NUMA Configuration settings on AMD EPYC 2nd Generation

Tech Note by

Mohan Rokkam

Summary

In multi-chip processors like the AMD-EPYC series, differing distances between a CPU core and the memory can cause Non-Uniform Memory Access (NUMA) issues.

AMD offers a variety of settings to help limit the impact of NUMA. One of the key options is called Nodes per Socket (NPS). This paper talks about some of the recommended NPS settings for different workloads.

Introduction

AMD Epyc is a Multi-Chip Module processor. With the 2nd generation AMD EPYC 7002 series, the silicon package was modified to make it a little simpler. This package is now divided into 4 quadrants, with up to 2 Core Complex Dies (CCDs) per quadrant. Each CCD consists of two Core CompleXes (CCX). Each CCX has 4 cores that share an L3 cache. All 4 CCDs communicate via 1 central die for IO called I/O Die (IOD).

There are 8 memory controllers per socket that support eight memory channels running DDR4 at 3200 MT/s, supporting up to 2 DIMMs per channel. See Figure 1 below:
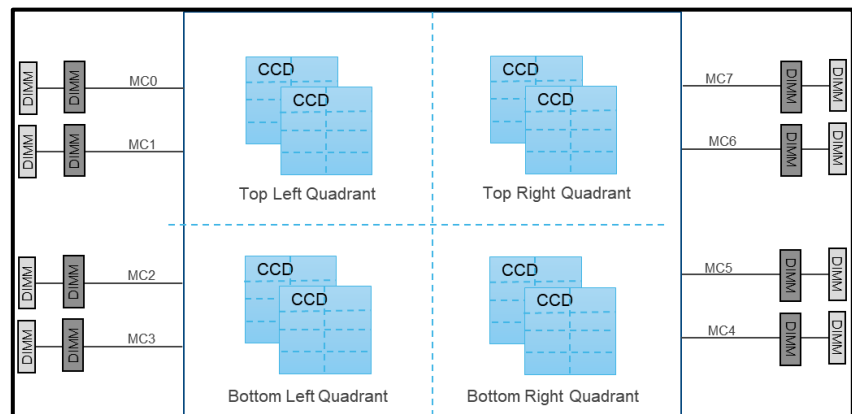

Figure 1 - Illustration of the ROME Core and memory architecture

With this architecture, all cores on a single CCD are closest to 2 memory channels. The rest of the memory channels are across the IO die, at differing distances from these cores. Memory interleaving allows a CPU to efficiently spread memory accesses across multiple DIMMs. This allows more memory accesses to execute without waiting for one to complete, maximizing performance.

NUMA and NPS

Rome processors achieve memory interleaving by using Non-Uniform Memory Access (NUMA) in Nodes Per Socket (NPS). The below NPS options can be used for different workload types:

DELL Technologies

- **NPS0** – This is only available on a 2-socket system. This means one NUMA node per system. Memory is interleaved across all 16 memory channels in the system.
- **NPS1** – In this, the whole CPU is a single NUMA domain, with all the cores in the socket, and all the associated memory in this one NUMA domain. Memory is interleaved across the eight memory channels. All PCIe devices on the socket belong to this single NUMA domain.
- **NPS2** – This setting partitions the CPU into 2 NUMA domains, with half the cores and memory in each domain. Memory is interleaved across 4 memory channels in each NUMA domain.
- **NPS4** – This setting partitions the CPU into four NUMA domains. Each quadrant is a NUMA domain, and memory is interleaved across the 2 memory channels in each quadrant. PCIe devices will be local to one of the 4 NUMA domains on the socket, depending on the quadrant of the IOD that has the PCIe root for the device.

Note: Not all CPUs support all NPS settings

Recommended NPS Settings

Depending on the workload type, different NPS settings might give better performance. In general, NPS1 is the default recommendation for most use cases. Highly parallel workloads like many HPC use cases might benefit from NPS4. Here is a list of recommended NPS settings for some key workloads. In some cases, benchmarks are listed to indicate the kind of workloads.

Workload Name	NPS Setting
General-Purpose Workloads	**NPS1**
SPEC CPU 2017	NPS1
SPECjbb 2015	NPS4/2
SPECpower ssj 2008	NPS4
I/O Intensive	NPS1
Virtualization	**NPS1/4**
Vmmark 3	NPS1
TPCx-V	NPS4
Containers	NPS4
Database and Analytics	**NPS1/4**
HammerDB	NPS1
Hadoop	NPS4
TPCx-IOT	NPS4
HPC and Telco	**NPS4/2**
HPC	NPS4/2
OpenStack NFV	NPS2
OpenStack for RealTime Kernel (NFV)	NPS2
EDA	NPS4

Figure 2 - Table of recommended NPS Settings depending on workload

For additional tuning details, please refer to the Tuning Guides shared by AMD [here](#).
For detailed discussions around the AMD memory architecture, and memory configurations, please refer to the [Balanced Memory Whitepaper](#)

DELLTechnologies

In Conclusion

The Dell AMD EPYC based servers offer multiple configuration options to optimize memory performance. Based on workload, choosing the appropriate NPS setting can help maximize performance.

PowerEdge DfD Repository
For more technical learning

Contact Us
For feedback and requests

Follow Us
For PowerEdge news

DELLTechnologies