Configuring and tuning HPE ProLiant Servers for low-latency applications
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>3</td>
</tr>
<tr>
<td>What’s new</td>
<td>3</td>
</tr>
<tr>
<td>Recommended hardware configurations</td>
<td>4</td>
</tr>
<tr>
<td>Preparing for low-latency configuration</td>
<td>6</td>
</tr>
<tr>
<td>Taking inventories or snapshots</td>
<td>6</td>
</tr>
<tr>
<td>Upgrading BIOS and firmware</td>
<td>6</td>
</tr>
<tr>
<td>Obtaining the Scripting Utilities</td>
<td>8</td>
</tr>
<tr>
<td>Recommended platform tuning</td>
<td>9</td>
</tr>
<tr>
<td>System requirements</td>
<td>9</td>
</tr>
<tr>
<td>Tuning recommendations and explanations</td>
<td>9</td>
</tr>
<tr>
<td>Tuning with HPE RESTFUL Interface Tool (Gen9)</td>
<td>15</td>
</tr>
<tr>
<td>Recommended operating system tuning</td>
<td>19</td>
</tr>
<tr>
<td>Linux</td>
<td>19</td>
</tr>
<tr>
<td>Windows</td>
<td>22</td>
</tr>
<tr>
<td>HPE-TimeTest</td>
<td>23</td>
</tr>
<tr>
<td>Frequently asked questions</td>
<td>24</td>
</tr>
<tr>
<td>Support and other resources</td>
<td>25</td>
</tr>
<tr>
<td>Resources and documentation</td>
<td>25</td>
</tr>
<tr>
<td>Before you contact HPE</td>
<td>25</td>
</tr>
<tr>
<td>HPE contact information</td>
<td>25</td>
</tr>
<tr>
<td>Acronyms and abbreviations</td>
<td>27</td>
</tr>
</tbody>
</table>
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 other industries such as 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 latency, speed, and capacity; network components; and storage subsystem, including SSDs
- **OS selection**—Operating system kernels specifically designed and tuned for minimum latency and, in some cases, real-time preemption
- **BIOS configuration**—BIOS support configured for minimum latency and maximum performance
- **Networking fabric**—Network technology (1/10/40 Gigabit Ethernet, InfiniBand, Fibre Channel)
- **Middleware**—Messaging and database services on the network designed for minimum latency and maximum throughput with reliability
- **End-user applications**—Designed to perform multicast messaging accelerated via kernel bypass and RDMA techniques
- **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 tuning to obtain the lowest-latency performance from HPE ProLiant BL c-Class server blades and HPE ProLiant DL, ML, SL, and XL servers. While this document contains information pertaining to G7 and earlier ProLiant servers, the primary focus is Gen8 servers and later.

The recommendations to disable System Management Interrupts (SMIs) are intended only for extreme latency-sensitive use cases. Most customers benefit from the power savings, monitoring, and notifications that the SMIs enable. These SMIs consume less than 0.1% of the server's processing capability, and HPE continues to reduce their impact with each new generation of ProLiant servers.

---

**Important**

The information in this document is accurate as of the document's release date but is subject to change based on updates made by HPE.

---

**What's new**

The current edition of the Configuring and Tuning HPE ProLiant Servers for Low-Latency Applications White Paper, 581608-009, includes the following additions and updates:

- **“Recommended hardware configurations” on page 4**
  - Updated information to new E5-2600 v4 versions of processors
  - Updated recommended memory speed to 2400 MHz

- **Updated the following tuning procedures:**
  - “Tuning with HPE RESTful Interface Tool (Gen9)” on page 15
  - “Tuning with conrep (Gen8 and older)” on page 16
  - “Recommended operating system tuning” on page 19
Recommended hardware configurations

HPE recommends the following HPE ProLiant Gen9 hardware configuration when low-latency is required. This information is subject to change and is valid as of the date of publication. For the latest information, see the server QuickSpecs on the HPE website (https://www.hpe.com/us/en/support.html).

- The HPE Trade and Match Server is optimized for applications that perform better at high frequency with lower core count. The base server configuration contains:
  - HPE Apollo 2000 Gen9 server
  - 4 HPE ProLiant XL170r Gen9 Server nodes powered by E5-1680 v3 processors
  - 128 GB DDR4-2133MT/s memory
  - Custom BIOS and configuration tools to enable overclocking processors
- Processor
  - E5-2637 v4 (4c 3.5 GHz), E5-2643 v4 (6c 3.4 GHz), E5-2667 v4 (8c 3.2 GHz) and E5-2697A v4 (16c 2.6 GHz) in HPE ProLiant DL, ML, and BL servers, and select Apollo System servers
  - E5-2690 v4 (14c 2.6 GHz) in select Apollo System servers that do not support the higher-wattage processors
- Memory
  - 8 GB Single Rank DDR4-2400MT/s CAS-17 RDIMMs
  - If installing only one DIMM per channel, consider using 8 Dual-Rank 2400MT/s 16 GB RDIMMs for improved memory interleaving.
  - Each channel should be populated with at least one DIMM.
- PCIe Gen3 architecture
  - The HPE ProLiant DL380 Gen9 Server offers three x8 slots that communicate with processor 1 and three x8 or higher slots that communicate with processor 2. Two additional option slots communicate with processor 1, one x8 FlexibleLOM slot for network options and an x8 Flexible SA slot for storage controller options.
  - The HPE ProLiant DL360 Gen9 Server offers two x8 or higher slots that communicate with processor 1 and one x16 slot that communicates with processor 2. Two additional option slots communicate with processor 1, one x8 FlexibleLOM slot for network options and an x8 Flexible SA slot for storage controller options.

Consider a single processor configuration if your workload does not benefit from a second processor. The benefits of using a single processor are as follows:

- Yields automatic PCI-to-core affinity (no application rewrite).
- DDIO performs optimally.
- Cache snooping is eliminated.
- No QPI latency is experienced.
- Simplified CPU core mapping is achieved.
- Even with one processor, there are still three x8 PCIe slots for NICs, timing cards, Fusion-io, and so forth.

- The HPE ProLiant BL460c Gen9 Server Blade has one x16 mezzanine slot that communicates with processor 1 and one x16 mezzanine that communicates with processor 2, plus a FlexibleLOM off processor 1.
- The HPE ProLiant XL230a Gen9 offers an optional x16 Riser Kit (Part #788126-B21) that provides a low-profile PCIe slot with direct connectivity to the second processor. Additionally, an optional HPE Apollo 6000 Dual FlexibleLOM Riser kit (Part # 757401-B21) is available to provide a second x8 FlexibleLOM slot.
- PCIe NIC
  - Mellanox ConnectX-4 Lx-based adapters offer ultra-low latency and are designed specifically for HPE servers in two form factors: PCIe card (640SFP28 and FlexibleLOM (640FLR-SFP28). They are sold, integrated, and directly supported by HPE. The Mellanox ConnectX-4 Fx NIC
offers native Gen3 x8 performance (25GbE). Mellanox Connect-IB based adapters offer even greater throughput by offering native Gen3 x16 performance.

