Prepare images for machine learning faster with servers powered by AMD EPYC™ 75F3 processors

A server cluster with 3rd Gen AMD EPYC processors achieved higher throughput and took less time to prepare images for classification than a server cluster with 3rd Gen Intel Xeon Platinum 8380 processors

Kubernetes® environments can make it easy to deploy, scale, and manage the image preprocessing phase of machine learning workflows. Choosing 3rd Gen AMD EPYC™ processors could help you save time during that phase and do more work by completing more image processing tasks.

We compared the performance of two four-node clusters: one comprising Supermicro AS-1124US-TNRP servers with 32-core 3rd Gen AMD EPYC 75F3 processors, and one comprising Supermicro SYS-620U-TNR servers with 40-core 3rd Gen Intel® Xeon® Platinum 8380 processors. We used a synthetic workload that we designed for Kubernetes containers, which emulates image processing tasks typical of the early phases of an image-based machine learning workflow. The clusters used VMware® vSphere® 7.0 Update 2, VMware vSAN™ pooled storage, and VMware Tanzu™ Kubernetes Grid Service, which integrates Tanzu Kubernetes support directly into vSphere. The AMD EPYC processor-based cluster handled 1,720 images in 7.2 percent less time than the cluster powered by Intel Xeon processors, processing images at a 7.9 percent higher frames per second (FPS) rate. We also found that the AMD EPYC processor-based cluster could offer a 30.2 percent lower hardware and support cost than the Intel Xeon Scalable processor-based cluster.

Achieve a 7.9% higher throughput rate
while processing more frames per second*

Prepare images in 7.2% less time
while running a workload of 1,720 images*

Get a 30.2% lower cost
on hardware and support*

**on the AMD EPYC processor-powered servers we tested vs. the Intel Xeon processor-powered servers we tested*

How we approached testing

We compared the following four-node clusters:

- Supermicro SYS-620U-TNR servers powered by Intel Xeon Platinum 8380 processors
    - For one server, the total cost of hardware plus three years of labor and support and a one-year CRS warranty was $29,908.90—a total of $119,635.60 for a four-node cluster[1]
- Supermicro AS-1124US-TNRP servers powered by AMD EPYC 75F3 processors
    - For one server, the total cost of hardware plus three years of support and labor and a one-year CRS warranty was $20,870.90—a total of $83,483.60 for a four-node cluster[2]

Other than the processors, we configured the server clusters identically. Each of the servers in both clusters had a 240GB 6Gbps SATA SSD to use for the hypervisor and three PCIe® 4.0 NVMe™ SSDs for the vSAN storage. We also equipped each server with 1,024 GB of PC4-3200 RAM across 16 memory modules.

We configured a vSAN datastore on each cluster with a single disk group comprising one 3.84TB NVMe SSD for cache and two 3.84TB NVMe SSDs for capacity per server. The vSAN datastore served as shared storage for our Tanzu Kubernetes environment. The vSphere with Tanzu environments had one worker node per server. To account for the core count differences between the two processors in our comparison—that is, 32 cores per AMD EPYC processor and 40 cores per Intel Xeon processor—we assigned a different number of CPU resources to each worker based on the CPU architecture. In the cluster powered by 3rd Gen AMD EPYC processors, each worker node had 56 vCPUs and 512 GB of memory fully reserved. In the cluster with 3rd Gen Intel Xeon Scalable processors, each worker node had 70 vCPUs and 512 GB of memory fully reserved.

About VMware vSAN

For organizations looking to reduce the complexity and footprint of their data center, hyperconverged infrastructure (HCI) can help. As part of their HCI portfolio, VMware offers software-defined storage with vSAN that eliminates the need for bulky, expensive, external arrays and instead brings compute and storage resources together.

According to VMware, vSAN is "an enterprise-class storage virtualization software that provides the easiest path to HCI and hybrid cloud."[3] To learn more about VMware vSAN, visit https://www.vmware.com/products/vsan.html.

During testing, both clusters averaged 86 percent CPU utilization. For more details about our configurations, testing methodologies, and CPU utilization, see the science behind the report.

Less time to process images

To use image-based machine learning algorithms, organizations must first prepare their data for analysis. For manufacturing organizations using the algorithms during assembly, for example, this might include images of various products taken at multiple angles. Our workload mimicked simple image processing tasks that a company might run on Kubernetes with vSphere with Tanzu during this preparation phase, such as reading images from storage, scaling the images to a machine-learning-friendly size, transposing the images, and converting the images to grayscale. (For more on our custom workload, see the science behind the report.)

By sorting and preparing images more quickly, organizations could get insights sooner. As Figure 1 shows, the cluster of AMD EPYC 75F3 processor-based servers took 107.2 seconds to process 1,720 images. That was 7.2 percent less than the cluster of servers backed by the Intel Xeon Platinum 8380 processors, which needed 115.6 seconds to complete the workload.

