

# Performance Tuning Best Practices Guide for Cisco UCS M7 Platforms

Using 4<sup>th</sup> Gen Intel Xeon Scalable Processors

---

# Contents

Purpose and scope	3
What you will learn	3
BIOS tuning scenarios	3
<b>Tuning for general-purpose workloads</b>	<b>3</b>
<b>Tuning for enterprise workloads</b>	<b>4</b>
Cisco UCS BIOS options	4
<b>Processor settings</b>	<b>4</b>
<b>Memory settings</b>	<b>10</b>
<b>Power and performance configuration</b>	<b>11</b>
<b>Fan Control Policy (CIMC/IMM)</b>	<b>13</b>
<b>BIOS settings for Cisco UCS M7 servers</b>	<b>15</b>
<b>BIOS recommendations for various general-purpose workloads</b>	<b>16</b>
<b>Additional BIOS recommendations for enterprise workloads</b>	<b>19</b>
Conclusion	23
For more information	23

---

## Purpose and scope

The Basic Input-and-Output System (BIOS) tests and initializes the hardware components of a system and boots the operating system from a storage device. A typical computational system has several BIOS settings that control the system's behavior. Some of these settings are directly related to the performance of the system.

This document explains the BIOS settings that are valid for the Cisco Unified Computing System™ (Cisco UCS®) M7 server generation of the following servers: the Cisco UCS C220 M7 Rack Server, Cisco UCS C240 M7 Rack Server, Cisco UCS X210c M7 Compute Node, and Cisco UCS X410c M7 Compute Node. All servers use fourth-generation (4<sup>th</sup> Gen) Intel® Xeon® Scalable Processors. This document describes how to optimize the BIOS settings to meet requirements for general-purpose workloads that are optimal for best performance and energy efficiency for the Cisco UCS M7 generation of blade and rack servers. This document also describes the BIOS recommendations for industry-standard enterprise workloads.

With the release of the 4<sup>th</sup> Gen Intel Xeon Scalable Processor family (architecture code-named Sapphire Rapids), Cisco released seventh-generation UCS servers to take advantage of the increased number of cores, higher memory speeds, and PCIe Gen 5.0 features of the new processors, thus benefiting CPU-, memory-, and I/O-intensive workloads.

Understanding the BIOS options will help you select appropriate values to achieve optimal system performance. This document does not discuss the BIOS options for specific firmware releases of Cisco UCS M7 servers. The settings demonstrated here are generic.

## What you will learn

The process of setting performance options in your system BIOS can be daunting and confusing, and some of the options you can choose are obscure. For most options, you must choose between optimizing a server for power savings or for performance. This document provides some general guidelines and suggestions to help you achieve optimal performance from your Cisco UCS blade and rack M7 servers that use 4<sup>th</sup> Gen Intel Xeon Scalable Processor family CPUs.

## BIOS tuning scenarios

This document focuses on two main scenarios: how to tune the BIOS for general-purpose workloads and how to tune the BIOS for enterprise workloads.

### **Tuning for general-purpose workloads**

With the latest multiprocessor, multicore, and multithreading technologies in conjunction with current operating systems and applications, the new Cisco UCS M7 servers based on the 4<sup>th</sup> Gen Intel Xeon Scalable Processor family deliver the highest levels of performance, as demonstrated in numerous industry-standard benchmark publications.

Cisco UCS servers with standard settings already provide an optimal ratio of performance to energy efficiency. However, through BIOS settings you can further optimize the system with higher performance and less energy efficiency. Basically, this optimization operates all the components in the system at the maximum speed possible and prevents the energy-saving options from slowing down the system. In general, optimization to achieve greater performance is associated with increased consumption of electrical power. This document explains how to configure the BIOS settings to achieve optimal computing performance.

---

## Tuning for enterprise workloads

With the evolution of computer architecture, performance has reached results that were unimaginable a few years ago. However, the complexity of modern computer architectures requires end users and developers to know how to write code. It also requires them to know how to configure and deploy software for a specific architecture to get the most out of it.

Performance tuning is difficult and general recommendations are problematic. This document tries to provide insights into optimal BIOS settings and OS tunings that have an impact on overall system performance. This document does not provide generic rule-of-thumb (or values) to be used for performance tuning. The finest tuning of the parameters described requires a thorough understanding of the enterprise workloads and the Cisco UCS platform on which they run.

### Cisco UCS BIOS options

This section describes the options you can configure in the Cisco UCS BIOS.

#### Processor settings

This section describes processor options you can configure.

##### Intel Hyper-Threading Technology

You can specify whether the processor uses Intel Hyper-Threading Technology, which allows multithreaded software applications to process threads in parallel within each processor. Multithreading is a form of parallelization or dividing up work for simultaneous processing. Instead of giving a large workload to a single core, threaded programs split the work into multiple software threads. These threads are processed in parallel by different CPU cores to save time. You should test the CPU hyperthreading option both enabled and disabled in your specific environment. If you are running a single-threaded application, you should disable hyperthreading.

The setting can be either of the following:

- Disabled: the processor does not permit hyperthreading.
- Enabled: the processor allows parallel processing of multiple threads.
- Platform default: the BIOS uses the value for this attribute contained in the BIOS defaults for the server type and vendor.

By default, this BIOS setting is Enabled.

##### Intel SpeedStep Technology

Intel SpeedStep Technology is designed to save energy by adjusting the CPU clock frequency up or down depending on how busy the system is. Intel Turbo Boost Technology provides the capability for the CPU to adjust itself to run higher than its stated clock speed if it has enough power to do so.