- Additional popular third-party PCIe Ethernet cards for ultra-low latency are available from Solarflare, Myricom, Chelsio, and Exablate and can be installed in HPE industry-standard ProLiant DL, ML, and SL servers.

- Storage
  - New HPE Smart Array P44x and P840 storage controllers offer 12 Gb/s SAS performance when used in Gen9 servers with 12 Gb/s devices (SSDs or HDDs), which can deliver as much as 60% more IOPS vs. 6Gb/s devices.
  - The B140i is the base storage controller on the HPE ProLiant DL360 Gen9 and HPE ProLiant DL380 Gen9. It utilizes a special OS driver to provide RAID functionality to provide a low cost solution. However, the driver-based RAID solution may introduce slight OS overhead which could affect latency. For latency-sensitive applications, it is recommended that the B140i be run in AHCI mode and for RAID-based configurations a hardware-based RAID controller such as the HPE Smart Array P44x be used.
  - HPE I/O Accelerator now supports up to 16 TB MLC in server blade mezzanine cards. For more information, see the HPE IO Accelerator for HPE BladeSystem c-Class QuickSpecs on the HPE website (http://h18004.www1.hpe.com/products/quickspecs/13220_div/13220_div.pdf).
Preparing for low-latency configuration

Taking inventories or snapshots
Before you configure servers for low-latency applications, HPE recommends that you take an inventory or snapshot of the following items. This will enable you to track changes during the optimization process.

- dmidecode
  For RHEL before 6.2 and SLES before 12, obtain v. 2.11 from the nongnu website (http://www.nongnu.org/dmidecode).
- lspci -vv
- conrep (for ProLiant Gen9 and earlier servers)
- hpdiscovery
To obtain the latest versions of conrep or hpdiscovery, see “Obtaining the Scripting Utilities” on page 8.

- sysctl -a
- HP-timeTest7.3
- HPE-TimeTest is a utility distributed by HPE that enables customers to test for jitter in a server. To obtain the HPE-TimeTest utility, contact HPE by emailing to: (lowlatency@hpe.com). Please include the name of your HPE contact.

- cat /proc/cmdline

Upgrading BIOS and firmware
Before making changes for low-latency operation, be sure that all platform firmware is up-to-date. For low latency, it is especially important to upgrade the BIOS, iLO 4 and network card firmware to the latest versions. HPE offers the Service Pack for ProLiant (SPP) as a comprehensive solution to maintain all of the firmware and software for a server in a single ISO image. Use of the HPE SPP is the recommended method of upgrading the platform firmware. Refer to the “HPE Service Pack for ProLiant Contents Report” for the selected HPE SPP on www.hpe.com to verify latest versions of BIOS, iLO 4, and network card firmware are included.

Important
The HPE SPP and BIOS images both require an active warranty or support agreement covering HPE ProLiant servers to be linked to the HPE Support Center profile used to download the components. Please refer to:

To obtain the most recent HPE Service Pack for ProLiant (HPE SPP) upgrade for HPE ProLiant servers:
2. Select HPE Servers, Storage, and Networking
3. Enter the server model number, and then click Search.
4. Select Get drivers, software, & firmware.
5. Select the appropriate product link.
6. Select your operating system.
7. Select the Application - System Management category.
8. To obtain the HPE SPP upgrade, do the following:
   o Download the latest HPE SPP ISO image, and then upgrade the firmware using the instructions included with the ISO. The HPE SPP can be used in online mode from either a Windows- or Linux-hosted operating system, or in offline mode by booting to the ISO.

To obtain the most recent BIOS upgrade for HPE ProLiant servers, if not in HPE SPP:
2. Select HPE Servers, Storage and Networking.
3. Enter the server model number, and then click Search.
4. Select Get drivers, software, & firmware.
5. Select the appropriate product link.
6. Select your operating system.
7. Select the BIOS - System ROM category.
8. To obtain the BIOS upgrade, do one of the following:
   - Download the latest ROMPaq firmware, and then upgrade the firmware using the instructions included with the ROMPaq.
   - Select Online ROM Flash Component, click the Installation Instructions tab, and then follow the instructions on the Online ROM Flash Component page.

To obtain the latest network card firmware, if not in HPE SPP:
2. Select HPE Servers, Storage, and Networking.
3. Enter the server model number, and then click Search.
4. Select Get drivers, software, & firmware.
5. Select the appropriate product link.
6. Select your operating system.
7. Select Firmware - Network.
8. Download the appropriate NIC firmware.

**Important**
Version 1.40 of the iLO 4 firmware has been found to experience an increased number of periodic SMIs that cause some latency jitter on Gen8 systems. It is therefore strongly recommended to update the iLO 4 firmware to the v150 or greater release to address the problem.

Version 2.20 of the iLO 4 firmware has been found to experience periodic SMIs every 15 seconds that cause some latency jitter. It is therefore recommended to use the iLO 4 v230 or later release of the firmware to address the problem.

To obtain the latest iLO4 firmware, if not in HPE SPP:
2. Select HPE Servers, Storage, and Networking.
3. Enter the server model number, and then click Search.
4. Select Get drivers, software, & firmware.
5. Select the appropriate product link.
6. Select your operating system.
8. Click Obtain software, and then click the executable file to download it.
Obtaining the Scripting Utilities

For Gen9 servers:

The HPE RESTful Interface Tool can be used to configure BIOS options on Gen9 servers and is the preferred tool for future systems.

To install the HPE RESTful Interface Tool:

2. Select HPE Servers, Storage, and Networking.
3. Enter the server model number, and then click Search.
4. Select Get drivers, software, & firmware.
5. Select the appropriate product link.
6. Select your operating system.
7. Select Utility - Tools.
8. Click Download, next to the appropriate executable file to save it.

For Gen8 and older servers:

The conrep utility can be used to configure Processor Power and Utilization Monitoring or Memory Pre-Failure Notification for minimum latency, and is included in Smart Start Scripting Toolkit (SSSTK) 9.10 or later. For Gen8 servers, SSSTK is now called STK.

conrep is the only method available for configuring these options on HPE ProLiant G5 servers and HPE ProLiant G6 servers that utilize AMD Opteron processors. The utility is one method available for configuring HPE ProLiant G6, G7, and Gen8 servers that utilize Intel Xeon processors.

To install the STK:

2. Select HPE Servers, Storage, and Networking.
3. Enter the server model number, and then click Search.
4. Select Get drivers, software, & firmware.
5. Select the appropriate product link.
6. Select your operating system.
7. Select Utility - Tools.
8. Click Download, next to the appropriate executable file to save it.
Recommended platform tuning

System requirements
The HPE BIOS configuration options described in this document include options in HPE ProLiant servers to disable the generation of periodic System Management Interrupts (SMIs) used for Power Monitoring and for Memory PreFailure Notification, with their attendant latency impact. BIOS options are generally independent of the OS, and a properly tuned low-latency operating system is also required to achieve deterministic performance.

