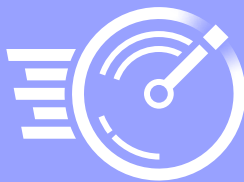


# Reduce Time to Complete Databricks Queries by up to 80% and Save up to 61% with Microsoft Azure Edsv4 VMs Featuring 2<sup>nd</sup> Gen Intel<sup>®</sup> Xeon<sup>®</sup> Scalable Processors



Databricks



**Up to 80% Less Time to Run Decision Support Queries with Photon-Enabled E8ds\_v4 VMs**

*vs. Older E8s\_v3 VMs*



**Up to 61% Lower Cost to Run Decision Support Queries with Photon-Enabled E8ds\_v4 VMs**

*vs. Older E8s\_v3 VMs*

## These VMs, with the Photon Vectorized Query Engine Enabled, Delivered Stronger Decision Support Workload Performance than Older VMs Featuring Previous-Generation Processors

Choosing the right hardware is essential to getting optimal performance for your decision support workloads. It might seem obvious that updated hardware delivers better performance, but it's not always clear just how much of an improvement your organization can expect by choosing the newer option and how much this extra performance costs. To explore these questions, we tested a decision support workload on a 20-node E8s\_v3 cluster with Databricks Runtime 9.0 to get a baseline performance metric. The older Esv3 series from Azure offer VMs with processors ranging from the Intel<sup>®</sup> Xeon<sup>®</sup> E5-2673 v4 to the Intel Xeon 8272CL. Each time you spin up a VM, Azure randomly assigns a processor. This means that a 20-node cluster could use a mix of CPU types, with some being as much as three Intel CPU generations behind the newest processors. For consistency, we ensured that all the E8s\_v3 VMs listed the same Intel Xeon Platinum 8171M processor when we started our tests. We then tested the same workload on a 20-node E8ds\_v4 cluster. Azure guarantees that every Edsv4 VM uses an Intel Xeon Platinum 8272CL processor, which delivers reliable performance. On the newer VMs, we enabled Photon, a vectorized query engine that can speed SQL query performance.

## Improve Data Warehouse Performance By Using Photon

The TPC-DS decision support benchmark measures data warehouse performance in terms of time to run a set of queries. Shorter times mean gaining insights earlier and reducing the VM uptime you must pay for. Figure 1 makes the performance advantages upgrading to the newer E8ds\_v4 VMs with Photon enabled very clear. With the 1TB data set, the E8ds\_v4 cluster query reduced completion time to only 26% of that of the E8s\_v3 cluster. With the 10TB data set, the E8ds\_v4 cluster query completion time was even lower, one-fifth that of the E8s\_v3 cluster.

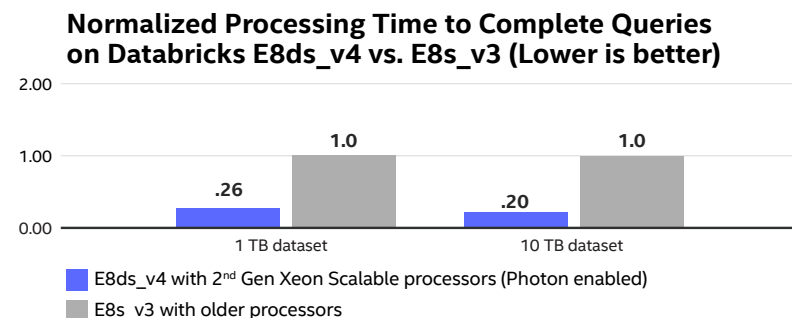


Figure 1. The relative processing time to complete the 99 decision support benchmark queries on a Photon-enabled E8ds\_v4 VM cluster with 2<sup>nd</sup> Gen Intel Xeon Scalable processors compared to an older E8s\_v3 VM cluster on both 1TB and 10TB datasets.

## Get Better Value By Upgrading

With the dramatically improved query times that we show on the previous page, one might assume that it would be well worth paying extra for the newer VMs. Figure 2 confirms this. Using the public price per hour at the time of testing, we determined the cost to execute each workload scenario. We converted the total query processing time from milliseconds to hours, combined the hourly cost of the instances and storage, and calculated the price per TB run for all four scenarios. We found that running a decision support workload with a 1TB dataset would cost almost twice as much on the older E8s\_v3 cluster as it would on the Photon-enabled E8ds\_v4 cluster. Even more impressive, running the 10TB dataset on the E8ds\_v4 cluster would cost well under half as much as it would on the older E8s\_v3 cluster, a savings of 61%.

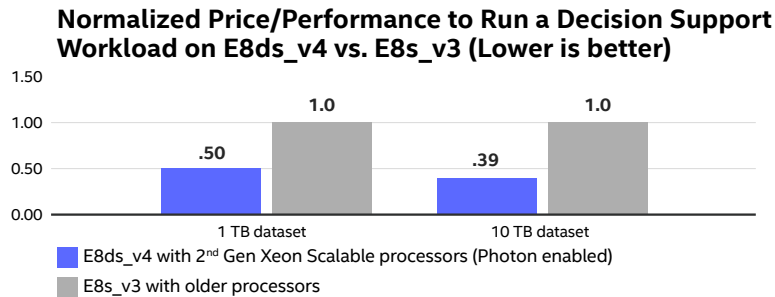


Figure 2. Normalized price/performance to run a decision support workload against a Databricks environment on Photon-enabled Azure E8ds\_v4 VMs compared to E8s\_v3 VMs on both 1TB and 10TB datasets.

## Conclusion

We found that decision support workloads completed as little as one-third the time on Photon-enabled eight-vCPU E8ds\_v4 VMs featuring 2<sup>nd</sup> Gen Intel® Xeon® Scalable processors compared to older E8s\_v3 VMs. This performance improvement led to a cost savings of up to 61%. This makes Photon-enabled E8ds\_v4 VMs featuring 2<sup>nd</sup> Gen Intel® Xeon® Scalable processors a great choice for your data analytics workloads.

## Learn More

To begin running your Databricks clusters on Photon-enabled Microsoft Azure Edsv4 VMs with 2<sup>nd</sup> Gen Intel Xeon Scalable processors, visit <https://docs.microsoft.com/en-us/azure/virtual-machines/edv4-edsv4-series>.

To read more about the results discussed here as well as see how the Microsoft Azure Edsv4 VMs performed compared to similar AMD VMs, read the report at <https://www.intel.com/content/www/us/en/partner/workload/microsoft/enhance-databricks-with-azure-vm-photon-benchmark.html>.



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 0122/JO/PT/PDF US001

