Monitoring Application Performance and Availability with Service Availability

r6.1
This documentation and any related computer software help programs (hereinafter referred to as the "Documentation") is for the end user's informational purposes only and is subject to change or withdrawal by CA at any time.

This Documentation may not be copied, transferred, reproduced, disclosed, modified or duplicated, in whole or in part, without the prior written consent of CA. This Documentation is confidential and proprietary information of CA and protected by the copyright laws of the United States and international treaties.

Notwithstanding the foregoing, licensed users may print a reasonable number of copies of the Documentation for their own internal use, and may make one copy of the related software as reasonably required for back-up and disaster recovery purposes, provided that all CA copyright notices and legends are affixed to each reproduced copy. Only authorized employees, consultants, or agents of the user who are bound by the provisions of the license for the product are permitted to have access to such copies.

The right to print copies of the Documentation and to make a copy of the related software is limited to the period during which the applicable license for the product remains in full force and effect. Should the license terminate for any reason, it shall be the user's responsibility to certify in writing to CA that all copies and partial copies of the Documentation have been returned to CA or destroyed.

EXCEPT AS OTHERWISE STATED IN THE APPLICABLE LICENSE AGREEMENT, TO THE EXTENT PERMITTED BY APPLICABLE LAW, CA PROVIDES THIS DOCUMENTATION "AS IS" WITHOUT WARRANTY OF ANY KIND, INCLUDING WITHOUT LIMITATION, ANY IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NONINFRINGEMENT. IN NO EVENT WILL CA BE LIABLE TO THE END USER OR ANY THIRD PARTY FOR ANY LOSS OR DAMAGE, DIRECT OR INDIRECT, FROM THE USE OF THIS DOCUMENTATION, INCLUDING WITHOUT LIMITATION, LOST PROFITS, BUSINESS INTERRUPTION, GOODWILL, OR LOST DATA, EVEN IF CA IS EXPRESSLY ADVISED OF SUCH LOSS OR DAMAGE.

The use of any product referenced in the Documentation is governed by the end user's applicable license agreement.

The manufacturer of this Documentation is CA.

Provided with "Restricted Rights." Use, duplication or disclosure by the United States Government is subject to the restrictions set forth in FAR Sections 12.212, 52.227-14, and 52.227-19(c)(1) - (2) and DFARS Section 252.227-7014(b)(3), as applicable, or their successors.

All trademarks, trade names, service marks, and logos referenced herein belong to their respective companies.

Copyright © 2008 CA. All rights reserved.
CA Product References

This document references the following CA product:

- CA eHealth® (eHealth)
- CA eHealth® for Voice
- CA AdvantEDGE View™ (AdvantEDGE View)
- CA eHealth® SystemEDGE™ (SystemEDGE)
- CA eHealth® Live Health® Application (Live Health)

Contact Technical Support

For online technical assistance and a complete list of locations, primary service hours, and telephone numbers, contact Technical Support at http://ca.com/support.
Contents

Chapter 1: Introducing Service Availability (SA)
  eHealth Service Availability Overview ................................................................. 7
  Service Availability Features .................................................................................. 8
    Active Testing of Application Services ................................................................. 8
    Visibility into the n-Tier Architecture ................................................................... 8
    Real-Time Notifications ....................................................................................... 9
    Performance Confirmation Against Service Level Agreements ....................... 10
    Historical Data Maintenance and Reports ......................................................... 10
      Standard Reports ............................................................................................... 10
      Historical Analysis Reports .............................................................................. 10

Chapter 2: Managing Service Availability
  Case Study: How to Manage the Availability of a Corporate Web Site ............. 11
    Step 1 – Place Test Systems in All Key Locations ............................................. 11
    Step 2 – Configure the Active Tests and Associate them with Agents .......... 12
    Step 3 – Run Historical Reports ....................................................................... 13
    Step 4 – Drill Down for Details ......................................................................... 13
    Step 5 – Use the BSC to Obtain an Overview of Your Services by Location ...... 15
  Test and Service Coverage .................................................................................... 16
  Application Response and WinTask ...................................................................... 17
Chapter 1: Introducing Service Availability (SA)

This chapter contains the following topics:
- eHealth Service Availability Overview
- Service Availability Features

**eHealth Service Availability Overview**

eHealth Service Availability provides information and real-time alarms on the availability and response times of critical business service applications. Through continuous active testing, it provides the information to monitor service levels, identify and report application outages, and document service performance.