The tuning recommendations described in this document are based on testing and customer interactions. But no single "recipe" can be prescribed. 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.

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

- Take an inventory or snapshot See “Taking inventories or snapshots” on page 6.
- HP-TimeTest7.3
HP-TimeTest is a utility distributed by HPE that enables customers to test for jitter in a server. To obtain the HP-TimeTest utility, contact HPE by emailing to: (lowlatency@hpe.com). Please include the name of your HPE contact.

- Capture kernel boot settings
  o For non-UEFI systems (Gen8 and earlier)
    • cat /boot/grub/grub.conf (for RHEL)
    • cat /boot/grub/menu.lst (for SLES)
  o For UEFI systems (DL580 Gen8 and Gen9)
    • cat /boot/efi/EFI/redhat/grub.conf (for RHEL)
  cat /boot/efi/EFI/SUSE/efilo.conf (for SLES) on page.
- Upgrade the BIOS. See “Upgrading BIOS and firmware” on page 6.
- If using a Linux-based server, prepare the server for low-latency tuning. See “Preparing Linux-based servers for low-latency tuning” on page 19.
- Make the recommended changes to the BIOS.
- For tuning recommendations and instructions, see the following sections:
  o “Tuning with the ROM-based Setup Utility (RBSU)” on page 15.
  o “Tuning with HPE RESTful Interface Tool (Gen9)” on page 15.
  o “Preparing for low-latency configuration” on page 6.

HPE 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 and hprcu utilities, configuration tools that are provided by HPE.

All HPE ProLiant G6 and later Intel-based servers, regardless of the ROM version, support setting Intel Turbo Boost and C-States. For G7 and earlier servers, HPE ProLiant 100 Series servers do not support advanced features for iLO Performance Monitoring and Memory Pre-Failure notification.
The following table provides descriptions of the recommended low-latency settings for Linux environments. For recommended Windows settings, see "Windows" on page 22.

<table>
<thead>
<tr>
<th>PARAMETER</th>
<th>VALUE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intel Virtualization Technology</td>
<td>Disabled</td>
<td>Allows Virtual Machine Managers to utilize virtualization hardware capabilities</td>
</tr>
<tr>
<td>(pre-UEFI menu)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virtualization Technology (UEFI menu)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intel Hyperthreading Options</td>
<td>Disabled</td>
<td>Allows Hyperthreading, which adds logical cores but increases computational jitter</td>
</tr>
<tr>
<td>Intel Turbo Boost Technology</td>
<td>Enabled</td>
<td>Allows processors to make a transition to a frequency that is higher than its rated speed. For more information, see &quot;Turbo mode information and considerations&quot; on page 12.</td>
</tr>
<tr>
<td>Intel VT-d</td>
<td>Disabled</td>
<td>Enables virtualized Directed I/O</td>
</tr>
<tr>
<td>Thermal Configuration</td>
<td>First try Optimal Cooling, then repeat with Increased Cooling and then Max Cooling (If available)*</td>
<td>Steps through the different available cooling settings available in RBSU. Use the one that provides the desired performance for the lowest power consumption. For more information, see &quot;Thermal considerations&quot; on page 14.</td>
</tr>
<tr>
<td>HPE Power Profile</td>
<td>Maximum Performance</td>
<td>Disables all power management options that may negatively affect performance</td>
</tr>
<tr>
<td>HPE Power Regulator</td>
<td>HPE Static High Performance Mode</td>
<td>Keeps processors in their maximum power/performance state (automatically set by HPE Power Profile for Gen8 and Gen9 servers)</td>
</tr>
<tr>
<td>Intel QPI Link Power Management</td>
<td>Disabled</td>
<td>Precludes placing unutilized QPI links into low power state</td>
</tr>
<tr>
<td>Minimum Processor Idle Power Core State</td>
<td>No C-States</td>
<td>Precludes processor transitions into low-power core C-States (automatically set by HPE Power Profile for Gen8 and Gen9 servers)</td>
</tr>
<tr>
<td>Minimum Processor Idle Power Package State</td>
<td>No Package State</td>
<td>Precludes processor transitions into low-power package C-States (automatically set by HPE Power Profile for Gen8 and Gen9 servers)</td>
</tr>
<tr>
<td>Energy/Performance Bias</td>
<td>Maximum Performance</td>
<td>Configures processor subsystems for high-performance/low-latency (automatically set by HPE Power Profile for Gen8 and Gen9 servers)</td>
</tr>
<tr>
<td>Collaborative Power Control</td>
<td>Disabled</td>
<td>Precludes the OS from changing clock frequency (automatically set by HPE Power Profile for Gen8 and Gen9 servers)</td>
</tr>
<tr>
<td>DIMM Voltage Preference</td>
<td>Optimized for Performance</td>
<td>Runs DIMMs at a higher voltage if it increases performance. (Gen8 only)</td>
</tr>
<tr>
<td>Dynamic Power Capping Functionality</td>
<td>Disabled</td>
<td>This option allows for disabling System ROM Power Calibration during the boot process. Doing so accelerates boot times but precludes enabling of a Dynamic Power Cap. (Gen8 and Gen9 only)</td>
</tr>
</tbody>
</table>
### Table 1. Recommended low-latency settings for Linux environments continued

<table>
<thead>
<tr>
<th>PARAMETER</th>
<th>VALUE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory Power Savings Mode</td>
<td>Maximum Performance</td>
<td>This option configures several memory parameters to optimize the memory subsystems performance and is configured to Balanced by default. (Gen8 only)</td>
</tr>
<tr>
<td>QPI Snoop Configuration</td>
<td>Early Snoop, Home Snoop, or Cluster on Die**</td>
<td>This option allows for the configurations of different snoop modes that impact the QPI interconnect. Changing this option may improve performance in certain workloads. Home Snoop provides high memory bandwidth in an average NUMA environment (default setting). Cluster on Die may provide increased memory bandwidth in highly optimized NUMA workloads. Early Snoop may decrease memory latency but may also result in lower overall bandwidth as compared to other modes. (Gen9 only)</td>
</tr>
<tr>
<td>QPI Home Snoop Optimization</td>
<td>Directory + OSB Enabled</td>
<td>This option allows the disabling of the Directory and Opportunistic Snoop Broadcast (OSB) functionality that is available when the QPI snoop mode is set to Home Snoop. Directory and OSB provides better memory latency and increased bandwidth and is recommended for the vast majority of workloads that benefit from Home Snoop and is enabled by default (Gen9 only)</td>
</tr>
<tr>
<td>ACPI SLIT Preferences</td>
<td>Enabled</td>
<td>This ACPI SLIT describes the relative access times between processors, memory subsystems, and I/O subsystems. Operating systems that support the SLIT can use this information to improve performance by allocating resources and workloads more efficiently. This option is disabled by default on most ProLiant Gen8 and Gen9 servers.</td>
</tr>
<tr>
<td>Processor Power and Utilization Monitoring</td>
<td>Disabled***</td>
<td>Disables ILO Processor State Mode Switching and Insight Power Manager Processor Utilization Monitoring, and its associated SMI</td>
</tr>
<tr>
<td>Memory Pre-Failure Notification</td>
<td>Disabled***</td>
<td>Disables Memory Pre-Failure Notification and its associated SMI</td>
</tr>
<tr>
<td>Memory Patrol Scrubbing</td>
<td>Disabled***</td>
<td>Memory Periodic Patrol Scrubber corrects memory soft errors so that, over the length of the system runtime, the risk of producing multibit and uncorrectable errors is reduced.</td>
</tr>
<tr>
<td>Memory Refresh Rate</td>
<td>HPE Static High Performance Mode***</td>
<td>Keeps processors in their maximum power/performance state (automatically set by HPE Power Profile for Gen8 and Gen9 servers)</td>
</tr>
</tbody>
</table>

*If Turbo mode is enabled, then step through the available cooling settings described in “Thermal considerations” on page 14. Otherwise, the default Optimal Cooling setting is adequate.

