Bringing AMD to the Datacenter

Tech Note by

Michael Bennett

Summary

IT administrators are excited to reap the benefits of these high core count processors but are unsure of how to best incorporate a second x86 architecture in their datacenter.

This Direct from Development tech note discusses compatibility between Intel and AMD processors, workload migrations and scheduler support for heterogenous environments.

Introduction

Today we have more capabilities than ever before coming from our IT departments. They operate datacenters and manage workloads that provide collaboration, insight and enable operations. As customers look to expand these services, they often ask about the 2nd generation of AMD EPYC processors (Rome), how their workload will perform, and what it means to operate a datacenter where two different x86 vendors are present. In this Direct from Development tech note we briefly cover the AMD EPYC 7002 series of processors before going over best practices and key considerations operating two x86 instruction sets in your datacenter.

AMD EPYC 7002 series

AMD EPYC 7002 series processors have the highest core density currently offered in the x86 market with the AMD EPYC 7742 containing 64 cores. In addition to this extreme core count, the EPYC CPU lineup also offers several SKUs that have configurations optimized for a specific workload such as the recently announced 7Fxx series which support up to 32 cores at boost frequencies of 3.9Ghz, a half gigahertz increase over what is offered by other EPYC CPUs. The 7Fxx lineup of CPUs is targeted at hyper converged infrastructure, high performance computing and relational database management systems such as SQL.

Deploying and Managing EPYC in the Data Center

Introducing new hardware in the datacenter requires careful consideration. IT operations teams will likely want to first test and validate several workloads as a prototype before deploying into the production environment, and tooling + procedures will need to be put in place to manage the lifecycle of these servers. For Dell EMC customers this is a relatively seamless process – both the Intel and AMD line of 14G PowerEdge servers contain iDRAC9. This enables operations teams to use the same familiar interface to deploy, manage and secure all PowerEdge servers. Customers who use OpenManage Enterprise will find that it puts all of the systems in a single pane of glass for management and updates and provides granularity to support different firmware baselines.

Even with a universal management framework through iDRAC9 there are other considerations for how to cluster these systems, the impact on

workload scheduling and how to migrate workloads. Before discussing that, a bit on iDRAC telemetry. As you deploy new systems and manage a diverse set of applications it becomes increasingly difficult to monitor the performance and health of these workloads. With the iDRAC9 Data Center edition you can stream telemetry information into databases such as Graphana or Prometheus. With all systems streaming telemetry into a single database you can better understand the health of workloads in your datacenter and respond to issues quicker. As you make decisions about which applications to deploy on AMD systems, this data can be used to compare metrics that are relevant to application performance and allow you to make data driven decisions on where to deploy.

Mixed CPU Cluster Compatibility

Most customers who are purchasing their first AMD system have an existing footprint of x86 processors. This drives concerns and questions from our customers about the compatibility of their clusters across the two different x86 instruction sets. Services that operate in a clustered fashion are generally a homogenous configuration for ease of management and to provide consistent performance. As we move to a world that consists of not just the core datacenter but also the cloud and edge, it is becoming increasingly common for services to operate a collection of clusters, the configuration of which will be optimized for both site (edge, core, cloud) and function.

There are several reasons IT administrators may want to avoid a mixture of processor generations and vendors in a clustered system. One is that migration tools such as vMotion Live Migration do not work due to differences in the instruction set architecture between x86 vendors. This doesn't apply to generations of processors from the same vendor if supported by Enhanced vMotion Compatibility (EVC) mode but enabling this has a performance cost [1]. Another reason is that maintenance windows will be hard to keep up with when you have BIOS and microcode updates coming from two different vendors. Multiple generations of processors can make updates difficult because of the additional effort that is needed to complete increased amount of testing that must be done, which is at least one for each generation of processor.

IT leaders now have more options for x86 processors and should evaluate each new system deployment and consider if Intel or AMD processors would be optimal for the workload and environment. For example, a large network service provider has deployed AMD due to the benefits they saw in the large L3 cache present on AMD EPYC processors. Conversely a major retail customer of Dell EMC looking to deploy AI services at 400 edge locations selected Intel Xeon Cascade Lake processors for their support of DLBoost, a technology that allows twice as many AI inference operations at 8-bit precision.

Workload Migration

IT administrators move workloads between systems and clusters for a variety of reasons, some planned and some unplanned. As touched on earlier, some workload migration techniques such as VMware vMotion Live Migration are not compatible between Intel and AMD systems. This is an important limitation that must be accounted for when considering high availability, fault capacity and how you conduct general day to day activities such as cluster balancing and planned maintenance.

In the event it is necessary to migrate a workload between Intel and AMD processor-based systems, there are a few options. For VMware environments, while vMotion Live Migration does not work, you can do a cold migration after shutting down the virtual machine. In other cases, you can use application specific migration techniques. Most applications support backup/restore functionality that can be used in conjunction with tools like load balancers to allow migration with little or no downtime.

D≪LL Technologies

Heterogenous Scheduling

Workload schedulers such as Slurm, kube-scheduler and Hadoop all support various methods of gaining awareness and preference when determining which CPU to schedule a workload. With Slurm, nodes are divided into partitions and it is trivial to separate Intel and AMD systems into different partitions. Hadoop 3 and kube-scheduler support a variety of features that enable exposing CPU information and grouping similar systems using labels, namespaces and roles.

For now, schedulers still require explicit definition of whether to run a workload AMD or another processor if you want consistent performance, though policy-based rules can provide default definitions. Capacity planning is also challenging when attempting to do heterogenous scheduling, and because of this complexity it is a good idea to avoid heterogenous scheduling for real time applications and instead only use this for non-real-time applications such as data processing and batch workloads.

In Conclusion

Dell EMC offers several PowerEdge servers that support AMD EPYC 7002 series processors. The PowerEdge R6515 and R7515 support a single 7002-series processor and the PowerEdge C6525, R6525 and R7525 support two processors. The C6525 for those unfamiliar with it is a 2U server with 4 compute sleds. Configured with two 64-core AMD processors per sled this chassis can provide 512 CPU cores in a 2U footprint. All the PowerEdge models listed also have support for PCIe 4.0, though the number of expansion slots varies by model.

R6515	R7515	R6525	R7525	C6525
Single-socket 1U rack server offers peak performance and excellent TCO	Highly scalable 2U rack server delivers performance and outstanding TCO	Highly configurable 1U rack server delivers outstanding balanced performance for dense compute	Highly adaptable 2U rack server brings powerful performance and flexible configuration	Compute-dense server sled accelerates data center performance to tackle diverse HPC applications
1S RACKS		2S RACKS		C-SERIES

To effectively make use of these systems' capacity, administrators should use homogenous configurations for clustered services and ensure they have a tested procedure to migrate their workloads. Heterogenous collections of systems are best suited for batch workloads. Finally, customers can optimize the performance of their workload by using processor SKUs that have specific features and capabilities for enhanced performance.

https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/techpaper/vmware-vsphere-evc-performance-white-paper.pdf

PowerEdge DfD Repository
For more technical learning

Contact Us
For feedback and requests

Follow Us
For PowerEdge news

D≪LL Technologies