

Configuring the HP ProLiant Server BIOS for Low-Latency Applications White Paper

Abstract

This document is for the person who installs, administers, and troubleshoots servers and storage systems. HP assumes you are qualified in the servicing of computer equipment and trained in recognizing hazards in products with hazardous energy levels.



Part Number: 581608-003
August 2011
Edition: 3

© Copyright 2011 Hewlett-Packard Development Company, L.P.

The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

AMD and Opteron are trademarks of Advanced Micro Devices, Inc.

Intel and Xeon are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Windows Server is a U.S. registered trademark of Microsoft Corporation.

Contents

Introduction	4
Operating systems	5
Red Hat Enterprise MRG	5
SUSE Linux Enterprise Real Time Extension	5
HP support for real-time operating systems	5
Configuring BIOS	6
System requirements	6
BIOS tuning recommendations for optimum performance	6
BIOS low-latency settings	7
Upgrading BIOS	7
Obtaining SSSTK	8
BIOS configuration methods	8
Using RBSU to reconfigure BIOS	9
Using conrep to reconfigure BIOS	9
Determining PowerManagement state	9
Restoring management functionality	10
Restoring management functionality using RBSU	10
Restoring management functionality using conrep	10
Frequently asked questions	11
Technical support	13
Before you contact HP	13
HP contact information	13
Acronyms and abbreviations	14

Introduction

Low-latency, deterministic system performance is a required system characteristic in the financial services market, where it enables high frequency trading, market data distribution, and exchange data processing. It is also required in defense and data acquisition applications for real-time signal and image processing.

These systems must respond rapidly to external events in a predictable manner. They must do so under heavy workloads, sometimes reaching millions of transactions per second. To achieve this level of performance, system designers must consider the following factors during system design and configuration:

- Hardware — System design, processor type and speed, memory capacity, network components
- OS selection — Real-time operating system kernels specifically designed and tuned for minimum latency and real-time preemption
- BIOS configuration — BIOS support configured for minimum latency and maximum performance
- Networking fabric — Network technology (1 and 10 Gigabit Ethernet, InfiniBand, Fibre Channel)
- Middleware — Messaging and database services on the network designed for minimum latency and maximum reliability
- End-user applications — Designed to perform multicast messaging and RDMA
- Physical distances — Physical separation between the information sources and clients affects overall system performance.

This document presents suggestions and best practice recommendations on BIOS configuration, and on OS selection and tuning to obtain the best performance from HP ProLiant BL c-Class server blades and HP ProLiant DL, ML, and SL servers in the financial service market.



IMPORTANT: The configuration changes described in this document apply only to systems with a low-latency OS kernel installed. Overall system throughput is not affected, whether or not the system management features are enabled.

Operating systems

The choice of operating system is a key decision in low-latency applications. General purpose operating systems such as SUSE Linux Enterprise Server, Red Hat Enterprise Linux, and Microsoft® Windows Server® are designed so that some operating system processes cannot be interrupted once started. Operating systems that are designed for real-time, low-latency applications are designed to permit some operating system processes to be interrupted by the critical application. This guarantees that the critical application has the resources necessary without interruption from the operating system.

Red Hat Enterprise MRG

Red Hat Enterprise MRG is a next-generation IT infrastructure incorporating Messaging, Realtime, and Grid functionality. MRG's realtime kernel provides the highest levels of predictability for consistent low-latency response times to meet the needs of time-sensitive workloads. For more information about Red Hat Enterprise MRG, see the Red Hat MRG website (<http://www.redhat.com/mrg/realtime/>).

SUSE Linux Enterprise Real Time Extension

SLERT is an enhanced version of SUSE Linux Enterprise Server that provides reduced latency and predictable performance in critical, time-sensitive applications. For more information about SLERT, see the SLERT website (<http://www.novell.com/products/realtime/>).