**QPI Snoop Configuration selection depends on the processor and workload used. See “QPI Snoop Configuration information and considerations (Gen9 only)” on page 11.

***These options are under the Service Options menu. See “Tuning with ROM Based Setup Utility (RBSU)” on page 15, “Tuning with conrep” on page 16, or “Tuning with HPE RESTful Interface Tool” on page 12 for details on how to set these options.

### QPI Snoop Configuration information and considerations (Gen9 only)

The QPI Snoop Configuration setting will control how cache snoops are handled. When using the “Early Snoop” option the snoops will be sent by the caching agents; this will provide better cache latency for processors when the snoop traffic is low. The “Home Snoop” option will cause the
snoops to be sent from the home agent, this provides optimal memory bandwidth balanced across local and remote memory access. The “Cluster on Die” option will snoop the directory cache first and then the home agent. Using this option will also cause the processor to appear as two NUMA nodes within operating systems, one for each memory controller. This option provides optimal performance for highly NUMA-aware workloads. Note that the “Cluster on Die” option is available only on processors with 10 or more cores.

Xeon E5-2600 v4 series processors add further optimization to the Home Snoop configuration with the “Directory and Opportunistic Snoop Broadcast (OSB)” option under the ‘QPI Home Snoop Optimization’ RBSU setting. Each home agent has a small cache that holds the directory state of migratory cache lines. The home agent will speculatively snoop the remote socket in parallel with the directory read. This enables low cache-to-cache latencies, low memory latencies, and higher memory bandwidths. Home Snoop with Dir+OSB generally will be the best snoop mode for most workloads. See the table below for a summary of the QPI Snoop options.
Table 2. QPI Snoop modes supported in two-socket configurations

<table>
<thead>
<tr>
<th></th>
<th>EARLY SNOOP</th>
<th>HOME SNOOP (DEFAULT RBSU OPTION)</th>
<th>HOME SNOOP with DIRECTORY + OSB</th>
<th>CLUSTER ON DIE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previously available on</td>
<td>E5-2600 (SNB)</td>
<td>E5-2600 v2 (VB)</td>
<td>E5-2600 v4 (BDW)</td>
<td>E5-2600 v3 (HSW)</td>
</tr>
<tr>
<td>Snoop sent by</td>
<td>Caching Agent</td>
<td>Home Agent</td>
<td>Home Agent + speculative remote socket snooping</td>
<td>Directory Cache, then Home Agent</td>
</tr>
<tr>
<td>Best used for</td>
<td>Memory latency-sensitive workloads*</td>
<td>NUMA workloads that need maximum local and remote bandwidth</td>
<td>NUMA workloads that need maximum local and remote bandwidth</td>
<td>Highly NUMA-optimized workloads</td>
</tr>
</tbody>
</table>

*For v4 series processors, Home Snoop with Dir+OSB provides better local NUMA latencies with higher memory bandwidth and is recommended over Early Snoop.

Core frequencies for AVX vs. non-AVX applications information and considerations (Gen9 only)

With the Intel Xeon E5-2600 v3 series and later processors, Advanced Vector Extensions version 2.0 (AVX2) allow applications to perform 256-bit wide operations for integer and floating-point operations, providing an opportunity for increased performance. However, the power requirements for running AVX instructions are higher than for non-AVX instructions. Therefore the CPU’s core frequency range will change depending upon whether AVX instructions are executing or not. Cores that are executing AVX instructions will be constrained to a lower frequency range (AVX base and AVX Turbo) while running the instructions. The CPU’s core frequency will return to the non-AVX frequency range ~1msec after the AVX instructions have completed. The table below shows both the AVX and non-AVX frequency ranges for three segment-optimized E5-2600 v4 processors. Note that the processor is still also governed by the power/thermal characteristics of the system, so the actual frequency will be determined by both the type of instructions used and the power/thermal conditions.

Turbo mode information and considerations

Intel Turbo Boost can be used to increase the processor's operating clock frequency, but at the risk of computational jitter if the processor changes its turbo frequency. When that happens, processing stops for a small period of time, introducing uncertainty in application processing time. Turbo operation is a function of power consumption, processor temperature, and the number of active cores. Carefully managing these factors, however, can result in consistent turbo operation without jitter. The maximum turbo frequencies for various numbers of active cores for two selected processors are given in the following table.

Table 3. Turbo frequency ranges for certain E5-2600 v4 Series processors

<table>
<thead>
<tr>
<th>PROCESSOR</th>
<th>POWER</th>
<th>BASE FREQUENCY</th>
<th>NUMBER OF ACTIVE CORES</th>
<th>TURBO-ENABLED FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>AVX</td>
<td>Non-AVX</td>
<td>AVX</td>
</tr>
<tr>
<td>E5-2687W v4</td>
<td>160 W</td>
<td>2.6 GHz</td>
<td>3.0 GHz</td>
<td>4-12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>32 GHz</td>
<td>3.3 GHz</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-2</td>
<td>34 GHz</td>
<td>3.5 GHz</td>
</tr>
<tr>
<td>E5-2667 v4</td>
<td>135 W</td>
<td>2.6 GHz</td>
<td>3.2 GHz</td>
<td>3-8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-2</td>
<td>3.5 GHz</td>
<td>3.6 GHz</td>
</tr>
<tr>
<td>E5-2637 v4</td>
<td>135 W</td>
<td>3.2 GHz</td>
<td>3.5 GHz</td>
<td>3-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-2</td>
<td>3.6 GHz</td>
<td>3.7 GHz</td>
</tr>
</tbody>
</table>

If the penalty of computational jitter is too severe and you are unable to control temperature and TDP, you should disable Turbo Mode or reduce the maximum turbo frequency.
Power consumption
Pushing the processor's TDP limit will result in the processor changing its turbo frequency if the processor consumes too much power. Because of the risk of processor failure, Intel offers no method to lock a processor into Turbo Mode. Most applications will not consume enough power to exceed the processor's TDP. If you are concerned that yours might, then you can disable a core per processor from within the BIOS or set a core offline from the OS, reducing power consumption and providing TDP headroom.

