardware	Peak CAPS	Peak Ports	Comment
MSML Play Announcement (Codec H263 and AMR, 4CIF, 512Kbps 10fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	23	1380 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H263 and AMR, 4CIF, 2Mbps 30fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	8	480 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, CIF, 128Kbps 10fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	25	1500 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, CIF, 256Kbps 15fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	17	1000 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, CIF, 512Kbps 30fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	8.5	500 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, 4CIF, 512Kbps 10fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	22	1300 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, 4CIF, 1Mbps 15fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	16	960 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.

Table 104: Media Control Platform / Media Server Capacity (Virtual Servers)

Application Type (Windows)	Hardware	Peak CAPS	Peak Ports	Comment
MSML Play Announcement (Codec H264 and AMR, 4CIF, 2Mbps 30fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	7.5	450 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, 720P, 1Mbps 10fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	19	1100 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, 4CIF, 2Mbps 15fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	9	540 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, 4CIF, 4Mbps 30fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	4	240 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement, (Codec H264 and AMR, 720P, 4Mbps 30fps, 60 seconds duration, no transcoding)	2x Hex-Core Xeon X5675 3.06GHz	2.5	150 calls	Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement, (Codec H264 and AMR, 3gp file 720P, 4Mbps 30fps, high profile level 3, transcoding, cache disabled, 60 seconds duration)	2x Hex-Core Xeon X5675 3.06GHz	0.2	12 calls	Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.

Table 104: Media Control Platform / Media Server Capacity (Virtual Servers)

Application Type (Windows)	Hardware	Peak CAPS	Peak Ports	Comment
MSML Play Announcement (Codec H264 and AMR, 3gp file 720P, 4Mbps 30fps, high profile level 3, transcoding to main profile level 2 CIF, cache enabled, 60 seconds duration)	2x Hex-Core Xeon X5675 3.06GHz	16.6	1000 calls	Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement, (Codec VP8 and G.711, avi file, VGA, 30fps 60 seconds duration, non-transcoding)	2x Hex-Core Xeon X5675 3.06GHz	20	2400 calls	Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement, (Codec VP8 and G.711, avi file, CIF, 30fps 60 seconds duration, non-transcoding)	2x Hex-Core Xeon X5675 3.06GHz	30	3600 calls	Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement, (Codec VP8 and G.711, avi file, QCIF, 20fps 60 seconds duration, non-transcoding)	2x Hex-Core Xeon X5675 3.06GHz	40	4800 calls	Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play and Digit Connect (Codec G711 and SIP INFO Digit, 34 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	50	1700 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Record (MP3, 96Kbit, 32KHz, 120 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	3	360 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Record (MP3, 320Kbit, 48KHz, 120 seconds duration)	1x Hex-Core Xeon X5670 2.93GHz	2	240 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.

Table 104: Media Control Platform / Media Server Capacity (Virtual Servers)

Application Type (Windows)	Hardware	Peak CAPS	Peak Ports	Comment
MSML Conference (32 participants per conference, all speakers. Each participant stays and speaks (300 secs in the conference. Codec G.711)	2x Quad-Core Xeon E5620 2.40GHz	2.6	768 participants (24 conference sessions)	Tested on 4 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Application Type (Linux)		**Maximum CAPS**	**Tested Ports**	**Comment**
MSML Play Announcement, (Codec G711, 120 seconds duration)	2x Hex-Core Xeon X5675 3.06GHz	60	7200 calls	Tested on 6 VMs of EXSi 5.0, Guest OS RHEL 5.8 x64. One MCP per VM.
MSML Play Announcement, (Codec G711, 120 seconds duration)	2x Hex-Core Xeon X5675 3.06GHz	24	2880 calls	Tested on 6 VMs of EXSi 5.0, Guest OS RHEL 6.4 x64. One MCP per VM. Play cache enabled as default. GVP 8.1.7 or later.
MSML Play Announcement, (Codec G711, 120 seconds duration)	2x Hex-Core Xeon X5675 3.06GHz	42	5040 calls	Tested on 6 VMs of EXSi 5.0, Guest OS RHEL 6.4 x64. One MCP per VM. Play cache disabled. GVP 8.1.7 or later.
MSML Play Announcement, (MP3, Any KHz, Any Kb, Cache enabled, Negotiated codec G711, 120 seconds duration)	2x Hex-Core Xeon X5675 3.06GHz	60	7200 calls	Tested on 6 VMs of EXSi 5.0, Guest OS RHEL 5.8 x64. One MCP per VM.

Table 104: Media Control Platform / Media Server Capacity (Virtual Servers)

Application Type (Windows)	Hardware	Peak CAPS	Peak Ports	Comment
MSML Play Announcement, (MP3, 320Kbit, 44.1KHz, Cache disabled, Negotiated codec: G.711, 120 seconds duration)	2x Hex-Core Xeon X5675 3.06GHz	16	1920 calls	Tested on 6 VMs of EXSi 5.0, Guest OS RHEL 5.8 x64. One MCP per VM.
MSML Play Announcement (MP3, 92Kbit, 32KHz, Cache disabled, Negotiated codec: G.711, 120 seconds duration)	2x Hex-Core Xeon X5675 3.06GHz	23	2760 calls	Tested on 6 VMs of EXSi 5.0, Guest OS RHEL 5.8 x64. One MCP per VM.

Table 105 describes the capacity testing for overall system performance when the Resource Manager and MRCP Proxy (Windows only) are tested with multiple Media Control Platform instances.

Table 105: Resource Manager and MRCP Proxy Capacity Testing

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
Resource Manager (Windows)				
SIP Call (Resource Manager performance)	2x Core 2 Quad Xeon x5355, 2.66 GHz	800	Any number	Using both TCP and UDP. Results occur regardless of the port density or the type of calls routed. Multiple Media Control Platform instances are required to achieve the peak CAPS. **With the Reporting Server configured in one of two ways:** • Enabled and in No-DB mode—Without the DB (all data is dropped), the Reporting Server can handle much higher capacities. If both Reporting Server and DB are enabled, a peak CAPS bottleneck would occur. See SIP Call (Reporting Server in partitioning mode with Microsoft SQL 2008 Enterprise Server) in Table 106 on page 279. • Disabled

Table 105: Resource Manager and MRCP Proxy Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
SIP Call (Resource Manager performance)	2x Core 2 Quad Xeon x5355, 2.66 GHz	200	Any number	Tested on UDP only on RHEL 6.4 x64. GVP 8.1.7 or later. Results occur regardless of the port density or the type of calls routed. Multiple Media Control Platform instances are required to achieve the peak CAPS. **With the Reporting Server configured in one of two ways:** • Enabled and in No-DB mode—Without the DB (all data is dropped), the Reporting Server can handle much higher capacities. If both Reporting Server and DB are enabled, a peak CAPS bottleneck would occur. See SIP Call (Reporting Server in partitioning mode with Microsoft SQL 2008 Enterprise Server) in Table 106 on page 279. • Disabled

Table 105: Resource Manager and MRCP Proxy Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
SIP Call (Resource Manager performance with 1000 tenants configured.)	2x Core 2 Quad Xeon x5335, 2.66 GHz	600	Any number	Results occur regardless of the port density and the type of calls being routed. To achieve the peak CAPS, multiple Media Control Platforms might be required. **With the Reporting Server configured in one of two ways:** • Enabled and in No-DB mode—Without the DB (all data is dropped), the Reporting Server can handle much higher capacities. If both Reporting Server and DB are enabled, a peak CAPS bottleneck would occur. See SIP Call (Reporting Server in partitioning mode with Microsoft SQL 2008 Enterprise Server) in Table 106 on page 279. • Disabled
SIP Call (Resource Manager performance with MSML embedded in SIP INFO messages.)	2x Core 2 Quad Xeon x5335, 2.66 GHz	300	Any number	To achieve the peak CAPS, multiple Media Control Platforms might be required. **With the Reporting Server configured in one of two ways:** • Enabled and in No-DB mode—Without the DB (all data is dropped), the Reporting Server can handle much higher capacities. If both Reporting Server and DB are enabled, a peak CAPS bottleneck would occur. See SIP Call (Reporting Server in partitioning mode with Microsoft SQL 2008 Enterprise Server) in Table 106 on page 279. • Disabled

Table 105: Resource Manager and MRCP Proxy Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
Resource Manager (Linux)				
SIP Call (Resource Manager performance)	2x Core 2 Quad Xeon x5355, 2.66 GHz	800	Any number	Using both TCP and UDP. Results occur regardless of the port density or the type of calls routed. Multiple Media Control Platform instances are required to achieve the peak CAPS. **With the Reporting Server configured in one of two ways:** • Enabled and in No-DB mode—Without the DB (all data is dropped), the Reporting Server can handle much higher capacities. If both Reporting Server and DB are enabled, a peak CAPS bottleneck would occur. See SIP Call (Reporting Server in partitioning mode with Microsoft SQL 2008 Enterprise Server) in Table 106 on page 279. • Disabled

Table 105: Resource Manager and MRCP Proxy Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
SIP Call (Resource Manager performance)	2x Core 2 Quad Xeon x5355, 2.66 GHz	600	Any number	In this scenario, 100 K of DID numbers are configured and mapped to 262 IVR applications, and defined without wild cards or ranges—In other words, ordinary one-to-one mappings. Results occurs regardless of the port density or the type of calls routed. Multiple Media Control Platforms required to achieve the peak CAPS. **With the Reporting Server configured in one of two ways:** • Enabled and in No-DB mode—Without the DB (all data is dropped), the Reporting Server can handle much higher capacities. If both Reporting Server and DB are enabled, a peak CAPS bottleneck would occur. See SIP Call (Reporting Server in partitioning mode with Microsoft SQL 2008 Enterprise Server) in Table 106 on page 279. • Disabled
SIP Call (Resource Manager performance)	2x Core 2 Quad Xeon x5355, 2.66 GHz	800	Any number	In this scenario, 1 million DID numbers are configured and mapped to 262 IVR applications, and defined in a multi-tenant environment (32 tenants with 30~35 K of DIDs per tenant), without wildcards or ranges—In other words, simple one-to-one mappings. Results occurs regardless of the port density or the type of calls routed. Multiple Media Control Platforms required to achieve the peak CAPS. **Reporting Server disabled** (due to the fact that the Reporting Server is unable to support 1 million DIDs).

Table 105: Resource Manager and MRCP Proxy Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
SIP Call (Resource Manager performance with MSML embedded in SIP INFO messages.)	2x Core 2 Quad Xeon x5355, 2.66 GHz	350	Any number	Multiple Media Control Platforms required to achieve the peak CAPS. **With the Reporting Server configured in one of two ways:** • Enabled and in No-DB mode—Without the DB (all data is dropped), the Reporting Server can handle much higher capacities. If both Reporting Server and DB are enabled, a peak CAPS bottleneck would occur. See SIP Call (Reporting Server in partitioning mode with Microsoft SQL 2008 Enterprise Server) in Table 106 on page 279. • Disabled
MRCP Proxy (Windows)				
MRCPv1 requests (MRCP Proxy performance)	2x Core 2 Quad Xeon x5355, 2.66 GHz	1600	N/A	Tested with simulated MRCP servers and clients, calculation is based on MRCP sessions. Tested on Windows 2008 R2.

Reporting Server Capacity Testing Table

Table 106 describes the capacity testing for overall system performance when the Reporting Server is tested with multiple Media Control Platform instances.

Tables 106, 107 and 108 show the performance of other GVP components individually. Use these tables to see determine if you encountered any performance limits beyond those already defined in Table 100 through Table 105.

Use these tables if you are interested in determining the overall system limits, which may occur in VoiceXML, media services, reporting, RM, or other functions.

Table 106: Reporting Server Capacity Testing

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
SIP Call (Reporting Server in partitioning mode with Microsoft SQL 2008 Enterprise Server)	Reporting Server: 2x Core 2 Quad Xeon x5355, 2.66 GHz Microsoft SQL Server DB: 2x Core 2 Quad Xeon x5355, 2.66 GHz	270	Any number 32,400 (~30,000 based on a 120 seconds call duration)	Results occur regardless of the port density or the type of calls processed. Resource Manager and Media Control Platform log information to the Reporting Server using default settings. Increased reporting and logging can reduce Reporting Server capacity. Microsoft SQL database installed on Windows 2008 Server with the database files residing on a 15k rpm HDD Disk Array.
SIP Call (Reporting Server in partitioning mode with Oracle 10g R2 Server)	Reporting Server: 2x Core 2 Quad Xeon x5355, 2.66 GHz Oracle DB: 2x Core 2 Quad Xeon x5355, 2.66 GHz	270	Any number	Results occur regardless of the port density or the type of calls processed. Resource Manager and Media Control Platform log information to the Reporting Server using default settings. Increased reporting and logging can reduce Reporting Server capacity. Oracle database installed on Windows 2003 Server with the database files residing on a 15k rpm HDD Disk Array.

Table 106: Reporting Server Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
SIP Call - Reporting Server /w MS SQL Server 2008 R2 Enterprise (partitioning mode)	Reporting Server - 2x Core2Quad Xeon x5355 2.66GHz MSSQL DB - 2x Core2Quad Xeon x5355 2.66GHz /w 15 HDD Disk Array	300	Any	Regardless of the port density and the type of calls being processed With official architecture which RM and MCP are both logging information to RS and using default setting. Heavier reporting/logging can reduce the RS capacity. MSSQL on Windows 2008 R2 with DB data files reside on a 15 HDD Disk Array (15k rpm)
SIP Call (Reporting Server in No DB mode)	Reporting Server 2x Core 2 Quad Xeon E5504, 2.0 GHz, 8 GB RAM	800	Any number	When Reporting Server is configured in No DB mode, data that is sent to it, is dropped. Tested with an actual Resource Manager instance (not a VM) without the Media Control Platform.
SIP Call (Reporting Server in partitioning mode with Oracle 11g Server)	Reporting Server: 2x Core 2 Quad Xeon x5355, 2.66 GHz Oracle DB: 2x Core 2 Quad Xeon x5355, 2.66 GHz /w 15 HDD Disk Array	300	Any number	Results occur regardless of the port density and the type of calls being processed. RM and MCP both log information to the Reporting Server using default settings. Increased reporting/logging can reduce RS capacity. Oracle DB on Windows 2008 R2 x64 with DB data files reside on a 15 HDD Disk Array (15k rpm).

CTI Connector and CTI Connector with ICM Capacity Testing Table

Table 107 describes the capacity testing for overall system performance when the CTI Connector is tested with multiple Media Control Platform instances. Results are provided for CTI applications and treatments using both GVPi and the NGI. In addition, CPUs of varying types and speeds were used for testing on Windows, and are specified for each application.

Table 107: CTI Connector with IVRSC and CTI Connector with ICM Capacity Testing

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
Genesys CTI (Windows)				
CTI treatments and bridge transfer application.	2x Quad Core Xeon E5335, 2.0 GHz	25 (MCPs w/ GVPi)	800	A call that starts with a prompt-play and route request with 3 treatments and then bridge transfers to an agent. Tested with 5 Media Control Platform instances (GVP 8.1.4 only).
CTI treatments and one-step transfer application with GVPi	2x Quad Core Xeon E5335, 2.0 GHz	25 (MCPs w/ GVPi)	800	A call that starts with a prompt-play and route request with 3 treatments, and then transfers to an agent. Tested with 5 Media Control Platform instances. (GVP 8.1.4 or earlier)
CTI treatments and Bridge transfer application with NGI	2x Quad Core Xeon E5335 2.0GHz, 4 GB RAM, 270 GB SAS hdd	15 (MCPs w/ NGI)	480	A call that starts with a prompt play and route request with 3 treatments, and then a bridge transfer to an agent. Tested with 5 Media Control Platform instances (GVP 8.1.3 or later).
CTI Treatments and One-Step Transfer application with NGI	2x Quad Core Xeon E5335 2.0GHz, 4 GB RAM, 270 GB SAS hdd	25 (MCPs w/ NGI)	800	A call that starts with a prompt play and route request with 3 treatments, and then transfers (in one step via CTIC) out to an agent. Tested with 5 Media Control Platform instances (GVP 8.1.3 or later).

Table 107: CTI Connector with IVRSC and CTI Connector with ICM Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
Cisco CTI (Windows)				
CTI - ICM treatments, followed by a bridge transfer. (Call variable event is set to ICM.)	2x Quad Core Xeon E5335 2.0GHz, 2.53 GHz	22 (MCPs w/ NGI)	440	Transfer with CED, Call and ECC variable events passing from two MCP instances to a single ICM. Tested on Windows 2008 R2 (GVP 8.1.4 and 8.1.5 releases), with CTIC installed in CRI mode.
CTI - ICM (CRI mode) Treatments, followed by a bridge transfer. Set Call Variable event to ICM (overall system performance, /w multiple MCPs, NGi)	2x Quad Core Xeon E5630 2.53GHz	25 CAPS overall system (MCP/NGi)	500 (overall system)	Bridge Transfer with CED, Call and ECC Variable passing from MCP to ICM. Only one ICM is configured Tested on Windows 2008 R2. GVP 8.1.6+
CTI - ICM scripts treatments, followed by a cancellation and blind transfer. (Call Variable event is set to ICM.)	2x Quad Core Xeon E5335 2.0GHz	30 (MCPs w/ NGI)	600	Different tenants associated with two VRU-PGs; A blind transfer with CED, Call and ECC variable passing from two MCP instances to two ICMs. Tested on Windows 2008 R2 (GVP 8.1.4 or later releases), with CTIC installed in SCI mode.
Genesys CTI (Linux)				
CTI treatments and bridge transfer.	2x Quad Core Xeon E5630, 2.53 GHz	25 (MCPs w/ GVPi)	800	A call that starts with a prompt play and route request with 3 treatments, and then transfers to an agent. Tested with 5 Media Control Platform instances. Supported with Linux x86 platforms only (GVP 8.1.4 only).

Table 107: CTI Connector with IVRSC and CTI Connector with ICM Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
CTI treatments and one-step transfer	2x Quad Core Xeon E5630 2.5 GHz	25 (MCPs w/ GVPi)	800	A call that starts with a prompt play and route request with 3 treatments, and then transfers to an agent. Tested with 5 Media Control Platform instances. Supported with Linux x86 platforms only (GVP 8.1.4 only).
CTI Treatments and Bridge Transfer application (overall system performance, /w multiple MCPs, NGi)	2x Quad Core Xeon E5410 2.33GHz	20 CAPS overall system (MCP/NGi)	640 (overall system)	A call that starts with a prompt play and route request with 3 treatments, and then a bridge transfer to an agent. Tested with 5 MCPs on EL5.x - x64, GVP 8.1.5 or later.
CTI Treatments and Bridge Transfer application (overall system performance, /w multiple MCPs, NGi)	2x Quad Core Xeon E5335 2.33GHz	20 CAPS overall system (MCP/NGi)	640 (overall system)	Call starts with a prompt play and route request with 3 treatments and then bridge transfer to an agent. Tested with 5 MCPs on EL 6.4 - x64, GVP 8.1.7 or later.
CTI treatments and bridge transfer.	2x Quad Core Xeon E5630 2.53GHz	20 (MCPs w/ NGI)	640	A call that starts with a prompt play and route request with 3 treatments, and then a bridge transfer to an agent. Tested with 4 MCP instances. Supported with Linux x86 platforms only (GVP 8.1.4 or later releases).

Table 107: CTI Connector with IVRSC and CTI Connector with ICM Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
CTI treatments and one-step transfer	2x Quad Core Xeon E5410 2.33GHz	25 (MCPs w/ NGI)	800	A call that starts with a prompt play and route request with 3 treatments, and then transfers 2x Quad Core Xeon E5410 2.33GHz to an agent. Tested with 5 MCPs on EL 6.4 - x64. Supported with Linux x86 platforms only (GVP 8.1.5 or later releases).
CTI Treatments and One Step Transfer application using INFO + INFO model (overall system performance, /w multiple MCPs, NGi)	2x Quad Core Xeon E5335 2.33GHz	25 CAPS overall system (MCP/NGi)	800 (overall system)	A call that starts with a prompt play and route request with 3 treatments (using INFO + INFO model) and then transfer out to an agent. Tested with 5 MCPs on RHEL 6.4 x64, GVP 8.1.7 or later.
CTI Treatments and One Step Transfer application (overall system performance, /w multiple MCPs, NGi)	2x Quad Core Xeon E5335 2.33GHz	25 CAPS overall system (MCP/NGi)	800 (overall system)	A call that starts with a prompt play and route request with 3 treatments and then transfer out to an agent. Tested with 5 MCPs on EL 6.4 x64, GVP 8.1.7 or later.
CTI treatments and one-step transfer	2x Quad Core Xeon E5630 2.33GHz	25 CAPS (MCPs w/ NGI)	800	A call that starts with a prompt play and route request with 3 treatments, and then transfers 2x Quad Core Xeon E5410 2.33GHz to an agent. Tested with 5 MCP instances. Supported with Linux x86 platforms only (GVP 8.1.4 or later releases).

Table 107: CTI Connector with IVRSC and CTI Connector with ICM Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
Cisco CTI (Linux)				
CTI - ICM treatments, followed by a bridge transfer in CRI mode. (Call variable event is set to ICM.)	2x Quad Core Xeon E5335 2.00GHz	18 (MCPs w/ NGI)	270	Transfer with CED, Call and ECC variable events passing from two MCP instances to a single ICM. (GVP 8.1.4 only)
CTI - ICM treatments, followed by a bridge transfer in CRI mode. (Call variable event is set to ICM.)	2x Quad Core Xeon E5335 2.00GHz	22 (MCPs w/ NGI)	440	Transfer with CED, Call and ECC variable events passing from the Media Control Platform two MCP instances to a single ICM. (GVP 8.1.5 or later)
CTI - ICM treatments, followed by a bridge transfer in CRI mode, set call variable event to ICM (overall system performance, /w multiple MCPs, NGi)	2x Quad Core Xeon E5335 2.0GHz	25 CAPS overall system (MCP/NGi)	500 (overall system)	Bridge Transfer with CED, Call and ECC Variable passing from MCP to ICM. Only one ICM is configured. Tested on GVP 8.1.6 EL 5.x x64
CTI - ICM Treatments, followed by a bridge transfer in CRI mode, set call variable event to ICM (overall system performance, /w multiple MCPs, NGi)	2x Quad Core Xeon E5335 2.0GHz	30 CAPS overall system (MCP/NGi)	600 (overall system)	Bridge Transfer with CED, Call and ECC Variable passing from MCP to ICM. Only one ICM is configured. Tested on GVP 8.1.7 EL 6.4 x64

Table 107: CTI Connector with IVRSC and CTI Connector with ICM Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
CTI - ICM scripts treatments, followed by a cancellation and blind transfer. CTIC is configured in SCI mode. (Call variable event is set to ICM.)	2x Quad Core Xeon E5630 2.53GHz, 2.53 GHz	30 CAPS (MCPs w/ NGI)	600 (overall system)	Different tenants associated with two VRU-PGs; A blind transfer with CED, Call and ECC variable passing from two MCP instances to two ICMs. (GVP 8.1.4 or later releases, with CTIC installed in SCI mode.)
CTI - ICM Scripts Treatments, followed by cancellation and blind transfer. Set Call Variable event to ICM (overall system performance, /w multiple MCPs, NGi)	2x Quad Core Xeon E5335 2.0GHz	30 CAPS overall system (MCP/NGi)	600 (overall system)	Different Tenant tied to two VRU-PGs, Blind Transfer with CED, Call and ECC Variable passing from MCP to ICM. Two ICMs are configured. Tested with 3 MCPs on RHEL 6.4 x64, GVP 8.1.7 or later.

PSTN Connector and SSG Capacity Testing Table

Table 108 describes the capacity testing for overall system performance when the PSTN Connector or Supplementary Services Gateway components are tested with multiple Media Control Platform instances. In addition, CPUs of varying types and speeds were used for testing on Windows, and are specified for each application.

Table 108: PSTN Connector and SSG Capacity Testing

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
Supplementary Services Gateway (Windows)				
Supplementary Services Gateway outbound call application	2x Core2Quad Xeon E5335 2.53GHz, 2.53 GHz	65	N/A	The Supplementary Services Gateway makes outbound calls through SIP Server, which becomes the overall system bottleneck. Multiple Media Control Platform instances are required to achieve peak capacity. GVP 8.1.5 with SIP Server 8.1.0 or later.
Supplementary Services Gateway outbound call application	2x Core2Quad Xeon E5335 2.53GHz, 2.53 GHz	50	N/A	The Supplementary Services Gateway makes outbound calls through SIP Server which become the overall system bottleneck. Multiple Media Control Platform instances are required to achieve peak capacity. GVP 8.1.3 or 8.1.4 with SIP Server 8.0.4 or later.
Supplementary Services Gateway outbound call application	2x Core 2 Quad Xeon x5355, 2.66 GHz	40	N/A	The Supplementary Services Gateway makes outbound calls through SIP Server which become the overall system bottleneck. Multiple Media Control Platform instances are required to achieve peak capacity. Pre-GVP 8.1.3 with SIP Server 8.0.3.

Table 108: PSTN Connector and SSG Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
Supplementary Services Gateway (Linux)				
Supplementary Services Gateway outbound call application	2x Quad Core Xeon E5335 @ 2.00GHZ, 4 GB RAM, 67 GB SAS hdd	66	N/A	The Supplementary Services Gateway makes outbound calls through SIP Server, which becomes the overall system bottleneck. Multiple MCP instances are required to achieve peak capacity. GVP 8.1.5 on RHEL 5.x with SIP Server 8.1.000.54.
SSG outbound call application (overall system performance, /w multiple MCPs)	2x Core2Quad Xeon E5335 2.00GHz	64 CAPS (overall system)	N/A	SSG makes outbound calls via SIP Server which becomes the bottleneck overall system. Multiple MCPs are required to achieve peak capacity. GVP 8.1.7 on RHEL 6.4 x64 with SIP Server 8.1.1.
Supplementary Services Gateway outbound call application	2x Quad Core Xeon E5335 @ 2.53GHZ, 4 GB RAM, 67 GB SAS hdd	50	N/A	The Supplementary Services Gateway makes outbound calls through SIP Server which become the overall system bottleneck. Multiple Media Control Platform instances are required to achieve peak capacity. GVP 8.1.3 and 8.1.4 releases with SIP Server 8.0.4.
Supplementary Services Gateway outbound call application	2x Core 2 Quad Xeon x5355, 2.66 GHz	40	N/A	The Supplementary Services Gateway makes outbound calls through SIP Server which become the overall system bottleneck. Multiple Media Control Platform instances are required to achieve peak capacity. Pre-GVP 8.1.3 with SIP Server 8.0.3.

Table 108: PSTN Connector and SSG Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
PSTN Connector (Windows)				
PSTN Connector VoiceXML_App1	2x Xeon 3.0 GHz	N/A	8 T1/E1 spans	Tested with two DMV boards.
PSTN Connector VoiceXML_App3	2x Xeon 3.0 GHz	N/A	8 T1/E1 spans	Tested with two DMV boards.
PSTN Connector (Linux)				
PSTN Connector VoiceXML_App1	2x Xeon 3.0 GHz	N/A	8 T1/E1 spans (ISDN only)	Tested with two DMV boards. (RHEL 5.8 x86 only)
PSTN Connector VoiceXML_App3	2x Xeon 3.0 GHz	N/A	8 T1/E1 spans (ISDN only)	Tested with two DMV boards. (RHEL 5.8 x86 only)

In the following test cases, maximum capacity was achieved within the constraints of specific thresholds. However, the system was also tested beyond the recommended capacity to determine the extent of performance degradation.

Call Setup Latency Test Case

The test case in Figure 85 uses the VoiceXML_App1 profile (in Table 98 on page 246) to show how the CSL increases as the PD increases. The rate at which the CSL increases is relatively constant until the system reaches a bottleneck—for example, when the system load is beyond peak capacity.

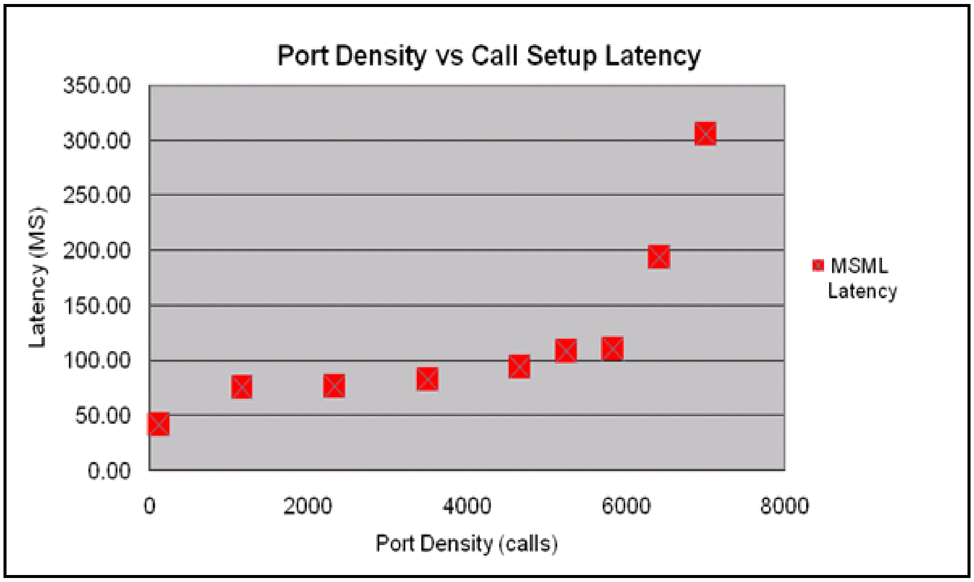


Figure 85: PD Versus CSL

Caller Perceived Latency Test Case

The graph in Figure 86 shows the DTMF response-to-audio-prompt latency at various port densities (relative to the peak capacity indicated in Table 100 on page 250). Notice that the TTS prompts produce ~300 ms more latency than the audio file prompts. This is due to the beginning silence played by the TTS engine.

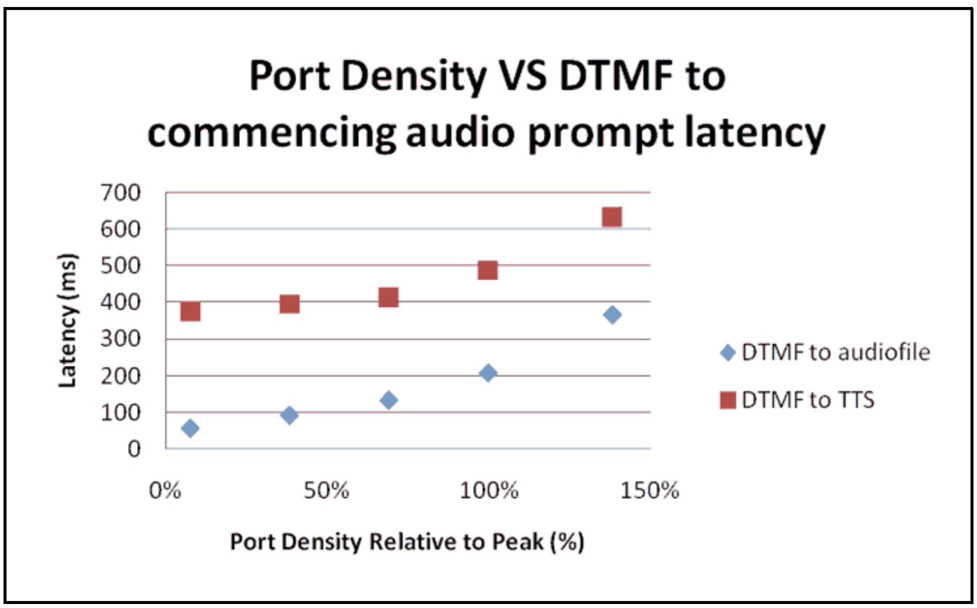


Figure 86: PD Versus DTMF

When there is speech input, additional latency is usually caused by the ASR engine. In Figure 87, the latency result is from 1000 words of grammar using the Nuance OSR3 MRCP version 1 (MRCPv1) engine. The result can vary, depending on the type of MRCP engine used, the type of speech grammar used, and the load on the speech engine.

The performance results in Figure 87 were obtained from isolated ASR engines supporting the same number of recognition sessions at all Media Control Platform port densities; the MRCP engines did not cause a bottleneck. Therefore, depending on the load on the Media Control Platform, it can add as much as ~100 ms of latency.

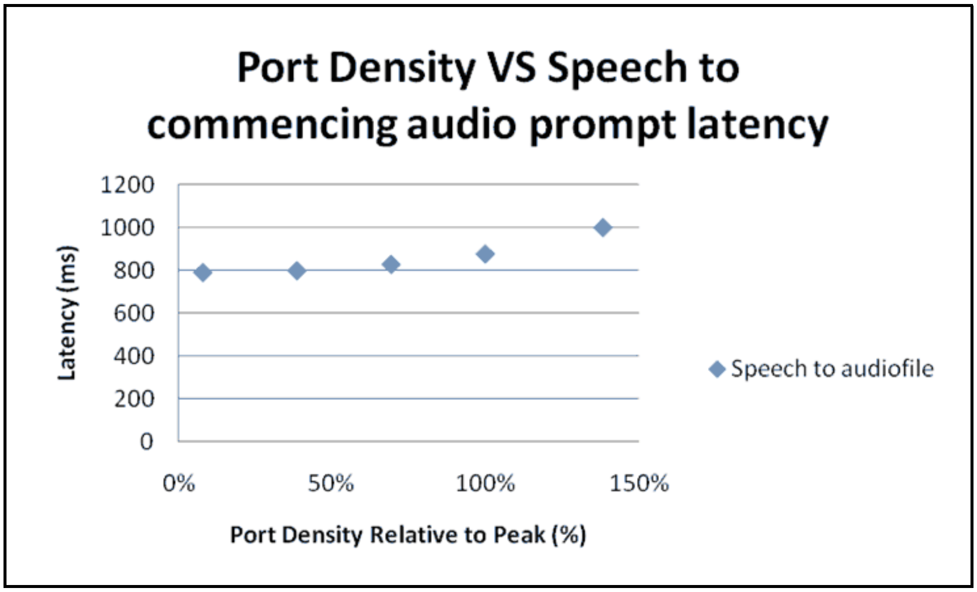


Figure 87: PD Versus Speech

Cachable VoiceXML Content Test Cases

GVP can cache internal, compiled VoiceXML objects. Caching VoiceXML objects saves a significant amount of compilation time, which results in less CPU consumption. The VoiceXML_App1 application is used for the test case in Figure 88 and is based on the peak capacity indicated in Table 100 on page 250.

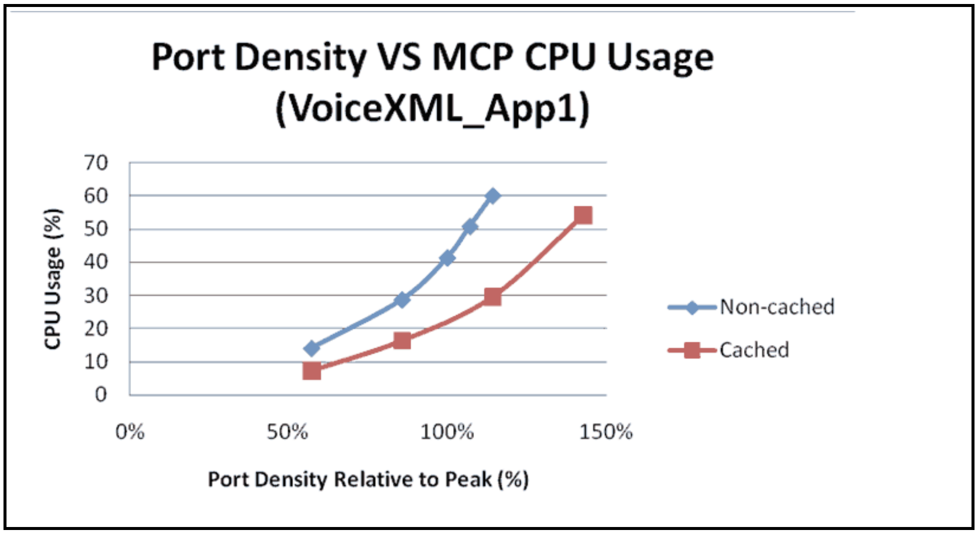


Figure 88: PD Versus CPU (VoiceXML_App1)

The more complex the VoiceXML content, the greater the benefit of having cachable content. The test case in Figure 89 is similar to the one in Figure 88 except the more complex VoiceXML_App2 application is used (see Table 98 on page 246).

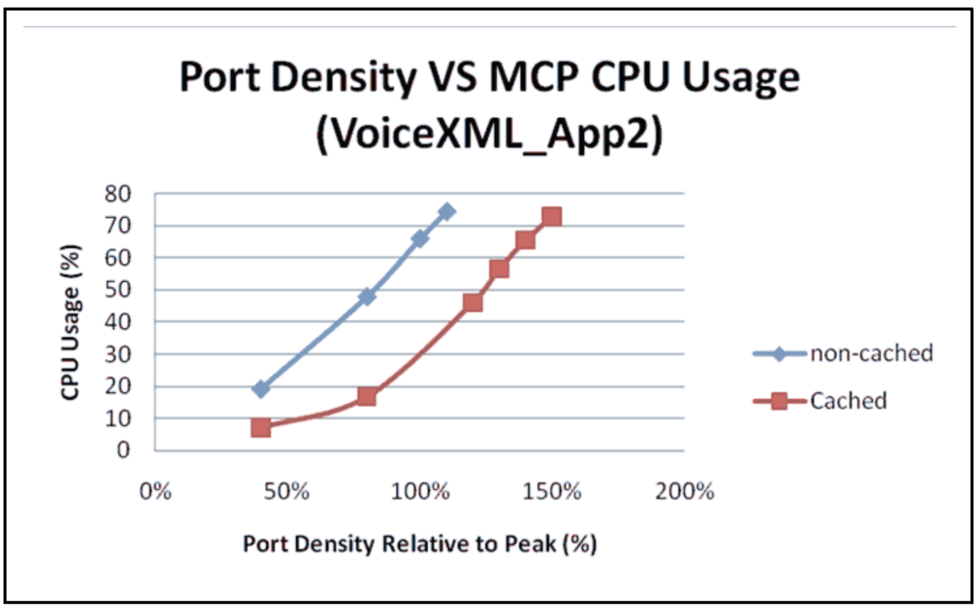


Figure 89: PD Versus CPU (VoiceXML_App2)

In Figures 88 and 89 the processing of cachable and non-cachable content are compared with the Media Control Platform using the same level of CPU consumption for both applications. The following results clearly show the benefits of using cachable content:

- CPU consumption—Media Control Platform at peak capacity.
  - 15% less consumption than non-cached content using VoiceXML_App1.

- ~30% less consumption than non-cached content using VoiceXML_App2.
- Port density—CPU consumption at same level for both applications:
  - ~30-35% greater than non-cached content using VoiceXML_App1,
  - ~50% greater than non-cached content using VoiceXML_App2.

Performance Planning and Scalability

Use the information in the following test cases and performance comparisons to determine the performance requirements for your GVP 8.1 environment.

Performance Test Cases

The Genesys QA performance testing samples in this section include:

- "Application Test Cases"
- "Component Test Cases" on page 319
- "Single Server Test Cases" on page 334
- "Multiple MCP Instances and Virtual Machines Test Cases" on page 319

Application Test Cases

The following application test cases are described in this section:

- "NETANN Announcement" on page 293
- "MSML Announcement" on page 295
- "Conference Performance" on page 314
- "Transcoding" on page 303
- "Secure RTP" on page 307

NETANN Announcement

When the Media Control Platform acts as a media announcement server, high call rates can be sustained. Using a NETANN announcement application, it can sustain up to 200 CAPS (~1100 ports) for a typical audio playback of 3.3 seconds, however, call setup and tear down latency increases.

The graph in Figure 90 on page 294 shows call durations at various CAPS. When CAPS reaches its peak (200 CAPS), the setup and tear down latency can reach 3.3 seconds. Optimally, call setup and tear down latency should be maintained at <1sec (or 500 ms each) with CAPS at 150 (with 600 ports).

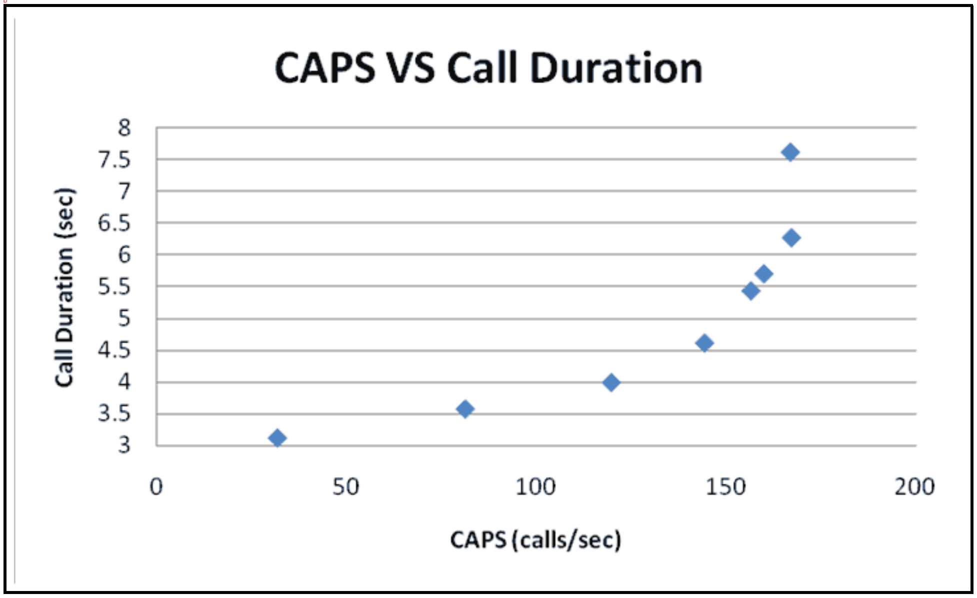


Figure 90: CAPS Versus CD (Announcement)

In Figure 91, as the call duration increases with higher port density, the additional call setup and tear down latency prevents the CAPS from scaling linearly in relation to the port density.

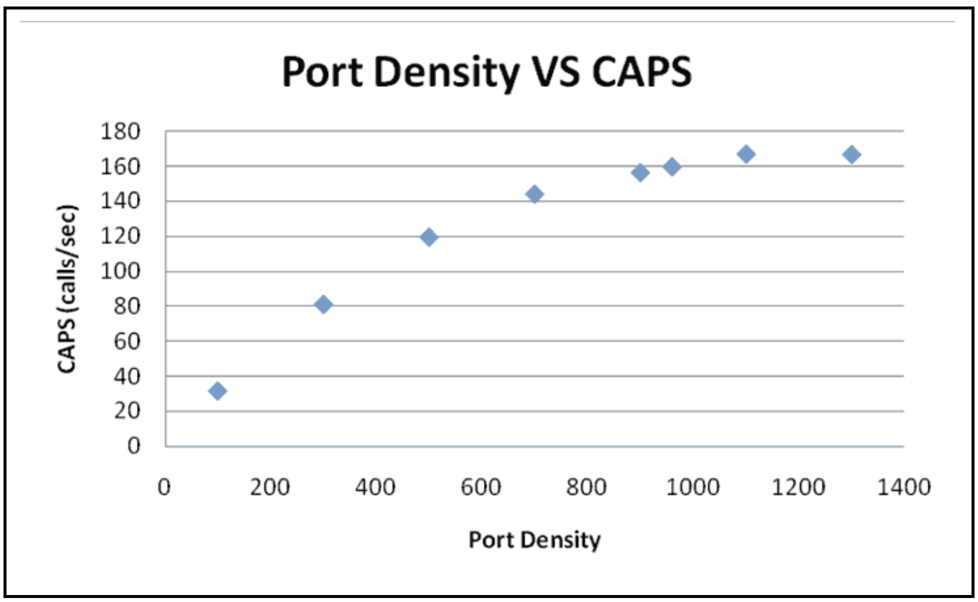


Figure 91: PD Versus CAPS (Announcement)

In Figure 92 on page 295, a bottleneck is caused by the media streaming. Shorter audio announcements increase the time spent on call setup and tear down and, although the load on the system decreases, shorter audio prompts cause the peak CAPS to increase.

The graph in Figure 92 on page 295 depicts a use case where a 1 second audio announcement drives the peak CAPS to ~235. Optimally, in this use case, call setup and tear down latency should be maintained at <1sec and CAPS at 200 (with ~500 ports).

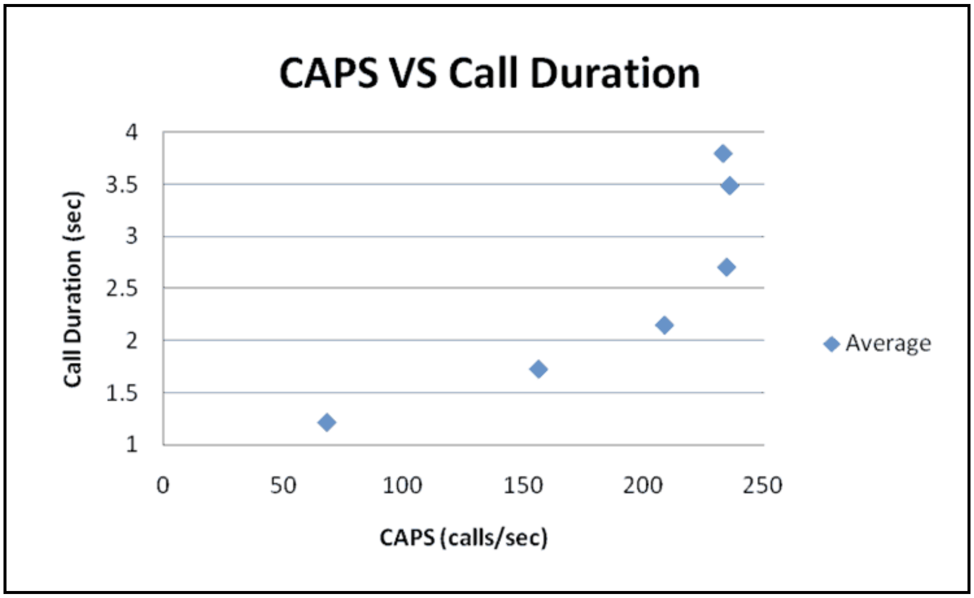


Figure 92: CAPS Versus CD (Average)

MSML Announcement

MSML announcement applications of 3, 10, and 20 second durations were tested on RHE Linux 5, Update 4, x64. Announcement applications were tested to compare, 1 prompt/1 request versus. 2 prompts/1 request versus. 2 prompts/2 requests versus 3 prompts/3 requests in the following scenarios:

MSML Announcement - 3 second duration

- 1 audio file
- 1 prompt (SIP INFO)

MSML Announcement - 10 second duration

- 1 audio file (10s)
- 2 audio files (4s and 6s)
- 2 audio files (per prompt)
- 1 prompt (SIP INFO)
- 1 prompt (SIP INFO)
- 2 prompts (SIP INFO)

MSML Announcement - 20 second duration

- 1 audio file (20s)
- 3 audio files (4s, 6s, 10s)
- 3 audio files (per prompt)
- with 1 prompt (SIP INFO)
- with 1 prompt (SIP INFO)
- with 3 prompts (SIP INFO)

A 3 second audio file with a single prompt (SIP INFO) and gvp:precheck turned on, resulted in a peak capacity of 80 CAPS or 260 ports, which is lower than the NETANN preferred capacity of 120 CAPS or 500 ports. The graph in Figure 93 provides a comparison of the test scenarios.

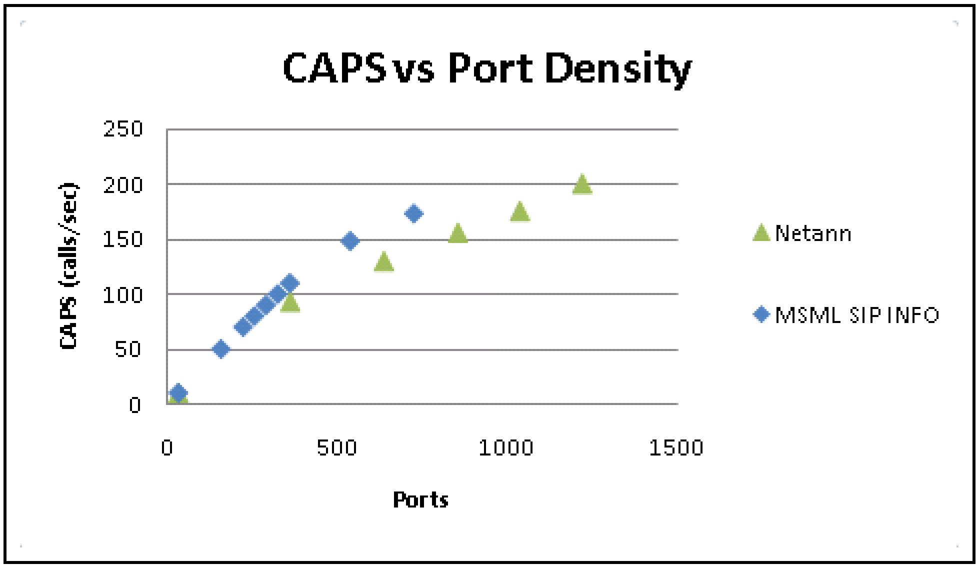


Figure 93: CAPS Versus PD—NETANN and MSML Announcement (Linux)

In Figure 93, we can see that testing starts to failed beyond 80 CAPS in the MSML test case. Even the call duration deviation is better than in the NETANN test case. See Figure 94.

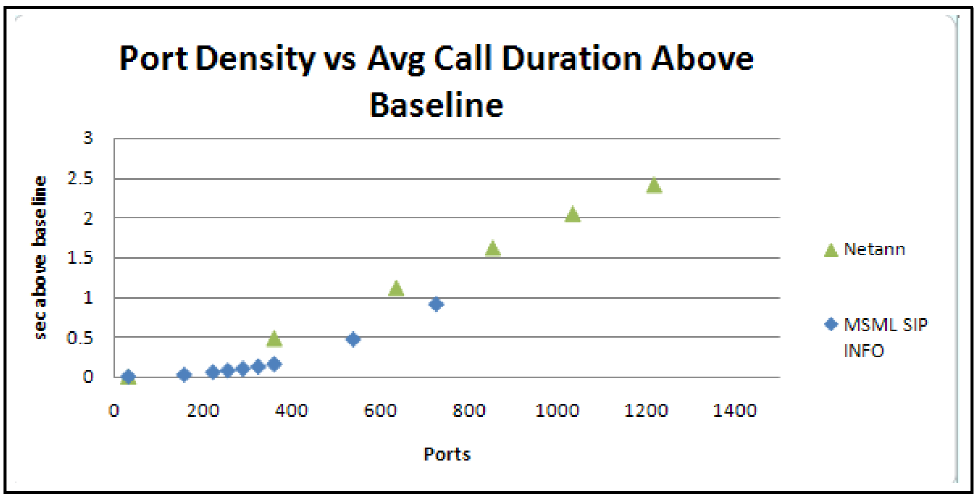


Figure 94: PD Versus ACD—NETANN and MSML Announcement (Linux)

Overall system CPU usage is illustrated in the graph in Figure 95. We can see that overall CPU usage is quite similar, but MSML test case is slightly higher than NETANN at high ports, which is beyond peak capacity

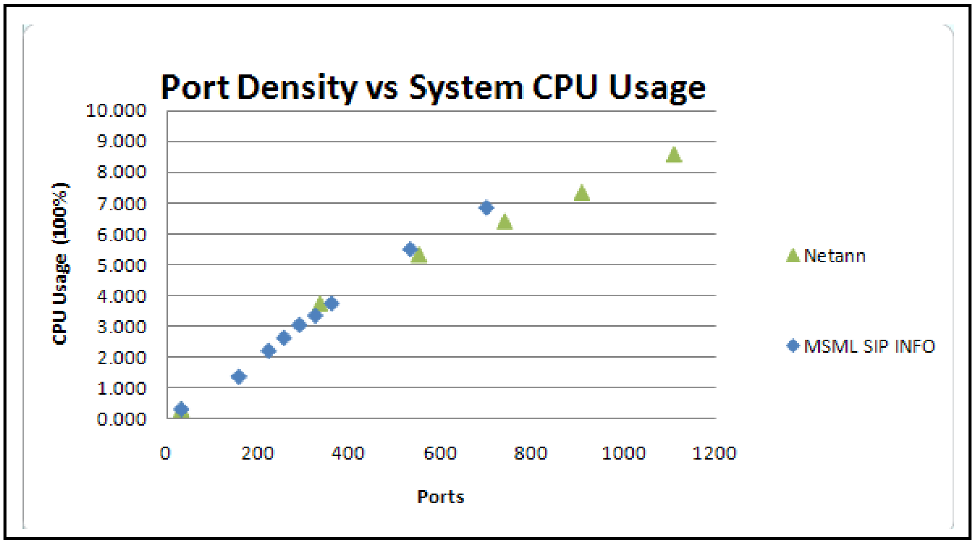


Figure 95: PD Versus System CPU Usage—MSML and NETANN (Linux)

As indicated by the graph in Figure 96, performance for the 10 seconds announcement application when gvp:precheck is turned off, is almost the same with 1 or 2 audio files in a single prompt (200 CAPS or 2000 ports) while two prompts (SIP INFO) only achieve 130 CAPS or 1400 ports.

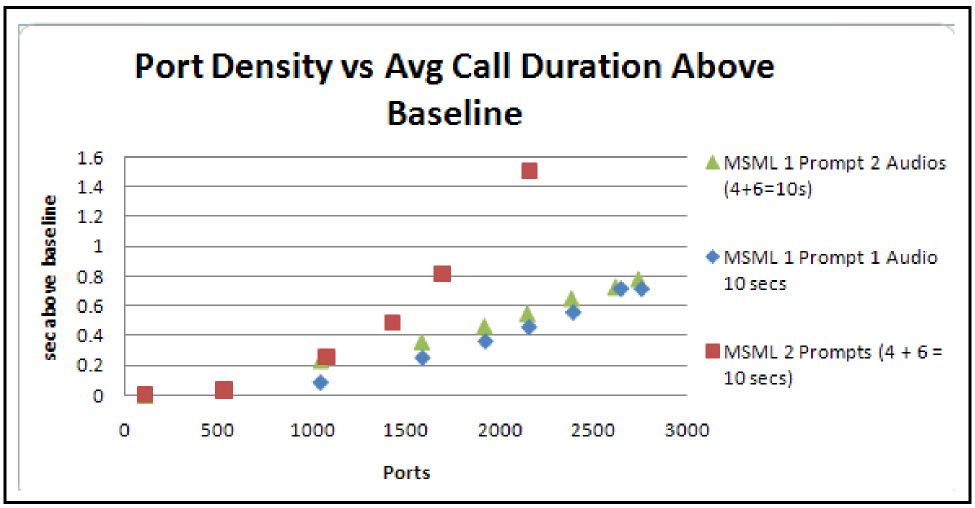


Figure 96: PD versus ACD—MSML Application (Linux)

In Figure 96 we can see that the performance of 2 single prompt test case results are quite similar, while the call duration increases significantly for the 2 prompts scenario. We can see some trending in the overall CPU usage in Figure 97.

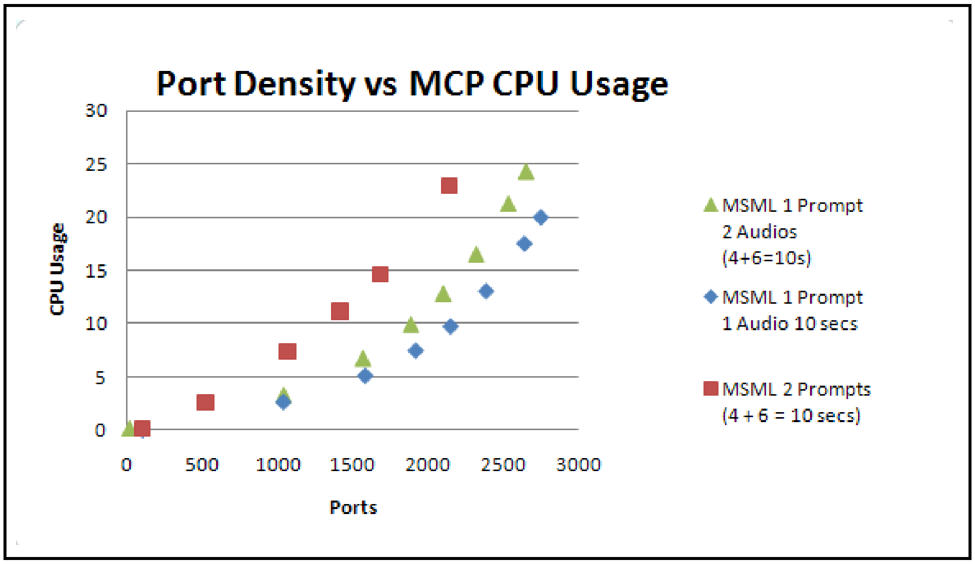


Figure 97: PD Versus MCP CPU Usage—MSML Application (Linux)

In the 2 prompts test case in Figure 97, the CPU usage is significantly higher than in the 2 single-prompt test.s

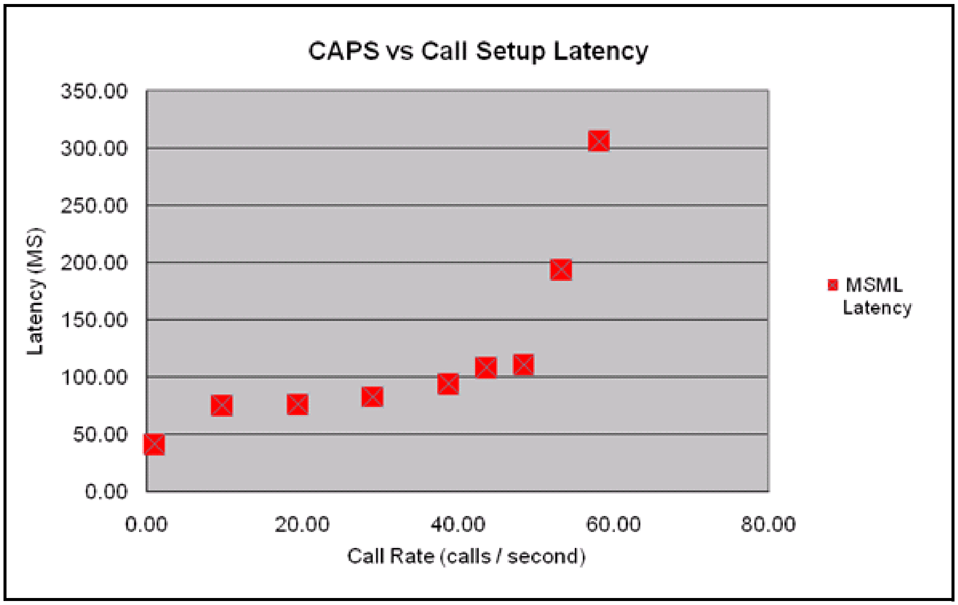
In a 20 seconds announcement scenario with gvp:precheck turned off, when there are more audio files, even in a single prompt, performance is impacted: 130 CAPS or 2600 ports for three audio files in a single prompt versus 150 CAPS or 3000 ports for one audio file in a single prompt.

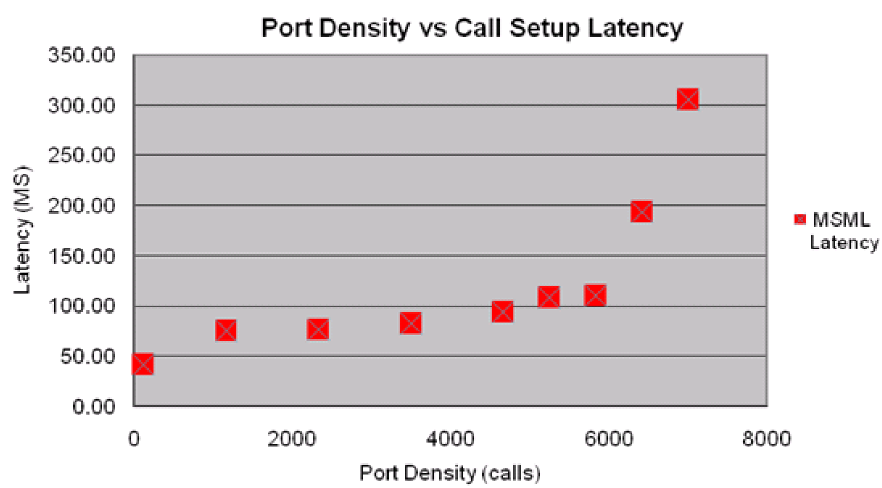
Multiple prompts can also downgrade the performance, in that only 100 CAPS, or 2000 ports are achieved in the 3 prompts test case (3 SIP INFO messages).

MSML Announcement Latency

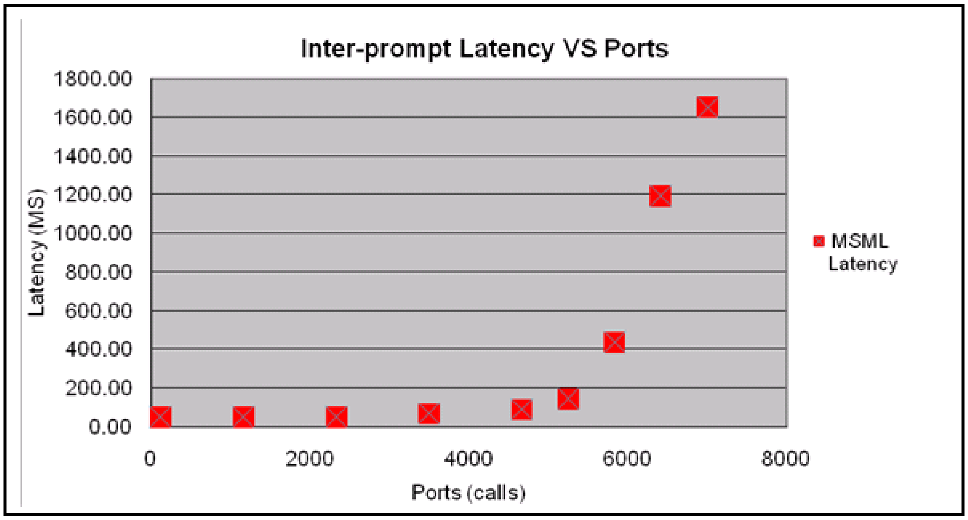
Latencies for MSML Play Announcement are measured using the following metrics: call setup latency (from INVITE to first RTP packet), SIP INFO (with MSML) response latency and Inter-prompt latency. The background load is a prompt (G.711), audio only and lasting 120 seconds.

Below are two graphs for call setup latency of MSML Play Announcement on a four-VM setup of vSphere using two Xeon E5620 CPUs (eight cores) with 16GB RAM. Each VM used two vCPUs and 4GB RAM. The guest OS is Windows 2008 Server R2 Enterprise. The first of the two graphs shows the latency (in milliseconds) based on call rate and the other shows port density. You can see a small jump in latency when the load goes above 50 cps or 6000 ports. The latency is still below acceptable criteria (500 ms). The overall CPU usage approaches 70% when the CAPS rate is 50.

