

Complete Databricks Queries in up to 77% Less Time and Spend up to 57% Less by Enabling Photon on AWS Instances with Intel® Xeon® Processors



Databricks



Run Decision Support Queries in up to 77% less time with Photon Enabled on AWS instances with Intel Xeon processors

compared to the same instances without Photon



Spend up to 57% Less to Run Decision Support Queries with Photon Enabled on AWS instances with Intel Xeon processors

compared to the same instances without Photon

Across AWS Instances with Various Processor Generations, Enabling Photon Vectorized Query Engine Improved Decision Support Workload Performance

With organizations collecting increasingly large amounts of data—both structured and unstructured—the challenge of storing and making sense of that data also increases. Databricks provides a Lakehouse platform equipped to govern and analyze data so that organizations can get quick insights from data. To speed up analysis, Databricks offers the Photon Engine, which is a vectorized query engine that can speed SQL query performance. If you're running decision support queries on data stored in AWS instances with Intel® Xeon® processors, how much difference does enabling Photon affect query completion times? To find out, we tested three sets of instances, each with a different generation of Intel Xeon processor:

- AWS i3 instances with Intel Xeon E5-2686 v4 processors
- AWS i3en instances with 1st Gen Intel Xeon Scalable processors
- AWS R5d instances with 2nd Gen Intel Xeon Scalable processors

We ran a benchmark that measures decision support system performance in terms of time to run a set of queries. Across all three instance types, we found that enabling Photon reduced the time it took to analyze data by as much as 77% and could reduce the cost to run analysis workloads by up to 57% compared to the same instances without Photon.

Boost Decision Support Performance By Enabling Photon

Figure 1 shows the performance advantages of enabling Photon on AWS instances with Intel Xeon processors. Enabling Photon shrunk data analysis times significantly for all generations of processors we tested.

Normalized Processing time to complete Databricks Queries With vs. Without Photon

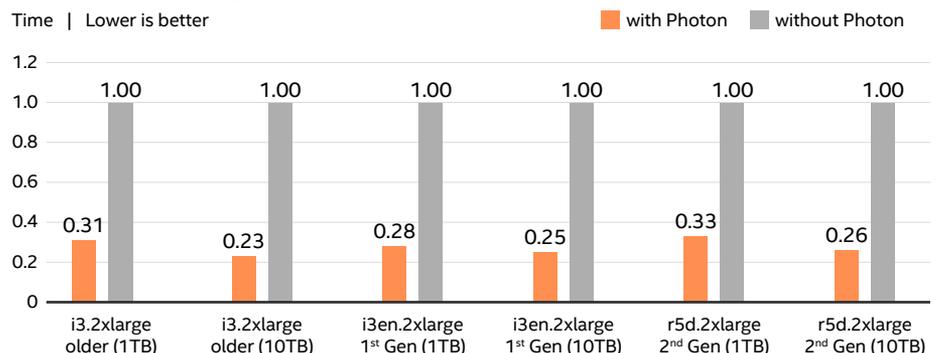


Figure 1. The relative processing time to complete the 99 decision support benchmark queries with Photon enabled and disabled on both 1TB and 10TB datasets for three AWS instance types with various Intel Xeon processors.

Enable Photon on Databricks Workloads and Improve Value

As Figure 2 shows, completing Databricks queries in less time also has value in price/performance that can help keep cloud costs down. 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, storage, and Databricks DBUs, and calculated the price per TB run for all four scenarios. Across instance types, running decision support workloads with Photon enabled could save as much as 57%.

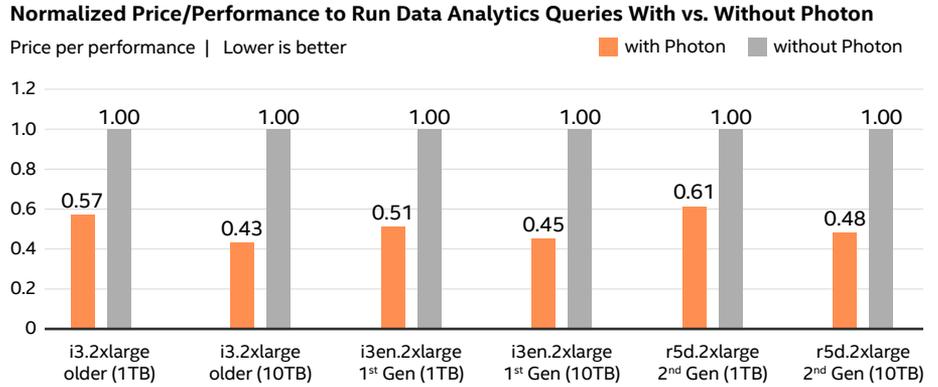


Figure 2. Normalized price/performance to run a decision support workload against a Databricks environment on Photon-enabled and Photon disabled AWS instances using both 1TB and 10TB datasets.

Conclusion

If you're running Databricks workloads on AWS, it makes sense to enable Photon regardless of which AWS instance type and Intel® Xeon® processor combination you use. These results show that for three different generations of processors, enabling Photon can help you get answers from data in less time and make better use of your cloud budget.

Learn More

To begin running your Databricks clusters on Photon-enabled AWS instances, visit <https://aws.amazon.com/quickstart/architecture/databricks/>.

To learn more about Databricks' Photon Vectorized Query Engine, visit <https://databricks.com/product/photon> and <https://docs.databricks.com/runtime/photon.html>.

For all of the results in this report, we used a decision support workload derived from TPC-DS. All tests were conducted in December 2021 on the us-east-1 AWS region. All tests used 20-node clusters with Ubuntu 18.04.1, kernel version 5.4.0-1059-AWS, Databricks 9.0, Apache Spark 3.1.2, Scala 2.12. All VMs had 8 vCPUs and 64GB RAM. The i3.2xlarge instances had a 1900GB NVMe SSD, 10Gbps Network BW and 4,750 Mbps Storage BW. The i3en.2xlarge had a 5,000GB NVMe SSD, 25 Gbps Network BW, and 4,750 Mbps Storage BW. The r5d.2xlarge had a 300GB NVMe SSD, 10 Gbps Network BW, and 4,750 Mbps Storage BW.



Performance varies by use, configuration and other factors. Learn more at www.intel.com/PerformanceIndex.

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See backup for configuration details. No product or component can be absolutely secure. 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

