



Microsoft Windows® Server Tuning Guide for AMD EPYC™ 7002 Processors

Publication # **56782** Revision: **1.0**

Issue Date: **November 2019**

Author: **Chris Karamatas**

© 2019 Advanced Micro Devices, Inc. All rights reserved.

The information contained herein is for informational purposes only and is subject to change without notice. While every precaution has been taken in the preparation of this document, it may contain technical inaccuracies, omissions and typographical errors, and AMD is under no obligation to update or otherwise correct this information. Advanced Micro Devices, Inc. makes no representations or warranties with respect to the accuracy or completeness of the contents of this document, and assumes no liability of any kind, including the implied warranties of noninfringement, merchantability or fitness for particular purposes, with respect to the operation or use of AMD hardware, software or other products described herein. No license, including implied or arising by estoppel, to any intellectual property rights is granted by this document. Terms and limitations applicable to the purchase or use of AMD's products are as set forth in a signed agreement between the parties or in AMD's Standard Terms and Conditions of Sale.

Trademarks

AMD, the AMD Arrow logo, AMD EPYC, and combinations thereof are trademarks of Advanced Micro Devices, Inc. Other product names and links to external sites used in this publication are for identification purposes only and may be trademarks of their respective companies.

Contents

Chapter 1	Introduction.....	5
Chapter 2	AMD EPYC™ 7002 Series Microarchitecture.....	6
2.1	Overview.....	6
2.2	Zen 2 Core	6
2.3	Core Complex Die (CCD) and Core-Complex (CCX).....	6
2.4	I/O Die (IOD).....	7
2.5	Memory and I/O.....	7
2.6	NUMA Topology.....	8
Chapter 3	Hardware Configuration Best Practices.....	10
3.1	Processors	10
3.2	Memory.....	10
3.3	I/O Subsystem.....	11
3.3.1	Storage	11
3.3.2	Network Interfaces Adapters	11
Chapter 4	BIOS Tuning Guidelines	13
4.1	Collaborative Processor Performance Control (CPPC).....	14
4.2	NUMA	15
4.3	x2APIC	15
Chapter 5	Supported Microsoft Windows Server Versions.....	17
Chapter 6	OS Tuning and Tools	18
6.1	OS Settings	18
6.1.1	Setting Processor Affinity.....	18
6.2	Managing Microsoft Hyper V.....	19
6.3	Other Tools & Resource	20
6.3.1	Windows Performance Monitoring Tools & Resources.....	20
6.3.2	AMD Tools	20

Revision History

Date	Revision	Description
November 18, 2019	1.0	Initial public release.

Purpose

This guide provides information on tuning servers with AMD EPYC 7002 Processors running Microsoft Windows Server Operating Systems.

This document includes:

- AMD EPYC 7002 Series Micro-architecture details
- Information relating to Microsoft Windows Server OSes on EPYC 7002 Processors
- BIOS settings that may impact performance
- Hardware configuration best practices
- Supported versions of operating systems and support details
- OS commands that relate to optimization
- Information on further resources to assist with performance and analysis

Chapter 1 Introduction

The AMD EPYC™ 7002 Series Processors are built with leading-edge 7nm technology, Zen 2 core microarchitecture. The AMD EPYC™ SoC offers a consistent set of features across 8 to 64 cores, including 128 lanes of PCIe® Gen 4, 8 memory channels and access to up to 4 TB of high-speed memory. AMD EPYC 7002 processors are built with the following specifications:

AMD EPYC™ 7002 Series	
Process technology	7nm
Max Processor speed	3.4GHz
Max number of cores	64
Max memory speed	3200MHz
Max memory capacity	4TB
Peripheral Component Interconnect	128 lanes (max) PCIeGen4

Chapter 2 AMD EPYC™ 7002 Series Microarchitecture

2.1 Overview

Processor cores, memory controllers, I/O controllers, and security are incorporated into a Multi-Chip Module (MCM).