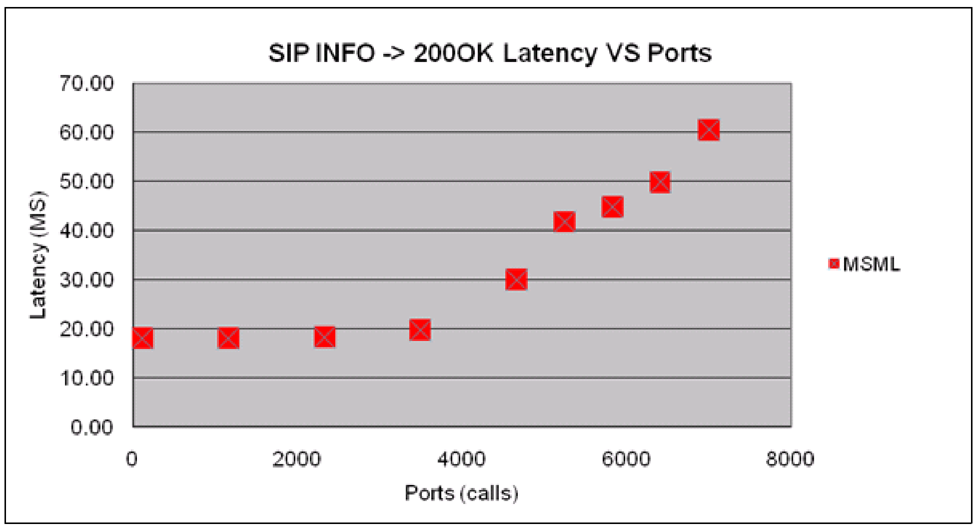




You can also see a small jump in inter-prompt latency (using audio files) when the load goes beyond 6000 ports.



The following graph shows the SIP INFO (with MSML embedded) response (200 OK) latency.

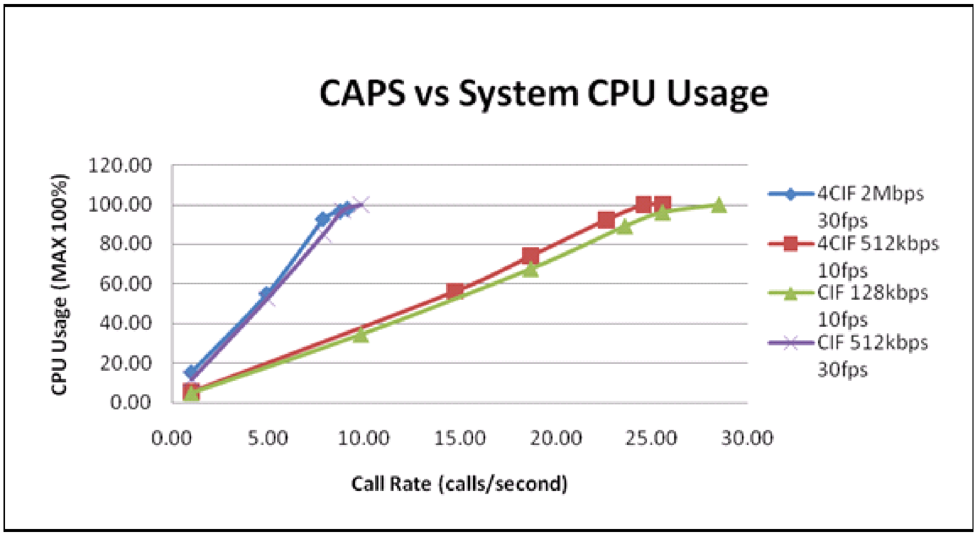


MSML Video Play Performance

Several factors can affect the performance of Video Play using MSML Announcements, such as resolution, bit rate and frame rate. For H.263 video (tested using the AMR audio codec), the following tests are chosen for analysis and comparison:

- CIF, 512 Kbps bit rate (high), 30 fps (high).
- CIF, 128 Kbps bit rate (low), 10 fps (low)
- 4CIF, 2 Mbps bit rate (high), 30 fps (high).
- 4CIF, 512 Kbps bit rate (low), 10 fps (low)

Tests were conducted using three VMs under vSphere 5.0 on a single hex-core machine using Xeon X5670. Each VM was assigned two vCPUs with only one MCP installed on each VM. The results of each test follow.



In the graphs below, the dominant factor of peak capacity is the frame rate, while the impact from both bit rate and resolution is small. The CPU is the apparent bottleneck in these tests, meaning that additional capacity can be reached with more powerful CPUs.

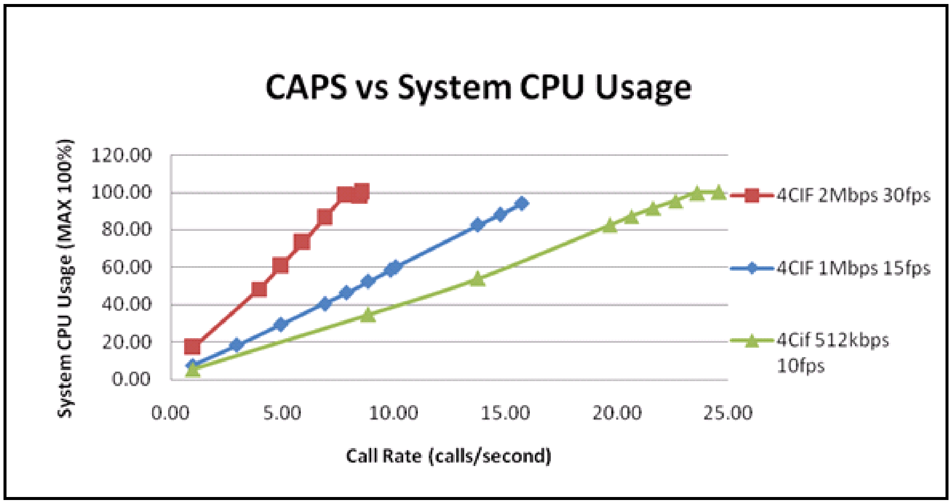
For H.264 video (with AMR audio), using finer granularity resulted in more groups of combinations:

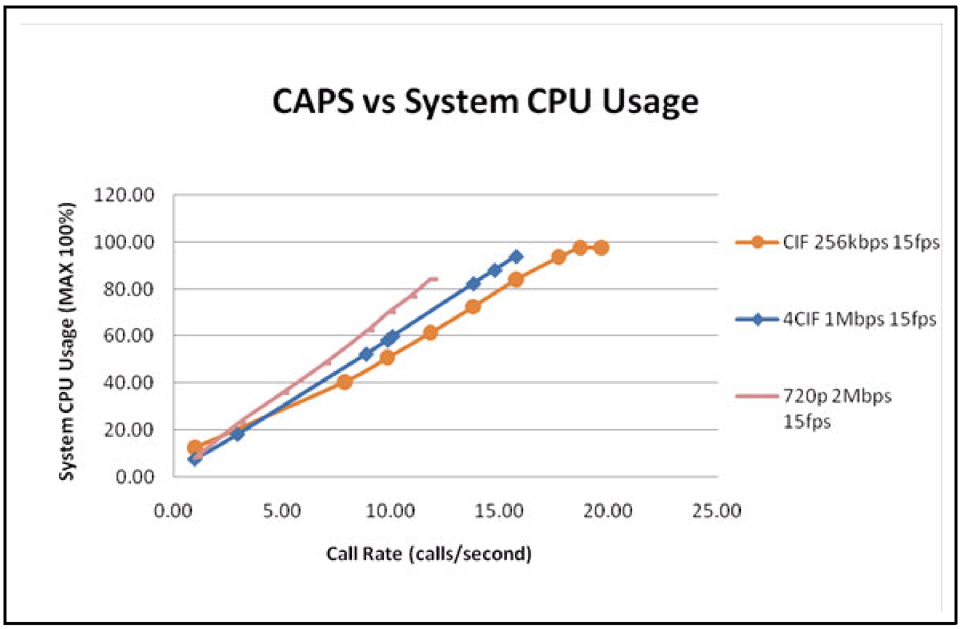
- CIF, 512 Kbps bit rate (high), 30 fps frame rate (high)
- CIF, 256 Kbps bit rate (middle), 15 fps frame rate (middle)
- CIF, 128 Kbps bit rate (low), 10 fps frame rate (low)

- 4CIF, 2 Mbps bit rate (high), 30 fps frame rate (high)
- 4CIF, 1 Mbps bit rate (middle), 15 fps frame rate (middle)
- 4CIF, 512 Kbps bit rate (low), 10 fps frame rate (low)

- 720P, 4 Mbps bit rate (high), 30 fps frame rate (high)
- 720P, 2 Mbps bit rate (middle), 15 fps frame rate (middle)
- 720P, 1 Mbps bit rate (low), 10 fps frame rate (low)

A similar trend is seen when testing H.264. The first of the following graphs shows how varying bit rate and frame rate, while keeping the resolution constant (4CIF), affects CPU usage. The second graph shows how varying the resolution and bit rate, while keeping the frame rate constant (15 fps), affects CPU usage.





Keeping the resolution constant, and varying the frame rate and bit rate, causes larger variations in CPU usage and peak capacity, while keeping the frame rate constant and varying the resolution and bit rate does not.

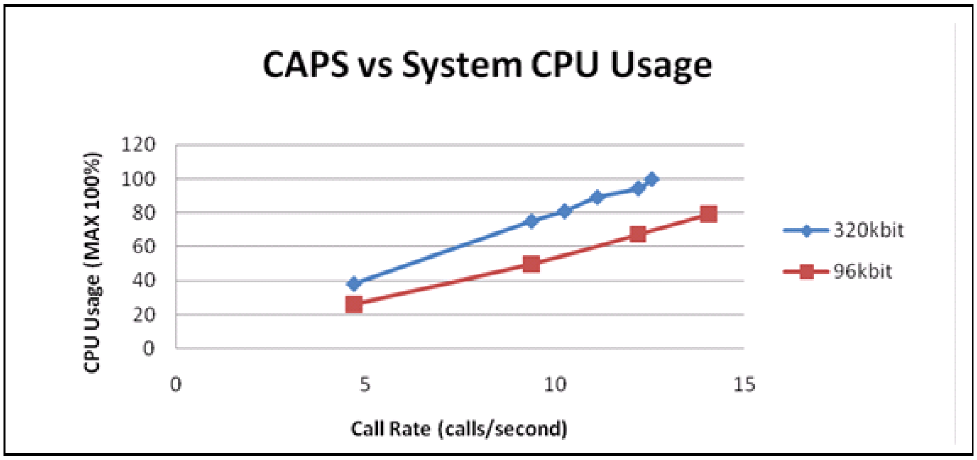
MSML MP3 Play & Recording Performance

MP3 playback using MSML Announcements was tested using the G.711 ulaw audio codec on the RTP channel. The MCP was required to transcode from MP3 to G.711 ulaw during this test. Two types of MP3 files were used in these tests:

- 96K bit rate, 32KHz sampling stereo
- 320K bit rate, 44.1KHz sampling stereo

Testing was conducted with three VMs running under vSphere 5.0 on a single hex-core Xeon X5670 processor machine. Each VM was assigned two vCPUs and each ran a single MCP instance.

Transcoding involves additional CPU resources. Lower bit rate and lower sampling rates will use fewer CPU resources and achieve a higher peak capacity. The following graph depicts overall CPU usage vs. call rate for the above mentioned MP3 files:



MP3 recording using the same two bit rates resulted in a lower peak capacity, since a disk speed bottleneck was reached before CPU saturation occurred.

Transcoding

The Media Control Platform can transcode various media codecs in real time. The impact on performance from transcoding overhead varies, depending on the codec the Media Control Platform is transcoding to or from. Other variables that contribute to transcoding overhead are the number of audio prompts played by GVP and the amount of customer input received.

The worst case scenario occurs when the Media Control Platform is constantly transcoding between two codecs during the entire call. (Most VoiceXML applications require minimal encoding [G711u to AMR]). In Figure 98, the least amount of transcoding overhead is between G711u and G711a codecs, where the peak capacity drops by ~25%.

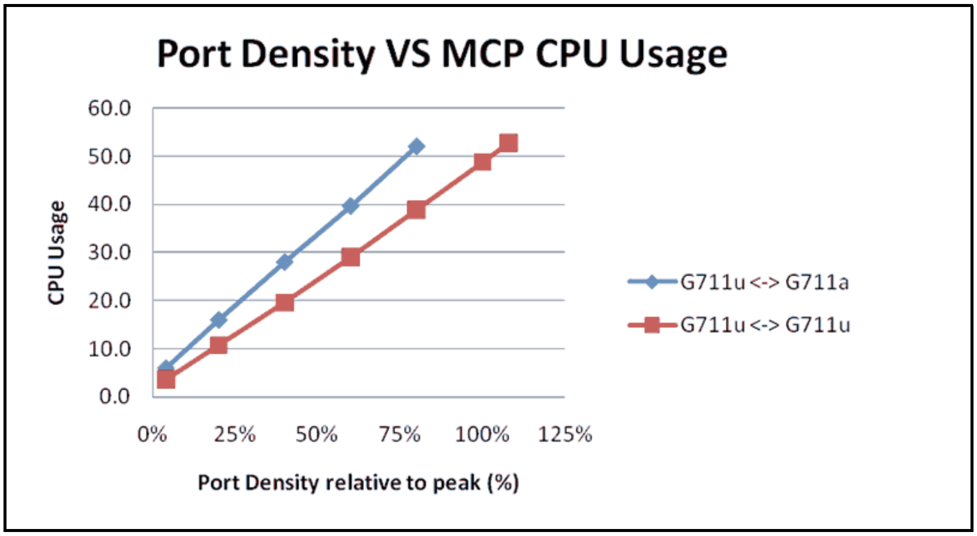


Figure 98: PD Versus CPU Usage (G711u and G711a)

The graph in Figure 99 illustrates the impact of transcoding overhead. It is greater between the G711u and AMR codecs, where the peak capacity drops by ~75%.

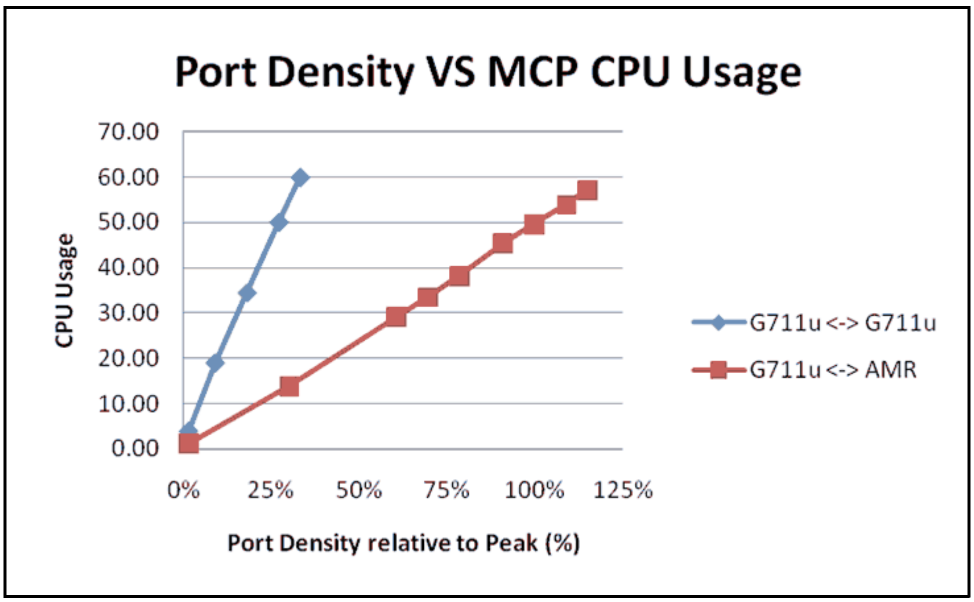


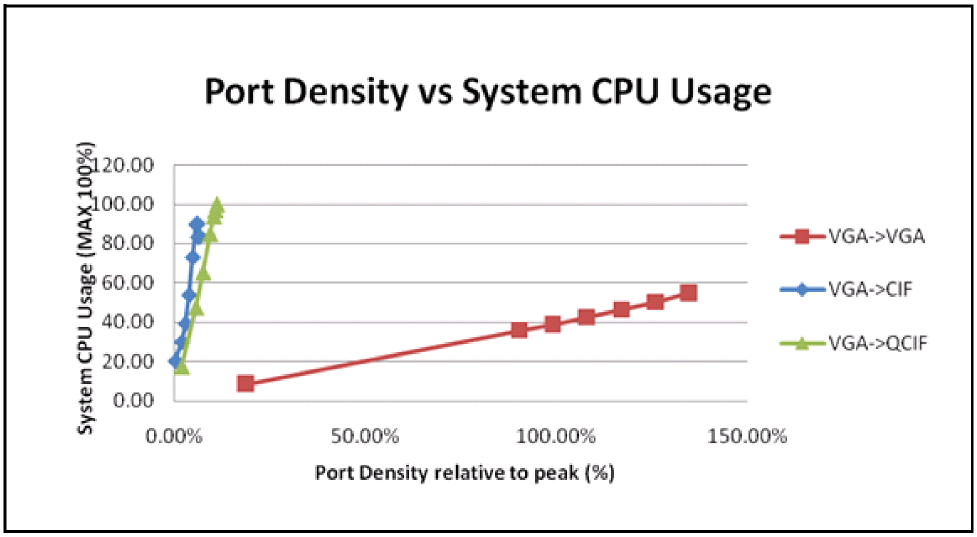
Figure 99: PD Versus CPU Usage (G711u and AMR)

As mentioned previously, the transcoding test cases in this section depict the worst case scenario involving constant transcoding between two codecs. However, in a call flow scenario where the audio stream was suppressed or silent, and the application was waiting for user input 50% of the time, transcoding overhead would be reduced by 50%.

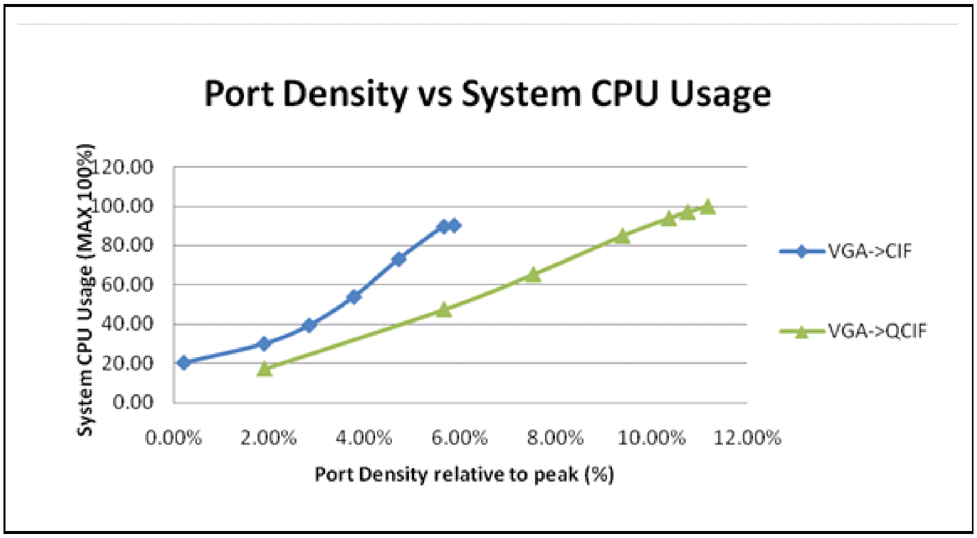
Video Transcoding Performance

Beginning with version 8.1.5, the MCP is capable of performing video transcoding. Video Transcoding Performance was measured using bridge transfers containing video and audio streams. Video transcoding requires more CPU resources than audio-only transcoding. A video bridge transfer without transcoding can achieve hundreds of ports on a machine containing 3 VMs running on vSphere 5.0, with a single hex-core Xeon X5675 processor. On the same machine, video bridge transfers that involve video transcoding can range from single digit port numbers to a few dozen. Peak capacity is affected by resolution, frame rate and bit rate.

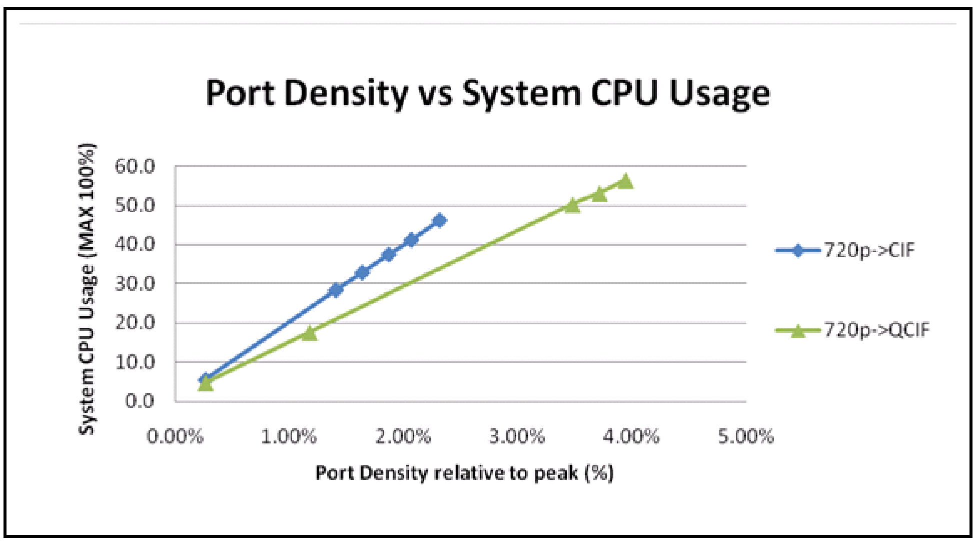
Transcoding was tested with video codec H.264 and AMR audio codec. The testing was divided into groups that consist of resolution downscaling, frame rate downscaling and bit rate downscaling. Below is the group of resolution downscaling from VGA to CIF and QCIF, respectively, with the same frame rate of 30. It shows the transcoding performance drops up to ~80%.



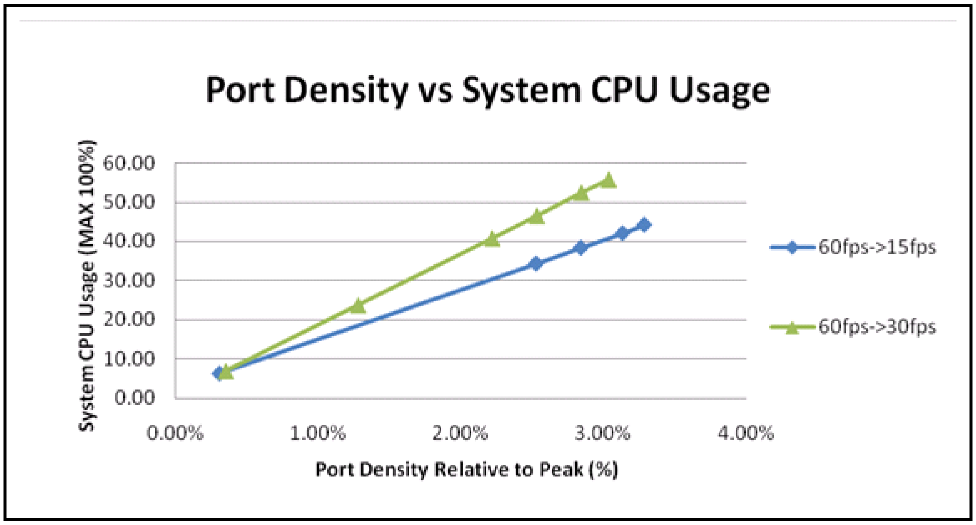
Below is the focus for transcoding only:



With higher resolution, downscaling transcoding of the performance drops even further. Below is a graph for transcoding from 720P to CIF and QCIF. You can see that performance drops up to ~90% for VGA to QCIF while ~95% for VGA to CIF transcoding.



As for a downscaled frame rate, the same resolution (VGA) was used for testing, and performance dropped more than 95%.



Secure RTP

Secure Real Time Protocol (SRTP) performance testing was conducted by using two bridge transfer scenarios with unidirectional RTP/SRTP streams; One included PCMU audio only, and the other, a 3gp container of H.264 video plus AMR audio. The PCMU audio only transfer was tested on both Windows and Linux, while the video plus audio transfer was tested on Linux only.

Tests were conducted with SRTP in the following scenarios (and one with RTP only) to provide comparison of results:

1. Baseline of RTP only (without SRTP)
2. Default SRTP mode (encrypted and authenticated) encryption
3. Default SRTP mode (encrypted and authenticated) decryption
4. Unencrypted SRTP mode (authenticated only) encryption
5. Unencrypted SRTP mode (authenticated only) decryption
6. Unauthenticated SRTP mode (encrypted only) encryption
7. Unauthenticated SRTP mode (encrypted only) decryption

Based on the test results, we can conclude that peak capacity is almost the same for SRTP and RTP, regardless of the SRTP mode used. The audio only tests resulted in 1200 ports achieved on Windows and 1500 ports on Linux, and 400 ports for the video + audio test case (on Linux only).

Capacity results were also the same, regardless of SRTP mode, however, CPU utilization results varied.

The graph in Figure 100 depicts the audio only test case on Windows, which compares CPU usage in testing scenarios 1, 2, and 3.

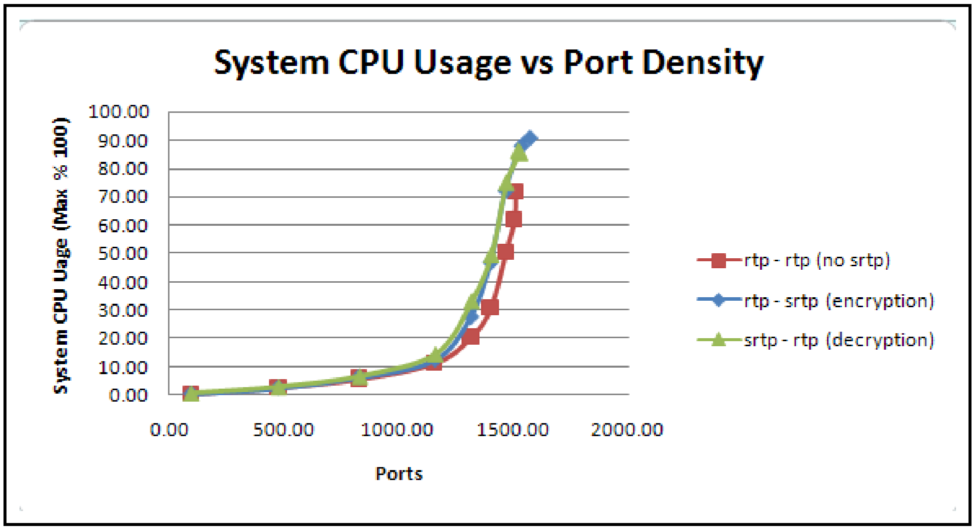


Figure 100: System CPU Usage Versus PD—Test Cases 1, 2, 3

The graph in Figure 101 depicts the audio only test case on Windows, which compares CPU usage in testing scenarios 2, 4, and 6.

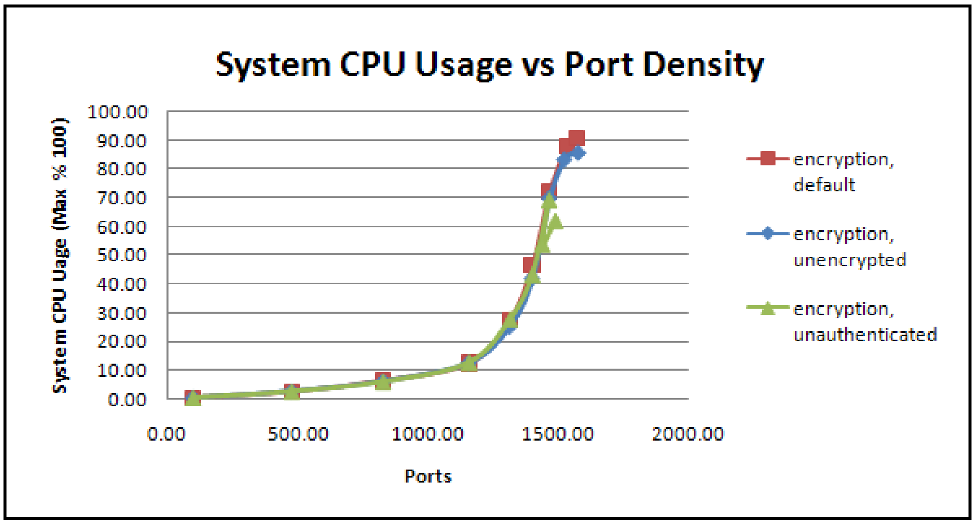


Figure 101: System CPU Usage Versus PD—Test Cases 2, 4, 6

In Figures 101, you can see that CPU usage increases beyond 1200 ports (launching 1300 ports), meaning the peak capacity is also 1200 ports. The following additional results were observed:

- In the encryption scenario, the overall system CPU usage increased from 11.4% to 12.5%—a 10% increase at 1200 ports.
- In the decryption scenario, the overall system CPU usage increased from 11.4% to 14.4%—a 26% increase at 1200 ports.
- The difference in CPU usage is negligible whether SRTP is configured in default (encrypted and authenticated), unencrypted, or unauthenticated mode.

In Figure 102, the audio only test case (on Linux), the CPU usage is more linear than on Windows, therefore, 1500 ports (launching 1700 ports) is considered peak capacity. The following additional results were observed:

- In the encryption scenario, the overall CPU usage increased from 22.8% to 33.1%—a 45% increase.
- In the decryption scenario, the overall CPU usage increased from 22.8% to 31.4%—a 38% increase.
- The difference in CPU usage is negligible whether SRTP is configured in default (encrypted and authenticated), unencrypted, or unauthenticated mode.

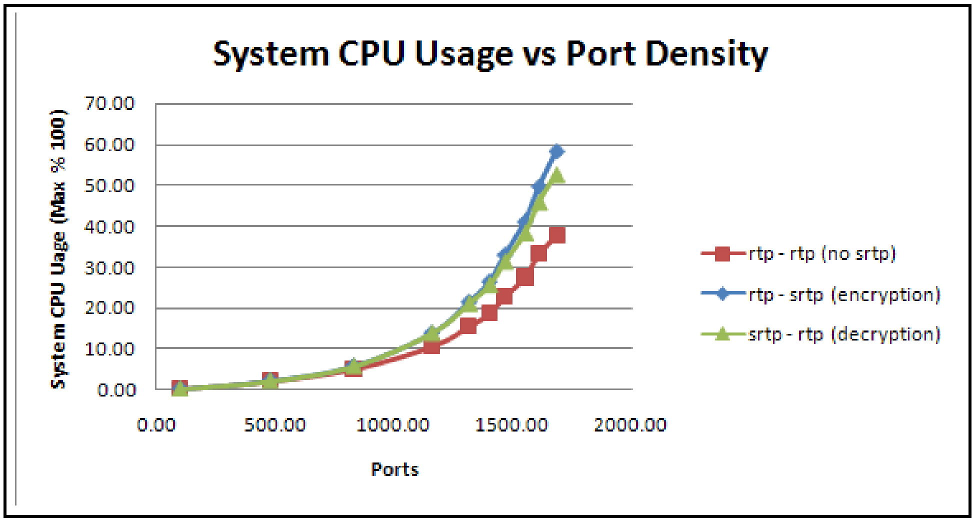


Figure 102: System CPU Usage Versus PD—Audio Only Test Case

In Figure 103, in the video + audio test case, you can see that memory usage is causing the bottleneck. The graph depicts a comparison of virtual memory usage when default encryption, default decryption SRTP mode, and an RTP only scenario is tested. All of these test case results approach the 3 GB limit when ports reach 400. Even in the RTP only test case, the virtual memory is only slightly lower. Therefore, 400 ports would be considered peak capacity.

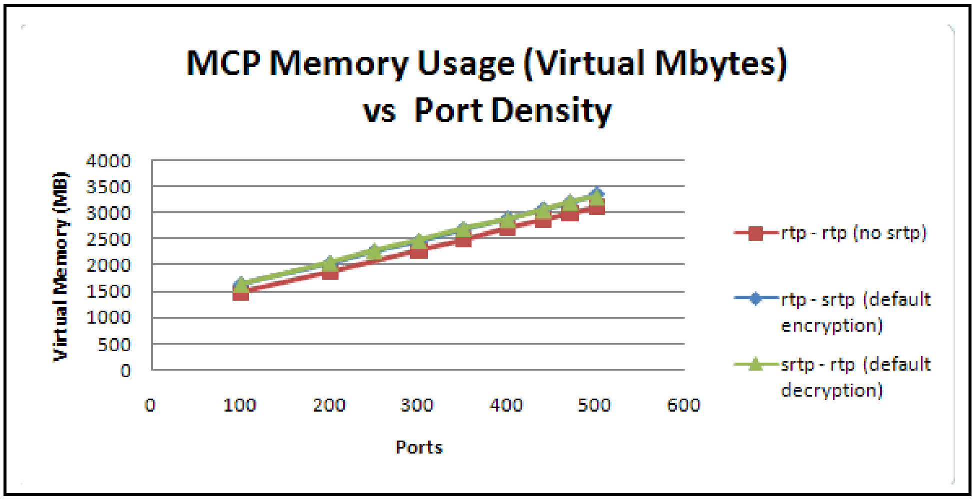


Figure 103: MCP Memory Usage Versus PD

The graphs in Figures 104, 105, and 106 provide comparisons of the system CPU usage in various encryption and decryption test case scenarios.

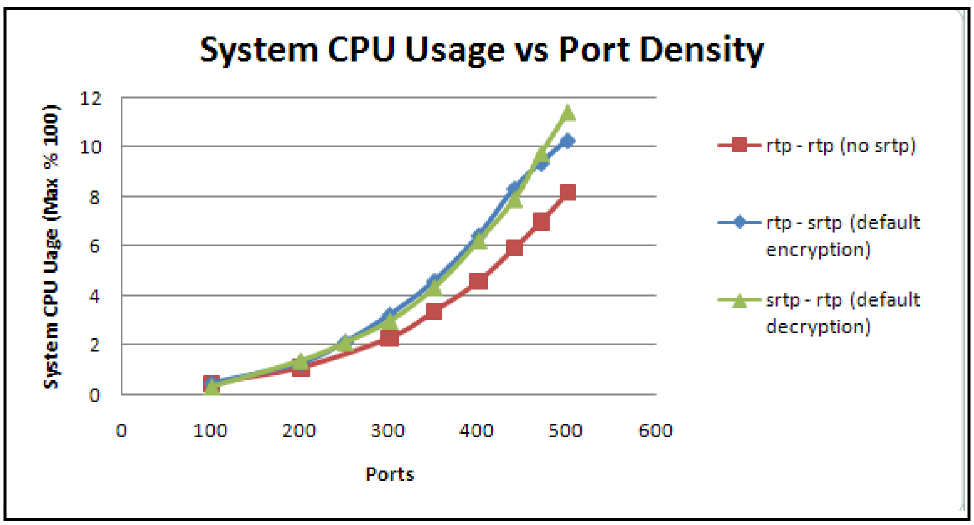


Figure 104: System CPU Usage Versus PD—Default Encryption, Decryption

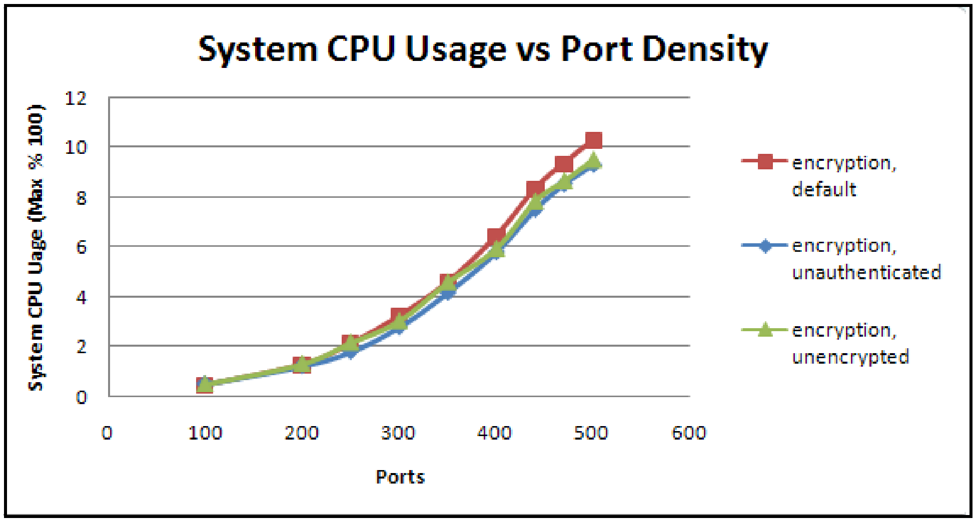


Figure 105: System CPU Usage Versus PD—Encryption

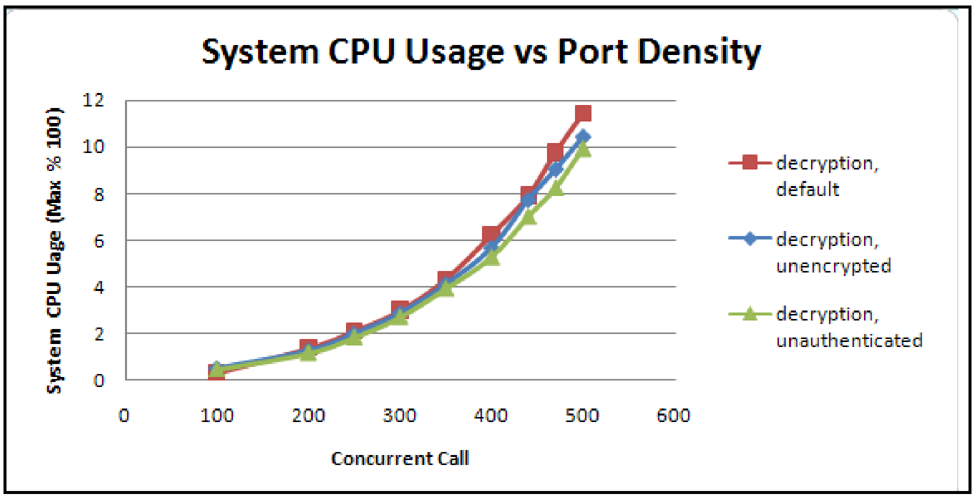


Figure 106: System CPU Usage Versus PD—Decryption

The following conclusions can be made, based on the SRTP/RTP test results:

- In the encryption scenario, the overall system CPU usage increased from 4.6% to 6.5%—a 41% increase at 400 ports.
- In the decryption scenario, the overall system CPU usage increased from 4.6% to 6.2%—a 35% increase at 400 ports.
- In the encryption scenario, the unencrypted and unauthenticated mode test cases indicates lower CPU usage than the default mode test cases—89% and 91% respectively at 400 ports.
- In the decryption scenario, the unencrypted and unauthenticated mode test cases indicated lower CPU usage than the default mode test cases—85% and 92% respectively at 400 ports.

Play Cache

Enabling the play cache functionality increases overall capacity. The transcoding occurs just once, during the first call. Those transcoded contents is cached and reused in all subsequent calls, and resources normally used for transcoding are no longer be needed.

Figures 107, 108, and 109 are graphs of data derived from MP3 file playback via MSML play.

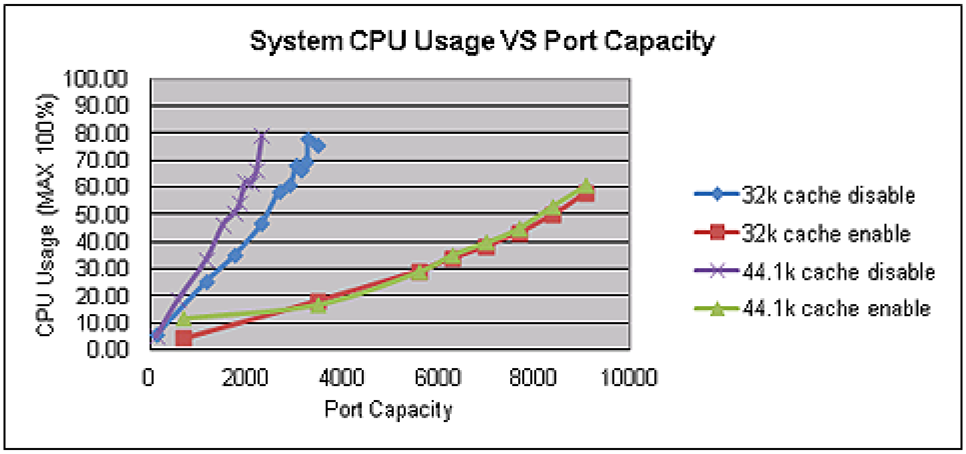


Figure 107: System CPU Usage vs. Port Capacity (audio only)

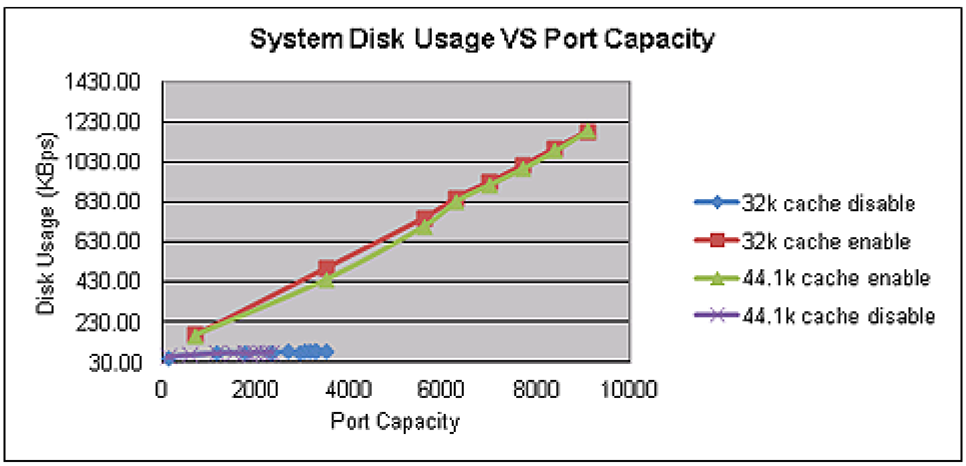


Figure 108: System Disk Usage vs. Port Capacity

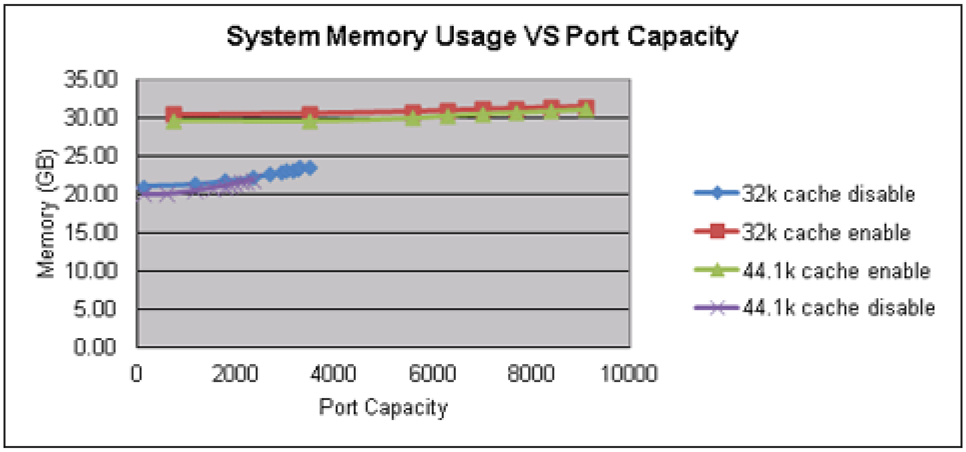


Figure 109: System Memory Usage vs. Port Capacity

Notes: By storing and then accessing a transcoded codec, the play cache both conserves and expends system resources. It is a mostly positive trade-off.

Play Cache Enabled

- Transcoding (and its strong demand on system capacity) occurs only during the first call; system performance improves noticeably on all subsequent calls.
- The play cache consumes system memory when enabled, which increases disk traffic and affects system performance.

Play Cache Disabled

- CPU usage is intensive during transcoding, which occurs for every call. System performance is noticeably affected.

The graph below compares the results for transcoding video and audio with the play cache enabled and disabled. The video stream is transcoded from 720p (30fps, 4Mbps, High profile and level 3) to CIF (Main profile and level 2). The audio is AMR. The source file is 3gp. Note that capacity is even further impacted than with audio only.

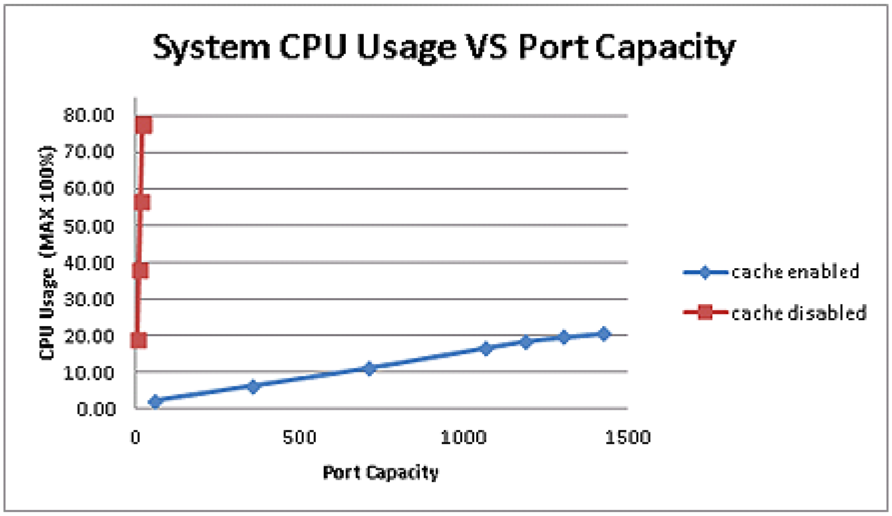


Figure 110: System CPU Usage vs. Port Capacity (video and audio)

Conference Performance

In the following conference use cases, three variables affect Media Control Platform performance:

- The number of simultaneous conferences.
- The number of participants per conference.
- The number of speaking participants.

As the graph in Figure 111 illustrates, the variable impacting performance the most is the total number of participant hosted by GVP (the number of conferences multiplied by the number of participants per conference).

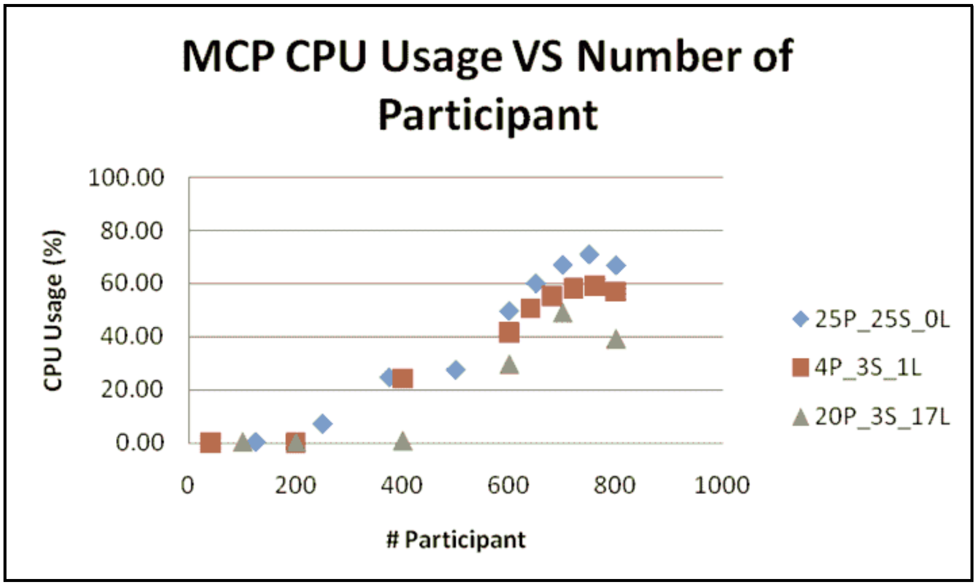


Figure 111: CPU Usage Versus Total Number of Participants

The symbols and legend in the graph in Figure 111 are explained below:

- 25P_25S_0L = 25 participants per conference (25 active speakers + 0 listen only)
- 4P_3S_1L = 4 participants per conference (3 active speakers + 1 listen only)
- 20P_3S_17L = 20 participants per conference (3 active speakers + 17 listen only)

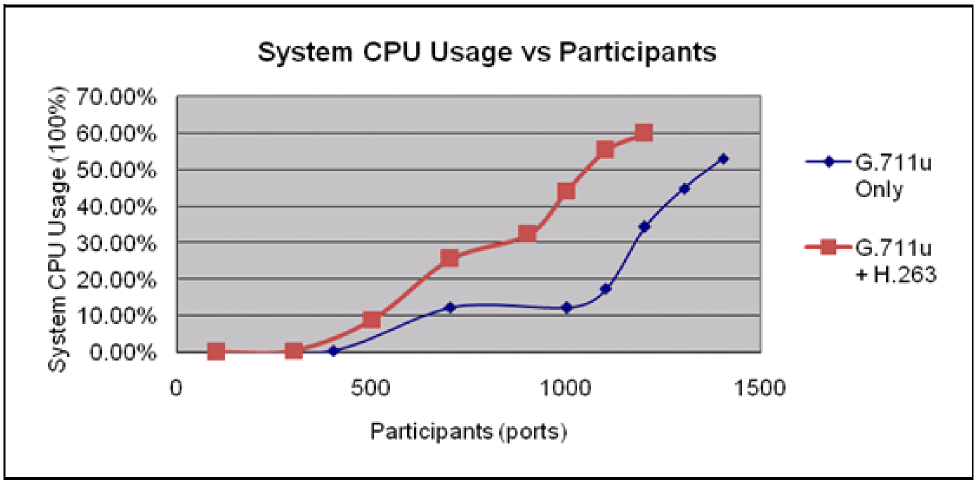
Overall, the CPU usage increases with a higher percentage of actively speaking participants. However, regardless of the conference configuration, the system bottleneck occurs when the total number of participants reaches ~600 (on a 2x Xeon5160 @3.0GHz server).

The test was conducted using pre-8.1.5 GVP versions (on a physical 2x Xeon5160 @3.0GHz server that NetAnn used to create and join the conference.).

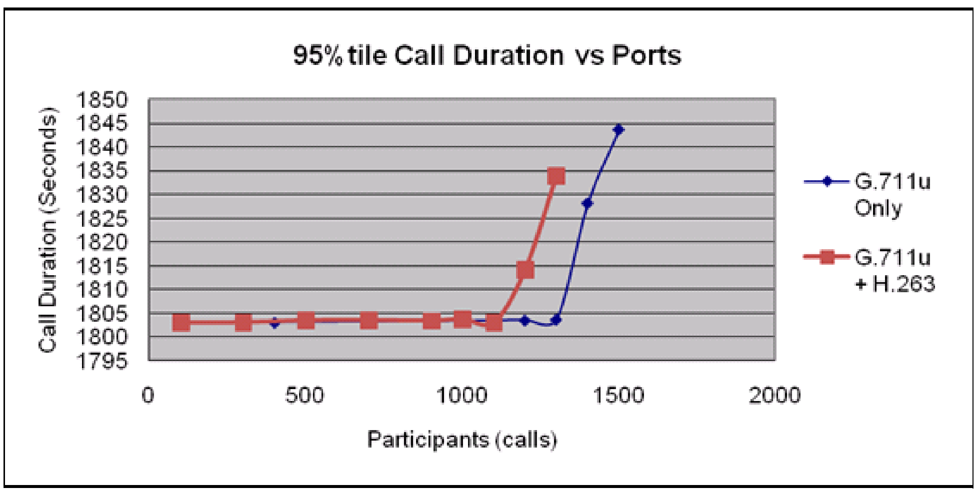
For GVP 8.1.5, there is no limit to the number of participants. Two types of testing were conducted: a conference with 32 participants and a giant conference with as many participants as possible. Both test types used MSML to create and join the conference.

The first test type (32 participants) used four VMs on vSphere 5.0 on 2x Xeon E5620 (8 cores). Each participant stayed an active speaker for five minutes (300 seconds). The slightly higher number of participants (768 participants from 24 conferences of 32 participants each) succeeded. The overall system CPU usage is not as high as before, since the bottleneck is the call to join MSML conference.

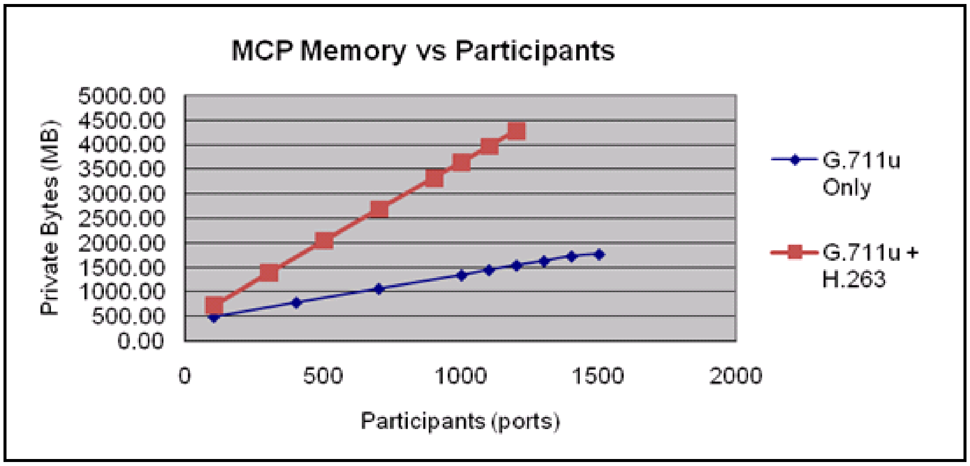
The second test used a physical machine—a Xeon X5675 @3.06GHz—since only one conference would be created. The testing was conducted with two types of codecs: G.711u audio only and H.263 video (video switching) + G.711u audio. The newly introduced MCP parameter conference.threadedoutput had to be enabled (default off) for single giant conference, otherwise, MCP could not achieve such a high number of participants. There were only three active speakers in the conference while all other participants were listeners. Each participant would stay 30 minutes (1800 seconds) in the conference. The graph below shows the overall system CPU usage:



The graph below shows that the CPU usage jumps beyond 1100 participants for the G.711u-only case and 900 for the G.711 + H.263 case, 48-hour load testing can survive 1300 participants for the G.711u-only case and 1100 participants for the G.711u + H.263 case. The ninety-fifth percentile of call duration shows that it would jump beyond 1300 participants for the G.711u-only case and 1100 for the G.711u + H.263 case.



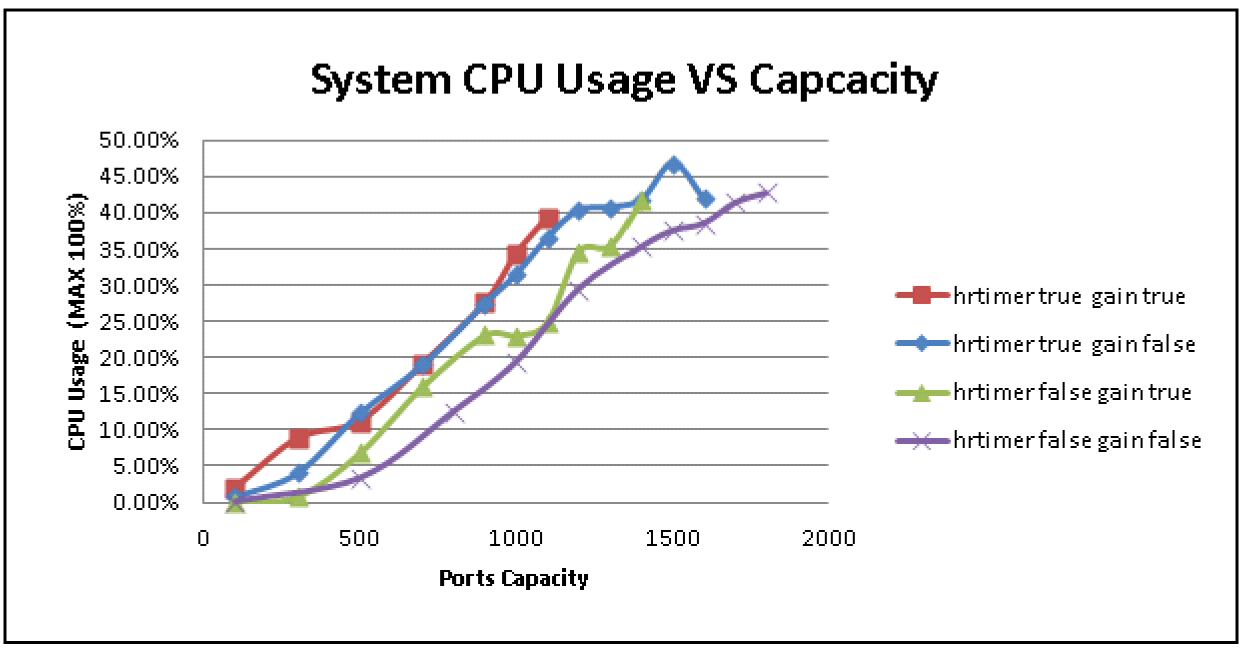
The graph below illustrates memory usage:

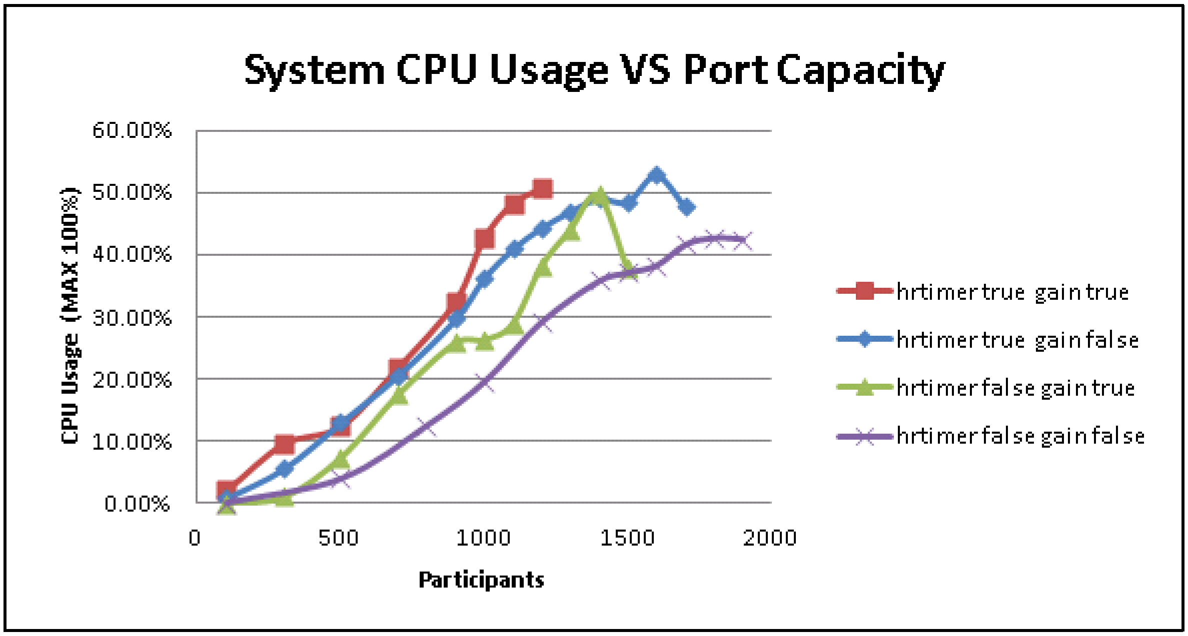


The G.711u + H.263 case shows that the memory usage exceeds 2GB beyond 500 participants (3GB virtual memory), since the test was conducted on a 64-bit OS. When a 32-bit OS is used, MCP crashed if virtual memory exceeded 3GB, so the peak capacity for G.711u + H.263 is 500.

HR Timer

Two parameters: HR Timer (specifically for Windows) and Gain Control, impact the performance of conference. The graphs below compare the performance in terms of system CPU usage from combinations of different values of these two parameters:





Note that the highest port capacity (measured in Participants) with least CPU usage can be achieved when both HR Timer and gain control turned OFF. Conversely, the lowest port capacity with the highest CPU usage is achieved when both parameters turned ON. The table below documents these results.

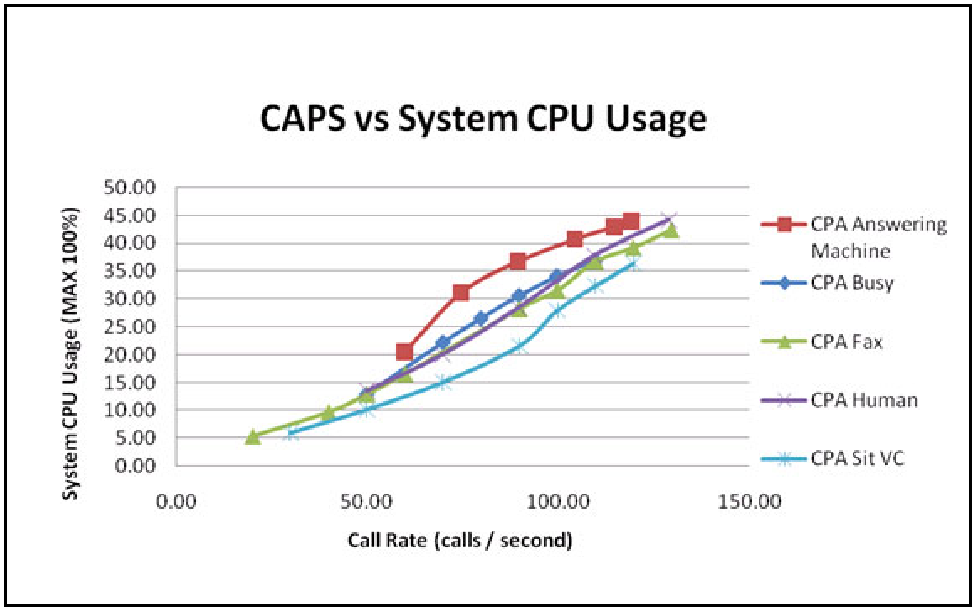
Table 109: HR Timer and Gain Control Settings vs. Port Capacity

		Gain Control	
		On	**Off**
HR	**On**	1000	1400
Timer	**Off**	1300	1800

Prior to 8.1.6, gain control was hard-coded to turned on and there is no HR timer. So in 8.1.6 gain control on and HR time off is compatible with previous releases.

MSML CPA/CPD Performance

CPA/CPD is requested through MSML, and so performance is measured by call rate. The testing was conducted with different tones such as Answering Machine, Busy, Fax, Human and SIT VC. Below is a plotted graph for all above tones tested:



The call duration varies, depending upon the type of tone and the length of recognition, and the peak call rates are quite close one another for each tone. In other words, call rate—and not ports—is a major factor determining peak capacity.

Component Test Cases

The following component test cases are described in this section:

- Media Control Platform on Windows, page 319
- Media Control Platform on Linux, page 321
- Resource Manager, page 323
- MRCP Proxy, page 328
- PSTN Connector, page 330
- CTI Connector, page 330
- CTI Connector/ICM, page 331
- Supplementary Services Gateway, page 332
- Reporting Server, page 333

Media Control Platform on Windows

Testing was conducted on Windows 2003, Windows 2008 (x86 and x64), and Windows 2008 R2, however, not all testing was executed in the same release. In general, performance results were similar when the Media Control Platform was installed on either version of Windows, however, as you can see in Figure 112, there were slight differences.

The graph in Figure 112 depicts the call setup latency metrics for each Windows version when the VoiceXML_App1 was executed as the background load.

