



## Configuring PowerEdge Servers for Low-Latency Environments

Tech Note by:  
Tad Walsh

### SUMMARY

While Dell EMC PowerEdge 14<sup>th</sup>-generation servers are shipped from the factory with settings optimized for maximum throughput, optimum power efficiency, etc., some workloads and applications set higher priority on lowest latency and real-time responsiveness. PowerEdge server settings can be modified to achieve these criteria.

This Direct from Development tech note summarizes best-practice modifications to settings that can minimize system latency and provide optimal jitter-free predictable performance and processing power.

This tech note is derived from a full-length white paper, *Configuring Low-Latency Environments on Dell 14<sup>th</sup>-Generation PowerEdge Servers*, available to customers under NDA terms from their Dell EMC Sales Representative. Please contact your Sales Representative to receive a copy

A key characteristic and benefit of Dell EMC PowerEdge servers is the ability to configure, tune and optimize them for specific workloads and applications. While a large number of PowerEdge servers are installed for general-purpose computing and are tuned for high, balanced performance delivering maximum throughput, many PowerEdge servers are customer-tuned for achievement along a specific dimension. A good example of this is tuning to achieve lowest-latency, fastest response time for real-time or near real-time performance. Applications where this is a key priority include High Frequency Trading (HFT), chemical process control, manufacturing process control, and signal processing. In these application areas, users will trade off e.g. highest overall throughput, or maximum power efficiency, to achieve the primary goal of low latency and fastest response time.

A full length white paper, *Configuring Low-Latency Environments on Dell 14<sup>th</sup> Generation PowerEdge Servers*, is available to customers through their Dell EMC sales representatives, under Non-Disclosure Agreement (NDA) terms. That NDA paper provides detailed recommendations for tuning PowerEdge servers for low latency. The information below in this Tech Note summarizes that full-length NDA white paper and presents key considerations for achieving lowest latency.

It is important to note that the low latency guidelines provided in the white paper are not to be applied “as-is” by general-computing customers desiring enhanced system performance. In some cases, the low-latency recommendations may not deliver the best possible performance for some computing scenarios. Follow the specific recommendations of the referenced NDA white paper *only* where your environment is primarily latency-sensitive, such as the application areas mentioned above.

### Reducing system latency

With today's multi-socket, multi-core, highly threaded CPUs in 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® 2017, SPECjbb® 2015, VMware® VMmark 2™ 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, number of transactions, input/output (I/O) or pages per second). However, for certain applications (e.g. High Frequency Trading and others mentioned above), users care more about reducing the time it takes to complete 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. More specifically, the requirement is for achieving predictable (i.e. “jitter-free”) and lowest possible system latency. Network latency improvements are also paramount to achieve overall system latency improvements, so tuning for these environments is similarly important.

## Consider the entire solution

To reduce system latency, the entire solution must be taken into consideration:

- **The server** — including processor, memory, I/O sub-system 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 performance I/O devices
- **Application tuning** — for example, affinizing processes/threads to local memory in a Non-Uniform Memory Access (NUMA) environment
- **Adapter slot placement** — consider using localized PCIe network adapters in PCIe slots associated with the processor socket planned to utilize that adapter.

## Guideline 1: Choose an Optimal Server/Processor Architecture

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

- PowerEdge server model
- Processor type, core count and frequency
- Balance of memory speed vs. memory capacity
- Appropriate memory configuration for the architecture
- PCIe slot architecture

The full-length NDA white paper previously referenced gives detailed recommendations for each of the criteria bulleted above. The point of listing them here is to give a taste of some of the key considerations about server configuration. For example, of the many flavors, or “bins”, in the Intel Processor Scalable Family, certain offerings are more appropriate for low latency applications than others. The full-length NDA white paper includes a table that identifies the leading candidates for lowest latency applications.

## Guideline 2: Update PowerEdge BIOS and Firmware

Continual improvements are made in the PowerEdge server BIOS and the embedded server management firmware which is onboard the Integrated Dell Remote Access Controller (iDRAC9). Dell recommends that you regularly check for the latest versions of BIOS and firmware. This is particularly relevant prior to tuning and optimizing your server, including tuning for low latency. The full-length NDA white paper provides step-by-step instructions on how to check for updated versions of BIOS and firmware, and how to download and install.