HP support for real-time operating systems

Tuning the operating system is critical and is dependent on application characteristics. HP works with real-time operating system vendors to ensure compatibility with HP servers and server blades. For more information, see the HP Realtime Linux website (<http://www.hp.com/go/realtimelinux/>).

Configuring BIOS

System requirements

The configuration options described in this document are based on recent versions of HP BIOS that enable customers to disable the generation of periodic System Management Interrupts used for power and CPU monitoring, with their attendant latency impact.

A properly tuned low-latency operating system is also required to achieve deterministic performance. No single "recipe" can be documented. Customers needing a low-latency environment often perform exhaustive testing of the latency impact of various tuning parameters with their application and systems to determine the optimum settings for their environment.

BIOS tuning recommendations for optimum performance

HP servers are configured by default to provide the best balance between performance and power consumption. These default settings may not provide the lowest latency. The first step in tuning for low latency is to examine these additional settings that may assist in obtaining optimal low-latency performance. These settings are accessible through RBSU and with the `conrep` script, a configuration tool provided by HP.

Consider the following options as part of any deployment in low-latency OS kernel environments:

- Set the HP ProLiant Power Regulator Mode to Static High Mode.
- Disable Processor C-State Support. Options depend on the HP ProLiant server model and generation. C-State Support options include the following:
 - C1E Support
 - AMD C1 Halt Support
 - Any other C-State modes, such as C3 or C6, that are enabled by the BIOS
- On HP ProLiant G6 servers utilizing Intel® Xeon® processors, disable QPI Power Management.
- Disable Intel® Turbo Boost on HP ProLiant G6 servers and server blades based on Intel® Xeon® processors.

Starting with HP ProLiant G6 servers that utilize Intel® Xeon® processors, setting the HP Power Profile Option in RBSU to Maximum Performance Mode sets these recommended additional low-latency options for minimum BIOS latency.

All HP ProLiant G6 and later servers, regardless of the ROM version, support setting Intel® Turbo Boost and C-States. HP ProLiant 100 Series and HP ProLiant SL servers do not support advanced features for iLO Performance Monitoring and Memory Pre-Failure notification.

BIOS low-latency settings

To configure BIOS for minimum latency, you can disable Processor Power and Utilization Monitoring, Memory Pre-Failure Notification, or both. Disabling each option causes some server features to become unavailable. Before reconfiguring BIOS, be sure that none of the features described below are required.

Disabling Processor Power and Utilization Monitoring disables the following features:

- iLO 2 Processor State Monitoring
- Insight Power Manager CPU Utilization Reporting
- HP Dynamic Power-Savings Mode

Disabling Memory Pre-Failure Notification has the following effects:

- Disables Memory Pre-Failure Warranty Support
- Disables notification when correctable memory errors occur above a pre-defined threshold
- Forces the system to run in Advanced ECC Mode, regardless of the mode configured in RBSU



IMPORTANT: Online Spare Mode, Mirroring Mode, and Lock-step Mode are not supported when Memory Pre-Failure Notification support is disabled. Supported AMP modes depend on the generation and model of the ProLiant server.

Disabling Processor Power and Utilization Monitoring provides the greatest benefit in low-latency environments. Disabling Memory Pre-Failure Notification has a much smaller effect because this feature generates an interrupt at a very low frequency — from once every several minutes to as seldom as once an hour on HP ProLiant G6 servers and server blades with Intel® Xeon® processors. Because of the very low frequency of this SMI, only the most latency-sensitive environments should have Memory Pre-Failure Notification disabled.

Some low-latency kernels provide features that result in system latency greater than that caused by Memory Pre-Failure Notification. For example, Red Hat Enterprise Linux 5 kernel soft lockup detection may generate an interrupt that results in an approximately 1-second latency each time it occurs. This latency is greater than that caused by Memory Pre-Failure Notification.

In any case, disabling Memory Pre-Failure Notification does not disable the Advanced ECC mode or correction of errors. Uncorrectable errors are still flagged, logged, and bring the system down. The only difference when this SMI is disabled is that there is no early notification if the uncorrectable error threshold is exceeded.

