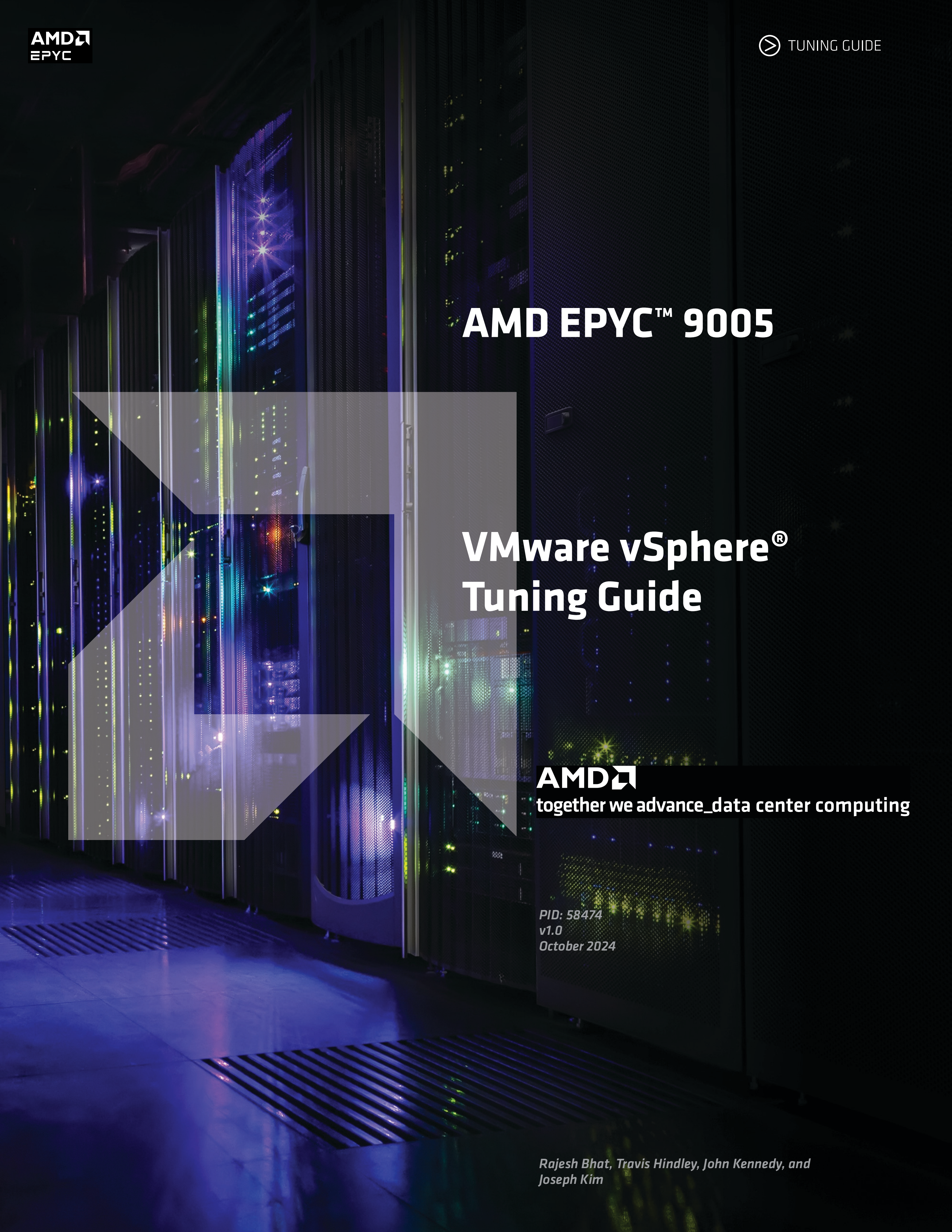

TUNING GUIDE

AMD EPYC™ 9005

VMware vSphere® Tuning Guide

AMD
together we advance_data center computing

PID: 58474
v1.0
October 2024

Rajesh Bhat, Travis Hindley, John Kennedy, and Joseph Kim

DATE	VERSION	CHANGES
June, 2024	0.1	Initial NDA release
October, 2024	1.0	Initial public release

AUDIENCE

This document is intended for a technical audience. You should:

- Be familiar with your server's BIOS management interface.
- Have admin access to vSphere/VMware ESXi™ and be familiar with configuring virtual environments.
- Be familiar with OS-specific configuration, monitoring, and troubleshooting tools.