You can specify whether the processor uses Enhanced Intel SpeedStep Technology, which allows the system to dynamically adjust processor voltage and core frequency. This technology can result in decreased average power consumption and decreased average heat production.

The setting can be either of the following:

- Disabled: the processor never dynamically adjusts its voltage or frequency.
- Enabled: the processor uses Intel SpeedStep Technology and enables all supported processor sleep states to further conserve power.
- Platform default: the BIOS uses the value for this attribute contained in the BIOS defaults for the server type and vendor.

By default, this BIOS setting is Enabled.

### Processor prefetchers

Intel Xeon processors have several layers of cache. Each core has a tiny Layer-1 cache, sometimes referred to as the Data Cache Unit (DCU), that has 32 KB for instructions and 32 KB for data. Slightly bigger is the Layer-2 cache, with 256 KB shared between data and instructions for each core. In addition, all cores on a chip share a much larger Layer-3 cache, which is about 10 to 45 MB in size (depending on the processor model and number of cores).

The prefetcher settings provided by Intel primarily affect the Layer-1 and Layer-2 caches on a processor core (Table 1). You will likely need to perform some testing with your individual workload to find the combination that works best for you. Testing on the Intel Xeon Scalable Processor has shown that most applications run best with all prefetchers enabled. See tables 2 and 3 for guidance.

**Table 1.** Processor prefetcher options

Processor prefetcher options	Cache affected
Hardware prefetcher	Layer 2
Adjacent-cache-line prefetcher	Layer 2
DCU prefetcher	Layer 1
DCU instruction pointer (DCU-IP) prefetcher	Layer 1

### Hardware prefetcher

The hardware prefetcher prefetches additional streams of instructions and data into the Layer-2 cache upon detection of an access stride. This behavior is more likely to occur during operations that sort sequential data, such as database table scans and clustered index scans, or that run a tight loop in code.

You can specify whether the processor allows the Intel hardware prefetcher to fetch streams of data and instructions from memory into the unified second-level cache when necessary.

The setting can be either of the following:

- Disabled: the hardware prefetcher is not used.
- Enabled: the processor uses the hardware prefetcher when cache problems are detected.

By default, this BIOS setting is Enabled.

---

### **Adjacent-cache-line prefetcher**

The adjacent-cache-line prefetcher always prefetches the next cache line. Although this approach works well when data is accessed sequentially in memory, it can quickly litter the small Layer-2 cache with unneeded instructions and data if the system is not accessing data sequentially, causing frequently accessed instructions and code to leave the cache to make room for the adjacent-line data or instructions.

You can specify whether the processor fetches cache lines in even or odd pairs instead of fetching just the required line.

The setting can be either of the following:

- Disabled: the processor fetches only the required line.
- Enabled: the processor fetches both the required line and its paired line.

By default, this BIOS setting is Enabled.

### **Data cache unit streamer prefetcher**

Like the hardware prefetcher, the Data Cache Unit (DCU) streamer prefetcher prefetches additional streams of instructions or data upon detection of an access stride; however, it stores the streams in the tiny Layer-1 cache instead of the Layer-2 cache.

This prefetcher is a Layer-1 data cache prefetcher. It detects multiple loads from the same cache line that occur within a time limit. Making the assumption that the next cache line is also required, the prefetcher loads the next line in advance to the Layer-1 cache from the Layer-2 cache or the main memory.

The setting can be either of the following:

- Disabled: the processor does not try to anticipate cache-read requirements and fetches only explicitly requested lines.
- Enabled: the DCU prefetcher analyzes the cache-read pattern and prefetches the next line in the cache if it determines that it may be needed.

By default, this BIOS setting is Enabled.

### **Data cache unit-IP prefetcher**

The DCU-IP prefetcher predictably prefetches data into the Layer-1 cache on the basis of the recent instruction pointer load instruction history.

You can specify whether the processor uses the DCU-IP prefetch mechanism to analyze historical cache-access patterns and preload the most relevant lines in the Layer-1 cache.

The setting can be either of the following:

- Disabled: the processor does not preload any cache data.
- Enabled: the DCU-IP prefetcher preloads the Layer-1 cache with the data it determines to be the most relevant.

By default, this BIOS setting is Enabled.

---

### **Last-level cache prefetch**

This BIOS option configures the processor's last-level cache (LLC) prefetch feature as a result of the noninclusive cache architecture. The LLC prefetcher exists on top of other prefetchers that can prefetch data into the core DCU and midlevel cache (MLC). In some cases, disabling this option can improve performance.

The setting for this BIOS option can be either of the following:

- Disabled: the LLC prefetcher is disabled. The other core prefetchers are not affected.
- Enabled: the core prefetcher can prefetch data directly to the LLC.

By default, this BIOS setting is Enabled.

### **Intel Virtualization Technology**

Virtualization abstracts hardware that allows multiple workloads to share a common set of resources. On shared virtualized hardware, a variety of workloads can colocate while maintaining full isolation from each other, freely migrate across infrastructures, and scale as needed. Intel Virtualization Technology allows a platform to run multiple operating systems and applications in independent partitions.

Intel Virtualization Technology (Intel VT) represents a growing portfolio of technologies and features that make virtualization practical by eliminating performance overheads and improving security. Intel VT provides hardware assistance to the virtualization software, reducing the size, cost, and complexity of the software. Special attention is also given to reduce the virtualization overheads occurring in cache, I/O, and memory.

By default, this BIOS setting is Enabled.

### **Intel VT for Directed I/O**

You can specify whether the processor uses Intel Virtualization Technology (VT) for Directed I/O (VT-d), which allows a platform to run multiple operating systems and applications in independent partitions.

The setting can be either of the following:

- Disabled: the processor does not permit virtualization.
- Enabled: the processor allows multiple operating systems in independent partitions.