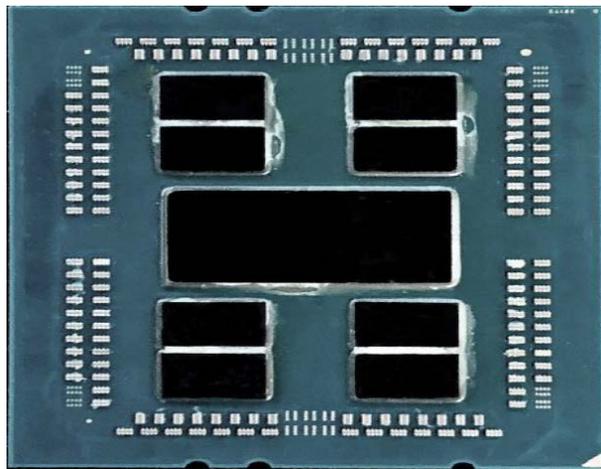


Figure 1 EPYC 7002 Configuration with 8 Core Complex Dies (CCDs) and central I/O Die (IOD)

2.2 Zen 2 Core

The EPYC 7002 series processor is based the new Zen 2 processor core, which includes an L1 write-back cache. Each core can support Simultaneous Multi-threading (SMT), allowing 2 execution threads to run simultaneously per core. Each core also includes a private 512KB L2 cache.

2.3 Core Complex Die (CCD) and Core-Complex (CCX)

Up to four Zen 2 cores share a 16MB (last level) L3 cache. These four cores and their associated caches are referred to as a Core-Complex (CCX). Each Core Complex Die (CCD) contains 2 CCXs. While the two CCXs and corresponding L3 Caches are on the same CCD or chiplet, they are distinct, and not directly connected.



Figure 2 Two Core Complexes (CCXs) on a Core Complex Die (CCD)

The SoC package can include 2, 4, 6 or 8 (see Figure 4) CCDs or chiplets and any number of cores activated to create the EPYC 7002 Series product line.

2.4 I/O Die (IOD)

CCDs may be abstracted as a quadrant, as shown in figure below. The CCDs connect to memory, I/O, and each other through the I/O Die (IOD). This IOD supports dual DDR4 memory channels, PCIe Gen4, and Infinity Fabric links. All dies, or chiplets, interconnect with each other via AMD’s Infinity Fabric Technology.

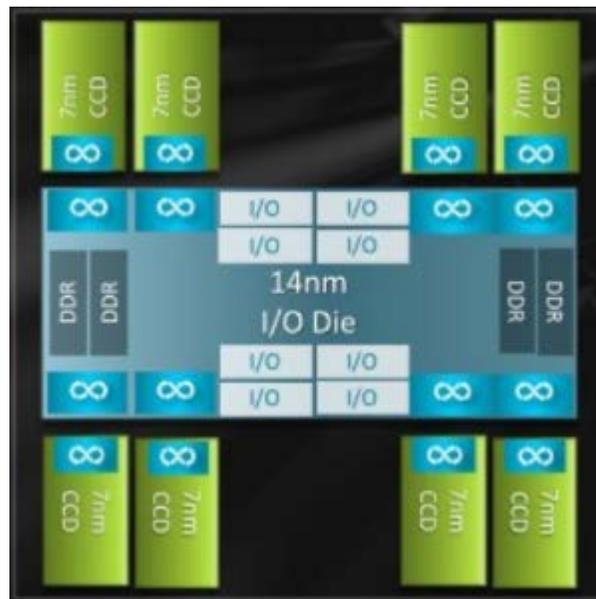


Figure 3 Single socket EPYC 7002 Processor internal connection between CCDs and Memory through the IOD via Infinity Fabric

2.5 Memory and I/O

Each EPYC 7002 processor supports 8 memory channels. Each memory channel supports up to 2 DIMMs. Based upon BIOS settings these channels can be interleaved across a quadrant (2-way interleave), all the way through 16-channel interleave, that is interleaved across all memory

channels of a 2-socket system. The system may have access to 4TB of DDR4 memory running at 3200MHz.