To operate any business effectively, you must know whether an application is available and how much time that application takes to process user activity so that you can minimize the impact to users if any problems develop. Service Availability provides automated, real-time monitoring of availability and response time, immediate notification of delays and downtime, and (when used with eHealth) historical records of availability and response time.

The following graphic shows the deployment of a Service Availability test system.
Service Availability Features

With eHealth Service Availability, you can do the following:

- Actively test availability and response time for services 24x7.
- Gain visibility into the complex, n-tier infrastructure so that you can pinpoint problems and fix them before users are affected.
- Obtain real-time notification of delays, outages, or performance problems.
- Confirm that your services are performing well against service level agreements.
- Maintain historical data for capacity planning, troubleshooting, or analyzing trends in long-term behavior.

Active Testing of Application Services

An application service consists of an application, the systems on which it runs, and the network connections between the client systems and the servers. Service Availability considers the service to be available when all components are working and respond within acceptable service thresholds.

Service Availability provides automatic, 24x7, active tests to verify that all the components of a service are working. Active testing enables continuous coverage for your business applications, even during low activity times such as nights, holidays, off-peak hours, and weekends. When any of the active tests fail, or when transactions take too long, Service Availability notifies you of the problem so that you can take action to resolve the problem before your employees begin their workday or customers notice problems in service.

You can use Service Availability to run automatic, continuous tests, such as the following:

- Ping a server to ensure that it is running and reachable.
- Perform an HTTP test to verify that a Web server is responding.
- Run a custom script to verify that a specific feature is working.

When a test fails, Service Availability notifies you immediately that the service is unavailable. When tests succeed, Service Availability provides response time data to help you evaluate performance and watch for degradations in response time.

Visibility into the n-Tier Architecture

In a typical environment, applications are delivered through a complex array of load-balanced servers that exist in multiple tiers. For example, Web servers may make up the first tier, application servers make up the next, and database servers make up the next. Service Availability provides visibility into each of these tiers and into the services that provide communication between the tiers.
If you are using a complete eHealth solution, you can monitor your applications with Application Response and your systems on which the applications reside with the SystemEDGE agent. Application Response can alert you if the response time for a particular application begins to increase. It can also indicate whether the problem is with the network, client, or server. If Application Response indicates a server problem, you can use the SystemEDGE agent to monitor that server to determine whether issues exist with resources, such as amount of disk space or CPU, or performance.

Perhaps, however, the delay arises from a problem with one of the services that the application is using to communicate with the servers at other tiers of the infrastructure. Application Response and SystemEDGE cannot pinpoint such a problem, which may hinder your awareness of the problem and your ability to address it. Service Availability can pinpoint that type of problem because it provides visibility into the inter-tier communications. This feature offers a more complete and specific view of the problem so that you can address it before users are affected.

Real-Time Notifications

Because Service Availability is a module of the SystemEDGE agent, you can configure it to take advantage of SystemEDGE features. For example, you can configure Service Availability to perform any of the following actions:

- Send traps.
- Restart processes, service, or applications.
- Send e-mail, traps, or pager messages.
- Run a script or command.

Service Availability can perform these actions to notify you immediately when any of the following occur:

- A test fails (service outage).
- Response times exceed their thresholds.
- A test variable, such as failed attempts, reaches a new maximum value.
- Specified text, such as error or warning, appears in a log file.

For example, if response tests to your Web server fail, SystemEDGE can run an action script to page the Web server manager.
Performance Confirmation Against Service Level Agreements

When you have defined service profiles to address the requirements of your service level agreements, you can run Health and Service Level reports to monitor your performance against those profiles.

- **Health reports** analyze polled data against the thresholds, ranges, and percentiles specified in the service profile and identify situations that result from errors, unusual utilization rates, or excessive volume.
- **Service Availability Health reports** identify slow or unavailable services and applications that are hindering the productivity of users.
- **Service Level reports for Response Paths** provide summaries of the performance of your services by enterprise, region, department, or business unit.

Historical Data Maintenance and Reports

When you use Service Availability with eHealth, the response time and availability data that Service Availability collects is stored in the eHealth database. This feature lets you to run reports on data from the last 24 hours to the last year, depending on how long you are storing the data.