**Note:** If you change this option, you must power the server off and on before the setting takes effect.

By default, this BIOS setting is Enabled.

### **Intel Ultra Path Interconnect link enablement**

The Intel Ultra Path Interconnect (UPI) BIOS option allows you to change the number of UPI links. Use this option to configure the UPI topology to use fewer links between processors, when available. Changing this option from the default can reduce UPI bandwidth performance in exchange for less power consumption.

The values for this BIOS setting are 1, 2, and Auto.

By default, this BIOS setting is Auto.

---

## Intel UPI power management

Intel UPI power management is used to conserve power on a platform. Low-power mode reduces UPI frequency and bandwidth. This option is recommended to save power; however, UPI power management is not recommended for high-frequency, low-latency, virtualization, and database workloads.

This BIOS option controls the link L0p enable and link L1 enable values.

L1 saves the most power but has the greatest impact on latency and bandwidth. L1 allows a UPI link to transition from the full-link-down state. L1 is the deepest power-savings state.

L0p allows a partial-link-down state. A subset of all of the lanes will remain awake.

By default, this BIOS setting is Disabled.

## UPI link speed

The UPI link speed BIOS option allows you to set the UPI link speed. Running the UPI link speed (frequency) at a lower rate can reduce power consumption, but it can also affect system performance.

UPI link frequency determines the rate at which the UPI processor interconnect link operates. If a workload is highly Nonuniform Memory Access (NUMA) aware, sometimes lowering the UPI link frequency can free more power for the cores and result in better overall performance.

By default, this BIOS setting is set to Auto.

## Sub-NUMA clustering

The sub-NUMA clustering (SNC) BIOS option provides localization benefits similar to the Cluster-on-Die (CoD) option, without some of the disadvantages of CoD. SNC is a replacement for the CoD feature found in previous processor families. SNC (two-way sub-NUMA) divides the LLC into two disjointed clusters called NUMA nodes, and it is based on address range, with each cluster bound to a subset of the memory controllers in the system. SNC improves average latency to the LLC and memory. For a multi-socket system, all SNC clusters are mapped to unique NUMA domains. Integrated memory controller interleaving must be set to the correct value to correspond with the SNC setting. OS support that recognizes each cluster and a separate NUMA node is necessary to take advantage of SNC.

The setting for this BIOS option can be either of the following:

- Disabled: the LLC is treated as one cluster when this option is disabled.
- Enabled: the LLC capacity is used more efficiently, and latency is reduced as a result of the core and integrated memory controller proximity. This setting may improve performance on NUMA-aware operating systems.
- Enable SNC2(2-clusters): SNC2 partitions LLC into two NUMA domains containing an equal number of cores, an equal number of LLC slices, and an equal amount of socket address space, and each node is bound to a subset of the memory controller on the socket.
- Enable SNC4(4-clusters): SNC4 is an extension of SNC2, where it divides the die into four NUMA domains and divides the memory range equally between the domains. Both SNC2 and SNC4 trade peak IO bandwidth for the best core-to-memory latencies.

**Note:** When SNC is selected, the operating system discovers each physical CPU socket as two NUMA nodes, except for 4<sup>th</sup> Gen Intel Xeon Scalable Processors with fewer than 12 cores, for which SNC is not supported. Refer to your OS documentation to determine whether SNC is supported.



---

By default, this BIOS setting is Disabled.

#### **Extended prediction table prefetch**

Extended prediction table (XPT) prefetch is a new capability that is designed to reduce local memory access latency. This prefetcher exists on top of other prefetchers that can prefetch data in the core DCU, MLC, and LLC. The XPT prefetcher will issue a speculative DRAM read request in parallel with an LLC lookup. This prefetch bypasses the LLC, reducing latency. You can specify whether the processor uses the XPT prefetch mechanism to fetch the data into the XPT.

The setting can be either of the following:

- Disabled: the processor does not preload any cache data.
- Enabled: the XPT prefetcher preloads the Layer-1 cache with the data it determines to be the most relevant.

By default, this BIOS setting is set to Auto.

#### **Intel UPI prefetch**

Intel UPI prefetch is a new capability that is designed to reduce remote memory access latency. The UPI controller issues a UPI prefetch command, also in parallel with an LLC lookup, to the memory controller when a remote read request arrives in the home socket.

By default, this BIOS setting is set to Auto.

#### **XPT remote prefetch**

The XPT remote prefetch (extended prediction table) BIOS option configures the XPT remote prefetcher processor performance option. When it is enabled, this feature can improve remote read request latency from a processor core by directly accessing the UPI. Values for this BIOS setting can be auto, enabled, or disabled.

By default, this BIOS setting is set to Auto.

#### **Last-level cache dead line**

With the Intel Xeon Scalable Processors' non-inclusive cache scheme, MLC evictions are filled into the LLC if the data is shared across processor cores. When cache lines are evicted from the MLC, the processor core can flag them as "dead," meaning that they are not likely to be read again. With this option, the LLC can be configured to drop dead lines and not fill them in the LLC.

Values for the LLC dead line BIOS option can be either of the following:

- Disabled: if this option is disabled, dead lines will be dropped from the LLC. This setting provides better utilization in the LLC and prevents the LLC from evicting useful data.
- Enabled: if this option is enabled, the processor determines whether to keep or drop dead lines. By default, this option is enabled.

By default, this BIOS setting is Enabled

---

## Memory settings

You can use several settings to optimize memory performance.

