Complete artificial intelligence workloads faster using Microsoft Azure virtual machines featuring 3rd Gen Intel Xeon Scalable processors

The Edsv5-series VMs achieved greater AI performance than Edsv4-series VMs featuring older Intel Xeon processors

The number of companies using artificial intelligence is growing, with a 2022 survey finding that more than 95 percent of organizations have launched AI initiatives.[1] They are using AI to generate automated insights that drive decision-making, to improve the customer experience, to perform pattern and image recognition, and to solve many other business problems. With many organizations opting to run compute-intensive AI applications in the cloud, cloud service providers (CSP) offer a range of powerful virtual machine options. Because performance can vary dramatically from one VM to another—even those from the same CSP—customers who want to maximize their cloud investment must do their homework.

Using a schema we derived from TPCx-AI benchmark that includes use cases targeting multiple aspects of artificial intelligence, we measured the performance of two types of Microsoft Azure VMs: Edsv5-series VMs featuring 3rd Gen Intel® Xeon® Scalable processors and Edsv4-series VMs enabled by older Intel Xeon processors. The latest-generation Edsv5-series VMs outperformed the previous-generation VMs by 34 percent overall. This improvement could allow your organization to execute more AI work with a given number of VMs, which could mean getting valuable insights earlier. There is also a financial advantage. At the time of writing, the cost of the two VM series is the same, which means the stronger performance of the newer VMs could reduce both VM compute time and costs.[2]

This project was commissioned by Intel.

About our testing

We tested two generations of Azure VMs:

- Edsv5-series VMs featuring 3rd Gen Intel Xeon Platinum 8370C processors
- Edsv4-series VMs featuring 2nd Gen Intel Xeon Platinum 8272CL processors

We used five 48vCPU VMs from each series as worker nodes. We also used a 16vCPU VM from each series as a head node. We ran a workload derived from the TPCx-AI benchmark to show the performance customers can expect to see using VMs with newer Intel processors vs. VMs with previous-generation processors. The benchmark provides one metric that combines machine learning and deep learning for 10 end-to-end use cases.[3] We tested all VMs in the East US region. For additional configuration information, see the science behind the report.

About the TPCx-AI-like workload we used

To measure the artificial intelligence performance of the VMs, we used a workload we derived from the TPCx-AI benchmark. TPCx-AI runs on CDP (Cloudera data platform), a Hadoop-based platform. According to documentation, the benchmark "measures the performance of an end-to-end machine learning or data science platform. The benchmark development has focused on emulating the behavior of representative industry AI solutions that are relevant in current production datacenters and cloud environments."[4] Because we derived our workload from the TPCx-AI benchmark, our results are not comparable to published TPCx-AI results.

The workload includes 10 use cases and generates both individual and overall scores. We report only the overall scores in this document. For more information, visit https://www.tpc.org/tpcx-ai/default5.asp.

What we learned

We evaluated each VM with the benchmark tool three times, and we report the median result. In every run, the Azure Edsv5-series VM with 3rd Gen Intel Xeon Scalable processors outperformed the Edsv4-series VM with 2nd Gen Intel Xeon Scalable processors. As Figure 1 shows, the median overall performance score on the TCPx-AI-derived benchmark was 34 percent higher for the 48vCPU Edsv5-series VM than for its Edsv4-series counterpart.

