



Run Your Databricks Queries in up to 38% Less Time and Reduce Costs by up to 39% by Selecting Microsoft® Azure® Edsv4 VMs Featuring 2nd Gen Intel® Xeon® Scalable Processors

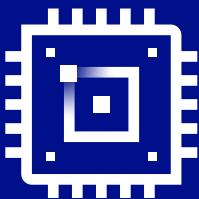


Databricks



Run Decision Support Queries in up to 38% Less Time with E8ds_v4 VMs enabled by 2nd Gen Intel Xeon Scalable processors

vs. L8s_v2 VMs with AMD EPYC processors



Spend up to 39% Less to Run Decision Support Queries with E8ds_v4 VMs enabled by 2nd Gen Intel Xeon Scalable processors

vs. L8s_v2 VMs with AMD EPYC processors

With Photon Vectorized Query Engine Enabled, These VMs Delivered Stronger Decision Support Workload Performance than Storage-Optimized L8s_v2 VMs Featuring AMD EPYC™ Processors

Databricks and Databricks Lakehouse Platform store and analyze the great volumes of structured and unstructured data that organizations gather. If you run these workloads in the cloud, you can speed the time necessary to conduct queries by selecting instances based on hardware that performs well. Speedier queries means implementing the provided insights sooner.

To help companies choosing cloud VMs for data warehousing/decision support, we tested two Microsoft Azure VM series that are well suited to such workloads: Edsv4 VMs enabled by 2nd Gen Intel® Xeon® Scalable processors and storage-optimized Lsv2 VMs with AMD EPYC processors. We tested a decision support workload on clusters of these two VM series enabled by Databricks Runtime 9.0. We enabled Photon, a vectorized query engine designed to improve SQL query performance, on both.

The Edsv4 VMs with 2nd Gen Intel Xeon processors outperformed the storage-optimized Lsv2 VMs by completing the queries more quickly. Furthermore, when we calculated price/performance of the two series on this workload, we found that the Edsv4 VMs delivered better value as well.

Enjoy Speedier Data Warehouse Performance with Edsv4 VMs

Our tests used a decision support benchmark based on TPC-DS, which delivers a lower-is-better metric that reflects the time necessary to conduct a given set of queries. Shorter times not only get actionable insights into the hands of decision-makers earlier, but can also translate to savings by reducing VM uptime and associated costs. As Figure 1 shows, E8ds_v4 VMs with 2nd Gen Intel Xeon Scalable processors completed queries on a 1TB data set in 38% less time than L8s_v2 VMs with AMD EPYC processors did. With a 10TB data set, query completion time of the E8ds_v4 cluster was 36% shorter than that of the L8s_v2 cluster.

Normalized Processing Time to Complete Queries on Databricks E8ds_v4 vs. L8s_v2 (Lower is better)

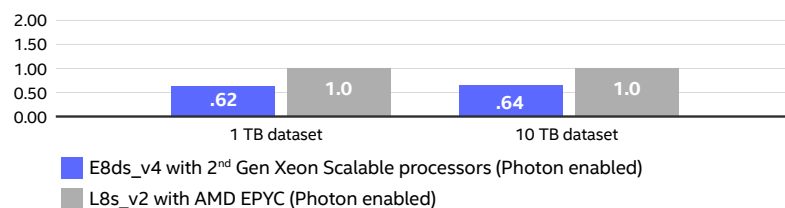


Figure 1. Relative processing time to complete a set of benchmark queries on a Photon-enabled E8ds_v4 VM cluster with 2nd Gen Intel Xeon Scalable processors and an L8s_v2 cluster with AMD EPYC processors on both 1TB and 10TB data sets.

Faster Query Time Translates to Better Value

As you shop for the right VMs for your Databricks workloads, pricing can be an important factor. To calculate the price of carrying out the test scenarios we describe on the previous page, we started with price per hour for each VM at time of testing. We used that rate and the times in Figure 1 to determine the price per TB run for all four scenarios. As Figure 2 shows, we could run decision support workloads on Edsv4 VMs provides at a lower cost for a given amount of performance. For the 1TB dataset, the E8ds_v4 cluster enabled by 2nd Gen Intel® Xeon® Scalable processors offered 39% lower price/performance than the storage-optimized L8s_v2 cluster with AMD EPYC processors did. For the 10TB dataset, the E8ds_v4 cluster delivered price/performance savings of 22%.

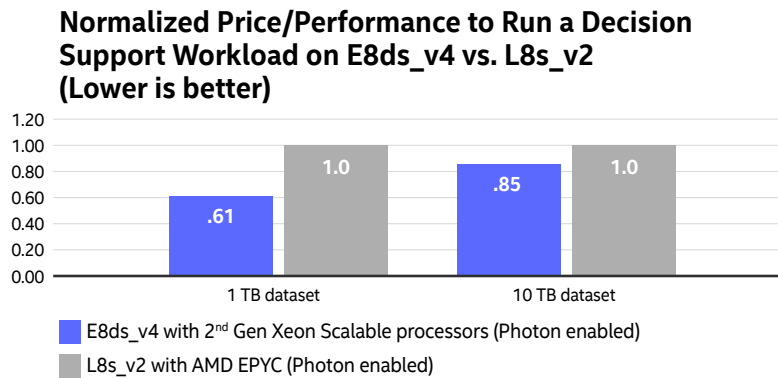


Figure 2. Normalized price/performance to run a decision support workload against a Databricks environment on Photon-enabled Azure E8ds_v4 VMs compared to L8s_v2 VMs on both 1TB and 10TB datasets.

Conclusion

We investigated two metrics—the time to complete a set of Databricks queries and the price/performance—for two different data set sizes on Microsoft Azure E8ds_v4 VMs featuring 2nd Gen Intel Xeon Scalable processors and storage-optimized L8s_v2 VMs with AMD EPYC processors. The E8ds_v4 VMs completed sets of queries in up to 38% less time. Combined with hourly pricing, these VMs delivered cost savings as high as 39%. By selecting E8ds_v4 VMs featuring 2nd Gen Intel Xeon Scalable processors, your organization could gain insights earlier while also spending less.

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 0422/JO/PT/PDF US003

Please Recycle