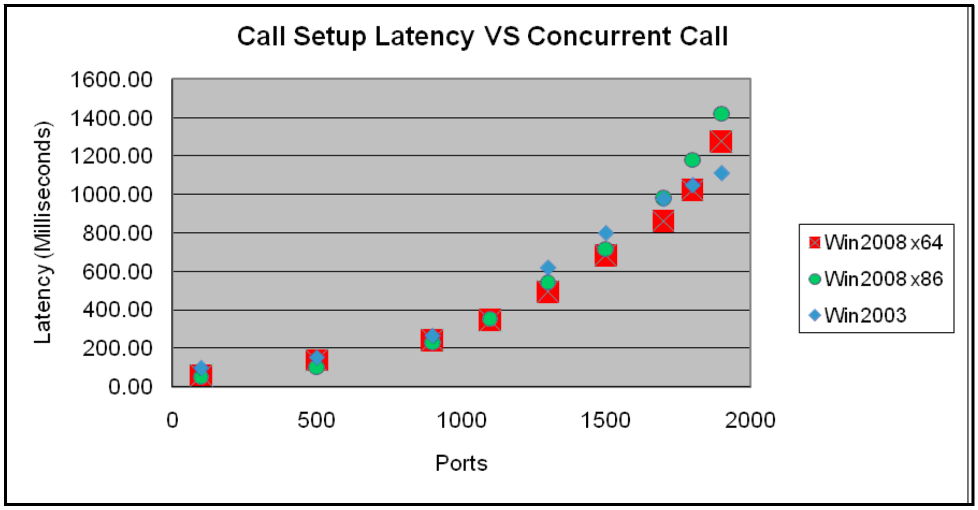


Figure 112: Call Setup Latency Versus Concurrent Calls (MCP on Windows)

The latency numbers are quite close for ports below 1300, while they are slightly higher on Windows 2003 when ports are 1300 or higher. This trend continues up to 1800 ports, at which point latency on Windows 2008 exceeds those on Windows 2003.

In Figure 113, the graph depicts CPU usage for the overall system on each Windows version, when testing was performed by using VoiceXML_App1 and the results were scaled.

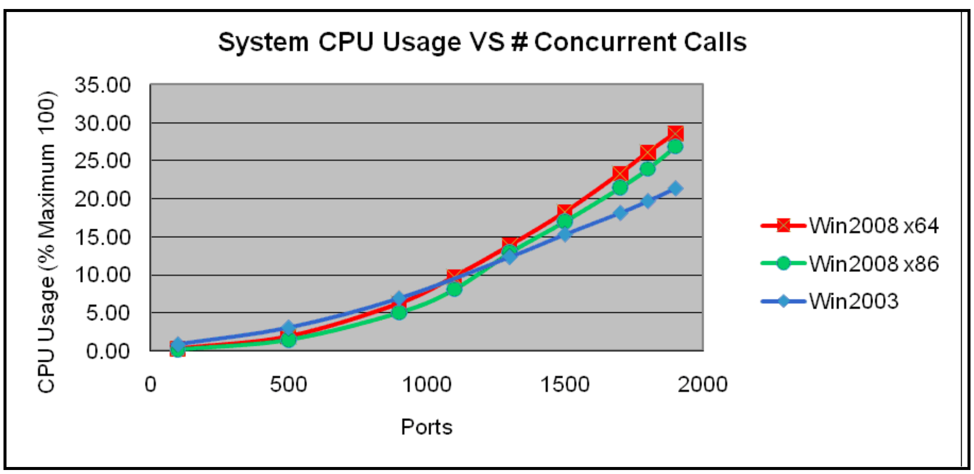


Figure 113: System CPU Usage Versus Concurrent Calls (MCP on Windows)

Notice that the CPU usage for both versions of Windows 2008 (x86 and x64) are almost in line with one another, trending a little higher than Windows 2003. It was also noted that Windows 2003 can sustain higher ports than the *preferred* 1300, (without considering other factors, such as call setup latency). Beyond 1300 ports, the Windows 2008 call pass rate drops below Genesys QA pass criteria of 99.99%. However, on Windows 2003, 1800 ports can be achieved within the pass criteria if call setup latency is ignored.

Media Control Platform on Linux

Testing was conducted on Red Hat Enterprise Linux 4, RHEL 5 (x86 and x64), and RHEL 6 x64. However, not all testing was executed in the same release. There were no significant differences in performance on RHEL 4 & RHEL 5 (both x86 & x64), while the performance was better on RHEL 5 than on RHEL 6. RedHat was alerted regarding this.

Below is the graph of call setup latency measured in different Red Hat Linux systems on physical servers except RHEL 6.x x64 (VXML_App1 was executed as background load):.

The graph in Figure 114 depicts the call setup latency metrics for RHEL 4 and RHEL 5 on physical servers when VoiceXML_App1 was executed as the background load.

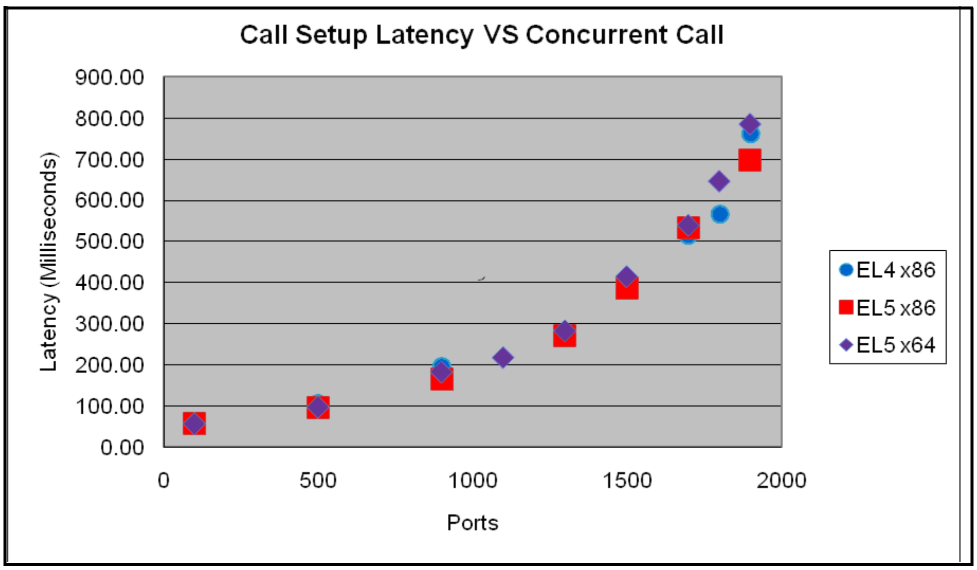


Figure 114: Call Setup Latency Versus Concurrent Calls (MCP on Linux)

All three Linux versions showed latency results that were almost in line with one another at 1700 ports or lower. Above 1700 ports, which is beyond Genesys QA *preferred* peak capacity, there were some differences.

In Figure 115, the graph provides a comparison of the overall CPU usage when VoiceXML_App1 is executed.

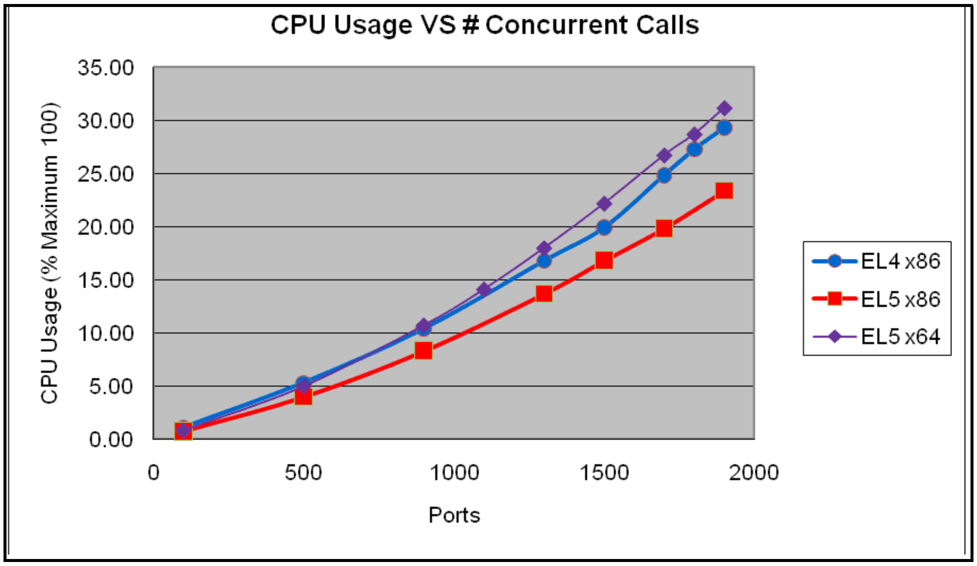


Figure 115: CPU Usage Versus Concurrent Calls (MCP on Linux)

Notice that there are no significant differences between Linux versions for overall CPU usage, and overall performance is similar, even when multiple simultaneous factors, such as call setup latency and call pass rate, are considered.

There are performance differences between Linux and Windows, depending on specific use cases. Performance (the maximum number of concurrent ports) on on Linux is slightly better than Windows in some test cases in which there are higher ports, such as MSML with CPD, but worse for other test cases, such as those in which G.711 and G.729 transcoding were used.

GVP overall performance on Linux and Windows is quite similar, and although the test cases performed on both Windows and Linux were not identical, the peak capacity overall is not significantly different.

Because performance suffers on RHEL 6.4 x64, the virtual environment was used only to test MCP on RHEL 6.x x64 as a guest OS on ESXi 5.0. Below is the comparison of call setup latency between EL5 x64 and EL6 x64 while both were on virtual environment as guest OS:

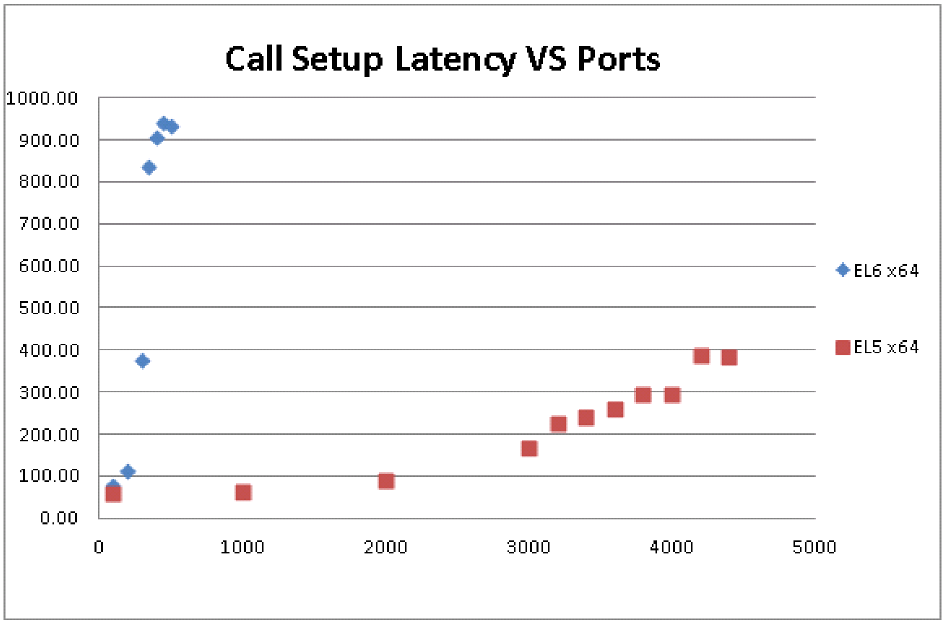


Figure 116: Call Setup Latency vs. Ports

Note that call setup latency increased significantly on EL6 x64, compared with EL5 x64.

Resource Manager

Resource Manager performance is measured in terms of CAPS. Performance is not affected by the number of simultaneous calls held by Resource Manager. Resource Manager performs most efficiently when multiple Media Control Platforms are used.

The effect on Resource Manager performance differs, depending on the type of call being processed (for example, conference versus announcement calls), but generally, a peak of 800 CAPS can be sustained for call types such as call routing and conference, and regardless of whether its in an HA or non-HA configuration. This applies to all Windows versions and most RHEL servers except RHEL 6 x64.

CPU consumption on the Resource Manager is very low. The 800 CAPS limit mentioned previously is due to the use of virtual memory, which exceeds the 3 GB limit (configured at the OS level) when Resource Manager is running consistently beyond 800 CAPS.

The same capacity results were achieved when the Resource Manager was tested using both UDP and TCP due to a bottleneck when it reached the 3 GB virtual memory limit.

The graph in Figure 117 depicts CPU usage when Resource Manager is installed on Red Hat Enterprise Linux 5 x86.

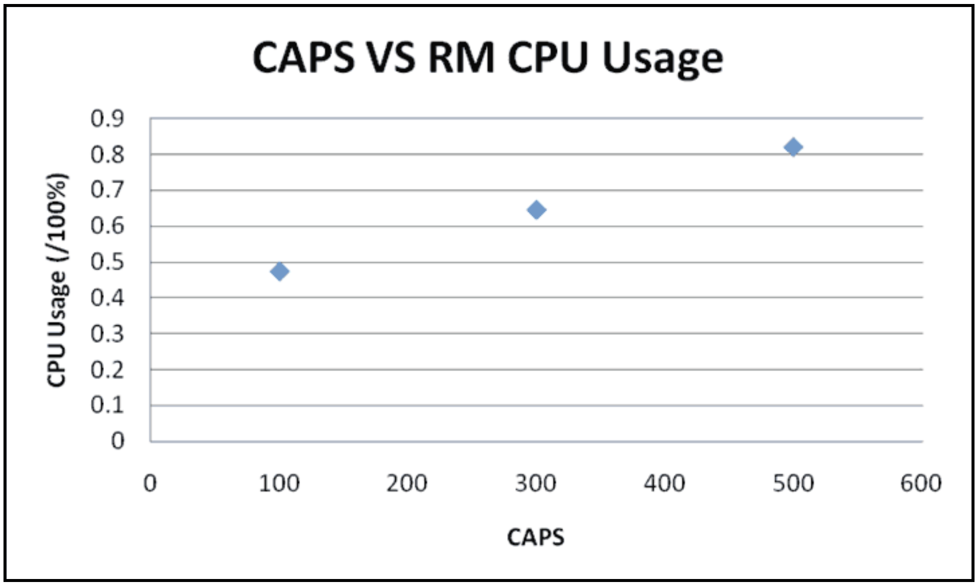


Figure 117: CAPS Versus CPU Usage (Resource Manager)

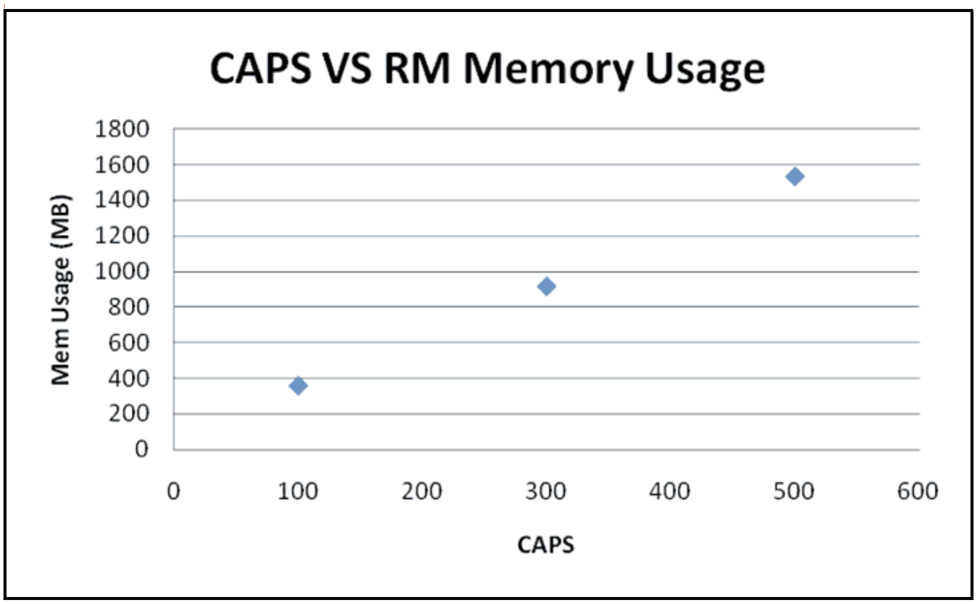


Figure 118: CAPS Versus Memory Usage (Resource Manager)

In Figure 118, we can see that the Resource Manager can sustain 900 CAPS, however, the memory usage almost reaches the 3 GB limit and a 800 CAPS peak capacity is more appropriate.

When configured with the Reporting Server, Resource Manager sustained 800 to 900 CAPS, but the Reporting Server performance invariably caused a bottleneck to occur. See "Reporting Server" on page 333.

When a single tenant (Environment by default) is used, 800 CAPS is achieved (see Figure 118). When multiple tenants are configured, the Resource Manager performance is slightly degraded. When tested with 1000 tenants, each configured with 1 child tenant, the Resource Manager performance achieves 600 CAPS of peak capacity.

When a large number of Direct Inward Dialing (DID) numbers are configured in a single tenant, the Resource Manager performance, again, is slightly degraded. When 100 K of DID numbers are tested with 262 IVR Profiles (without the use of DID ranges or wild cards, for example, a long, simple mapping list), peak capacity is at 600 CAPS.

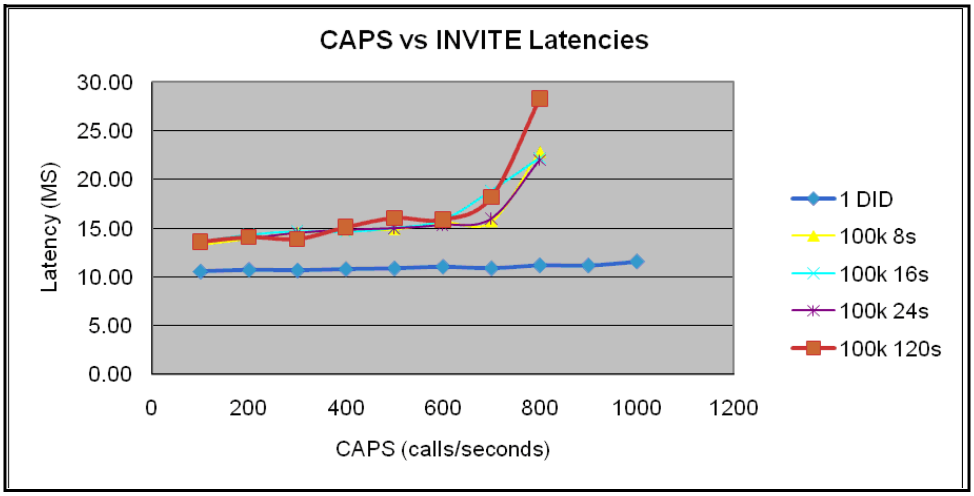


Figure 119: CAPS Versus SIP INVITE Latencies (Resource Manager)

Figure 119 depicts the propagation of SIP message latencies between 1 and 100 K of DID numbers with various call durations.

The graph in Figure 119 shows higher latencies for SIP INVITE messages for 100 K of DID numbers versus the 1 DID baseline, while there is not a lot of difference in latencies with SIP ACK messages (see Figures 121 and 122). This makes sense, as the delay likely occurs when Resource Manager searches for mappings, upon receiving SIP INVITE messages. The testing also indicates that call duration is not relevant to Resource Manager performance.

The two previous scenarios (1000 tenants with one DID entry each and a100 K DID in a single tenant) actually produce the worst results. Resource Manager can achieved better performance results when multiple tenants are configured with a small number of DID entries per tenant. Resource Manager was tested with the requirement of 1 million DIDs distributed among 32 tenants, each containing 30~35 K of DID entries. (A 4 MB size limitation exists for Management Framework objects). Even in this configuration, the Resource Manager still achieved 800 CAPS.

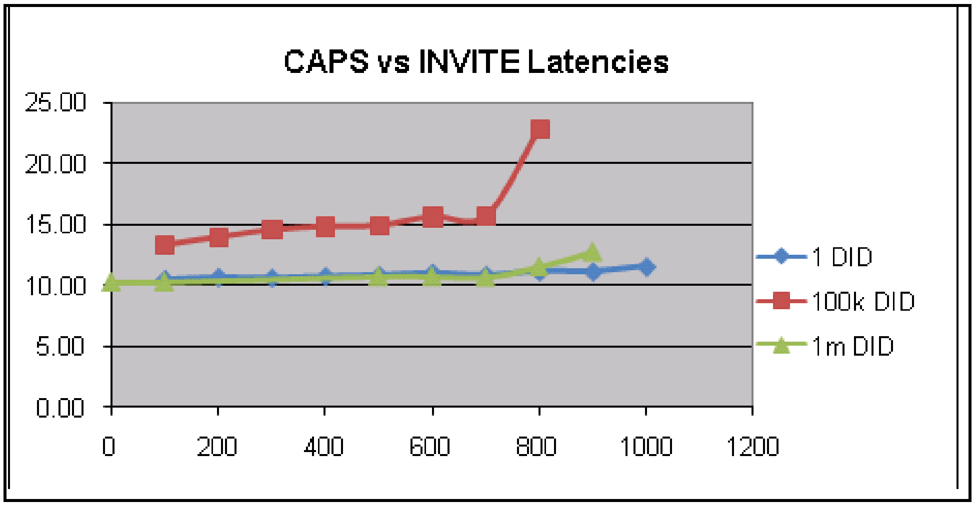


Figure 120: CAPS versus SIP INVITE Latencies (Resource Manager)

In Figure 120 the SIP INVITE latency is almost in line with one DID entry until capacity reaches 900 CAPS, then it increases. The virtual memory is also close to the 3 GB limit.

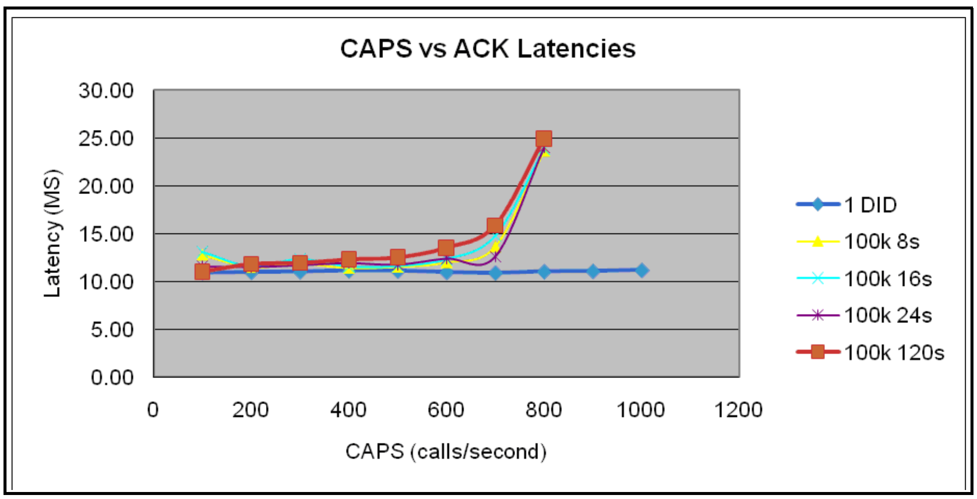


Figure 121: CAPS Versus SIP ACK Latencies (Resource Manager)

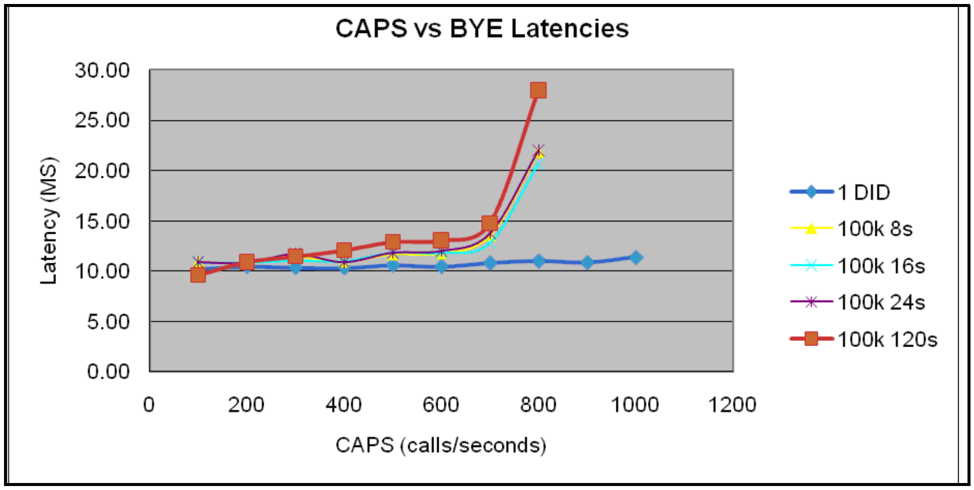


Figure 122: CAPS Versus SIP BYE Latencies (Resource Manager)

Performance superiority of RHEL 5 x86 over RHEL 6 x64 was observed during RM testing. Below is the graph for propagating latency of INVITE to compare RHEL 6 x64 and RHEL 5 x86:

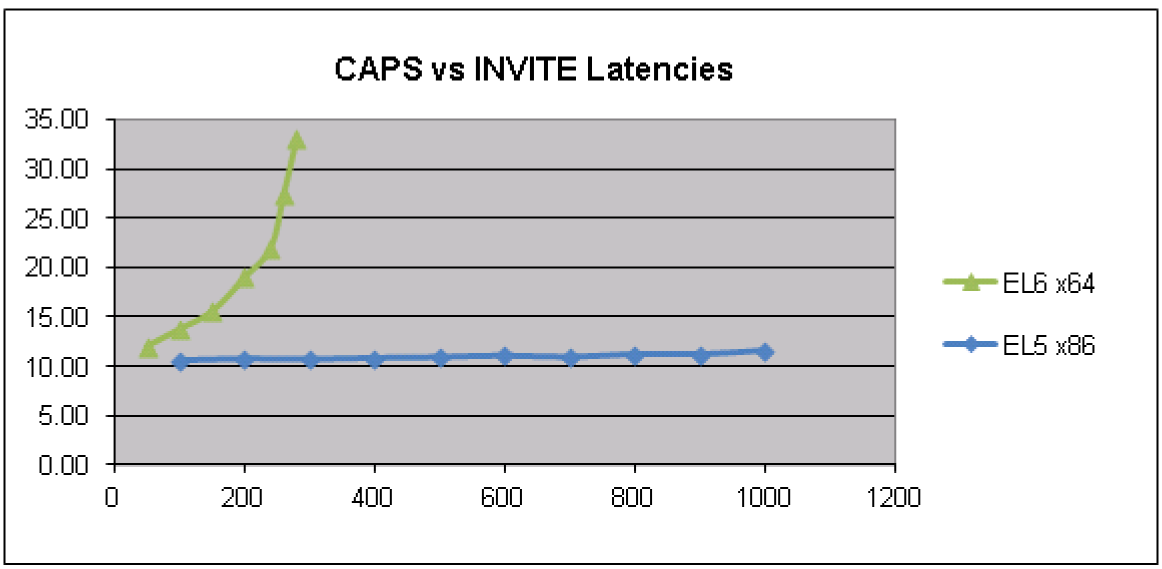


Figure 123: CAPS vs. INVITE Latencies for RHEL6 and RHEL5

It can be observed that latency increased significantly on RHEL 6 x64 even at low CAPS.

MRCP Proxy

The MRCP Proxy is a new component in GVP 8.1.4. Its performance is measured in terms of MRCP sessions and is benchmarked by using simulated MRCP v1 servers and clients. A typical 10 second MRCP session for either an ASR or TTS request is used for testing.

The peak capacity is achieved at 1,600 concurrent MRCP requests per second (half ASR and half TTS) in CAPS, but the MCRP Proxy can hold 16,000 MRCP sessions in total. Beyond 1600 CAPS, it might still respond, however, the entire session becomes quite lengthy and will eventually time out. The graph in Figures 124 to 125 depict the ASR and TTS 95 percentile of call duration. The results indicate that the call duration beyond 800 CAPS more than doubles.

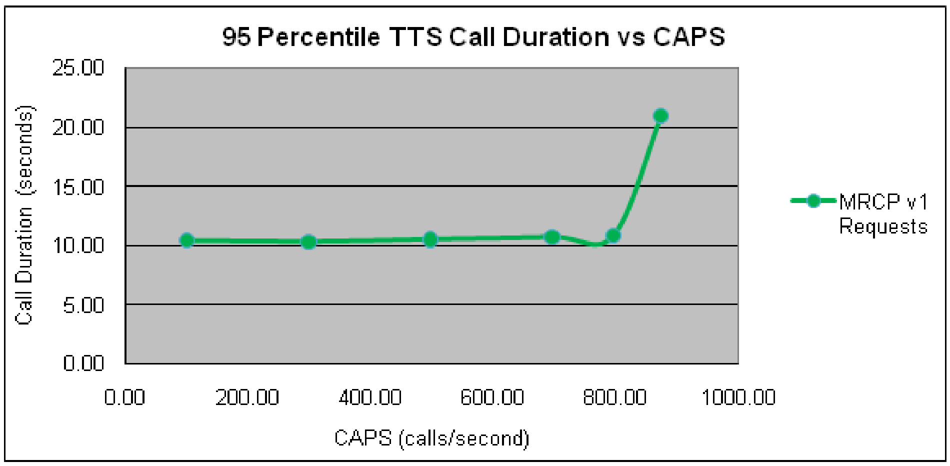


Figure 124: ASR Call Duration Versus CAPS—MRCP Proxy (MRCPv1)

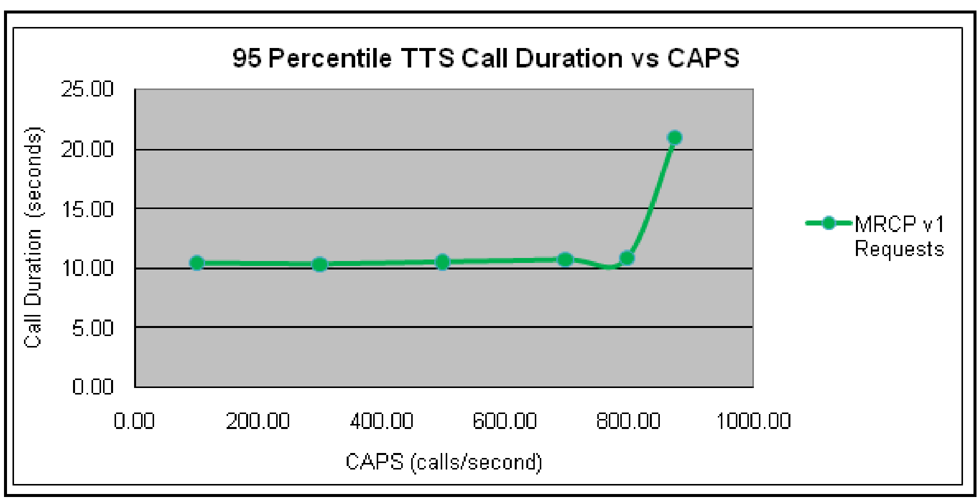


Figure 125: TTS Call Duration Versus CAPS—MRCP Proxy (MRCPv1)

The graph in Figures 125 and 126 depict the overall CPU usage for the MRCP Proxy. Note that the CPU usage increases substantially beyond 1600 CAPS.

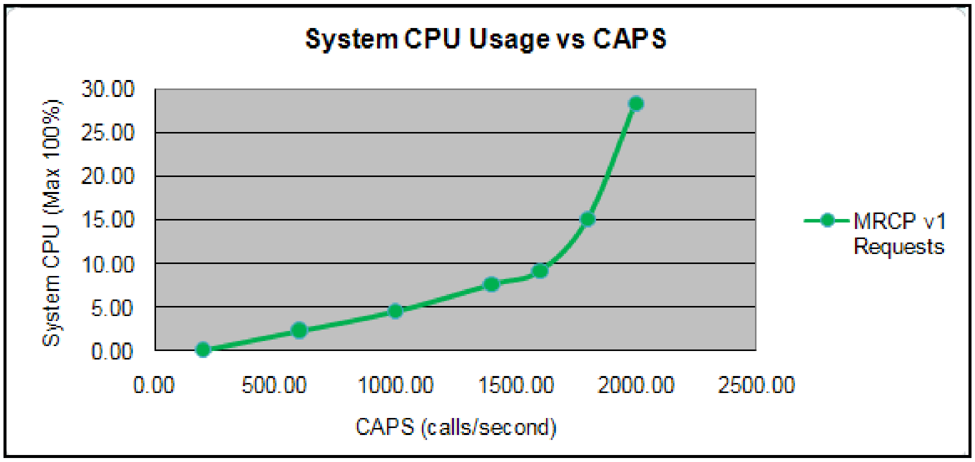


Figure 126: System CPU Usage Versus CAPS—MRCPv1 (MRCP Proxy)

As shown in Figures 127 and 128, the same significant increase in memory (private bytes) consumption beyond 1600 CAPS, is indicated.

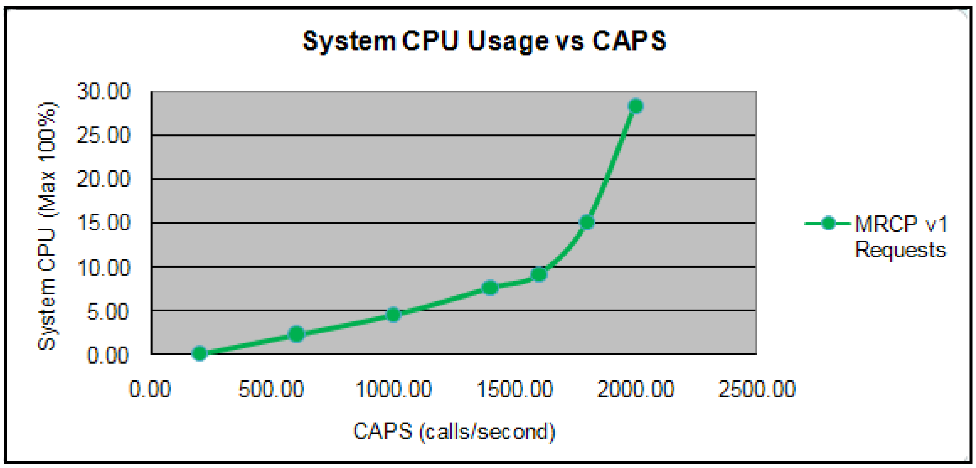


Figure 127: System CPU Usage Versus CAPS—MRCPv1 (MRCP Proxy)

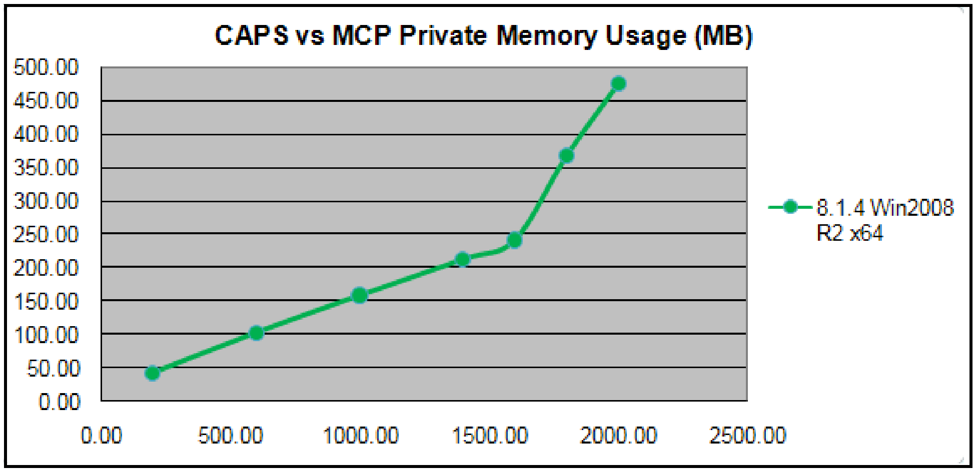


Figure 128: Memory Usage Versus CAPS (MRCP Proxy)

PSTN Connector

The performance of PSTN Connector is measured in terms of T1/E1 spans. Two Dialogic DMV Media Boards, which provide 8 T1/E1 spans, were tested in a machine with a Dual Xeon, 3.0GHz CPU and with the following additional conditions:

- All other components, such as, the Media Control Platform, Resource Manager, and SIP Server were installed off board.
- Two different protocols were used—ISDN and CAS/RB Wink.
- Two application profiles were used—VoiceXML_App1 and VoiceXML_App3.

The overall CPU idle was approximately 80% for both applications.

CTI Connector

CTI Connector performance is tested in two scenarios, in which a Play Treatment application is used with two different transfer types (a bridge transfer and a blind transfer). The Media Control Platforms are configured to use NGI and GVPi, respectively.

Two test cases, in which GVPi was used, produced 25 CAPS on Windows and Linux. A test case in which the NGI was used in a blind transfer scenario produced 25 CAPS on Windows and Linux, while a bridge transfer produced only 15 CAPS on all supported Windows versions and 20 CAPS on Linux.

Beyond peak capacity, the overall call-pass-rate dropped below the 99.95% criteria. The graph in Figure 129 is a sample of the CTI Connector overall CPU usage versus port density CAPS when NGI is used in a blind transfer scenario:

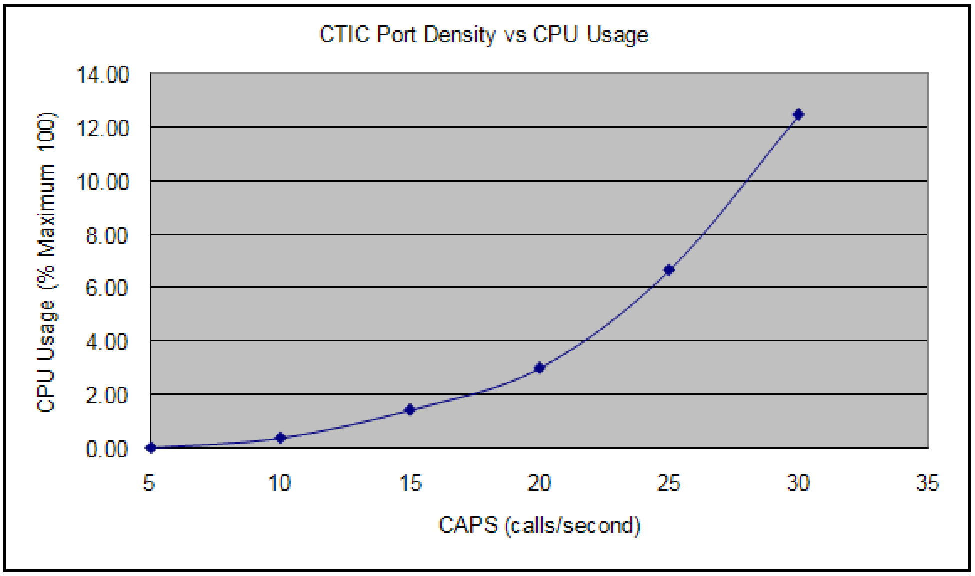


Figure 129: PD Versus CPU Usage (CTI Connector)

CTI Connector/ICM

The CTI Connector 8.1.4 is integrated with Cisco Intelligent Contact Management (ICM), enabling customers to choose between two Computer Telephony Integration (CTI) deployment modes—Genesys CTI or Cisco CTI.

In these test cases, CTI Connector/ICM performances measured in CAPS and testing was conducted by using two ICM modes of operation—Service Control Interface (SCI) and Call Routing Interface (CRI).

Two bridge transfer scenarios and a blind transfer scenario were tested with CED, Call, and ECC variables passing from the Media Control Platform to ICM. Multiple Media Control Platform instances configured to use NGI only, achieved the following results:

- One ICM configured in CRI mode achieved 22 CAPS on Windows and 18 CAPS on Linux with GVP 8.1.4 or earlier, and 22 CAPS with GVP 8.1.5.
- Two ICMs configured in SCI mode on both Windows and Linux achieved 30 CAPS.

Test results indicated a bottleneck on the ICM side of the platform. The graphs in Figures 130 and 131 depict a sample of the CPU usage when CTI Connector/ICM performs a blind transfer in SCI mode on Windows 2008 R2.

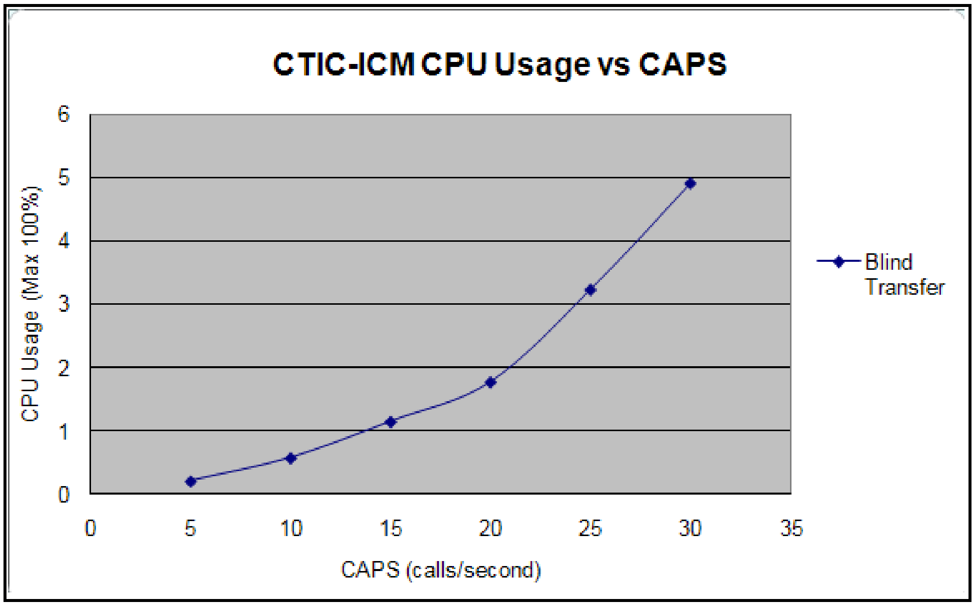


Figure 130: CPU Usage Versus CAPS—Blind Transfer (CTI Connector/ICM)

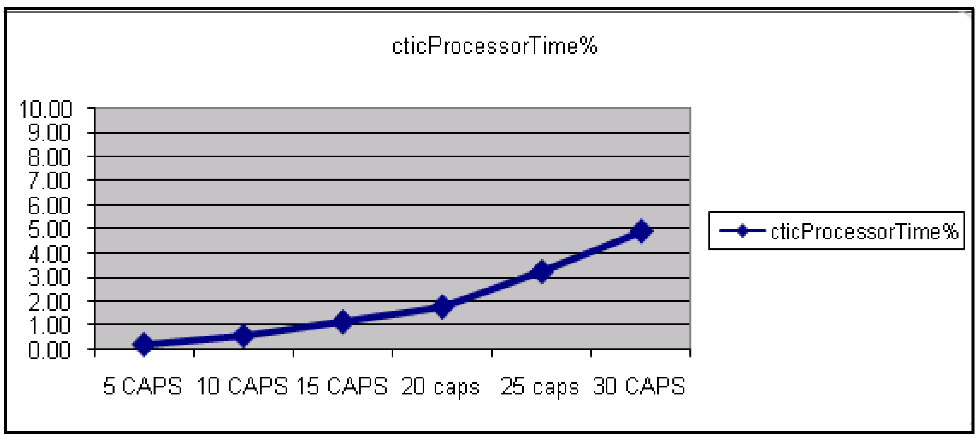


Figure 131: CPU Usage During Blind Transfer (CTI Connector/ICM)

Supplementary Services Gateway

The Supplementary Services Gateway makes outbound calls through SIP Server, therefore, the call rate (or CAPS) is used to measure the Supplementary Services Gateways performance. Figure 132 provides a comparison of launching call rates (HTTP Requests or targets) and notification of completed calls: (real or achieved CAPS).

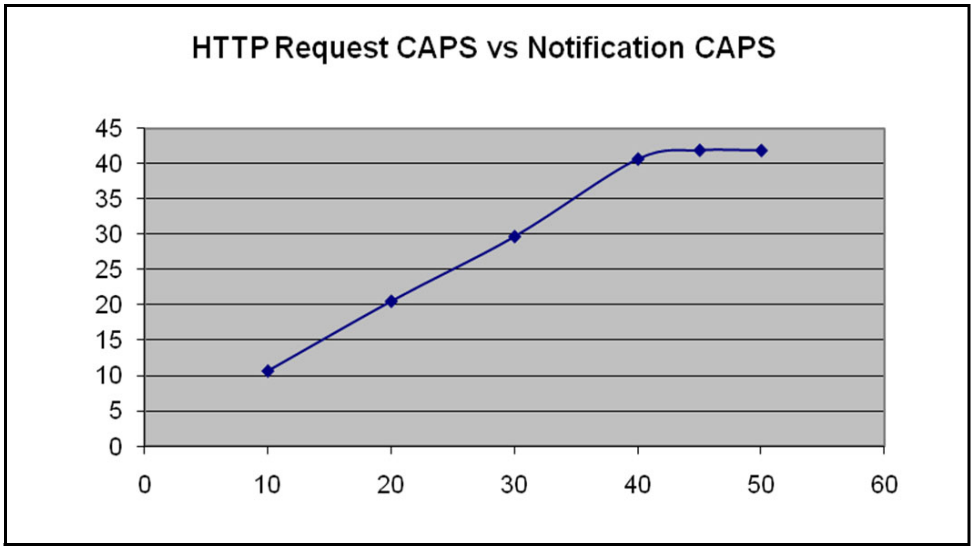


Figure 132: HTTP Request CAPS Versus Notification CAPS

The Supplementary Services Gateways peak capacity result is 65 CAPS (using GVP 8.1.5 with SIP Server 8.1.0), 50 CAPS (using GVP 8.1.3 or later and SIP server 8.0.4) and 40 CAPS (using pre-GVP 8.1.3 and pre-SIP Server 8.0.4). These results are due to a bottleneck on the SIP Server side of the network—multiple Media Control Platforms are used to provide a sufficient number of ports to handle VoiceXML applications, regardless of their complexity. The call rate can exceed SIP Servers peak capacity, but the requests (which are stored on the database server) tend to pile up. If egress rate is not high enough, the stored request records can easily reach the database limit of 100,000 records.

Reporting Server

Like the Resource Manager, Reporting Server performance is measured in terms of CAPS. The number of simultaneous calls being processed by GVP does not affect performance and there are no known performance bottlenecks with the Reporting Server software, however, performance can be affected by the database setup. When the Reporting Server is tested without the Media Control Platform in No DB mode (Reporting Server drop all call data), it can achieve 800 CAPS.

Capacity reached 800 CAPS when the Reporting Server was tested in No DB mode with the Resource Manager only (without the Media Control Platform), In No DB mode, the Reporting Server drops all received call data.

A use case was conducted on Microsoft SQL 2008 and Oracle 10g R2 Enterprise Servers, with the Resource Manager and the Media Control Platform streaming information (including CDR, upstream, and SQA data) to the Reporting Server for each call with a default log level. The result was a peak capacity of 270 CAPS.

The same use case was conducted on an Oracle 11g Enterprise Server only; the result was a peak capacity of 300 CAPS.

Note: For this test case, the Reporting Server was installed on a Dual Quad Core Xeon computer with a 2.66 GHz CPU, separate from the database server. The Microsoft SQL and Oracle Database Servers were installed on a 15-disk Disk Array computer with a Dual Quad Core Xeon, 2.66 GHz CPU. For practicality, a simulator was used instead of a real MCP (or a Reporting Server client inside MCP) to run the tests. Also, simulated data (five MCP metrics per call) was used for the MCP simulator to submit the data to the Reporting Server.

The Reporting Server can connect to other GVP components by using TCP or TLS.

Single Server Test Cases

The single-server performance testing that is described in this section is conducted on hardware that is slightly different from the suggested hardware requirements. The servers used for the performance test cases have the following hardware specifications: 1x Intel Xeon 5160, with a 3.0 GHz CPU, 8 GB of RAM, and a 73 GB SAS HD.

The following software components are installed:

- Windows 2008 Enterprise Server, SP2, x86 or Windows 2008 Enterprise Server R2, x64
- Microsoft SQL Server 2008 Standard version
- Microsoft Internet Information Server (IIS) configured as a Web Application Server (WAS)
- Management Framework 8.0.3 (Database Server, Configuration Server, Solution Control Server, Message Server)
- Genesys Voice Platform 8.1.3 or 8.1.4 (Resource Manager, Media Control Platform [Squid], Reporting Server)
- SNMP Master Agent 8.0.2
- Genesys Administrator 8.0.3
- SIP Server 8.0.4
- An ASR/TTS Server (Nuance Recognizer 9.0.12, RealSpeak 4.5, Nuance Speech Server 5.0.9)

The following test results indicate higher performance metrics than GVP 7.6 with 48 ports achieved:

- 600 ports - VoiceXML_App1 (DTMF)
- 100 ports - VoiceXML_App2 (ASR with MRCP v1)
- 160 ports - VoiceXML_App3 (AS ASR/TTS with MRCP v1)

- 120 ports - VoiceXML_App3 (AS ASR/TTS with MRCP v2)

The graph in Figure 133 depicts the trend of overall CPU usage versus ports density of the VoiceXML_App1 profile.

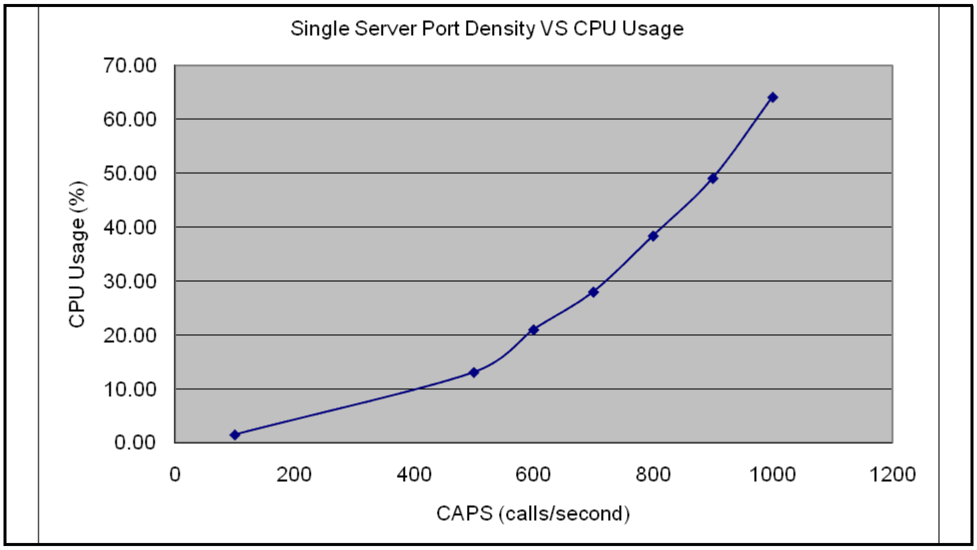


Figure 133: Port Density Versus CPU Usage (Single Server)

The graph in Figure 134 depicts the call setup latency versus concurrent calls in a Media Control Platform only configuration.

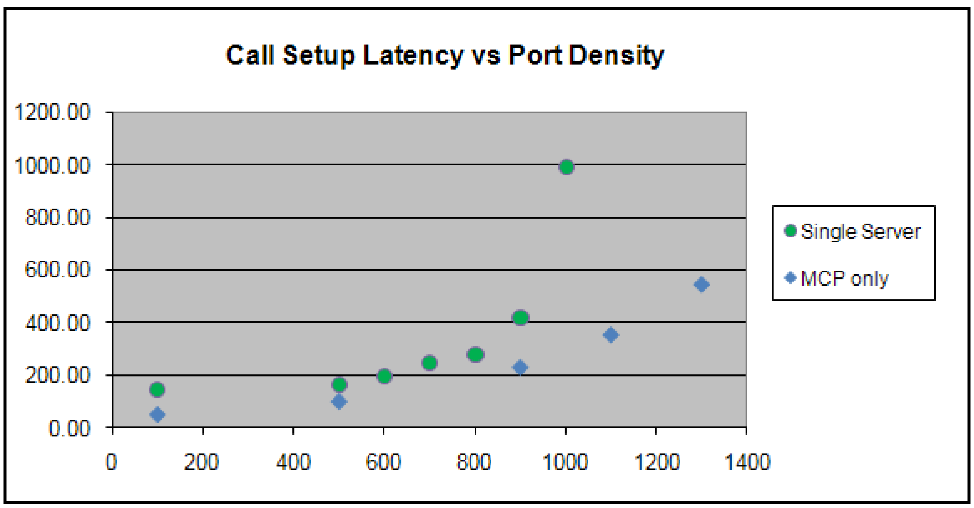


Figure 134: Call Setup Latency Versus Port Density (MCP only)

In this test case, the latency aligns with the Media Control Platform only configuration with fewer ports configured. Here, the latency is slightly than higher, because in a single-server configuration, the Resource Manager and SIP Server are configured before Media Control Platform. Notice that the latency jumps beyond the peak capacity after 800 ports.

Multiple MCP Instances and Virtual Machines Test Cases

The Media Control Platform only configuration is also measured with ESXi 4.1 Hypervisor VMs and multiple instances on the same host to compare performance metrics. The testing is conducted on a 2x Core 2Quad Xeon x5355, with a 2.66 GHz CPU, and 8 Cores, with 12 GB of RAM (higher than recommended).

First Series Testing of Multiple Media Control Platforms

The first series of performance tests were conducted on servers with 1, 2, 4, and 8 VM images installed respectively, and only one Media Control Platform on each VM with the following hardware configuration:

- 1 VM 8 virtual CPUs 12 GB RAM for the VM
- 2 VM 4 virtual CPUs 6 GB RAM for each VM
- 4 VM 2 virtual CPUs 3 GB RAM for each VM
- 8 VM 1 virtual CPUs 1.5 GB RAM for each VM

To provide comparisons, the operating system used for the VMs is Windows 2008 Enterprise SP2, x86, which is also the operating system that is installed on the host used to test multiple Media Control Platform instances (1, 2, 4, and 8, respectively).

The VoiceXML_App1 (DTMF) is used as the standard application profile. The graph in Figure 135 depicts the peak capacity that is obtained from the configurations that are described in the previous paragraphs.

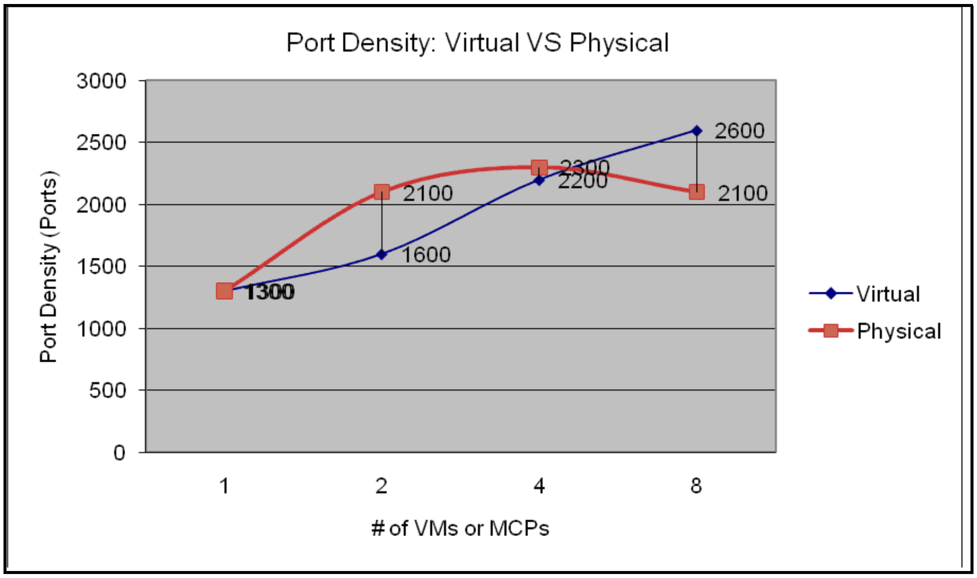


Figure 135: Port Density—Virtual Machines Versus Media Control Platforms

The graph in Figure 136 shows the highest peak capacity when 8 VMs (2600 ports) are configured, while the highest peak capacity when multiple Media Control Platform instances (2300 ports) are configured is at four instances (peak capacity is actually lower when eight Media Control Platform instances are configured). You can utilize the greatest number of ports when the number of VMs corresponds to the number of CPUs (Cores). However, using the multiple VM configurations results in a higher percentage of CPU usage. Figure 136 depicts the number ports when CPU usage is measured during testing.

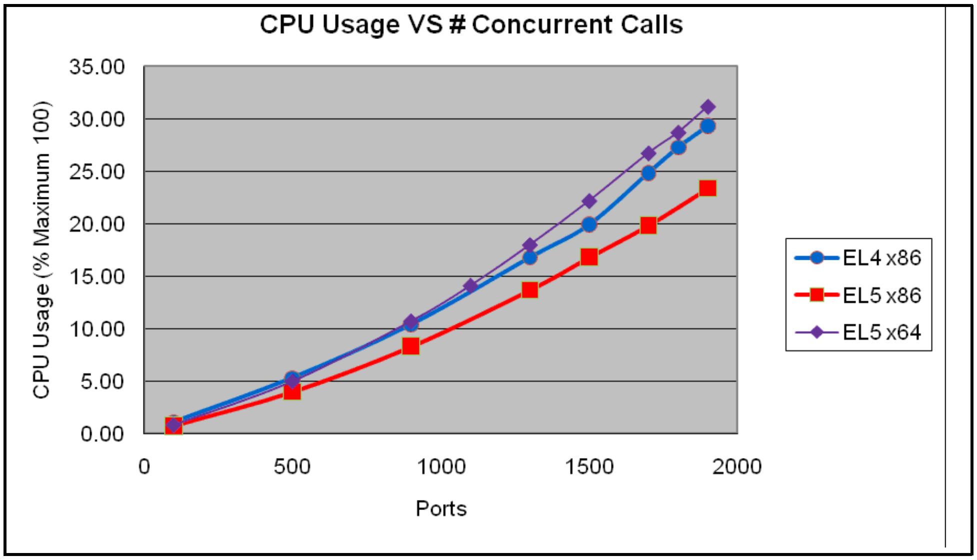


Figure 136: CPU Usage—Virtual Machines Versus Media Control Platforms

The graphs in Figures 137 and 138 provide a comparison of the call setup latency when multiple VMs are configured versus multiple Media Control Platform instances.

The latency is lower when more VMs or more Media Control Platform instances at same port density because more VMs or more Media Control Platforms, less calls distributed to each VM or Media Control Platform instance.

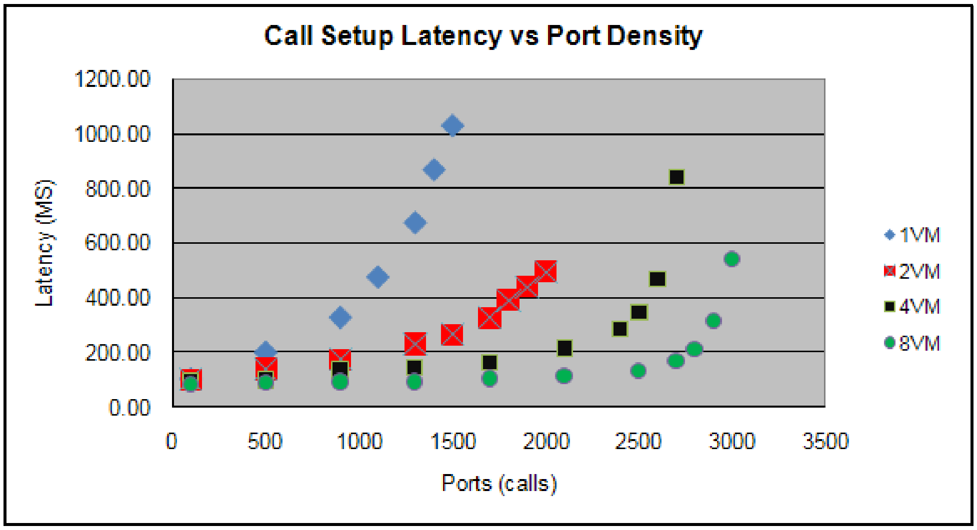


Figure 137: Call Setup Latency Versus Port Density (Virtual Machines)

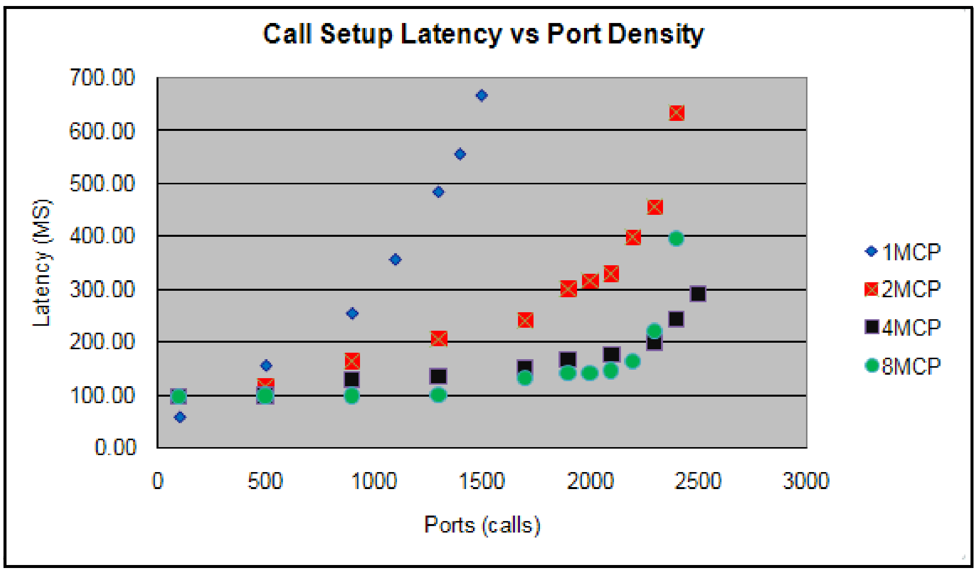


Figure 138: Call Setup Latency Versus Port Density (Actual MCPs)

The graphs in Figures 139 to 142 provide a comparison between multiple VMs and multiple Media Control Platform instances. The graphs depict 1-to-1, 2-to-2, 4-to-4, and 8-to-8 comparisons, respectively.

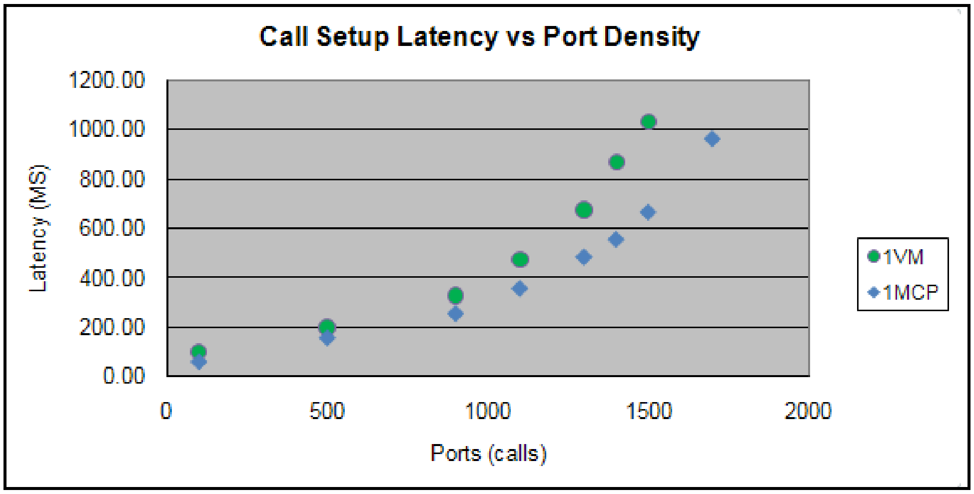


Figure 139: Call Setup Latency Versus Port Density (1 VM-to-1 MCP)

In Figure 139 the same peak ports is achieved when comparing 1 Media Control Platform instance to 1 VM, however, the Media Control Platform produced lower call setup latency.