Tuning Guide for AMD EPYC 9005 Processors on vSphere

CONTENTS

Chapter 1: Introduction

This Tuning Guide describes parameters you can adjust on servers using AMD EPYC™ 9005 processors running vSphere and helps to optimize performance on a per-workload level. Default configurations of hardware and BIOS from different OEM vendors may not provide the best possible performance on all operating system platforms and for all workloads. Use this guide as a reference for:

- Supported versions of vSphere software and their optimization options.
- System configuration best practices.
- BIOS settings that can impact performance.
- Workload-specific optimizations in BIOS and operating systems for a variety of workloads.

1.1 - Supported Software Versions

vSphere is the software stack that allows customers to run their virtual machines and manage their hypervisor. VMware® ESXi™ is the Hypervisor/OS within vSphere that abstracts compute, memory, network, and storage to provide an infrastructure to run virtual machines. VMware® vSAN™ is the Software Defined Storage solution. vSphere is planned to be supported on AMD EPYC 9005 processor-based platforms starting with an updated version of vSphere.

Please see AMD EPYC™ Processors Minimum Operating System (OS) Versions for detailed OS support information. AMD recommends using the latest available OS version. Updated versions of vSphere are frequently optimized for AMD EPYC architecture after initial support and can provide better performance on a variety of workloads. Not all operating systems or hypervisors support all of the features that are introduced with a new generation of processors. Please refer to the vSphere documentation on specific releases to identify support for these features. Please also refer to the latest version of the *AMD64 Architecture Programmer's Manuals or Processor Programming Reference (PPR) for AMD Family 1Ah* for more details on the features.

1.2 - Important Reading

Please be sure to read the following guides (available from the AMD Documentation Hub), which contain important foundational information about 5th Gen AMD EPYC processors:

- *AMD EPYC™ 9005 Processor Architecture Overview*
- *BIOS & Workload Tuning Guide for AMD EPYC™ 9005 Series Processors*
- *Memory Population Guidelines for AMD EPYC™ 9005 Series Processors*

THIS PAGE INTENTIONALLY LEFT BLANK.

Chapter 2: BIOS Configuration

This chapter provides high-level lists of the default AMD EPYC 9005 BIOS settings that are relevant for running vSphere. Please see Chapter 4 of the *BIOS & Workload Tuning Guide for AMD EPYC™ 9005 Series Processors* (available from the AMD Documentation Hub) for detailed descriptions. Later chapters in this Tuning Guide discuss the BIOS options as they relate to vSphere and associated set of workloads.

Note: The default setting name and values described in this chapter are the AMD default names and values that serve as recommendations for OEMs. End users must confirm their OEM BIOS setting availability and options.

Name	Description
SMT Control	• **Disabled:** Single hardware thread per core. • **Enabled/Auto:** Two hardware threads per core.
Determinism Control	• **Auto:** Use default determinism settings. • **Manual:** Specify custom determinism settings.
Determinism Enable	• **0:** Disables performance determinism. • **1/Auto:** Enables performance determinism.
C-States	• **Enabled:** Allows the idle cores to enter lower power states to improve power savings. Can allow for additional performance boost on other active cores. • **Disabled:** Prevents idle cores from entering lower power states. Provides a more consistent response time from the cores at the expense of additional power consumed.
LLC as NUMA Node (CCX as NUMA)	• **Enabled:** Overrides the **NPS** setting for the number of NUMA nodes by mapping each LLC as a NUMA node. This does not impact the memory interleaving • **Disabled:** NUMA nodes are determined by the **NPS** setting.
Memory Target Speed	• **Values 3200–6000 MT/s:** Run the DRAM memory Target clock Speed at the specified speed. The DRAM memory target is the DDR rate.