### Nonuniform Memory Access (NUMA)

Most modern operating systems, particularly virtualization hypervisors, support NUMA because in the latest server designs a processor is attached to a memory controller: therefore, half the memory belongs to one processor, and half belongs to the other processor. If a core needs to access memory that resides in another processor, a longer latency period is needed to access that part of memory. Operating systems and hypervisors recognize this architecture and are designed to reduce such trips. For hypervisors such as those from VMware and for modern applications designed for NUMA, keep this option enabled.

By default, this BIOS setting is Enabled

### Virtual NUMA

When virtual NUMA is enabled, two NUMA nodes are created per physical CPU socket without changing memory-controller and channel interleaving and LLC grouping. Virtual NUMA mode provides a potential memory bandwidth advantage. The latency between these two virtual NUMA nodes is identical to its local latency. The BIOS options are enabled and disabled. By default, this option is disabled.

By default, this BIOS setting is Enabled.

### Memory refresh rate

This BIOS option controls the refresh rate of the memory controller and may affect the performance and resiliency of the server memory. This option sets the memory refresh rate to either 1x refresh or 2x refresh. By default, 2X refresh is enabled.

By default, this BIOS setting is set to 1X refresh.

### Adaptive double device data correction sparing

Adaptive double device data correction (ADDDC) is a memory RAS feature that enables dynamic mapping of failing DRAM by monitoring corrected errors and taking action before uncorrected errors can occur and cause an outage. It is now enabled by default.

After ADDDC sparing remaps a memory region, the system could incur marginal memory latency and bandwidth penalties on memory-bandwidth-intensive workloads that target the affected region. Cisco recommends scheduling proactive maintenance to replace a failed DIMM after an ADDDC RAS fault is reported.

**Note:** For the optimal balance of performance and system stability, you should use the platform default. ADDDC sparing will incur a small performance penalty for memory-intensive workloads.

**Note:** By default, this BIOS setting is Enabled.

---

## Patrol scrub

You can specify whether the system actively searches for, and corrects, single-bit memory errors even in unused portions of the memory on the server.

The setting can be either of the following:

- **Disabled:** the system checks for memory Error-Correcting Code (ECC) errors only when the CPU reads or writes a memory address.
- **Enabled:** the system periodically reads and writes memory, searching for ECC errors. If any errors are found, the system attempts to fix them. This option may correct single-bit errors before they become multiple-bit errors, but it may adversely affect performance when the patrol-scrub process is running.

By default, this BIOS setting is Enabled.

## Power and performance configuration

### Enhanced CPU performance

This BIOS option helps users modify the enhanced CPU performance settings. When it is enabled, this option adjusts the processor settings and enables the processor to run aggressively, which can improve overall CPU performance, but may result in higher power consumption. Values for this BIOS option can be Auto or Disabled. By default, the enhanced CPU performance option is disabled.

**Note:** This BIOS feature is applicable for benchmark purposes only and is not recommended for any production workloads. When this option is enabled, we highly recommend setting the fan policy at maximum power.

By default, this BIOS setting is Disabled.

### Energy-efficient turbo mode

The energy-efficient turbo mode BIOS option allows you to control whether the processor uses an energy-efficiency-based policy. In this operation mode, a processor's core frequency is adjusted within the turbo-mode range based on workload.

When energy-efficient turbo is enabled, the CPU's optimal turbo frequency will be tuned dynamically based on CPU utilization. The power performance bias setting also influences energy-efficient turbo.

By default, this BIOS option is disabled.

### Intel Turbo Boost Technology

Intel Turbo Boost Technology depends on Intel SpeedStep: if you want to enable Intel Turbo Boost, you must enable Intel SpeedStep first. If you disable Intel SpeedStep, you lose the capability to use Intel Turbo Boost.

Intel Turbo Boost is especially useful for latency-sensitive applications and for scenarios in which the system is nearing saturation and would benefit from a temporary increase in the CPU speed. If your system is not running at this saturation level and you want the best performance at a utilization rate of less than 90 percent, you should disable Intel SpeedStep to help ensure that the system is running at its stated clock speed at all times.

By default, this BIOS setting is Enabled.

---

## Processor C6 report

The C6 state is a power-saving halt and sleep state that a CPU can enter when it is not busy. Unfortunately, it can take some time for the CPU to leave these states and return to a running condition. If you are concerned about performance (for all but latency-sensitive single-threaded applications), and if you can do so, disable anything related to C-states.

You can specify whether the BIOS sends the C6 report to the operating system. When the OS receives the report, it can transition the processor into the lower C6 power state to decrease energy use while maintaining optimal processor performance.

The setting can be either of the following:

- Disabled: the BIOS does not send the C6 report.
- Enabled: the BIOS sends the C6 report, allowing the OS to transition the processor to the C6 low-power state.

By default, this BIOS setting is Disabled.

## Processor C1E

Enabling the C1E option allows the processor to transition to its minimum frequency upon entering the C1 state. This setting does not take effect until after you have rebooted the server. When this option is disabled, the CPU continues to run at its maximum frequency in the C1 state. Users should disable this option to perform application benchmarking.

You can specify whether the CPU transitions to its minimum frequency when entering the C1 state.

The setting can be either of the following:

- Disabled: the CPU continues to run at its maximum frequency in the C1 state.
- Enabled: the CPU transitions to its minimum frequency. This option saves the maximum amount of power in the C1 state.

By default, this BIOS setting is Disabled.

## Package C-state control

When the power technology is set to Custom, use this option to configure the lowest processor idle power state (C-state). The processor automatically transitions into package C-states based on the core C-states to which cores on the processor have transitioned. The higher the package C-state, the lower the power use of that idle package state. The default setting, Package C6 (nonretention), is the lowest power idle package state supported by the processor.

You can specify the amount of power available to the server components when they are idle.

