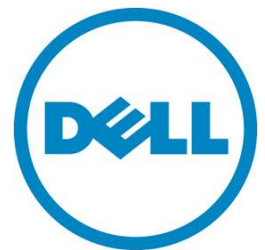

Configuring Low-Latency Environments on Dell PowerEdge 12th Generation Servers

Dell PowerEdge 12th generation servers can be optimized for maximum throughput or lowest latency. This technical white paper discusses best practices for low-latency environment optimization.

John Beckett | Solutions Performance Analysis

David J. Morse | Solutions Performance Analysis

Mukund Khatri | Server Advanced Engineering



Contents

Introduction	3
Recommendation 1: Choose an Optimal Server/Processor Architecture	4
Recommendation 2: Update the PowerEdge BIOS and Firmware	5
Recommendation 3: Tune the PowerEdge BIOS for Low Latency	5
Summary	7
References	8

This document is for informational purposes only and may contain typographical errors and technical inaccuracies. The content is provided as is, without express or implied warranties of any kind.

© 2012 Dell Inc. All rights reserved. Dell and its affiliates cannot be responsible for errors or omissions in typography or photography. Dell, the Dell logo, OpenManage, and PowerEdge are trademarks of Dell Inc. Intel and Xeon are registered trademarks of Intel Corporation in the U.S. and other countries. Other trademarks and trade names may be used in this document to refer to either the entities claiming the marks and names or their products. Dell disclaims proprietary interest in the marks and names of others.