Tests have shown that the E5-2690 v4 processor under heavy computational load is able to stay at the maximum Turbo frequency indefinitely when the system is properly configured, as outlined in this document. However, this is not guaranteed behavior and you should verify this with your workload. Also be aware that due to processor manufacturing variations, some CPUs may be able to stay at full Turbo frequency while other samples of the same processor model may not.

Thermal considerations
The processor's thermal limits are another consideration in maintaining consistent turbo operation. Ensure that the server's inlet temperature meets the specification in the associated QuickSpecs. Beyond that, there is a BIOS parameter that can be used to regulate the amount of cooling delivered by the fans, but before changing it, note that most configurations will maintain the preferred operating state with the default Optimal Cooling setting. If the system requires more cooling, the server will respond by increasing the fan speed to deliver the necessary cooling.

However, some demanding environments may require a greater base level of cooling. If testing shows that your server's turbo frequency varies in response to exceeding temperature limits due to varying system load, evaluate the Increased Cooling option, which carries a penalty of increased system power consumption, acoustics, and airflow demand.

The third setting for this parameter is Maximum Cooling, which causes the fans to operate always at their highest speed. Use this setting only if your environment requires it, as it has significantly higher power consumption, acoustic noise, and facility airflow demand.

Keep in mind that different processors have different requirements. The E5-2687W v4 has a notably higher TDP than the E5-2690 v4, but the TCase for the E5-2687W v4 is 13°C (23°F) lower than for the E5-2690 v4, making proper cooling especially important.

Active cores
In addition to TDP and thermals, the amount of frequency boost obtained is a function of the number of active cores, which is never more than the number of operational cores as specified by a BIOS setting. Active cores are cores in C0, C1, or C1E State, and HPE recommends disabling C-States in order to keep the number of active cores constant and avoid the attendant latency jitter of changing turbo frequencies.

Other considerations for Turbo Mode
As noted in “Active cores” page 14, C-States must be disabled in the BIOS. However, some versions of Linux ignore the BIOS setting and must be configured to disable C-States. For more information, see “Recommended Linux boot-time settings” on page 21.

Disabling processor power and utilization monitoring and memory pre-failure notification SMI
Disabling System Management Interrupts to the processor provides one of the greatest benefits to low-latency environments. Disabling the Processor Power and Utilization Monitoring SMI has the greatest effect because it generates a processor interrupt eight times a second in G6 and later servers. Disabling the Memory Pre-Failure Notification SMI has a much smaller effect because it generates an interrupt at a low frequency: once per hour on G6 and G7 servers, once every five minutes on Gen8 servers, and once every minute on the DL580 Gen8 and all Gen9 servers.

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 Processor State Monitoring
• Insight Power Manager CPU Utilization Reporting
• HPE 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 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 correctable error threshold is exceeded.

**Disabling Dynamic Power Capping Functionality**

Disabling Dynamic Power Capping Functionality prevents the ability to enable a Power Cap via iLO. When this parameter is disabled, the option to enable a Power Cap via iLO is no longer available. Since low-latency installations are unlikely to set power caps, the Dynamic Power Capping Functionality option may be safely disabled in the BIOS. This option accelerates the boot process but does not have any impact on latency when the platform is operating.

**Disabling Patrol Scrubbing**

Patrol Scrubbing is a feature that scans memory to correct soft memory errors. On the HPE ProLiant Gen9 Server, the Patrol Scrubber re-arms itself through an SMI. The frequency of this event is roughly once per day, but varies based on the amount of installed memory. Low-latency installations can avoid this SMI by disabling Patrol Scrubbing, which is an option in the Service Options menu. On other platforms, Patrol Scrubbing does not require SMI functionality and does not need to be disabled.

**Setting the Memory Refresh Rate**

An extremely rare potential for memory errors is eliminated by the default memory refresh rate of 2x. Decreasing the rate to 1x will improve memory performance, but with a vanishingly small potential for memory errors. This affects G6 and later servers. This option is available in the Service Options menu.

**Tuning with the ROM-based Setup Utility (RBSU)**

For new system testing or in environments with a small number of systems where script-based maintenance is not used, the RBSU in the UEFI Systems utilities is the recommended method to configure the BIOS.

To configure BIOS low-latency options using RBSU:

1. Power the server on.
2. When prompted during POST, press **F9** to enter RBSU (Gen8 and earlier) or System Utilities (Gen9).
3. For Gen9 servers, select **System Configuration → BIOS/Platform Configuration (RBSU)**.
4. Browse through the menus to change the parameters. For more information, see "Tuning recommendations and explanations" on page 2.

**Important**

Do not change the other options in the Services Options menu.

5. For the parameters marked with "****" in the "Tuning recommendations and explanations" table on page 2, go into the Service Options menu:
   a. For Gen8 and earlier, while in the top level of the RBSU menu, press **CTRL-A** to display the option for the Service Options menu. Select **Service Options**
   b. For Gen9 and DL580 Gen8, press **CTRL-A**. You will be immediately redirected to the Service Options menu.
6. Verify that the parameters are set as indicated in "Tuning recommendations and explanations" on page 2.

**Tuning with HPE RESTful Interface Tool (Gen9)**

The HPE RESTful Interface tool is useful for scripting the deployment of BIOS options across multiple servers from a common profile file. For complete details on how to utilize the HPE RESTful Interface tool for scripting, please refer to the HPE RESTful Interface Tool 1.50 User Guide on the HPE website (See: Support and other resources * on page 25 for links).
This section will provide details on how to download the service menu options to a json profile file and modify those settings via hprest. To configure BIOS low-latency options using the HPE RESTful Interface Tool:

1. Edit the /etc/hprest/hprest.conf file and update the following fields:
   - `url = https://[iLO IP address]`
   - `username = [iLO user account name]`
   - `password = [iLO user password]`

2. Change the current directory to a convenient working directory:
   - `cd /home/user/hprest`

3. Capture a snapshot of your current Service menu settings:
   - `hprest rawget /rest/v1/systems/1/bios/service/settings > service.opts.txt`

4. To disable Processor Power and Utilization Monitoring, disable Memory Pre-Failure Notification, and set Memory Refresh Rate to 1x, and create a patch file, `service.patch`, with the following format:
   ```json
   {
   "path": "/rest/v1/systems/1/bios/Service/Settings",
   "body": {
      "ProcPwrUtilMonitor": "Disabled",
      "MemPreFailureNotification": "Disabled",
      "MemRefreshRate": "1xRefresh"
   }
   }
   ```

5. To disable Patrol Scrubbing, add to the `service.patch` patch file the following markup in the “body” stanza:
   ```json
   "MemPatrolScrubbing": "Disabled",
   ```

---

**Important**
The “body” stanza in the `service.patch` file can contain multiple BIOS settings. Each setting should be separated by a comma, the final setting should not have a comma following it. Also, you cannot include settings from the Service RBSU menu and the other menus in the same patch file.