The possible settings are as follows:

- C0/C1 State: when the CPU is idle, the system slightly reduces power consumption. This option requires less power than C0 and allows the server to return quickly to high-performance mode.
- C2 State: when the CPU is idle, the system reduces power consumption more than with the C1 option. This option requires less power than C1 or C0, but the server takes slightly longer to return to high-performance mode.

- 
- C6 Nonretention: when the CPU is idle, the system reduces power consumption more than with the C3 option. This option saves more power than C0, C1, or C3, but the system may experience performance problems until the server returns to full power.
  - C6 Retention: when the CPU is idle, the system reduces power consumption more than with the C3 option. This option consumes slightly more power than the C6 Nonretention option, because the processor is operating at Pn voltage to reduce the package's C-state exit latency.

By default, this BIOS setting is set to C0/C1 State.

### **Energy performance tuning**

This BIOS option determines how aggressively the CPU is power-managed and placed into turbo mode. If you select BIOS Control, the system controls the setting. If you select OS Control, the operating system controls the setting. By default, OS Control is enabled.

By default, this BIOS setting is set to OS Control.

### **Workload configuration**

You can tune the system's I/O bandwidth between balanced and I/O sensitive by adjusting the processor's core and uncore frequencies. This configuration allows users to set a parameter to optimize workload characterization.

This setting can be either of the following:

- Balanced: the balanced setting is used for optimization.
- I/O Sensitive: the I/O-sensitive setting is used for optimization. By default, I/O Sensitive is enabled.

By default, this BIOS setting is set to I/O Sensitive.

### **Fan Control Policy (CIMC/IMM)**

Fan policy enables you to control the fan speed to reduce server power-consumption and noise levels. Prior to the introduction of fan policy, fan speed increased automatically when the temperature of any server component exceeded the set threshold. To help ensure that fan speeds were low, the threshold temperatures of components were usually set to high values. Although this behavior suited most server configurations, it did not address the following situations:

- Maximum CPU performance: for high performance, certain CPUs must be cooled substantially below the set threshold temperature. This cooling requires very high fan speeds, which results in increased power consumption and noise levels.
- Low power consumption: to help ensure the lowest power consumption, fans must run very slowly and, in some cases, stop completely on servers that allow this behavior. But slow fan speeds can cause servers to overheat. To avoid this situation, you need to run fans at a speed that is moderately faster than the lowest possible speed.

---

You can choose the following fan policies:

- **Balanced:** this is the default policy. This setting can cool almost any server configuration, but it may not be suitable for servers with PCI Express (PCIe) cards, because these cards overheat easily.
- **Low Power:** this setting is well suited for minimal-configuration servers that do not contain any PCIe cards.
- **High Power:** this setting can be used for server configurations that require fan speeds ranging from 60 to 85 percent. This policy is well suited for servers that contain PCIe cards that easily overheat and have high temperatures. The minimum fan speed set with this policy varies for each server platform, but it is approximately in the range of 60 to 85 percent.
- **Maximum Power:** this setting can be used for server configurations that require extremely high fan speeds ranging between 70 and 100 percent. This policy is well suited for servers that contain PCIe cards that easily overheat and have extremely high temperatures. The minimum fan speed set with this policy varies for each server platform, but it is approximately in the range of 70 to 100 percent.
- **Acoustic:** the fan speed is reduced to reduce noise levels in acoustic-sensitive environments. Rather than regulating energy consumption and preventing component throttling as in other modes, the Acoustic option could result in short-term throttling to achieve a lowered noise level. Applying this fan control policy might result in short-duration and transient performance impacts.

By default, this FAN control policy is set to Acoustic.

To configure FAN policy:

- To configure FAN policy on CIMC, Click on Compute ▢ Power policy ▢ Configure Fan policy ▢ Fan Policy.
- To configure FAN policy on IMM, create Thermal policy and set the Fan Control Mode of the System.

**Note:** High Power, Maximum Power and Acoustic modes are only supported for Cisco UCS X series Chassis.

## BIOS settings for Cisco UCS M7 servers

Table 2 lists the BIOS token names, defaults, and supported values for the Cisco UCS M7 blade and rack servers for 4<sup>th</sup> Gen Intel Xeon Scalable Processors.

**Table 2.** BIOS token names and supported values

BIOS token	Platform default	Supported values
<b>Processor configuration</b>		
Intel Hyper-Threading Technology	Enabled	Enabled and Disabled
Hardware prefetcher	Enabled	Enabled and Disabled
Adjacent cache line prefetcher	Enabled	Enabled and Disabled
DCU IP prefetcher	Enabled	Enabled and Disabled
DCU streamer prefetch	Enabled	Enabled and Disabled
LLC prefetch	Enabled	Enabled and Disabled
Intel Virtualization Technology	Enabled	Enabled and Disabled
<b>Uncore configuration</b>		
UPI link enablement	Auto	Auto, 1, 2, and 3
UPI power management	Disabled	Enabled and Disabled
UPI link speed	Auto	Auto, 10.4 GTs, 11.2 GTs, and 16 GTs
XPT remote prefetch	Auto	Auto, Enabled, and Disabled
UPI prefetch	Auto	Auto, Enabled, and Disabled
XPT prefetch	Auto	Auto, Enabled, and Disabled
Sub-NUMA clustering	Disabled	Auto, SNC2, SNC4 and Disabled
LLC dead line	Enabled	Auto, Enabled, and Disabled
<b>Memory configuration</b>		
NUMA	Enabled	Enabled and Disabled
Virtual NUMA	Disabled	Enabled and Disabled
Memory refresh rate	1x Refresh	1x Refresh and 2x Refresh