## Guideline 3: Tune PowerEdge BIOS for Low Latency

The default settings shipped with PowerEdge servers are optimal for many workloads as a good balance between performance and power efficiency. However, different workloads require optimization along different vectors: Specifically, optimizing for low latency will likely have tradeoffs with vectors around e.g. throughput, and power efficiency. BIOS settings for low latency, detailed in a table in the NDA white paper, can help users tune latency-sensitive workloads, including suppressing any/all sources of system management interrupts (SMIs).

## Guideline 4: Evaluate the Sub-NUMA Clustering Option

PowerEdge 14<sup>th</sup>-generation (14G) servers using Intel Xeon Scalable processors include a new architectural option called Sub-NUMA Clustering (SNC) for higher core count CPUs. By enabling Sub-NUMA Clustering, cores and memory controllers are split between two NUMA nodes per socket. Experimentation with Sub-NUMA Clustering can reveal if advantages can be gained for your highly NUMA-affinitized workload. Note that Dell EMC PowerEdge core enumeration patterns differ from competitor systems in Linux environments. Thus, when evaluating Sub-NUMA Clustering, and particularly when considering core affinities/thread pinning, it is recommended that users refer to the Core Enumeration tables in the NDA white paper.



### Guideline 5: Dell Controlled Turbo

**Dell Controlled Turbo** is the BIOS setup option for controlling **Dell Processor Acceleration Technology (DPAT)**, an innovative, patented Dell technology designed to deliver predictable, jitter-free Turbo operation. DPAT is available to all Skylake processor bins that support Turbo Mode. This feature is primarily targeted for low latency applications with sensitivity to the latency jitter induced by turbo transitions in Intel architectures. Dell highly recommends that you enable this feature for minimizing jitter with turbo-enabled configurations for your specific low-latency workloads. Figure 1 below illustrates the issues that can arise with uncontrolled Turbo mode and how the **Dell Controlled Turbo** feature in DPAT resolves these issues.