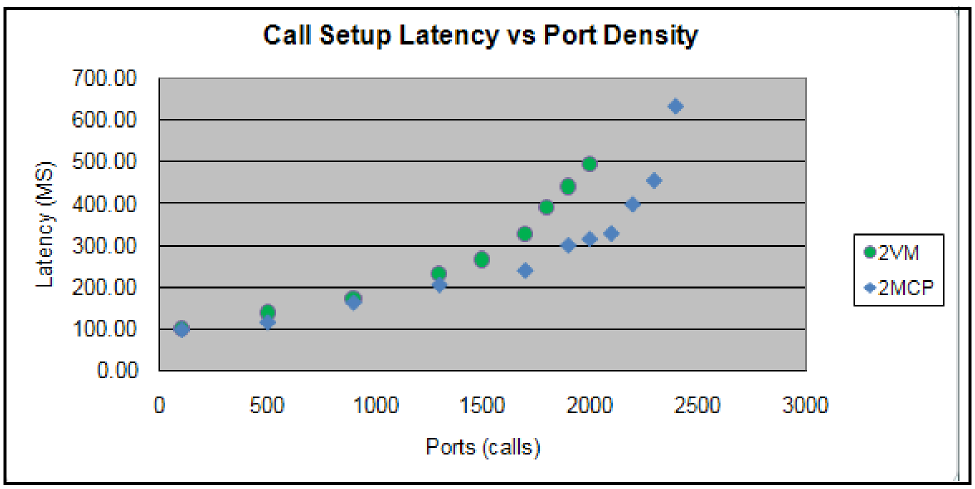


Figure 140: Call Setup Latency Versus Port Density (2 VM-to-2 MCP)

The graph in Figure 140 indicates that 2 Media Control Platform instances perform better than 2 VMs with lower latency and higher peak ports.

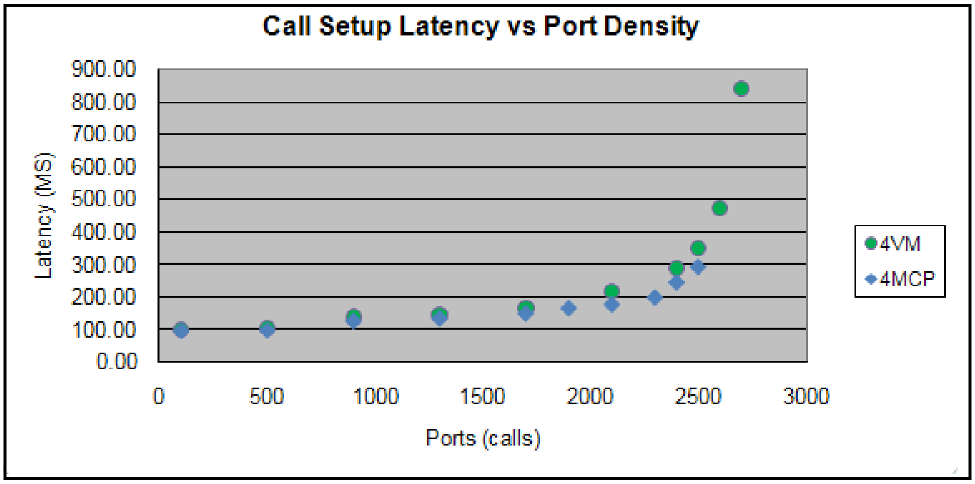


Figure 141: Call Setup Latency Versus Port Density (4 VM-to-4 MCP)

The graph in Figure 141 indicates that the performance of the 4 VMs and 4 Media Control Platform instances are quite close. The peak capacity and the latency trends are almost the same.

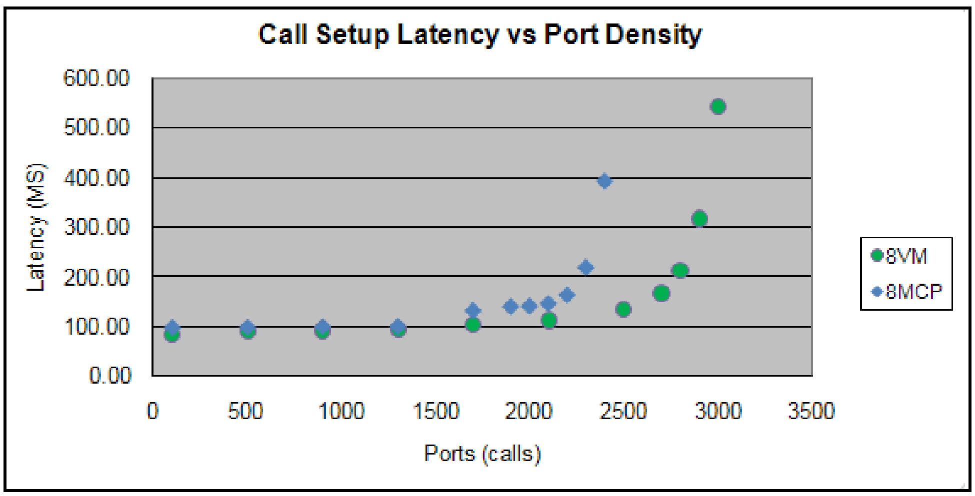


Figure 142: Call Setup Latency Versus Port Density (8 VM-to-8 MCP)

The graph in Figure 142 indicates that 8 VMs perform better than 8 Media Control Platform instances. The VMs produce lower latency and higher peak ports.

Second Series Testing of Multiple Media Control Platforms

The second series of performance tests were conducted on servers with 1 VM with 1 Media Control Platform instance using 1 CPU and then, the same configuration using 2 CPUs. These tests were executed by using 8 VMs and a

325-port load as the baseline, which is the highest peak capacity we could attain.

As shown in Table 110, the test conducted by using 2 CPUs (per VM) resulted in slightly higher CPU usage than the test with 1 CPU, while both results were only about 1/8 of the overall CPU usage when 8 VMs were configured.

Table 110: CPU Usage—1 Media Control Platformper VM

VMs	MCPs per VM	CPUs per VM	Ports	Overall CPU Usage
8	1 (8 total)	1 (8 total)	2600	83.69%
1	1 (1 total)	1 (1 total)	325	9.45%
1	1 (1 total)	2 (2 total)	325	9.76%

Third Series Testing of Multiple Media Control Platforms

The third series of performance tests were conducted on servers with 1 VM with 2 Media Control Platform instances and then, the same configuration with 2 VMs and 4 VMs, respectively. See Table 111.

Table 111: CPU Usage—2 Media Control Platforms per VM

VMs	MCPs per VM	CPUs per VM	Ports	Overall CPU Usage
2	1 (2 total)	4 (8 total)	1600	69.14%
2	2 (4 total)	4 (8 total)	1600	69.22%
4	1 (4 total)	2 (8 total)	2200	80.10%
4	2 (8 total)	2 (8 total)	2200	77.94%

Test results do not indicate higher ports as capacity peaks, because CPU usage is already high, however, the call setup time gets shorter. See Figures 143 and 144.

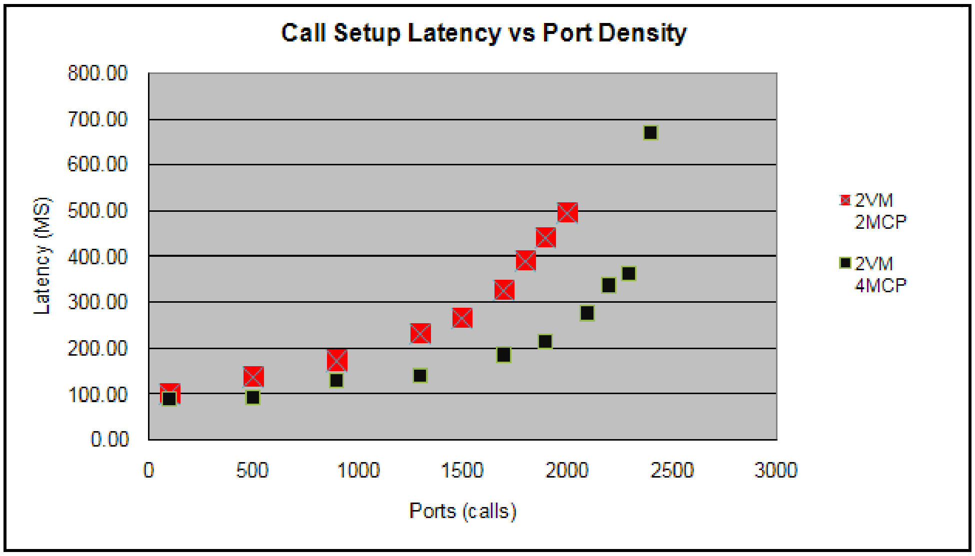


Figure 143: Call Setup Latency Versus Port Density—2 VMs

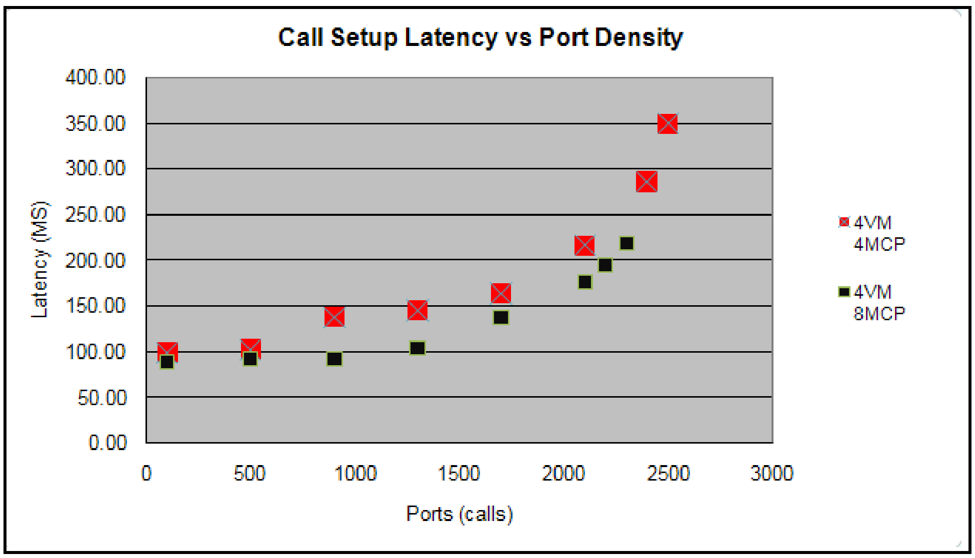


Figure 144: Call Setup Latency Versus Port Density—4 VMs

Throughout testing results indicated somewhat shorter call durations, which is indicated in Figures 145 and 146 showing the 95 percentile of call duration.

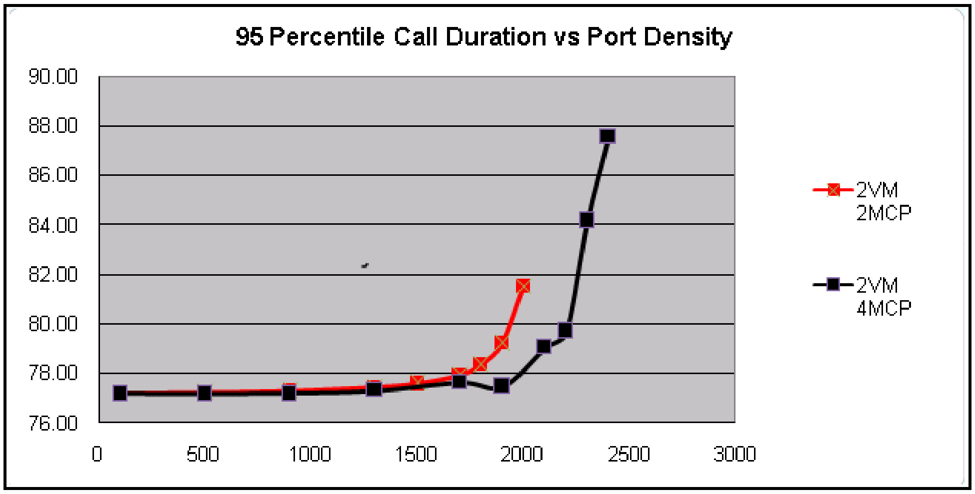


Figure 145: Call Duration Versus Port Density—2 VMs

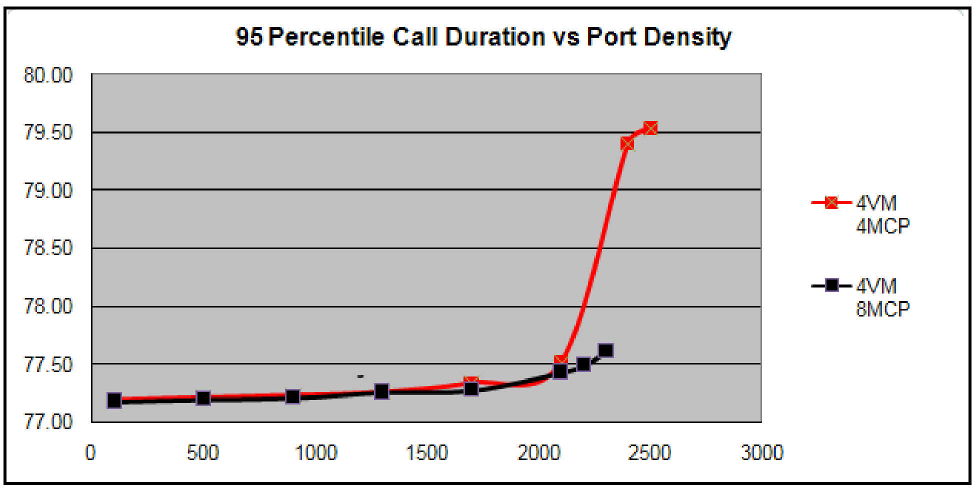
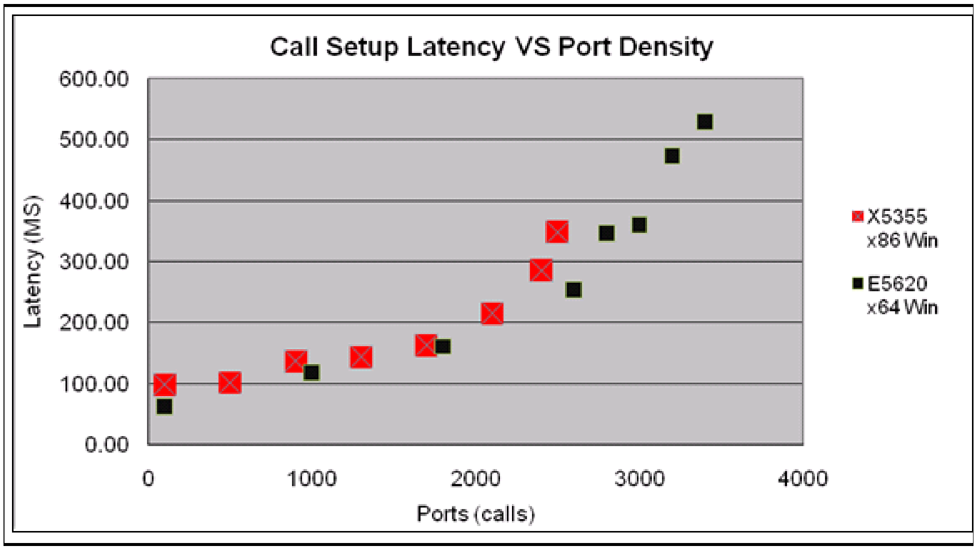


Figure 146: Call Duration Versus Port Density—4 VMs

The same test was repeated for GVP 8.1.5 on Windows 2008 Server R2 x64, with the latest hardware, which obviously achieves higher peak capacity. From the graph below, you can see that the latencies of x64 Win and x86 Win systems are quite in line with one another at lower ports, but a marked difference appeared at higher ports around peak capacity.



Jitter Quality on Virtual and Actual Media Control Platform Machines

Two metrics are used to measure jitter quality—Jitter Average (the weighted average of a stream's packets) and Jitter Max (the maximum number streamed packets). Two VMs and 2 Media Control Platform instances were used to test jitter quality. As expected, the results revealed some differences between the virtual and actual machines: See Figures 147 and 148.

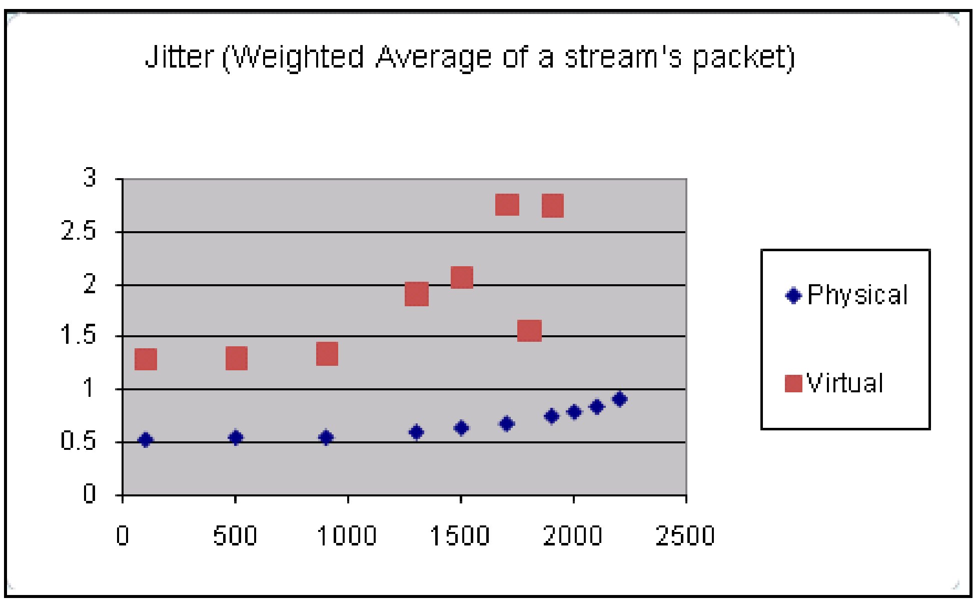


Figure 147: Jitter (Weighted Average)

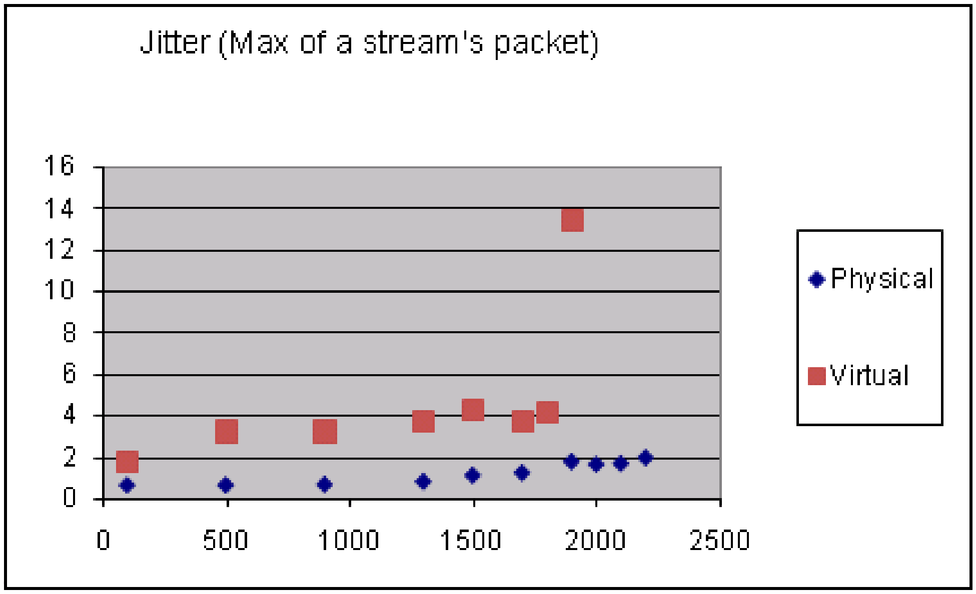


Figure 148: Jitter (Maximum of Streamed Packets)

From the perspective of media latency, the difference is minor (less than or equal to 5%). In Figure 149, speech response latency was tested with 1000 words of grammar.

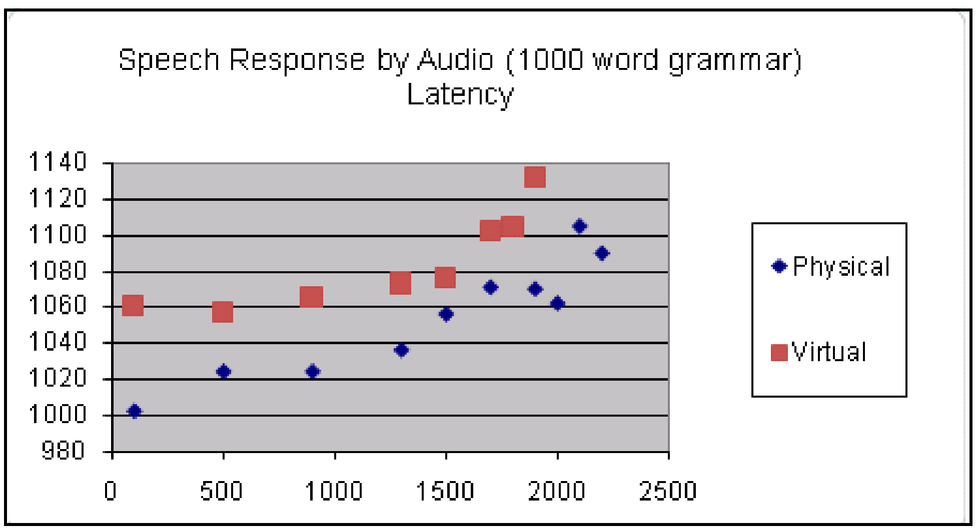


Figure 149: Speech Resource by Audio Latency

In Figure 150, when speech bargein is compared to TTS latency the difference is between 20 and 50 milliseconds.

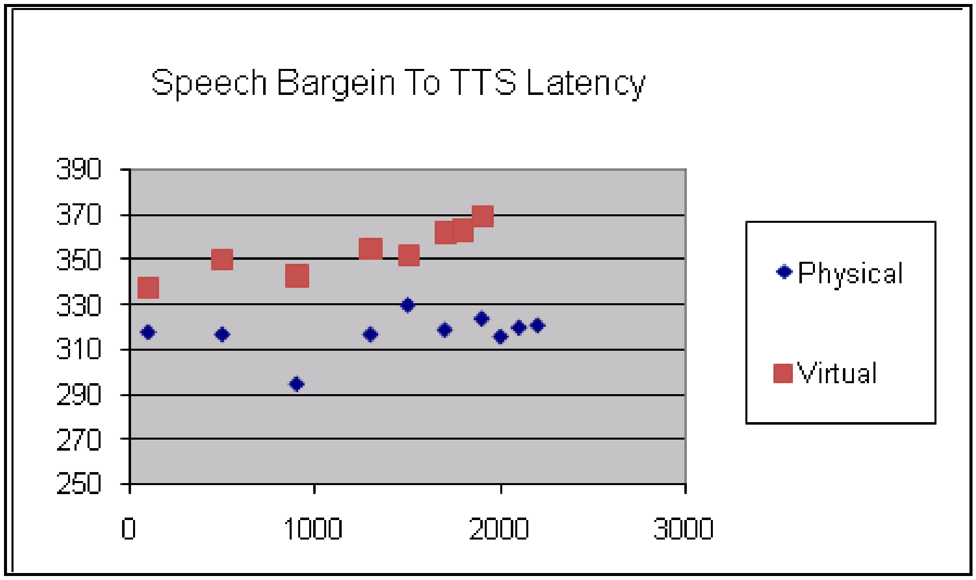


Figure 150: Speech BargeIn to TTS Latency

Performance and Scalability Comparisons

GVP 8.x contains many performance enhancements that are not included in GVP 7.6 or VoiceGenie (VG) 7.2.

Performance Comparisons

In this section, performance and scalability comparisons are made between GVP 8.x and previous releases by using the application profiles in Table 98 on page 246.

Tested with VoiceXML_App1

Peak capacity of GVP 8.x:

- with NGI:
  - ~50% higher than VG 7.2
  - ~90% higher than GVP 7.6
- with GVPi, equivalent to GVP 7.6

Comparing GVP 8.x and GVP 7.6 (with GVPi) to GVP 8.1:

- 8.1 uses significantly fewer CPU cycles (relatively 30%)
- 8.1 uses less memory (relatively 30%)

In 8.1, the peak capacity is identical to previous releases (using identical temp file management mechanisms), as the bottleneck is due to disk I/O.

Tested with VoiceXML_App2

Peak capacity of GVP 8.x:

- with NGI:
  - ~66% higher than VG 7.2
  - ~100% higher than GVP 7.6
- with GVPi, equivalent to GVP 7.6

In the use case with GVPi, the peak capacity for GVP 8.x is identical to GVP 7.6 (using identical temp file management mechanisms), because the bottleneck is due to disk I/O.

Scalability Comparisons

For applications that are CPU-dependent (or applications in which bottlenecks occur due to CPU cycles) GVP 8.x can use additional CPU cycles and cores. Use case results showed that peak port densities scaled upward linearly relative to an increase in CPU clock speed.

Table 112 contains examples of peak capacity when VoiceXML_App1 is used:

Table 112: Peak Capacity—VoiceXML_App1

Processor	Total Clock Speed	Peak Port Density
2x Core 2 Quad, 2.66 GHz	21.28 GHz	1300
2x Core 2 Dual, 3.00 GHz	12 GHz	700
1x Core 2 Dual, 3.00 GHz	6 GHz	400

Figure 151 is a graphical depiction of the peak port density in Table 116 on page 360.

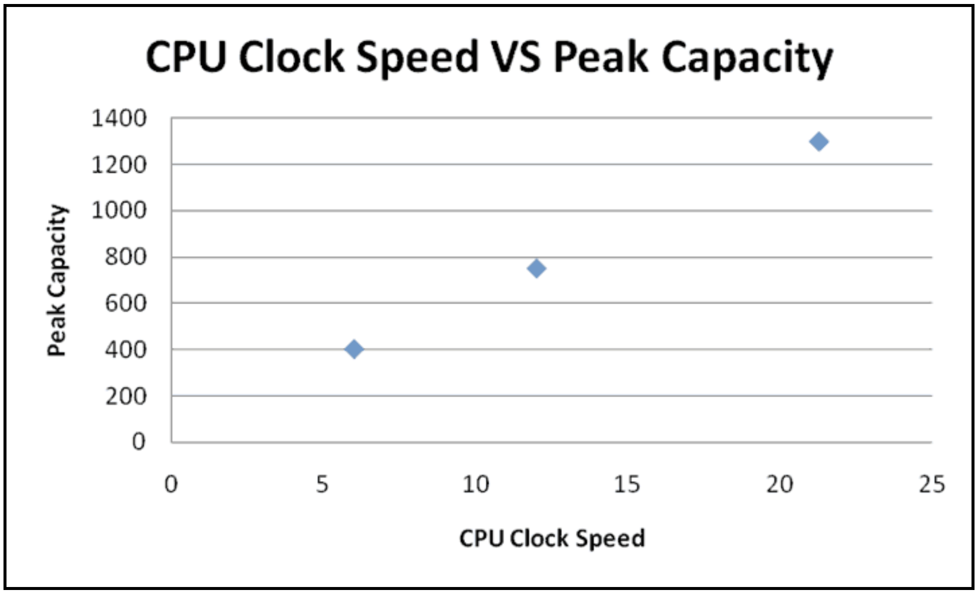


Figure 151: CPU Clock Speed Versus Peak Capacity

To increase the total clock speed by 100%, the peak capacity would have to increase by ~90 to 100%, assuming:

- The type of CPUs are the same as the ones in Table 116 on page 360.
- The VoiceXML_App1 application is used.
- The overall system bottleneck CPU cycles remain the same.

High Performance Configuration

The Media Control Platform can support more than 400 ports on a single host, however, some configuration changes are required. Use Genesys Administrator to configure the Media Control Platform for high performance by modifying the options and default values in Table 113, and configure the Windows Registry on the Media Control Platform to support either the NGI, GVPi, or both.

Table 113: High Performance Configuration for Media Control Platform

Section	Option/Key	Default Value	High Performance Value
Media Control Platform with NGI			
mpc	maxmediathreads	32	16
vxmli	max_num_documents	5000	10,000 (> 1000 ports)

Table 113: High Performance Configuration for Media Control Platform (Continued)

Section	Option/Key	Default Value	High Performance Value
Windows Registry key: HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\Tcpip\Parameters\ TCPTimedWaitDelay		None	Type = DWORD Value = 30 or 1e (hex)
Media Control Platform with GVPi			
mpc	maxmediathreads	32	32
PageCollector	maxpoolthreads	512	>= Port Density
PopGateway1	maxpoolthreads	512	>= Port Density
Windows Registry key: HKEY_LOCAL_MACHINE\Software\CallNet\ CnlnetSettings\MaxThreadPool		None	Type = DWORD Value >= Port Density /2

Hardware and Bandwidth Usage

This section contains hardware (or disk space) usage and bandwidth estimates for the Reporting Server, and bandwidth usage estimates for the Media and Call Control Platforms.

Reporting Server Hardware Usage

The amount of disk space that is required for the Reporting Server depends on many factors:

- Retention period
- Call rate
- Number of IVR Profiles, Tenants, and DNs

Table 114 on page 350 provides the information necessary to estimate the disk space that is required for the Reporting Server data types.

For more information about data retention, and data types, see the *Genesys Voice Platform 8.1 User's Guide*.

Table 114: Reporting Server Disk Space Estimates

Data type	Usage	Estimated disk storage in bytes	Required estimates	Retention periods
Resource Manager				
CDR	Very High	600	Number of calls per day	retention.cdr
Calculation: 600 * number of calls per day * retention.cdr				
Operational Reporting (5 minutes)	Medium	300	Number of: • DNs • IVR Profiles • Tenants • RM, CTIC, PSTNC	retention.operations.5min
Calculation: 300 * (number of DNs + number of IVR Profiles + number of tenants + number of CTIC, PSTNC +1) * (number of RMs) * 2 * 1440 * retention.operations.5min				
Operational Reporting (30 minutes)	Medium	300	Number of: • DNs • IVR Profiles • Tenants • RM, CTIC, PSTNC	retention.operations.30min
Calculation: 300 * (number of DNs + number of IVR Profiles + number of tenants + number of CTIC, PSTNC +1) * (number of RMs) * 2 * 48 * retention.operations.30min				
Resource Manager				
Operational Reporting (hourly)	Medium	300	Number of: • DNs • IVR Profiles • Tenants • RM, CTIC, PSTNC	retention.operations.hourly
Calculation: 300 * (number of DNs + number of IVR Profiles + number of tenants + number of CTIC, PSTNC +1) * (number of RMs) * 2 * 24 * retention.operations.hourly				

Table 114: Reporting Server Disk Space Estimates (Continued)

Data type	Usage	Estimated disk storage in bytes	Required estimates	Retention periods
Operational Reporting (daily)	Medium	300	Number of: • DNs • IVR Profiles • Tenants • RM, CTIC, PSTNC	retention.operations.daily
Calculation: 300 * (number of DNs + number of IVR Profiles + number of tenants + number of CTIC, PSTNC +1) * (number of RMs) * 2 * retention.operations.daily				
Operational Reporting (weekly)	Medium	300	Number of: • DNs • IVR Profiles • Tenants • RM, CTIC, PSTNC	retention.operations.weekly
Calculation: 300 * (number of DNs + number of IVR Profiles + number of tenants + number of CTIC, PSTNC +1) * (number of RMs) * 2 * retention.operations.weekly/7				
Operational Reporting (monthly)	Medium	300	Number of: • DNs • IVR Profiles • Tenants • RM, CTIC, PSTNC	retention.operations.monthly
Calculation: 300 * (number of DNs + number of IVR Profiles + number of tenants + number of CTIC, PSTNC +1) * (number of RMs) * 2 * retention.operations.monthly/30				
Media Control Platform				
CDR	Very High	600	Number of calls per day	retention.cdr
Calculation: 600 * number of calls per day * retention.cdr				
Operational Reporting (5 minutes)	Medium	300	Number of: • IVR Profiles • MCPs	retention.operations.5min

Table 114: Reporting Server Disk Space Estimates (Continued)

Data type	Usage	Estimated disk storage in bytes	Required estimates	Retention periods
Calculation: 300 * (number of IVR Profiles + 1) * (number of MCPs) * 1440 * retention.operations.5min + 100 * (number of MCPs) * 1440 * retention.operations.5min **Note:** The first product is for the arrivals that are stored per IVR Profile for each MCP. The second product is for the peaks that are stored for each MCP.				
Operational Reporting (30 minutes)	Medium	300	Number of: • IVR Profiles • MCPs	retention.operations.30min
Calculation: 300 * (number of IVR Profiles + 1) * (number of MCPs) * 48 * retention.operations.30min + 300 * (number of MCPs) * 48 * retention.operations.30min				
Media Control Platform				
Operational Reporting (hourly)	Medium	300	Number of: • IVR Profiles • MCPs	retention.operations.hourly
Calculation: 300 * (number of IVR Profiles + 1) * (number of MCPs) * 24 * retention.operations.hourly + 300 * (number of MCPs) * 24 * retention.operations.hourly				
Operational Reporting (daily)	Medium	300	Number of: • IVR Profiles • MCPs	retention.operations.daily
Calculation: 300 * (number of IVR Profiles + 1) * (number of MCPs) * retention.operations.daily + 300 * (number of MCPs) * retention.operations.daily				
Operational Reporting (weekly)	Medium	300	Number of: • IVR Profiles • MCPs	retention.operations.weekly
Calculation: 300 * (number of IVR Profiles + 1) * (number of MCPs) * retention.operations.weekly/7 + 300 * (number of MCPs) * retention.operations.weekly/7				
Operational Reporting (monthly)	Medium	300	Number of: • IVR Profiles • MCPs	retention.operations.monthly
Calculation: 300 * (number of IVR Profiles + 1) * (number of MCPs) * retention.operations.monthly/30 + 300 * (number of MCPs) * retention.operations.monthly/30				

Table 114: Reporting Server Disk Space Estimates (Continued)

Data type	Usage	Estimated disk storage in bytes	Required estimates	Retention periods
Events	Very High	500	Number of: • events per call • calls per day	retention.events
Calculation: 500 * number of events per call * number of calls per day * retention.events				
VAR CDR	Very High	200 per VAR CDR 150 per VAR custom variable	Number of: • calls per day • custom variables per call	retention.cdr
Calculation: (200 +150 * number of custom variables per call) * number of calls per day * retention.cdr				
Media Control Platform				
VAR Summary (5 minutes)	Medium	300	Number of: • IVR Profiles • Tenants • MCPs • IVR Actions • unique call-end reasons	retention.var.5min
Calculation: 300 * (number of IVR Profile + number of tenants) * number of MCPs * (number of IVR Actions +1) * number of unique call-end reasons * 1440 * retention.var.5min				
VAR Summary (30 minutes)	Medium	300	Number of: • IVR Profiles • Tenants • MCPs • IVR Actions • unique call-end reasons	retention.var.30min
Calculation: 300 * (number of IVR Profile + number of tenants) * number of MCPs * (number of IVR Actions +1) * number of unique call-end reasons * 48 * retention.var.30min				

Table 114: Reporting Server Disk Space Estimates (Continued)

Data type	Usage	Estimated disk storage in bytes	Required estimates	Retention periods
VAR Summary (hourly)	Medium	300	Number of: • IVR Profiles • Tenants • MCPs • IVR Actions • unique call-end reasons	retention.var.hourly
Calculation: 300 * (number of IVR Profile + number of tenants) * number of MCPs * (number of IVR Actions +1) * number of unique call-end reasons * 24 * retention.var.hourly				
Media Control Platform				
VAR Summary (daily)	Medium	300	Number of: • IVR Profiles • Tenants • MCPs • IVR Actions • unique call-end reasons	retention.var.daily
Calculation: 300 * (number of IVR Profile + number of tenants) * number of MCPs * (number of IVR Actions +1) * number of unique call-end reasons * retention.var.hourly				
VAR Summary (weekly)	Medium	300	Number of: • IVR Profiles • Tenants • MCPs • IVR Actions • unique call-end reasons	retention.var.weekly
Calculation: 300 * (number of IVR Profile + number of tenants) * number of MCPs * (number of IVR Actions +1) * number of unique call-end reasons * retention.var.weekly/7				

Table 114: Reporting Server Disk Space Estimates (Continued)

Data type	Usage	Estimated disk storage in bytes	Required estimates	Retention periods
VAR Summary (monthly)	Medium	300	Number of: • IVR Profiles • Tenants • MCPs • IVR Actions • unique call-end reasons	retention.var. monthly
Calculation: 300 * (number of IVR Profile + number of tenants) * number of MCPs * (number of IVR Actions +1) * number of unique call-end reasons * retention.var.monthly/30				
SQA Latency (hourly)	Medium	600	Number of components	retention.latency. hourly
Calculation: 600 * (number of components) * retention.latency.hourly * 24				
Media Control Platform				
SQA Latency (daily)	Medium	600	Number of components	retention.latency. daily
Calculation: 600 * (number of components) * retention.latency.daily				
SQA Latency (weekly)	Medium	600	Number of components	retention.latency. weekly
Calculation: 600 * (number of components) * retention.latency.weekly/7				
SQA Latency (monthly)	Medium	600	Number of components	retention.latency. monthly
Calculation: 600 * (number of components) * retention.latency.monthly/30				
SQA Failure Details	Medium	500	Number of calls per day Failure rate percentage	retention.sq. failures
Calculation: 500 * number of calls per day * failure rate percentage * retention.sq.failures				
SQA Failure Summary (hourly)	Medium	200	Number of: • MCPs • IVR Profiles	retention.sq.hourly

Table 114: Reporting Server Disk Space Estimates (Continued)

Data type	Usage	Estimated disk storage in bytes	Required estimates	Retention periods
Calculation: 200 * number of MCPs * number of IVR Profiles * retention.sq.hourly * 24				
SQA Failure Summary (daily)	Medium	200	Number of: • MCPs • IVR Profiles	retention.sq.daily
Calculation: 200 * number of MCPs * number of IVR Profiles * retention.sq.daily				
SQA Failure Summary (weekly)	Medium	200	Number of: • MCPs • IVR Profiles	retention.sq.weekly
Calculation: 200 * number of MCPs * number of IVR Profiles * retention.sq.weekly/7				
Media Control Platform				
SQA Failure Summary (monthly)	Medium	200	Number of: • MCPs • IVR Profiles	retention.sq. monthly
Calculation: 200 * number of MCPs * number of IVR Profiles * retention.sq. monthly/30				
Call Control Platform				
CDR	Very High	600	Number of calls per day	retention.cdr
Calculation: 600 * number of calls per day * retention.cdr				
Operational Reporting (5 minutes)	Medium	300	Number of: • CCPs • IVR Profiles	retention.operations. 5min
Calculation: 300 * (number of IVR Profiles +1) * number of CCPs * 1440 * retention.operations.5min + 300 * number of CCPs * 1440 * retention.operations.5min **Note:** The first product is for the arrivals that are stored per IVR Profile for each CCP. The second product is for the peaks that are stored for each CCP.				
Operational Reporting (30 minutes)	Medium	300	Number of: • CCPs • IVR Profiles	retention.operations. 30min

Table 114: Reporting Server Disk Space Estimates (Continued)

Data type	Usage	Estimated disk storage in bytes	Required estimates	Retention periods
Calculation: 300 * (number of IVR Profiles +1) * number of CCPs * 48 * retention.operations.30min + 300 * number of CCPs * 48 * retention.operations.30min				
Operational Reporting (hourly)	Medium	300	Number of: • CCPs • IVR Profiles	retention.operations.hourly
Calculation: 300 * (number of IVR Profiles +1) * number of CCPs * 24 * retention.operations.hourly + 300 * number of CCPs * 24 * retention.operations.hourly				
Call Control Platform				
Operational Reporting (daily)	Medium	300	Number of: • CCPs • IVR Profiles	retention.operations.daily
Calculation: 300 * (number of IVR Profiles +1) * number of CCPs * retention.operations.daily + 300 * number of CCPs * retention.operations.hourly				
Operational Reporting (weekly)	Medium	300	Number of: • CCPs • IVR Profiles	retention.operations.weekly
Calculation: 300 * (number of IVR Profiles +1) * number of CCPs * retention.operations.weekly / 7+ 300 * number of CCPs * retention.operations.weekly / 7				
Operational Reporting (monthly)	Medium	300	Number of: • CCPs • IVR Profiles	retention.operations.monthly
Calculation: 300 * (number of IVR Profiles +1) * number of CCPs * retention.operations.monthly / 30 + 300 * number of CCPs * retention.operations.monthly / 30				
Events	Very High	500	Number of: • events per call • calls per day	retention.events
Calculation: 500 * number of events per call * number of calls per day * retention.cdr				

Bandwidth Usage

The following tables describe the bandwidth usage for the following components:

- Media Control Platform (Table 115)
- Call Control Platform (Table 116 on page 360)
- Reporting Server (Table 117 on page 360)

Table 115 describes the bandwidth usage when bi-directional traffic exists between the Media Control Platform and other servers.

Table 115: Media Control Platform Bandwidth Usage

Protocol	Estimated bi-directional traffic	Criticality	Comments
Between Media Control Platform and SIP components			
SIP	Simple inbound call: 5KB per call Outbound with Supplementary Services Gateway: 10KB per call	Very High	SIP traffic can vary, depending on the call flow, the amount of user data, and number of treatments applied to the call.
Between Media Control Platform and MRCPv1			
RTSP MRCP RTP	ASR: 8 KB per recognition, and 8 KB/sec of RTP traffic TTS: 3 KB per prompt, and 8 KB/sec of RTP traffic	Very high	RTP traffic is uni-directional only.
Between Media Control Platform and MRCPv2			
SIP MRCP RTP	ASR: 15 KB per recognition, and 10 KB/sec of RTP traffic TTS: 6 KB per prompt, and 8 K/sec of RTP traffic	Very high	RTP traffic is uni-directional only.

Table 115: Media Control Platform Bandwidth Usage (Continued)

Protocol	Estimated bi-directional traffic	Criticality	Comments
Between Media Control Platform and RTP components			
RTP	PCMU/PCMU/G.722: 20 KB/sec per call leg G.729: 6 KB/sec per call leg G.729d: 5.6 KB/sec per call leg G.729e: 7 KB/sec per call leg G.729-16: 8 KB/sec per call leg G.726-24: 10 KB/sec per call leg G.726-32: 12 KB/sec per call leg G.726-40: 14 KB/sec per call leg GSM: 7.3 KB/sec per call leg AMR: 2-7.3 KB/sec per call leg AMR-WB: 5-10 KB/sec per call leg (the rate varies, depending on the audio data) H.263/H.264-1998: 10-70 KB/sec per call leg (the rate varies, depending on video data) H.264: 20-90 KB/sec per call leg (the rate varies, depending on video data)	Very high	Examples of RTP components are: • RTSP software • Soft phone • Media gateway
Between Media Control Platform and HTTP Server/Proxy Server			
HTTP	1 KB per request and content size of the VoiceXML page or audio file in the HTTP request and response.	Very high	HTTP traffic can vary, based on the number of files that are used by the VoiceXML application, the maxage and maxstale settings of the VoiceXML application, and the expiry settings on the HTTP server.

Table 116 describes the bandwidth usage when bi-directional traffic exists between the Call Control Platform and other servers.

Table 116: Call Control Platform Bandwidth Usage

Protocol	Estimated bi-directional traffic	Criticality	Comments
Between Call Control Platform and SIP components			
SIP	Simple inbound call without join: ~7 KB per session Inbound call starting a simple dialog: ~20 KB per session	Very high	Significantly dependent on call flow and network conditions. If the network connection is poor, messages could be resent according to the SIP protocol.
Between Call Control Platform and HTTP Server/Proxy Server			
HTTP	1 KB per request and content size of the CCXML page in the HTTP request and response.	Very high	HTTP traffic can vary, based on the number of files that are used by the CCXML application, the maxage and maxstale settings of the CCXML application, and the expiry settings on the HTTP server.

For information about bandwidth usage for the Management Framework components, see the Management Framework chapter in this guide.

Table 117 on page 360 describes the bandwidth usage when bi-directional traffic exists between the Reporting Server and other servers.

Table 117: Reporting Server Bandwidth Usage

Protocol	Estimated bi-directional traffic	Criticality	Comments
Between Reporting Server and Media Control Platform			
Proprietary (per call)	CDR: 1 KB per call Events: 1 KB per call	Very high	CDR: 2 updates per call, 400 bytes per update. Events: 10 events per call, 100 bytes per event. **Note:** The number of updates per call depends on the application that is used.
Proprietary (Operational Reporting)	OR: 100 bytes/min. OR: 100 bytes per IVR Profile per minute.	Low	One update per minute for peak (~50 bytes), and one update per minute for arrivals (~50 bytes).

Table 117: Reporting Server Bandwidth Usage (Continued)

Protocol	Estimated bi-directional traffic	Criticality	Comments
Proprietary (SQA)	SQA: 50 KB per 15 min. SQA: 3 KB per IVR Profile per minute	Low	This depends on the frequency at which the SQA is configured to send data upstream to the Reporting Server. The default is 15 minutes. If the deployment is configured differently, the estimate must be adjusted.
Between Reporting Server and Resource Manager			
Proprietary (per call)	CDR: 3 KB per call	Very high	CDR: 7 updates per call, 400 bytes per update. **Note:** The number of updates per call depends on the application that is used.
Proprietary (OR)	OR:100 bytes per IVR Profile per minute OR: 100 bytes per tenant per minute OR:100 bytes per DN per minute OR: 100 bytes per CTI Connector or PSTN Connector component per minute **Note:** These data usage results are only for the IVR Profile, Tenant, Component, and DN that are invoked during each 5-minute period.	Medium	Two updates per minute per IVR Profiles, 50 bytes per update. Two updates per minute per tenant, 50 bytes per update. Two updates per minute per CTI Connector/PSTN Connector component, 5 bytes per update. Two updates per minute per DN, 50 bytes per update.
Between Reporting Server and Call Control Platform			
Proprietary (per call)	CDR: 1 KB per call Events: 0.5 KB per call	Very high	CDR: 2 updates per call, 400 bytes per update. Events: 5 events per call, 100 bytes per event. **Note:** The number of updates per call depends on the application that is used.
Proprietary (OR)	OR: 100 bytes per minute OR: 100 bytes per IVR Profile per minute	Low	One update per minute for peak (~50 bytes), and one update per minute for arrivals (~50 bytes).

Table 117: Reporting Server Bandwidth Usage (Continued)

Protocol	Estimated bi-directional traffic	Criticality	Comments
Between Reporting Server and an Off-board Reporting Database			
Proprietary (database vendor)	The sum of all estimates between the Reporting Server and all the Media Control Platform, Call Control Platform, and Resource Manager servers.	Very high	This bandwidth estimate applies when the database is off-board only.

Chapter

12 Genesys SIP Server Solution

This chapter provides hardware sizing guidelines and the basic information required for deploying and capacity planning of a Voice over Internet Protocol (VOIP) solution based on SIP Server releases 7.6, 8.0, and 8.1. It also describes the factors that affect SIP Server performance, and lists sample performance measurements for reference platforms for both Windows and Linux. The chapter contains the following sections:

- SIP Server 8.1.1 and SIP Proxy 8.1.1: Sizing and Performance, page 363
- SIP Server 8.1.0: Sizing and Performance, page 365
- SIP Server 8.1.0: Reference Call Flows, page 382
- SIP Server 8.0: Sizing and Performance, page 389
- SIP Server 8.0: Reference Call Flows, page 402
- Genesys Media Server Sizing with SIP Server, page 408
- Genesys SIP Voicemail 8.1, page 416

SIP Server 8.1.1 and SIP Proxy 8.1.1: Sizing and Performance

Starting with release 8.1.1, SIP Server offers the Sizing Tool to evaluate SIP Server and SIP Proxy application CPU load and network traffic in your environment. It provides the following functionality:

- Handles input of customer deployment projected call activities.
- Allows selection of the software platform on which SIP Server and SIP Proxy applications will be running.
- Allows input of the number of SIP Server HA pairs and SIP Proxies in the deployment.

- Calculates CPU load of each application thread based on projected call activities and number of applications.
- Calculates separately the projected network bandwidth for SIP signaling and T-Library interface.

The Sizing Tool is available on the SIP Server Documentation web page as a Microsoft Excel spreadsheet. The embedded User's Guide provides the structure and workflow of the Tool.

SIP Server and SIP Proxy sizing is calculated simultaneously. If SIP Proxy is not deployed in your environment, ignore SIP Proxy sizing results. SIP Proxy does not affect SIP Server load.

Note: Use the Sizing Tool for SIP Server version 8.1.1 only. For previous versions of SIP Server, see "SIP Server 8.1.0: Sizing and Performance" on page 365 and "SIP Server 8.0: Sizing and Performance" on page 389 in this document.

Covered Scenarios

The Sizing Tool provides calculations for the following scenarios:

- Inbound and outbound (non-predictive) routed calls
- Single-step transfer
- Two-step transfer
- Call recording on a DN
- Single-step conference
- Two-step conference
- Consultation
- Internal calls using third-party call control (3pcc)
- Treatment on a Routing Point
- Predictive calls
- Predictive calls in ASM (Active Switching Matrix) mode
- ISCC transactions (with ISCC type direct-uui and route)

In scenarios involving media services, sizing is done using the MSML protocol, more efficient protocol comparing to NETANN used previously.

Platform and Log Verbosity

The Sizing Tool contains four similar sets for sizing coefficients, as follows:

- Windows Server 2008 (64-bit) and SIP Server with the application option verbose set to all.

- Windows Server 2008 (64-bit) and SIP Server with the application option verbose set to trace.
- Linux Red Hat Server (64-bit) and SIP Server with the application option verbose set to all.
- Linux Red Hat Server (64-bit) and SIP Server with the application option verbose set to trace.

SIP Server and SIP Proxy General Sizing Guidelines

SIP Server 8.1.1 sizing is calculated using the same basic guidelines as SIP Server 8.1.0:

- One SIP Server HA pair can handle up to 15000 simultaneous calls and the same number of T-Library connections. To determine the ability to handle a particular load, enter user requirements in the Sizing Tool.
- A single instance of SIP Proxy is usually loaded less than the SIP Server it serves. Use the Sizing Tool to determine the SIP Proxy load.

SIP Server 8.1.0: Sizing and Performance

This section provides hardware sizing guidelines and capacity planning of a VOIP solution based on SIP Server release 8.1. It covers the following topics:

- Solution Capacities
- Required hardware resources (the number of physical host computers)
- Number of applications (SIP Servers and Stream Managers)
- Recommended network architecture
- Application placement across the host computers
- Expected system loads (CPU, memory, and traffic per network interface)

Benchmarks are provided for the following basic scenarios:

- SIP inbound call via Routing Point to a DN
- Single-step transfer
- Two-step transfer
- Call recording on a DN
- Single-step conference
- Two-step conference
- Consultation call
- Internal call using third-party call control (3pcc)
- Treatment on a Routing Point or ACD Queue

Running a particular scenario involves multiple system components. For a reference system architecture, see “Reference Architecture” on page 367.

For the benchmark data obtained for the reference scenarios, see "Scenarios and Benchmarks" on page 370.

SIP Server performance is measured for variable call rates. Other important parameters are:

- scenario complexity
- number of concurrent calls
- number of open connections
- number of monitored clients
- size and update rate of attached data

The results presented in this section are based on lab benchmarks obtained on the operating systems, as described in Table 118. The SIP Server is benchmarked in multi-threaded mode (the sip-link-type configuration option is set to 3). SIP Server was configured in HA mode and had two monitoring T-Library clients (Stat Server and URS).

All performance data is represented for the most demanding threads and for the whole computer. In multi-threaded mode, two threads consume the most CPU resources:

- *Call Manager* handles SIP messages processing.
- *Main* handles T-Library processing.

To estimate CPU utilization in single thread mode, use CPU utilization for the whole computer in multi-thread mode. To match or exceed the performance levels, Genesys recommends that you use hardware that is at least as powerful as that in the lab environment.

Hardware Details

Table 118 provides details of the hardware platform that was used for benchmarks on Windows and Linux operating systems.

Table 118: Hardware Details

Feature	Details	Notes
Operating System	• Microsoft Windows Server 2008 R2 64-bit • Linux Red Hat 5.4 64-bit	
Hardware Platform	Dell T7500	
CPU	Intel Xeon E5520 2.26 GHz	Total number of processors per box: 2
RAM	12 GB	

Solution Capacities

The SIP Server solution was characterized for variable call rates using typical call flows found in contact centers. In general, a single instance of the SIP Server would provide the following capacities:

- Windows 2008 64-bit:
  - 15,000 simultaneous calls
  - 15,000 registered agents (direct or aggregated T-Library client connections)
  - Up to 150 calls per second
- Red Hat Linux 5 64-bit:
  - 15,000 simultaneous calls
  - 15,000 registered agents (direct or aggregated T-Library client connection)
  - Up to 150 calls per second
- Windows 2003 or 2008 32-bit and Red Hat Linux 5 32-bit:
  - 4,500 simultaneous calls
  - 10,000 registered agents (aggregated T-Library client connection) or 4,000 registered agents (direct T-Library connections)
  - Up to 150 calls per second

It should be noted that the stated capacities may be used only as indicators of the practical limits, and are not substitutes for careful capacity planning.

The subsequent sections describe the basic assumptions about the Solution's architecture and call flows. A detailed step-by-step procedure explains how to estimate system resource usage including CPU, memory, and network. Included in this section are several examples that show how to apply the suggested methodology to typical situations.

Reference Architecture

Figure 152 depicts the generic architecture of a VOIP solution that was deployed for benchmarking.

Incoming SIP calls that are to be balanced across multiple instances of SIP Server are processed by the Network SIP Server. Media processing (RTP traffic) is handled by multiple Stream Managers. Typically, Stream Managers support treatments (for example, Music On Hold), conferences, and call recording. Media gateways provide media conversion between the TDM (Time-Division Multiplexing) and VOIP domains. The media gateway is a third-party component.

High-availability (HA) capability of the VOIP solution is achieved through the introduction of redundant backup components (shown as shadow boxes in Figure 152). The backup components should be deployed on separate physical hosts, different from the hosts on which the primary components are running.

For proper sizing, this architecture assumes the following constraints:

- For optimum performance, a single instance of SIP Server requires at least four CPU cores. For example, a dual Quad-Core Intel Xeon® host computer runs no more than two server applications.
- SIP Server and Stream Manager must be placed on separate physical host computers.
- TCP/IP transport is used for SIP communication between SIP Server and Stream Manager(s).
- The sip-link-type configuration option is set to 3. (It forces SIP Server to work in multi-threaded mode.)
- SIP Server is configured in HA mode.
- SIP Server has two monitoring T-Library clients (Stat Server and URS).

The quality of the network interface cards (NIC) and NIC drivers is important for achieving optimal performance. The hardware buffer size of the NIC should be at least 64 KB. For example, Genesys has seen good performance with the following NICs on Windows platforms:

- Broadcom BCM5708C NetXtreme II GigE
- Intel E1000 (Intel PRO/1000 Family)

Genesys recommends deploying the latest drivers that are available from the NIC vendor. The default operating system (OS) drivers may not be optimal.

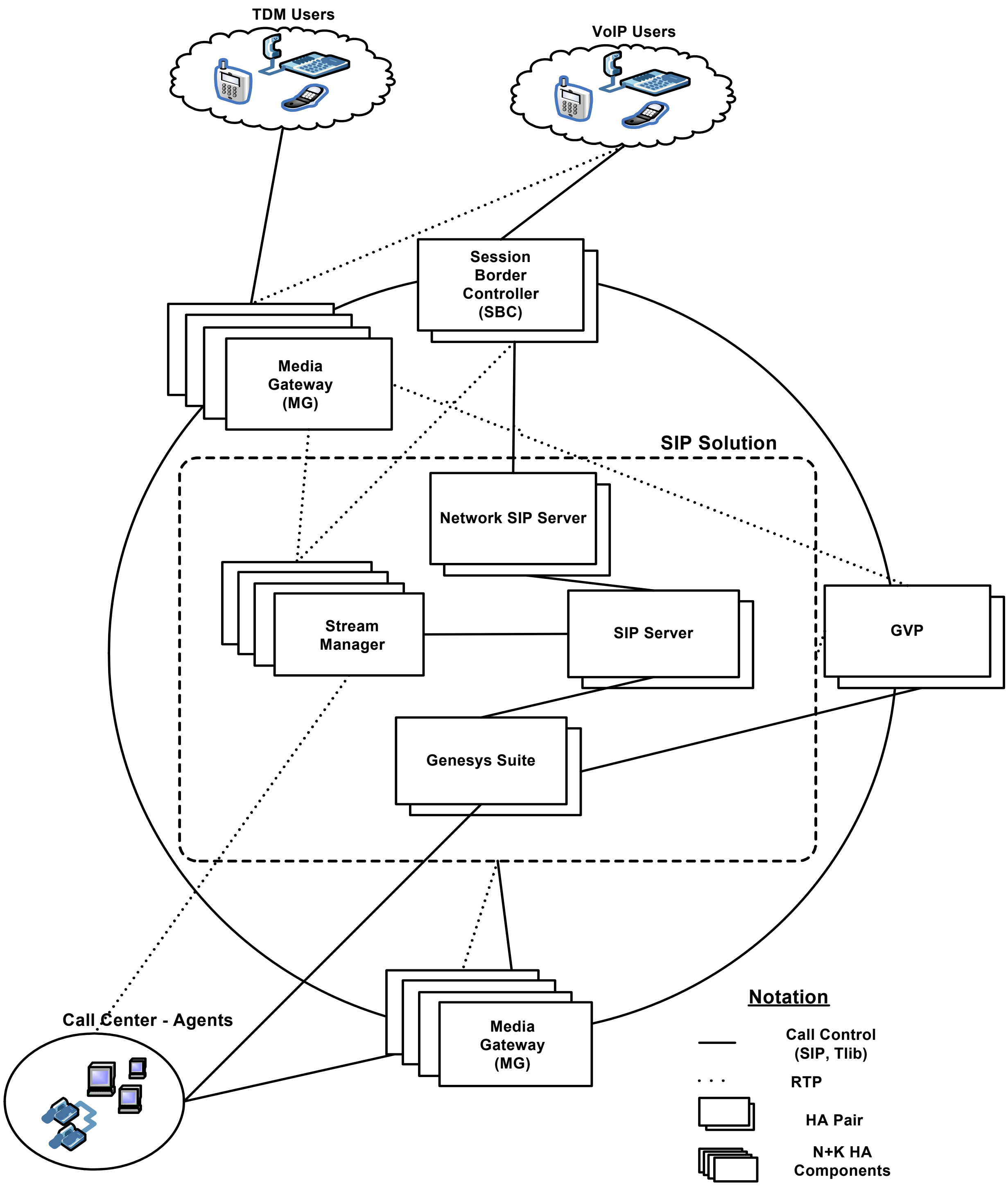


Figure 152: Reference Architecture for VOIP Solution

Scenarios and Benchmarks

This section discusses benchmarks for the following reference scenarios:

- SIP inbound call via Routing Point to a DN
- Single-step transfer
- Two-step transfer
- Call recording on a DN
- Single-step conference
- Two-step conference
- Consultation call
- Internal call using third-party call control (3pcc)
- Treatment on a Routing Point or ACD Queue

These scenarios are considered to be typical for a VOIP solution. For each scenario, this chapter provides the dependencies between the call rate (measured in calls per second [calls/sec]) and CPU usage. These dependencies were evaluated for several SIP Server operating conditions, including the number of concurrent sessions, the number of T-Library client connections, and network traffic. SIP Server performance was evaluated for the following set of values for this parameter:

- 500 concurrent sessions
- 1,000 concurrent sessions
- 1,500 concurrent sessions
- 4,000 concurrent sessions
- 15,000 concurrent sessions (for Windows 2008 64-bit and Red Hat Linux 64-bit for SIP Server 8.1.100.74 and later)

Table 119 indicates the estimated number of SIP dialogs, SIP messages, and T-Library messages for each of the reference scenarios. The actual number of dialogs and messages depend on the options that control SIP Server functionality.

Table 119: Reference Call Flow Details

Scenario Type	Number of SIP Dialogs	Number of SIP Messages	Number of T-Library Requests/Events
SIP Inbound Call	3	20	15
SIP Inbound Call with Single-Step Transfer	4	29	21

Table 119: Reference Call Flow Details (Continued)

Scenario Type	Number of SIP Dialogs	Number of SIP Messages	Number of T-Library Requests/Events
SIP Inbound Call with Two-Step Transfer	6	52	28
SIP Inbound Call with Recording	4	36	15
SIP Inbound Call with Single-Step Conference	6	53	24
SIP Inbound Call with Two-Step Conference	9	76	32
SIP Inbound Call and Consultation Call	4	47	32
Internal Call (using 3pcc)	2	15	13
SIP Inbound Call with Treatment	3	21	17

See also “SIP Server 8.1.0: Reference Call Flows” on page 382.

Subsequent sections describe how to estimate the CPU usage, memory usage and network traffic associated with each scenario, and provide detailed message flows for each scenario.

CPU Usage Estimation

Based on the lab measurements, CPU usage for a given call scenario appears to be a linear function of call rate. Due to the effective implementation of a caching mechanism in SIP Server, the dependency between CPU usage, the number of concurrent sessions, and the number of T-Library client connections is relatively small. CPU utilization for the busiest thread usually depends on the call rate and call complexity. That is, CPU utilization can be estimated by using reference scenarios benchmarking or by using the INVITE message rate (this document uses the reference scenarios benchmarking to estimate the CPU utilization). There are a number of coefficients that can be used to help you account for factors such as the number of connections, the number of monitoring clients, and penalties incurred by the processing of attached data.

The following is the recommended procedure for evaluating SIP Server performance when running mixed call flows.

Procedure: Estimating CPU Usage

Purpose: To estimate the required number of SIP Server instances and CPU loads, given an arbitrary mix of reference scenarios.

Start of procedure

1. Express the specific call flows in terms of the reference scenarios.

   Usually, a customer scenario can be represented as a combination of reference scenarios and additional actions, such as transfer, conference, or treatments on a Routing Point or ACD Queue (in this case, a CPU utilization will be the sum of each scenario and related actions). You should also take into account other treatments that could be applied to a call indirectly, for example—ringback, music in queue, or music on hold.

   For the details of the scenarios, see “SIP Server 8.1.0: Reference Call Flows” on page 382.

2. Specify the average call rate (in calls/sec) and average call duration (in seconds) for inbound, internal, and outbound calls (without consultation calls). For each call type, calculate the average number of concurrent calls by using this formula:

   $ConcurrentCalls_i = CallRate_i * CallDuration_i$

3. Calculate the total number of concurrent calls for all scenarios:

   $TotalConcurrentCalls = \Sigma\ ConcurrentCalls_i$

   Check that the calculated value for TotalConcurrentCalls does not exceed 15,000. If it does, you must use multiple instances of SIP Server, and a Network SIP Server for load balancing, to support your requirements for call volume.

   To estimate the number of SIP Server instances required, use this formula:

   nSIPServers = Max Integer (TotalConcurrentCalls / 15,000)

   As shown in the formula, the result must be rounded to the next highest integer value.

4. If you are using more than one SIP Server, calculate the effective call rate per instance:

   EffectiveCallRate = CallRate / nSIPServers

5. For each scenario, calculate effective call rate and estimate CPU usage for the Call Manager and Main threads by using these formulas:

   $CPUcm_i = Kcm_i * EffectiveCallRate_i$

   $CPUmain_i = Kmain_i * EffectiveCallRate_i$

   where Kcm_i and $Kmain_i$ are the CPU usages for the Call Manager and Main threads, respectively, from Table 120, which contains CPU usage coefficients for most demanding threads (Call Manager and Main) for different scenarios and actions.

6. Calculate total CPU utilization for the Call Manager thread (TotalCPUcm) and total CPU utilization for the Main thread (TotalCPUmain) by using these formulas:

   $TotalCPUmain = \Sigma\ CPUmain_i$

   where $CPUmain_i$ equals the Main thread coefficient values from Table 120 that apply to the customer call flow scenario; and

   $TotalCPUcm = \Sigma\ CPUcm_i$

   where $CPUcm_i$ equals the Call Manager thread coefficient values from Table 120 that apply to the customer call flow scenario.