BIOS token	Platform default	Supported values
<b>Memory RAS configuration setup</b>		
<b>ADDDC sparing</b>	Enabled	Enabled and Disabled
<b>Patrol scrub</b>	Enabled	Enabled and Disabled
<b>Intel VT for Directed I/O</b>	Enabled	Enabled and Disabled
<b>Power and performance configuration</b>		
<b>Enhanced CPU performance</b>	Disabled	Auto and Disabled
<b>Energy-efficient turbo mode</b>	Disabled	Enabled and Disabled
<b>Intel Turbo Boost Technology</b>	Enabled	Enabled and Disabled
<b>Processor C6</b>	Disabled	Enabled and Disabled
<b>Processor C1E</b>	Disabled	Enabled and Disabled
<b>Package C-state control</b>	C0/C1 State	No Limit, Auto, C0/C1 State, C2, C6 Retention, and C6 Nonretention
<b>Uncore frequency scaling</b>	Enabled	Enabled and Disabled
<b>Energy performance Tuning</b>	OS control	OS controls EPB; BIOS controls EPB.
<b>Workload configuration</b>	I/O Sensitive	Balanced and I/O Sensitive

## BIOS recommendations for various general-purpose workloads

This section summarizes the BIOS settings recommended to optimize general-purpose workloads:

- CPU-intensive workloads
- Energy-efficient workloads
- Low-latency workloads

The following sections describe each workload.

### CPU-intensive workloads

For CPU-intensive workloads, the goal is to distribute the work for a single job across multiple CPUs to reduce the processing time as much as possible. To do this, you need to run portions of the job in parallel. Each process-, or thread, handles a portion of the work and performs the computations concurrently. The CPUs typically need to exchange information rapidly, requiring specialized communication hardware. CPU-intensive workloads generally benefit from processors that achieve the maximum turbo frequency for any individual core at any time. Processor power management settings can be applied to help ensure that any component frequency increase can be readily achieved.

The BIOS option helps users modify the enhanced CPU performance settings. When it is enabled, this option adjusts the processor settings and enables the processor to run aggressively, which can improve overall CPU performance, but may result in higher power consumption.



## Energy-efficient workloads

Energy-efficient optimizations are the most common balanced-performance settings. They benefit most application workloads while also enabling power-management settings that have little impact on overall performance. The settings that are applied for energy-efficient workloads increase general application performance rather than power efficiency. Processor power management settings can affect performance when virtualization operating systems are used. Hence, these settings are recommended for customers that do not typically tune the BIOS for their workloads.

## Low-latency workloads

Workloads that require low latency, such as financial trading and real-time processing, require servers to provide a consistent system response. Low-latency workloads are for customers who demand the least amount of computational latency for their workloads. Maximum speed and throughput are often sacrificed to lower overall computational latency. Processor power management and other management features that might introduce computational latency are disabled.

To achieve low latency, you need to understand the hardware configuration of the system under test. Important factors affecting response times include the number of cores, the processing threads per core, the number of NUMA nodes, the CPU and memory arrangements in the NUMA topology, and the cache topology in a NUMA node. BIOS options are generally independent of the OS, and a properly tuned low-latency operating system is also required to achieve deterministic performance.

## Summary of BIOS settings optimized for general-purpose workloads

**Table 3.** BIOS recommendations for CPU-intensive, energy-efficient, and low-latency workloads

BIOS token	Platform default	CPU intensive	Energy efficient	Low latency
<b>Processor Configuration</b>				
<b>Intel Hyper-Threading Technology</b>	Enabled	Platform default	Platform default	<b>Disabled</b>
<b>Hardware prefetcher</b>	Enabled	Platform default	Platform default	Platform default
<b>Adjacent cache line prefetcher</b>	Enabled	Platform default	Platform default	Platform default
<b>DCU IP prefetcher</b>	Enabled	Platform default	Platform default	Platform default
<b>DCU streamer prefetch</b>	Enabled	Platform default	Platform default	Platform default
<b>LLC prefetch</b>	Enabled	<b>Disabled</b>	Platform default	Platform default
<b>Intel Virtualization Technology</b>	Enabled	Platform default	Platform default	<b>Disabled</b>

BIOS token	Platform default	CPU intensive	Energy efficient	Low latency
<b>Uncore configuration</b>				
UPI link enablement	Auto	<b>2</b>	Platform default	Platform default
UPI power management	Disabled	<b>Enabled</b>	Platform default	Platform default
UPI link speed	Auto	Platform default	Platform default	Platform default
XPT remote prefetch	Auto	Platform default	Platform default	Platform default
UPI prefetch	Auto	<b>Enabled</b>	Platform default	Platform default
XPT prefetch	Auto	<b>Enabled</b>	Platform default	Platform default
Sub-NUMA clustering	Disabled	<b>SNC4</b>	Platform default	Platform default
LLC dead line	Enabled	<b>Disabled</b>	Platform default	Platform default
<b>Memory configuration</b>				
NUMA	Enabled	Platform default	Platform default	Platform default
Virtual NUMA	Disabled	Platform default	Platform default	Platform default
Memory refresh rate	1x Refresh	Platform default	Platform default	Platform default
<b>Memory RAS configuration setup</b>				
ADDDC sparing	Enabled	<b>Disabled</b>	Platform default	Platform default
Patrol scrub	Enabled	<b>Disabled</b>	Platform default	Platform default
Intel VT for Directed I/O	Enabled	Platform default	Platform default	<b>Disabled</b>
<b>Power and performance configuration</b>				
Enhanced CPU performance*	Disabled	Platform default	Platform default	Platform default
Energy-efficient turbo mode	Disabled	Platform default	<b>Enabled</b>	Platform default
Intel Turbo Boost Technology	<b>Enabled</b>	Platform default	Platform default	<b>Disabled</b>
Processor C6	Disabled	<b>Enabled</b>	<b>Enabled</b>	Platform default
Processor C1E	Disabled	Platform default	<b>Enabled</b>	Platform default