---

6. Update the BIOS with the modified settings:
   - `hprest rawpatch service.patch`

7. Log out of hprest’s iLO session:
   - `hprest logout`

8. Reboot the server:
   - `reboot`

**Tuning with conrep (Gen8 and older)**

conrep is useful for scripting the deployment of BIOS options across multiple servers from a common profile file. For complete details in how to utilize conrep for scripting, please refer to the HP Scripting Toolkit for Linux User Guide on the HPE website. (See “Support and other resources” on page 23 for links).

---

**Important**
Using conrep to modify BIOS settings may result in different behavior than using RBSU or the HPE RESTful interface tool. In particular, caution must be made when changing the HPE Power Profile with conrep as it will not propagate changes to other BIOS settings like RBSU and the HPE RESTful interface tool will. When changing the HPE Power Profile using conrep, you will also need to change the settings shown in Table 1 on page 10 that are linked to the HPE Power Profile.
This section will provide details on how to add the service menu options to the default profile file, conrep.xml and modify those setting via conrep. To configure low-latency Service Menu BIOS options using the conrep utility in STK:

1. Change the current directory to the STK/Utilities directory:
   ```
   cd STK/Utilities
   ```

2. Edit the conrep.xml file to include the following stanzas before `</Conrep>` at the end of the file:

   ```xml
   <Section name="PowerMonitoring">
     <helptext><![CDATA[This setting determines if Pstate logging and utilization is supported.]]></helptext>
     <ev>CQHGV3</ev>
     <length>1</length>
     <value id="0x00">Enabled</value>
     <value id="0x10">Disabled</value>
     <mask>0x10</mask>
     <byte>0</byte>
   </Section>
   <Section name="DisableMemoryPrefailureNotification">
     <helptext><![CDATA[This setting allows the user to disable Memory Pre-Failure Notification support, which will remove the periodic SMI associated with this support. Not recommended for anyone except for those who absolutely need every periodic SMI removed.]]></helptext>
     <ev>CQHGV3</ev>
     <length>1</length>
     <value id="0x00">No</value>
     <value id="0x20">Yes</value>
     <mask>0x20</mask>
     <byte>0</byte>
   </Section>
   <Section name="Memory_Refresh_Rate_Gen9">
     <helptext><![CDATA[This setting allows the user to change the Memory Refresh Rate setting on Gen9 servers.]]>
     <platforms>
       <platform>Gen9</platform>
     </platforms>
     <nvram>0x257</nvram>
     <value id="0x00">1x_Refresh</value>
     <value id="0x10">2x_Refresh</value>
     <value id="0x20">3x_Refresh</value>
     <mask>0x30</mask>
   </Section>
   <Section name="Memory_Refresh_Gen8">
     <helptext><![CDATA[This setting allows the user to change the Memory Refresh setting on Gen8 servers.]]>
     <platforms>
       <platform>Gen8</platform>
     </platforms>
     <nvram>0x261</nvram>
     <value id="0x00">1x_Refresh</value>
     <value id="0x01">2x_Refresh</value>
     <value id="0x02">3x_Refresh</value>
     <mask>0x03</mask>
   </Section>
   <Section name="Memory_Patrol_Scrubbing_Gen9">
   ```
<helptext><![CDATA[This setting allows the user to enable or disable the Memory Patrol Scrubbing setting on Gen9 servers.]]></helptext>

<platforms>
    <platform>Gen9</platform>
</platforms>

<nvram>0x257</nvram>

<value id="0x08">Disabled</value>
<value id="0x00">Enabled</value>

<mask>0x08</mask>

3. Capture a snapshot of your current settings:

   ./conrep -s -x conrep.xml -f conrep_settings.xml

4. To disable Intel Turbo Boost Technology, verify that the conrep_settings.xml file contains the following markup (G7 servers):

   <Section name="Intel_Processor_Turbo_Mode" helptext="Allows Intel processors to transition to a higher frequency than its rated speed if the processor has available headroom and is within temperature specification.">Disabled</Section>

5. To disable Intel Turbo Boost Technology, verify that the conrep_settings.xml file contains the following markup (Gen8 and Gen9 servers):

   <Section name="Intel_Turbo_Boost_Optimization_Gen8" helptext="Optimize Turbo Boost heuristics for different situations. For Gen8 or later servers only.">Disabled</Section>

6. To disable Processor Power and Utilization Monitoring, verify that the conrep_settings.xml file contains the following markup:

   <Section name="PowerMonitoring" helptext="This setting determines if Pstate logging and utilization is supported.">Disabled</Section>

7. To disable Memory Pre-Failure Notification, verify that the conrep_settings.xml file contains the following markup:

   <Section name="DisableMemoryPrefailureNotification" helptext="This setting allows the user to disable Memory Pre-Failure Notification support, which will remove the periodic SMI associated with this support. Not recommended for anyone except for those who absolutely need every periodic SMI removed.">Yes</Section>

8. To disable Memory Patrol Scrubbing (Gen9 servers), verify that the conrep_settings.xml file contains the following markup:

   <Section name="Memory_Patrol_Scrubbing" helptext="This setting allows the user to enable or disable the Memory Patrol Scrubbing setting on Gen9 servers.">Disabled</Section>

9. Update the BIOS with the modified settings:

   ./conrep -l -x conrep.xml -f conrep_settings.xml

10. Reboot the server:

    reboot
Recommended operating system tuning

Linux

Preparing Linux-based servers for low-latency tuning
Before configuring a ProLiant Gen8 or later server for low latency, do the following:

1. Make the following edits:
   o For non-UEFI configurations (Gen8):
     - Red Hat (EL 6x): Edit /boot/grub/grub.conf and add "nosoftlockup intel_idle.max_cstate=0 mce=ignore_ce" to the kernel line
     - SLES: Edit /boot/grub/menu.lst and add "nosoftlockup intel_idle.max_cstate=0 mce=ignore_ce" to the kernel line
   o For UEFI configurations (Gen9 and DL580 Gen8):
     - Red Hat (EL 6x): Edit /boot/efi/EFI/redhat/grub.conf and add "nosoftlockup intel_idle.max_cstate=0 mce=ignore_ce" to the kernel line
     - SLES: Edit /boot/efi/efi/SuSE/elilo.conf and add "nosoftlockup intel_idle.max_cstate=0 mce=ignore_ce" to the kernel line
   o For Red Hat Enterprise Linux Server 7.0 or greater:
     - Edit /etc/default/grub file and add "nosoftlockup intel_idle.max_cstate=0 mce=ignore_ce" to the "GRUB_CMDLINE_LINUX" value.