7. Adjust for the number of T-Library connections, the number of monitoring clients, and penalties incurred by the processing of attached data. These adjustments mostly impact the Main thread CPU utilization.

8. Select the most demanding thread and adjust values for customer hardware. Total CPU utilization for the most demanding thread should not exceed 75%.

Table 120: Calibration Table for CPU Usage (for Intel Xeon CPU E5520 2.26 GHz)

Scenario or Action Type	Coefficient for CPU Usage of Call Manager Thread vs. Call Rate on Windows 2008	Coefficient for CPU Usage of Main Thread vs. Call Rate on Windows 2008	Coefficient for CPU Usage in Single Thread Mode vs. Call Rate on Windows 2008	Coefficient for CPU Usage of Call Manager Thread vs. Call Rate on RHEL 5.4	Coefficient for CPU Usage of Call Main Thread vs. Call Rate on RHEL 5.4	Coefficient for CPU Usage in Single Thread Mode vs. Call Rate on RHEL 5.4
SIP Inbound Call	0.46	0.34	0.87	0.62	0.47	1.2
Single-Step Transfer	0.4	0.32	0.77	0.4	0.32	0.77
Two-Step Transfer	0.97	0.54	1.63	0.97	0.54	1.63

Table 120: Calibration Table for CPU Usage (for Intel Xeon CPU E5520 2.26 GHz) (Continued)

Scenario or Action Type	Coefficient for CPU Usage of Call Manager Thread vs. Call Rate on Windows 2008	Coefficient for CPU Usage of Main Thread vs. Call Rate on Windows 2008	Coefficient for CPU Usage in Single Thread Mode vs. Call Rate on Windows 2008	Coefficient for CPU Usage of Call Manager Thread vs. Call Rate on RHEL 5.4	Coefficient for CPU Usage of Call Main Thread vs. Call Rate on RHEL 5.4	Coefficient for CPU Usage in Single Thread Mode vs. Call Rate on RHEL 5.4
Recording	0.2	0	0.2	0.2	0	0.2
Single-Step Conference	0.9	0.39	1.39	0.9	0.39	1.39
Two-Step Conference	1.6	0.7	2.66	1.6	0.7	2.66
Consultation Call	0.9	0.4	1.4	0.9	0.4	1.4
Internal Call (using 3pcc)	0.46	0.3	0.8	0.46	0.3	0.8
Treatment	0.1	0.02	0.12	0.1	0.02	0.12

Notes:
- When running in multi-threaded mode, CPU usage may exceed 100% on a multicore platform.
- The typical margin of error is 10 percent.
- For the same family of Intel Xeon CPUs, performance of the application is proportional to the CPU clock frequency.
- Estimating CPU usage for a complex call flow simply requires you to combine values for multiple scenarios. For example, to estimate CPU usage for a SIP Inbound Call with a Two-Step Transfer you need to combine CPU calculations for both scenarios.

If the value of parameter EffectiveCallRate is more than the SIP Server can handle, you must deploy additional instances of SIP Server and repeat Steps 4–8 for a new number of SIP Server instances. If a Network SIP Server is present in the configuration, use the following formulas:

$TotalCPUcm = 1.2 * \Sigma CPUcm_i$

$TotalCPUmain = 1.2 * \Sigma CPUmain_i$

These formulas take into account the additional overhead associated with the additional messaging that the Network SIP Server introduces. If TotalCPU exceeds 75 percent for the busiest thread (this value is chosen to accommodate some load fluctuation associated with sparks of incoming rate, endpoint registration, and subscription activity), you must deploy additional instances of SIP Server. Increment the number of SIP Servers (parameter nSIPServers), recalculate EffectiveCallRate, and then use the formulas in Table 120 to find the CPU usage, repeating this process until TotalCPU drops below 75 percent for the most demanding thread.

End of procedure

Estimating CPU Usage for Multiple Client Connections

The preceding estimates are valid when all SIP Server clients share a single T-Library connection to the server (for example, Genesys Desktop is deployed). If clients connect to SIP Server individually (multiple connections to SIP Server), a connection correction factor (CCF) must be applied. T-Library connections impose performance penalties mostly on the Main thread. If the Main thread becomes the most demanding thread, calculate the corrected value for its CPU utilization and use the value for performance calculation.

On Windows

Main thread

On the Windows platform, CPU usage for the Main thread is calculated as follows:

CorrectedTotalCPUmain = TotalCPUmain * (1 + CCFmain * NumberOfConnections)

where:

- TotalCPUmain is the CPU utilization for the Main thread calculated using Table 120
- CCFmain equals one of the following values:
  - 0.000078 for SIP Server 8.1.0 to, but not including, 8.1.100.74
  - 0.000022 for SIP Server 8.1.100.74 and later

Call Manager thread

CPU usage for the Call Manager thread is calculated as follows:

CorrectedTotalCPUcm = TotalCPUcm * (1 + CCFcm * NumberOfConnections)

where:

- TotalCPUcm is the CPU utilization for the Call Manager thread calculated using Table 120
- CCFcm equals one of the following values:
  - 0.000024 for SIP Server 8.1.0 to, but not including, 8.1.100.74
  - 0.000020 for SIP Server 8.1.100.74 and later

On Linux

Main thread

On the Linux platform, CPU usage for the Main thread is calculated as follows:

CorrectedTotalCPUmain = TotalCPUmain * (1 + CCFmain * NumberOfConnections)

where:

- TotalCPUmain is the CPU utilization for the Main thread calculated using Table 120
- CCFmain equals one of the following values:
  - 0.00042 for SIP Server 8.1.0 to, but not including, 8.1.100.74
  - 0.00012 for SIP Server 8.1.100.74 and later

Call Manager thread

CPU usage for the Call Manager thread is calculated as follows:

CorrectedTotalCPUcm = TotalCPUcm * (1 + CCFcm * NumberOfConnections)

where:

- TotalCPUcm is the CPU utilization for the Call Manager thread calculated using Table 120
- CCFcm equals one of the following values:
  - 0.00012 for SIP Server 8.1.0 to, but not including, 8.1.100.74
  - 0.0000 for SIP Server 8.1.100.74 and later

On both Windows and Linux platforms, you must use the CorrectedTotalCPU parameter instead of the TotalCPU parameter in the calculations for estimating the number of the SIP Server instances (nSIPServers).

Note: SIP Server 8.1.100.74 or later is required to obtain better performance results.

Estimating CPU Usage for Number of Attached Data Updates and Attached Data Size

Attached data processing impose performance penalties mostly on the Main thread. Performance penalties depend on attached data update rate and attached data size. If the Main thread becomes the most demanding thread, calculate the corrected value for its CPU utilization and use the value for performance calculation.

To estimate performance penalties for attached data processing, use the formula:

CorrectedTotalCPUmain = TotalCPUmain * (1 + CCupdates * NUpdates) *
(1 + CCattach * EffectiveAttachedDataSize)

where:

- TotalCPUmain is a CPU utilization for the Main thread calculated using Table 120 and adjusted by formulas above
- NUpdates is the number of attached data updates
- CCupdates = 0.028
- CCattach = 0.000015
- EffectiveAttachedDataSize = 0.5 * NUpdates * AttachedDataSize (bytes)

EffectiveAttachedDataSize calculation depends on specific customer call flow. The formula describes a scenario where each attached data update adds a new key-value pair.

Use the CorrectedTotalCPUmain value as TotalCPUmain for further calculations.

For example:

TotalCPUmain = 41%, NUpdates= 20, KeyValue size= 30 bytes

AttachedDataSize= 20 * 30 = 600

EffectiveAttachedDataSize = 1/2 * 20 * 600 = 6000

CorrectedTotalCPUmain = 41 * (1 + 0.028 * 20) * (1 + 0.000015 * 6000) = 69%

Estimating CPU Usage for Multiple Monitoring T-Library Clients

CPU utilization depends on the number of clients that are registered for all T-Library notifications. If the Main thread becomes the most demanding thread, calculate the corrected value for its CPU utilization and use the value for performance calculation. To estimate performance penalties for an additional monitoring client, use the formula:

CorrectedTotalCPUmain = TotalCPUmain * (1 + 0.02 * NumberOfClients)

CPU Clock Scaling

The results for CPU load estimation are given for the maximum number of concurrent sessions, with two monitoring T-Library clients in HA configuration. For the same family of Intel Xeon CPUs, performance of the application is proportional to the clock frequency. For example, if you are upgrading from Intel Xeon model E5520 (four cores, 8 MB L3 cache, HT, 2.26

GHz to Xeon model E5530 (four cores, 8 MB L3 cache, HT, 2.4 GHz), you should expect a performance boost of about 6 percent (2.4 GHz vs. 2.26 GHz).

Note: CPU load estimation that is based on the CPU frequency is valid only for the same family of CPU. For example, Intel Xeon E7220 2.92 GHz shows 20% more CPU utilization in comparison with Intel Xeon E5520 2.26 GHz.

Memory Usage

Based on lab measurements, a single instance of SIP Server uses up to 4 GB of memory on a 64-bit Windows Server 2008 operating system.

This value assumes the maximum number of concurrent sessions (15,000) and the maximum call rate. Less memory is used under a lighter load. An idle server with 4,000 DNs in a switch configuration takes about 190 MB of RAM on idle.

Network Traffic Estimation

Network traffic (T-Library only) depends on T-Event rate, size of attached data, and number of monitoring clients. Use the formula below to estimate T-Library network traffic.

$$TX = \Sigma(CallRate_i * TEventPerCall_i) * (TEventSize + AverageDataSize) * NMonitoredClients$$

$$RX = \Sigma(CallRate_i * TRequestPerCall_i) * TRequestSize$$

where:

- TX—T-Library outbound traffic.
- RX—T-Library inbound traffic.
- $CallRate_i$—The call rate for each scenario.
- $TEventPerCall_i$—The number of TEvents distributed to clients for a single call in a certain scenario.

  Table 119 shows the number of TEvents in reference scenarios. In real deployments, scenarios are usually customized. In this case, TEventPerCall should account for all additional events and requests such as TUpdateUserData and EventAttachedDataChanged.
- TEventSize—The size of a TEvent; 250 bytes is average (the value does not include attached data).
- TRequestSize—The size of a T-Library request; 250 bytes is average (the value does not include attached data).
- NMonitoredClients—The number of T-Library clients getting all T-Library events.
- AverageDataSize—The average size of attached data in one TEvent.

Example

This example illustrates how to use the proposed methods to estimate CPU usage and network traffic on Windows 2008 when no Network SIP Server is deployed. For the purpose of this example, the system specifications are as follows:

- The system employs four basic scenarios, which are reasonably close to the following reference call flows:
  - SIP Inbound Call
  - Two-Step Transfer (50% of all inbound calls will be transferred to another agent)
  - Recording (80% of all inbound calls will get be recorded)
  - Treatment (100% of all inbound calls will get a treatment on a Routing Point)

  Each call has Attached Data with 10 keys of 80 bytes each. Each key is attached sequentially.
- The common call rates and average call durations are:
  - For the SIP Inbound Call scenario: 15 calls/second, 180 seconds
  - For the Two-Step Transfer: 7.5 transfers per second (50% out of inbound call rate)
  - For the Recording scenario: 12 calls/second (80% of inbound call rate)
  - For the Treatment on Routing Point: 15 treatments per second (100% out of inbound call rate)
- The CPU type is Quad Core Intel Xeon E5530, 2.4 GHz.
- The contact center has 1,500 agents, and each agent application has an individual T-Library connection to a SIP Server on Windows 2008.
- The environment includes Stat Server, URS, and ICON.

CPU Usage Estimation

Using the procedure "Estimating CPU Usage" on page 372, the calculations are as follows:

Step 1: Express call flows in terms of reference scenarios.

This step is not required, because the input data provides this information.

Step 2: Calculate the number of concurrent calls per scenario:

a. Per SIP Inbound Call scenario:

ConcurrentCalls = 15.0 calls/sec * 180 sec = 2700 calls

b. Per Two-Step Transfer, Treatment, and Recording scenarios, there are no new calls for our calculation.

Step 3: Calculate the maximum number of concurrent calls in the system:

TotalConcurrentCalls = 2700 + 0 + 0 + 0 = 2700

In this step, we simply check that the number of concurrent calls is below 15,000, which means that one instance of SIP Server is sufficient (and no Network SIP Server is required).

Step 4: Calculate the effective call rate per scenario:

Because we are dealing with a single SIP Server instance, the effective call rates are the same as the original call rates:

a. Per SIP Inbound Call scenario:

EffectiveCallRate = 15.0 calls/sec

b. Per Two-Step Transfer scenario:

EffectiveCallRate = 7.5 call/sec

c. Per Recording scenario:

EffectiveCallRate = 12.0 calls/sec

d. Per Treatment scenario:

EffectiveCallRate = 15.0 calls/sec

Step 5: Estimate CPU usage per scenario:

Using the calibration table for CPU usage (Table 120 on page 373), we get the following estimates for CPU usage per scenario:

a. Per SIP Inbound Call scenario:

CPUcm = 0.46 (% * sec/call) * 15.0 calls/sec = 6.9%

CPUmain = 0.34 (% * sec/call) * 15.0 calls/sec = 5.1%

b. Per Two-Step Transfer scenario:

CPUcm = 0.97 (% * sec/call) * 7.5 call/sec = 7.28%

CPUmain = 0.54 (% * sec/call) * 7.5 call/sec = 4.05%

c. Per Recording scenario:

CPUcm = 0.2 (% * sec/call)* 12.0 calls/sec = 2.4%

CPUmain = 0%

d. Per Treatment scenario:

CPUcm = 0.1 (% * sec/call)* 15.0 calls/sec = 1.5%

CPUmain = 0.02 (% * sec/call)* 15.0 calls/sec = 0.3%

Thus:

TotalCPUcm = 6.9 + 7.28 + 2.4 + 1.5 = 18.08%

TotalCPUmain = 5.1 + 4.05 + 0 + 0.3 = 9.45%

Next, adjust for the number of client connections (1,500 clients) by factoring in the CPU usage:

CorrectedTotalCPUcm = 18.08 * (1+ 0.000024 * 1500) = 18.73%

CorrectedTotalCPUmain = 9.45 * (1+ 0.000078 * 1500) = 10.56%

Adjust for Attached Data processing. In this case, we have 10 key-value pairs that are attached sequentially (in other words, 10 Attached Data updates). Each key-value pair is 80 bytes.

EffectiveAttachedDataSize = 1/2*10*800 = 4000

CorrectedTotalCPUmain = 10.56 * (1 + 0.028 * 10) * (1 + 0.000015 * 4000) = 14.33%

To adjust for the number of monitoring T-Library client connections (where the data in the calibration table, Table 120, is collected with active Stat Server and URS, and our environment includes ICON), we must adjust for one additional client:

CorrectedTotalCPUmain = 14.33 * (1 + 0.02 * 1) = 14.62%

To this point our calculations yield:

CorrectedTotalCPUcm = 18.73%

CorrectedTotalCPUmain = 14.62%

Because the most demanding thread is the Call Manager thread, we will use CorrectedTotalCPUcm for further calculations.

Before we move on, we must adjust for a clock speed difference. The calibration table for CPU usage assumes an Intel CPU running at 2.26 GHz, but we are using a 2.4 GHz CPU. Therefore, we must scale the result for the lower clock frequency:

CPU = 18.73 * (2.26/2.4) = 17.64%

Network Traffic Estimation

To estimate the amount of network traffic, we need to estimate T-Event rate. As defined above:

- Inbound Call rate = 15 calls per second
- Two-Step Transfer rate = 7.5 transfers per second
- Treatment rate = 15 treatments per second

We use the following facts taken from the log:

- Inbound Call scenario generates 12 T-Events
- Two-Step Transfer adds 11 T-Events
- Treatment adds 3 T-Events
- Attached Data updates 10 T-Events

Calculate the T-Library outbound traffic:

$TX = \sum (CallRate_i * TEventPerCall) * (TEventSize + AverageDataSize) * NMonitoredClients =$

$((15 * 12) + (7.5 * 11) + (15 * 3) + (15 * 10)) * (250 + 400) * 3$

= 892 KB

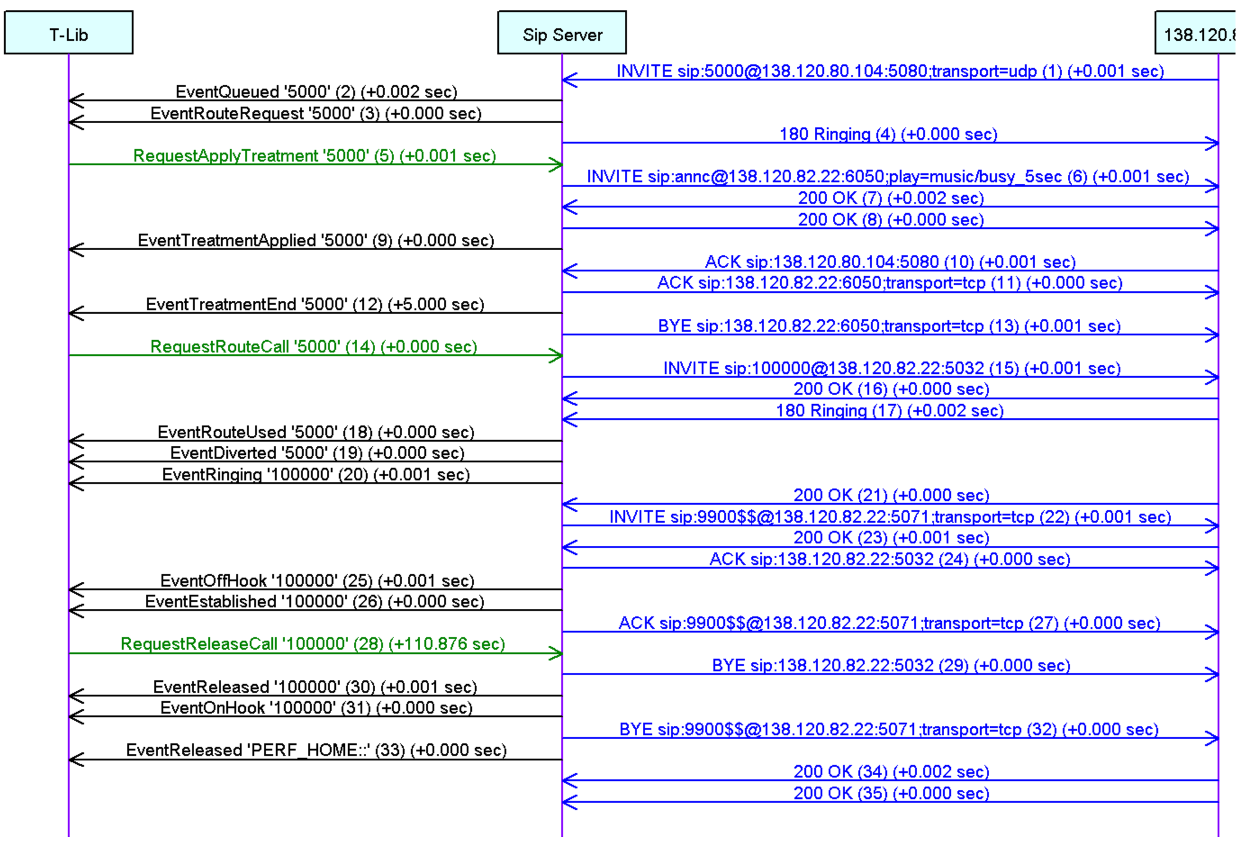
Note: The amount of traffic via host network interfaces depends on the particular placement of SIP Server instances.

SIP Server 8.1.0: Reference Call Flows

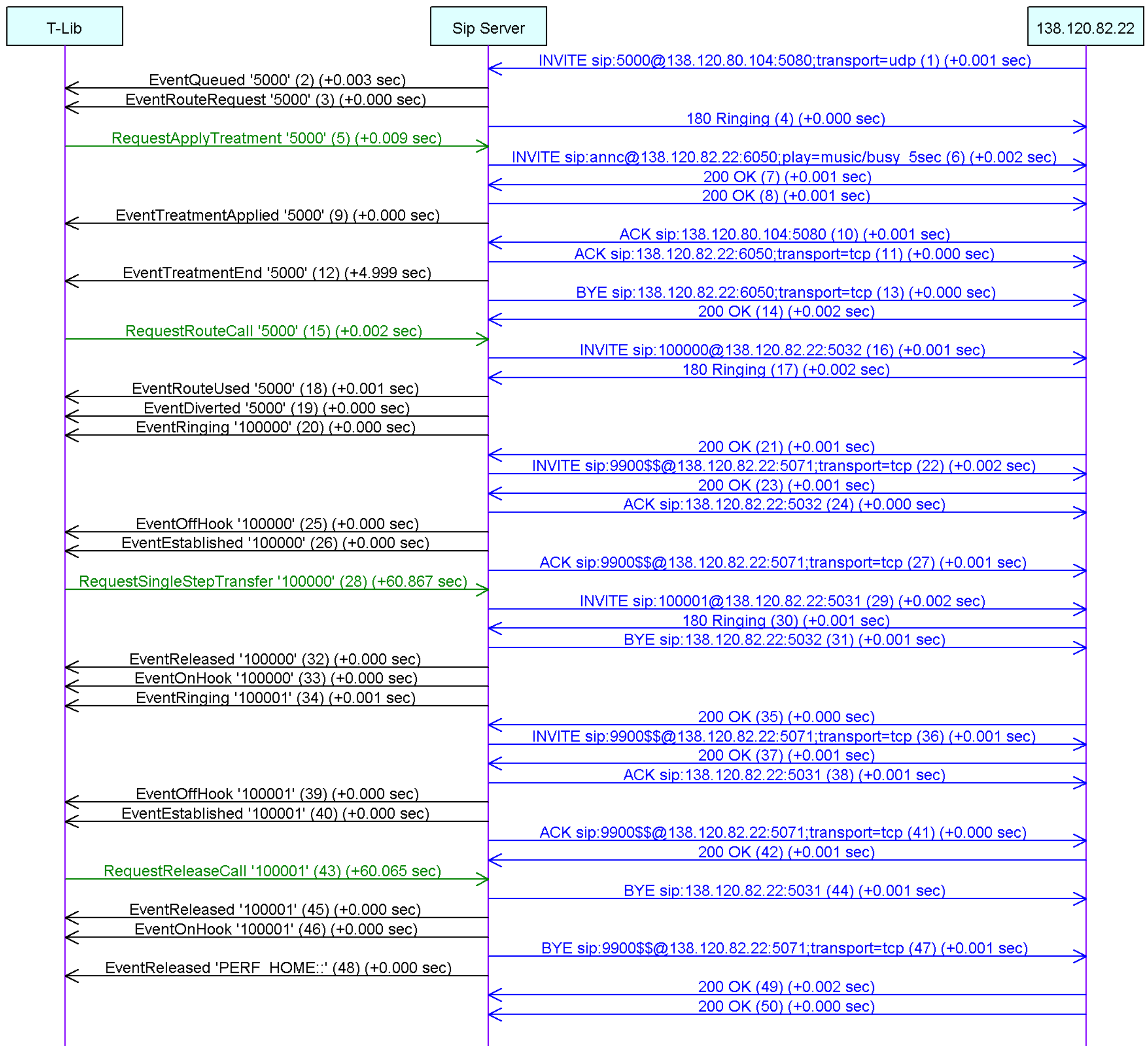
This section provides reference call flows and details of SIP messages that were used for benchmarking of the following call scenarios:

- "SIP Inbound Call with Treatment" on page 382
- "SIP Inbound Call with Single-Step Transfer" on page 383
- "SIP Inbound Call with Two-Step Transfer" on page 384
- "SIP Inbound Call with Recording" on page 385
- "SIP Inbound Call with Single-Step Conference" on page 386
- "SIP Inbound Call with Two-Step Conference" on page 387
- "SIP Inbound Call and Consultation Call" on page 388
- "Internal Call" on page 389

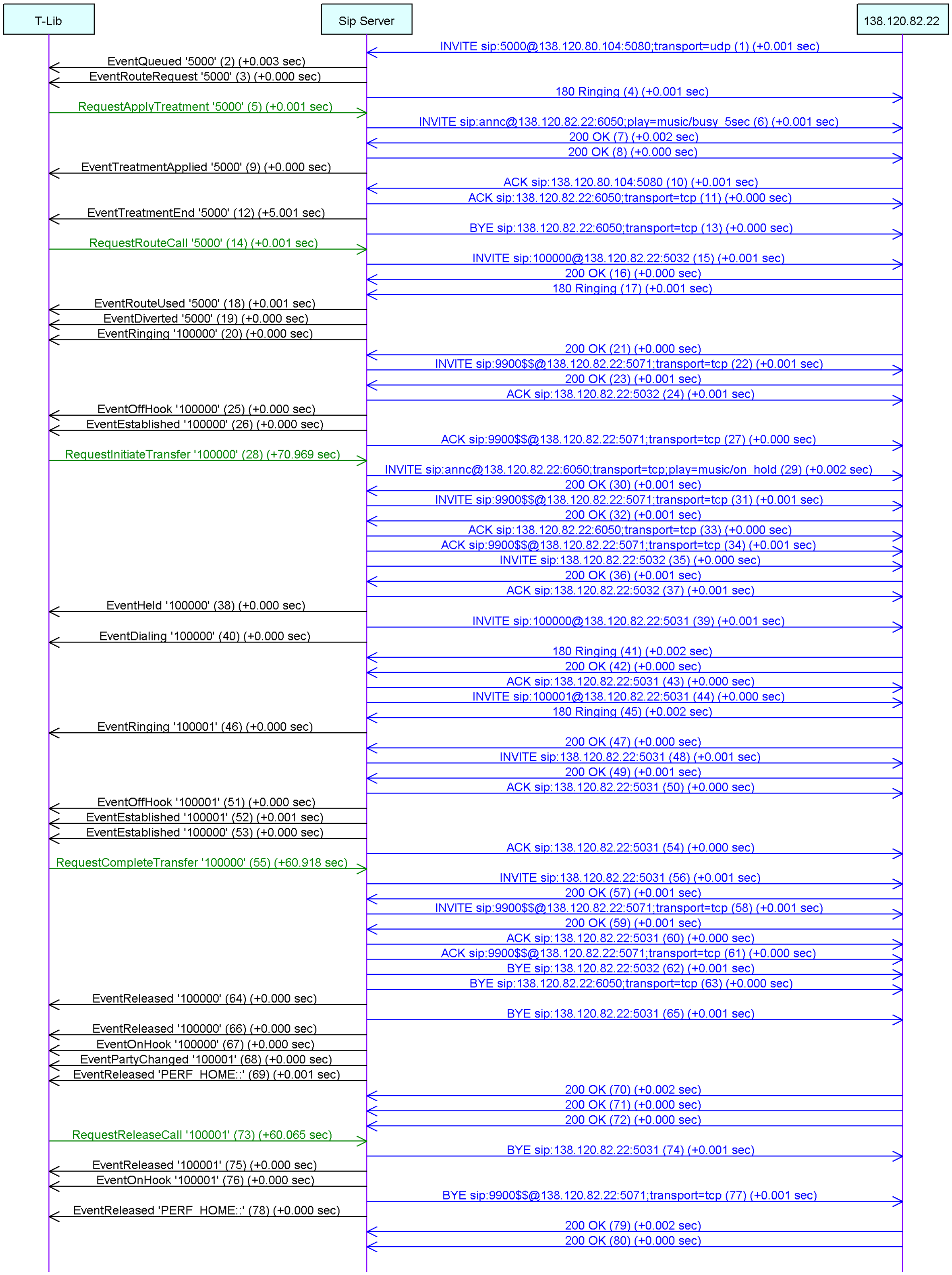
SIP Inbound Call with Treatment



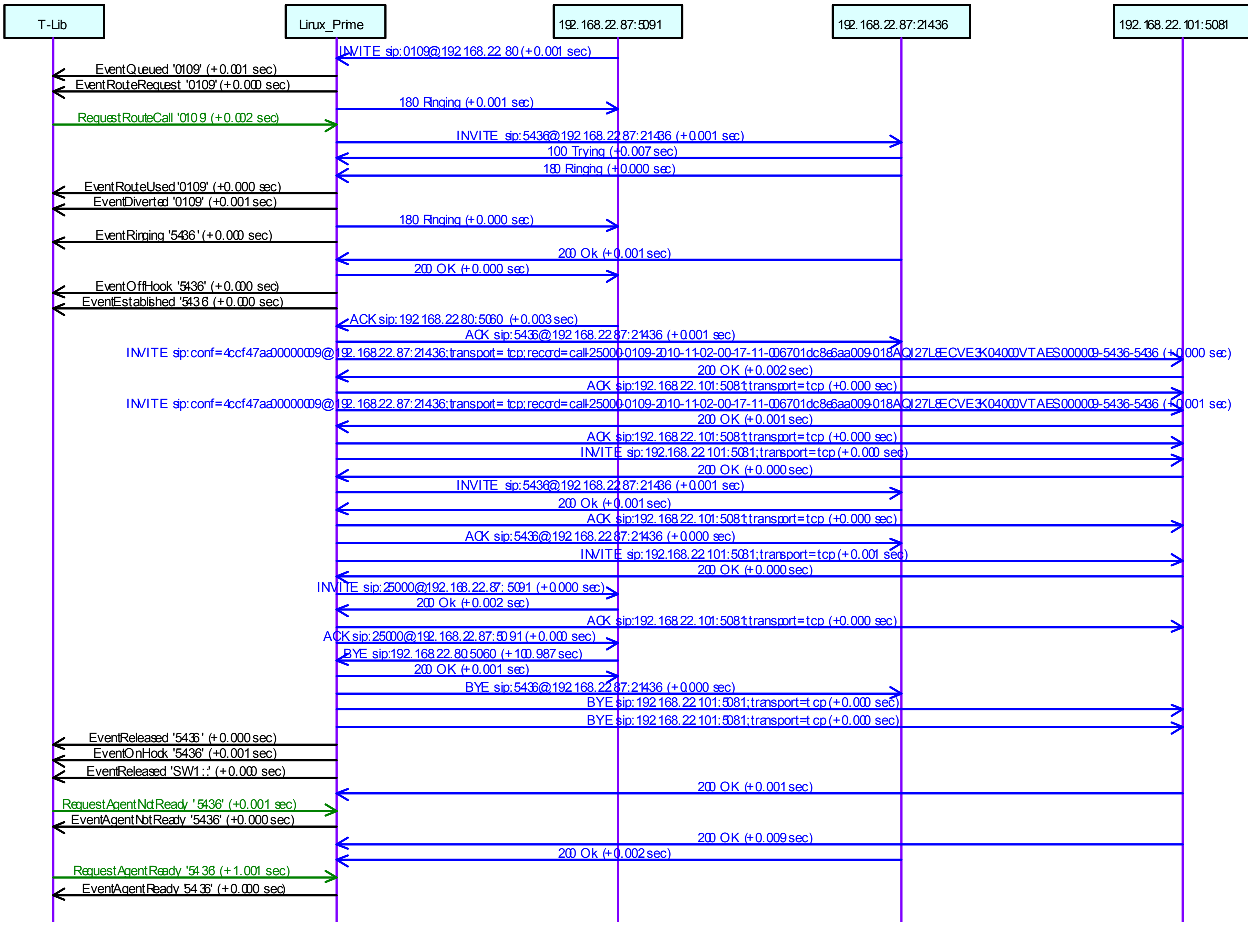
SIP Inbound Call with Single-Step Transfer



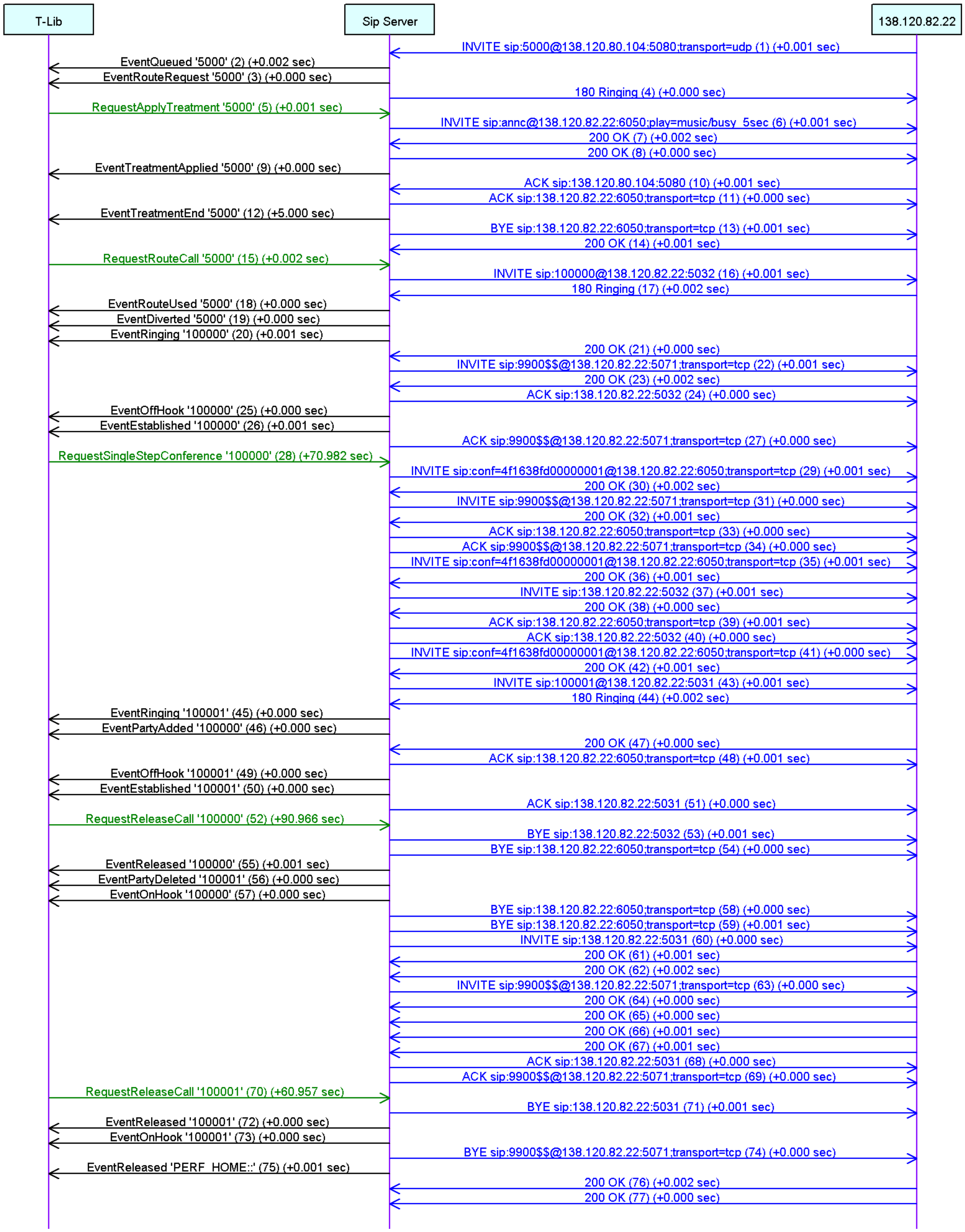
SIP Inbound Call with Two-Step Transfer



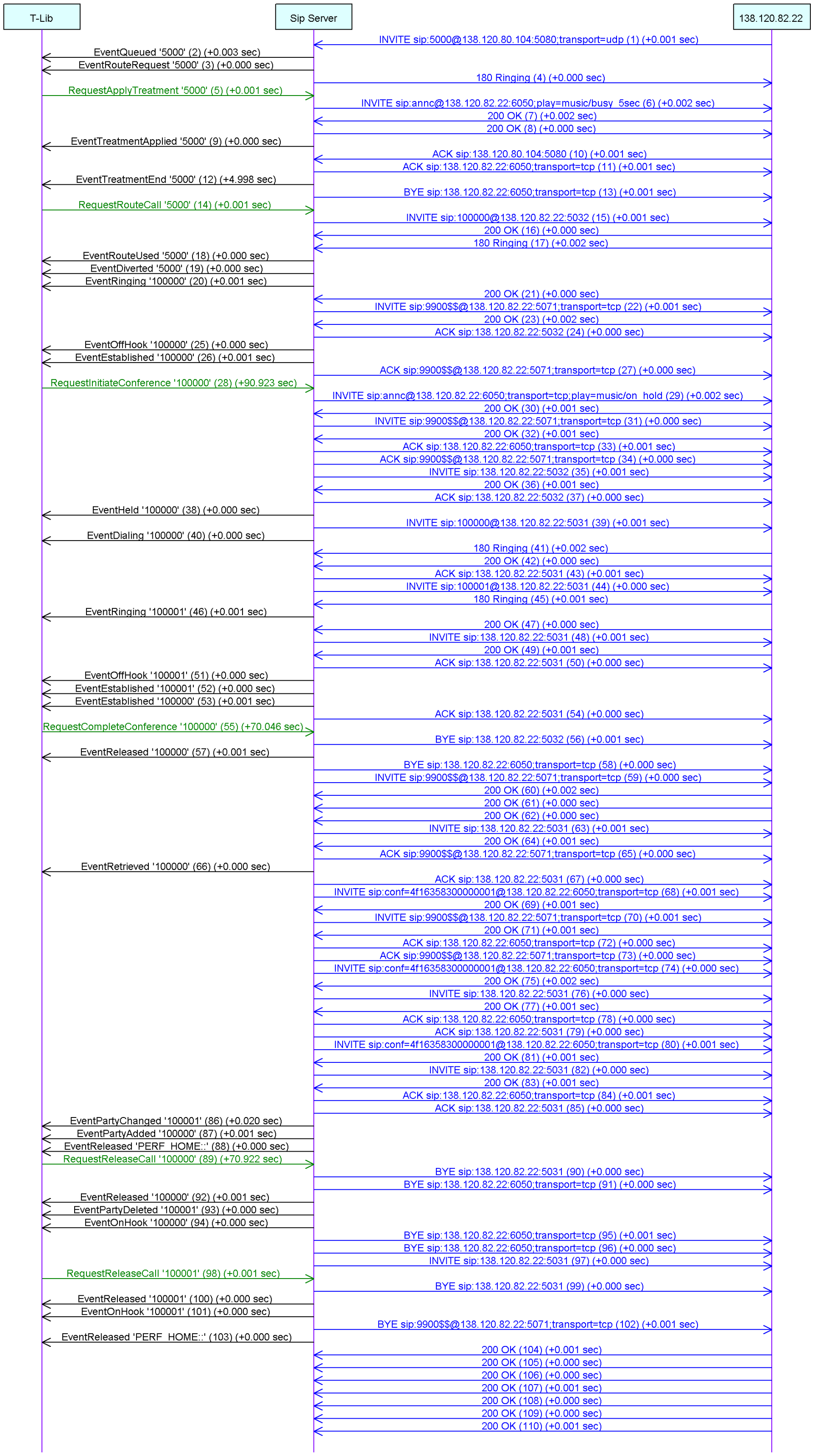
SIP Inbound Call with Recording



SIP Inbound Call with Single-Step Conference

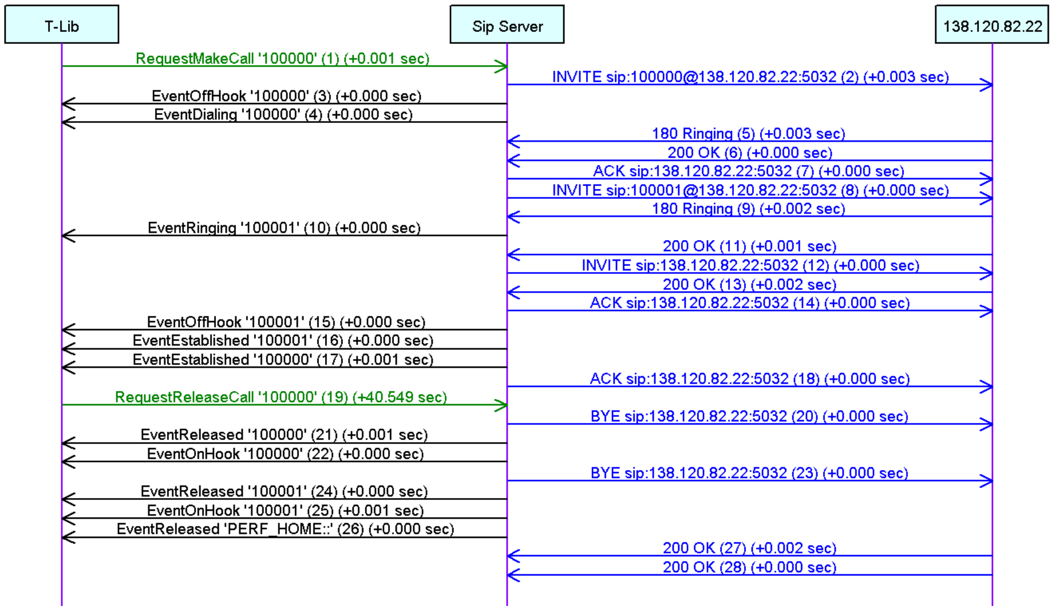


SIP Inbound Call with Two-Step Conference



SIP Inbound Call and Consultation Call

Internal Call



SIP Server 8.0: Sizing and Performance

This section provides hardware sizing guidelines and the basic information required for deploying and capacity planning of a Voice over Internet Protocol (VOIP) solution based on SIP Server releases 7.6 and 8.0. It covers the following topics:

- Solution Capacities
- Required hardware resources (the number of physical host computers)
- Number of applications (SIP Servers and Stream Managers)
- Recommended network architecture
- Application placement across the host computers
- Expected system loads (CPU, memory, and traffic per network interface)

This section also provides benchmarks for the following basic scenarios:

- SIP Inbound Call
- Single-Step Transfer
- Recording
- Single-Step Conference
- Consultation Call
- Internal Call (using third-party call control [3pcc])
- Treatment

Running a particular scenario involves multiple system components. For a reference system architecture, see "Reference Architecture" on page 391.

For the benchmark data obtained for the reference scenarios, see "Scenarios and Benchmarks" on page 392.

SIP Server performance is measured for variable call rates. Other important parameters are the number of concurrent sessions and the complexity of the scenarios. The results presented in this section are based on lab benchmarks obtained on the operating systems, as described in Table 121. SIP Server is benchmarked in multi-threaded mode.

To match or exceed the performance levels, Genesys recommends that you use hardware that is at least as powerful as that in the lab environment.

Hardware Details

Table 121 provides details of the hardware platform that was used for benchmarks on Windows and Linux.

Table 121: Hardware Details

Feature	Details	Notes
Operating System	• Microsoft Windows Server 2003 Enterprise Edition, 32-bit SP2 • Linux Red Hat 5.4 32-bit	Default socket buffer size adjusted to 64 KB
Hardware Platform	Dell 2950	
CPU	Dual Dual-Core Intel Xeon X5160, 3.0 GHz clock frequency	Total number of cores: 4 L2 cache size: 2x4 MB
RAM	8 GB	2x4 GB, 667 MHz (PAE is used)
HDD	Two (2) 73 GB 15K RPM Serial-Attach SCSI 3 Gbps 3.5-in	Controller Type: Integrated SAS/SATA, PERC 6/I, no RAID configuration
NIC	Two (2) Broadcom BCM5708C NetXtreme II GigE	1 GB Full Duplex, TCP Offload Engine (TOE) disabled

Default Socket Buffer Size

For optimal SIP Server and Stream Manager performance, the default socket buffer size (for both TCP and UDP sockets) must be increased to at least 64

KB. The default buffer size is OS-specific, and is typically less than 64 KB. For example, on Windows Server 2003 with SP2, the default settings could be 8,096 bytes.

Solution Capacities

The SIP Server 8.0 solution was characterized for variable call rates using typical call flows found in contact centers. In general, a single instance of the SIP Server would provide the following capacities:

- Up to 4,500 concurrent calls
- Up to 100 calls/sec call rate
- Up to 10,000 SIP endpoints and agent desktops using an aggregated connection for T-Library clients
- Up to 4,000 SIP endpoints and agent desktops using dedicated connections for T-Library clients

It should be noted that the stated capacities may be used only as indicators of the practical limits, and are not substitutes for careful capacity planning.

The subsequent sections describe the basic assumptions about the Solution's architecture and call flows. A detailed step-by-step procedure explains how to estimate system resource usage including CPU, memory, and network. Included in this section are several examples that show how to apply the suggested methodology to typical situations.

Reference Architecture

Figure 152 depicts the generic architecture of a VOIP solution that was deployed for benchmarking.

Incoming SIP calls that are to be balanced across multiple instances of SIP Server are processed by the Network SIP Server. Media processing (RTP traffic) is handled by multiple Stream Managers. Typically, Stream Managers support treatments (for example, Music On Hold), conferences, and call recording. Media gateways provide media conversion between the TDM (Time-Division Multiplexing) and VOIP domains. The media gateway is a third-party component.

High-availability (HA) capability of the VOIP solution is achieved through the introduction of redundant backup components (shown as shadow boxes in Figure 152). The backup components should be deployed on separate physical hosts, different from the hosts on which the primary components are running. This chapter does not discuss the impact of HA.

For proper sizing, this architecture assumes the following constraints:

- For optimum performance, a single instance of SIP Server requires at least four CPU cores. For example, a dual Quad-Core Intel Xeon® host computer runs no more than two server applications.

- SIP Server and Stream Manager must be placed on separate physical host computers.
- The total number of open TCP/UDP ports per host should not exceed 10,000.
- The default socket buffer size is set to at least 64 KB.
- To prevent voice quality degradation, the amount of network traffic (the sum of the sent/received bytes) per network interface should be limited to 100 Mbit/sec, and to 250 Mbit/sec per network segment.
- TCP/IP transport is used for SIP communication between SIP Server and Stream Manager(s).

The quality of the network interface cards (NIC) and NIC drivers is important for achieving optimal performance. The hardware buffer size of the NIC should be at least 64 KB. For example, Genesys has seen good performance with the following NICs on Windows platforms:

- Broadcom BCM5708C NetXtreme II GigE
- Intel E1000 (Intel PRO/1000 Family)

Genesys recommends deploying the latest drivers that are available from the NIC vendor. The default operating system (OS) drivers may not be optimal.

Scenarios and Benchmarks

This section discusses benchmarks for the following reference scenarios:

- SIP Inbound Call (first-party call control [1pcc])
- Single-Step Transfer
- Recording
- Single-Step Conference
- Consultation Call
- Internal Call (using 3pcc)
- Treatment

These scenarios are considered to be typical for a VOIP solution. For each scenario, this chapter provides the dependencies between the call rate (measured in calls per second [calls/sec]) and CPU usage. These dependencies were evaluated for several SIP Server operating conditions, with the major parameter being the number of concurrent sessions. SIP Server performance was evaluated for the following set of values for this parameter:

- 500 concurrent sessions
- 1,000 concurrent sessions
- 1,500 concurrent sessions
- 4,000 concurrent sessions (for a SIP Inbound Call scenario)

Table 122 indicates the typical number of SIP dialogs, SIP messages and T-Library messages for each of the reference scenarios. Subsequent sections describe how to estimate the CPU usage, memory usage and network traffic associated with each scenario, and provide detailed message flows for each scenario.

Table 122: Reference Call Flow Details

Scenario Type	Number of SIP Dialogs	Number of SIP Messages	Number of T-Library Requests/Events
SIP Inbound Call	2	14	18
Single-Step Transfer	3	25	28
Recording	4	36	15
Single-Step Conference	6	48	31
Consultation Call	4	47	32
Internal Call (using 3pcc)	2	17	25
Treatment	3	21	17

See also “SIP Server 8.0: Reference Call Flows” on page 402.

CPU Usage Estimation

Based on the lab measurements, CPU usage for a given call scenario appears to be a linear function of call rate. Due to the effective implementation of a caching mechanism in SIP Server, the dependency between CPU usage and the number of concurrent sessions is relatively small. The following is the recommended procedure for evaluating SIP Server performance when running mixed call flows.

Procedure: Estimating CPU Usage

Purpose: To estimate the required number of SIP Server instances and CPU loads, given an arbitrary mix of reference scenarios.

Start of procedure

1. Express the specific call flows in terms of the reference scenarios.

   For the details of the scenarios, see “SIP Server 8.0: Reference Call Flows” on page 402.

2. For each scenario, specify the average call rate (in calls/sec) and average call duration (in seconds). Calculate the average number of concurrent calls per scenario by using this formula:

   $ConcurrentCalls_i = CallRate_i * CallDuration_i$

3. Calculate the total number of concurrent calls for all scenarios:

   $TotalConcurrentCalls = \Sigma\ ConcurrentCalls_i$

   Check that the calculated value for TotalConcurrentCalls does not exceed 4,000. If it does, you must use multiple instances of SIP Server, and a Network SIP Server for load balancing, to support your requirements for call volume.

   To estimate the number of SIP Server instances required, use this formula:

   nSIPServers = Max Integer (TotalConcurrentCalls / 4,000)

   As shown in the formula, the result must be rounded to the next highest integer value. The number of SIP Server instances per host must not exceed the number of CPU cores. At maximum load, each SIP Server requires up to 500 MB of resident memory, so the latter constraint is bound by the amount of memory that is available to applications on a 32-bit version of Windows Server 2003 (default user space of 2 GB, assuming 4 GB of physical RAM).

4. If you are using more than one SIP Server, calculate the effective call rate per instance:

   $EffectiveCallRate_i = CallRate_i / nSIPServers$

5. For each scenario and effective call rate, estimate CPU usage by using an approximation formula for the corresponding call flow. Table 123 contains these formulas, which take form $CPU_i = K_i * EffectiveCallRate_i$, where the index ($_i$) corresponds to a specific call flow (scenario).

Table 123: Calibration Table for CPU Usage (for Intel Xeon CPU, 2.33 GHz)

Scenario Type	Formula for CPU Usage CPU(%) vs. Call Rate (Calls/Sec) on Windows 2003	Formula for CPU Usage CPU(%) vs. Call Rate (Calls/Sec) on RHEL 5.4	Maximum Estimated Call Rate (Calls/Sec)
SIP Inbound Call	CPU = 0.80 * EffectiveCallRate	CPU = 0.74 * EffectiveCallRate	100
Single-Step Transfer	CPU = 1.49 * EffectiveCallRate	CPU = 1.55 * EffectiveCallRate	50
Recording	CPU = 1.04 * EffectiveCallRate	CPU = 0.9 * EffectiveCallRate	100
Single-Step Conference	CPU = 2.50 * EffectiveCallRate	CPU = 2.62 * EffectiveCallRate	30
Consultation Call	CPU = 2.17 * EffectiveCallRate	CPU = 2.07 * EffectiveCallRate	50

Table 123: Calibration Table for CPU Usage (for Intel Xeon CPU, 2.33 GHz) (Continued)

Scenario Type	Formula for CPU Usage CPU(%) vs. Call Rate (Calls/Sec) on Windows 2003	Formula for CPU Usage CPU(%) vs. Call Rate (Calls/Sec) on RHEL 5.4	Maximum Estimated Call Rate (Calls/Sec)
Internal Call (using 3pcc)	CPU = 0.918 * EffectiveCallRate	CPU = 1.01 * EffectiveCallRate	100
Treatment	CPU = 0.99 * EffectiveCallRate	CPU = 0.99 * EffectiveCallRate	100

Notes:
- When running in multi-threaded mode, the CPU usage may exceed 100% on a multicore platform.
- The typical margin of error is 10 percent.
- For simplicity, the calibration table (Table 123 on page 394) assumes the maximum number of concurrent sessions and a CPU clock frequency of 2.33 GHz. For Intel Xeon CPUs with an L2 cache size above 2 MB, performance of the application scales linearly with the CPU clock frequency.

If the value of parameter EffectiveCallRate is out of the applicable range indicated in Table 123, you must deploy additional instances of SIP Server and recalculate EffectiveCallRate as follows:

a. Using the calibration table (Table 123), calculate the total CPU usage by summing individual values:

$TotalCPU = \Sigma\ CPU_i$

b. If a Network SIP Server is present in the configuration, use the following formula:

$TotalCPU = 1.20 * \Sigma\ CPU_i$

This formula takes into account the additional overhead associated with the additional messaging that the Network SIP Server introduces. If TotalCPU exceeds 70 percent, you must deploy additional instances of SIP Server. Increment the number of SIP Servers (parameter nSIPServers), recalculate EffectiveCallRate, and then use the formulas in Table 123 to find the CPU usage, repeating this process until TotalCPU drops below 70 percent.

End of procedure

Estimating CPU Usage for Multiple Client Connections

The preceding estimates are valid when all SIP Server clients share a single T-Library connection to the server (for example, Genesys Desktop is deployed). If clients use individual multiple connections to SIP Server, a connection correction factor (CCF) must be applied to TotalCPU.

On Windows

On Windows platform the resulting CPU usage is calculated using the connection factor from Table 124, as follows:

CorrectedTotalCPU = CCF * TotalCPU

Table 124: Correction Factor for Multiple Connections on Windows

Number of Client Connections	Connection Correction Factor (CCF)
1–99	1.00
100–499	1.02
500–999	1.05
1000–1999	1.10
2000–4000	1.30

On Linux

On Linux platform the resulting CPU usage can be calculated as follows:

CorrectedTotalCPU = 30% + TotalCPU

The formula is valid for a number of connections between 1000 and 2000 and for a call rate range between 5 to 100 call/sec.

On both Windows and Linux platforms, the CorrectedTotalCPU parameter must be used instead of parameter TotalCPU in the calculations for estimating the number of the SIP Server instances (nSIPServers).

CPU Clock Scaling

The results for CPU load estimation are given for the maximum number of concurrent sessions and a CPU clock frequency of 3.0 GHz. For Intel Xeon CPUs with an L2 cache size above 2 MB, performance of the application scales linearly with the clock frequency. For example, if you are upgrading from Intel Xeon model X5160 (Dual Core, 2x4 MB cache, 3.0 GHz, 1333 MHz FSB (Front-Side Bus) to Xeon model X5260 (Dual Core 2x6 MB cache, 3.33 GHz, 1333 MHz FSB), you should expect a performance boost of about 11 percent (3.33 GHz/3.0 GHz). In this example, the scaling coefficients in Table 123 on page 394 must be *reduced* by 11 percent accordingly.

Memory Usage

Based on lab measurements, a single instance of SIP Server uses up to 500 MB of resident memory on a 32-bit Windows Server 2003 operating system. This value assumes the maximum number of concurrent sessions (4,000) and the maximum call rate. Less memory is used under a lighter load. An idle server with 4,000 DNs in a switch configuration takes about 190 MB of RAM on idle.

Network Traffic Estimation

Table 125 provides formulas for estimating the amount of incoming (RX) and outgoing (TX) traffic per SIP Server instance and specific scenario. When using Table 125, you must use parameter $EffectiveCallRate_i$ for each scenario (denoted by index $_i$), as calculated in Step 4 on page 394.

Table 125: Calibration Table for Network Traffic

Scenario Type	Formula for Incoming Traffic (RX): KB Received vs. EffectiveCallRate (Calls/Sec)	Formula for Outgoing Traffic (TX): KB Sent vs. EffectiveCallRate (Calls/Sec)	Applicable Call Rate (Calls/Sec)
SIP Inbound Call	RX = 7.3 * EffectiveCallRate	TX = 51.8 * EffectiveCallRate	0–100
Single-Step Transfer	RX = 14.2 * EffectiveCallRate	TX = 75.8 * EffectiveCallRate	0–50
Recording	RX = 7.8 * EffectiveCallRate	TX = 48.6 * EffectiveCallRate	0–100
Single-Step Conference	RX = 24 * EffectiveCallRate	TX = 132 * EffectiveCallRate	0–25
Consultation Call	RX = 20.4 * EffectiveCallRate	TX = 112.2 * EffectiveCallRate	0–40
Internal Call (using 3pcc)	RX = 9.75 * EffectiveCallRate	TX = 71.48 * EffectiveCallRate	0–100
Treatment	RX = 9.07 * EffectiveCallRate	TX = 51.3 * EffectiveCallRate	0–100

Note: The typical margin of error is 5 percent.

The resulting network traffic is the sum of individual components (calculated per scenario):

$$TotalRX = \sum RX_i$$

$$TotalTX = \sum TX_i$$

Examples

This section provides several examples to show how the described approach works in practice.

Simple Example

This example illustrates how to use the proposed methods to estimate CPU usage and network traffic on Windows 2003 when no Network SIP Server is deployed. Typically, this means that the aggregated call rate and the number of concurrent sessions are relatively low.

For the purpose of this example, the system specifications are as follows:

- The system supports three basic scenarios, which are reasonably close to the following reference call flows:
  - SIP Inbound Call
  - Single-Step Transfer
  - Recording
- The common call rates and average call durations are:
  - For the SIP Inbound Call scenario: 2 calls/sec, 180 sec
  - For the Single-Step Transfer scenario: 1 call/sec, 240 sec
  - For the Recording scenario: 2 calls/sec, 300 sec
- The CPU type is Quad Core Intel Xeon E5405, 2x6 MB cache, 2.0 GHz, 1333 MHz FSB.
- The contact center has 1,500 agents, and each agent application has an individual T-Library connection to a SIP Server on Windows 2003.

CPU Usage Estimation

Using the procedure "Estimating CPU Usage" on page 393, the calculations are as follows:

Step 1: Express call flows in terms of reference scenarios.

This step is not required, because the input data provides this information.

Step 2: Calculate the number of concurrent calls per scenario:

a. Per SIP Inbound Call scenario:

$ConcurrentCalls_1$ = 2.0 calls/sec * 180 sec = 360 calls

b. Per Single-Step Transfer scenario:

$ConcurrentCalls_2$ = 1.0 call/sec * 240 sec = 240 calls

c. Per Recording scenario:

$ConcurrentCalls_3$ = 2.0 calls/sec * 300 sec = 600 calls

Step 3: Calculate the maximum number of concurrent calls in the system:

TotalConcurrentCalls = 360 + 240 + 600 = 1,200

In this step, we simply check that the number of concurrent calls is below 4,000, which means that one instance of SIP Server is sufficient (and no Network SIP Server is required).

Step 4: Calculate the effective call rate per scenario:

Because we are dealing with a single SIP Server instance, the effective call rates are the same as the original call rates:

a. Per SIP Inbound Call scenario:

$EffectiveCallRate_1$ = 2.0 calls/sec

b. Per Single-Step Transfer scenario:

$EffectiveCallRate_2$ = 1.0 call/sec

c. Per Recording scenario:

$EffectiveCallRate_3$ = 2.0 calls/sec

Step 5: Estimate CPU usage per scenario:

Using the calibration table for CPU usage (Table 123 on page 394), we get the following estimates for CPU usage per scenario:

a. Per SIP Inbound Call scenario:

CPU_1 = 0.8 (% * sec/call) * 2.0 calls/sec = 1.6%

b. Per Single-Step Transfer scenario:

CPU_2 = 1.49 (% * sec/call) * 1.0 call/sec = 1.49%

c. Per Recording scenario:

CPU_3 = 1.04 (% * sec/call)* 2.0 calls/sec = 2.08%

So, TotalCPU = 1.6 + 1.49 + 2.08 = 7.3%

Because we have multiple client connections to SIP Server on Windows 2003 (a maximum of 1,500), we must factor the CPU usage according to Table 124 on page 396:

CorrectedTotalCPU = 1.10 * 7.3 = 8.03%

However, the calibration table for CPU usage assumes an Intel CPU running at 3.0 GHz, and we are using a 2.0 GHz CPU. Therefore, we must scale the result for the lower clock frequency:

CPU = 5.69 * (3.0/2.0) = 8.53%

Network Traffic Estimation

To estimate the amount of network traffic, use the calibration table for network traffic (Table 125 on page 397), summing the values for individual scenarios to get the following results (for SIP traffic):

- For received (RX) traffic:

  RX = 7.3 * 2.0 + 14.2 * 1.0 + 7.8 * 2.0 = 44.4 (KB/sec)

- For transmitted (TX) traffic:

  TX = 51.8 * 2.0 + 75.8 * 1.0 + 48.6 * 2.0 = 276.6 (KB/sec)

Complex Example

This example shows how to use the proposed procedure to estimate CPU usage and network traffic when a Network SIP Server is deployed. This means that the aggregated call rate and the number of concurrent sessions are relatively high.

For the purpose of this example, the system specifications are as follows:

- The system supports three basic scenarios, which are reasonably close to the following reference call flows:
  - SIP Inbound Call
  - Single-Step Transfer
  - Consultation Call
- The common call rates and average call durations are:
  - For the SIP Inbound Call scenario: 20 calls/sec, 180 sec
  - For the Single-Step Transfer scenario: 10 calls/sec, 240 sec
  - For the Consultation Call scenario: 2 calls/sec, 300 sec
- The CPU type is Quad Core Intel Xeon E5430, 2x6 MB cache, 2.66 GHz, 1333 MHz FSB.
- The contact center has 8,000 agent desktops, with individual T-Library client connections aggregated by means of Genesys Desktop Server.

CPU Usage Estimation

Using the procedure “Estimating CPU Usage” on page 393, the calculations are as follows:

Step 1: Express call flows in terms of reference scenarios.

This step is not required, because the input data provides this information.

Step 2: Calculate the number of concurrent calls per scenario:

a. Per SIP Inbound Call scenario:

$ConcurrentCalls_1$ = 20 calls/sec * 180 sec = 3,600 calls

b. Per Single-Step Transfer scenario:

$ConcurrentCalls_2$ = 10 calls/sec * 240 sec = 2,400 calls

c. Per Consultation Call scenario:

$ConcurrentCalls_3$ = 2.0 calls/sec * 300 sec = 600 calls

Step 3: Calculate the maximum number of concurrent calls in the system:

TotalConcurrentCalls = 3,600 + 2,400 + 600 = 6,600 calls

To keep the number of sessions for a single instance of SIP Server below the acceptable maximum, we would need two SIP Servers and a Network SIP Server:

nSIPServers = Max Integer (6,600 / 4,000) = 2

The average number of concurrent sessions per instance is:

(6,600 / 2) = 3,300.