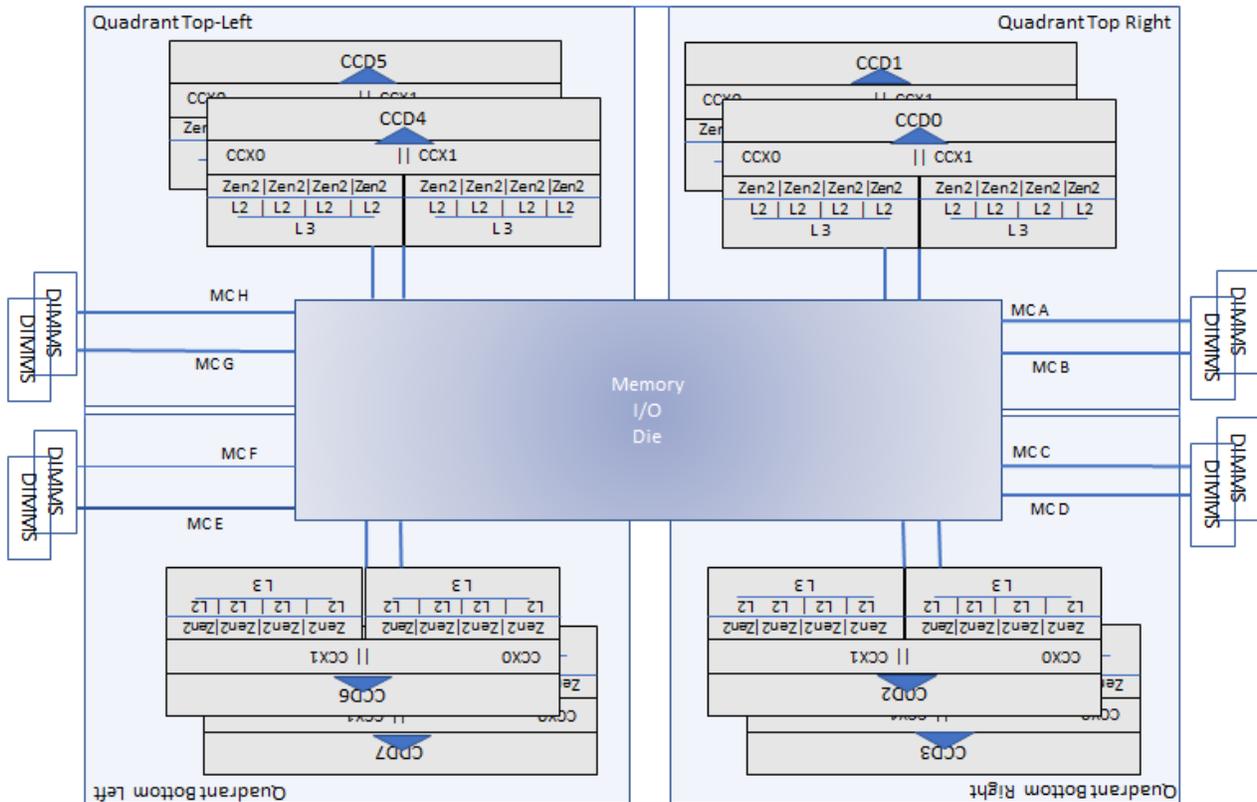


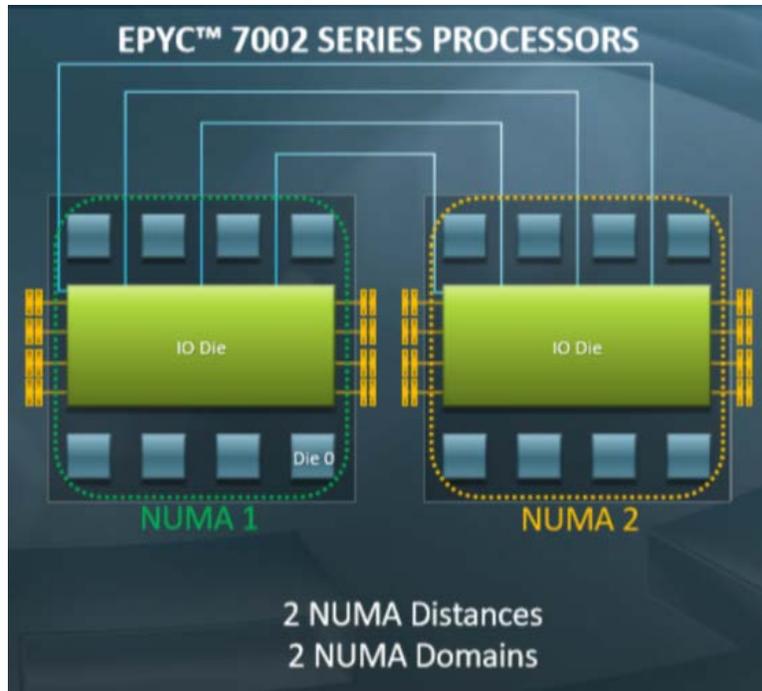
Figure 4. 64core EPYC 7002

The PCI subsystem provides up to 128 lanes of high speed I/O. While all Memory and I/O connects to the single I/O Die, they can be abstracted in to separate quadrants each with 2 DIMM channels and 32 I/O lanes.

Two EPYC 7002 SoC's are interconnected through an external Global Memory Interconnect (xGMI2) links, as part of the Infinity Fabric.

2.6 NUMA Topology

The EPYC 7002 Series processors use a Non-Uniform Memory Access (NUMA) Micro-architecture. Using the NUMA Nodes Per Socket (NPSx) BIOS settings, a system may potentially be configured as 1, 2, 4, or 8 NUMA domain system. (See Section 4.2 NUMA below).



**Figure 4 One potential NUMA Configuration for a 2-socket EPYC 7002 System –
 NUMA nodes per socket (NPSx) = 1, gives us 2 NUMA Domains**

Figure 4 shows a dual socket or dual SoC Processor, and one of the possible NUMA configurations (NPS1). Here each socket is defined as a separate NUMA node. Via BIOS each socket can be further configured into 2 (or even 4) NUMA domains.

The closest processor-memory distance is between a core and memory (controller) within the same quadrant. The furthest distance is between a core and memory (controller) in separate sockets, opposite quadrants.

For more details, see:

Socket SP3 Platform NUMA Topology for AMD Family 17h Models 30h–3Fh