Table 2-1: BIOS settings relevant to vSphere

NUMA Nodes Per Socket (NPS)	Memory Interleaving: The NPS setting always determines the memory interleaving regardless of whether LLC as NUMA is Enabled or Disabled. Number of NUMA nodes per socket (if **LLC as NUMA Node** is **Disabled**): • **NPS1:** One NUMA node per socket (provides consistent average memory latency to all the accesses within a socket). • **NPS2:** Two NUMA nodes per socket. • **NPS4:** Four NUMA nodes per socket • **NPS0:** Only applicable for dual-socket systems. A single NUMA node is created for the whole two- socket platform.
APBDIS	• **0/Auto:** Dynamically switch the Infinity Fabric P-state based on link usage. • **1:** Enabled fixed Infinity Fabric P-state control.
IOMMU	• **Disabled:** Disables IOMMU. • **Enabled:** Enables IOMMU.
TSME	• **Disabled:** Disables transparent secure memory encryption. • **Enabled:** Enables transparent secure memory encryption.
SEV-ES	Secure Encrypted Virtualization-Encrypted State (SEV-ES) mode extends SEV protection to the contents of the CPU registers by encrypting them when a virtual machine stops running. Combining SEV and SEV-ES can reduce the attack surface of a VM by helping protect the confidentiality of data in memory. • **Disabled:** SEV-ES is disabled. • **Enabled:** SEV-ES is enabled.
SEV-SNP	Secure Encrypted Virtualization-Secure Nested Paging (SEV-SNP) mode builds on SEV and SEV-ES by adding strong memory integrity protection to create an isolated execution environment that helps prevent malicious hypervisor-based attacks such as data replay and memory re-mapping. SEV-SNP also introduces several additional optional security enhancements that support additional VM use models, offer stronger protection around interrupt behavior, and increase protection against recently-disclosed side channel attacks. • **Disabled:** SEV-SNP is disabled. • **Enabled:** SEV-SNP is enabled. *Note: SEV-SNP is not currently enabled in vSphere.*

Table 2-1: BIOS settings relevant to vSphere (Continued)

Chapter 3: System Configuration Best Practices

3.1 - Memory Configuration

For optimal performance, populate 12 DIMMs per processor for a 1 DIMM per channel (1DPC) configuration. AMD EPYC 9005 Series processors support a **Maximum Memory Bus Frequency** of 6000 MHz.

3.2 - PCIe Subsystem

The ESXi scheduler automatically allocates CPU and memory resources to VMs on appropriate NUMA nodes to allow for optimal performance. In some cases, such as for I/O-intensive workloads, you can improve performance by manually pinning the workload on the same socket that connects to the I/O device used. For example, for a networking-intensive operation, you can manually pin the workload on the socket that the NIC connects to. Tools such as `esxcfg-info` (`-n` for NIC and `-s` for HBA) can help determine the connectivity between PCI devices and sockets. *VMware vSphere® Network Tuning Guide for AMD EPYC™ 9005 Series Processors* (available from the [AMD Documentation Hub]) provides additional details on tuning the network configuration on your virtual infrastructure.

3.3 - BIOS Settings

Supported vSphere versions include an enhanced scheduler to provision workloads on cores within a CCX/CCD to best utilize the shared L3 cache. This feature optimizes performance while drastically simplifying the out-of-the-box BIOS configuration needed by Virtual Infrastructure administrators on AMD EPYC-based platforms. This section describes additional options to help you evaluate the impact on your workload.

3.3.1 - Prerequisites

- Upgrade to the latest available vendor BIOS to get the most current fixes, features, and performance benefits.
  - Refer to the server vendor's browser compatibility guidelines and need for maintenance windows if you are using a web browser to upgrade the BIOS.
  - Refer to the server vendor's guidelines on state of the machine and need for maintenance windows if you are using the OS to upgrade the BIOS.
- Ensure that all the desired settings remain in place after every BIOS upgrade.

3.3.2 - BIOS Settings for Performance

The following BIOS options provide optimal performance for a variety of virtual applications on vSphere deployments. They may not achieve additional objectives, such as optimal power consumption.