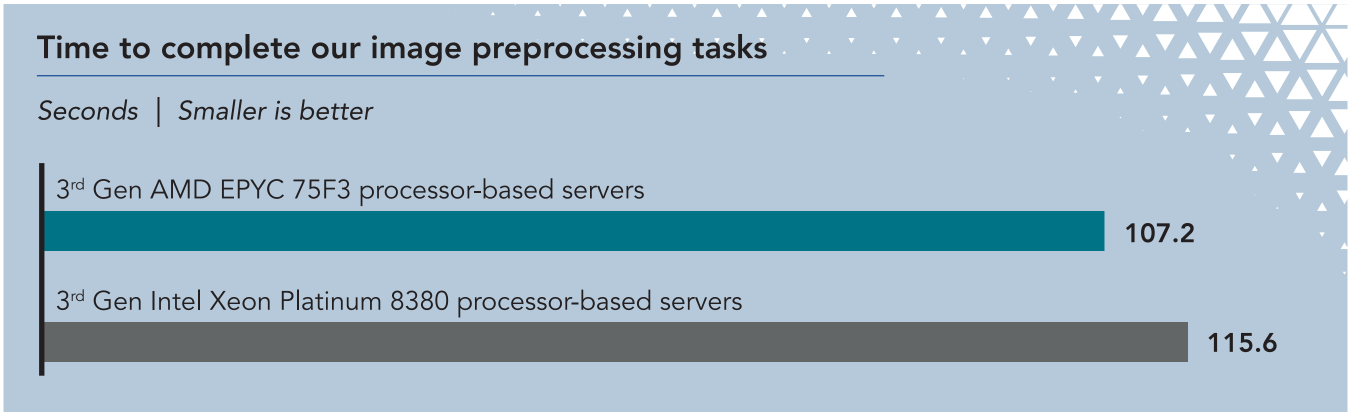

Figure 1: The time in seconds that each cluster needed to complete the image preprocessing tasks on 1,720 images. Less time is better. Source: Principled Technologies.

About VMware vSphere with Tanzu

We used vSphere with Tanzu to run our containerized image processing workload. VMware states that vSphere with Tanzu "bridges the gap between IT and developers for cloud-native apps on-premises and in the cloud."[4] Tanzu Kubernetes Grid Service is part of the Kubernetes-focused VMware Tanzu portfolio, which enables organizations to "build, run and manage modern apps on any cloud—and continuously deliver value to your customers," as well as to "simplify multi-cloud operations and free developers to move faster with easy access to the right resources," according to VMware.[5] For more information, visit https://tanzu.vmware.com/tanzu.

Real-world benefits for manufacturers

Machine learning algorithms can help companies perform quality control inspections, ensure product safety compliance, and verify correct assemblage. Saving time during any part of those workflows could translate to reaching business goals sooner, improving productivity, and reducing production costs. During the image preprocessing phase for any part of the manufacturing process, your company could save time by running workloads in a vSphere, vSAN, and Tanzu Kubernetes environment on 3rd Gen AMD EPYC 75F3 processor-based servers. We saw better performance from a cluster of these servers using this environment, processing images in 7.2 percent less time and at a 7.9 percent higher FPS rate than the servers powered by Intel Xeon Platinum 8380 processors.

About VMware vSphere 7.0 Update 2

We used VMware vSphere 7.0 Update 2 for our test environment. According to VMware, this vSphere 7.0 update includes a CPU scheduler that is optimized for AMD EPYC processors, as it is "designed to take advantage of the multiple last-level caches (LLCs) per CPU socket offered by the AMD EPYC processors." VMware also states that vSphere 7.0 Update 2 with the CPU scheduler can achieve near-optimal performance on most applications and benchmarks on AMD EPYC processors.[6]

For more information, visit https://www.vmware.com/products/vsphere.html.

A higher FPS processing rate

Figure 2 shows the image processing rate that each solution achieved during our testing. The AMD EPYC 75F3 processor-based cluster offered a rate of 15,537 FPS. That's an increase of 7.9 percent compared to the Intel Xeon Platinum 8380 processor-powered cluster, which processed the images at a rate of 14,388 FPS.

