Medium instances: Finish data warehouse analysis up to 1.22 times as fast*

Large instances: Finish data warehouse analysis up to 1.39 times as fast**

*M5n instances with 16 vCPUs vs. M4 instances with 16 vCPUs

**M5n instances with 64 vCPUs vs. M4 instances with 64 vCPUs

Finish Microsoft SQL Server data analysis faster with new M5n series instances for Amazon Web Services powered by 2nd Generation Intel Xeon Scalable Processors – Cascade Lake

Newer 2nd Generation Intel Xeon Scalable (8259CL) Cascade Lake processor-powered M5n instances versus older M4 series instances powered by Intel Xeon (E5-2686 v4) Broadwell processors

If you've decided to use cloud computing to power your company's business intelligence applications, you'll need fast cloud instances to help you analyze data to discover insights, plan and forecast budgets, and get a better overall picture of your business's health.

At Principled Technologies, we tested an online analytical processing (OLAP) workload on two series of instances for Amazon Elastic Cloud Compute (Amazon EC2) running Microsoft SQL Server:

- M5n series instances powered by 2nd Generation Intel® Xeon® Scalable (8259CL) Cascade Lake processors
- M4 series instances powered by Intel Xeon (E5-2686 v4) Broadwell processors

We found that the newer instances completed the database queries faster than the older instances, and that this speed advantage held true across the database and instance sizes we tested. This could enable your business to reduce the amount of time it spends on analysis and to reach critical decisions sooner.

How we tested

Our cloud solution

We purchased instances from two series of general-purpose instances for Amazon EC2: the newer M5n series powered by 2nd Generation Intel Xeon Scalable (8259CL) Cascade Lake processors, and older M4 series powered by Intel Xeon (E5-2686 v4) Broadwell processors. Note that M4 series instances are available in a few different CPU configurations, but we used only Intel Xeon (E5-2686 v4) Broadwell processors for our testing. We ran all instances in the us-east1-a region.

Determining the size of the instances and databases

To represent different customer needs, we tested two sizes for each instance series: medium and large. Medium instances contained 16 vCPUs and a 30GB database size. Large instances contained 64 vCPUs and a 100GB database. We determined database size based on two factors: we wanted to represent a realistic workload scenario for each compute level, and we also wanted to ensure the processors were working harder than the storage drives by limiting database size to the amount of allocated RAM.

Figure 1 shows the specifications for each of the instances we tested. Note that for each test run, we ran a single-stream test to preload the database into memory before running the number of streams we wished to test.