3.3.2.1 - Performance and Power Profiles

- OEM BIOS implementations tend to have profiles that initialize several BIOS settings to match workload requirements. To improve performance, pick the profile that favors performance in virtualized environments.
- Platforms often present a trade-off between power savings and performance. Turning off power saving schemes in the BIOS may provide consistent performance for certain workloads at the expense of additional power consumption.
- Platforms may also provide the ability to control determinism. Power Determinism allows you to extract the maximum performance from your processor by disabling artificial limits on the processor to establish a standardized performance level. Hence, select "Power Determinism" over "Performance Determinism." See Power/Performance Determinism for more information.
- Providing additional cooling to the CPU helps drive higher workloads more effectively. Set BIOS options that enhance cooling. The server may consume more power as a result.
- Always provide adequate power to the server by plugging in all of its redundant power supply units (PSUs). This also helps keep the server running if one of the PSUs fails.

3.3.2.2 - Processor

- Ensure that AMD SMT is enabled. This enables multiple hardware threads per core, which effectively doubles the number of logical processors available to your vSphere workloads.
- Ensure that AMD Virtualization Technology is turned on. This setting is essential for virtual machines running on vSphere.

3.3.2.3 - NUMA Configuration

Supported versions include an enhanced scheduler that provisions workloads on cores within a CCX/CCD to best utilize the shared L3 cache, as described in Performance Optimizations in VMware vSphere 7.0 U2 CPU Scheduler for AMD EPYC Processors*. This simplifies the BIOS configuration needed for NUMA (shown below) for optimal performance on a variety of workloads to essentially overlap with the out of the box settings delivered by various OEMs. Workloads using VMs with 4 or fewer vCPUs may see improved performance from the following settings:

- NPS (NUMA per socket) = 1
- LLC As NUMA = Disabled

Workloads using VMs larger than 8 vCPUs may span NUMA boundaries and tend to have lower performance. Ensure that workloads fit into NUMA boundaries for best performance. This section provides alternate NUMA settings that may help align your workloads to NUMA boundaries. Please see *AMD EPYC™ 9005 Series Architecture Overview* (available from the AMD Documentation Hub) for detailed information about NUMA settings.

Workloads with few vCPUs per virtual machine (that is, less than a quarter of the number of cores per socket) tend to benefit from the following settings:

- NPS (NUMA per socket) = 4
- LLC As NUMA = Enabled

Workloads with many vCPU per virtual machine (that is, greater than half the number of cores per socket) benefit from the following settings:

- NPS (NUMA per socket) = 1
- LLC As NUMA = Disabled

The settings above are also good defaults if your workloads contain either dissimilarly-sized VMs or VMs whose sizes are not known in advance. ESXi scheduler also works best with these options.

3.3.2.4 - Other Settings

The following additional BIOS settings can also impact performance:

- **C-States:** If the server cores are not saturated running workloads, then enabling C-States on the platform may boost the performance of the executing cores while saving power on unused cores that have transitioned to deeper C-States.
- **IOMMU:** Ensure that AMD IOMMU is turned on. IOMMU is essential for compatibility with I/O devices, system performance, and for technologies such as SR-IOV.
- **Network/Storage Endpoint (NIC/HBA) specific settings:** Settings such as Relaxed Ordering may continue to provide additional performance benefit if supported by the endpoint vendor. Consult with the respective endpoint vendor on support and configuration guidelines for such features.

3.4 - ESXi (OS) Advanced Configuration Parameters

Application software can have varying hardware resource requirements. The Operating System (OS) attempts to fulfill these requirements by interacting with hardware, abstracting the hardware resource and making it available to the application software. The following list captures some of the OS advanced configuration parameters and the conditions under which each parameter can improve application and workload performance.

- **Numa.LocalityWeightActionAffinity=0:** See NUMA nodes are heavily load imbalanced causing high contention for some virtual machines*.
- **Numa.PreferHT=1:** Use hyperthreads with fewer NUMA nodes instead of full physical cores spread over multiple NUMA nodes for memory latency sensitive workloads with low processor utilization or high inter-thread communication. See Configure virtual machines to use hyper-threading with NUMA in VMware ESXi 4.1.x / 5.0.x*.
- **Power.CpuPolicy=HighPerformance:** This influences the ESXi power profile as an alternative to setting high-performance profile in the BIOS.

THIS PAGE INTENTIONALLY LEFT BLANK.