https://developer.amd.com/wp-content/resources/56338_1.00_pub.pdf

Chapter 3 Hardware Configuration Best Practices

3.1 Processors

AMD provides a variety of EPYC 7002 Series Processors to support a range of workloads and environments. Processors with higher core counts will provide more computational performance. Processors that run at higher frequencies also will increase performance.

For more details on available OPNs/models, see:

<https://www.amd.com/system/files/documents/AMD-EPYC-7002-Series-Datasheet.pdf>

3.2 Memory

EPYC 7002 Series Processors provide access to 4TB of RAM. In general, it is important to have enough memory for corresponding workloads, to avoid excessive paging which may degrade performance.

AMD recommends all eight memory channels per CPU socket be populated and each channel have equal and identical capacity DIMM.

In systems that provide 2 DIMM slots per channel, it is suggested to populate open channels before populating two DIMMs on a given channel.

Maximum memory speeds would be attained by populating a single DIMM slot with maximum frequency DDR4 memory:

- The maximum DDR frequency for platforms that support previous generations of AMD EPYC would be 2666MT/sec.
- The maximum DDR frequency for platforms specifically designed for EPYC 7002 is 3200 MT/sec.

For workloads sensitive to memory throughput, use the maximum memory frequency supported by your DIMMs. This is usually the system BIOS default.