May 2012 | Rev 1.0

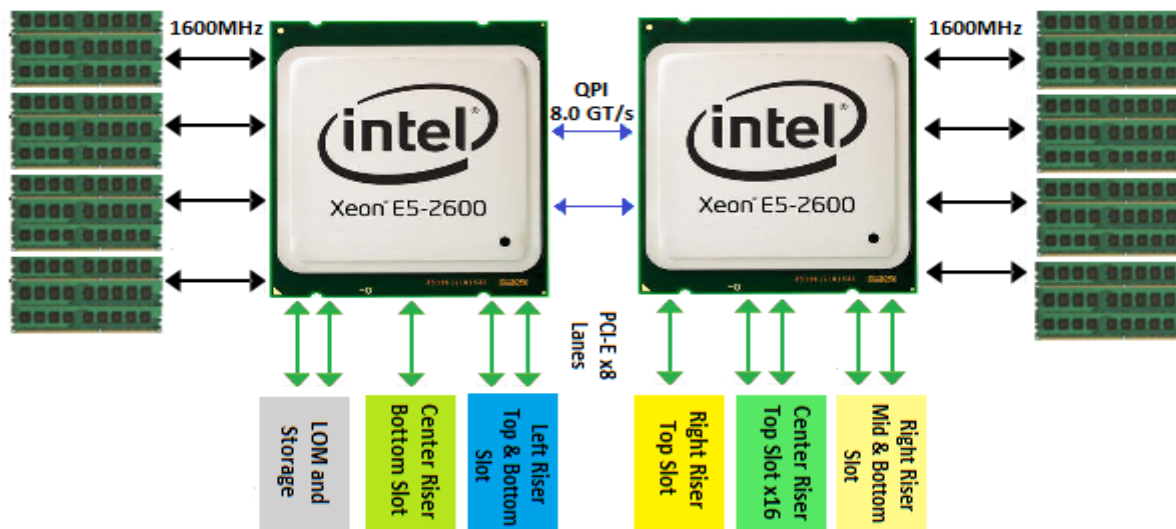
Introduction

This white paper focuses on best practices for reducing latency within Dell™ PowerEdge™ 12th generation server hardware. With today's multi-socket, multi-core, highly-threaded [PowerEdge servers](#), the operating system, applications, and drivers are expected to be written to take advantage of this massively parallel architecture. While most industry-standard benchmarks and tools (for example, [SPECrate®](#), [SPECjbb®2005](#), VMware® [VMmark™](#), and database benchmarks from the [Transaction Processing Performance Council](#)) can be configured and optimized to saturate all the processing power of these servers, these benchmarks typically measure **throughput** (for example, transactions, input/output (I/O), or pages per second). However, many organizations, especially in the financial industry (where high-frequency trading occurs), still care about reducing the time it takes to solve a single task. In these cases, the focus must be on reducing system **latency** (typically measured in nanoseconds, microseconds, or milliseconds) rather than increasing throughput. Network latency improvements are also partially tied to system latency improvements, so tuning for these environments is similar. To reduce system latency, the entire solution must be taken into consideration:

- The server — including processor and memory architecture and BIOS tuning
- The network stack — especially the choice of network controllers and network driver tunings such as coalesce settings
- Operating system (OS) selection and tuning — (for example, kernel/registry settings and binding/pinning interrupts of high-I/O devices)
- Application tuning — (for example, affinitizing processes/threads to local memory in a Non-Uniform Memory Access ([NUMA](#)), environment)
- Adapter slot placement — consider using multiple PCIe network adapters and localize each in PCIe slots associated with the processor socket planned to utilize that adapter. Each Intel® Xeon® processor E5-2600 product family has its own PCIe controller, so each PCIe slot is local to only one processor socket. Refer to Figure 1 for an illustration of PCIe slot/socket localization.

Figure 1 illustrates the PowerEdge R720 processor, memory, and I/O interconnects.

Figure 1. PowerEdge R720 processor, memory, and I/O interconnects



Recommendation 1: Choose an Optimal Server/Processor Architecture

Selecting a low-latency solution when purchasing your PowerEdge server is an optimum first step. Key options to keep in mind when configuring your PowerEdge server at purchase include:

- Processor frequency
- Balance of memory speed versus memory capacity
- Appropriate memory configuration for the architecture

As of May 2012, the lowest-latency server processor architecture is the Intel Xeon processor [E5-2600](#) product family. There are many processor model choices within this architecture, but the lowest-latency offering is currently the Intel Xeon processor [E5-2643](#), as it offers the highest combination of processor frequency (3.3 GHz), Intel QPI link speed (8.0 Giga-Transfers/second), and DDR3 memory speed (up to 1600 MT/s); to achieve this frequency, the E5-2643 consists of four cores, which limits its parallel capabilities but achieves high single-threaded performance. For customers who need higher core counts for parallel applications, or expanded platform availability, the six-core Intel Xeon processor [E5-2667](#), or the eight-core Intel Xeon processor [E5-2690](#) are alternatives, albeit with lower clock speeds (2.9 GHz for both).

When populating memory in a two-socket Dell server such as the [PowerEdge R620](#) or [PowerEdge R720](#), each of the four memory channels should be populated by one or two 1600MHz DIMMs for optimal memory bandwidth. Two DIMMs per channel is often the best choice for lowest memory latency, using 1600MHz Dual Rank RDIMMs. Going to three DIMMs per memory channel will reduce the memory speed and could negatively impact system latency.

Recommendation 2: Update the PowerEdge BIOS and Firmware

Continual improvements are made in the PowerEdge server BIOS and embedded server management firmware, which is the Integrated Dell Remote Access Controller (iDRAC). Dell recommends that you always check for the latest versions of BIOS and firmware by performing the following steps.

1. Navigate to <http://support.dell.com/>.
2. Select Drivers and Downloads.
3. Click Choose from a list of all Dell Products.
4. Select Servers, Storage, Networking.
5. Select PowerEdge Server.
6. Select your server product model (for example, R720 for a rack server or M1000e for a blade chassis).
7. Change the Operating System (if needed).
8. Click ESM (Embedded Server Management) and download a newer BMC/iDRAC file if the version available is newer than the version currently installed.¹
9. Click BIOS and download the file if the version posted is newer than the version currently installed.
10. Follow the installation instructions to upgrade each component. Dell recommends the system firmware be upgraded before the server BIOS.