**Standard Reports**

eHealth provides the following standard historical reports:

- Trend reports
- At-a-Glance reports
- Top N reports
- What-If reports

**Historical Analysis Reports**

eHealth also provides two reports that perform historical analysis, comparing current performance with past performance:

- Health reports
- Service Level reports
Chapter 2: Managing Service Availability

This chapter contains the following topics:

- Case Study: How to Manage the Availability of a Corporate Web Site
- Test and Service Coverage
- Application Response and WinTask

Case Study: How to Manage the Availability of a Corporate Web Site

Many organizations provide customer information through a Web site or portal. This case study shows how you can use Service Availability to create active tests that continuously monitor the availability of a Web site and the average response time for the users who access it.

In this example, you have a Web server that resides in the Chicago headquarters. Members of the field sales regions in Phoenix, Tampa Bay, and Baltimore must access the server to record sales and obtain the latest product information.

Step 1 – Place Test Systems in All Key Locations

Install Service Availability on test systems in Phoenix, Tampa Bay, and Baltimore (the sales regions where your users are located) to establish locations from which you can duplicate user activity. The following diagram shows a Service Availability (SA) test system located near a series of user terminals within each sales region.
Step 2 – Configure the Active Tests and Associate them with Agents

Using the Service Availability page of the eHealth Web interface, create HTTP tests and ping tests to the Web server to verify that the Web server is accessible. Next, associate the HTTP and ping tests that you created with an agent on each of your test systems. After you configure test profiles and commit your changes, Service Availability runs the tests for the associated agents.

For example, create one test profile that associates the agent on your Baltimore system with the ping and HTTP tests you created. Then monitor the agent on that system to obtain a quick glimpse of the response and availability for all tests on that agent, as shown on the Test Monitor page below.
When you layer the tests by creating both HTTP and ping tests for the same system, you can obtain more insight into problems that may occur. For example, if the HTTP test is failing and the ping test is succeeding, you can deduce that the system is running, but the Web server application is down.

**Step 3 – Run Historical Reports**

After Service Availability has been running the tests for several intervals, run a Top N report to see the response times and availability of the Web site from each test system. Specify 100% as the availability goal value against which you want to evaluate the response times.

<table>
<thead>
<tr>
<th>Service Avail (%)</th>
<th>Avg Resp Time (msec)</th>
<th>Max Response (msec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>All Deviation</td>
<td>All</td>
</tr>
<tr>
<td>Element</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phoenix-Chicago-wr</td>
<td>100.00</td>
<td>82.68</td>
</tr>
<tr>
<td>Tampa-Chicago-wr</td>
<td>100.00</td>
<td>33.37</td>
</tr>
<tr>
<td>Baltimore-Chicago</td>
<td>100.00</td>
<td>32.90</td>
</tr>
</tbody>
</table>

This Top N report shows that Service Availability for all three paths was 100%, which indicates that the Web server was available and responding for the entire report period. The report also shows the average response times for each test and the deviation in response time from the goal of 100 milliseconds.

The response times show that the Baltimore and Tampa regions have almost identical average response times (32–33 msec). The Phoenix region, however, is more than twice as slow (83 msec). Phoenix is still operating below the 100 msec goal, so users may not notice any problems. It is still important to investigate the reason for the longer response time from Phoenix before the problem results in user complaints and interruptions to business.

**Step 4 – Drill Down for Details**

To obtain details about the response performance for the Phoenix region, drill down to an At-a-Glance report for that response path.

The Standard At-a-Glance report for a Service Availability Response Path includes the following charts:
- Average Response
- Minimum and Maximum Response
- Attempts
- Availability
- Total Bytes (Sent and Received)
- Total Throughput
- Total Errors
For this case study, the attempts and service availability were both 100% for the report period. Also, the total bytes sent and received and the throughput were consistent throughout the report period. The Average Response and Minimum and Maximum Response charts indicate when the response times were peaking and what levels they reached.

The Avg. Response chart from the At-a-Glance report for the Phoenix region shows that the average response spiked close to the limit of 1 second but did not exceed the limit during the report period. It also shows that the average response times increased most during the time from 8:00 AM to 12:00 PM, and then again from 4:00 PM to 8:00 PM.

The Minimum and Maximum Response chart shows a range of response times for individual tests. At times, the maximum response reaches 5 seconds, which is significantly higher than the 1-second response limit.