Upgrading BIOS

Before attempting to disable Processor Power and Utilization Monitoring or Memory Pre-Failure notification, upgrade the BIOS to the most recent version.



IMPORTANT: Before upgrading the BIOS in HP ProLiant G1 through G5 server blades with InfiniBand mezzanine cards installed, upgrade the InfiniBand mezzanine card firmware.

To obtain the mezzanine card firmware:

1. Search the HP website (<http://www.hp.com/go/bladesystem>) for the mezzanine card part number.
2. Select the installed OS.
3. Download the upgrade.

4. Install the upgrade.

To obtain the most recent BIOS upgrade for HP ProLiant servers:

1. Go to the HP website (<http://www.hp.com/go/support>).
2. Select **Download drivers and software (and firmware)**.
3. Enter the server model number and then click >>.
4. Select the appropriate product link.
5. Select an operating system.
6. Select the **BIOS - System ROM** category.
7. To obtain the BIOS upgrade, do one of the following:
 - o Download the latest ROMPaq firmware, and then upgrade the firmware using the instructions included with the ROMPaq.
 - o Select **Online ROM Flash Component**, click the **Installation Instructions** tab, and then follow the instructions on the Online ROM Flash Component page.

Obtaining SSSTK



IMPORTANT: SSSTK v.2.20c or later is required to reconfigure BIOS. The `conrep` utility provided in earlier versions does not provide the required functionality.

The `conrep` utility, a part of SSSTK, can be used to configure Processor Power and Utilization Monitoring or Memory Pre-Failure Notification for minimum latency. This is the only method available for configuring these options on HP ProLiant G5 servers and HP ProLiant G6 servers that utilize AMD Opteron™ processors. The utility is one method available for configuring HP ProLiant G6 servers that utilize Intel® Xeon® processors.

To install the SSSTK:

1. Go to the HP website (<http://www.hp.com/go/support>).
2. Enter SmartStart Scripting Toolkit Linux Edition in the **Search:** field, and then click >>.
3. Download the latest SmartStart Scripting Toolkit Linux Edition.
4. Create a new directory.
5. Unpack the archive in the new directory.

BIOS configuration methods

The appropriate tools to reconfigure the BIOS for low latency in HP ProLiant 300 Series or above G5, G6, or G7 servers and server blades depend on the server generation and on the processor. Use the following table to determine the appropriate method.

Generation	AMD™ processors	Intel® processors
G5	<code>conrep</code>	<code>conrep</code>
G6	<code>conrep</code>	<code>conrep</code> RBSU
G7	<code>conrep</code> RBSU	<code>conrep</code> RBSU

SSSTK utilities, including `conrep`, are designed for mass deployment scenarios. A combination of the utilities may be used to configure RBSU, Smart Array, and iLO settings, and then install the operating system.

Where the option to use either RBSU or `conrep` exists, RBSU may be quicker than setting up the toolkit if only a few servers are being configured.

When deploying servers in a low-latency environment, an additional benefit of using `conrep` is that it can be used during periodic maintenance to verify that the RBSU settings are maintained in their proper state.

The best practice for getting the data file to be used with `conrep` is to boot a system into RBSU and make all the changes, then boot the system with the toolkit and capture RBSU settings with the capture script.

Using RBSU to reconfigure BIOS

To configure BIOS low-latency options using RBSU:

1. Press **F9** during POST to enter RBSU.
2. Press **CTRL-A** to open the menu.
3. Select **Service Options**.
4. Disable either or both of the following options:
 - o Processor Power and Utilization Monitoring
 - o Memory Pre-Failure Notification

Using `conrep` to reconfigure BIOS



IMPORTANT: SSSTK v2.20c or later is required to reconfigure BIOS. The `conrep` utility provided in earlier versions does not provide the required functionality.