Run command:

```
# grub2-mkconfig -o /boot/grub2/grub.cfg (non-UEFI configurations) or
# grub2-mkconfig -o /boot/efi/EFI/redhat/grub.cfg (UEFI configurations)
```

nosoftlockup prevents the kernel from logging an event when a high-priority thread executes continuously on a core for longer than the soft lockup threshold.

intel_idle.max_cstate=0 prevents the kernel from overriding the BIOS C-State setting.

mce=ignore_ce prevents Linux from initiating a poll every five minutes of the Machine Check Banks for correctable errors, which can cause latency spikes. For more information, see the Linux Kernel Archives website (http://www.kernel.org/doc/Documentation/x86/x86_64/boot-options.txt).

2. Set tuned profile (RHEL only). Tuned is a utility introduced in RHEL 6 that allows the user to implement a set of OS optimizations as part of a profile. Red Hat provides a set of pre-defined profiles that can be used. For RHEL 6, it is recommended that the "latency-performance" profile is used for latency-sensitive applications. For RHEL 7, "network-latency" is recommended for low latency environments. For details on the performance options being set, please see the /usr/lib/tuned/[performance profile]/tuned.conf file for the desired performance profile. To set the desired profile, run the command:

```
# tuned-adm profile latency-performance (RHEL 6) or
# tuned-adm profile network-latency (RHEL 7)
```

3. Reboot the server.
4. After reboot, run the `stop-services.sh` script to stop extraneous services. The following example stops the services shown and prevents them from starting on subsequent boots:

```bash
for SERVICE in \n  acpid alsasound autofs avahi-daemon bluetooth \n  conman cups cupsrenice \n  dhcdbd dnsmsq dund firstboot hidd \n  ip6tables irida kudzu libvirtd \n  lvm2-monitor mcstrans mdmonitor mdmpd messagebus \n  multipathd netconsole netfs netplugd nsed \n  oddjobd pand pcscd postfix powersaved \n  psacct rdsc readahead_early readahead_later restoresecond \n  rhnsd rpcgssd rpscmdmapd rpscvgssd saslauthd \n  sendmail slpd smartd smbfs suseRegister \n  sysstat wpa_supplicant xfs vpbind yum-updatesd \n
do
  chkconfig --level 2345 $SERVICE off
  service $SERVICE stop
done
```

**Note**
For RHEL 7 systems, use the following script to disable services:

```bash
for SERVICE in \n  avahi-daemon.service crond.service dnsmsq.service \n  firewalld.service lvm2-monitor.service postfix.service \n  rpcgssd.service rpscmdmapd.service rpscvgssd.service \n  wpa_supplicant.service \n
do
  systemctl disable $SERVICE
  systemctl stop $SERVICE
done
```

5. Use the irqbalancer to preclude some cores from servicing software IRQs:

a. Enter the following command:

b. `# service irqbalance stop`

c. Do a one-time run of the irq balancer, where `CoreMask` is a mask of the CPUs on which the OS should not schedule interrupts

d. `# IRQBALANCE_ONESHOT=1 IRQBALANCE_BANNED_CPUS=${CoreMask} irqbalance`

e. Wait until the command returns "irqbalance is stopped."

f. On SLES, the name of the IRQ balancer service is `irq_balancer`.

g. On RHEL 7, use `systemctl` instead of `service` command to stop `irqbalance`.

**Red Hat MRG Realtime**

Red Hat resolved scaling issues for the MRG 2.3 operating system for ProLiant servers with large core counts, such as the DL580 G7 server with four 10-core E7-4870 processors. If you are using MRG 2.3 on servers with a large number of cores, be sure to use a release with a kernel version equal to or greater than the following:
kernel-rt-3.6.11-rt30.25.el6rt

In addition to having a large number of cores, if your server is running the MRG 23 (or later) Realtime kernel, it is using the SLUB memory allocator. The SLUB memory allocator requires additional tuning for real-time performance. The SLUB allocator has pseudo-files named "cpu Partial" in the `/sys/kernel/slab` file system. To get the best real-time performance from the allocator, these files should be set to 0, disabling the cpu partial logic. This can be done with the following command:

```bash
# find /sys/kernel/slab -name 'cpu Partial' -exec sh -c 'echo 0 > {}' \
```  

**Recommended Linux boot-time settings**

The Linux boot parameter "idle=poll" keeps the processing cores in C0 state when used in conjunction with "intel_idle.max_cstate=0". Without it, the processor will enter C1 state.

- For RHEL systems:
  
  For RHEL 6, edit `/boot/grub/grub.conf` (or `/boot/efi/EFI/redhat/grub.conf` for UEFI systems) and add "idle=poll" to the kernel line. This is in addition to the "nosoftlockup intel_idle.max_cstate=0 mce=ignore_ce" parameters that should have been added previously.

  For RHEL 7, edit `/etc/default/grub.cfg` and add "idle=poll". This is in addition to the "nosoftlockup intel_idle.max_cstate=0 mce=ignore_ce" parameters that should have been added previously. After the edit, run the command:

  ```bash
  # grub2-mkconfig -o /boot/grub2/grub.cfg (non-UEFI configurations) or
  # grub2-mkconfig -o /boot/efi/EFI/redhat/grub.cfg (UEFI configurations)
  ```

- For SLES systems:
  
  Edit `/boot/grub/menu.lst` (or `/boot/efi/EFI/SuSE/elilo.conf` for UEFI systems) and add "idle=poll" to the kernel line. This is in addition to the "nosoftlockup intel_idle.max_cstate=0 mce=ignore_ce" parameters that should have been added previously.

**Verifying the configuration**

To verify your ProLiant server is properly configured for low-latency operation, clear one core (selected at random) of the operating system IRQs, and then run the HP-TimeTest utility on the randomly selected core:

```bash
Core=5
CoreMask=`echo "16 o 2 $Core ^ p" | dc`

service irqbalance stop
until [ "service irqbalance status" = "irqbalancer is stopped" ] ; do sleep 1 ; done
IRQBALANCE_ONESHOT=1 IRQBALANCE_BANNED_CPUS=${CoreMask} irqbalance
sleep 1
until [ "service irqbalance status" = "irqbalancer is stopped" ] ; do sleep 1 ; done
numactl --physcpubind=${Core} --localalloc nice -n -20 ./HP-TimeTest7.3 -v -f csv -o smi_count
```

On SLES, the name of the IRQ balancer service is irq_balancer.

On RHEL 7, use "systemctl" to disable and monitor the status of the "irqbalance service" process.

Consider the following:

- Consider changing the smp.affinity for the IRQs. For example, on a server on which you want to leave core 0 for the OS, the following masks off the other processors for all IRQs:

  ```bash
  for MF in `find /proc/irq -name *smp.affinity` ; do awk -F, \
  '{for(i=1;i<NF;i++)printf("00000000,",);}printf("%8.8x\n",and(0x00000001, \
  strtonum("0x"$NF))))' \n  $MF > $MF ; done
  ```
Consider using `cset` (http://code.google.com/p/cpuset/) to shield cores from the OS. For example, on a server on which you want to keep the OS from all cores except for 0, use the following command:

```
# cset shield --cpu 1-15 --kthread=on
```

If running as root, the following command can then be used to move the current PID to the "user" set of cores:

```
# cset proc --move --pid=$$ --threads --toset=user
```

**Windows**

HPET BIOS low-latency options are supported in Windows Server 2008 and 2012 environments.