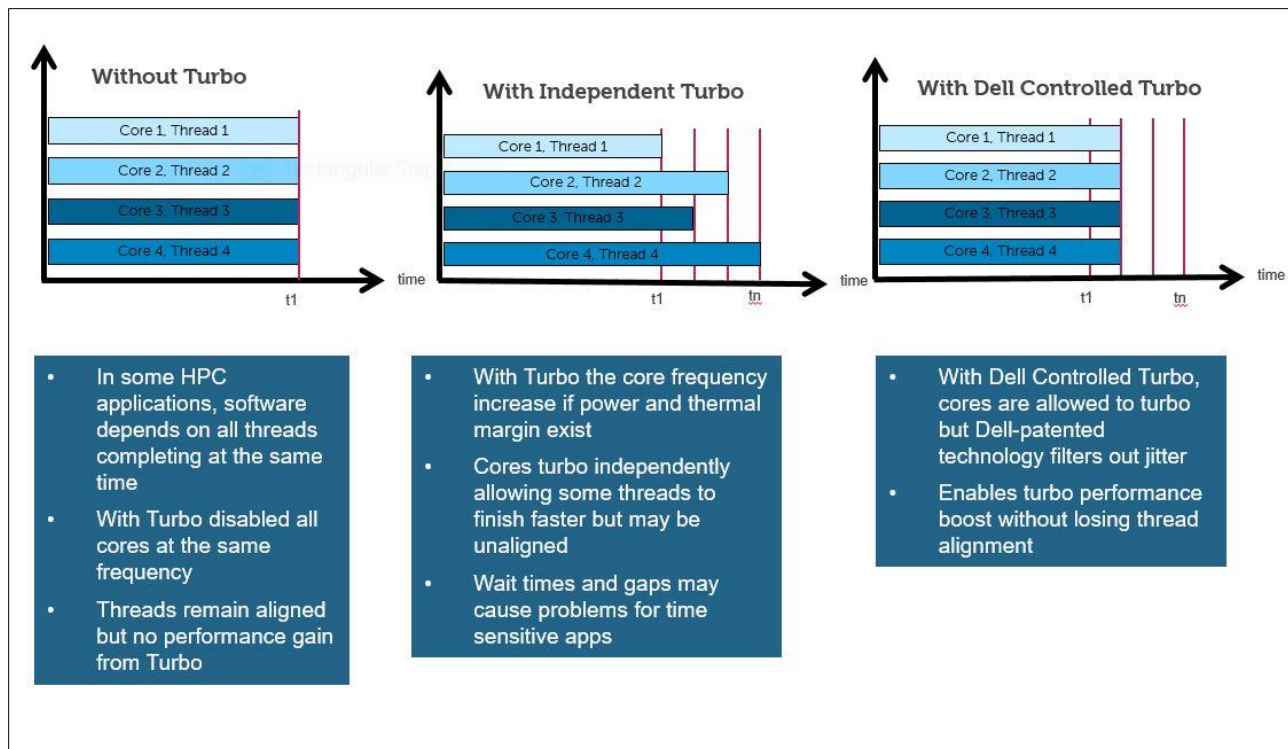


Figure 1: Dell Controlled Turbo enables turbo boost without losing thread alignment, overcoming problems for time-sensitive applications caused by wait times and gaps.

### Guideline 6: Linux Tuning Parameters

The *intel\_idle* driver present in most recent Enterprise-class Linux distributions is capable of enforcing some C-states, even if BIOS C-states/C1E options are disabled. To optimize the system for low latency workloads, an option identified in the full-length NDA white paper should be appended to the kernel boot line. Adding this line to the kernel boot options and rebooting the system will prevent the *intel\_idle* driver from loading, and should ensure that no C-states will be present in the system that will drive unwanted latency characteristics.

The referenced white paper also suggests *pstate* and *balancing* parameters to disable unwanted power savings and NUMA balancing effects, respectively, in more recent Linux distributions.

### Guideline 7: Thermal Configuration

The guidelines described above will help to lower latency and drive faster response times for critical applications. When implemented, some of these changes could also result in greater heat being produced by the system. Higher temperatures are a natural result of driving a server faster and harder. PowerEdge servers are designed for high

thermal efficiency, thus if implementing the guidelines above results in thermal challenges, a straightforward fix is readily available and is detailed in the NDA white paper: An option exists in the iDRAC settings to increase fan speed to more aggressively cool the system. This does so at a higher cost to power consumption and perhaps to acoustics (fan noise), but again, this is a natural trade-off to achieve the primary criteria of attaining lower latency.

#### Guideline 8: Quick Latency Tuning Option

Using the Dell OpenManage RACADM tool from the Dell OpenManage Deployment Toolkit (DTK), customers can now set their system BIOS to conform to Dell's recommended Low Latency BIOS options *with a single command*. Once the OpenManage RACADM tool has been installed on the system, run this command to set your BIOS options to conform to Dell recommended settings for Low Latency:

**`racadm set bios.sysprofilesettings.WorkloadProfile LowLatencyOptimizedProfile`**

BIOS options may be further modified to tune for your specific environment, if necessary.

#### Summary

Dell's 14<sup>th</sup>-generation of PowerEdge servers are optimized from the factory with BIOS defaults that strike an excellent balance between performance and power efficiency for general-purpose environments. There are, however, environments where the server may need to be tweaked and optimized for other criteria, e.g. maximum throughput, or optimal power efficiency, or lowest latency. Taking into account the considerations detailed in this paper and the eight tuning guidelines presented, PowerEdge server system latency can be significantly reduced, to provide optimal responsiveness where real-time responses are needed.

While users can implement the guidelines above manually, the Dell OpenManage™ Deployment Toolkit (DTK) can streamline this process by providing users the ability to apply needed changes programmatically, with a single command.

#### Notes:

1. To receive the full-length white paper **Configuring Low-Latency Environments on Dell 14th Generation PowerEdge Servers** for use under Non-Disclosure Agreement (NDA), contact your Dell EMC sales representative.
2. Users interested in thermal configuration for low-latency workloads can download the Direct from Development tech note **PowerEdge Multi-Vector Cooling**, available on Dell EMC TechCenter at [http://en.community.dell.com/techcenter/extras/m/white\\_papers/20444244](http://en.community.dell.com/techcenter/extras/m/white_papers/20444244). This 3-page tech note describes Dell EMC's multi-faceted approach to server thermal control, which addresses and resolves not only cooling but also power consumption, airflow consumption and acoustics.