Recommendation 3: Tune the PowerEdge BIOS for Low Latency

The defaults shipped with PowerEdge servers are optimal for many customers as a good balance between performance and power efficiency. Different workloads require optimization along different vectors; specifically, optimizing for low latency will likely have tradeoffs with vectors around performance and power efficiency. For low-latency optimization, there are some settings that can improve response times, as shown in Table 1. To access these options, enter the System Setup Program as detailed in the Using the System Setup Program in the *Hardware Owner's Manual* for your specific server model.

Changing the settings detailed in Table 1 may help in latency-sensitive workloads (as we have observed in our lab environment) and should also benefit real-time environments by suppressing System Management Interrupts (SMIs).

¹ For the PowerEdge M1000e blade chassis, the chassis management controller (CMC) firmware should also be checked. The preceding steps apply, except Chassis System Management should be clicked rather than BIOS/Embedded Server Management.

Table 1. BIOS Settings for Low Latency

System Setup screen	Setting	Default	Recommended alternative for low-latency environments
System Profile Settings	System Profile	Performance Per Watt ¹	Custom
System Profile Settings	CPU Power Management	System DBPM	Maximum Performance
System Profile Settings	Memory Frequency	Maximum Performance	Maximum Performance
System Profile Settings	Turbo Boost	Enabled	Disabled ²
System Profile Settings	C1E	Enabled	Disabled
System Profile Settings	C States	Enabled	Disabled
System Profile Settings	Monitor/Mwait	Enabled	Disabled ³
System Profile Settings	Memory Patrol Scrub	Enabled	Enabled ⁴
System Profile Settings	Memory Refresh Rate	1x	1x
Memory Settings	Memory Mode	Advanced ECC or Optimizer	Optimizer
Memory Settings	Memory Node	Disabled	Disabled
Processor Settings	Logical Processor	Enabled	Disabled
Processor Settings	Virtualization Technology	Enabled	Disabled
Processor Settings	QPI Speed	Maximum Data Rate	Maximum Data Rate
Processor Settings	Alternate RTID Setting	Disabled	Disabled
Processor Settings	Adjacent Cache Line Prefetch	Enabled	Enabled
Processor Settings	Hardware Prefetcher	Enabled	Enabled
Processor Settings	DCU Streamer Prefetcher	Enabled	Enabled
Processor Settings	DCU IP Prefetcher	Enabled	Enabled

The available BIOS options may vary, depending upon server model, processor/memory architecture, and BIOS revision. Consult your *Hardware Owner's Manual* for more details.

¹ Depends on how system was ordered. Other System Profile defaults driven by this choice, and may differ than the examples listed.

² Turbo Boost Technology 2.0 is substantially better than previous generations for latency-sensitive environments, but specific Turbo residency cannot be guaranteed under varying conditions. Evaluate Turbo Boost Technology in your own environment to choose which setting is most appropriate for your workload.

³ Monitor/Mwait should be disabled in parallel with disabling Logical Processor.

⁴ You can test your own environment to determine whether disabling Memory Patrol Scrub is helpful.

The [Dell OpenManage Deployment Toolkit \(DTK\)](#) version 4.0 and above can be used to reliably deploy optimal settings (using a script) to large numbers of Dell PowerEdge 12th generation servers without dramatically changing current deployment processes. Specifically, the DTK's system configuration utility, SYSCFG, can be used to configure server settings. These settings can also be changed while booted into a supported operating system with DTK, for details please consult the [DTK OpenManage Deployment Toolkit Version 4.0 User Guide](#).

DTK can also be used to disable **Memory Pre-Failure Notification**, which is another recommended tuning parameter to optimize for lower system latency.

Disabling the Memory Pre-Failure Notification feature will have the following impacts:

- Correctable ECC memory errors will not be reported. This does not disable correction of memory ECC errors, but only disables system logging and user notification if the correctable error threshold is exceeded.
- The Memory Operating Mode must be set to **Optimizer Mode**. Redundant Memory modes (**Mirror Mode** and **Spare Mode**) are not supported.