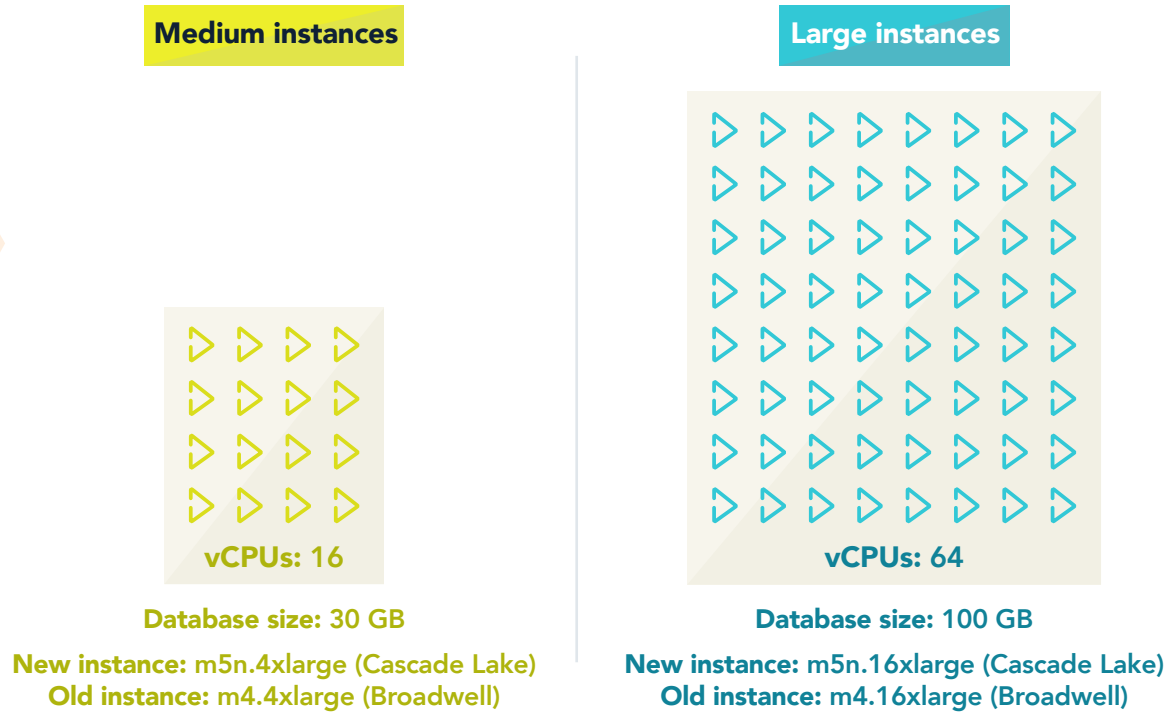

Figure 1: Configuration details for the instances we tested. Source: Principled Technologies.

M5n series instances for Amazon EC2

In 2019, Amazon introduced M5n instances to their EC2 offerings, extending 100Gbps networking and adding improved packet-processing performance capabilities to its general-purpose M family.[2] In addition to adding 2nd Generation Intel Xeon Scalable processors to the M-series options, M5n instances support the following:

- Support for Intel DL Boost Vector Neural Network Instructions, which Intel claims can "deliver a significant performance improvement" for deep learning applications[3]
- 25 Gbps of peak bandwidth for small and medium instances, 75Gbps for large instances
- AWS Nitro System, which provides better security and performance compared to older AWS hypervisors[4]

OLAP workload: HammerDB

We tested the instances using a data warehouse workload from the HammerDB benchmarking suite. The HammerDB developers derived this workload from TPC-H benchmark specifications. However, because this workload is not a full implementation of the official TPC-H benchmark, our results are not directly comparable to published TPC-H results.

This workload measures the time required for an instance to finish analyzing streams of database queries. Each stream in this workload is made of 22 serialized database queries. We tested each instance size multiple times, increasing the number of concurrent streams up to the maximum number recommended by the Transaction Processing Performance Council (TPC).[1] We tested up to four streams on the medium instances, and up to five streams on the large instances. When testing at the maximum stream count, we ensured that each instance's processors achieved full saturation.

Our results: Faster data analysis with new instances

When analyzing just a single stream of 22 data warehouse queries, the new M5n series instances were up to 1.39 times as fast as the M4 series instances.

After completing the single-stream test, we wanted to determine how increasing amounts of work would affect completion time. To do that, we performed additional tests, increasing the number of streams by one until we reached the maximum recommended by TPC—four in the case of the medium instances, and five in the case of the large instances. Figures 2 and 3 show the full results of our query stream testing for medium and large Amazon EC2 instances.