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

1. Obtain the STK. See “Obtaining the Scripting Utilities” on page 8.
2. Run the SmartComponent for the most recent version of the STK, note the directory it is in, and then change to it in Windows Explorer or a command window.
3. Run conrep. See “Tuning with conrep (Gen8 and older)” on page 16.

For other low-latency tuning recommendations in a Windows environment, do the following:


For more information or assistance, contact Microsoft to be put in touch with one of their low-latency experts.
**HPE-TimeTest**

The original behavior of HPE-TimeTest has been maintained through its many edits, but this behavior is not optimal. For example, it runs at real-time priority 99, but should be run at no higher than 80. On an otherwise idle system, a real-time priority of "1" is adequate for HPE-TimeTest to run properly.

The following provides an example of running HPE-TimeTest with an explanation of each component of the command:

```bash
time numactl --physcpubind=3 \ Bind to core 3 and use local memory
--localalloc
nice -n -20 \ nice; probably not necessary
/HP-TimeTest/HP-TimeTest7.3 \ HP-TimeTest7.3 executable
-f csv \ output in Comma Separated Variable (csv) format
-o smi \ print SMI_count at the beginning and end
-o date \ print a timestamp at the beginning and end
-m cycles \ latency is determined by cycles (instead of time)
-t `echo '.000005 2900000000 * 0 k \ threshold is 5 μsec on 2.90 GHz processor
1 / p' | dc` \ run for ~30 minutes on 2.90 GHz processor
44` \ ("44" is # of cycles per loop iteration I get)
-p RR,1,-20 \ Use RR scheduling at priority 1; use “nice”
\ of -20 (I suspect irrelevant for RT policies)
```

Generating the output in CSV format allows for easy import into a spreadsheet for plotting.

To provide additional suggestions, contact the HPE low-latency team.
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. HPE 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 Processor State Monitoring in the HPE 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. Use one of the following options to verify that the low-latency option is set:

- See the information in “Tuning recommendations and explanations” on page 2.
- Run HPE-TimeTest to see if you are getting spikes. For more information, contact HPE by emailing to low_latency@hpe.com. Please provide the name of your local HPE representative and region of origin so that we can better serve your request.

Q. Can I interrogate or confirm the memory operating speed?

A. To interrogate or confirm the memory operating speed, ensure your SMBIOS is 2.7 or later and use dmidecode 2.11 or later with the following command:

```
dmidecode -t 17
```

Q. How do I tune a network adapter for optimum low latency?

A. This white paper does not address this topic. Refer to the supplier of the network adapter's controller technology. For example, tuning advice for Mellanox ConnectX-3 adapters integrated and supported by HP is available on the Mellanox website (http://www.mellanox.com/reledocs/prod_software/Performance_Tuning_Guide_for_Mellanox_Network_Adapters.pdf).

Q. How does HP recommend I disable cores in ProLiant Gen8 servers?

A. Do the following:

1. From the RBSU menu, navigate to **System Options > Processor Options > Processor Core Disable (Intel Core Select).**
2. Enter the number of cores per processor that you want to enable.
   For example, if you have 8-core processors and want to disable 1 core, enter “7” in this field.
3. Boot the server. Verify that the correct information appears during POST; for example, “2 Processor(s) detected, 14 total cores enabled.”
   The number of enabled cores can also be modified with hprest or conrep. To modify the number of enabled cores with conrep, use version available from STK for Linux 10.0 or later. To modify the number with hprest, use HPE RESTful Interface Tool 1.10 or later.

Q. How do I verify what turbo frequency my cores are running at?

A. There are a number of utilities that track the real time frequency of each CPU core. For example, for Linux:

1. **i7z** is an open source utility that provides information on the Intel Core i3, i5, i7, and corresponding Xeon processors. Pre-compiled versions of this utility can be found for most Linux distributions, including Red Hat and SLES.
2. Red Hat Enterprise Linux 6.4 and later provides the utility turbostat as part of its cpupowerutils package. Both of these utilities will provide real-time information about each cores’ frequency and percent time in each C-state.
Support and other resources

Resources and documentation
The following resources are available:

- HPE ROM-Based Setup Utility User Guide on the HPE website (http://www.hpe.com/support/tbsu)
- iLO documentation:
  - HPE iLO 4 User Guide (for Gen8 and Gen9 servers) on the HPE website (http://h10032.www1.hp.com/ctg/Manual/c03334051)

  The conrep, hrcpu, and hdpscovery utilities are available as part of the STK. For more information on downloading STK, see “Obtaining the Scripting Utilities” on page 8.

- HP-TimeTest 7.3 utility. To obtain the utility, contact HPE by emailing: low.latency@hpe.com. Please provide the name of your local HPE representative and region of origin so that we can better serve your requests.

Before you contact HPE
Be sure to have the following information available before you call HPE:

- Active Health System log
- Download and have available an Active Health System log for three days before the failure was detected. For more information, see the HPE iLO 4 User Guide or HPE Intelligent Provisioning User Guide on the HPE website at http://www.HPE.com/go/ilo/docs
- Onboard Administrator SHOW ALL report (for HPE BladeSystem products only)
  
  For more information on obtaining the Onboard Administrator SHOW ALL report, see the HPE website http://h20628.www2.hp.com/km-ext/kmc directs/emr_na-c00705292-47.pdf
- 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

HPE contact information
For United States and worldwide contact information, see the Contact HPE website at http://www8.hp.com/us/en/hpe/contact/contact.html.

In the United States:
• To contact HPE 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 at http://h20565.www2.hpe.com/portal/site/hpsc. If the problem cannot be resolved at the website, call 1-800-633-3600. For more information about Care Packs, see the HPE website at https://ssc.hpe.com/portal/site/ssc/.

On a best-effort basis only, HPE offers technical assistance on low-latency tuning to customers who have followed this guide and still have questions. For more information, contact HPE by emailing low_latency@hpe.com. Please provide the name of your local HPE representative and region of origin so that we can better serve your request.
Acronyms and abbreviations

**ACPI**
Advanced Configuration and Power Interface specification

**AMP**
Advanced Memory Protection

**AVX**
Intel Advanced Vector Extension

**BIOS**
Basic Input/Output System

**DDIO**
Distributed Discrete Input/Output

**HPRCU**
HP ROM Configuration Utility

**ILO**
Integrated Lights-Out

**LOM**
LAN on Motherboard

**MRG**
Red Hat Enterprise Messaging Realtime Grid platform

**POST**
Power-On Self Test

**QPI**
Intel QuickPath Interconnect

**RBSU**
ROM-Based Setup Utility

**SLES**
SUSE Linux Enterprise Server

**SLIT**
System Locality Information Table

**SLUB**
Unqueued slab memory allocator

**SMI**
System Management Interrupt

**STK**
Scripting Toolkit

**TDP**
Thermal Design Power

**UEFI**
Unified Extensible Firmware Interface