To configure BIOS low-latency options using the `conrep` utility in SSSTK:

1. Change the current directory to the SSSTK/utilities directory:
`cd SSSTK/utilities`
2. To disable Processor Power and Utilization Monitoring, verify that the `conrep.dat` file contains the following markup:
`<Conrep>`
`<PowerMonitoring>0x10</PowerMonitoring>`
`</Conrep>`
3. To disable Memory Pre-Failure Notification, verify that the `conrep.dat` file contains the following markup:
`<Conrep>`
`<DisableMemoryPrefailureNotification>1</DisableMemoryPrefailureNotificat`
`ion>`
`</Conrep>`
4. Enter the following commands:
`./conrep -l -fconrep.dat`
`reboot`

Determining PowerManagement state

To use the System Maintenance Utility to determine the state of the Power Monitoring bit:

1. Press **F10** when **Press F10 for System Maintenance Utility** appears.
2. Select **Inspect Utility**.
3. Select **System EV Data**.

4. Scroll down to **CQHGV3**.

CQHGV3 values have the following meanings:

- 0x04 — Processor Power and Utilization Monitoring SMI ENABLED, Memory Pre-Failure Notification SMI ENABLED
- 0x14 — Processor Power and Utilization Monitoring SMI DISABLED, Memory Pre-Failure Notification SMI ENABLED
- 0x34 — Processor Power and Utilization Monitoring SMI DISABLED, Memory Pre-Failure Notification SMI DISABLED

Restoring management functionality

To restore BIOS management functionality, see "BIOS configuration methods (on page 8)," and then select one of the following methods.

Restoring management functionality using RBSU

To restore Processor Power and Utilization Monitoring and Memory Pre-Failure Notification using RBSU:

1. Press **F9** during POST to enter RBSU.
2. Press **CTRL-A** to open the menu.
3. Select **Service Options**.
4. Enable either or both of the following options:
 - Processor Power and Utilization Monitoring
 - Memory Pre-Failure Notification

Restoring management functionality using conrep

To restore management functionality using the `conrep` utility in SSSTK, complete the following steps:

1. Change the current directory to the SSSTK/utilities:
`cd SSSTK/utilities`
2. To enable management functionality, verify that the `conrep.dat` file contains the following markup:
`<Conrep>`
`<PowerMonitoring>0x00</PowerMonitoring>`
`</Conrep>`
3. To enable Memory Pre-Failure Notification, verify that the `conrep.dat` file contains the following markup:
`<Conrep>`
`<DisableMemoryPrefailureNotification>0</DisableMemoryPrefailureNotificat`
`ion>`
`</Conrep>`
4. Enter the following commands:
`./conrep -l -fconrep.dat`
`reboot`

Frequently asked questions

Q. Does disabling Memory Pre-Failure Notification disable memory error correction?

A. Memory errors are still corrected, but notification that the error rate has exceeded a pre-set threshold is disabled. The latency impact of this feature is very small. HP recommends disabling Memory Pre-Failure Notification only if absolutely necessary.

Q. What memory features are lost if Memory Pre-Failure Notification is disabled?

A. If Memory Pre-Failure Notification is disabled, Online Spare and Mirroring memory modes become unavailable. The system is forced to run in Advanced ECC mode, regardless of the mode set in BIOS. Memory Pre-Failure Warranty Support also becomes unavailable because there is no notification of errors exceeding the programmed threshold.

Q. How does disabling iLO 2 Processor State Monitoring in the HP ProLiant c-Class enclosure affect power management?

A. Disabling state monitoring does not affect power management.

Q. How can I verify that a server has the low-latency option set?

A. There are three methods to verify that the low-latency option is set:

- Use RBSU:
 - a. Boot the system and press **F10**, and then select **Inspect Utility**.
 - b. Select **System EV Data**.
 - c. Select **CQHGV3**. The value of CQHGV3 should be 0x10 (Bit[4] is set).
- Use a special configuration file, available on request from Lee Fisher (lee.fisher@hp.com), that displays these hidden low-latency settings.
- Write and run a test script to see if you are getting spikes. For more information, contact Lee Fisher (lee.fisher@hp.com).