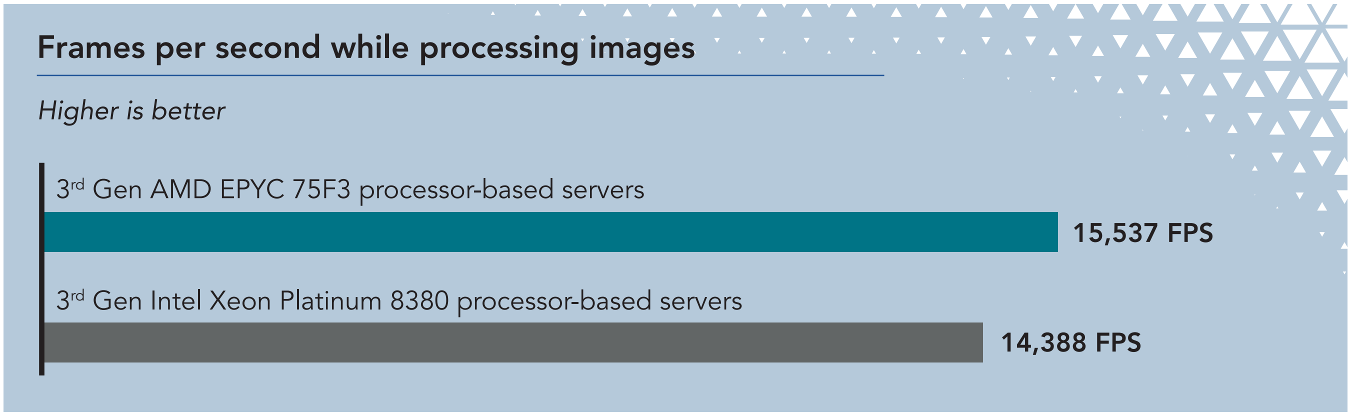

Figure 2: The rate of FPS each solution delivered during our 1,720-image preprocessing workload. Higher numbers are better. Source: Principled Technologies.

About AMD EPYC 75F3 processors

Part of the third-generation EPYC 7003 Series, the AMD EPYC 75F3 processor has 32 cores and 64 threads of computing power. According to AMD, the processor features PCIe 4.0 I/O connectivity, supports up to eight DDR4 memory channels per socket, and is well-suited for workloads such as VDI and HCI.[7] The third generation of AMD EPYC processors can also offer AMD Infinity Guard security features, such as Secure Encrypted Virtualization (SEV), Secure Nested Paging (SEV-SNP), Secure Memory Encryption (SME), and more.[8] To learn more about the EPYC 75F3 processor, visit https://www.amd.com/en/products/cpu/amd-epyc-75f3.

Conclusion

We compared the image preprocessing for machine learning performance of two server solutions powered by different processors: the AMD EPYC 75F3 and the Intel Xeon Platinum 8380. When we ran our containerized image preprocessing workload on each cluster, we found that the 3rd Gen AMD EPYC processor-powered cluster needed 7.2 percent less time to completely process 1,720 images while handling 7.9 percent more frames per second. This kind of performance shows that an organization using vSphere with Tanzu and vSAN to run their Kubernetes workloads could get better image processing performance with the servers powered by AMD EPYC 75F3 processors. Our price analysis also demonstrated that the hardware and support costs for the AMD EPYC 75F3 processor-powered Supermicro AS-1124US-TNRP server cluster was 30.2 less than the Intel Xeon Platinum 8380 processor-powered Supermicro SYS-620U-TNR server cluster.

---

1. We received a quote from Supermicro on February 2, 2022 for the hardware and support cost of the server minus drive costs. To arrive at the total cost, we added this amount to a drive cost quote we had received from Supermicro on August 9, 2021.
2. We received a quote from Supermicro on February 2, 2022 for the hardware and support cost of the server minus drive costs. To arrive at the total cost, we added this amount to a drive cost quote we had received from Supermicro on August 9, 2021.
3. VMware, "What is vSAN?" accessed February 15, 2022, https://www.vmware.com/products/vsan.html.
4. VMware, "vSphere with Tanzu," accessed February 7, 2022, https://www.vmware.com/products/vsphere/vsphere-with-tanzu.html.
5. VMware, "VMware Tanzu Solution Brief," accessed February 7, 2022, https://d1fto35gcfffzn.cloudfront.net/tanzu/VMware-Tanzu-Solution-Brief-0121.pdf.
6. VMware, "Performance Optimizations in VMware vSphere 7.0 U2 CPU Scheduler for AMD EPYC Processors," accessed Feburary 7, 2022, https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/techpaper/performance/vsphere70u2-cpu-sched-amd-epyc.pdf.
7. AMD, "AMD EPYC™ 75F3," accessed February 7, 2022, https://www.amd.com/en/products/cpu/amd-epyc-75f3.
8. Server OEMs and cloud providers must enable AMD Infinity Guard features for use. In addition, security features can vary by AMD EPYC processor generations. Learn more about Infinity Guard at https://www.amd.com/en/technologies/infinity-guard.

Read the science behind this report at https://facts.pt/pagvE8O ▶

Principled Technologies®

Facts matter.®

Principled Technologies is a registered trademark of Principled Technologies, Inc.
All other product names are the trademarks of their respective owners.
For additional information, review the science behind this report.

This project was commissioned by AMD.