



Reduce Time to Complete Databricks Queries by up to 65% and Save up to 35% by Enabling the Databricks Photon Query Engine on New Microsoft Azure Eight-vCPU Edsv4 VMs Featuring 2nd Gen Intel[®] Xeon[®] Scalable Processors

Enable Photon to Get the Best out of the Latest Intel Processors for Data Analytics and AI Workloads

As companies contend with increasing amounts of data, finding effective ways of storing and analyzing that data becomes increasingly important. Data lakes and data warehouses provide large-scale storage infrastructure for unstructured data and structured data, respectively. Databricks combines features from both of these to store and analyze vast amounts of structured and unstructured data with their Lakehouse Platform. Photon Engine, included in the Lakehouse Platform, is a vectorized query engine that can speed SQL query performance, delivering business insights even sooner and reducing costs.

The decision support benchmark derived from TPC-DS measures the performance of data warehousing by running a set number of queries and recording the time to complete. Faster queries translate to less VM uptime to pay for. A Photon-enabled Microsoft Azure E8ds_v4 VM cluster featuring 2nd Gen Intel Xeon Scalable processors, for example, finished querying a 1TB Databricks cluster in 65% less time, at 35% lower cost, than the same cluster with Photon disabled. Photon works on larger datasets too; the same E8ds_v4 cluster with Photon finished querying a 10TB Databricks cluster in 62% less time, 30% lower cost, than without Photon.

Improve Data Warehouse Performance By Using Photon

The sooner data analytics queries complete, the faster you can implement the insights to improve and expand your business. To demonstrate how well Photon can enhance query performance, we tested our eight-vCPU Edsv4 cluster with Photon disabled and enabled. Figure 1 shows how the E8ds_v4 cluster with Photon enabled completed a 1TB dataset in 65% less time than the same cluster without Photon, and completed a 10TB dataset in 62% less time.

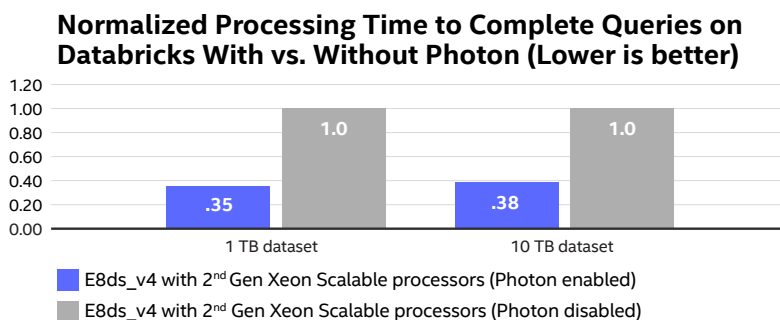
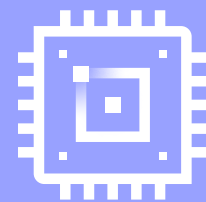


Figure 1. The relative processing time to complete the 99 decision support benchmark queries with Photon compared to without Photon on E8ds_v4 clusters on 1TB and 10TB datasets.



Databricks



Up to 65% Less Time to Run Decision Support Queries with Photon Enabled vs. Without Photon on E8ds_v4 VMs

vs. Without Photon on E8ds_v4 VMs



Up to 35% Lower Cost to Run Decision Support Queries with Photon Enabled

vs. Without Photon on E8ds_v4 VMs

Get a better value with Photon

Not only does using Photon accelerate the time to insights, but this speed means less VM uptime for which you must pay. As Figure 2 shows, the E8ds_v4 cluster with Photon enabled would cost 35% less to run a 1TB dataset than the same cluster with

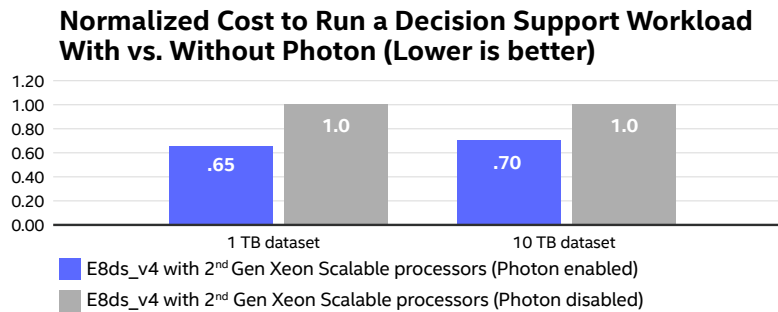


Figure 2. Normalized HammerDB test results comparing performance (in new orders per minute) achieved by the Edsv4 VM to Esv3 VM with 16 vCPUs.

Photon, and 30% less to run a 10TB dataset. As you can see, shorter run times translate to savings.

Conclusion

When you want the best decision support performance from your E8ds_v4 VMs, use the Databricks Photon query engine to reduce query completion time. These eight-vCPU VMs featuring 2nd Gen Intel Xeon Scalable processors finished a TPC-DS job on a Databricks cluster in up to 65% less time with Photon enabled, which led to a cost savings of up to 35%. When it comes to data analytics, make the smart choice and choose E8ds_v4 VMs featuring 2nd Gen Intel® Xeon® Scalable processors with Photon enabled.

Learn More

To begin running your Databricks clusters with Photon enabled on Microsoft Azure Edsv4 VMs with 2nd 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-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