Configuring Low-Latency Environments on Dell PowerEdge Servers

A Dell Technical White Paper

David J. Morse | Systems Performance Analysis

John Beckett | Systems Performance Analysis

Mukund Khatri | Server Advanced Engineering

THIS WHITE PAPER 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.

December 2010

Contents

Tables

Figures

Introduction

This white paper focuses on best practices for reducing latency within Dell™ PowerEdge™ 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 (e.g., SPECrate®, SPECweb®, 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** (i.e., transactions, 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 network driver tunings such as coalesce settings
- Operating system (OS) selection and tuning (e.g., kernel/registry settings and binding/pinning interrupts of high-I/O devices)
- Application tuning (e.g., affinitizing processes/threads to local memory in a Non-Uniform Memory Access, or NUMA, environment)

Figure 1 illustrates processor, memory, and input/output (I/O) interconnects.

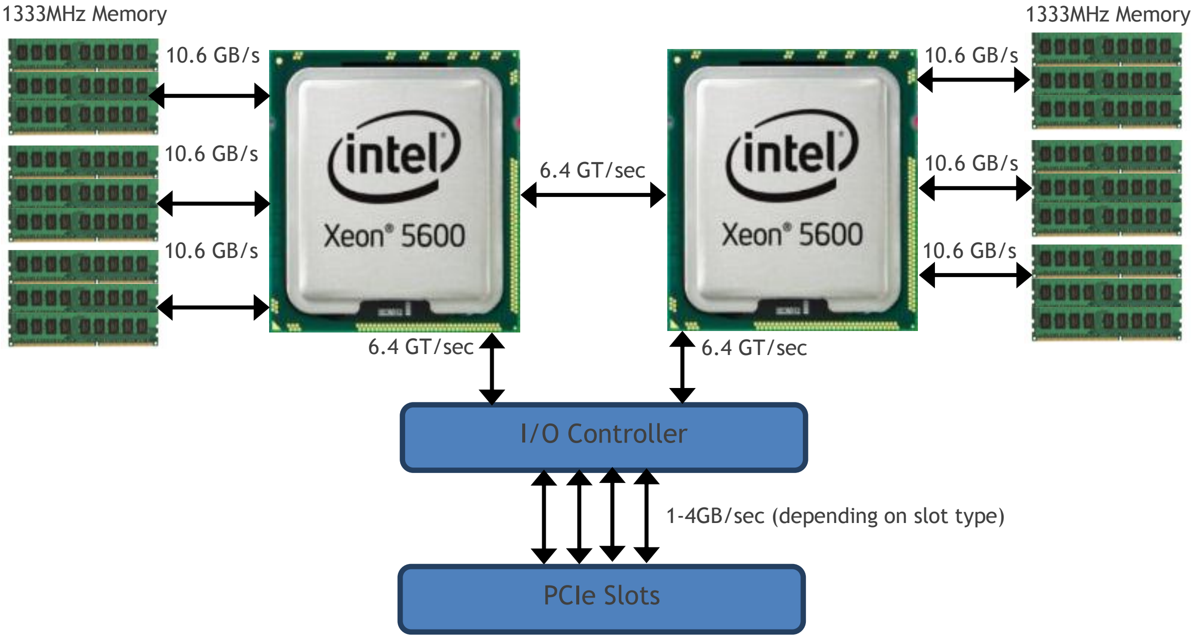

Figure 1. Illustration of Processor, Memory, and I/O Interconnects

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 December 2010, the lowest-latency server processor architecture is the Intel® Xeon® processor 5600 series. There are many processor model choices within this architecture, but the Intel Xeon processor X5677 may well be the current lowest-latency offering, as it offers the highest combination of CPU frequency (3.46 GHz, with a maximum Turbo Boost Technology frequency of 3.73 GHz), Intel QPI Link Speed (6.4 Giga-Transfers/second), and DDR3 memory speed (up to 1333 MHz). For customers who need maximum core count, the six-core Intel Xeon processor X5680 offers two additional cores per socket at a slightly lower frequency (3.33 GHz); the lower frequency may slightly increase latency.

When populating memory in a two-socket Dell server such as the PowerEdge R610 or R710, each of the three memory channels should be populated by one or two 1333MHz DIMMs for optimal memory bandwidth. Going to three DIMMs per memory channel will reduce the memory speed to 800MHz and could negatively impact system latency. For more information, see the whitepaper Memory Selection Guidelines for High Performance Computing with Dell PowerEdge 11G Servers.

Recommendation 2: Update the PowerEdge BIOS and Firmware

Continual improvements are made in the PowerEdge server BIOS and server firmware (also called BMC, Baseboard Management Controller, or iDRAC [Integrated Dell Remote Access Controller]). Dell recommends that you always check for the latest versions of BIOS and firmware as follows:

1. Navigate to http://support.dell.com/.
2. Click Start Here under the Small Businesses or Enterprise IT area.
3. Select Drivers and Downloads.
4. Click Select Model.
5. Select Servers, Storage, Networking.
6. Select PowerEdge Server.
7. Select your server product model (e.g., R710, or M1000e for a blade chassis).
8. Click Confirm.
9. Change the Operating System (if needed)
10. Click Embedded Server Management and download a newer BMC/iDRAC file if the version available is newer than the version currently installed.[1]
11. Click BIOS and download the file if the version posted is newer than the version currently installed.
12. Follow the Installation Instructions to upgrade each component. Dell recommends the system firmware be upgraded before the server BIOS.

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

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 and UEFI Boot Manager section in the *Hardware Owner's Manual* for your specific server model.

Changing the settings detailed below 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).

Table 1. BIOS Settings for Low Latency

System Setup Screen	Setting	Default	Recommended Alternative for Low-Latency Environments
Processor Settings	Logical Processor	Enabled	Disabled
Processor Settings	Turbo Mode	Enabled	Disabled[2]
Processor Settings	C-States	Enabled	Disabled
Processor Settings	C1E	Enabled	Disabled
Power Management	Power Management	Active Power Controller	Maximum Performance

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.

The Dell OpenManage™ Deployment Toolkit (DTK) can be used to reliably deploy optimal settings (using a script) to large numbers of PowerEdge 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.

DTK can also be used to disable **Memory Pre-Failure Notification** (currently only supported on the PowerEdge R610, R710, and T610), 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.

[2] You can test your own environment to determine whether Turbo Mode benefits your workload or not.

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

```
syscfg --token=0x02F7
```

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

```
syscfg --token=0x02F8
```

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 X.X Command Line Interface Reference Guide* in the appropriate version of the DTK documentation set on Support.Dell.com.

Summary

Dell PowerEdge 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 following the recommendations, you can considerably reduce the system latency of 11th-generation PowerEdge servers to provide optimal responsiveness where real-time responses are needed. The OpenManage Deployment Toolkit can ease this process by providing you with the ability to apply needed changes programmatically.