Step 4: Calculate the effective call rate per scenario:

Because we are using multiple SIP Servers, we must calculate the effective call rate per instance (assuming even call distribution across instances):

a. Per SIP Inbound Call scenario:

EffectiveCallRate1 = 20.0/2 = 10.0 calls/sec

b. Per Single-Step Transfer scenario:

EffectiveCallRate2 = 10.0/2 = 5.0 calls/sec

c. Per Consultation Call scenario:

EffectiveCallRate3 = 2.0/2 = 1.0 call/sec

Step 5: Estimate CPU usage per scenario:

Using the calibration table for CPU usage (Table 123 on page 394), we get the following estimates for CPU usage per scenario:

a. Per SIP Inbound Call scenario:

CPU1 = 0.8 (% * sec/call) * 10.0 calls/sec = 8.0%

b. Per Single-Step Transfer scenario:

CPU2 = 1.49 (% * sec/call) * 5.0 calls/sec = 7.45%

c. Per Consultation Call scenario:

CPU3 = 2.17 (% * sec/call)* 1.0 call/sec = 2.17%

So, the expected TotalCPU = 8.0 + 7.45 + 2.17 = 17.62%

Because we are using a Network SIP Server for call distribution, we must factor the CPU usage according to Table 124 on page 396:

CorrectedTotalCPU = 1.20 * 17.62 = 21.14%

However, the calibration table for CPU usage assumes an Intel CPU running at 3.0 GHz, and we are using a 2.66 GHz CPU. Therefore, we must scale the result for the higher clock frequency:

CPU = 21.14 * (3.0/2.66) = 23.84%

Network Traffic Estimation

To estimate the amount of network traffic (per instance), use the calibration table for network traffic (Table 125), summing the values for individual scenarios to get the following results (for SIP traffic per SIP Server instance):

- For received (RX) traffic:

  RX = 7.3 * 10.0 + 14.2 * 5.0 + 20.4 * 1.0 = 164.4 (KB/sec)

- For transmitted (TX) traffic:

  TX = 51.8 * 10.0 + 75.8 * 5.0 + 132.0 * 1.0 = 1042.5 (KB/sec)

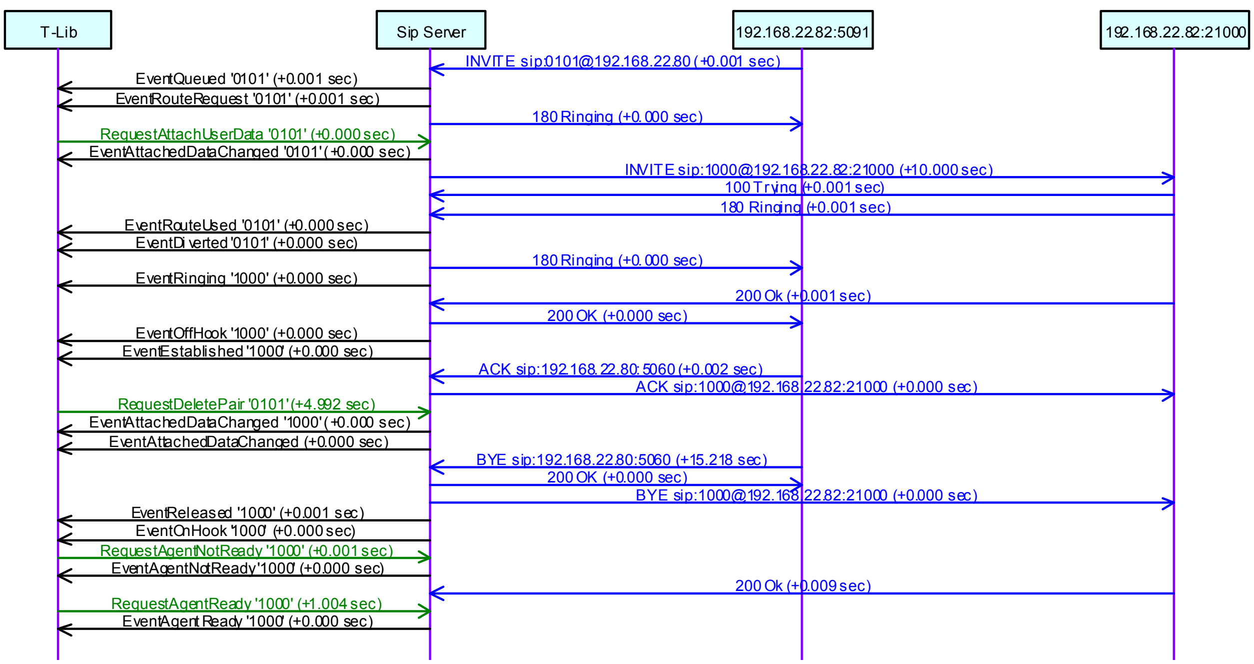
Note: The amount of traffic via host network interfaces depends on the particular placement of SIP Server instances.

SIP Server 8.0: Reference Call Flows

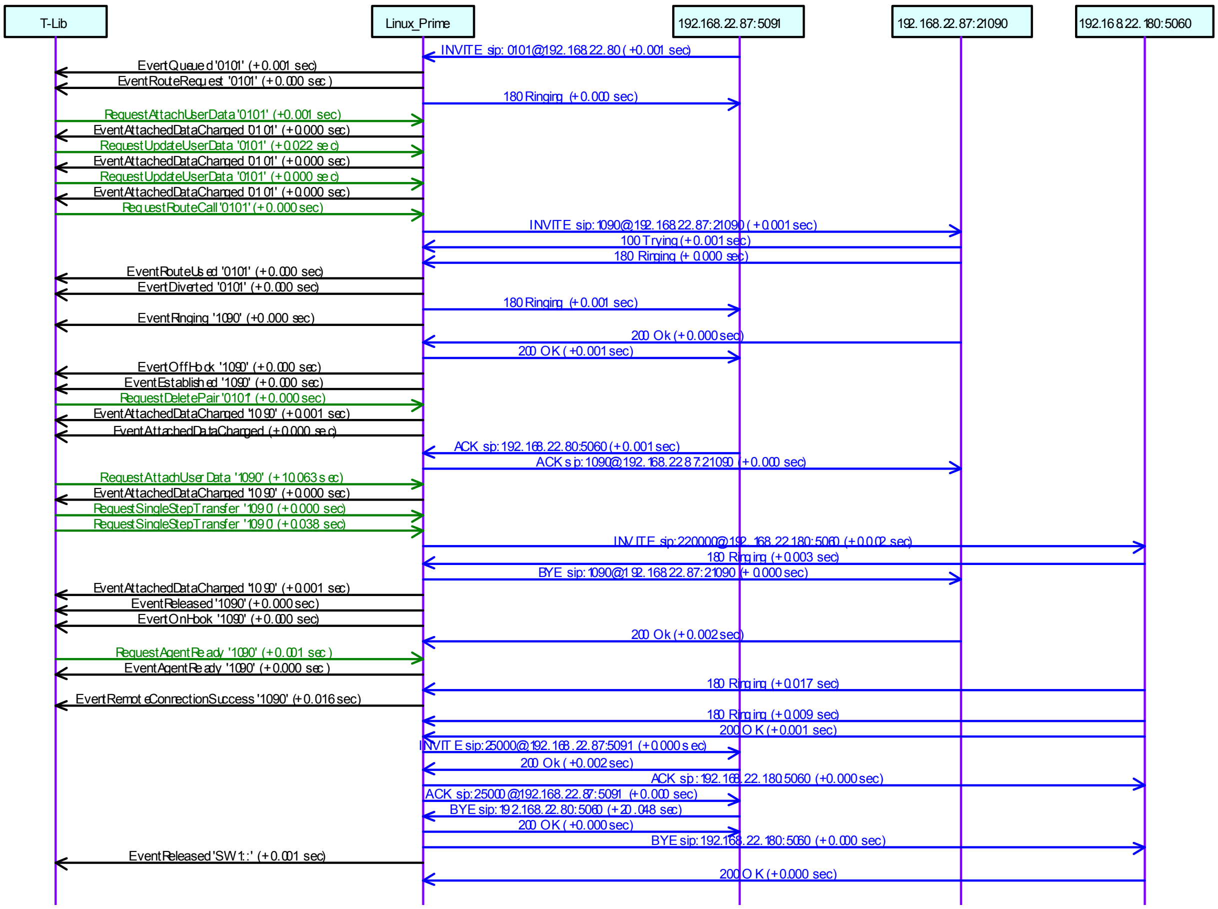
This section provides reference call flows and details of SIP messages that were used for benchmarking of the following call scenarios:

- "SIP Inbound Call" on page 402
- "SIP Inbound Call with Single-Step Transfer" on page 403
- "SIP Inbound Call with Recording" on page 404
- "SIP Inbound Call with Single-Step Conference" on page 405
- "SIP Inbound Call and Consultation Call" on page 406
- "SIP Inbound Call with Treatment" on page 407
- "Internal Call" on page 408

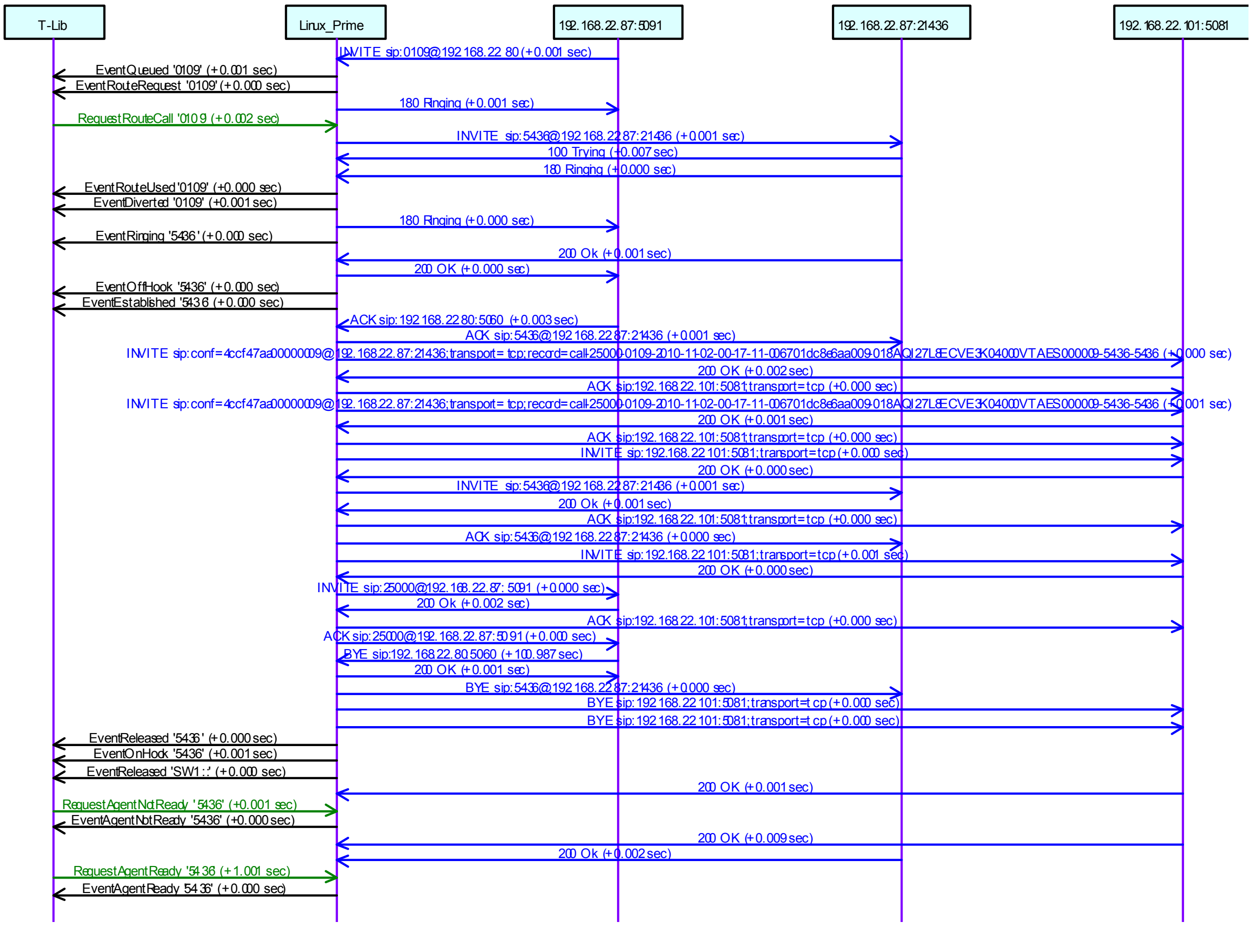
SIP Inbound Call



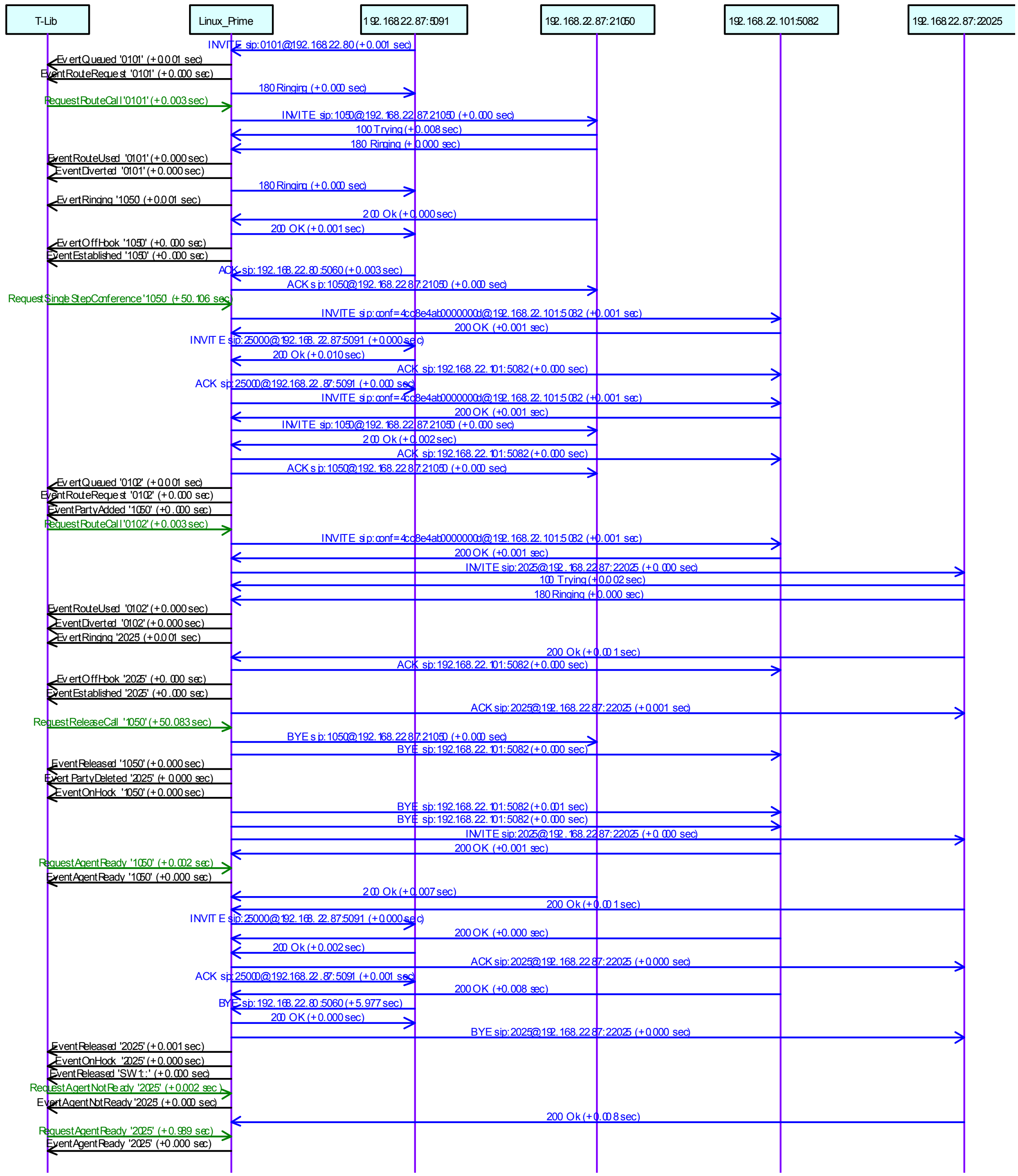
SIP Inbound Call with Single-Step Transfer



SIP Inbound Call with Recording

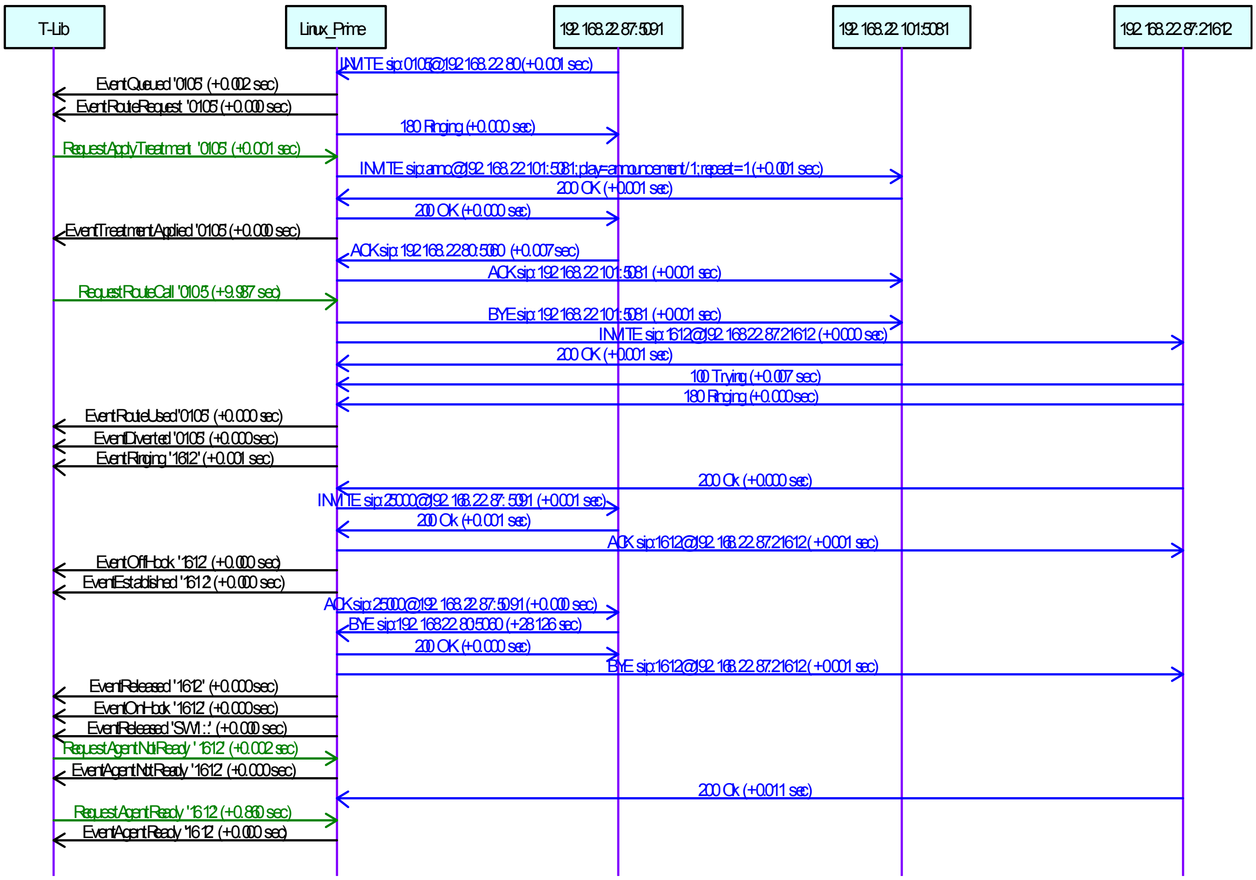


SIP Inbound Call with Single-Step Conference

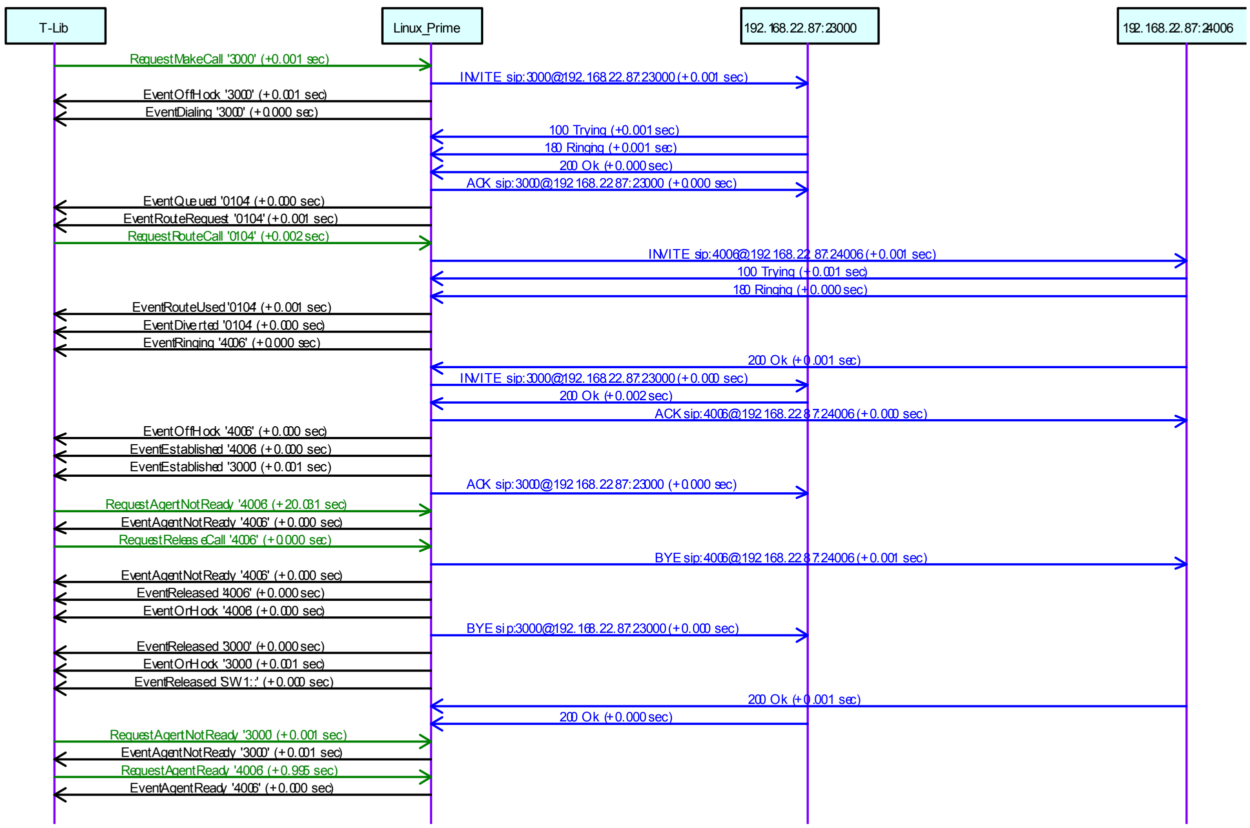


SIP Inbound Call and Consultation Call

SIP Inbound Call with Treatment



Internal Call



Genesys Media Server Sizing with SIP Server

Table 126 describes the capacity testing for Genesys Media Server 8.1.3 or later. Tests were performed by using a single instance of the Media Server on Windows and Linux systems with 2x Core 2 Quad, Xeon x5355, 2.66 GHz CPUs.

For GVP sizing information, see the chapter “Genesys Voice Platform 8.1,” in this guide.

Table 126: Genesys Media Server Capacity Testing

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
Windows				
NETANN Call Recording 2 participants Codec G.711u ~60 second duration	2x Core2Quad Xeon x5355 2.66GHz	12	720 call legs 360 recording sessions	Tested on Windows 2003.
NETANN Call Recording 2 participants Codec G.729 60 second duration	2x Core2Quad Xeon x5355 2.66GHz	9	540 call legs 270 recording sessions	Tested on Windows 2003.
NETANN Call Recording 2 participants Codec GSM ~60 second duration	2x Core2Quad Xeon x5355 2.66GHz	8	480 call legs 240 recording sessions	Tested on Windows 2003.
MSML Play Announcement (Codec G711, 120 seconds duration)	2x Quad-Core Xeon E5620 2.40GHz	50	6000 calls	Tested on 4 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (MP3, 92Kbit, 32KHz 120 seconds duration)	1x Hex-Core Xeon X5670 2.93GHz	13	1560 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (MP3, 320Kbit, 44.1KHz 120 seconds duration)	1x Hex-Core Xeon X5670 2.93GHz	12	1440 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.

Table 126: Genesys Media Server Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
MSML Play Announcement (Codec H263 and AMR, CIF, 128Kbps 10fps, 60 seconds duration)	1x Hex-Core Xeon X5670 2.93GHz	25	1500 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec AMR and H263 CIF, 512Kbps 30fps, 60 seconds duration)	1x Hex-Core Xeon X5670 2.93GHz	8.5	500 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H263 and AMR, 4CIF, 512Kbps 10fps, 60 seconds duration)	1x Hex-Core Xeon X5670 2.93GHz	23	1380 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H263 and AMR, 4CIF, 2Mbps 30fps, 60 seconds duration)	1x Hex-Core Xeon X5670 2.93GHz	8	480 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, CIF, 128Kbps 10fps, 60 seconds duration)	1x Hex-Core Xeon X5670 2.93GHz	25	1500 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, CIF, 256Kbps 15fps, 60 seconds duration)	1x Hex-Core Xeon X5670 2.93GHz	17	1000 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.

Table 126: Genesys Media Server Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
MSML Play Announcement (Codec H264 and AMR, CIF, 512Kbps 30fps, 60 seconds duration)	1x Hex-Core Xeon X5670 2.93GHz	8.5	500 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, 4CIF, 512Kbps 10fps, 60 seconds duration)	1x Hex-Core Xeon X5670 2.93GHz	22	1300 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, 4CIF, 1Mbps 15fps, 60 seconds duration)	1x Hex-Core Xeon X5670 2.93GHz	16	960 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, 4CIF, 2Mbps 30fps, 60 seconds duration)	1x Hex-Core Xeon X5670 2.93GHz	7.5	450 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, 720P, 1Mbps 10fps, 60 seconds duration)	1x Hex-Core Xeon X5670 2.93GHz	19	1100 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, 4CIF, 2Mbps 15fps, 60 seconds duration)	1x Hex-Core Xeon X5670 2.93GHz	9	540 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.

Table 126: Genesys Media Server Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
MSML Play Announcement (Codec H264 and AMR, 4CIF, 4Mbps 30fps, 60 seconds duration)	1x Hex-Core Xeon X5670 2.93GHz	4	240 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play and Digit Connect (Codec G711 and SIP INFO Digit, 34 seconds duration)	1x Hex-Core Xeon X5670 2.93GHz	50	1700 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Conference 3 participants using the same codec Codec G.711, G.729, or GSM ~60 second duration	2x Core2Quad Xeon x5355 2.66GHz	6	360 call legs 120 conference sessions	The capacity is the same for all three codecs. Tested on Windows 2003.
MSML Conference 3 participants using different codecs Codec G.711 and G.729 ~60 second duration	2x Core2Quad Xeon x5355 2.66GHz	6	360 call legs 120 conference sessions	Tested on Windows 2003.
MSML Conference 3 participants using different codecs Codec G.711 and GSM ~60 second duration	2x Core2Quad Xeon x5355 2.66GHz	6	360 call legs 120 conference sessions	Tested on Windows 2003.
MSML Conference (32 participants per conference, all speakers. Each participant stays and speaks (300 secs in the conference. Codec G.711)	2x Quad-Core Xeon E5620 2.40GHz	2.6	768 participants (24 conference sessions)	Tested on 4 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.

Table 126: Genesys Media Server Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
MSML Conference (One giant conference that only 3 speakers and all other participants are listeners. Each participant stays 1800 secs (30 mins) in the conference. Codec G.711)	1x Hex-Core Xeon X5675 3.06GHz	0.72	1300 participants (1 conference session)	Tested on Windows 2008 Server R2 x64 SP1 with only one MCP instance. Threaded outputs enabled (conference.threadedoutputs = true).
MSML Conference (One giant conference that only 3 speakers and all other participants are listeners. Each participant stays 1800 secs (30 mins) in the conference. Codec H263 + G.711)	1x Hex-Core Xeon X5675 3.06GHz	0.61	1100 participants (1 conference session)	Tested on Windows 2008 Server R2 x64 SP1 with only one MCP instance. Threaded outputs enabled (conference.threadedoutputs = true).
Linux				
NETANN Play Treatment Codec G.711u ~60 second duration	2x Core2Quad Xeon x5355 2.66GHz	30	1800	No transcoding.
NETANN Play Treatment Video Codec H.263 (+) ~120 seconds duration	2x Core2Quad Xeon x5355 2.66GHz	10	1200	No transcoding.
NETANN Play Treatment Video 3 gp or .avi Codec H.263 ~120 seconds duration	2x Core2Quad Xeon x5355 2.66GHz	8.3	1000	No transcoding.

Table 126: Genesys Media Server Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
NETANN Recording Single Call Codec G.711u (raw, .au & .wav), G.722, or G.726 ~120 seconds duration	2x Core2Quad Xeon x5355 2.66GHz	8.3	1000	The capacity is the same for G.711u, G.722, and G.726.
NETANN Recording Single Call Codec G.729 or AMR ~120 seconds duration	2x Core2Quad Xeon x5355 2.66GHz	5.8	700	The capacity is the same for G.729 and AMR.
NETANN Recording Single Call Codec AMR-WB ~120 seconds duration	2x Core2Quad Xeon x5355 2.66GHz	6.6	800	
NETANN Recording Single Call Video (raw) Codec H.263 (+) ~120 seconds duration	2x Core2Quad Xeon x5355 2.66GHz	4.2	500	
NETANN Recording Single Call Video avi Codec H.263 (+) or G.711u ~120 seconds duration	2x Core2Quad Xeon x5355 2.66GHz	4	480	

Table 126: Genesys Media Server Capacity Testing (Continued)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
NETANN Recording Single Call Video 3 gp Codec H.263 (+), or AMR ~120 seconds duration	2x Core2Quad Xeon x5355 2.66GHz	2	240	
NETANN Recording Single Call Video (raw) Codec H.264 ~120 seconds duration	2x Core2Quad Xeon x5355 2.66GHz	2	250	
NETANN Call Recording 2 participants Codec G.711u ~60 second duration	2x Core2Quad Xeon x5355 2.66GHz	11	660 call legs 330 recording sessions	
MSML Play Announcement 1 prompt (SIP INFO) 1 audio file Codec N/A ~3 second duration	2x Core2Quad Xeon x5355 2.66GHz	80	260	Total call duration = 3.13 seconds (gvp.precheck is on).
MSML Conference 3 participants using the same codec Codec G.711u ~60 second duration	2x Core2Quad Xeon x5355 2.66GHz	6	360 call legs 120 conference sessions	The capacity is the same for G.711u, G.729, or GSM.
Note: *Preferred* means the highest capacity that the system can sustain while maintaining optimal user experience. *Peak* means the highest capacity that the system can sustain regardless of the user experience.				

Genesys SIP Voicemail 8.1

Genesys SIP Voicemail 8.1.0 is recommended for deployment on a single dedicated server.

Sizing guidance is provided below for two configurations.

Table 127: Genesys SIP Voicemail 8.1.0 Sizing

Configuration	Required CPU [a]	Required RAM	Suggested Storage for Messages and Operations [b]	Mailbox Limit	Simultaneous Voice Connections Limit
1	1 x Quad-Core	8 GB	80 GB	5,000	100
2	2 x Quad-Core	16 GB	240 GB	15,000	150

a. CPU should be 2+ GHz on Intel Xeon E5405 or similar, to achieve stated limits.

b. Suggested storage for messages and operations is an estimate based on 15 messages per mailbox and 30 second average message length.

Chapter

13 Performance Management Advisors

This chapter provides recommendations for hardware sizing for typical contact center scenarios. It contains the following sections:

- Overview, page 417
- Performance Considerations, page 418
- General Guidelines for Contact Center Sizing Categories, page 420
- Stat Server Sizing, page 424
- Advisors Genesys Adapter Performance Information, page 424
- Apache Tuning Tips, page 429
- Desktop, page 430
- Capacity, Measurement, and Sample Architecture, page 430

Overview

The following acronyms are used:

- CCAdv—Contact Center Advisor
- WA—Workforce Option/Workforce Advisor
- FA—Frontline Advisor
- AA—Agent Advisor
- AGA—Advisors Genesys Adapter
- AM—Administration Module
- ACA—Advisors Cisco Adapter
- SDS—Supervisor Desktop Service
- RMC—Resource Management Console
- ActionMgt—Action Management
- CCAdv-ME—Contact Center Advisor - Mobile Edition

For more information about terminology and concepts used in this chapter, see:

- *Genesys Performance Management Advisors Deployment Guide*
- *Genesys Performance Management Advisors Frontline Advisor Administration User's Guide*
- *Genesys Performance Management Advisors Contact Center Advisor and Workforce Advisor Administrator User's Guide*.

Performance Considerations

A key performance measurement is the number of concurrent dashboard users (that is, the load-carrying capacity) on a specific deployment architecture. The hardware requirements of the different products within the suite depend on a number of factors that impact this performance measurement.

Contact Center Advisor and Workforce Advisor

The performance of Contact Center Advisor (CCAdv) and Workforce Advisor (WA) are not tied directly to the number of calls handled by the underlying platform(s). Instead, their performance depends on the complexity of the configured hierarchy and the number of statistics handled. The number of underlying base objects (queues, agent groups, and agents) that are being monitored, and their relationships to each other, determine the performance of these applications. This is further complicated when you use filters to segment the data for a given base object.

Table 128 shows the default number of statistics that are requested by the Advisors Genesys Adapter (AGA) for each type of base object (if the base objects are not segmented by filters) when CCAdv and WA are deployed on a Genesys platform. Note the following:

- These are the numbers of statistics that are requested by default (out-of-box). Additional statistics can be enabled for a specific deployment. There might be fewer default metrics in later releases because, with improvements to the Metrics Manager, you can create more custom metrics that better reflect the needs of your enterprise.
- WA contact group metrics are not counted in this type of stat server load sizing.

Table 128: CCAdv/WA Source Metrics

Release	Agent Group Voice	Agent Group Multimedia	Application Voice	Application Multimedia	Agent
8.1.5	24	50	41	35	3
8.5.0	39	29	49	16	3
8.5.1	39	37	50	23	3

Frontline Advisor and Agent Advisor

The performance of Frontline Advisor (FA) and Agent Advisor (AA) are not tied directly to the number of calls handled by the underlying platform(s). Instead, their performance depends on the number of agents that are being monitored, the number of rules that have been activated for each agent, and the depth of the organizational hierarchy.

Table 129 shows the default number of statistics that are requested by the Advisors Genesys Adapter (AGA) when FA is deployed on a Genesys platform. Note the following:

- These are the numbers of statistics that are requested by default (out-of-box). Additional statistics can be enabled for a specific deployment. There might be fewer default metrics in later releases because, with improvements to the Metrics Manager, you can create more custom metrics that better reflect the needs of your enterprise.
- FA rule metrics are not enabled out-of-box; if you enable rule metrics, then you can have up to 12.
- The count of Agent Performance voice and multimedia source metrics in Table 129 reflect one time profile enabled out-of-box. If you enable additional FA time profiles, performance and rule metrics are multiplied by the number of enabled time profiles to get the total number.

Table 129: FA Source Metrics

Release	Agent State	Agent Performance Voice	Agent Performance Multimedia
8.1.5	5	19	16
8.5.0	6	19	14
8.5.1	6	20	22

Advisors Genesys Adapter

The performance of the Advisors Genesys Adapter (AGA) depends mostly on the number of statistics it is handling and the number of base objects (queues, agent groups, and agents) configured in the Configuration Server. In releases earlier than 8.5.1, the AGA extracts these objects from the Configuration Server on start-up and stores them in its memory, therefore, a large configuration requires the AGA be allocated extensive amounts of memory.

The default value for the AGA maximum heap space size is 1 GB. Consider increasing this size for the larger deployments.

The number of statistics that the AGA is handling depends on the set of selected base objects and whether the AGA is serving CCAdv/WA or FAAA. (Note that a single instance of AGA cannot serve both CCAdv and FA.)

The performance of the AGA also partly depends on the call volume since the number of T-Events being generated in the Platform drives the number of updates being sent from the Stat Server to the AGA.

Cisco Adapter The performance of the Advisors Cisco Adapter (ACA) depends on both the call volume and the number of agents that are being monitored. ACA is designed to work only with FA/AA, hence the number of base objects being monitored in CCAdv has no effect on ACA. The ACA works off the call records retrieved from the underlying Cisco HDS database. The more calls going through the Platform, the more records the ACA must process to extract the statistics required by FAAA.

Summary Table 130 summarizes the factors that impact performance for each Advisors component.

Table 130: Factors Impacting Performance of Advisors Components

Product	Hierarchy Complexity	Statistics	Base Objects	Filters	Agents	Rules	Call Volume	Metric Graphing
CCAdv/WA	X	X	X	X	X			X
FAAA	X				X	X		
AGA		X	X		X		X	
ACA					X		X	

Keeping all these considerations in mind, the information is organized according to the size of the contact center as a function of the number of base objects being monitored and the number of calls that are flowing through the platform on a daily basis.

General Guidelines for Contact Center Sizing Categories

Table 131 shows the contact center sizing categories based on the number of base objects being monitored and the daily call volume.

Table 131: Contact Center Sizing Categories

Size Category	Number of Agents	Number of Agent Groups	Number of Queues	Daily Call Volume
Small	Fewer than 500	Fewer than 50	Fewer than 50	Of the order of tens of thousands

Table 131: Contact Center Sizing Categories (Continued)

Size Category	Number of Agents	Number of Agent Groups	Number of Queues	Daily Call Volume
Medium	Fewer than 5000	Fewer than 400	Fewer than 1000	Up to 500,000
Large	Fewer than 30,000	Fewer than 1000	Fewer than 8000	Up to 4 million

Example Configurations for Contact Centers based on Size

The following are examples of possible configurations based on contact center size. You can use these examples as general guidelines when deploying the full Advisors suite, particularly for Advisors releases prior to 8.1.5. The examples are based on servers running Windows operating systems, but – starting with Performance Management Advisors release 8.5.0 – you can deploy Advisors components on Red Hat Enterprise Linux 5. See the Performance Management Advisors page in the *Genesys Supported Operating Environment Reference Guide* for a list of supported operating systems.

See also "Capacity, Measurement, and Sample Architecture" on page 430, which provides performance information from tested environments running Advisors release 8.1.5 software. "Capacity, Measurement, and Sample Architecture" discusses each Advisors component separately (CCAdv, WA, and FA) and provides specific deployment architectures for each to successfully achieve 1500 concurrent dashboard users.

Note: In the configurations listed below, FAAA running on a Cisco platform using the ACA has not been shown. If you have a Cisco environment and wish to use FAAA, a separate instance of FAAA needs to be installed along with an instance of the ACA. Hence, the hardware requirements shown in this section for FA and the AGA (for FA) will need to be duplicated.

Contact Center Size—Small

Table 132 shows an example of the Contact Center Size-Small architecture.

Note: In this architecture, there is no separate server for the Web tier. Apache is deployed on one of the servers hosting the applications.

Table 132: Contact Center Size—Small

Server Number	Application Component(s)	Processor(s)	Memory	Hard Drive Space
1	Platform, AM, CCAdv, WA, XMLGen, AGA, ActionMgt, RMC	Quad-core 2.0 GHz+	4 GB	10 GB
2	CCAdv-ME	Dual-core 2.0 GHz+	4 GB	5 GB
3	Apache, Platform, FA, AGA	Quad-core 2.0 GHz+	4 GB	10 GB
4	Databases	Dual Quad-core 2.0 GHz+	4 GB+	30 GB
5	SDS	Quad-core 2.0 GHz+	4 GB+	10 GB

Contact Center Size—Medium

Table 133 shows an example of the Contact Center Size-Medium architecture.

Note: In this architecture, you separate the major application, database, and Apache installations.

Table 133: Contact Center Size—Medium

Server Number	Application Component(s)	Processor(s)	Memory	Hard Drive Space
1	Apache Web Server	Dual-core 1.86 GHz+	512+ MB	5 GB
2	Platform, AM, CCAdv, ActionMgt, RMC	Dual Quad-core 2.0 GHz+	4 GB	10 GB

Table 133: Contact Center Size—Medium (Continued)

Server Number	Application Component(s)	Processor(s)	Memory	Hard Drive Space
3	Platform, XMLGen, WA	Dual Quad-core 2.0 GHz+	4 GB	10 GB
4	CCAdv-ME	Dual-core 2.0 GHz+	2 GB	5 GB
5	AGA (for CCAdv), AGA (for FA)	Dual Quad-core 2.0 GHz+	4 GB	10 GB
6	Platform, FA	Dual Quad-core 2.0 GHz+	4 GB	10 GB
7	Databases	Dual Quad-core 2.0 GHz+	4 GB+	50 GB
8	SDS	Quad-core 2.0 GHz+	6 GB+	10 GB

Contact Center Size—Large

Table 134 shows an example of the Contact Center Size-Large architecture.

Table 134: Contact Center Size—Large

Server Number	Application Component(s)	Processor(s)	Memory	Hard Drive Space
1	Apache Web Server	Dual-core 1.86 GHz+	8 GB	5 GB
2	Platform, AM, CCAdv, ActionMgt, RMC	Dual Quad-core 2.83 GHz+	16 GB	20 GB
3	Platform, XMLGen, WA	Dual Quad-core 2.83 GHz+	16 GB	20 GB
4	CCAdv-ME	Dual-core 2.0 GHz+	4 GB	5 GB

Table 134: Contact Center Size—Large (Continued)

Server Number	Application Component(s)	Processor(s)	Memory	Hard Drive Space
5	AGA (for CCAdv)	Dual Quad-core 2.83 GHz+	16 GB	20 GB
6	AGA (for FA)	Dual Quad-core 2.83 GHz+	16 GB	20 GB
7	Platform, FA	Dual Quad-core 2.83 GHz+	16 GB	10 GB
8	Databases	Dual Quad-core 3.0 GHz+	32 GB	80 GB
9	SDS	Quad-core 3.0 GHz+	8 GB+	10 GB

Stat Server Sizing

The processing part of the Stat Server application is single-threaded; therefore, Genesys recommends that you allocate one complete core on a server to each Stat Server application, without any other processes also using that core.

On a multi-core server machine, you can install more than one Stat Server application, but not more than the number of cores. Genesys recommends that you allow one extra core for the operating system activities. You can install additional components on the same multi-core machine only if additional cores and additional resources are available. Allow at least 2 GB of memory for each deployed Stat Server.

Advisors Genesys Adapter Performance Information

The following performance results were achieved with Advisors Genesys Adapter software release 8.5.1. To avoid AGA performance issues, both the statistics load and maximum message rate must be below the thresholds described in the following Table. See “Determining Message Rates” on page 426 and “Estimating the Number of Requested Statistics for Frontline Advisor” on page 427 for more information.

Table 135: AGA Performance Information

Advisors Application	Configuration	Message Rate	Stats Load	Notes and Recommendations
Contact Center Advisor	One AGA and one Stat Server	Maximum message rate under 45 000 messages/ second Average message rate under 6500 messages/second	AGA can process up to 836 000 statistics without performance degradation. To generate 836 000 statistics, Genesys used the following configuration during testing: • Agent monitoring: On • Number of queues: 550 (with 3 time profiles) • Number of agent groups: 23 352	If your statistics load is more than 836 000, then add one more Stat Server. If your maximum message rate exceeds 45 000 messages/ second, then add one more AGA.
Contact Center Advisor	One AGA and two Stat Servers	Maximum message rate under 45 000 messages/ second Average message rate under 6500 messages/second	AGA can process up to 891 000 statistics without performance degradation. To generate 891 000 statistics, Genesys used the following configuration during testing: • Agent monitoring: On • Number of queues: 550 (with 3 time profiles) • Number of agent groups: 25 396	If your statistics load is more than 891 000, then add one more AGA. If your maximum message rate exceeds 45 000 messages/ second, then add one more AGA.

Table 135: AGA Performance Information (Continued)

Advisors Application	Configuration	Message Rate	Stats Load	Notes and Recommendations
Frontline Advisor	One AGA and one Stat Server	Maximum message rate under 70 000 messages/ second Average message rate under 18 000 messages/second	AGA can process up to 1 200 000 statistics without performance degradation. To generate 1 200 000 statistics, Genesys used the following configuration during testing: • Three time profiles enabled • FA hierarchy: 11 320 agents	If your statistics load is more than 1 200 000, then add one more Stat Server. If your maximum message rate exceeds 70 000 messages/ second, then add one more AGA.
Frontline Advisor	One AGA and two Stat Servers	Maximum message rate under 70 000 messages/ second Average message rate under 18 000 messages/second	AGA can process up to 1 450 000 statistics without performance degradation. To generate 1 450 000 statistics, Genesys used the following configuration during testing: • Three time profiles enabled • FA hierarchy: 13 680 agents	If your statistics load is more than 1 450 000, then add one more AGA. If your maximum message rate exceeds 70 000 messages/ second, then add one more AGA.

Determining Message Rates

It is important to know how many messages arrive per second at the AGA from Stat Server to ensure you have a sufficient number of adapters deployed to handle the load. Use the following procedure to determine the message rates for your AGA.

Procedure:
Calculating the Message Rates

Start of procedure

1. Locate your log4j.properties file in the <*AGA installation*>\conf folder.
2. Change the log4j.logger.timing log entry to:
   log4j.logger.timing=debug, timingLog
3. Check for new entries in the timing.log file. You should see the following type of entries:
   ... DEBUG timing - For object type: XXXX received xxxx messages from SS:xxx ...
4. Wait until AGA accumulates sufficient data, and then you can determine the maximum number of messages received per second from Stat Server. You can also calculate the average number of messages received per second from Stat Server.

End of procedure

Estimating the Number of Requested Statistics for Frontline Advisor

You can evaluate the number of statistics requested for each agent in Frontline Advisor. When you have an estimate of the number of requested statistics, you can also estimate the following:

- the number of Stat Servers to deploy
- the FA statistics load on each Stat Server

The number of statistics requested for Frontline Advisor depends on the following:

- the number of time profiles currently enabled
- the number of report metrics currently enabled, including all dependencies that are not enabled
- the number of agents currently logged on

Initially, estimate the load based on the enabled default metrics, or based on the metrics that are enabled in the migrated environment. If you change the number of enabled metrics after Advisors installation or migration, or if you make changes to time profiles, then use the same process to re-evaluate the load.

If necessary, you can do a post-installation adjustment of the Stat Server configuration to achieve optimal performance for your enterprise. See the *Genesys Performance Management Advisors Deployment Guide* for information about Advisors Stat Server configuration.

Procedure: Estimating the Number of Statistics Requested for Each Agent in Frontline Advisor

Start of procedure

1. Use the following queries to determine the number of statistics requested for each agent in Frontline Advisor:

   SELECT COUNT (*) "Perf. metrics" from fa_vw_performance_source;

   SELECT COUNT (*) "Metrics for rules p/agent" from fa_vw_rule_source;

   SELECT COUNT (*) "State metrics p/agent" from fa_vw_state_source;

   The first query (SELECT COUNT (*) "Perf. metrics" from fa_vw_performance_source;) provides the number of performance metrics. The second query (SELECT COUNT (*) "Metrics for rules p/agent" from fa_vw_rule_source;) provides the number of rules metrics. The third query (SELECT COUNT (*) "State metrics p/agent" from fa_vw_state_source;) provides the number of state metrics.

2. Use the following calculation to estimate the overall, real-time number of statistics that will be sent to the configured Stat Servers:

   *<Number of statistics obtained from running the queries> * <Average number of agents typically logged on to the system>*

End of procedure

Next Steps

For additional, detailed information, use the following queries.

- Number of Performance metrics grouped by time profile:

  SELECT COUNT (*) "Perf. metrics p/tprofile",type "Time Profile Type",interval "Time Profile Interval" from fa_vw_performance_source group by type,interval;

- Number of Performance metrics grouped by channel:

  SELECT COUNT (*) "Perf. metrics p/channel",m1.Channel "Channel" from fa_vw_performance_source vws

  JOIN fa_vw_metrics vwm

  ON vws.AppSpecificId=vwm.AppSpecificId

  JOIN METRICS_1 m1

  ON m1.METRIC_ID=vwm.BaseMetricId

  GROUP BY m1.Channel;

- Number of Performance metrics grouped by time profile and channel:

  SELECT COUNT (*) "Perf. m p/channel/tprofile",m1.Channel "Channel",type "Time Profile Type",interval "Time Profile Interval" from fa_vw_performance_source vws

  JOIN fa_vw_metrics vwm

  ON vws.AppSpecificId=vwm.AppSpecificId

```
JOIN METRICS_1 m1
ON m1.METRIC_ID=vwm.BaseMetricId
GROUP BY m1.Channel,vws.type,vws.interval;
```

Apache Tuning Tips

There are some useful Apache tuning tips available at http://www.devside.net/articles/apache-performance-tuning.

XML Generator Compression Options for Apache

Depending on the complexity of your Contact Center Advisor (CCAdv) hierarchy, enabling Apache compression of XML files might be necessary to improve performance. Genesys Professional Services can advise you on the need for such compression in your enterprise. If required, use the following procedure.

Procedure: Apache Compression for XML Files

Purpose: Enable compression of XML files to reduce payload size and download times for the CCAdv dashboard updates.

Start of procedure

In the httpd.conf file of each Apache server installation (the httpd.conf file is located in the conf folder of the Apache Web Server installation), do the following:

1. Make sure that the following line is not commented out (that is, the line must not have the # preceding it):
   ```
   LoadModule deflate_module modules/mod_deflate.so
   ```
2. Deflate (compress) only the files associated with the ca-xml context root by specifying the /ca-xml/ location after the ProxyPass statement. See the “Example” on page 429.

End of procedure

Example

Add the following section after the /ca-xml/ ProxyPass statement to enable Apache compression of the XML files:

```
<Location "/ca-xml/">
  SetOutputFilter DEFLATE
```

```
</Location>
```

For example:

```
ProxyPass /ca-xml/ ajp://host:8009/ca-xml/
<Location "/ca-xml/">
  SetOutputFilter DEFLATE
</Location>
```

Desktop

The recommended requirements for a system running the Advisors dashboards are as follows:

- Processor—Dual-core 2.0 GHz
- Memory—2 GB RAM
- Hard Drive—1 GB

Prior to Advisors release 8.5.0, users access Advisors dashboards using a client application called the Advisors browser, which you install on each user's computer. You must also install the Flash player plug-in for non-IE browsers (for example, Firefox) to use the Advisors browser.

Starting with Advisors release 8.5.0, users access the application using any of the supported standard web browsers. See *Genesys Supported Operating Environment Reference Guide* for information about specific supported browsers and browser versions. You must also install the Flash Player plug-in appropriate for your particular browser.

Note: Contact center size makes no difference to desktop usage because the Advisors dashboards are accessed on the end-user’s local machine.

Capacity, Measurement, and Sample Architecture

Genesys tested the information in this section using Advisors release 8.1.5 on Microsoft Windows servers. This section includes the following:

- Performance Measurement Environment, page 431
- Load-Carrying Capacity, page 431
  - CCAdv Deployment Architecture and Recommendations for Optimal Performance, page 432
  - WA Deployment Architecture and Recommendations for Optimal Performance, page 440
  - FA Deployment Architecture and Recommendations for Optimal Performance, page 449

- Improving Supervisor Desktop Service Performance, page 457

Performance Measurement Environment

The key measurement for Performance Management Advisors products is the number of concurrent dashboard users (that is, the load-carrying capacity) on a specific deployment architecture.

The *deployment architecture* is a combination of the following:

- Object count
- Metric count
- Hardware capacity
- Product Configuration

Object count for a given product includes a subset of the number of Geographic Regions, Reporting Regions, Operating Units, Contact Centers, Agent Groups, Application Groups, and Applications. The layout of the object hierarchy (that is, the number and nature of the objects, as well as the nature of the relationships) may impact the performance of Advisor products, but the number of objects is the main factor affecting performance.

The metric count includes both default and defined custom metrics. Metrics may affect Advisors performance, depending on the metrics definitions, but like object count, the variance is assumed to be minimal.

Unless otherwise specified, performance measurements rely on an environment in which each individual Advisor component is running in its own operating environment, which includes specific hardware (physical or virtualized) and operating system.

The sizing information provided is applicable only to the specific deployment architecture described in this section. For each of the Advisor products, there is also a *best practices* section that contains general guidelines for performance optimization.

Load-Carrying Capacity

Load-carrying capacity is the number of concurrent dashboard users without significant performance degradation in the deployment architecture.

You can scale up the load-carrying capacity of Advisor products by increasing the number of presentation instances (presentation instances service the dashboard requests).

Table 136, “Presentation Load-Carrying Capacity,” on page 432 lists the load-carrying capacity for each product based on the following configurations:

- Single-Presentation instance load carrying capacity: The user load that a single independent (with distributed cache) presentation instance can carry in Genesys’ recommended deployment architecture.

- Dual-Presentation instance load carrying capacity: The user load that a pair of independent (with distributed cache) presentation instances can carry in Genesys' recommended deployment architecture.
- Triple-Presentation instance load carrying capacity: The user load that three independent (with distributed cache) presentation instances can carry in Genesys' recommended deployment architecture.

The information in the following Table is limited to demonstrating the capability to support at least 1500 concurrent users for each product. The deployment architecture that was used to collect the key performance measurement for each product is described in the following sections:

- "CCAdv Deployment Architecture and Recommendations for Optimal Performance" on page 432
- "WA Deployment Architecture and Recommendations for Optimal Performance" on page 440
- "FA Deployment Architecture and Recommendations for Optimal Performance" on page 449

Table 136: Presentation Load-Carrying Capacity

Product	Single-Presentation Instance Load-carrying Capacity	Dual-Presentation Instance Load-carrying Capacity	Triple-Presentation Instance Load-carrying Capacity
CCAdv	600	1300	1600
WA	700	1200	1500
FA	1500	Not required; 1500 concurrent users achieved with a single-presentation instance.	Not required; 1500 concurrent users achieved with a single-presentation instance.

CCAdv Deployment Architecture and Recommendations for Optimal Performance

CCAdv Object Configuration Information

The following Table describes the high-level dimensions controlling the environment used to achieve the results described in "Load-Carrying Capacity" on page 431.

Table 137: CCAdv Presentation Object Configuration

Object	Count
Geographic Regions	1
Contact Centers	40

Table 137: CCAdv Presentation Object Configuration (Continued)

Object	Count
Reporting Regions	20
Operating Units	1
Application Groups	150
Applications	1600
Agent Groups	3200 (2 for each application)

CCAdv Environment Topology

The following diagram shows the environment topology used to successfully achieve 1500 concurrent users of CCAdv. VM is a *virtual machine*. The *Aggregation Instance* performs data aggregation; the *Presentation Instance* services the dashboard.

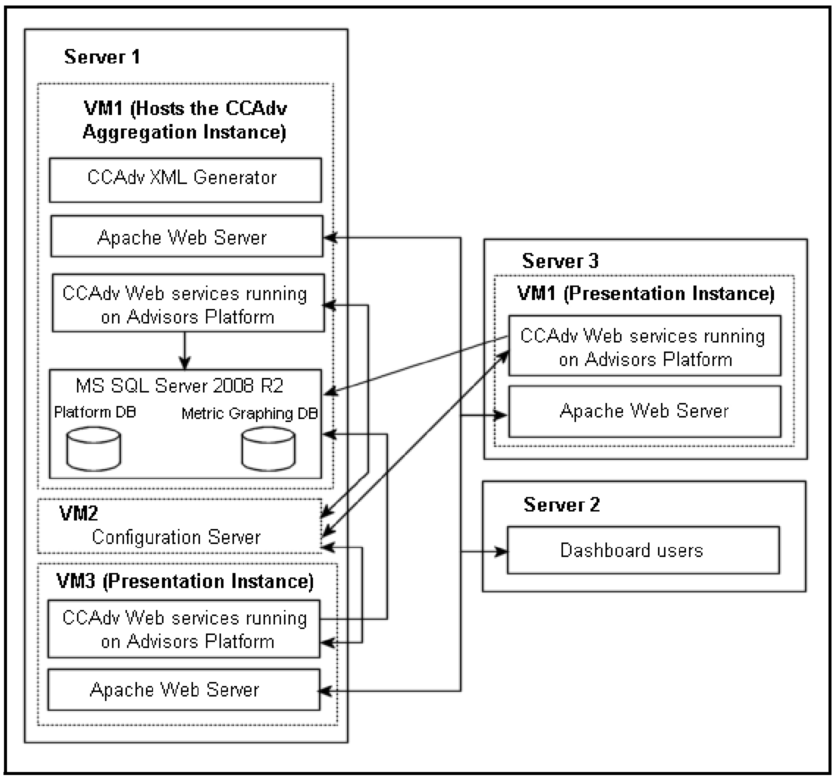


Figure 153: Contact Center Advisor Environment Topology

CCAdv Presentation Machine and VM Information

The following Table describes the characteristics of the hardware and virtualization environment shown in Figure 153, which supports 1500 concurrent users of CCAdv.

Table 138: CCAdv Presentation Machine and VM Information

Server	# of Processors	Processor Type	# of Cores	Total Cores	RAM	OS	Application
Server 1 VM-Host	2	Intel Xeon X5675 @ 3.07GHz	6	24 logical cores (with hyper-threading)	32 GB	VMWare ESXi 5.0 Standard 64bit	
Server-1 VM-1	2	(same as host)	4	8	12 GB	Windows Server 2008 Standard SP2 64bit	MS SQL (Advisor database), Apache Geronimo (CAXMLon Advisors Platform), XML Generator Simulator, Apache
Server-1 VM-2	2	(same as host)	1	2	1.5 GB	Windows Server 2003 Standard SP2 64bit **NOTE:** At the time of testing, Genesys supported Windows Server 2003. In your enterprise, be sure to use a Genesys-supported Windows Server operating system.	Configuration Server
Server-1 VM-3	2	(same as host)	4	8	8 GB	Windows Server 2008 R2 Standard SP1 64bit	Apache Geronimo (CAXML on Advisors Platform), Apache
Server 2	1	Intel Xeon X3440 @ 2.53GHz	4	4	8 GB	Windows Server 2008 R2 Standard SP1 64bit	Dashboard simulator

Table 138: CCAdv Presentation Machine and VM Information (Continued)

Server	# of Processors	Processor Type	# of Cores	Total Cores	RAM	OS	Application
Server 3 VM-Host	2	Intel Xeon X5675 @ 3.07GHz	6	24 logical cores (with hyper-threading)	32 GB	VMWare ESXi 5.0 Standard 64bit	
Server-3 VM-1	2	(same as host)	4	8	8 GB	Windows Server 2008 R2 Standard SP1 64bit	Apache Geronimo (CAXML on Advisors Platform), Apache

CCAdv Configuration for High Performance

The following Table describes settings you can change to improve CCAdv performance.

Table 139: Recommendations for Performance Improvement

Location	Sub-directory or File, where applicable	Settings
On each CCAdv node	<*CCAdv home*>\geronimo-tomcat6-minimal-2.2.1\bin\setenv.bat	Change GERONIMO_OPTS=-ms128m -mx1024m -XX:MaxPermSize=128m To GERONIMO_OPTS=-Xms6g -Xmx6g -XX:MaxPermSize=256m
	<*CCAdv home*>\geronimo\var\catalina\server.xml	Under the <Connectorname=”TomcatAJPConnector”> section, add maxThreads=”1600”
On CCAdv presentation nodes only	<*CCAdv home*>\geronimo\ var\config\config-substitutions.properties	Set MaxThreadPoolSize to “3000”
Dashboard Administration setting		For optimal performance: • Select independent configuration mode (not integrated configuration mode) • Set Show Totals and Averages to No

Table 139: Recommendations for Performance Improvement (Continued)

Location	Sub-directory or File, where applicable	Settings
Advisors Platform database		For optimal performance of CCAdv/WA, you can turn off agent monitoring. For release 8.1.5, see the *Disabling the agent level statistics templates for CCAdv* procedure in the *Performance Management Advisors 8.1 Deployment Guide*. For release 8.5.0 and later, see Enable and Disable Agent-level Monitoring in the *Performance Management Advisors Deployment Guide*.
On AGA	Releases 8.1.5 and 8.5.0 <*AGA home*>\conf\wrapper.conf Release 8.5.1 and later conf\run.bat (Windows) bin\setenv.sh (Linux)	Releases 8.1.5 and 8.5.0 Change Wrapper.java.initmemory=128 Wrapper.java.maxmemory=1024 To Wrapper.java.initmemory=4096 Wrapper.java.maxmemory=14336 Release 8.5.1 and later Change the following values in the JAVA_OPTS parameter: Change -ms128m -mx1024m To -ms4096m -mx14336m

Table 139: Recommendations for Performance Improvement (Continued)

Location	Sub-directory or File, where applicable	Settings
On AGA	<*AGA home*>\conf\ inf_genesys_adapter.properties	Change informiam.genesys_connector.stat Server.maxOpenReqsPerGroup = 1000 informiam.genesys_connector.stat Server.messages.queuesize = 500000 informiam.genesys_connector.GC StatisticsObjectDao.batchSize Overide = 100 informiam.genesys_connector.stats issue.pausechecklimit = 5000 To informiam.genesys_connector.stat Server.maxOpenReqsPerGroup = 6000 informiam.genesys_connector.stat Server.messages.queuesize = 350000 informiam.genesys_connector.GC StatisticsObjectDao.batchSize Overide = 1000 informiam.genesys_connector.stats issue.pausechecklimit = 10000 **Note:** The informiam.genesys_connector. statsissue.pausechecklimit parameter is applicable only to releases earlier than 8.5.1.

Table 139: Recommendations for Performance Improvement (Continued)

Location	Sub-directory or File, where applicable	Settings
On each Apache HTTP proxy	httpd.conf	• Uncomment or add the following modules: ◆ LoadModule deflate_module modules/mod_deflate.so ◆ LoadModule headers_module modules/mod_headers.so ◆ LoadModule proxy_module modules/mod_proxy.so ◆ LoadModule proxy_ajp_module modules/mod_proxy_ajp.so ◆ LoadModule proxy_balancer_module modules/mod_proxy_balancer.so ◆ LoadModule proxy_http_module modules/mod_proxy_http.so • Add the following block to increase the number of Apache worker threads (note that this is for a Windows-based server): <IfModule mpm_winnt_module> ThreadsPerChild 512 MaxConnectionsPerChild 0 </IfModule> If you use a Linux server, add the following block: <IfModule mpm_event_module> StartServer 6 ServerLimit 32 MinSpareThreads 150 MaxSpareThreads 250 ThreadsPerChild 25 MaxRequestWorkers 800 MaxConnectionsPerChild 0 </IfModule>

Table 139: Recommendations for Performance Improvement (Continued)

Location	Sub-directory or File, where applicable	Settings
On each Apache HTTP proxy		• Add the following to enable a request response proxy: ◆ ProxyPass /am/ ajp://localhost:8009/am/ ◆ ProxyPass /admin/ ajp://localhost:8009/admin/ ◆ ProxyPass /am-admin/ ajp://localhost:8009/am-admin/ ◆ ProxyPass /ca/ ajp://localhost:8009/ca/ ◆ ProxyPass /ca-ws/ ajp://localhost:8009/ca-ws/ ◆ ProxyPass /ea-ws/ ajp://localhost:8009/ea-ws/ ◆ ProxyPass /base-ws/ ajp://localhost:8009/base-ws/ ◆ ProxyPass /dashboard/ ajp://localhost:8009/dashboard/ ◆ ProxyPass /nav-service/ ajp://localhost:8009/nav-service/ ◆ ProxyPass /prefs-service/ ajp://localhost:8009/prefs-service/ ◆ ProxyPass /wu/ ajp://localhost:8009/wu/ ◆ ProxyPass /rmc/ ajp://localhost:8009/rmc/ ◆ ProxyPass /gc-admin/ ajp://localhost:8009/gc-admin/ ◆ ProxyPass /ca-xml/ ajp://localhost:8009/ca-xml/

Best Practices for CCAdv Sizing

Use the following notes and best practices for optimizing CCAdv performance:

- Use Gigabit connectivity between the CCAdv aggregation node (runs CCAdv XML Generator) and CCAdv presentation node(s).
- Enable an Apache JServ Protocol (AJP) connection between the Apache HTTP proxy and CCAdv presentation node(s).
- Allocate as much CPU resource to CCAdv as possible; CCAdv performance is improved if you provide multiple CPU cores and faster clock speeds.
- Allocate sufficient memory for CCAdv components (Genesys recommends 6GB).
- Genesys recommends increasing the number of presentation nodes if the dashboard request response time exceeds acceptable thresholds.
- Apply role-based access control to minimize the number of hierarchy objects and metrics that each user can access.

- Regarding Stat Server performance:
  - Stat Server is a single threaded process. Carefully monitor the CPU usage of your Stat Server(s).
  - Consider adding more Stat Server pairs if a Stat Server is saturating a CPU. You may require up to four pairs of Stat Servers for best performance.
- Regarding XML Generator performance:
  - Increasing the "Thirty Mins And Today" metrics processing cycle duration reduces XML Generator processing overhead. The configuration parameter name is generationForThirtyMinsAndToday (default=120s) and it is located in the following file:

    <*XML Generator home*>\conf\xmlgen.properties
  - Performance improves with a small number of objects and degrades with a large number of objects, however a large number of reporting regions, geographic regions, and/or contact centers causes less degradation than a large number of operating units and/or application groups.
  - The number of columns displayed on the dashboard does not impact XML Generator performance.
- Regarding metrics graphing:
  - The greatest impact to load for the metrics graphing feature is against the aggregation node (XML Generator), not the presentation nodes.
  - The key scaling factor is the number of graphable metrics:
    - The XML Generator CPU usage scales up linearly with the number of graphable metrics.
    - Up to 15 graphable metrics are supported.
  - The number of users and number of requests for distinct graphs has minimal impact on performance.