However, for workloads sensitive to decreased memory latencies, synchronizing your memory frequency with that of the Infinity Fabric will perform best.

Platforms that support previous generations of AMD EPYC have a maximum Infinity Fabric frequency of 1333 MHz, so running memory that is multiple of this i.e. 2667 Mhz (or lower), will result in lower latency.

Platforms specifically designed for EPYC 7002 have a maximum Infinity Fabric frequency of 1467 MHz. Having memory at 2933 Mhz (or lower) will result in lower latency, as this synchronizes with the Data Fabric clock.

Lowering the memory clock speed may also result in power savings in the unified memory controller. This allows other SoC components power to potentially consume for performance boost elsewhere – depending on the workload.

3.3 I/O Subsystem

For IO intensive workloads, performance can be improved by having the workload execute on the cores that correspond to the same socket that also connects to the PCIe slot used by the IO device.

3.3.1 Storage

EPYC based systems provide access to a variety of devices which incorporate performant technologies. NVMe SSDs provide performance advantages with respect to I/O queues, interrupt processing, etc. in addition to power efficiencies.

Providing adequate number of disks, appropriately distributed to avoid bottlenecks is important. Placing the paging file on it's own disk will avoid contention from other processes attempting to access a shared disk.

3.3.2 Network Interfaces Adapters

For a networking intensive operation, place the workload on the cores of the socket that the NIC also connects through. Many modern day NICs support Receive Side Scaling (RSS), which allows for distribution of network workloads across multiple logical processors. Based upon this support we can determine nearest processor-PCIe slot distance.

- [Start Windows PowerShell](#)
- Run `Get-NetAdapterHardwareInfo` cmdlet, this will provide you with the name of the connection for each NIC

```
PS C:\Users\Administrator> Get-NetAdapterHardwareInfo
```

Name	Segment	Bus	Device	Function	Slot	NumaNode	PcieLinkSpeed	PcieLinkWidth	Version
Ethernet 2	0	33	0	0	238	9	8.0 GT/s	8	1.1
Ethernet	0	33	0	1	238	9	8.0 GT/s	8	1.1

- Run `Get-NetAdapterRss -Name <Connection Name>`
This will provide you with the `RssProcessorArray` elements, each of which shows the distance between processor cores and the PCIe slot where the NIC is currently plugged in.

Chapter 4 BIOS Tuning Guidelines

Windows Server performance can be impacted by various BIOS settings. Vendor's default BIOS settings are intended to provide general high performance for their servers. You might want to further optimize performance for specific workloads and contexts.

The following areas may be available within certain vendor BIOSes and are listed for awareness as they may affect performance. Not all settings below may be available on all vendor platforms or models. Also note that modifying a setting may come with other consequences, such as higher power consumption, etc.

- **I/O Die related settings**

- *xGMI Link Max Speed*: increasing speed between SoC communication link (which increases power to SoC which may have unintended consequence of preventing core frequency boost).
- *xGMI Link Width*: widening the communication link between sockets.
- *SoC P-state*: forcing Data Fabric power state to highest performing state.
- *Data Fabric C-states*: preventing I/O Die from going into low power state

- **NUMA related settings**

- *ACPI SRAT L3 Cache as NUMA Domain*: Use the ACPI SRAT table to define the NUMA domain based on CCX boundary, so number of NUMA domains is equal to the number of Last Level Caches or CCXs
- *NUMA Nodes per Socket (NPS)*: This setting relates to memory interleaving of the 8 memory channels per socket. For example, NPS1 implies a single NUMA domain, with all the cores within the socket and all the corresponding memory in this one NUMA domain. Memory is interleaved across the socket's eight memory channels. All PCIe devices on the socket belong to this single NUMA domain. NPS2 i.e. 2 nodes per socket, interleaves memory across 4 channels. With Hyper-V and 2P-64c\SMT part, NPSx = 2 or 4 is suggested. See section [4.2 NUMA](#) below. NPSx setting may be independent of the ACPI SRAT L3 Cache as NUMA Domain BIOS setting. NPS modes depend on the part you use, for example a 6 CCD part would not support NPS4.

