

Perform Databricks Queries in Less Time and Get a Better Value with Google Cloud Platform n2-highmem-8 VMs Featuring 2nd Gen Intel[®] Xeon[®] Scalable Processors

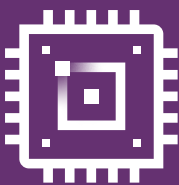


Databricks



Reduce the time to complete decision support workloads by up to 3.3 times with Photon and N2 VMs backed by Intel processors

compared to N2D VMs with AMD processors



Running decision support databases on N2D VMs with AMD processors cost up to 80% more

than N2 VM instances with Intel processors and Photon enabled

With Photon Query Engine Enabled, These VMs with Intel Xeon Scalable Processors Outperformed n2d-highmem-8 VMs with AMD EPYC[™] Processors

The sooner data analytics queries complete, the faster you can get relevant data to make business-critical decisions. Combining data warehouse and data lake features, the Lakehouse Platform from Databricks enables organizations to store and analyze structured and unstructured data. Photon, a feature of the Lakehouse Platform, is a vectorized query engine that can help speed SQL query performance. According to a summary from Databricks, other Photon benefits include:

- “Supports SQL and equivalent DataFrame operations against Delta and Parquet tables.
- Expected to accelerate queries that process a significant amount of data (100GB+) and include aggregations and joins.
- Faster performance when data is accessed repeatedly from the Delta cache.
- More robust scan performance on tables with many columns and many small files.
- Faster Delta and Parquet writing using UPDATE, DELETE, MERGE INTO, INSERT, and CREATE TABLE AS SELECT, especially for wide tables (hundreds to thousands of columns).
- Replaces sort-merge joins with hash-joins.”¹

We tested two types of Google Cloud Platform (GCP) VMs: n2-highmem-8 with 2nd Gen Intel Xeon processors and Photon enabled, and n2d-highmem-8 with 2nd Gen AMD EPYC processors. Photon was not available for the N2D VMs. To measure data warehousing performance, we ran a decision support benchmark that recorded the time to complete a set number of queries. Results indicated that the N2 VMs with Photon reduced time to complete 1TB and 10TB datasets, which also means that the N2 VMs delivered a better value.

Less Time to Complete Queries, Faster Time to Insight

We ran the decision support benchmark on eight-vCPU n2-highmem-8 VMs with Photon with a 1TB dataset and a 10TB dataset, and we did the same for the eight-vCPU n2d-highmem-8 VMs. As Figure 1 shows, the N2 VM cluster with Intel Xeon Scalable processors and Photon completed queries 3.1 times as fast as the N2D cluster on the 1TB dataset; on the 10TB dataset, it completed queries 3.3 times as fast as the N2D cluster.

Normalized Processing Time to Complete Queries on Databricks

Relative time to complete | Lower is better

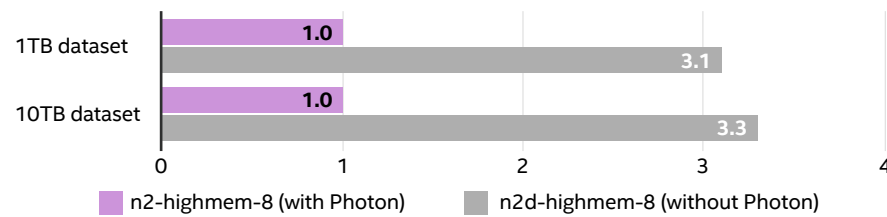


Figure 1. The relative processing time to complete decision support benchmark queries on GCP n2-highmem-8 VMs and n2d-highmem-8 VMs on 1TB and 10TB datasets.

Less VM Uptime Necessary, More Cost Savings

Your business can benefit from performance improvements to decision support workloads, but value is another important consideration. Using the VM price per hour at the time of testing and the amount of time to complete each dataset, we calculated the price per TB run for each cluster across both datasets. Figure 2 shows that running Databricks workloads on N2 VMs delivered a better value than N2D VMs at both dataset sizes. For the 1TB dataset, the n2d-highmem-8 VMs with AMD EPYC processors cost 70% more than the n2-highmem-8 VMs with Intel® Xeon® Scalable processors. Similarly, the n2d-highmem-8 VMs cost 80% more than the n2-highmem-8 VMs to complete a 10TB dataset.

Normalized Cost to Run a Decision Support Workload

Relative cost | Lower is better

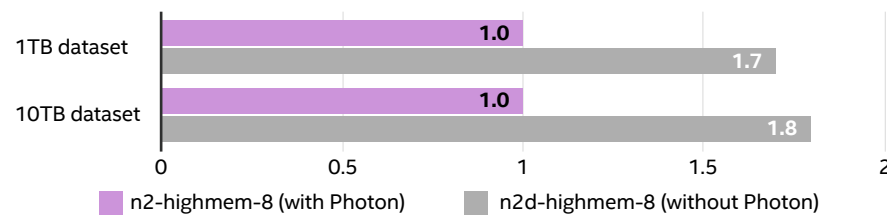


Figure 2. Normalized price/performance to run a decision support workload against a Databricks environment on GCP n2-highmem-8 VMs and n2d-highmem-8 VMs on both 1TB and 10TB datasets.

Conclusion

Backed by 2nd Gen Intel Xeon processors, GCP n2-highmem-8 VMs with the Photon query engine completed decision support workloads up to 3.3 times as fast as n2d-highmem-8 VMs. Not only did they improve performance, but they also delivered a better value, as n2d-highmem-8 VMs cost up to 80% more to complete dataset queries. To equip your business with cost savings and the speedy insights you need to make informed decisions, choose n2-highmem-8 VMs featuring 2nd Gen Intel Xeon Scalable processors and Photon.

Learn More

To begin running your Databricks clusters with Photon enabled on GCP N2 VMs with 2nd Gen Intel Xeon Scalable processors, visit, <https://cloud.google.com/compute/docs/general-purpose-machines>.

¹Databricks, "Photon," accessed April 12, 2022, <https://docs.databricks.com/runtime/photon.html>.

Tests by Intel in March 2021 for Intel VM testing, and March 2022 for AMD VM testing; both on GCP us-central1 (Iowa). All configurations: 21 instances (20 workers + 1 master), 8 vCPUs, 128GB RAM, 25 Gbps, 500GB remote SSD+0.75TB local SSD, 240-1200/240-1200 (R/W remote SSD) 9360/4680 (R/W local SSD) Ubuntu 20.04.3 LTS kernel 5.4.170+, Databricks 10.3. Spark config: spark.databricks.passthrough.enabled true, spark.databricks.adaptive.autoOptimizeShuffle.enabled true, spark.databricks.io.cache.maxMetadataCache 10g, spark.databricks.io.cache.maxDiskUsage 100g, spark.databricks.delta.preview.enabled true. N2-highmem-8: Intel Cascade Lake CPU. N2d-highmem-8: AMD Rome CPU. Total cluster cost per run as of Mar 2022: w/Photon 1TB Intel: \$6.44; w/Photon 10TB Intel: \$33.11; w/o Photon 1TB AMD: \$11.17; w/o Photon 10TB AMD: \$61.53



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 0522/JO/PT/PDF US002