WA Deployment Architecture and Recommendations for Optimal Performance

WA Object Configuration Information

The following Table describes the high-level dimensions controlling the environment used to achieve the results described in "Load-Carrying Capacity" on page 431.

Table 140: WA Presentation Object Configuration

Object	Count
Contact Groups	1552[1]
Contact Centers	20
Application Groups	200

Table 140: WA Presentation Object Configuration (Continued)

Object	Count
Reporting Regions	20
Operating Units	1
Applications	600
Agent Groups	1300
[1] In this environment, the forecast data for all 1552 contact groups is updated every 10 minutes.	

WA Environment Topology

The following diagram shows the environment topology used to successfully achieve 1500 concurrent users of WA. VM is a *virtual machine*. The *Aggregation Instance* performs data aggregation; the *Presentation Instance* services the dashboard.

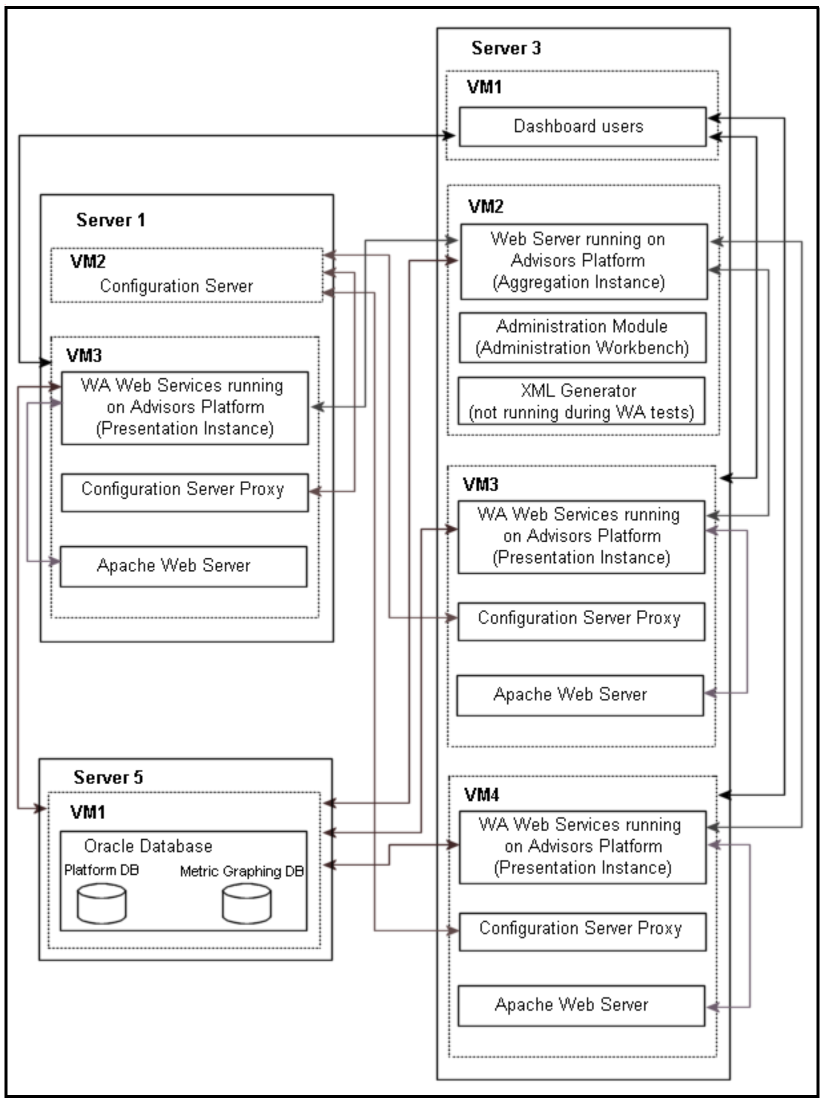


Figure 154: Workforce Advisor Environment Topology

WA Presentation Machine and VM Information

The following Table describes the characteristics of the hardware and virtualization environment shown in Figure 154, which supports 1500 concurrent users of WA.

Table 141: WA Presentation Machine and VM Information

Server	# of Processors	Processor Type	# of Cores	Total Cores	RAM	OS	Application
Server 3 VM-Host	2	Intel Xeon X5675 @ 3.07GHz	6	24 logical cores with hyper-threading	32 GB	VMWare ESXi 5.0 Standard 64bit	
Server 3 VM-1	1	(same as host)	1	1	8 GB	Windows Server 2008 Standard SP2 64bit	Dashboard simulator
Server 3 VM-2	2	(same as host)	2	4	8 GB	Windows Server 2003 Standard SP2 64bit **NOTE:** At the time of testing, Genesys supported Windows Server 2003. In your enterprise, be sure to use a Genesys-supported Windows Server operating system.	Apache Geronimo (WA Aggregation instance on Advisors Platform)
Server 3 VM-3	2	(same as host)	4	8	8 GB	Windows Server 2008 R2 Standard SP1 64bit	Apache Geronimo (WA Presentation instance on Advisors Platform), Configuration Server Proxy, Apache

Table 141: WA Presentation Machine and VM Information (Continued)

Server	# of Processors	Processor Type	# of Cores	Total Cores	RAM	OS	Application
Server 3 VM-4	2	(same as host)	4	8	8 GB	Windows Server 2008 R2 Standard SP1 64bit	Apache Geronimo (WA Presentation instance on Advisors Platform), Configuration Server Proxy, Apache
Server 1 VM-Host	2	Intel Xeon X5675 @ 3.07GHz	6	12	32 GB	VMWare ESXi 5.0 Standard 64bit	
Server-1 VM-2	2	(Same as host)	1	2	1.5 GB	Windows Server 2003 Standard SP2 64bit **NOTE:** At the time of testing, Genesys supported Windows Server 2003. In your enterprise, be sure to use a Genesys-supported Windows Server operating system.	Configuration Server
Server-1 VM-3	2	(Same as host)	4	8	8 GB	Windows Server 2008 R2 Standard SP1 64bit	Apache Geronimo (WA Presentation instance on Advisors Platform), Configuration Server Proxy, Apache

Table 141: WA Presentation Machine and VM Information (Continued)

Server	# of Processors	Processor Type	# of Cores	Total Cores	RAM	OS	Application
Server 5 VM-Host	2	AMD Opteron 2439SE @ 2.8GHz	6	12	32 GB	VMWare ESXi 5.0 Standard 64bit	
Server-5 VM-1	2	(Same as host)	1	2	8 GB	RHEL Server 5.7 64bit	Oracle database

WA Configuration for High Performance

The following Table describes settings you can change to improve WA performance.

Table 142: Recommendations for Performance Improvement

Location	Sub-directory or File, where applicable	Settings
On each WA node	<*WA home*>\geronimo-tomcat6-minimal-2.2.1\bin\setenv.bat	Change GERONIMO_OPTS=-ms128m -mx1024m -XX:MaxPermSize=128m To GERONIMO_OPTS=-Xms6g -Xmx6g -XX:MaxPermSize=256m
	<*WA home*>\geronimo\var\catalina\server.xml	Under the <Connectorname=”TomcatAJPConnector”> section, add maxThreads=”800”
On WA presentation nodes only	<*WA home*>\geronimo-tomcat6-minimal-2.2.1\var\config\config-substitutions.properties	Set MaxThreadPoolSize to “3000”
Dashboard administration setting		For optimal performance: • Select independent configuration mode (not integrated configuration mode) • Set Show Totals and Averages to No

Table 142: Recommendations for Performance Improvement (Continued)

Location	Sub-directory or File, where applicable	Settings
On AGA	Releases 8.1.5 and 8.5.0 <*AGA home*>\conf\wrapper.conf Release 8.5.1 and later conf\run.bat (Windows) bin\setenv.sh (Linux)	Releases 8.1.5 and 8.5.0 Change Wrapper.java.initmemory=128 Wrapper.java.maxmemory=1024 To Wrapper.java.initmemory=4096 Wrapper.java.maxmemory=14336 Release 8.5.1 and later Change the following values in the JAVA_OPTS parameter: Change -ms128m -mx1024m To -ms4096m -mx14336m

Table 142: Recommendations for Performance Improvement (Continued)

Location	Sub-directory or File, where applicable	Settings
On each Apache HTTP proxy	httpd.conf	• Uncomment or add the following modules: ◆ LoadModule deflate_module modules/mod_deflate.so ◆ LoadModule headers_module modules/mod_headers.so ◆ LoadModule proxy_module modules/mod_proxy.so ◆ LoadModule proxy_ajp_module modules/mod_proxy_ajp.so ◆ LoadModule proxy_balancer_module modules/mod_proxy_balancer.so ◆ LoadModule proxy_http_module modules/mod_proxy_http.so • Add the following block to increase the number of Apache worker threads (note that this is for a Windows-based server): <IfModule mpm_winnt_module> ThreadsPerChild 512 MaxConnectionsPerChild 0 </IfModule> If you use a Linux server, add the following block: <IfModule mpm_event_module> StartServer 6 ServerLimit 32 MinSpareThreads 150 MaxSpareThreads 250 ThreadsPerChild 25 MaxRequestWorkers 800 MaxConnectionsPerChild 0 </IfModule>

Table 142: Recommendations for Performance Improvement (Continued)

Location	Sub-directory or File, where applicable	Settings
On each Apache HTTP proxy		• Add the following to enable a request response proxy: ◆ ProxyPass /am/ ajp://localhost:8009/am/ ◆ ProxyPass /admin/ ajp://localhost:8009/admin/ ◆ ProxyPass /ca-ws/ ajp://localhost:8009/ca-ws/ ◆ ProxyPass /ea-ws/ ajp://localhost:8009/ea-ws/ ◆ ProxyPass /dashboard/ ajp://localhost:8009/dashboard/ ◆ ProxyPass /nav-service/ ajp://localhost:8009/nav-service/ ◆ ProxyPass /prefs-service/ ajp://localhost:8009/prefs-service/ ◆ ProxyPass /ca-xml/ ajp://localhost:8009/ca-xml/ ◆ ProxyPass /wu/ ajp://localhost:8009/wu/ ◆ ProxyPass /base-ws/ ajp://localhost:8009/base-ws/ ◆ ProxyPass /fa/ ajp://localhost:8009/fa/ ◆ ProxyPass /static/ ajp://localhost:8009/static/

Best Practices for WA Sizing

Use the following notes and best practices for optimizing WA performance:

- Use Gigabit connectivity between the WA aggregation node and WA presentation node(s).
- Enable an AJP connection between the Apache HTTP proxy and WA presentation layer(s).
- Allocate as much CPU resource to WA as possible; WA performance is improved if you provide multiple CPU cores and faster clock speeds.
- Allocate sufficient memory for WA components (Genesys recommends 6GB).
- Genesys recommends increasing the number of presentation nodes if the dashboard request response time exceeds acceptable thresholds.
- Apply role-based access control to minimize the number of hierarchy objects and metrics that each user can access.
- Avoid unnecessary updates to forecast data; that is, avoid calculations that consume processing power unnecessarily. For example, do not configure 10-minute updates of forecast data if hourly updates are sufficient.
- Regarding Stat Server performance:

- Stat Server is a single threaded process. Carefully monitor the CPU usage of your Stat Server(s).
- Consider adding more Stat Server pairs if a Stat Server is saturating a CPU. You may require up to four pairs of Stat Servers for best performance.

FA Deployment Architecture and Recommendations for Optimal Performance

FA Object Configuration Information

The following Table describes the high-level dimensions controlling the environment used to achieve the results described in “Load-Carrying Capacity” on page 431:

Table 143: FA Presentation Object Configuration

Object	Count
Agents	30 000
Depth (levels)	6
Multiplicity* * Refers to the average number of teams to which an agent belongs.	1
Agent Groups	5000 (with agents)
Time Profiles	3

FA Environment Topology

The following diagram shows the environment topology used to successfully achieve 1500 concurrent users of FA. VM is a *virtual machine*. The *Aggregation Instance* performs data aggregation; the *Presentation Instance* services the dashboard.

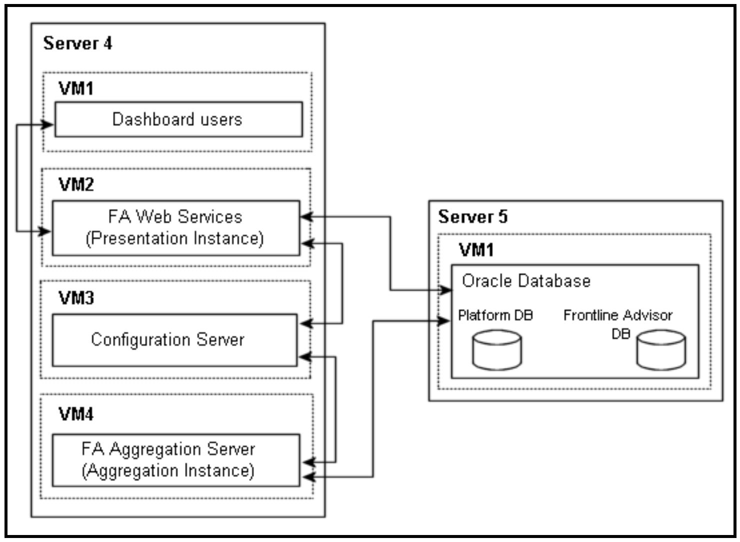


Figure 155: Frontline Advisor Environment Topology

FA Presentation Machine and VM Information

The following Table describes the characteristics of the hardware and virtualization environment shown in Figure 155, which supports 1500 concurrent users of FA.

Table 144: FA Presentation Machine and VM Information

Server	# of Processors	Processor Type	# of Cores	Total Cores	RAM	OS	Application
Server 4 VM-Host	2	Intel Xeon X5675 @ 3.07GHz	6	12	32 GB	VMWare ESXi 5.0 Standard 64bit	
Server 4 VM-1	2	(same as host)	4	8	8 GB	Windows Server 2008 Standard SP1 64bit	Dashboard simulator

Table 144: FA Presentation Machine and VM Information (Continued)

Server	# of Processors	Processor Type	# of Cores	Total Cores	RAM	OS	Application
Server 4 VM-2	2	(same as host)	4	8	8 GB	Windows Server 2003 Standard SP1 64bit **NOTE:** At the time of testing, Genesys supported Windows Server 2003. In your enterprise, be sure to use a Genesys-supported Windows Server operating system.	Geronimo (FA Presentation instance on Advisors Platform)
Server 4 VM-3	2	(same as host)	1	2	4 GB	Windows Server 2003 32bit **NOTE:** At the time of testing, Genesys supported Windows Server 2003. In your enterprise, be sure to use a Genesys-supported Windows Server operating system.	Configuration Server
Server 4 VM-4	2	(same as host)	4	8	8 GB	Windows Server 2008 R2 Standard SP1 64bit	FA Aggregation instance on Advisors Platform

Table 144: FA Presentation Machine and VM Information (Continued)

Server	# of Processors	Processor Type	# of Cores	Total Cores	RAM	OS	Application
Server 5 VM-Host	2	AMD Opteron 2439SE @ 2.8GHz	6	12	32 GB	VMWare ESXi 5.0 Standard 64bit	
Server-5 VM-1	2	(same as host)	1	2	8 GB	RHEL Server 5.7 64bit	Oracle database

FA Configuration for High Performance

The following Table describes settings you can change to improve FA performance.

Table 145: Recommendations for Performance Improvement

Location	Sub-directory or File, where applicable	Settings
On each FA node	<*FA home*>\geronimo-tomcat6-minimal-2.2.1\bin\setenv.bat	Change GERONIMO_OPTS=-ms128m -mx1024m -XX:MaxPermSize=128m To GERONIMO_OPTS=-Xms4g -Xmx8g -XX:MaxPermSize=512m
	<*FA home*>\geronimo\var\catalina\server.xml	Under the <Connectorname=”TomcatAJPConnector”> section, add maxThreads=”2000”

Table 145: Recommendations for Performance Improvement (Continued)

Location	Sub-directory or File, where applicable	Settings
On AGA	Releases 8.1.5 and 8.5.0 <*AGA home*>\conf\wrapper.conf Release 8.5.1 and later conf\run.bat (Windows) bin\setenv.sh (Linux)	Releases 8.1.5 and 8.5.0 Change Wrapper.java.initmemory=128 Wrapper.java.maxmemory=1024 To Wrapper.java.initmemory=4096 Wrapper.java.maxmemory=14336 Release 8.5.1 and later Change the following values in the JAVA_OPTS parameter: Change -ms128m -mx1024m To -ms4096m -mx14336m
On AGA	<*AGA home*>\conf\inf_genesys_ adapter.properties	Change informiam.genesys_connector.stat Server.addp.clienttimeout = 120 informiam.genesys_connector.timing. messagerate.numberofmessages.batch = 500 To informiam.genesys_connector.stat Server.addp.clienttimeout = 360 informiam.genesys_connector.timing. messagerate.numberofmessages.batch = 100000

Table 145: Recommendations for Performance Improvement (Continued)

<table>
<tr><th>Location</th><th>Sub-directory or File, where applicable</th><th>Settings</th></tr>
<tr><td>On each Apache HTTP proxy</td><td>httpd.conf</td><td>
• Uncomment or add the following modules:
◆ LoadModule deflate_module modules/mod_deflate.so
◆ LoadModule headers_module modules/mod_headers.so
◆ LoadModule proxy_module modules/mod_proxy.so
◆ LoadModule proxy_ajp_module modules/mod_proxy_ajp.so
◆ LoadModule proxy_balancer_module modules/mod_proxy_balancer.so
◆ LoadModule proxy_http_module modules/mod_proxy_http.so
• Add the following block to increase the number of Apache worker threads (note that this is for a Windows-based server):
<IfModule mpm_winnt_module>
ThreadsPerChild 512
MaxConnectionsPerChild 0
</IfModule>
If you use a Linux server, add the following block:
<IfModule mpm_event_module>
StartServer 6
ServerLimit 32
MinSpareThreads 150
MaxSpareThreads 250
ThreadsPerChild 25
MaxRequestWorkers 800
MaxConnectionsPerChild 0
</IfModule>
</td></tr>
</table>

Table 145: Recommendations for Performance Improvement (Continued)

Location	Sub-directory or File, where applicable	Settings
On each Apache HTTP proxy		• Add the following to enable a request response proxy: ◆ ProxyPass /fa/ ajp://localhost:8009/fa/ ◆ ProxyPass /am/ ajp://localhost:8009/am/ ◆ ProxyPass /admin/ ajp://localhost:8009/admin/ ◆ ProxyPass /am-admin/ ajp://localhost:8009/am-admin/ ◆ ProxyPass /ca/ ajp://localhost:8009/ca/ ◆ ProxyPass /ca-ws/ ajp://localhost:8009/ca-ws/ ◆ ProxyPass /ea-ws/ ajp://localhost:8009/ea-ws/ ◆ ProxyPass /base-ws/ ajp://localhost:8009/base-ws/ ◆ ProxyPass /dashboard/ ajp://localhost:8009/dashboard/ ◆ ProxyPass /nav-service/ ajp://localhost:8009/nav-service/ ◆ ProxyPass /prefs-service/ ajp://localhost:8009/prefs-service/ ◆ ProxyPass /wu/ ajp://localhost:8009/wu/ ◆ ProxyPass /rmc/ ajp://localhost:8009/rmc/ ◆ ProxyPass /gc-admin/ ajp://localhost:8009/gc-admin/ ◆ ProxyPass /ca-xml/ ajp://localhost:8009/ca-xml/

Best Practices for FA Sizing

Use the following notes and best practices for optimizing FA performance:

- Use Gigabit connectivity between the FA aggregation node and FA presentation node(s).
- Enable an AJP connection between the Apache HTTP proxy and FA presentation node(s).
- Allocate as much CPU resource to FA as possible; FA performance is improved if you provide multiple CPU cores and faster clock speeds.
- Allocate sufficient memory for FA components (Genesys recommends 6GB).
- You may require multiple AGAs (up to two).
- Regarding Stat Server performance:
  - Stat Server is a single threaded process. Carefully monitor the CPU usage of your Stat Server(s).

- Consider adding more Stat Server pairs if a Stat Server is saturating a CPU. You may require up to six pairs of Stat Servers for best performance.

FA Dashboard Age

Dashboard age is a performance measure used in assessing the state and performance/rule processing cycles. It represent the age of the statistics on the dashboard, which includes the following:

1. Pre-Rollup Delay: The time from the end of the last rollup until the scheduled start of the next rollup.
2. Rollup Duration: The duration of the rollup + the time to publish to the distributed cache.
3. Request Response Time (RRT): The 95th percentile of response time for a dashboard request. That is, 95% of the time, a response is returned to a dashboard request after X number of seconds, where X is a constant.

The following Table shows results from Genesys' performance testing. *State processing* refers to the state metric rollup cycle and *performance/rule processing* is the performance metric rollup cycle (state and performance metric cycles run independently).

Table 146: Dashboard Age Results from FA Performance Testing

Measure (seconds)	State Processing			Performance/Rule Processing		
	95th Percentile	Median	Average	95th Percentile	Median	Average
Dashboard age	18 seconds	12 seconds	12 seconds	70 seconds	62 seconds	62 seconds

Improving Supervisor Desktop Service Performance

Supervisor Desktop Service (SDS) is required only in installations where you are deploying the Resource Management Console (RMC).

To improve SDS performance, Genesys recommends that you make the following updates on the servers that host the SDS:

1. Update the ms and mx values.

On a Windows server that hosts your SDS service, update the GDesktopStarter.ini file, located in the bin directory.

On a Linux server that hosts your SDS service, update the /bin/setclasspath.sh file in the folder where SDS is installed.

Find the line in the file starting with [JavaArgs]. You can copy the block of text below that best suits your environment, and then paste it into the file to overwrite the default settings.

For a **small configuration** (the number of agents monitored by RMC is less than 1000), use the following text block:

```
Rem The following line should be used when only the Agents Desktop
Rem is used or for small to medium configurations using
Rem the Supervisor Desktop
-Xms1024M
-Xmx2048M
-XX:MaxPermSize=128M

Rem The line above should be modified for large configurations and
Rem when both the Agents and Supervisor desktop are used:
Rem -Xmx1024M
Rem -Xms768M
```

For a **medium configuration** (the number of agents monitored by RMC is between 1000 and 2000), use the following text block:

```
Rem The following line should be used when only the Agents Desktop
Rem is used or for small to medium configurations using
Rem the Supervisor Desktop
-Xms2048M
-Xmx4096M
-XX:MaxPermSize=128M

Rem The line above should be modified for large configurations and
Rem when both the Agents and Supervisor desktop are used:
Rem -Xmx1024M
Rem -Xms768M
```

For a **large configuration** (the number of agents monitored by RMC is more than 2000), use the following text block:

```
Rem The following line should be used when only the Agents Desktop
Rem is used or for small to medium configurations using
Rem the Supervisor Desktop
Rem -Xms128M
Rem -Xmx512M
-XX:MaxPermSize=128M

Rem The line above should be modified for large configurations and
Rem when both the Agents and Supervisor desktop are used:
-Xmx6144M
-Xms3096M
```

2. In addition to setting the ms and mx values, choose one of the following sets of options:

On a **single-processor system**:

```
Rem The following option should be added after the memory option in
Rem the line above when running on a single processor system:
-XX:+UseConcMarkSweepGC
Rem The following option should be added after the memory option in
Rem the line above when running on a multi-processor system:
Rem -XX:+UseConcMarkSweepGC
Rem -XX:+UseParNewGC
```

On a **multi-processor system**:

```
Rem The following option should be added after the memory option in
Rem the line above when running on a single processor system:
Rem -XX:+UseConcMarkSweepGC
Rem The following option should be added after the memory option in
Rem the line above when running on a multi-processor system:
-XX:+UseConcMarkSweepGC
-XX:+UseParNewGC
```

Chapter

14 Workspace Desktop Edition

This chapter describes the Genesys Interaction Workspace and Genesys Workspace Desktop Edition hardware and space requirements for various deployment methods for Agent Workstations, ClickOnce, and (Virtualized) Server for Interaction Workspace 8.1 and Workspace Desktop Edition 8.5.0

Note: For information about sizing for Workspace Desktop Edition 8.5.1 refer to the Genesys online documentation.

Workspace supports three main deployment methods:

- Local deployment in an agent workstation.
- The ClickOnce feature.
- A (virtualized) server, such as Citrix, or other. Refer to *Genesys Supported Operating Environment Reference Guide* at http://docs.genesys.com/ for information about supported virtualized environments.

This chapter includes the following sections:

- Agent Workstation Requirements, page 459
- ClickOnce Requirements, page 460
- (Virtualized) Server Requirements, page 460

Agent Workstation Requirements

Table 147 shows the recommended hardware requirements for the Workspace host machine.

Table 147: Recommended Hardware Requirements - Workspace Host Machine

Processor	Memory	Hard Drive	Graphic Card	Network
Intel Core 2 Duo CPU 2.6 GHz	2 GB [a]	200 MB	DirectX 9.0+	xDSL / Lan

a. 4 GB of RAM is recommended when non-Genesys applications are being run concurrently, or to improve performance.

ClickOnce Requirements

Table 148 shows the space that is recommended on the target workstations and the minimum network data transfer rate to deploy Workspace by using the ClickOnce feature.

Table 148: ClickOnce Target Workstations and Minimum Network Data Transfer Rate

Hard Drive	Network
500 MB Free Space	1 GB LAN, or more

The size of the Workspace downloads (without HTTP compression), including plugins, for each workstation are:

- Full-featured Workspace: 23 MB
- Workspace SIP Endpoint plugin: 23 MB
- Twitter, Facebook, and Genesys Agent Scripting plugins: 1 MB each

(Virtualized) Server Requirements

Table 149 shows the recommended hardware requirements for each session on the Centralized Workspace host machine.

Table 149: Recommended Hardware Requirements - Centralized Workspace Host Machine

Memory	Hard Drive	Network Bandwidth ICA / PCoIP	Network Bandwidth Audio
600 MB (x64 OS)	Sufficient space for Genesys Workspace and logs. 35 to 50MB for Workspace and approximately 1 GB for logs. The requirement for logs is configurable and so might be more or less than this value.	< 10 KB / Sec	8-16 KB / Sec

Note: For detailed information, please refer to the Workspace Deployment Guide: http://docs.genesys.com/Documentation/IW

Chapter

15 Unified Communications (UC) Connector

The Unified Communications (UC) Connector supports contact center integration with enterprise Unified Communications (UC) platforms of up to 2500 users. UC users are described as 'Knowledge Workers' in terms of the role these users play in the contact center.

Note: For integrations with UC platforms of more than 2500 Knowledge Workers, contact Genesys Customer Care or Professional Services.

- Server Requirements, page 461
- Desktop Requirements, page 462

Server Requirements

The recommended hardware requirements for the UC Connector host machine are as follows:

Table 150: Recommended Server Requirements, Max 2500 Knowledge Workers

Servers	Application	Processor(s)	Memory	Hard Drive	Operating System [a]
1	UC Connector	Dual-core 2.0GHz+	4 GB	10 GB	• Windows Server 2003 • Windows Server 2008 • Windows Server 2008 R2 **Note**: Both 32 and 64-bit versions are supported.

a. Support for virtual platforms is also available.

Desktop Requirements

For UC Connector integration with the Knowledge Worker desktop, the Knowledge Worker desktop requires the following:

- **Operating System**: Windows XP, Windows Server 2003, Windows Server 2008, Windows Vista, Windows 7

  Required for UC client installation. In release 8.0, UC Connector supports integration with Microsoft Office Communicator 2007 R2 and IBM Sametime Connect 8.5.

- **Browser**: Internet Explorer (IE) 8.0 or higher

  Required for UC client customization. In release 8.0, IE 8.0 is required for the custom Contact Center tab that appears in the Microsoft Office Communicator client interface.

- **Hardware**: Genesys does not have specific requirements for the host machine. The machine needs to be able to run both Microsoft Office Communicator 2007 R2 and Internet Explorer 8.0. Consult the third-party documentation for these products to determine their prerequisite hardware.

Chapter

16 Genesys Administrator

This chapter describes Genesys Administrator and Genesys Administrator Extension and provides hardware recommendations for their use. This chapter contains the following sections:

- Genesys Administrator, page 463
- Genesys Administrator Extension, page 463
- Network Connections, page 464
- Hardware Recommendations, page 466

Genesys Administrator

Genesys Administrator is a web-based application that combines the functionality of Configuration Manager, Solution Control Interface, and other Genesys GUIs. It also includes functionality for deploying Genesys Installation Packages on local and remote hosts.

Genesys Administrator runs on a web server, and provides a comprehensive browser-based user interface to:

- Configure, monitor, and control your Genesys environment.
- Deploy applications and solutions to remote hosts.
- Manage user access to your Genesys environment, particularly as it pertains to access permissions and Role-Based Access Control.

Genesys Administrator Extension

Genesys Administrator Extension (GAX) is the next-generation user interface for Genesys that reduces both the overall operating costs and the time to deployment, by providing user-friendly interfaces that perform complex operations, while at the same time preventing user error. This product is focused on the user experience for both Enterprise and Hosted customers, as well as by system administrators and line of business users.

Network Connections

Genesys Administrator

Figure 156 on page 464 illustrates the architecture of Genesys Administrator as described below the diagram.

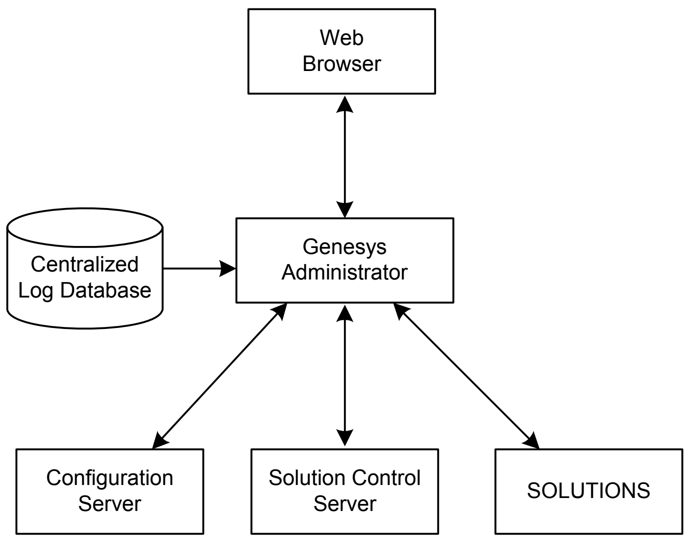


Figure 156: Genesys Administrator Architecture

Genesys Administrator connects to:

- Configuration Server (a Configuration Layer component) to exchange configuration information.
- Solution Control Server (a Management Layer component) to exchange status, operations, and control information.
- Other back end servers, depending on the solutions deployed in the system, to retrieve solution-specific information. These back-end servers may include:
  - For the Outbound Contact Solution—Outbound Contact Server
  - For the Voice Platform Solution—GVP Resource Manager, GVP Reporting Server, and IVR Solution

Genesys Administrator also reads logs from the Centralized Log Database, a Management Layer component.

Genesys Administrator Extension

Figure 157 illustrates the architecture of Genesys Administrator Extension.

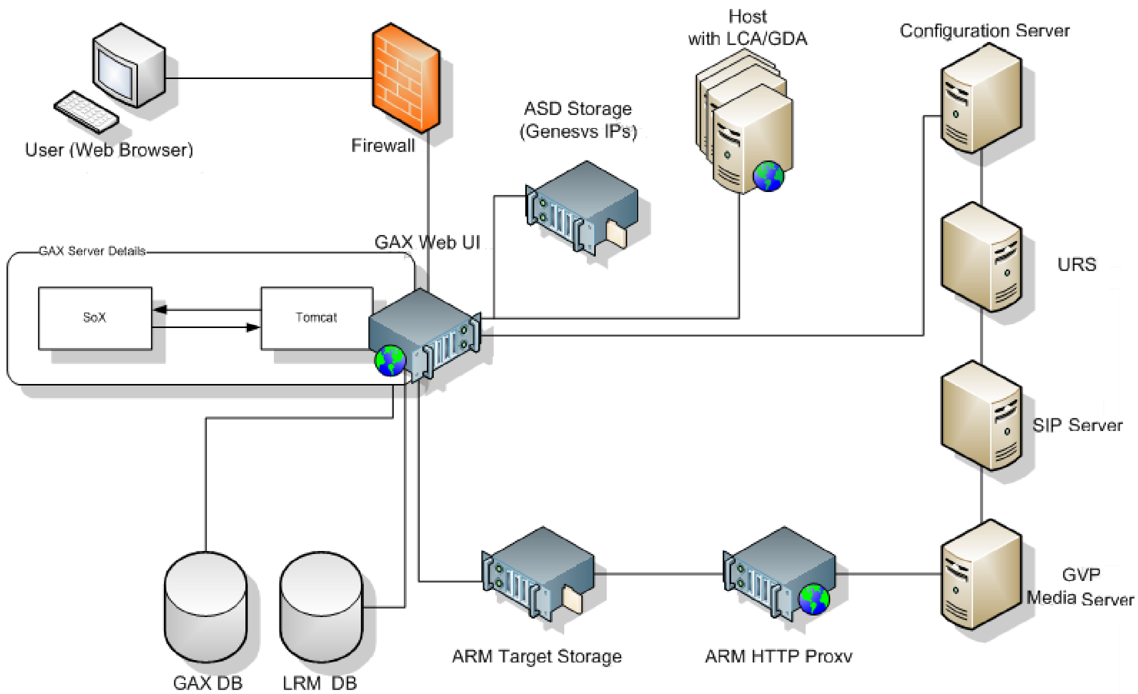


Figure 157: Genesys Administrator Extension Architecture

The browser-based GAX includes a comprehensive user interface to perform tasks that are related to Solution Deployment, Operational Parameter Management, Audio Resource Management, and Configuration Object Management.

Currently, Genesys Administrator and GAX are the only components in the User Interaction Layer.

- GAX:
  - Communicates with the Configuration Server (a Configuration Layer component) to exchange configuration data.
  - Communicates with the Solution Control Server (a Management Layer component) to exchange status, operations, and control information.
  - Depending on the solutions that are deployed in the system, Genesys Administrator Extension might also communicate with other back-end servers to retrieve solution-specific information.
  - Uses the GAX Database to store non-configuration information, such as operational parameter templates and audio resource metadata.

- Reads license utilization information from the LRM Database to generate License Usage reports.
- Uses Sound eXchange (SoX) to encode audio files.
- Sends encoded audio files to the ARM Storage, from where the ARM Web Server distributes them to GVP Media Servers.
- Uploads IPs to Solution Deployment storage.

Hardware Recommendations

Genesys Administrator

Minimum The Standard Server configuration will be sufficient for 10 Genesys Administrator users logged in simultaneously.

Recommended To support up to 120 Genesys Administrator users logged in simultaneously, the Standard Server requires the following:

- 64-bit operating system
- 8 GB RAM

Each instance of Genesys Administrator can support a maximum number of 120 users logged in simultaneously. Multiple deployments of Genesys Administrator, for configurations with more than 120 users logged in simultaneously, are supported.

Genesys Administrator Extension

To support up to 175 Genesys Administrator Extension users, the Standard Server requires the following:

- 64-bit operating system
- Four virtual processor cores
- 8 GB RAM
- Default configuration of 2 GB Java Heap Memory for GAX processes

Each instance of Genesys Administrator Extension can support a maximum number of 175 users, logged in at an average rate of one user per second. Multiple deployments of Genesys Administrator Extension, for configurations with more than 175 users logged in simultaneously, are supported.

Genesys

Chapter

17 Genesys Orchestration Platform

This chapter provides hardware sizing guidelines and basic information for deploying Genesys Orchestration Platform 8.x, and provides information about capacity and performance testing. This chapter covers the following topics:

- Genesys Orchestration Platform Overview, page 467
- Software Versions Utilized in Reference Architecture, page 469
- Call Model Scenarios and Benchmarks, page 470
- Performance Test Result Conclusions, page 483
- Hardware Specifications, page 485

Genesys Orchestration Platform Overview

This chapter provides performance data for Orchestration platform, which includes both Universal Routing Server (URS) and Orchestration Server (ORS) services.

The Orchestration Platform is measured for variable call rates and interaction type. Other important parameters are the number of services in a cluster, and its performance based on the complexity of the scenario.

The results presented in this chapter are based on lab benchmarks obtained on a 32-bit version of Microsoft Windows Server 2003 Service Pack 2 (SP2) on an Intel CPU. The benchmark is also performed utilizing a 32-bit Red Hat Enterprise Linux running on an Intel Machine.

For details of the hardware platform that was used, refer to the section “Hardware Specifications” on page 485.

To match or exceed the performance levels described in this chapter, Genesys recommends that you use hardware that is at least as powerful as that in the lab environment.

The Genesys Orchestration Platform chapter covers the following topics:

- "Genesys Orchestration Platform Reference Architecture" on page 468 includes different architecture options in deploying this solution, and Genesys recommendations.
- "Call Model Scenarios and Benchmarks" on page 470 includes call model sample scenarios and benchmark data obtained based on the reference architecture and scenarios. It discusses sample application placement across host computers; expected system loads with CPU and Memory usage-based use cases.
- "Performance Test Result Conclusions" on page 483 provides expected results with different architecture variations.
- "Hardware Specifications" on page 485 includes required hardware resources or the number of physical host computers required.

Genesys Orchestration Platform Reference Architecture

The Figure 158, "Genesys Orchestration Platform Reference Architecture: Benchmarking Sample," on page 469 depicts a generic architecture sample for Orchestration Platform deployed for benchmarking. The architecture involves multiple Orchestration Services working in a cluster in conjunction with a Universal Router. The persistence functionality is handled by Cassandra service, which is again deployed in a clustered architecture.

A different set of Orchestration services are used from this cluster in the performance testing, in conjunction with different call models. For more information regarding these call models, refer to the section"Call Model Scenarios and Benchmarks" on page 470.

High Availability and redundancy of Orchestration and Cassandra are achieved through clustering. Services in this cluster can be deployed on a single host or on different physical hosts. For simplicity, and to measure each service tied to a host, this reference architecture employs each Orchestration service on separate host machines. This chapter does not discuss the impact on having different service on a single host or in different VM images.

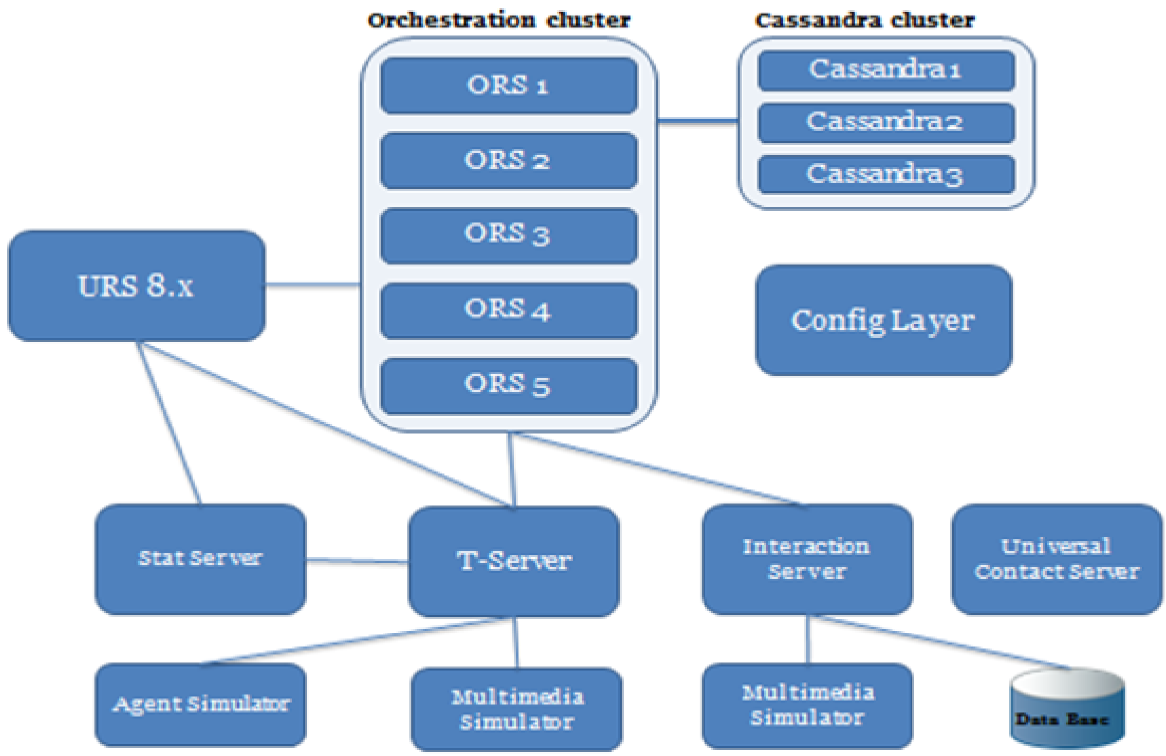


Figure 158: Genesys Orchestration Platform Reference Architecture: Benchmarking Sample

Please refer to the *Genesys Supported Operating Environment Reference Guide Wiki* at http://docs.genesys.com/ and product release notes on the Tech Support Website, for current support information for different Genesys solutions.

Software Versions Utilized in Reference Architecture

The reference architecture and the load results described in this section, based on different scenarios, are applicable to all Genesys services 8.x and later.

Table 151 provides the releases/versions utilized in the reference architecture.

Table 151: Software Versions Utilized in Reference Architecture

Application	Version
Configuration Server	8.0.300.07
Database Server	8.0.300.07
Solution Control Server	8.0.300.04

Table 151: Software Versions Utilized in Reference Architecture (Continued)

Application	Version
Message Server	8.0.300.05
T-Server Avaya	8.0.101.08
StatServer	8.0.000.30
Interaction Server	8.0.2000.12
Universal Contact Server	8.0.300.04
Data Base Server	MS SQL 2005
Universal Routing Server	8.0.100.17
Orchestration Server	8.1.000.19
WEB Server	Apache 2.2
Cassandra Server	0.6.12

Call Model Scenarios and Benchmarks

The following section describes various call model scenario examples and benchmarks. Please refer to the section "Hardware Specifications" on page 485 for details on the hardware utilized for each service in these scenarios.

The model scenarios in this section include:

- Voice Only Call Model Scenario, page 470
- Two ORS Applications in a Cluster on Windows, page 474
- Five ORS Applications in a Cluster on Windows, page 475
- Two ORS Applications in a Cluster on Linux, page 477
- Five ORS Applications in a Cluster on Linux, page 479
- Five ORS Nodes Cluster: Single WEB Server on Windows, page 481

Voice Only Call Model Scenario

In the Voice Only Call Model scenario, a single site call model is utilized with an in-front IVR architecture simulated through 800 DNs configured to route all inbound calls to a route point. Almost 100 bytes of data is attached to the call en-route, before routing the call to a route point.

An SCXML application associated with the route point then routes the incoming calls to Agents. The application utilized in this routing is complex in logic and also attaches 200 bytes of data to the interaction. Refer to Figure 159, "Genesys Orchestration Platform Voice Only Call Model," on page 472 for this call flow.

Call distribution algorithm provisioned in this SCXML application associated with this route point varies depending on data attached. In this environment, around 7000 Agents are configured to receive calls. Transfer time is 1 sec, and Agent Talk Time is 10 sec for each call.

For CAPS (Calls Per Second) distribution and CPU % information, please refer to the section "Performance Test Result Conclusions" on page 483.

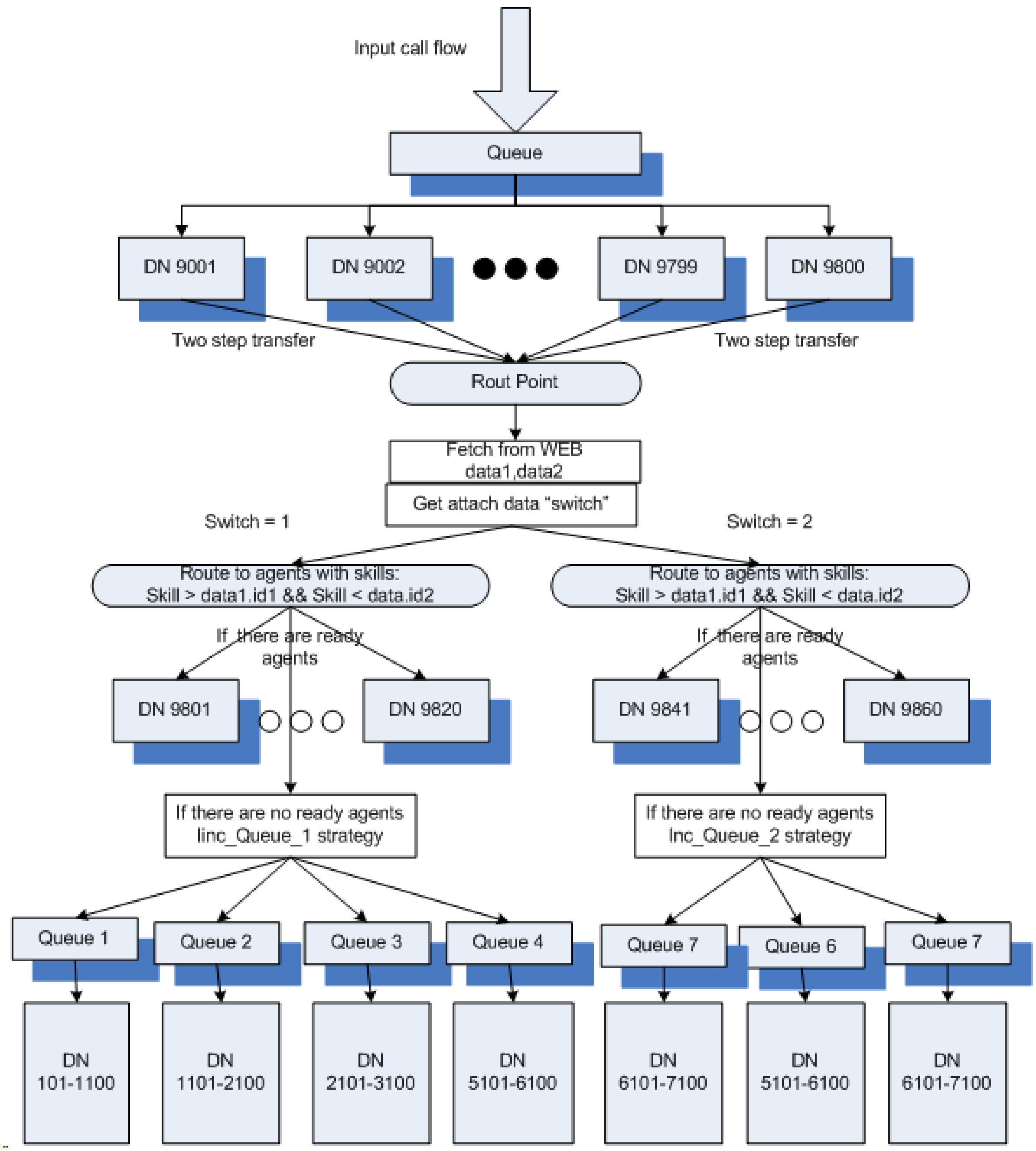


Figure 159: Genesys Orchestration Platform Voice Only Call Model

Voice Only Call Model Scenario Parameters

Table 152 shows the parameters considered for the voice only call model scenario.

Table 152: Software Versions Utilized in Voice Only Call Model Reference Architecture

Test Parameters	Parameter Values
SCXML Strategies	Complex SCXML application is utilized by provisioning it through a Route point. Refer to Orchestration deployment documentation regarding SCXML application provisioning
Reporting	Reporting clients are connected with requests made once in every 10 sec to get all active sessions. Please refer to Orchestration deployment documentation on how to configure Reporting services to view all active session in Genesys Administrator
Calls per second (CAPS)	The voice only scenario is tested with different loads at 40, 80, 120, 160, and 200 CAPS measuring all services for these loads
Tested configurations	Orchestration Platform, i.e. Single URS + cluster of ORS + Cassandra cluster: 1. 2 nodes ORS cluster + 3 Cassandra nodes cluster 2. 5 nodes ORS cluster + 3 Cassandra nodes cluster
Hardware platform	The voice only configurations are tested with different hardware platforms: • Windows - #1,#2 configurations • Linux - #1,#2 configurations The current version of Orchestration Platform supports Windows and Linux, only.

Two ORS Applications in a Cluster on Windows

In this sample call model, voice only interactions are routed utilizing two Orchestration services in a cluster, where both Orchestration service and Cassandra are running on a Windows machine. The CPU and memory utilization is measured for all services at different CAPS.

The CPU% is measured on all cores of the machine.

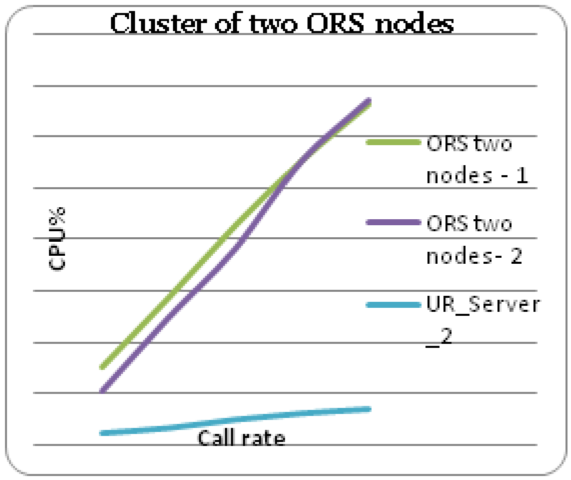


Figure 160: Genesys Orchestration Platform Voice Call Model with Two ORS Applications on Windows

Two ORS Applications in a Cluster on Windows Parameters

Table 153 shows the parameters considered for this scenario.

Table 153: Two ORS Applications in a Cluster on Windows

Application		40	80	120	160	200
ORS_0	CPU%	75	143	213	227	332
	Memory	360	400	441	459	470
ORS_1	CPU%	52	124	191	277	336
	Memory	343	392	413	458	499
Cassandra 1	CPU%	17	35	60	80	90
	Memory	350	600	1000	1700	1700

Table 153: Two ORS Applications in a Cluster on Windows (Continued)

Application		40	80	120	160	200
Cassandra 2	CPU%	17	35	55	87	90
	Memory	400	700	1000	1700	1700
Cassandra 3	CPU%	17	23	35	47	55
	Memory	350	500	900	1500	1700
Router	CPU%	11	16	24	30	34
	Memory	449	462	479	488	495
WEB server	CPU%	3	5	7	9	12
	Memory	62	62	62	62	62
WEB server	CPU%	2	4	7	10	12
	Memory	27	39	27	49	49

Five ORS Applications in a Cluster on Windows

In this call model, voice only interactions are routed utilizing five Orchestration services in a cluster, where both Orchestration service and Cassandra are running on a Windows machine. The CPU and memory utilization is measured for all services at different CAPS

The CPU% is measured on all cores of the machine.

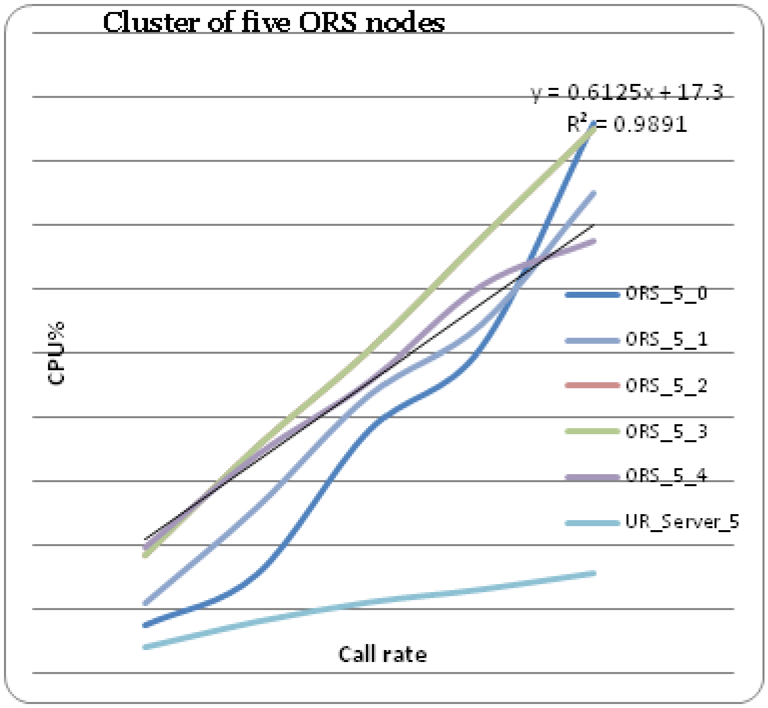


Figure 161: Genesys Orchestration Platform Voice Call Model with Five ORS Applications on Windows

Five ORS Applications on Windows Parameters

Table 154 shows the parameters considered for this scenario.

Table 154: Five ORS Applications on Windows

Application		40	80	120	160	200
ORS_0	CPU%	15	31	76	102	172
	Memory	323	348	371	399	409
ORS_1	CPU%	22	52	87	109	150
	Memory	326	346	375	397	402
ORS_2	CPU%	32	54	82	104	121
	Memory	335	352	370	377	395
ORS_3	CPU%	37	71	101	136	170
	Memory	341	357	393	425	421

Table 154: Five ORS Applications on Windows (Continued)

Application		40	80	120	160	200
ORS_4	CPU%	39	68	91	121	135
	Memory	344	358	381	409	414
Cassandra 1	CPU%	20	35	60	77	90
	Memory	650	1000	1400	1700	1700
Cassandra 2	CPU%	20	35	57	75	85
	Memory	650	900	1200	1700	1700
Cassandra 3	CPU%	10	17	30	43	47
	Memory	650	1000	1300	1600	1600
Router	CPU%	8	16	22	26	31
	Memory	451	464	477	490	495
WEB server	CPU%	0.2	0.6	2	2.8	5.4
	Memory	62	62	62	62	62
WEB server	CPU%	0.6	1.6	2.6	3.2	4.3
	Memory	49	49	49	49	49
WEB server	CPU%	1.1	2	2.7	3.2	3.9
	Memory	69	69	69	69	69
WEB server	CPU%	1.1	2.2	2.9	4.1	5
	Memory	55	55	55	55	55
WEB server	CPU%	1.2	1.9	2.7	3.8	4
	Memory	61	61	61	61	61

Two ORS Applications in a Cluster on Linux

In this call model, voice only interactions are routed utilizing two Orchestration services in a cluster, where both Orchestration service and Cassandra are running on Linux. The CPU and memory utilization is measured for all services at different CAPS.

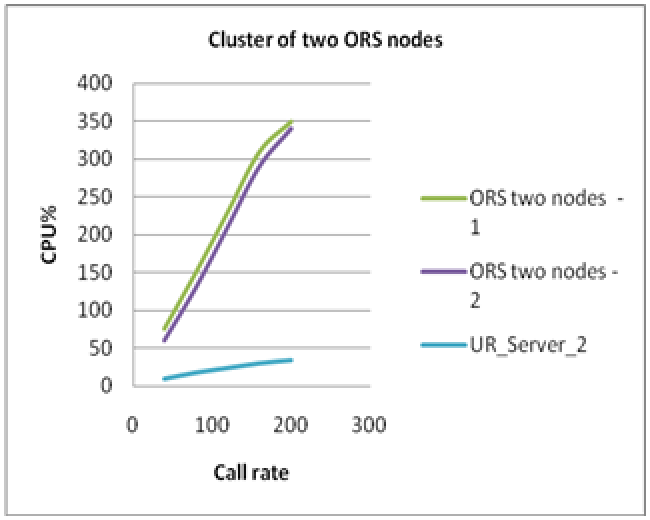


Figure 162: Genesys Orchestration Platform Voice Call Model with Two ORS Applications in a Cluster on Linux

Two ORS Applications in a Cluster on Linux Parameters

Table 155 shows the parameters considered for this scenario.

Table 155: Two ORS Applications in a Cluster on Linux

Application		40	80	120	160	200
ORS_0	CPU%	75	150	230	310	350
	Memory	550	600	650	680	730
ORS_1	CPU%	60	130	210	290	340
	Memory	530	620	650	730	740
Cassandra 1	CPU%	25	45	70	95	110
	Memory	300	600	900	1500	1700
Cassandra 2	CPU%	20	37	60	80	95
	Memory	400	550	950	1500	1700

Table 155: Two ORS Applications in a Cluster on Linux (Continued)

Application		40	80	120	160	200
Cassandra 3	CPU%	9	18	25	37	43
	Memory	300	390	410	1100	1700
T-Server	CPU%	10	20	29	39	48
	Memory	92	93	98	103	119
Router	CPU%	10	18	24	30	34
	Memory	449	463	476	489	493
StatServer	CPU%	10	21	32	44	53
	Memory	527	567	620	660	700

Five ORS Applications in a Cluster on Linux

In this call model, voice only interactions are routed utilizing five Orchestration services in a cluster, where both Orchestration service and Cassandra are running on Enterprise Red Hat Linux. The CPU and memory utilization is measured for all services at different CAPS.

The CPU% is measured on all cores of the machine.

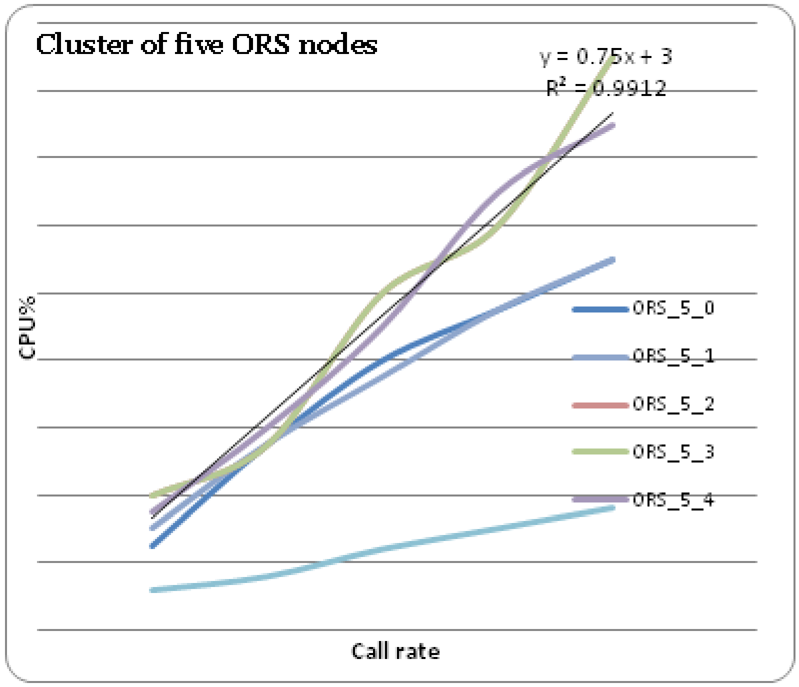


Figure 163: Genesys Orchestration Platform Voice Call Model with Five ORS Applications in a Cluster on Linux

Five ORS Applications in a Cluster on Linux Parameters

Table 156 shows the parameters considered for this scenario.

Table 156: Five ORS Applications in a Cluster on Linux

Application		40	80	120	160	200
ORS_0	CPU%	25	55	80	95	110
	Memory	510	550	560	590	600
ORS_1	CPU%	30	55	75	95	110
	Memory	520	540	570	590	600
ORS_2	CPU%	35	55	85	120	150
	Memory	320	340	350	370	400

Table 156: Five ORS Applications in a Cluster on Linux (Continued)

Application		40	80	120	160	200
ORS_3	CPU%	55	120	200	260	320
	Memory	550	600	620	670	700
ORS_4	CPU%	35	60	90	130	150
	Memory	320	350	370	380	400
Cassandra 1	CPU%	17	35	53	75	85
	Memory	550	820	1200	1700	1700
Cassandra 2	CPU%	18	35	55	75	87
	Memory	450	720	1100	1700	1700
Cassandra 3	CPU%	12	22	33	45	52
	Memory	450	800	1000	1600	1700
T-Server	CPU%	15	29	42	58	70
	Memory	93	96	102	108	115
Router	CPU%	12	16	24	30	36
	Memory	450	464	474	492	494
StatServer	CPU%	10	21	32	46	58
	Memory	525	564	605	665	705

Five ORS Nodes Cluster: Single WEB Server on Windows

In this scenario, there has been only a single Web Server utilized to provision SCXML applications. Though cache is enabled, constant changes were made on the SCXML application forcing a re-read of the application from the Orchestration Server. The resultant of this scenario did not made any changes to the expected outcome.

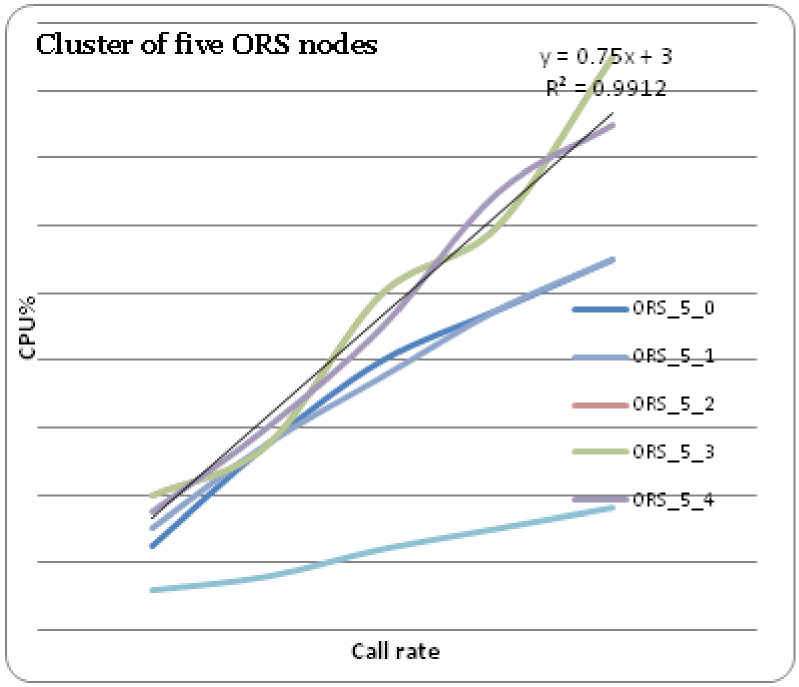


Figure 164: Five ORS Nodes Cluster with Single WEB Server on Windows

Five ORS Nodes Cluster: Single WEB Server on Windows Parameters

Table 157 shows the parameters considered for this scenario.

Table 157: Five ORS Nodes Cluster: Single WEB Server on Windows

Application		40	80	120	160
ORS_0	CPU%	30	65	125	140
	Memory	520	550	610	650
ORS_1	CPU%	50	90	130	175
	Memory	550	580	600	650
ORS_2	CPU%	25	60	95	125
	Memory	320	330	350	400

Table 157: Five ORS Nodes Cluster: Single WEB Server on Windows (Continued)

Application		40	80	120	160
ORS_3	CPU%	20	45	90	110
	Memory	290	330	350	400
ORS_4	CPU%	35	70	95	120
	Memory	320	350	370	400
Cassandra 1	CPU%	12	20	32	43
	Memory	320	500	1000	1500
Cassandra 2	CPU%	13	27	42	50
	Memory	300	500	900	1200
Cassandra 3	CPU%	15	33	60	72
	Memory	300	600	1000	1400
Router	CPU%	7	12	19	23
	Memory	434	448	461	465
WEB server	CPU%	5	9	14	18
	Memory	117	188	271	330

Performance Test Result Conclusions

The following list summarizes conclusions which are based on the performance test results from the scenarios and models in this chapter.