BIOS token	Platform default	CPU intensive	Energy efficient	Low latency
<b>Package C-state control</b>	C0/C1 State	Platform default	<b>C6 non-retention</b>	Platform default
<b>Uncore frequency scaling</b>	Enabled	Platform default	Platform default	Platform default
<b>Energy-performance tuning</b>	OS control	Platform default	Platform default	Platform default
<b>Workload configuration</b>	I/O Sensitive	Platform default	<b>Balanced</b>	<b>Balanced</b>

**Note:**

- From Table 3. Enhanced CPU performance\*: this BIOS feature is applicable for benchmark purposes only and is not recommended for any production workloads. When this option is enabled, we highly recommend setting the fan policy to Maximum power.
- Default BIOS options are generally selected to produce the best overall performance for typical workloads. However, typical workloads differ from end user to end user. Therefore, the default settings may not be the best choice for your specific workloads.

**Additional BIOS recommendations for enterprise workloads**

This section summarizes optimal BIOS settings for enterprise workloads:

- Relational database (Oracle and SQL) workloads
- Virtualization (virtual desktop infrastructure [VDI] and virtual server infrastructure [VSI]) workloads
- Data analytics (big data) workloads
- Analytical database systems (SAP HANA) workloads
- High-Performance Computing (HPC) workloads

The following sections describe each enterprise workload.

**Relational database workloads**

Relational database systems, also known as Online Transaction Processing (OLTP) systems, contain the operational data needed to control and run important transactional business tasks. These systems are characterized by their ability to complete various concurrent database transactions and process real-time data. They are designed to provide optimal data processing speeds.

These database systems are often decentralized to avoid single points of failure. Spreading the work over multiple servers can also support greater transaction processing volume and reduce response time. In a virtualized environment, when the OLTP application uses a direct I/O path, make sure that the Intel VT for Directed I/O option is enabled. By default, this option is enabled.

**Virtualization workloads**

Intel Virtualization Technology provides manageability, security, and flexibility in IT environments that use software-based virtualization solutions. With this technology, a single server can be partitioned and projected as several independent servers, allowing the server to run different applications on the operating system simultaneously. It is important to enable Intel Virtualization Technology in the BIOS to support virtualization workloads. The CPUs that support hardware virtualization allow the processor to run multiple operating systems in the virtual machines.

---

## **Data analytics workloads**

Data analytics applications are important because they help businesses optimize their performance. Implementing data analytics in the business model can help organizations reduce costs by identifying more efficient ways of doing business and by storing large amounts of data. A company can also use data analytics to make better business decisions and help analyze customer trends and satisfaction, which can lead to new—and better—products and services.

Big-data analytics is the use of advanced analytics techniques on very large, diverse big-data sets that include structured, semistructured, and unstructured data, from any source. These data sets can be defined as ones whose size or type is beyond the capability of traditional relational databases to capture, manage, and process with low latency. In addition, new capabilities include real-time streaming analytics and impromptu, iterative analytics on enormous data sets.

## **Analytical database systems workloads**

An analytical database, also called an analytics database, is a read-only system that stores historical data about business metrics such as sales performance and inventory levels. Business analysts, corporate executives, and other workers run queries and reports against analytics databases. The information is regularly updated to include recent transaction data from an organization's operations systems.

An analytics database is specifically designed to support Business Intelligence (BI) and analytics applications, typically as part of a data warehouse or data mart. This feature differentiates it from operational, transactional, and Online Transaction Processing (OLTP) databases, which are used to process transactions such as order entry and similar business applications.

The SAP HANA platform is a flexible, data-source-independent, in-memory data platform that enables analysis of large amounts of data in real time. Using the database services of the SAP HANA platform, users can store and access data in-memory and column-based, provision databases, scale system load, ensure high availability and disaster recovery, monitor and troubleshoot databases, test workloads, model data, and secure data and systems.

## **High-performance computing workloads**

High-Performance Computing (HPC) refers to cluster-based computing that uses multiple individual nodes that are connected and that work in parallel to reduce the amount of time required to process large data sets that would otherwise take exponentially longer to run on any one system. HPC workloads are computation intensive and typically also network-I/O intensive. HPC workloads require high-quality CPU components and high-speed, low-latency network fabrics for their Message Passing Interface (MPI) connections.

Computing clusters include a head node that provides a single point for administering, deploying, monitoring, and managing the cluster. Clusters also have an internal workload management component, known as the scheduler, that manages all incoming work items (referred to as jobs). Typically, HPC workloads require large numbers of nodes with nonblocking MPI networks so that they can scale. Scalability of nodes is the single most important factor in determining the achieved usable performance of a cluster.

## Summary of BIOS settings recommended for enterprise workloads

Table 4 summarizes the BIOS tokens and settings recommended for various enterprise workloads.