When the setting of "ACPI SRAT L3 As NUMA Domain" is Enabled; we will always get NUMA nodes equal to the number of L3's in the system i.e. 32 NUMA Nodes on a 2P EPYC 7422. Concurrently, we can also use "NUMA nodes per socket" (NPSx)" to define the memory interleave. However, if we have "ACPI SRAT L3 As NUMA Domain" Disabled, then in addition to "NUMA nodes per socket" (NPSx)" providing the memory interleave info, it will dictate the number of NUMA Nodes.

- **Memory related settings**

- *Memory Clock Speed*: Memory and IO Die each have corresponding clock. We can obtain lower latencies by adjusting memory speeds for minimum latency, as describe in [3.2 Memory](#) above.

- **Power related settings**

- *Power Determinism* - may increase power to die, to maximize core performance. More on power/performance determinism can also be found here: <https://www.amd.com/system/files/2017-06/Power-Performance-Determinism.pdf>
- *CPPC* - Collaborative Processor Performance Control allows OS control over processor boost. See [4.1 Collaborative Processor Performance Control \(CPPC\)](#) below.

- **Core related settings**

- *SMT*: Allowing for 2 execution threads per core. Turning on SMT will provide further gains. However, because there are resources within the core that are shared 2x performance is usually not expected.

- **I/O related settings**

- *Local APIC Mode* - depending upon Windows Server OS, x2APIC may be suggested. See [4.3 x2APIC](#), below.

For further information on BIOS Settings by workload, see:

Workload Tuning Guide for AMD EPYC™ 7002 Series Processor Based Servers
https://developer.amd.com/wp-content/resources/56745_0.80.pdf

4.1 Collaborative Processor Performance Control (CPPC)

Collaborative Processor Performance Control is defined in ACPI and provides a mechanism for the OS to request increased performance levels from the hardware. Windows provides 3 Power Plans:

- Balanced
- High Performance
- Power Saver - which will potentially exploit boost.

The default Windows power is Balanced. If you want to enable Collaborate Processor Performance Control, set the Power Option to High Performance.

CPPC also supports an autonomous mode, which AMD recommends, where hardware independently selects a performance level appropriate to the current workload.

For Performance Mode with CPPC use Windows' *Powercfg,exe tool*. Open Window Command as Administrator:

```
Powercfg /setacvalueindex scheme_current sub_processor PROCTHROTTLEMIN 100
```

```
Powercfg /setacvalueindex scheme_current sub_processor PROCTHROTTLEMAX 100
```

```
Powercfg /setacvalueindex scheme_current sub_processor perfautonomous 1
```

```
Powercfg /setacvalueindex scheme_current sub_processor PERFEPP 0
```

```
Powercfg /setacvalueindex scheme_current sub_processor PERFBOOSTMODE 4
```

```
Powercfg /setactive scheme_current
```

4.2 NUMA

When there are more than 64 processors, Windows associates processors into groups. The Windows scheduler sees each of the processor groups as a single entity. Currently, 64 logical processors are the maximum size of a Windows Processor Group.

EPYC 7002 series has SKUs that include 64 cores, or 128 logical processors with SMT. Since all the processors in a single-socket/128-logical-processor NUMA node cannot fit completely within a single Windows Processor Group, Windows creates a (virtual) secondary node to hold the additional processors. Windows will try and assign processors that are closest to each other into the same Processor Group.

Windows currently supports a maximum of 64 hardware threads per NUMA Node. Therefore, one may wish to be aware of BIOS settings which allow for more than 64 logical processors per node. For example, in a 2 socket EPYC 7742 system with SMT on, if (NPSx), is 0 or 1 this would mean 256 or 128 threads per NUMA Node respectively. While Windows Server natively may work fine with minor cosmetic issues, having such a configuration with Hyper-V can have more serious consequences, and not recommended.