Chapter 4: Workload Optimization on vSphere

Server vendors group different BIOS settings and default values into BIOS profiles to improve the out-of-the-box experience for customers. AMD EPYC-based platforms typically include BIOS profiles that help with virtualization workloads by defaulting to the highest memory speeds allowed by the DIMMs, NPS=1, LLC As NUMA disabled, etc. These settings provide good performance on a variety of workloads, but a few application workloads can benefit from further tuning, as discussed below.

4.1 - vSAN

vSAN is a hyperconverged storage virtualization solution that runs on vSphere. Please see the vSAN Planning and Deployment Guide* for detailed best practice information. The settings in this section may provide additional performance gains, as validated by the HCIBench synthetic I/O generating tool. HCIBench automates different synthetic I/O workloads using VMs and measures the resulting IOPS and latency. Settings that improve HCIBench results may not directly translate to similar gains on real-life workloads but may help vSphere administrators understand which BIOS settings help with which kinds of IO patterns. See the *vSAN Tuning Guide for AMD EPYC™ 9005 Processors* (available from the AMD Documentation Hub) for more detailed information about the settings described below.

4.1.1 - Optimize Latency

The following settings can help reduce latency in a vSAN cluster:

- Using lower DIMM speeds returns more power to the CPU, which can boost performance. Lower your DIMM speeds if your application does not need the highest DIMM speeds.
- xGMI is the communication link between the two sockets in a dual-socket server. Set xGMI Link Width to maximum (16), as described in Section 4.4.2 of the *BIOS & Workload Tuning Guide for AMD EPYC™ 9005 Series Processors* (available from the AMD Documentation Hub).
- Disable Algorithm Performance Boost (APB)=1 (disabled). Disabling APBDIS and specifying a fixed Infinity Fabric P-State of 0 forces the Infinity Fabric and memory controllers into full-power mode, which can help reduce latency jitters, as described in Section 4.4.3 of the *BIOS & Workload Tuning Guide for AMD EPYC™ 9005 Series Processors* (available from the AMD Documentation Hub).

4.1.2 - Power Profiles

- Certain processor C-States reduce the power consumed by inactive cores and redirect that power to active cores, which can help drive the frequency of those active cores closer to their maximum boost speeds. This can result in better application performance on some applications. Thus, applications that are not very sensitive to latency may benefit from setting **Power.CpuPolicy** to **Balanced** in the ESXi power profile and as OS-controlled in BIOS. These settings allow ESXi to transition the cores between various processor C-States and manage power/performance more efficiently.
- Setting **Power.CpuPolicy** to **High Performance** in the ESXi power profile restricts the use of processor C-states, which may lower latency at the cost of higher power consumption.
- Using NVMe-based drives as cache drives in disk group significantly improves throughput and reduces latency.
- Use at least a 25Gb NIC for vSAN I/O traffic to avoid any networking I/O bottlenecks.

4.2 - Data Management - RDBMS, Big Data Analytics

Databases such as Microsoft SQL Server and Big Data solutions such as Hadoop are increasingly being deployed in VMs and containers. Configuring the workloads within the guest OS can be different from how an administrator would configure the workload on a bare metal server. For more details, please see:

- Architecting Microsoft SQL Server on VMware vSphere*
- Virtualizing Hadoop on VMware vSphere*

Sizing the vCPUs and memory on the VMs to align with NUMA boundaries on the server while following the other best practices listed in these guides can optimize performance in both physical and virtual infrastructures. Please see *AMD EPYC™ 9005 Series Architecture Overview* (available from the AMD Documentation Hub) for additional information.

4.3 - Virtual Desktop Infrastructure (VDI)

VDI refers to a technology where VM sessions stream to end-users to provide the experience of a desktop computer. A centralized server deploys and manages these VMs. Several VDI solutions, such as Omnissa® Horizon® and Citrix® Virtual Desktops run on vSphere. The following settings may help improve the scale and performance of the deployed virtual desktops on AMD EPYC 9005-based platforms:

- High Performance profile with C-States disabled in the BIOS.
- ESXi power profile set to **High Performance**.

See *Virtual Desktop Infrastructure (VDI) Tuning Guide for AMD EPYC™ 9005 Series Processors* (available from the AMD Documentation Hub) for additional information.