Memory Pre-Failure Notification can be disabled with the following command:

```
syscfg --CorrEccSmi=Disabled
```

The following command can be used to re-enable Memory Pre-Failure Notification and allow correctable ECC errors to be reported:

```
syscfg --CorrEccSmi=Enabled
```

Finally, **Platform Power Capping** should not be enabled (it is disabled by default), as it could have a negative impact on latency-sensitive environments. For more information, see the [Dell OpenManage Deployment Toolkit Version 4.0 Command Line Interface Reference Guide](#) in the appropriate version of the [DTK documentation set](#) on <http://support.dell.com/>.

Summary

Dell PowerEdge 12th generation servers are optimized from the factory with BIOS defaults that strike a good balance between performance and power efficiency for general-purpose environments. However, there are environments where you may need to optimize a server for maximum throughput or lowest latency. Taking into account the considerations detailed in this paper and by the three recommendations presented, you can considerably reduce the system latency of PowerEdge servers to provide optimal responsiveness where real-time responses are needed. The Dell OpenManage™ Deployment Toolkit can ease this process by providing you with the ability to apply needed changes programmatically.

References

Following is a list of websites referenced in this document.

1. Dell PowerEdge 12th Generation Servers: <http://www.dell.com/powerededge>
2. Standard Performance Evaluation Corporation (Specrate): <http://www.spec.org/cpu2006/>
3. SPECjbb2005 (Java Server Benchmark): <http://www.spec.org/jbb2005/>
4. VMware VMmark 2.x: <http://www.vmmark.com/>
5. Transaction Processing Performance Council (TPC): <http://www.tpc.org/>
6. Non-Uniform Memory Access: http://en.wikipedia.org/wiki/Non-Uniform_Memory_Access
7. Intel Xeon Processor E5 Family:
<http://www.intel.com/content/www/us/en/processors/xeon/xeon-processor-5000-sequence.html>
8. Intel Xeon Processor E5-2643: [http://ark.intel.com/products/64587/Intel-Xeon-Processor-E5-2643-\(10M-Cache-3_30-GHz-8_00-GTs-Intel-QPI\)](http://ark.intel.com/products/64587/Intel-Xeon-Processor-E5-2643-(10M-Cache-3_30-GHz-8_00-GTs-Intel-QPI))
9. Intel Xeon Processor E5-2667: [http://ark.intel.com/products/64589/Intel-Xeon-Processor-E5-2667-\(15M-Cache-2_90-GHz-8_00-GTs-Intel-QPI\)](http://ark.intel.com/products/64589/Intel-Xeon-Processor-E5-2667-(15M-Cache-2_90-GHz-8_00-GTs-Intel-QPI))
10. Intel Xeon Processor E5-2690: [http://ark.intel.com/products/64596/Intel-Xeon-Processor-E5-2690-\(20M-Cache-2_90-GHz-8_00-GTs-Intel-QPI\)](http://ark.intel.com/products/64596/Intel-Xeon-Processor-E5-2690-(20M-Cache-2_90-GHz-8_00-GTs-Intel-QPI))
11. 12th Generation Dell PowerEdge R620 Rack Server:
<http://www.dell.com/us/business/p/powerededge-r620/pd>
12. 12th Generation Dell PowerEdge R720 Rack Server:
<http://www.dell.com/us/business/p/powerededge-r720/pd>
13. Dell OpenManage Deployment Toolkit: <http://support.dell.com/support/edocs/software/dtk/>
14. Dell OpenManage Deployment Toolkit Version 4.0 User's Guide:
http://support.dell.com/support/edocs/software/dtk/4_0/ug/pdf/OMDTUGMR.pdf
15. Dell OpenManage Deployment Toolkit Version 4.0 Command Line Interface Reference Guide:
http://support.dell.com/support/edocs/software/dtk/4_0/cli/pdf/DTKCLIMR.pdf