

Get Answers to Databricks Queries in 31% Less Time and Save up to 30% with Microsoft Azure Edsv4 VMs Featuring 2nd Gen Intel® Xeon® Scalable Processors

With Photon Vectorized Query Engine Enabled, These VMs Delivered Stronger Decision Support Workload Performance than Easv4 VMs Featuring AMD EPYC Processors

Databricks and Databricks Lakehouse Platform work to store and analyze the massive amounts of data—both structured and unstructured—that organizations collect. The faster you can analyze this data, the sooner your team can make solid business decisions with the facts at hand. For memory-intensive enterprise applications such as data warehousing/decision support workloads, Microsoft Azure offers multiple VM series: Edsv4 VMs enabled by 2nd Gen Intel® Xeon® Scalable processors and Easv4 VMs with AMD EPYC processors. To assist in selecting cloud VMs for decision support workloads, we tested a decision support workload on a 20-node E8ds_v4 cluster enabled by Databricks Runtime 9.0. We then tested the same workload on a 20-node E8as_v4 cluster, again assessing the time to complete queries as well as the price/performance to deliver insights. On both sets of VMs, we enabled Photon, a vectorized query engine that can speed SQL query performance.

We found that Edsv4 VMs with 2nd Gen Intel Xeon processors offered faster Databricks performance than Easv4 VMs, reducing the time to complete queries, while also offering better overall value.

Boost Data Warehouse Performance with Edsv4 VMs

We ran tests using a decision support benchmark derived from TPC-DS, which measures data warehouse performance in the amount of time it takes to complete a set of queries. Shorter times mean quicker answers, which can reduce ongoing costs for VM uptime. As Figure 1 shows, E8ds_v4 VMs with 2nd Gen Intel Xeon Scalable processors offered better Databricks workload performance than E8as_v4 VMs with AMD EPYC processors. With a 1TB data set, the E8ds_v4 cluster query reduced completion time by 31% over the E8as_v4 cluster. With a 10TB data set, the E8ds_v4 cluster reduced query completion time by 23% compared to the E8as_v4 cluster.

Normalized Processing Time to Complete Queries on Databricks E8ds_v4 vs. E8as_v4 (Lower is better)

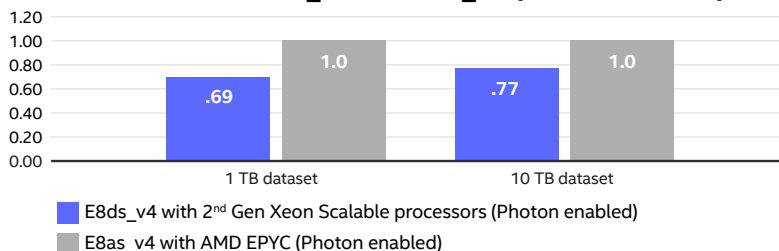
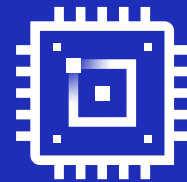


Figure 1. The relative processing time to complete the 99 decision support benchmark queries on a Photon-enabled E8ds_v4 VM cluster with 2nd Gen Intel Xeon Scalable processors compared to an E8as_v4 cluster with AMD EPYC processors on both 1TB and 10TB datasets.



Databricks



Run Decision Support Queries in up to 31% Less Time with E8ds_v4 VMs enabled by 2nd Gen Intel Xeon Scalable processors

vs. E8sv3 VMs with AMD EPYC processors



Spend up to 30% Less to Run Decision Support Queries with E8ds_v4 VMs enabled by 2nd Gen Intel Xeon Scalable processors

vs. E8s_v3 VMs with AMD EPYC processors

Better Performance and Better Value

Performance isn't the only thing to consider when selecting VMs to run your Databricks workloads. The ongoing cost to run them must also make business sense. We determined the cost to execute the workloads using the price per hour at time of testing. We converted the total query processing time from milliseconds to hours, combined the hourly cost of the instances, storage, and Databricks DBUs, and calculated the price per TB run for all four scenarios. As Figure 2 shows, running decision support workloads on Edsv4 VMs provides better value than Easv4 VMs. For a 1TB dataset, the E8ds_v4 cluster enabled by 2nd Gen Intel® Xeon® Scalable processors offered 30% lower price/performance than the E8as_v4 cluster with AMD EPYC processors. The price/performance for the 10TB dataset was similar, with the E8ds_v4 cluster reducing price/performance by 22% compared to the E8as_v4 cluster.

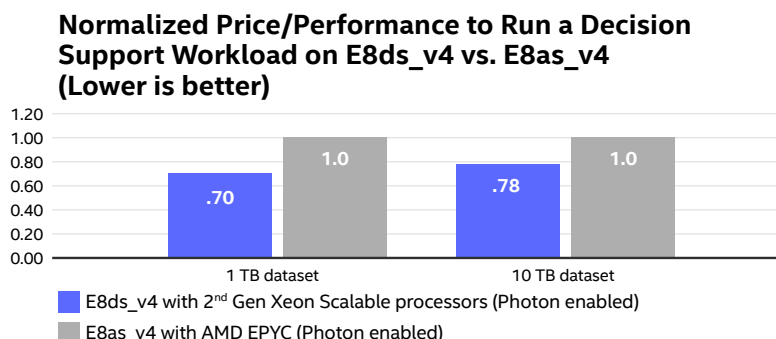


Figure 2. Normalized price/performance to run a decision support workload against a Databricks environment on Photon-enabled Azure E8ds_v4 VMs compared to E8as_v4 VMs on both 1TB and 10TB datasets.

Conclusion

Microsoft Azure E8ds_v4 VMs featuring 2nd Gen Intel® Xeon® Scalable processors finished decision support workloads in up to 31% less time than E8as_v4 VMs with AMD EPYC processors. This performance improvement led to a cost savings of up to 30%. These findings show that choosing E8ds_v4 VMs featuring 2nd Gen Intel® Xeon® Scalable processors offer a strong balance of performance and price for running Databricks decision support workloads, enabling your organization to process more data and gain insights sooner.

Learn More

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

For complete test details and results showing how these 2nd Gen Intel Xeon Scalable processor-enabled VMs fared against VMs with previous-generation processors, read the report at <https://www.intel.com/content/www/us/en/partner/workload/microsoft/enhance-databricks-azure-vm-benchmark.html>.



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

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 0322/JO/PT/PDF US002

Please Recycle