Q. Does Microsoft® Windows Server® 2008 support real-time operations?

A. Microsoft® Windows Server® 2008 is not available with a real-time extension.

Q. Do HP BIOS low-latency options work in Microsoft® Windows®?

A. HP BIOS low-latency options work in a Microsoft® Windows Server® 2008 environment. Because Microsoft® Windows Server® 2008 is not available with a real-time extension, only latencies associated with the BIOS SMIs will be affected.

To apply the low-latency options in a Microsoft® Windows® environment:

1. Obtain the SmartStart Scripting Toolkit ("[Obtaining SSSTK](#)" on page 8).

2. Run the SmartComponent for the most recent version of the SSSTK, note the directory that it is in, and then change to it in Windows® Explorer or a command window.
3. Build a Windows® PE boot image. For more information, see "Preparing the bootable media" in the *HP SmartStart Scripting Toolkit Windows Edition User Guide*.
4. Boot the image, and from Windows® PE, build an input CONREP.DAT.
5. Run `conrep` ("Using `conrep` to reconfigure BIOS" on page 9).

Q. Can I interrogate the memory operating speed?

A. Requested memory operating speed can be interrogated using RBSU or `conrep`. The information returned indicates the requested or programmed operating speed.

Q. How can the actual memory operating speed be confirmed?

A. Memory operating speed can be confirmed using memory benchmarking tests. On HP ProLiant servers utilizing Intel® Xeon® processors, RBSU can configure the memory operating speed.

Q. Why don't all the server blades boot when power is first applied to a fully populated enclosure?

A. The OA manages power based on the available power supply capacity in the enclosure and the power requested by each device that is installed in a bay. Before a server blade is installed for the first time, it is programmed to request the maximum possible amount of power that it is capable of using when fully configured with all possible options. During this first boot, the OA requests a stress test on the server blade and monitors the actual peak power used. This value is then retained for subsequent startups.

If all server blades in the enclosure request maximum power simultaneously, the OA may determine that the capacity of the power supplies could be exceeded and may deny power to one or more server blades, even though the power available is adequate to power all of the server blades under normal load. Sequentially powering on new server blades may address the issue.

OA power management is a complex subject and well beyond the scope of this document. The preceding discussion is greatly simplified. For more information about the Onboard Administrator and HP ProLiant BladeSystem enclosures, see the HP website (<http://www.hp.com/go/bladesystem>).

Q. Does setting the server blades to highest performance state affect the ability of all server blades in an enclosure to boot?

A. No, because the server blades always start in maximum performance state. The pre-set power request on first boot is based on the maximum performance state with all options installed.

Technical support

Before you contact HP

Be sure to have the following information available before you call HP:

- Technical support registration number (if applicable)
- Product serial number
- Product model name and number
- Product identification number
- Applicable error messages
- Add-on boards or hardware
- Third-party hardware or software
- Operating system type and revision level

HP contact information

For United States and worldwide contact information, see the Contact HP website (<http://www.hp.com/go/assistance>).

In the United States:

- To contact HP by phone, call 1-800-334-5144. For continuous quality improvement, calls may be recorded or monitored.
- If you have purchased a Care Pack (service upgrade), see the Support & Drivers website (<http://www8.hp.com/us/en/support-drivers.html>). If the problem cannot be resolved at the website, call 1-800-633-3600. For more information about Care Packs, see the HP website (<http://pro-aq-sama.houston.hp.com/services/cache/10950-0-0-225-121.html>).

Acronyms and abbreviations

iLO 2

Integrated Lights-Out 2

MRG

Red Hat Enterprise MRG

OA

Onboard Administrator

POST

Power-On Self Test

RBSU

ROM-Based Setup Utility

RDMA

Remote Direct Memory Access

SLERT

SUSE Linux Enterprise Real Time Extension

SMI

System Management Interrupt

SSSTK

SmartStart Scripting Tool Kit