**Table 4.** BIOS options recommended for enterprise workloads

BIOS token	Platform default	Relational database systems	Virtualization	Data analytics	Analytical database systems	High-performance computing
<b>Processor configuration</b>						
<b>Intel Hyper-Threading Technology</b>	Enabled	Platform default	Platform default	Platform default	Platform default	Platform default
<b>Hardware prefetcher</b>	Enabled	Platform default	Platform default	Platform default	Platform default	Platform default
<b>Adjacent cache line prefetcher</b>	Enabled	Platform default	Platform default	Platform default	Platform default	Platform default
<b>DCU IP prefetcher</b>	Enabled	Platform default	Platform default	Platform default	Platform default	Platform default
<b>DCU streamer prefetch</b>	Enabled	Platform default	Platform default	Platform default	Platform default	Platform default
<b>LLC prefetch</b>	Enabled	<b>Disabled</b>	Platform default	Platform default	Platform default	<b>Disabled</b>
<b>Intel Virtualization Technology</b>	Enabled	Platform default	Platform default	<b>Disabled</b>	<b>Disabled</b>	<b>Disabled</b>
<b>Uncore configuration</b>						
<b>UPI link enablement</b>	Auto	Platform default	Platform default	Platform default	Platform default	Platform default
<b>UPI power management</b>	Disabled	Platform default	Platform default	Platform default	Platform default	Platform default
<b>UPI link speed</b>	Auto	Platform default	Platform default	Platform default	Platform default	Platform default
<b>XPT remote prefetch</b>	Auto	Platform default	Platform default	Platform default	Platform default	Platform default
<b>UPI prefetch</b>	Auto	Platform default	Platform default	Platform default	Platform default	Platform default
<b>XPT prefetch</b>	Auto	<b>Enabled</b>	Platform default	Platform default	Platform default	Platform default
<b>Sub-NUMA clustering</b>	Disabled	<b>Auto</b>	Platform default	Platform default	Platform default	Platform default
<b>LLC dead line</b>	Enabled	<b>Disabled</b>	Platform default	Platform default	Platform default	<b>Disabled</b>

BIOS token	Platform default	Relational database systems	Virtualization	Data analytics	Analytical database systems	High-performance computing
<b>Memory configuration</b>						
<b>NUMA</b>	Enabled	Platform default	Platform default	Platform default	Platform default	Platform default
<b>Virtual NUMA</b>	Disabled	Platform default	Platform default	Platform default	Platform default	Platform default
<b>Memory refresh rate</b>	1x Refresh	Platform default	Platform default	Platform default	Platform default	Platform default
<b>Memory RAS configuration setup</b>						
<b>ADDDC sparing</b>	Enabled	Platform default	Platform default	Platform default	Platform default	Platform default
<b>Patrol scrub</b>	Enabled	<b>Disabled</b>	Platform default	Platform default	Platform default	Platform default
<b>Intel VT for Directed I/O</b>	Enabled	Platform default	Platform default	Platform default	Platform default	Platform default
<b>Power and performance configuration</b>						
<b>Enhanced CPU performance</b>	Disabled	Platform default	Platform default	Platform default	Platform default	Platform default
<b>Energy-efficient turbo mode</b>	Disabled	Platform default	Platform default	Platform default	Platform default	Platform default
<b>Intel Turbo Boost Technology</b>	Enabled	Platform default	Platform default	Platform default	Platform default	Platform default
<b>Processor C6</b>	Disabled	Platform default	<b>Enabled</b>	<b>Enabled</b>	<b>Enabled</b>	<b>Enabled</b>
<b>Processor C1E</b>	Disabled	Platform default	Platform default	Platform default	Platform default	Platform default
<b>Package C-state control</b>	C0/C1 State	Platform default	Platform default	Platform default	Platform default	Platform default
<b>Uncore frequency scaling</b>	Enabled	Platform default	Platform default	Platform default	Platform default	Platform default
<b>Energy performance tuning</b>	OS control	Platform default	Platform default	Platform default	Platform default	Platform default
<b>Workload configuration</b>	I/O Sensitive	Platform default	<b>Balanced</b>	Platform default	<b>Balanced</b>	<b>Balanced</b>

**Note:** Default BIOS options are generally selected to produce the best overall performance for typical workloads. However, typical workloads differ from end user to end user. Therefore, the default settings may not be the best choice for your specific workloads.

---

## Conclusion

When tuning system BIOS settings for performance, you need to consider a number of processor and memory options. If the best performance is your goal, be sure to choose options that optimize for performance in preference to power savings, and experiment with other options such as CPU prefetchers, CPU power management, and CPU hyperthreading.

## For more information

For more information about Cisco UCS C-Series, and X-Series M7 servers, see the following resources:

- Cisco UCS X210c M7 Compute Node: <https://www.cisco.com/c/dam/en/us/products/collateral/servers-unified-computing/ucs-x-series-modular-system/x210cm7-specsheet.pdf>
- Cisco UCS X410c M7 Compute Node: <https://www.cisco.com/c/dam/en/us/products/collateral/servers-unified-computing/ucs-x-series-modular-system/x410cm7-specsheet.pdf>
- Cisco UCS C220 M7 Rack Server: <https://www.cisco.com/c/dam/en/us/products/collateral/servers-unified-computing/ucs-c-series-rack-servers/c220m7-sff-specsheet.pdf>
- Cisco UCS C240 M7 Rack Server: <https://www.cisco.com/c/dam/en/us/products/collateral/servers-unified-computing/ucs-c-series-rack-servers/c240m7-sff-specsheet.pdf>
- 4<sup>th</sup> Gen Intel Xeon Scalable Processors brief: <https://www.intel.com/content/www/us/en/products/docs/processors/xeon-accelerated/4th-gen-xeon-scalable-processors-product-brief.html>

**Americas Headquarters**  
Cisco Systems, Inc.  
San Jose, CA

**Asia Pacific Headquarters**  
Cisco Systems (USA) Pte. Ltd.  
Singapore

**Europe Headquarters**  
Cisco Systems International BV Amsterdam,  
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at <https://www.cisco.com/go/offices>.

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: <https://www.cisco.com/go/trademarks>. Third-party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)