4.4 - Confidential Computing

AMD EPYC 9005 processors provide a variety of security technologies designed to guard against the security pitfalls of traditional virtualization. AMD Secure Memory Encryption (SME), AMD Secure Encrypted Virtualization (SEV), and AMD Secure Encrypted Virtualization Encrypted-State (SEV-ES) help secure customer data by:

- Encrypting individual memory pages using dedicated hardware in the memory controller.
- Generating and managing memory encryption keys inside the AMD Secure Processor.
- Encrypting the VM register state.
- See *Building Confidential Compute with VMware vSAN™ for Private Cloud* (available from the AMD Documentation Hub) for detailed information about these features and how to deploy them.

THIS PAGE INTENTIONALLY LEFT BLANK.

Chapter 5: Resources

5.1 - vSphere Performance Best Practices

- Performance Optimization in vSphere 7.0 U2 Scheduler for AMD EPYC Processors*
- Performance Best Practices for VMware vSphere 8.0*

5.2 - ESXi Commonly Used Tools

- Using the esxtop Utility* (performance monitoring tool)
- Collecting diagnostic information for VMware ESXi*
- ESXi 8.0 ESXCLI Command Reference*

5.3 - Blog/Videos on SEV-ES and vSphere

- Confidential vSphere Pods (using SEV-ES)*
- Confidential VMs on vSphere (using SEV-ES)*

5.4 - Other Resources

- Virtual Machine (VM) Migration Guide for AMD EPYC™ 7000 Series Processors
- HCIBench*
- Login VSI*
- vSAN Planning and Deployment*
- Omnissa Horizon Cloud Service*
- Citrix App and Desktop Virtualization*

Chapter 6: Glossary

- **APBDIS:** - Disable Algorithm Performance Boost (APB)
- **BMC:** Baseboard Management Controller
- **BIOS:** Basic Input/Output System
- **C-States:** Configurations of power consumption of a core if it does not have to execute any instructions
- **CCD:** Core/Cache Complex Die
- **CCX:** Core/Cache Complex
- **CPU:** Central Processing Unit (Processor)
- **CXL:** Compute Express Link™
- **DDR:** Double Data Rate (DRAM Technology)
- **DRAM:** Dynamic Random Access Memory
- **DIMM:** Dual In-line Memory Module
- **DPC:** DIMMs Per Channel
- **FMS:** Family/Model/Stepping of processor
- **Gb:** Gigabit
- **HBA:** Host Bus Adapter
- **I/O:** Input/Output
- **IOD:** I/O Die
- **IOMMU:** Input Output Memory Management Unit
- **IOPS:** Input/Output Operations Per Second
- **IPC:** Instructions per Cycle
- **LLC:** Last Level Cache
- **MCM:** Multi-Chip Module
- **NIC:** Network Interface Card
- **NMI:** Non-Maskable Interrupt
- **NPS:** NUMA Nodes Per Socket
- **NUMA:** Non-Uniform Memory Access
- **NVMe:** Non-Volatile Memory Host Controller Interface Specification

- **OEM:** Original Equipment Manufacturers
- **OPN:** Ordering Part Number
- **OS:** Operating System
- **P-States:** Configurations of voltage and CPU frequency during operation
- **PCI/PCIe:** Peripheral Component Interconnect Express
- **PSU:** Power Supply Unit
- **RAS:** Reliability, Availability, and Serviceability
- **RDBMS:** Relational Database Management System
- **SATA:** Serial ATA (Advanced Technology Attachment)
- **SME:** AMD Secure Memory Encryption
- **SEV:** AMD Secure Encrypted Virtualization
- **SEV-ES:** AMD Secure Encrypted Virtualization Encrypted-State
- **SMT:** Simultaneous Multi-Threading
- **SoC:** System on a Chip
- **SR-IOV:** Single-Root Input/Output Virtualization
- **TDP:** Thermal Design Point
- **UMC:** Universal Memory Controllers
- **vCPU:** Virtual CPUs (found inside VMs)
- **VDI:** Virtual Desktop Infrastructure
- **VM:** Virtual Machine
- **xGMI:** External (socket to socket) Global Memory Interconnect

THIS PAGE INTENTIONALLY LEFT BLANK.

VMware® vSphere® Tuning Guide for AMD EPYC™ 9005 Processors

PID: 58474

Rajesh Bhat, Travis Hindley, John Kennedy, and Joseph Kim