Regardless of NPS settings, applications will need to be multi-group aware to take advantage of all the processors (otherwise their affinity will be to a single processor group).

4.3 x2APIC

On systems with more than 255 logical processors such as 2 socket servers using the EPYC 7742, i.e. 2P 64c w/SMT and which are running Windows Server 2019; x2APIC mode must be used to take advantage of all the cores. x2APIC also provides a potentially more efficient mechanism for interrupt delivery and is necessary for 256 processor support.

Windows Server 2012.r2, and Windows Server 2016 currently support only xAPIC mode on EPYC 7002 series, and thus are slightly limited in their OPN/processor support.

Server Operating System	WS2012	WS2016	WS2019
Interrupt Mechanism Support	xAPIC	xAPIC	x2APIC
Max. OPN Core Count / Logical Processor Support	2 x 48c 192 LPs	2 x 48c 192 LPs	2 x 64c 256 LPs

Microsoft did a full media refresh for Windows Server 2019 in September. This includes x2APIC support for EPYC 7002. Use the latest updated Windows Server 2019 ISO (September 2019 or beyond) provided by Microsoft for new installs. This full media refresh version corresponds to the Windows Cumulative Update you would get for September.

However, if you are installing Windows Server 2019 GA or prior-to-September 2019 refresh, do the following:

1. Disable SMT in BIOS, also ensure xAPIC (not x2APIC) mode is enabled in BIOS.
2. Install WS2019 from original media.
3. Re-run Windows Update until all Cumulative Windows Updates have been installed. This action may require several reboots/updates.
4. Once all Windows Updates have been applied, and system boots into Windows without needing further updates, nor reboots. Shutdown system
5. Re-enable SMT in BIOS, re-enable x2APIC in BIOS.

Chapter 5 Supported Microsoft Windows Server Versions

Supported Windows Server Operating Systems include:

- Windows Server 2019, September 2019 full refresh media or latest update, which includes x2APIC support (see section 4.3 x2APIC). This is the recommended release.
- Windows Server 2016; Does not contain x2APIC support
- Windows Server 2012.R2; Mainstream support ended 10/2018

You can obtain the latest Windows Server versions from any one of the links listed below:

- Evaluation copy from the Evaluation Center [here](#).
- Visual Studio Subscription (formerly MSDN or Microsoft Developer Network) September 2019 or beyond [here](#).
- Microsoft Partner Network (MPN) [here](#).
- If you have valid Software Assurance, this new version of Windows Server 2019 can be downloaded from the Volume Licensing Service Center (VLSC) [here](#).

For more information on *[Windows Server support and installation instructions for the AMD Rome family of processors](#)*.

Chapter 6 OS Tuning and Tools

This section provides settings and links to tools that may help with respect to performance and analysis of Windows Server on Epyc 7002 Series Processors.

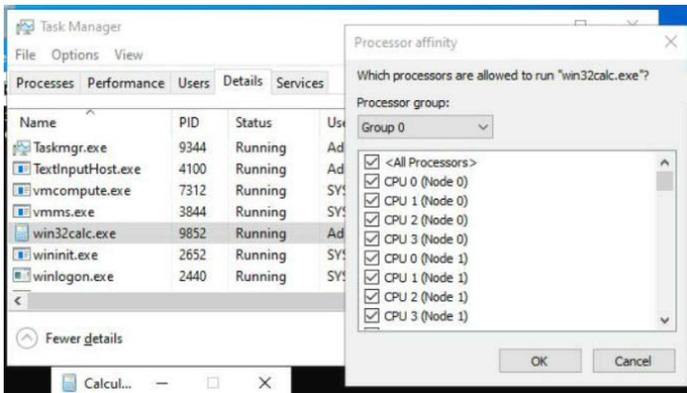
6.1 OS Settings

6.1.1 Setting Processor Affinity

Using Task Manager

To set processor affinity under Windows, run Task Manager as Administrator and do the following:

1. Select the App you wish to affinitytize.
2. Right-click and go to Details, to get process
3. Right-click and select 'Set Affinity'



Using PowerShell

- Obtaining affinity:
PowerShell "*Get-Process appname | Select-Object ProcessorAffinity*"
- Setting Affinity:
PowerShell "*\$Process = Get-Process app; \$Process.ProcessorAffinity=mask-value*"

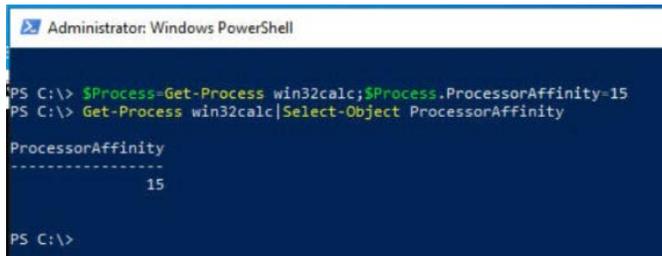
For example, running *get-process* will show you all currently running processes.

```
PS C:\> Get-Process wincalc32 | Select-Object ProcessorAffinity
```

ProcessorAffinity will return a bitmask representing the processors that the threads in the associated process can run on. So, to run a process on first CCX (assuming SMT is off)

```
PS C:\> $Process = Get-Process win32calc; $Process.ProcessorAffinity=15
```

```
PS C:\> Get-Process win32calc | Select-Object ProcessorAffinity
```



```
Administrator: Windows PowerShell
PS C:\> $Process=Get-Process win32calc;$Process.ProcessorAffinity-15
PS C:\> Get-Process win32calc|Select-Object ProcessorAffinity

ProcessorAffinity
-----
15

PS C:\>
```

Now Calculator will run only on first 4 Logical Processors, as this is a mask.

6.2 Managing Microsoft Hyper V

Hyper-V is a Type-1/bare-metal hypervisor, running directly on the underlying hardware. Hyper-V is the same hypervisor used for virtualization in Windows 10, Windows Server, and Azure.

Hyper-V provides isolated guest operating system environments, run on a single server platform. Each such isolated partition is given its own resources. Hyper-V has a parent, or management Root Partition, which is a virtual machine partition that has unique access and increased privileges. This root partition creates the requested isolated child partitions.

General recommendation is not to run other applications on Root Partition.

To enable Hyper-V on Windows Server via Powershell, the following cmdlet may be used:

```
Install-WindowsFeature -Name Hyper-V -IncludeManagementTools -Restart
```

Minroot allows user to specify the maximum number of processors used by the root partition, or specify the available number of processors per node to N using the following command:

```
C:\> bcdedit /set hypervisorrootproc  $N$ 
```

Cpugroups.exe is a tool from the Hyper-V Team which allows allocation of processing resources at a much more granular level. Learn more about VM CPU tools at:

<https://docs.microsoft.com/en-us/windows-server/virtualization/hyper-v/manage/manage-hyper-v-cpugroups>

OS Integration Services, provide enlightened guest OS drivers for Hyper-V and will improve performance of the guest OS while it is running on Hyper-V.

6.3 Other Tools & Resource

6.3.1 Windows Performance Monitoring Tools & Resources

The Operating System includes the following resources:

- **Windows Resource Monitor:** Monitor CPU, memory, disk, and network usage real-time
- **Windows Performance Monitor:** View OS\Hyper-V performance counters real time or saved from Performance recorder. Windows Performance Monitor feature is now also part of *Windows Admin Center* so multiple servers can be tracked (also see logman.exe).
- **Windows Assessment and Deployment Kit (ADK)** includes a variety of useful performance tools including:
 - **Windows Performance Recorder (WPR) / xPerf:** a recording utility based on Event Tracing for Windows (ETW).
 - **Windows Performance Analyzer** creates graphs and data tables of above recorded events.

6.3.2 AMD Tools

uProf for Windows: Is profiling tool used that can be used to monitor system metrics and performance counters