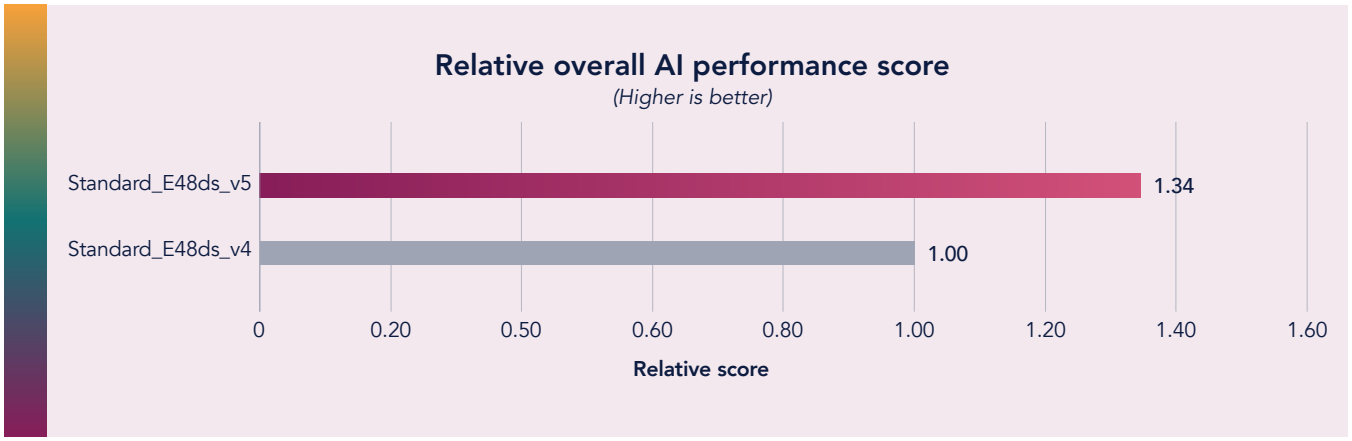

Figure 1: Relative overall AI performance scores for Edsv5-series and Edsv4-series Azure VMs. Higher is better. Source: Principled Technologies.

About Microsoft Azure Edsv5-series VMs

Azure Edsv5-series VMs feature 3rd Generation Intel Xeon Platinum 8370C processors. According to Microsoft, the VMs also offer the following specifications:[5]

- Up to 104 vCPUs and up to 672 GiB of RAM
- All-core turbo clock speed of up to 3.5GHz
- Intel Turbo Boost Technology
- Intel Advanced Vector Extensions 512 (Intel AVX-512)
- Intel Deep Learning Boost

To learn more, visit https://docs.microsoft.com/en-us/azure/virtual-machines/edv5-edsv5-series.

Conclusion

If your company is turning to the cloud for the compute power necessary to run artificial intelligence applications, selecting your virtual machines wisely is vital. Our testing compared the AI performance of two Microsoft Azure VM types. We learned that 48vCPU Edsv5-series VMs featuring 3rd Gen Intel Xeon Scalable processors outperformed their previous-generation counterparts by 34 percent overall despite the two VMs having the same price tag. The performance improvement you would experience by selecting the newer VMs could help your business by arriving at actionable insights more quickly and could help your bottom line by reducing VM uptime.

1. New Vantage Partners, "Data and AI Leadership Executive Survey 2022 Executive Summary of Findings," accessed September 16, 2022, https://c6abb8db-514c-4f5b-b5a1-fc710f1e464e.filesusr.com/ugd/e5361a_2f859f3457f24cff9b-2f8a2bf54f82b7.pdf.
2. Microsoft Azure, "Windows Virtual Machines Pricing," accessed September 20, 2022, https://azure.microsoft.com/en-us/pricing/details/virtual-machines/windows/#pricing.
3. TPC, TPCx-AI distinguishes itself with one metric combining machine learning and deep learning for end-to-end use cases (e.g., data sanitization/training/scoring). It also uses highly scalable data sets and includes a price/performance metric based on 1 year total cost of ownership [Tweet], Twitter, March 16, 2021, accessed September 20, 2022, https://twitter.com/tpcbenchmarks/status/1371937614303952896.
4. TPC, "TPCx-AI," accessed September 19, 2022, https://www.tpc.org/tpcx-ai/default5.asp.
5. Microsoft, "Edv5 and Edsv5-series," accessed September 19, 2022, https://docs.microsoft.com/en-us/azure/virtual-machines/edv5-edsv5-series.

Read the science behind this report at http://facts.pt/2Vl8uwU ▶

Principled Technologies®

Facts matter.®

This project was commissioned by Intel.