Longer response times can be a result of slower links or longer geographic distances from the test source system to the test destination. Response times also increase when problems begin (such as failed attempts to connect to the server), or when hardware problems cause interruptions or delays. Because the At-a-Glance report shows no problems with attempts or availability, the likely cause of the high response times is a slow interface for the Phoenix users. If these peak hours of usage are typical for the Phoenix users, consider increasing the speed of the links that they use to connect to the Web server.
Step 5 – Use the BSC to Obtain an Overview of Your Services by Location

With the Business Services Console (BSC), you can gain end-to-end visibility into the performance and availability of your business services. The BSC helps you determine where problems are occurring, whether someone is aware of any problems, and who is actively addressing them.

The following BSC diagram shows critical problems (red icons) for the Phoenix-Chicago test, which exceeded the maximum response time. (Yellow icons indicate major or minor problems, and green indicate that alarms or warnings have not occurred.) The check marks indicate that someone has acknowledged the alarms and is addressing them. The note provides additional information about the work that is being done to address the problem.

You can drill down from the BSC to Live Exceptions displays and eHealth reports. You can also associate groups of Service Availability tests with Live Exceptions profiles so that Live Exceptions can constantly monitor the tests and provide you with real-time observation of problems. When a problem occurs, the notifier can send an e-mail, run a command (such as a paging script to alert a user), or take another action that you specify.
Test and Service Coverage

Service Availability provides the following coverage for your business-critical services and applications:

- **Active Directory** – Verify that your Windows 2000 directory services are working properly to manage your shared files and resources.
- **Custom Tests** – Make sure that important custom services or other tasks are working efficiently.
- **Dynamic Host Configuration Protocol (DHCP)** – Confirm that your DHCP servers are responding to address requests.
- **Domain Name System (DNS)** – Verify that your DNS servers are processing hostname-to-address resolution requests.
- **File I/O** – Verify that operations such as read, write, and compare work across your file systems.
- **File Transfer Protocol (FTP) and TFTP** – Confirm that your users can log in to specified servers to upload and download files.
- **Hypertext Transfer Protocol (HTTP) and HTTP Secure (HTTPS)** – Verify that your users can connect to your business Web servers and determine whether specific text displays on a Web page.
- **Lightweight Directory Access Protocol (LDAP)** – Verify that you can connect to your LDAP servers and process user requests and LDAP queries.
- **Network Information System (NIS)/NIS+** – Confirm that NIS map requests are being processed.
- **Network News Transfer Protocol (NNTP)** – Make sure that your users can connect to their Usenet newsgroup servers and company bulletin boards.
- **Ping (ICMP echo)** – Make sure that your network devices exist and are reachable across the network.
- **E-mail services** – Confirm that the e-mail servers are available and are processing e-mail effectively for the following services: Internet Message Access Protocol (IMAP), Messaging Application Program Interface (MAPI), Post Office Protocol 3 (POP3), Simple Mail Transfer Protocol (SMTP), and round-trip e-mail that originates from an SMTP server and retrieves messages from IMAP, MAPI, or POP3 accounts.
- **Simple Network Management Protocol (SNMP)** – Confirm that SNMP agents are responding to SNMPv1 GET requests.
- **SQL Query** – Confirm that SQL database servers are available and processing short queries that you specify.
- **Transmission Control Protocol (TCP) Connect** – Verify that your systems are listening for and processing connection requests.
- **Virtual User** – Obtain continuous response time and availability data for actual user transactions (keyboard entry and mouse clicks) that you have recorded (typically with WinTask) to confirm that business tasks run successfully.
Application Response and WinTask

Using Application Response and WinTask with Service Availability provides additional application management and response time information about user activity as it happens.

*Application Response* monitors the response time and availability of critical applications as they are experienced by end users. It monitors only real user actions, in contrast to Service Availability, which runs tests continually and independently of user actions. Application Response provides response data for client, network, and server time to help you pinpoint the source of any problems or delays.

*WinTask* provides the bridge between Service Availability and Application Response to help you obtain 24x7 end-user data without requiring your users to perform transactions 24x7. With WinTask, you can record real user transactions and then schedule them to run continuously.

With these three products, you can access continuous, consistent data for all the user transactions and services that you care about from the systems that you are monitoring without waiting for actual user activity. This continuous data makes your eHealth reports more useful for capacity planning and longterm historical monitoring of the applications.