

# Get Real-Time Recommendations up to 3.48x Faster with Microsoft® Azure® Esv4 VMs

## Improve Wide & Deep Inference Performance with Azure Esv4 VMs featuring 2<sup>nd</sup> Gen Intel® Xeon® Scalable processors

Using a subset of machine learning—called deep learning—to make relationships between customer data can deliver real-time recommendations that help consumer locate what they seek. Utilizing wide linear models and deep neural networks to infer relationships between data, Wide & Deep workloads deliver real-time recommendations based on that data. Selecting Microsoft Azure Esv4 VMs enabled by 2<sup>nd</sup> Gen Intel® Xeon® Scalable processors over Esv3 VMs with previous-generation processors can improve Wide & Deep recommendation engine performance. The 2<sup>nd</sup> Gen Intel Xeon Scalable processor family features Intel Deep Learning Boost, which improves machine learning performance.

To determine which configuration offers better performance, independent third-party Principled Technologies tested Wide & Deep performance across three different VM sizes. Azure Esv4 VMs featuring Intel Xeon Platinum 8272CL processors handled up to 3.48x more samples per second than Esv3 VMs. With Esv4 VMs, organizations can deliver real-time recommendations based on the data they collect even faster, which can improve customer satisfaction and boost overall sales.

## Improve Deep Learning Performance on Small Instances

The faster your cloud VMs can infer meaningful relationships between data, the faster you can make recommendations to consumers. As Figure 1 shows, 8-vCPU Esv4 VMs enabled by 2<sup>nd</sup> Gen Intel Xeon Scalable processors outperformed 8-vCPU Esv3m VMs in a deep learning Wide & Deep benchmark test. The Intel Xeon processor-based VMs handled 3.48 times the samples per second that the previous-gen VMs did, which means they can process data and make recommendations faster.

### Relative Wide & Deep throughput at 8 vCPU

Samples per second | Higher is better

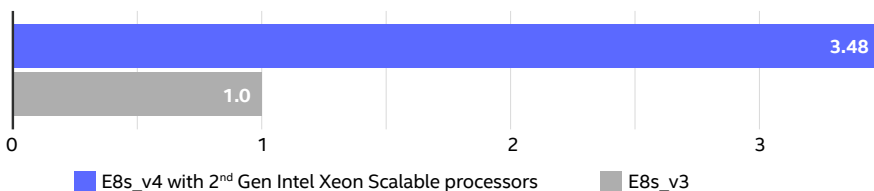


Figure 1. Relative results comparing the Wide & Deep benchmark performance of small (8-vCPU) Esv4 VMs vs. Esv3 VMs.

HiBench

**3.48x the samples per second with 8-vCPU Esv4 VMs**  
*vs. Esv3 VMs*

**3.23x the samples per second with 16-vCPU Esv4 VMs**  
*vs. Esv3 VMs*

**2.99x the samples per second with 64-vCPU Esv4 VMs**  
*vs. Esv3 VMs*

## Improve Deep Learning Performance on Medium Instances

Organizations with mid-sized datasets can also get improved deep learning inference performance by choosing VMs with newer processors. As Figure 2 shows, 16-vCPU Azure Esv4 VMs enabled by 2<sup>nd</sup> Gen Intel® Xeon® Scalable processors handled 3.23 times the samples per second in Wide & Deep tests compared to Esv3 VMs with previous-generation processors.

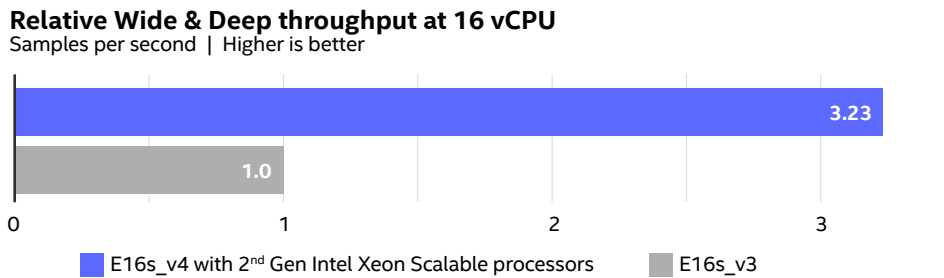


Figure 2. Relative results comparing the Wide & Deep benchmark performance of medium (16-vCPU) Esv4 VMs vs. Esv3 VMs.

## Improve Deep Learning Performance on Large Instances

Larger datasets that require larger VMs similarly benefit from choosing newer processor architecture for deep learning workloads. In tests, Esv4 VMs featuring 2<sup>nd</sup> Gen Intel Xeon Scalable processors handled 2.99 times the samples per second using the Wide & Deep benchmark test (see Figure 3).

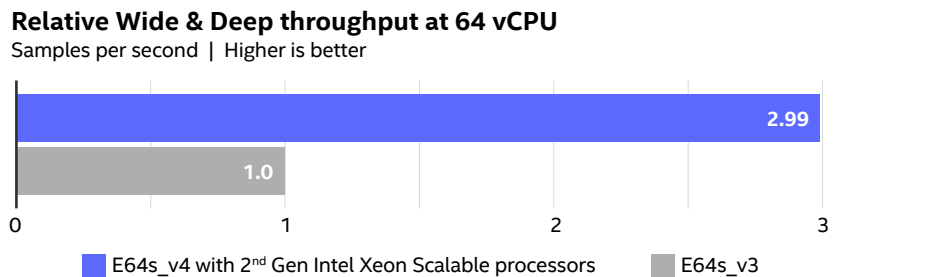


Figure 3. Relative results comparing the Wide & Deep benchmark performance of large (64-vCPU) Esv4 VMs vs. Esv3 VMs.

For datasets small, medium, and large, selecting Azure Esv4 VMs with 2<sup>nd</sup> Gen Intel Xeon Scalable processors over Esv3 VMs with previous-generation processors can boost deep learning performance to form meaningful relationships from data and make fast real-time recommendations to consumers.

### Learn More

To begin running your Wide & Deep workloads on Azure Esv4 Instances with 2<sup>nd</sup> Gen Intel Xeon Scalable processors, visit <http://intel.com/azure>.

For complete testing results, visit <http://facts.pt/YX3rsPQ>.



Performance varies by use, configuration and other factors. Learn more at <https://intel.com/benchmarks>.

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy. Your costs and results may vary.

Intel technologies may require enabled hardware, software or service activation.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others

Printed in USA 1021/JO/PT/PDF US001