- ORS and URS have almost the same performance on Windows and Linux OS.
- ORS and URS performance (CPU%) have linear dependence on input call rate.
- ORS performance has inverse dependence on number of cluster nodes, for example, each node in 5 node cluster has ~1/5 CPU% usage of single node ORS configuration. See Figure 165, "ORS on Windows," on page 484 and Figure 166, "ORS on Linux," on page 485.
- Maximum call rate has been limited by environment, and not by ORS or URS performance.

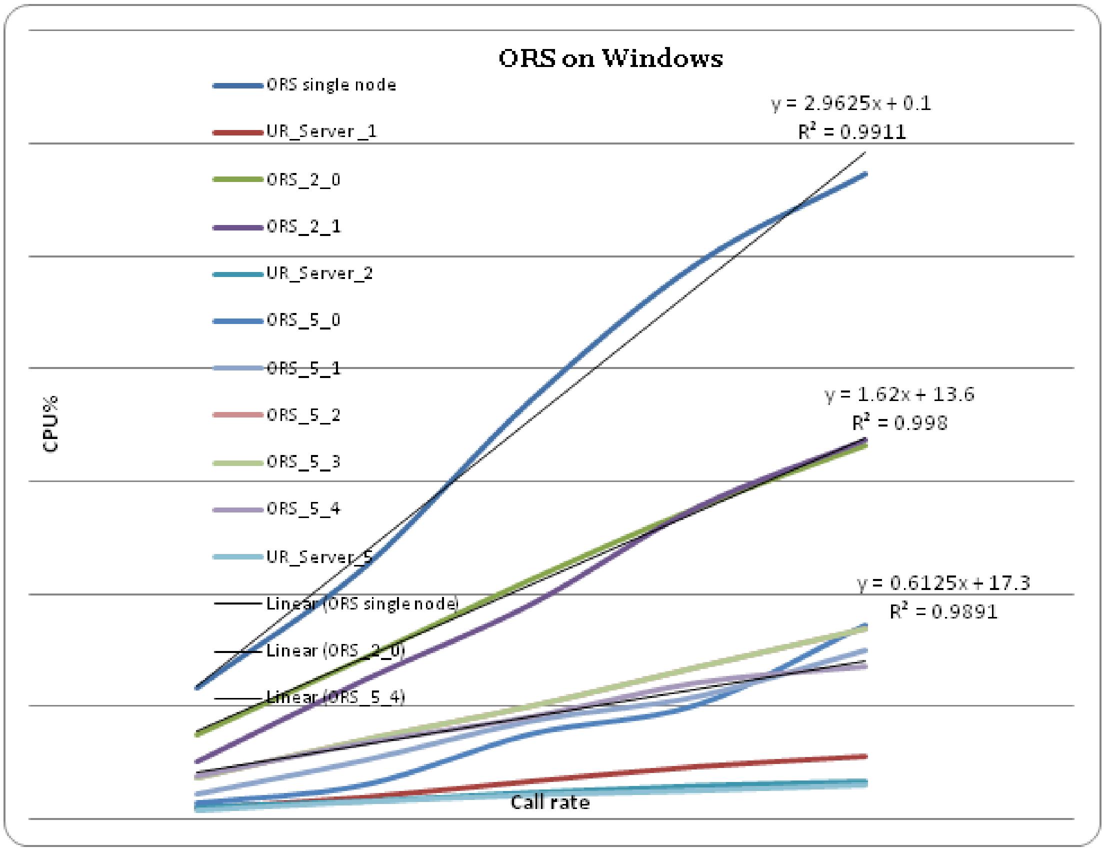


Figure 165: ORS on Windows

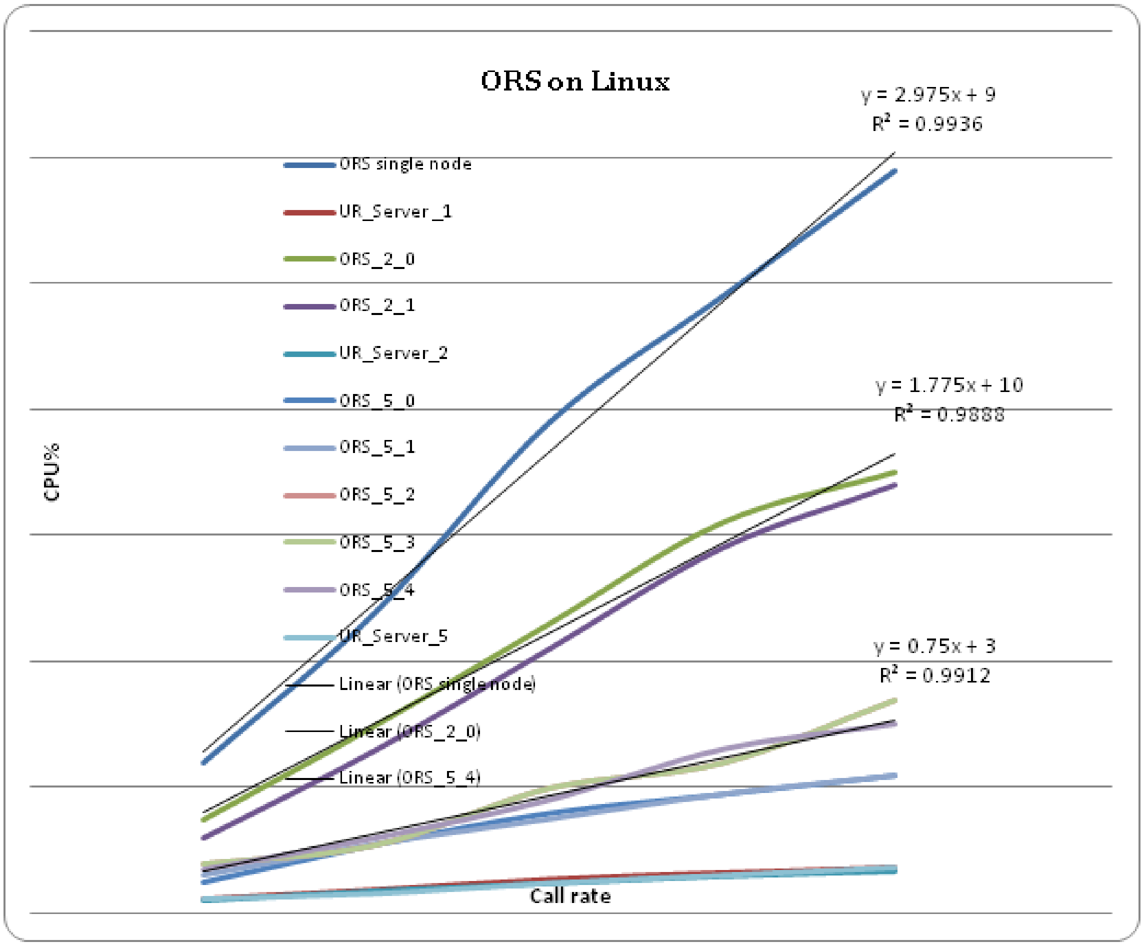


Figure 166: ORS on Linux

Hardware Specifications

Table 158, "Hardware Specifications," on page 486 provides the hardware specifications used for the load tests in the scenarios in this chapter. Each of the services was run on a separate machine with the configuration for Windows and Linux OS described in the table. Genesys recommends having hardware with this base line, or better, to achieve comparative results to those shown in the scenarios in this chapter.

Table 158: Hardware Specifications

Computer	Specifications
Windows	• Dell PowerEdge R710 • 2 Xeon X5560 * 4 cores, hyper-threading • 2.8 GHz, 8G • Windows 2003
Linux	• Dell PowerEdge R710 • 2 Xeon X5560 * 4 cores, hyper-threading • 2.8 GHz, 8G • RED HAT 5.4

Chapter

18 intelligent Workload Distribution

This chapter contains sizing guidelines and basic information required for deploying and capacity planning of Genesys intelligent Workload Distribution (iWD) release 8.0 on the Microsoft Windows operating system.

Note that this section uses the term *task*, which can be considered interchangeable with the term *interaction*.

This document should be used in conjunction with the *Genesys Database Sizing Estimator.*

- Methodology, page 487
- Sizing Calculations, page 489

Methodology

In order to understand how the system performance is dependent on multiple performance parameters, the iWD solution was tested with one test parameter changing while the rest of the parameters stayed constant, or had a small deviation. The system was brought up to the parameter point defined by the performance parameters and sustained for 1 hour. Statistical data was collected at each point of the performance parameter matrix. This data has been used to produce calibration coefficients for CPU usage estimation and database sizing.

To create the database sizing formulas, empirical data was collected based on Microsoft SQL Server. For iWD deployments using Oracle, the numbers will be approximately the same.

Performance Parameters

- Incoming task rate per second—7.5, 15, 30, 45, 60
- Number of tasks in backlog (in the iWD_Queued queue)—10000, 20000, 40000, 80000
- Number of agents in iWD Agent Group—500, 1000, 2000, 4000
- Attached data size in KB — 1, 2, 4, 8

Reference Scenario

The reference test scenario has 10 Departments configured. Each Department has 10 Processes.

In the *classification* phase, iWD rules are configured to set the business value and assign the task to the appropriate process using the IWD_ext_sourceProcessType task attribute (attached data value).

In the *prioritization* phase, linear rules were configured to set the priority and the re-prioritization interval.

Testing was carried out using a variation of the standard iWD business process (IWDBP).

Classification Rules

- 10 decision tables per Department
- 10 decision tables plus 2 linear rules per Business Process

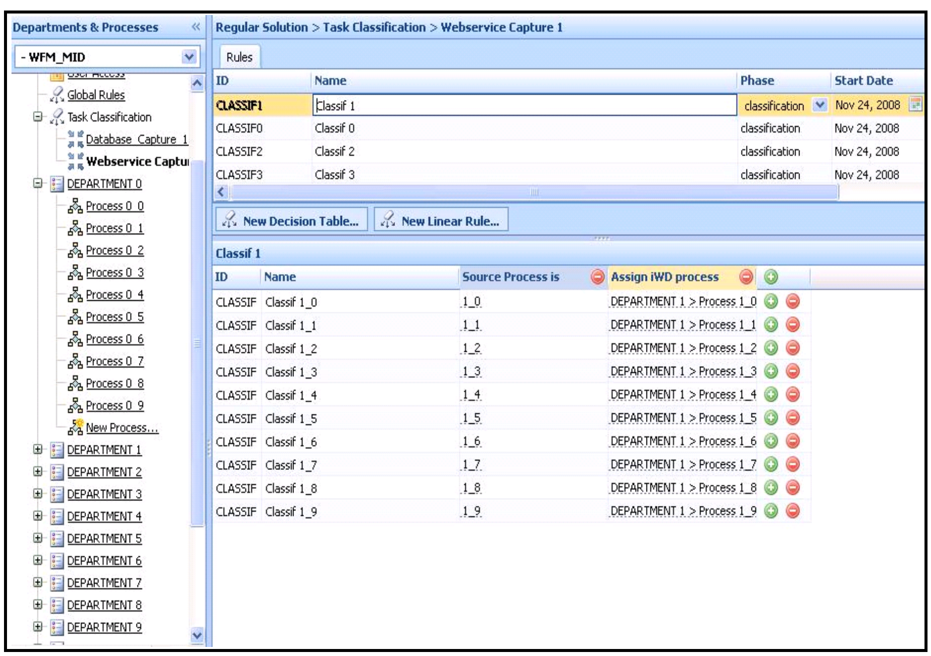


Figure 167: Classification Rules

Prioritization Rules

- 6 linear rules per Department plus 2 linear rules per Process

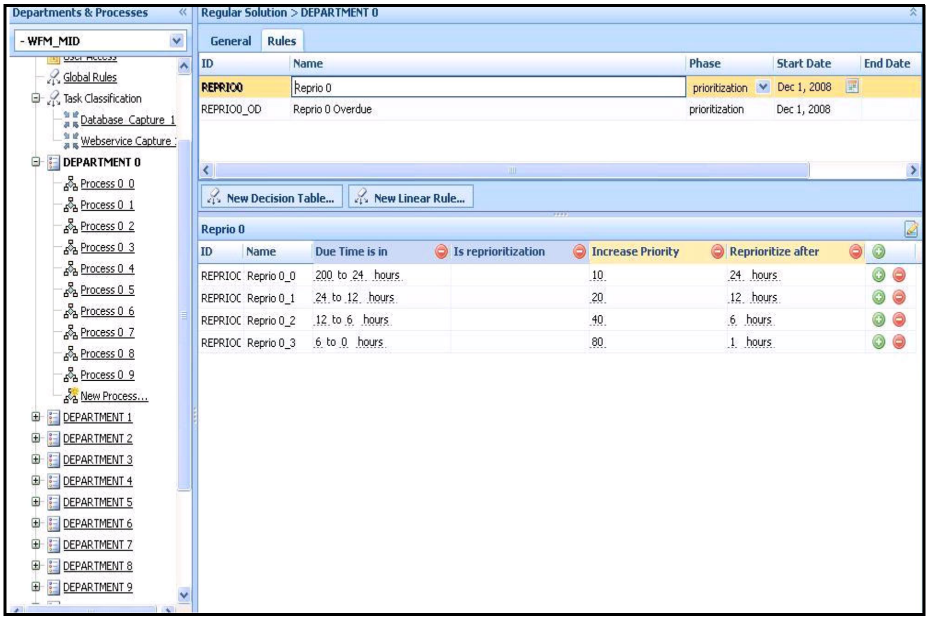


Figure 168: Prioritization Rules

Sizing Calculations

The purpose of the sizing calculations is to ensure that the system can handle the required peak incoming task rate and the databases have enough disk storage to store backlogged and archived tasks data to meet customer requirements.

iWD includes a number of services that work asynchronously. The sizing calculation takes into account that some services share the same resources and can start at the same time. Based on Genesys' measurement, Interaction Server defines the overall performance of the iWD solution.

For proper sizing the following constraints were used:

- Combined maximum CPU usage of all applications installed at a host does not exceed 80% of a host's total CPU.
- Maximum CPU usage does not exceed 80% of a host's total for multi-threaded applications and 80% of a single CPU core for single-threaded applications.
- Maximum memory usage does not exceed 80% of a host's total RAM.
- The iWD Runtime Node and iWD Data Mart node must be placed on separate hosts with at least 4 GB of RAM each.

- Number of agents with non-shared connections is limited to 4,000 per Interaction Server proxy.
- The total number of open TCP/UDP ports per host should not exceed 10,000.

Estimating Database File Size

Based on laboratory measurements, Genesys calculated the average data stored in each iWD database per task for the referenced scenario. To create the database sizing formulae, empirical data was collected based on Microsoft SQL Server. For iWD deployments using Oracle, the numbers will be approximately the same. Table 159 shows the average data per task stored in each iWD database.

Table 159: Average Data Size Stored in iWD Databases Per Processed Task

Database	KB Per Task
Interaction Server database	38.00
Interaction Server Event Log database	41.40
iWD Data Mart database	0.43

Tasks in the referenced iWD business process include about 20 keys of 50 bytes of attached data—about 1 KB in total. If attached data size is more than 1 KB, then the additional amount has to be added to the average record size for the Interaction Server database and the Interaction Server Event Log database sizing calculations.

Calculation Formulae

The following formulae can be used to calculate maximum file sizes (in KB) for each iWD database:

- *InteractionServerDatabaseSize= (IxnDbKbytesPerTask + UserAttachedDataSizeKbytes -1) *(TaskNumberInQueues + TasksExpirationTimeout * TaskRatePerDay)*
- *InteractionServerEventLogDatabaseSize = (IxnLogDbKbytesPerTask + UserAttachedDataSizeKbytes - 1) * IxnLogDbExpirationPeriod * TaskRatePerDay*
- *iWDDataMartDatabaseSize = iWDDataMartDbKbytesPerTask * DatamartDbRetentionPeriod * TaskRatePerDay*

Where:

- *TaskRatePerDay* is the average number of tasks processed by the iWD system (per day)
- *IxnDbKbytesPerTask* is the average number of bytes stored in the Interaction Server database per processed task (in KB)
- *IxnLogDbKbytesPerTask* is the average number of bytes stored in the Interaction Server Event Log database per processed task (in KB)
- *iWDDataMartDbKbytesPerTask* is the average number of bytes stored in the iWD Data Mart database per processed task (in KB)
- *TaskNumberInQueues* is the average number of tasks in backlog in all iWD queues except the iWD_Completed queue
- *TasksExpirationTimeout* is the average task expiration time set by iWD business rules (in days)
- *IxnLogDbExpirationPeriod* is the maximum period before the data in the Interaction Server Event Log database is pruned (in days)
- *DatamartDbRetentionPeriod* is the amount of time data is retained in the Data Mart database (in days)
- *UserAttachedDataSizeKbytes* is the amount of user attached data (in KB)

Estimating the Effective Task Rate

The number of tasks in the backlog (that is, in the queue), and the amount of attached data per task, impose additional load on iWD components.

The *effective task rate* is an integration parameter that accounts for:

- The task submission rate
- The size of the task backlog
- The task re-prioritization rate
- The attached data size per task

The following formula can be used to calculate the maximum effective task rate that iWD can handle for a specific re-prioritization scheme and attached data size requirements:

- *EffectiveTaskRate = (MaxInputTaskRate +TaskNumberInBacklog / ReprioritizationPeriodSec) * (1 + (K5 *(UserAttachedDataSizeKbytes -1)))*

Where:

- *EffectiveTaskRate* is the estimated effective task rate that is used for sizing calculations (in tasks per second)
- *MaxInputTaskRate* is the maximum incoming task rate that the system has to handle during peak hours (in tasks per second)
- *TaskNumberInBacklog* is the number of tasks in the backlog
- *ReprioritizationPeriodSec* is the average re-prioritization period (in seconds)

- *UserAttachedDataSizeKbytes* is the amount of user attached data (in KB)
- *K5* is the coefficient that characterizes the dependency of the attached data size and the effective task rate K5 = 0.061

Database Server Hardware Layout Selection

The overall iWD system performance is dependent on Interaction Server performance. Interaction Server performance sizing is described elsewhere in this document. However, we can state that Interaction Server performance depends on the complexity of the business process (mostly on the number of interaction queues) and on Interaction Server database performance.

Performance of the database server itself directly depends on physical data files access rate. Database tuning and choosing the right hardware (especially disk/RAID capabilities) is essential to achieving the desired performance of the iWD solution.

To characterize the dependence of the Interaction Server performance on physical data files access rate, Genesys conducted tests for "submit only" operations for different database hardware topologies:

- Single host for Interaction Server database and Interaction Server Event Log database on a single drive
- Separate host with a single drive for each database
- Separate host for each database with Interaction Server database hosting on a multi-disk RAID

A "submit only" operation test was performed by deactivating all iWD business process strategies and submitting tasks at the maximum rate.

A series of tests was performed with various values configured for Interaction Server's number-of-database-connections configuration option. Based on the test results, the optimum setting for this option was found to be 12.

Table 160 on page 493 below illustrates the estimated maximum effective task rate achieved both in the laboratory, and in "submit only" task rate for different reference test system layouts with number-of-database-connections set to value 12.

Table 160: Maximum Effective Task Rate for iWD System Depending on Database Layout

Deployment Type	Database Layout	Interaction Server "Submit Only" Tasks/Sec	iWD Maximum Effective Task Rate Task/Sec (estimated)
Small	Single host for Interaction Server database and Interaction Server Event Log database on a single drive	790	20
Medium	Separate host with single drive for each database	1,580	45
Large	Separate host for each database with Interaction Server database hosting on a multiple disk RAID	2,450	75

If the maximum effective task rate does not exceed 20 tasks/sec, all iWD databases can reside on one physical host. It is highly recommended that database files be allocated on separate physical disks. If the maximum effective task rate is 45 tasks/sec or higher, each iWD database has to be running on a separate physical host and the Interaction Server database data files have to be residing on a fast RAID.

System performance depends on:

- Database performance
- The condition of the physical database file
- Table space fragmentation

It is critical for the database administrator to perform periodic disk defragmentation and table space defragmentation.

Estimation of CPU Utilization

The purpose of this calculation is to check if CPU utilization is within the limits of the constraints.

The following formula can be used to calculate CPU utilization for each iWD Genesys component:

- *CPU = (CPUO + K1 * EffectiveTaskRate + K2 * ActiveTasksNumber + K3 * NumberOfAgents) * CPU_Normalization*

Where:

- *CPU0* is a CPU utilization offset compensating for non-linear behavior at idle and very low task rates
- *EffectiveTaskRate* is calculated in “Estimating the Effective Task Rate” on page 491.
- *K1* is the coefficient that characterizes the dependency of CPU utilization and tasks rate
- *K2* is the coefficient that characterizes the dependency of CPU utilization and tasks in backlog (in the iWD_Queued queue)
- *K3* is the coefficient that characterizes the dependency of CPU utilization and the number of active agents defined in the iWD solution
- *CPU_Normalization* is the coefficient that can be used to normalize hardware used at the customer site to the equipment used in the Genesys performance laboratory for these tests

Table 161: Calibration Coefficients for CPU Utilization (Dual Intel X Xeon x5560 CPU, 2.8 GHz)

Application	CPU0	K1	K2	K3
Interaction Server	1.1	1.28	2.23E-05	2.70E-03
iWD Runtime Node	0.87	0.68	4.03E-05	0
iWD Data Mart Node MAX	53	3.46	1.19E-05	1.42E-03
SQL Server	0.5	1.27	1.17E-05	0
Universal Routing Server	0	0.177	4.85E-05	0.00E+00
Stat Server	0.1	0.038	4.27E-05	0.00E+00

Memory Usage

iWD Runtime Node uses 580 MB when the system is idle and up to 1.5 GB during the inbound task rate of 60 tasks per second with 4,000 agents.

The iWD Data Mart node takes 460 MB at idle and up to 1.8 GB during the inbound task rate of 60 tasks per second with 4,000 agents.

Interaction Server takes up to 580 MB during the inbound task rate of 60 tasks per second with 4,000 agents.

Microsoft SQL Server takes up to 2.6 GB with 2 GB memory usage limits setting.

iWD Runtime Node, Data Mart node and SQL Server hosts must have at least 4 GB of RAM.

System Sizing Example

Using the above formulae and coefficients Genesys calculates sizing for the following system profile:

Table 162: Example System Profile

System Profile	Count/Units
Tasks Expiration Period	30 (days)
Data Mart DB Retention Period	365 (days)
Incoming Task Rate	200000 (per day)
Average re-prioritization Period	0.8 (hours)
Maximum Incoming Task Rate	30 (per second)
Average Number of Tasks in Backlog	40000
Number of Agents	2000
Attached Data Size	4 (KB)

Estimating the Maximum Effective Task Rate

- *EffectiveTaskRate = (MaxInputTaskRate -+ TaskNumberInBacklog/ ReprioritizationPeriodSec) * (1 + (K5 *(UserAttachedDataSizeKbytes -1)))*
- *EffectiveTaskRate = (30 + 40000/0.8/3600)) * (1 + 0.061 * (4-1)) = 51.92 tasks/sec*

Select the Database Server Hardware Layout

If the maximum effective task rate is 51.92 tasks per second, then each iWD database has to be running on a separate physical host and the Interaction Server database data files have to be residing on a fast RAID.

Estimating the CPU Utilization

Using formula and calibration coefficients for CPU utilization from Table 161 on page 494, we can calculate CPU utilization for each iWD component.

- Interaction Server
  1.1 + (1.28 * 51.92) + (2.23E-05 * 40000) + (2.70E-03 * 2000) = 91.63%
- iWD Runtime Node
  0.87+ (0.68 * 51.92) + (4.03E-05* 40000) = 47.23%
- iWD Data Mart
  53 + (3.46* 51.92) + (1.19E-05* 40000) + (1.42E-03*2000) = 284.20%
- Microsoft SQL Server
  0.5 + (1.27* 51.92) + (1.17E-05 * 40000) = 84.55%
- Universal Routing Server
  (0.177* 51.92) + (4.85E-05* 40000) = 13.59%
- Stat Server
  0.1 + (0.038* 51.92) + (4.27E-05* 40000) = 4.31%

Estimating the CPU Clock Scaling

The results for CPU load estimation are given for DELL® servers with Dual Quad-core Intel® Xeon® x5560 2.8 GHz CPU. For Intel Xeon CPUs with L2 cache size above 2 MB. The performance of the applications scales linearly with the clock frequency.

For example, if you are upgrading from Xeon® x5560 2.8 GHz CPU to Xeon® x5580 3.2 GHz, you should expect a performance boost of about 14% (3.2 GHz/2.8 GHz). In this example, CPU usage can be reduced by 12.5% accordingly (2.8 GHz/3.2 GHz).

- Interaction Server
  91.63% * 0.875 = 80.17%
- iWD Runtime Node
  47.23% * 0.875 = 41.33%
- iWD Data Mart
  284.20% * 0.875 = 248.51%
- Microsoft SQL Server
  84.55% * 0.875 = 73.98%
- Universal Routing Server
  13.59% * 0.875 = 11.89%
- Stat Server
  4.31% * 0.875 = 3.77%

Estimating the Database Sizing

- Interaction Server Database Size
  (38 KB + (4-1))*(40000 + 30 * 200000) = 236.19 GB
- Interaction Server Event Log Database Size
  (41.4 KB + (4-1))*(30 * 200000) = 255.78 GB
- iWD Data Mart Database Size
  0.42624 KB * 365 * 200000 = 29.67 GB

Note: Data stored in the iWD Data Mart database is highly configurable and dependent on a particular data mapping setting. That can increase the size of the data stored in the database per record.

Determining the Number of Hosts

When multiple Genesys components are running on one physical host, the maximum CPU usage should not exceed 80% of the total CPU power of the host, and the maximum memory usage should not exceed the physical RAM size. The total CPU power is defined by the following formula:

*TotalCPU = 100% * Number_of_CPUs * CPU_Number_of_Threads.*

For a host with Dual Intel Xeon® x5560 it is 1600%.

If the maximum effective task rate is 52 tasks per second, then each iWD database has to be running on a separate physical host and the Interaction Server database data files have to be residing on a fast RAID. Separate hossts are needed for the iWD Runtime and Data Mart nodes.

For the system deployment based on the input profile in Table 162 on page 495 6 hosts and a RAID for Interaction Server database are needed, one for each of the following:

- Configuration Server and Interaction Server
- iWD Runtime Node
- iWD Data Mart Node
- Interaction Server database with RAID
- iWD Data Mart database
- Interaction Server Event Log database

Reference Scenario - Hardware and Software Used

All computers used were DELL® R710. Table 163 gives their specification details.

Table 163: Reference Scenario - Hardware and Software Used

Item	Details
CPU	Dual Quad-core Intel® Xeon® x5560 2.8 GHz
Memory	8 GB
Network interface card	Dual port Broadcom BCM 5709
Hard Drive	X2 SAS(7.2K RPM) 67 GB (MSSQL Data RAID 0 4x300 GB)
OS	Windows 2003 SP2
Database server	MSSQL 2005

Chapter

19 Genesys Co-browse

This chapter provides sizing and capacity guidelines for Genesys Co-browse. It contains the following sections:

- Hardware, page 499
- Node Calculation, page 499
- Disk Space Usage Calculation, page 500
- Sizing Calculation, page 501

Hardware

The recommendations in this chapter are based on a deployment where Co-browse Server allocates 4 CPU, 4GB RAM on an Intel Xeon E5560 2.8 GHz server.

Note: This is a minimum recommended specification for node deployment in a VM environment. For dedicated hardware, Genesys recommends that you allocate 8GB RAM.

Node Calculation

The Co-browse solution performance can be scaled by the number of nodes and hardware performance. The node number calculation takes into account the number of agents, session duration and complexity (the number of page transitions), and the required reserved capacity for handling one node failure.

Genesys recommends that your Co-browse deployment includes a minimum of three nodes, which can support up to 2073 agents. Three nodes provide the ability to handle both node failure and the minimum recommended replication factor of 2 for stored session data and Co-browse operational data. To ensure higher reliability, the replication factor can be increased to 3 which will increase the required disk volume and network traffic, but will not decrease system throughput.

In Table 164, the number of nodes required to support the given number of agents is for typical session duration and complexity. You can use the formula in “Sizing Calculation” on page 501 to calculate the number of nodes for a particular deployment and usage profile.

Table 164: Dependency of active nodes and supported maximum number of agents

Required number of nodes	Maximum number of agents
3	<2073
4	4091
5	6729
6	9963
7	13770
8	18129
9	23020
10	28421

Disk Space Usage Calculation

Use Table 165 to estimate the required hard disk space needed to store session data. The replication factor should be 2 or more.

Table 165: Total database size (not compressed)

Replication factor	Size per session (MB)	Session rate	Hour (MB)	Day (GB)
2	0.027	5	480	11.52
3	0.04	5	720	17.28

Sizing Calculation

Use the formula below to estimate the maximum node CPU utilization. The number of nodes should satisfy both the solution quality criteria and the maximum CPU utilization limit.

$$CPUT \le CPUR + CPUN = \frac{NA \times KR}{TR \times N \times (N-1)} + \frac{NA \times KC}{(N-1) \times TT}$$

Note: In Table 166, the values in square brackets "[ ]" were used to calculate the number of nodes and maximum number of agents in "Node Calculation" on page 499.

Table 166: Formula values

Value	Description
N	Number of nodes
NA	Number of agents (total)
TR	t_reconnect (time interval for agents from the downed node to reconnect). **[15 seconds]**
TT	Talking time (average). **[300 seconds]**
CPUR	CPU consumption required to establish a session.
CPUN	CPU consumption required to process a session.
CPUT	Maximum total node CPU consumption. **[60%]**
KR	CPU consumption required on a single node to establish 1 session per second. **[2.5]**
KC	CPU consumption required on a single node to process 1 session. **[0.7]**

Chapter

20 Generated System Traffic

This chapter presents information on the traffic generated by the system. It covers the following topic:

- Traffic Generated by Components and Solutions, page 503

Traffic Generated by Components and Solutions

This section presents information on traffic generated by components and inbound solutions.

Traffic Generated by Framework 7.x Components

Table 167 provides basic data about link traffic among various Framework components. Use this information to help you determine the optimal component location on the network.

Table 167: Traffic Generated by Framework 7.x Components

Component 1	Component 2	Elements Determining Total Message Traffic	Primary Transaction Types	Average Message Length	Number of Messages per Transaction	Total Traffic Volume	Timelines of Message Delivery
Config Server	DB Server ICON	frequency of configuration changes	config changes	1 KB	one per change	very low	not critical
Config Server	Any client application	frequency of configuration changes	regular config updates	1 KB	one per update	low	not critical
Any client application	Message Server[2]	type of client application (see the log events documentation)	log events	0.5 KB	one per event	low (standard log output level)	not critical
Message Server	DB Server	number of message server clients	log records	0.5 KB	one per event	medium	not critical
Message Server	Message Server - Solution Control Server	• number of configured alarm conditions • overall system stability	alarms	0.5 KB	two per alarm condition	low	not critical

Table 167: Traffic Generated by Framework 7.x Components (Continued)

Component 1	Component 2	Elements Determining Total Message Traffic	Primary Transaction Types	Average Message Length	Number of Messages per Transaction	Total Traffic Volume	Timelines of Message Delivery
Solution Control Server	Solution Control Interface	• number of solutions • number of components per solution • number of computers serving the installation • number of configured alarm conditions • overall system stability	• solution control requests • changes in solution and application statuses • alarm information	0.5 KB	up to 20	low	not critical
DB Server for Log Database	Solution Control Interface	frequency of log view changes	log record blocks	50 KB	a block of 100 log records per single view change	medium	not critical
Solution Control Server	Local Control Agent	number of applications located on the same computer	• application startup • shutdown • redundancy arbitration commands	0.25 KB	one per control operation	very low	not critical

Traffic Generated by Inbound Solutions

Table 168 provides basic data about traffic among various inbound solutions with Reporting and IP Telephony Components included. The results shown in this table are based on Genesys performance test results.

Table 168: Traffic Generated by 7.5 Inbound Solution with Reporting Components

Component1	Component2	Protocol	Estimated bi-directional traffic	Criticality	Comments
T-Server	ICON	TLIB	9 Kbyte/call	Very high	Traffic was measured on generic average call flow.
T-Server	StatServer	TLIB	9 Kbyte/call	Very high	
T-Server	URS	TLIB	3.5 Kbyte/call	Very high	For one routing attempt.
URS	StatServer	StatLib	0.1 Kbyte/call	Very high	Could depend on strategy complexity.
StatServer	SS client	StatLib	0.1 Kbyte per statistic per update	High	
Interaction Concentrator (ICON) server	DBServer	DBLib	9 Kbyte/call	Very high	Traffic was measured for MS SQL on generic average call flow.
Genesys Info Mart	DataBase	JDBC	7 Kbyte/call	Very high	Aggregation and data transformation for supporting generic call flow.

Table 168: Traffic Generated by 7.5 Inbound Solution with Reporting Components (Continued)

Component1	Component2	Protocol	Estimated bi-directional traffic	Criticality	Comments
SIPServer	SIP Components	SIP	Simple incoming call: 18.5 Kbyte/call	Very high	Significantly depends on call flow and network conditions. If network connection is poor, messages could be re sent according SIP protocol.
CTI Link	T-Server	Link vendor appropriate	2 Kbyte /call	Critical	Given number is an estimation only and could vary for different switch vendors and configurations. Traffic was measured on generic average call flow.
Stream Manager	Stream component	RTP	• G711 codec: 22 Kbyte/sec per call leg • G729 codec: 7.67 Kbyte/sec per call leg • GSM codec: 6 Kbyte/sec per call leg	Critical	Examples of stream components are: Media Gateways, Genesys Voice Platform, SIP phones etc.
Genesys Integration Server (GIS)	Voice client	GSAP	12 Kbyte/call	Very high	Depends on call flow

Table 168: Traffic Generated by 7.5 Inbound Solution with Reporting Components (Continued)

Component1	Component2	Protocol	Estimated bi-directional traffic	Criticality	Comments
GIS	Voice client	SOAP	27 Kbyte/call	Very high	Depends on call flow
GIS	Stat client	SOAP	2.36 Kbyte per update per sec	Very high	Depends on call flow

Generic Call Flow

The generic call flow is as follows:

--> Queue --> 400 DNs -- ttransfer, attach ~110bytes --> Route Point --> 9 queues(100 DNs each) -->

--> 90% agents talk = 10 sec, release

--> 10% agents --ttransfer --> 2 queues(50 DNs each) -->

--> talk = 10 sec, release

Inbound Solution with Reporting Components Traffic Flow

Figure 169 shows the Inbound Solution with reporting components traffic flow.

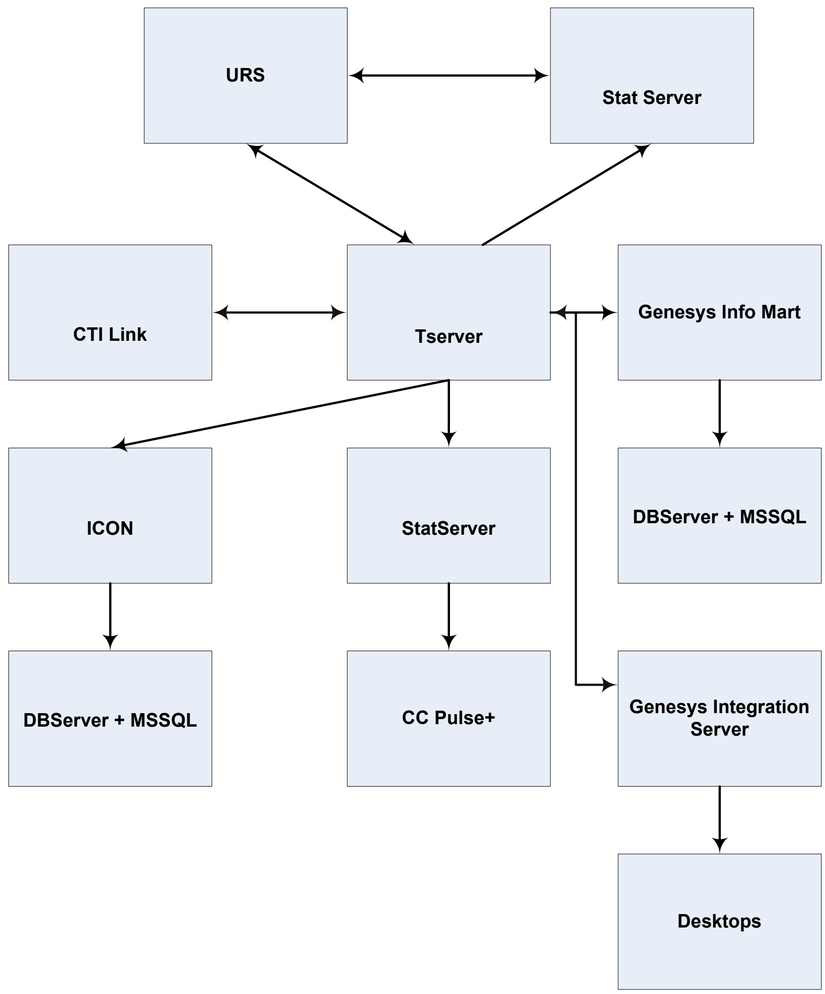


Figure 169: Inbound Solution with Reporting Components Traffic Flow

Chapter

21 Database Sizing Estimation

This chapter presents references to information that you may use to estimate the size of databases for Genesys components with convenient Microsoft Excel dynamic spreadsheets for various products. With theses spreadsheets, you may enter configuration data and automatically calculate an estimation of the database size for your environment.

In addition, this chapter provides references to other sections in this guide and other Genesys deployment guides that provide further information about database sizing estimation. It is recommended that you review the information in these additional resources before you begin any spreadsheet calculation. You might find information about alternative methods of estimation that could be more appropriate for your database.

When using a spreadsheet, keep in mind that it is a tool that is meant to help you determine an "approximate estimate" for the required disk memory volume. It is not providing the actual size of the database.

Some of the references provided in this chapter may contain more information on how the estimated size is relevant to the actual size.

This chapter contains:

- Database Sizing Estimator Spreadsheets, page 512
- Reference Information, page 513

Database Sizing Estimator Spreadsheets

This section provides information on accessing the Genesys database sizing estimator spreadsheets.

The *Genesys Database Sizing Estimator Spreadsheets* document is a Microsoft Excel file that provides tabs to access sizing spreadsheets for some of the Genesys products, by release.

Genesys 8.0

Genesys Database Sizing Estimator 8.0 document for the 8.0 release (80_DBSizing_WorkSheets.xls), is a system-level document, and is available on:

- Genesys Customer Care Website at http://www.genesys.com/customer-care. From the top menu, click Genesys Knowledge Base; Browse. Select the “system level documents by release" tab; System-Level Documents - 8.0: *Database Sizing Estimator 8.0.*
- Genesys Documentation Library DVD: see the System-Level Documents page. You can order this DVD by e-mail from Genesys Order Management at orderman@genesys.com.

Genesys 7.6

The *Genesys Database Sizing Estimator 7.0* document for the 7.6 release (760_DBSizing_WorkSheets.xls), is a system-level document, and is available on:

- Genesys Customer Care Website at http://www.genesys.com/customer-care. From the top menu, click Genesys Knowledge Base; Browse. Select the “system level documents by release" tab; System-Level Documents - 7.6: *Database Sizing Estimator 7.0.*
- Genesys Documentation Library DVD: see the System-Level Documents page. You can order this DVD by e-mail from Genesys Order Management at orderman@genesys.com.

Genesys 7.5

The *Genesys Database Sizing Estimator 7.0* document for the 7.5 release (750_DBSizing_WorkSheets.xls), is a system-level document, and is available on:

- Genesys Customer Care Website at http://www.genesys.com/customer-care. From the top menu, click Genesys Knowledge Base; Browse. Select the “system level documents by release" tab; System-Level Documents - 7.5: *Database Sizing Estimator 7.0.*

Genesys Documentation Library DVD: see the System-Level Documents page. You can order this DVD by e-mail from Genesys Order Management at orderman@genesys.com.

Reference Information

For more information, refer to the resources listed below. The Genesys product documentation is available on:

- Genesys Customer Care Website at http://www.genesys.com/customer-care.
- Genesys Documentation Library DVD, which you can order by e-mail from Genesys Order Management at orderman@genesys.com.

Resources

See the current versions of the following documentation:

- Configuration Database: see the *Management Framework Deployment Guide.*
- Data Mart and Operational Data Storage (ODS): see the *Reporting Deployment Guide*.
- Genesys Info Mart: see: "Genesys Info Mart 7.x Solution" on page 111 in this guide.
- Genesys Voice Platform: see "Genesys Voice Platform 7.x" on page 329 in this guide.
- Multimedia: see the *Multimedia Deployment Guide,* and "eServices (Multi-Channel Routing and Multimedia)" on page 21 in this guide.
- Operational Data Storage (ODS): see the *Reporting Deployment Guide.*
- Outbound Contact: see the *Outbound Contact Deployment Guide*, and"Framework, Reporting, Routing, Outbound, Voice Callback" on page 24 in this guide.

Chapter

22 Genesys WebRTC Service

The Genesys WebRTC Service allows your agents and customers to place voice or video calls from their web browser without downloading special plug-ins or apps.

For hardware sizing and performance information on this product, consult the Genesys Technical Documentation website at http://docs.genesys.com/Documentation/WRTC/8.5.2/Deployment/HardwareSizingInformation.

Genesys

Chapter

23 Genesys Web Engagement

Genesys Web Engagement provides the ability to monitor, identify, and proactively engage web visitors in conversations that match business objectives.

For hardware sizing and performance information about this product, consult the Genesys Technical Documentation website at http://docs.genesys.com/Documentation/GWE/8.1.2/Deployment/Sizing.

Appendix

A Performance Improvements

This section discusses the performance improvements for Genesys 7.x release and covers the following topics:

- Configuration Server Performance Improvements, page 519
- Logging Improvements, page 519
- Stat Server Performance Improvements 7.0 vs. 6.5, page 522
- Skill-Based Routing Performance Enhancement 7.0, page 523

Configuration Server Performance Improvements

Availability, Scalability, and Performance Improvements

- Startup time reduced up to 50%
- Enabling technology
  - New startup algorithm
    - Simultaneous initialization and data reading
  - Logical errors detection algorithm
    - Automatic correction or report of logical discrepancies

Note: Such errors no longer prevent Configuration Server Startup.

Logging Improvements

With the new options available in Genesys 7, the log performance can be up to five times better than in Genesys 6, depending on platform.

Log Test Description

Environment

- Two hosts are used in the test environment:
  - Dell Workstation PWS530, with 2 Xenon™ CPU 1.8GHz
  - Dell OptiPlex, Intel Pentium III CPU, 600MHz
- The network between workstations is 100 MB.
- Maximum disk write capacity on first workstation is about 12-13 MB/s.
- For all tests, the same configuration is used.

See Figure 170.

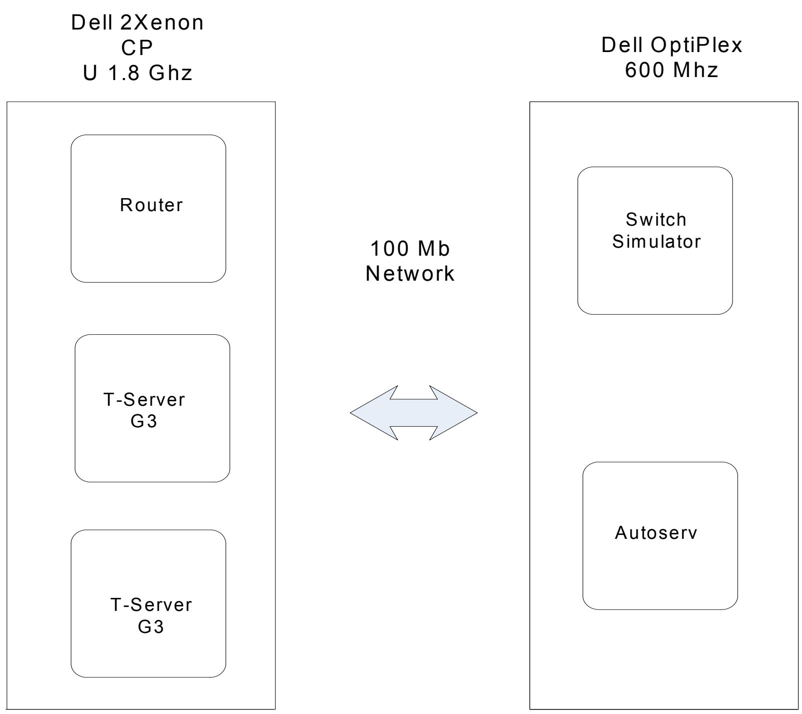


Figure 170: Dell 2 Xenon CPU 1.8 Ghz

Solution Performance Test

1. Two software environments are used: 6.5 and 7.0
   a. Environment 6.5:
      - Router 6.5.110.0
      - Stat server 6.5.101.06
      - T-server G3 6.5.306.01
   b. Environment 7.0:
      - Router 7.0.000.05
      - Stat server 7.0.000.11
      - T-server G3 7.0.011.00

2. Log level is set to the maximum for all applications and is not changed during any of the tests.
3. Log buffering is set to the value “true” for 6.5 applications.
4. All tests are started with 3 calls per second, and the number of calls are increased during test. Results are omitted because the test goal is to find upper limit for call volume.
5. Test is considered successful if there are no abandoned calls during a ten minute interval and are failed or stopped otherwise.
6. Number of calls, CPU, and disk usage are used for test interpretations.

See Table 169 for test results.

Table 169: Test Results

NN	Application	Output	Calls
1	T-Server 6.5	File	7
	Router 6.5	File	
	Stat Server 6.5	File	
2	T-Server 7.0	File	24
	Router 7.0	File	
	Stat Server 7.0	File	
3	T-Server 7.0	Memory map	36
	Router 7.0	Memory map	
	Stat Server 7.0	Memory map	

Results

The test shows significant performance improvement for Genesys 7.0 applications when compared to the 6.5 release. This is due to improved applications design and logging subsystem redesign: the relation is approximately 50-50.

According to the test results, this hardware configuration is recommended for call centers that have an average volume of about 12-18 calls per second and in cases where full logging is necessary. This volume is in middle of calculated results, which allows us to guarantee that the environment will be stable and that there will be almost double reserve for peak loading.

The memory-mapped log output is recommended for troubleshooting cases in a working environment as an output with minimum impact for performance; this output is also recommended for call centers with maximum activity.

If greater call volume is needed, use CPU with greater frequency and setup memory-mapped log output.

Stat Server Performance Improvements 7.0 vs. 6.5

Availability, Scalability, and Performance Improvements

- Ability to apply more statistical filters in a Stat Server instance because of improved filtering mechanism.
- Reduced start-up time, which increases real time statistics availability.
- Increased maximum amount of objects for which Stat Server can calculate statistics:
  - Example of usage: the same statistics can be ordered for more groups of agents.

Stat Server Performance

Improvements 7.0 vs. 6.5

All stat Server tests performed on Micron Netframe 2101 2 CPU Pentium 2 400 Mhz, 512 MB of RAM, Win2000

Performance is measured for 5, 10, and 20 seconds update interval for statistics. See Figure 171

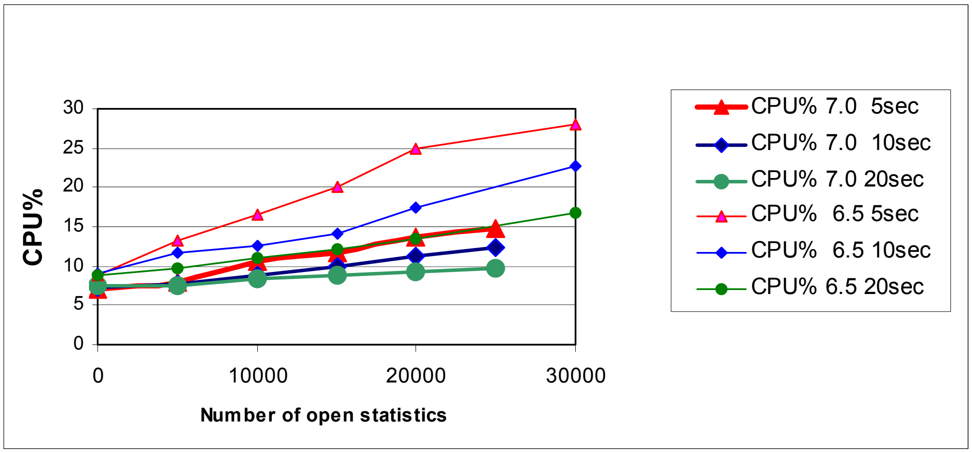


Figure 171: Stat Server Performance Improvements 7.0 vs. 6.5

Skill-Based Routing Performance Enhancement 7.0

- Performance for skill-based routing increased 4-5 times compared to release 6.x.

Note: Routing target is *skill expression*.

- For example:
  v. 6.5 on NT: 16 calls/sec takes 65.4% CPU
  v. 7.0 on NT: 16 calls/sec takes 13.4% CPU
- Hardware: Micron Netframe 2101 2 CPU Pentium

Note: Maximum number of calls per second will be determined later.

- Endurance run 120 hours at 24 calls/sec.

Note: Performance for ACD Queue (agent/virtual group) routing same as in 6.5.

See Figure 172 for Router CPU on List900 strategy.

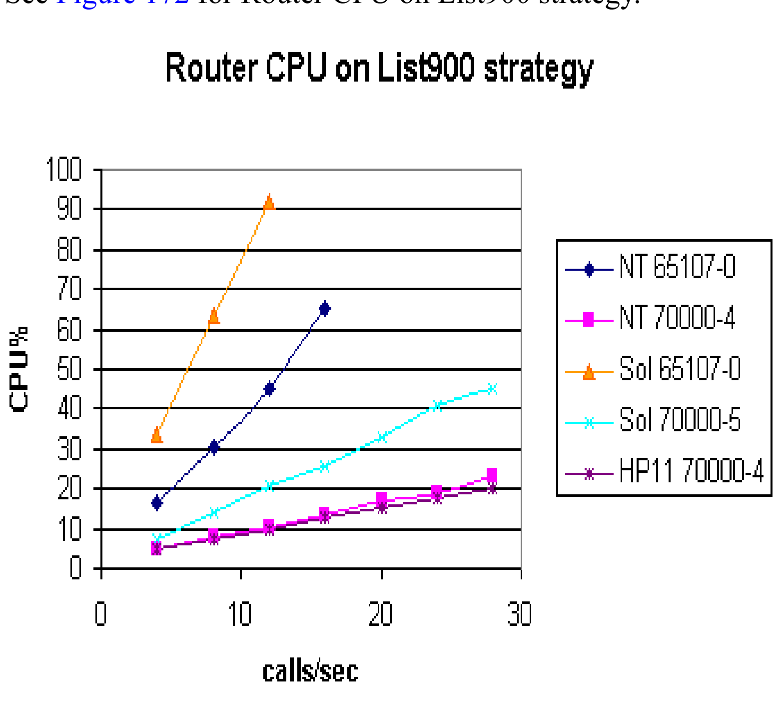


Figure 172: Router CPU on List900 Strategy

Appendix

B Log Database 7.0 Sizing Examples

This appendix contains:

- Examples, page 525
- Case 1, page 526
- Case 2, page 526
- Case 3, page 527
- Log Length Dependence for 7.0, page 528

Examples

Size of Log Record

Average size of the single Log Record (record in table G_LOG_MESSAGES), considering index space, is 500 bytes.

Log Record Attributes

Some of the Log Records (in particular, Log Records of level INTERACTION) would have several Attributes attached.

- Each Attribute attached to the Log Record is stored in Log Database as a record in table G_LOG_ATTRS.
- Average size of the single Log Record Attribute (record in table G_LOG_ATTRS) considering index space, is 300 bytes.

Log database space

Total amount of space, allocated by the log database, depends on many factors, in particular:

- How many applications are running.
- Whether or not Network log output is enabled for the applications.
- Log Messages of what level that are sent to the network log output.
- Call volume.

Consider the following simple environment as an example:

- DB Server
- Configuration Server
- Message Server
- Solution Control Server
- T-Server
- URS
- Stat Server

Case 1

Only STANDARD-level log messages are sent to the network log output

Average amount of Standard log messages produced by application could be estimated as approximately 100 per day. In our sample environment, 700 log messages would be written to the database each day. Most of the Standard log messages do not contain attached attributes. Therefore, the size of the log database needed to store log messages produced in 1 day could be estimated as ~340K. In this situation, the amount of space needed to store log records produced within 1 week is ~2.5 MB.

Note that this estimation assumes there is no error condition in our environment. If there is an error condition (such as one of the applications is not started, and other applications are repeatedly trying to connect to it), more log messages could be produced. Since error conditions do not usually persist for a long period of time, the temporary high volume of the log messages could be disregarded in our calculation.

Case 2

INTERACTION-level log messages are sent to the network log output

Interaction-level log messages contain information about ongoing interactions, and on the average has 5 attached attributes each. Therefore, in cases where Interaction-level log messages are sent to the network log output, the size of the log database would depend upon the call volume.

Not all applications produce Interaction log messages. For our sample environment, only T-Server, URS, and Stat Server produce such log messages.

The amount of Interaction log messages per call depends significantly upon the call scenario and the applications.

1. For example, let us assume that:
   - T-Server produces on the average 10 Interaction-level log messages per Interaction.
   - URS produces on the average 5 Interaction-level log messages per Interaction.
   - Stat Server produces on the average 5 Interaction-level log messages per Interaction.
2. Let us assume that:
   - The average call volume is 5 calls per second.
3. Therefore:

   (10 + 5 + 5)*5 = 100 Interaction log messages will be produced each second,

   with

   100 * 5 = 500 attached attributes records
4. These log records would require:

   100*500 + 500*300 = 195Kb of log database space per second

   Which for one hour is:

   195 * 60 * 60 = ~690 MB of log database space.

Note: The Interaction log messages are targeted for **test environments** in order to tune the interaction-related messaging between applications. It is not recommended to send Interaction-level log messages to the Network log output in the production environment.

Case 3

TRACE-level log messages are sent to the network log output

Applications produce even more Trace log messages than Interaction or Standard log messages. Therefore, it is also not recommended to send Trace log messages to the Network log output in the production environment.

Log Length Dependence for 7.0

- Call Flow—8 calls/sec, 5000 calls made.
- Attach data ~ 108 bytes
- Each application generates snapshot log file ~ 128 Kb always.

Table 170 specifies 7.0 applications:

Table 170: 7.0 Applications

Application	All		Trace		Standard		None		All/None Rating
	CPU%	Log (MB)	CPU%	Log (MB)	CPU%	Log (MB)	CPU%	Log (MB)	
T-Server	32.0	318.0	15.1	26.200	13.5	0.001	13.3	0.001	2.4
T-Server backup	21.0	269.0	6.3	0.001	5.8	0.001	6.0	0.001	3.5
HA-proxy	9.5	65.4	6.0	0.001	5.5	0.001	5.8	0.001	1.6
Ha-proxy backup	3.0	31.4	1.2	0.002	1.2	0.001	1.2	0.001	2.5
Router	4.7	51.3	2.8	0.776	2.8	0.001	1.8	0.001	2.6
Stat Server	7.3	62.0	4.0	8.977	3.2	0.001	3.4	0.001	2.1

Table 171 specifies 6.1 applications.

Table 171: 6.1 Applications

Application	All		Trace		Standard		None		All/None Rating
	CPU%	Log (MB)	CPU%	Log (MB)	CPU%	Log (MB)	CPU%	Log (MB)	
T-Server	26	200	7	0	7	0	7.5	0	3.5
T-Server backup	17.5	165	3.2	0	3.2	0	3.5	0	5.0
HA-proxy	13.5	61	5.2	0	5.2	0	3.5	0	3.9

Table 171: 6.1 Applications (Continued)

Application	All		Trace		Standard		None		All/None Rating
	CPU%	Log (MB)	CPU%	Log (MB)	CPU%	Log (MB)	CPU%	Log (MB)	
Ha-proxy backup	1.7	17	0.4	0	0.4	0	0.5	0	3.4
Router	17	120	11	1.9	11	0	3.3	0	5.2
Stat Server	9.2	80	6.2	0	6	0	3.7	0	2.5

Supplements

Related Documentation Resources

The following resources provide additional information that is relevant to Genesys software. Consult these additional resources as necessary.

Genesys

- *Genesys Glossary*, which is available on the Genesys documentation website, provides a comprehensive list of the Genesys and computer-telephony integration (CTI) terminology and acronyms used in this document.
- *Genesys Migration Guide*, which ships on the Genesys Documentation Library DVD, provides documented migration strategies for Genesys product releases. Contact Genesys Customer Care for more information.

Information about supported hardware and third-party software is available on the Genesys Customer Care website in the following documents:

- *Genesys Supported Operating Environment Reference Guide*
- *Genesys Support Media Interfaces Guide*

Consult the following additional resources as necessary:

- *Genesys Interoperability Guide,* which provides information on the compatibility of Genesys products with various Configuration Layer Environments; Interoperability of Reporting Templates and Solutions; and Gplus Adapters Interoperability.
- *Genesys Licensing Guide,* which introduces you to the concepts, terminology, and procedures that are relevant to the Genesys licensing system.
- *Genesys Database Sizing Estimator 8.x Worksheets,* which provides a range of expected database sizes for various Genesys products.

For additional system-wide planning tools and information, see the release-specific listings of System-Level Documents on the Genesys Documentation website (docs.genesys.com).

Genesys product documentation is available on the:

- Genesys Customer Care website at http://genesys.com/customer-care.
- Genesys Documentation site at http://docs.genesys.com/.
- Genesys Documentation Library DVD which you can order by e-mail from Genesys Order Management at orderman@genesys.com.

Document Conventions

This document uses certain stylistic and typographical conventions—introduced here—that serve as shorthands for particular kinds of information.

Document Version Number

A version number appears at the bottom of the inside front cover of this document. Version numbers change as new information is added to this document. Here is a sample version number:

80fr_ref_06-2008_v8.0.001.00

You will need this number when you are talking with Genesys Customer Care about this product.

Screen Captures Used in This Document

Screen captures from the product graphical user interface (GUI), as used in this document, may sometimes contain minor spelling, capitalization, or grammatical errors. The text accompanying and explaining the screen captures corrects such errors *except* when such a correction would prevent you from installing, configuring, or successfully using the product. For example, if the name of an option contains a usage error, the name would be presented exactly as it appears in the product GUI; the error would not be corrected in any accompanying text.

Type Styles

Table 172 describes and illustrates the type conventions that are used in this document.

Table 172: Type Styles

Type Style	Used For	Examples
Italic	• Document titles • Emphasis • Definitions of (or first references to) unfamiliar terms • Mathematical variables Also used to indicate placeholder text within code samples or commands, in the special case where angle brackets are a required part of the syntax (see the note about angle brackets on page 534).	Please consult the *Genesys Migration Guide* for more information. Do *not* use this value for this option. A *customary and usual* practice is one that is widely accepted and used within a particular industry or profession. The formula, $x + 1 = 7$ where x stands for . . .
Monospace font (Looks like teletype or typewriter text)	All programming identifiers and GUI elements. This convention includes: • The *names* of directories, files, folders, configuration objects, paths, scripts, dialog boxes, options, fields, text and list boxes, operational modes, all buttons (including radio buttons), check boxes, commands, tabs, CTI events, and error messages. • The values of options. • Logical arguments and command syntax. • Code samples. Also used for any text that users must manually enter during a configuration or installation procedure, or on a command line.	Select the Show variables on screen check box. In the Operand text box, enter your formula. Click OK to exit the Properties dialog box. T-Server distributes the error messages in EventError events. If you select true for the inbound-bsns-calls option, all established inbound calls on a local agent are considered business calls. Enter exit on the command line.
Square brackets ([ ])	A particular parameter or value that is optional within a logical argument, a command, or some programming syntax. That is, the presence of the parameter or value is not required to resolve the argument, command, or block of code. The user decides whether to include this optional information.	smcp_server -host [/flags]
Angle brackets (< >)	A placeholder for a value that the user must specify. This might be a DN or a port number specific to your enterprise. **Note:** In some cases, angle brackets are required characters in code syntax (for example, in XML schemas). In these cases, italic text is used for placeholder values.	smcp_server -host <confighost>

Index

Symbols

[] (square brackets) . . . 534
< > (angle brackets) . . . 534

A

angle brackets . . . 534

B

brackets
  angle . . . 534
  square . . . 534

C

chapter summaries . . . 12
commenting on this document . . . 18
conventions
  in document . . . 533
  type styles . . . 534

D

document
  conventions . . . 533
  errors, commenting on . . . 18
  version number . . . 533

F

font styles
  italic . . . 534
  monospace . . . 534

G

Genesys Agent Scripting . . . 31
Genesys Co-browse . . . 499
Genesys Voice Platform
  See GVP
GVP
  bandwidth usage
    Call Control Platform . . . 360
    Media Control Platform . . . 358
    Reporting Server . . . 360
  capacity
    VoIP . . . 245
  comparisons
    performance . . . 346
    scalability . . . 347
  disk space usage
    Reporting Server . . . 349
  formulas . . . 238
  hardware, recommended . . . 242
  high performance . . . 348
  metrics . . . 238
  operating systems, supported . . . 242
  performance . . . 293
  scalability . . . 293
  system capacity . . . 238
  terminology . . . 238
    busy hour call attempts (BHCA) . . . 238
    call arrivals per second (CAPS) . 238, 241, 239, 240, 238
    port density (PD) . . . 239
    response time latency (RTL) . . . 240
  test cases
    application . . . 293
    component . . . 319
    multiple instances (MCP) . . . 336
    single server . . . 334
    virtual machines . . . 336

H

hardware architecture diagrams 33
HP UX platform
  sample medium configuration 41
  sample small configuration 37

I

IBM/AIX platform
  sample medium configuration 40
  sample small configuration 36
Informiam/Performance Management Advisors . 417–458
intelligent Workload Distribution 487
Interaction Concentrator
  7.5 performance conclusions 104
  7.5 test results. 105
  7.6.0 performance conclusions 97
  7.6.0 test results. 89, 96
  7.6.1 performance conclusions 86
  7.6.1 test results. 76, 81
  8.0.0 performance conclusions 72
  8.0.0 test results. 67, 69, 71
  database configuration 63
  database optimization 63
  Database Size Estimator 63
  hardware architecture 58
  hardware considerations, ICON server . . .62
  hardware considerations, IDB 63
  performance considerations. 64
  recovery time 106
  sample database settings 99
  sample hardware specifications . 65, 74, 78, 87, 95, 98
  sample test environments 66, 70, 73, 77, 86, 94
italics 534

L

log database
  7.0 sizing examples 525

M

medium contact centers 23
  user interface hardware 32
monospace font 534

P

performance improvements
  configuration server 519
  logging. 519
    log length dependence 520
    solution performance test 520, 525
    test description 519, 520
  Routing 523
  Stat Server 522
Performance Management Advisors/Informiam . 417–430

S

sample medium configuration
  IBM/AIX platform 41
  Sun Solaris platform 39
  Windows platform. 38
sample small configuration
  IBM/AIX platform 36
  Sun Solaris platform 35
  Windows platform. 34
SIP Server
  estimating CPU usage 372, 393
  estimating network traffic 397
  estimating the number of instances. . 372, 394
  memory usage 378, 397
sizing recommendations. 12
  parameters 13
small contact centers 19
  Framework, Reporting, Routing, and OC . . 20
  user interface hardware. 22
square brackets 534
Sun Solaris platform
  sample medium configuration. 39
  sample small configuration 35

T

This 487
type styles
  conventions 534
  italic 534
  monospace 534
typographical styles 533, 534

V

version numbering, document 533

W

Windows platform
  sample medium configuration. 38
  sample small configuration 34