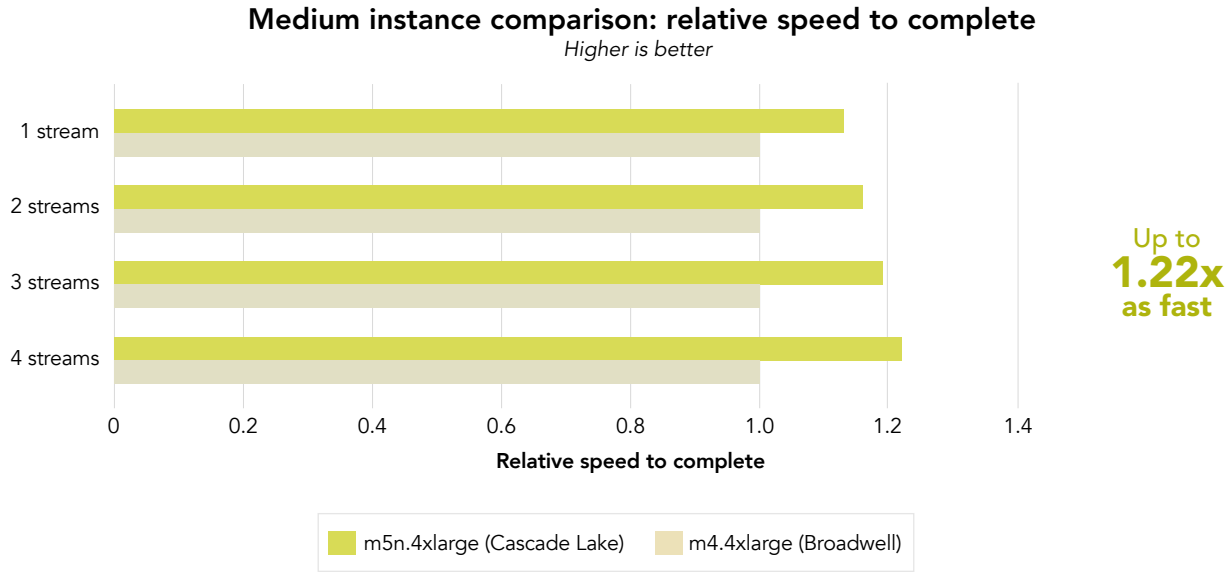

Figure 2: Normalized comparison of the speed at which each medium instance completed one, two, three, and four query streams from the TPC-H-like HammerDB workload. Higher speed is better. Source: Principled Technologies.

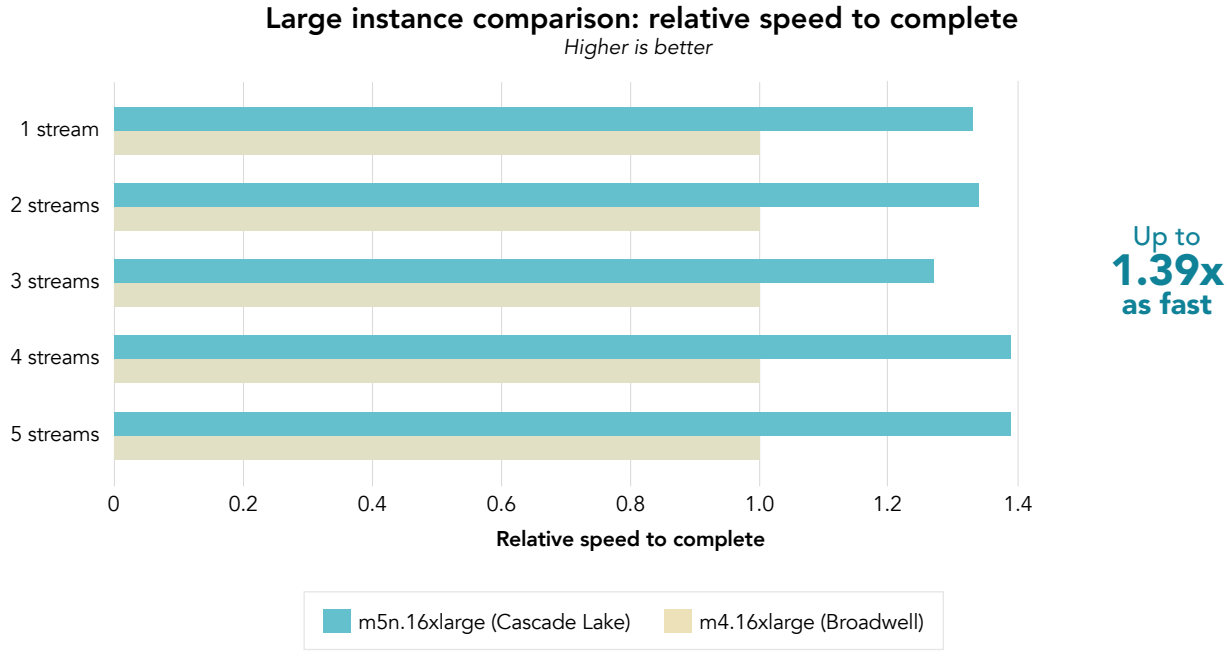

Figure 3: Normalized comparison of the speed at which each large instance completed one, two, three, four, and five query streams from the TPC-H-like HammerDB workload. Higher speed is better. Source: Principled Technologies.

Getting more value from your investment

In our data analytics performance tests, the new M5n instances finished their maximum-stream data warehouse work between 1.22 and 1.39 times as fast as the M4 instances. At the time of this writing, all sizes and specifications of the M5n instances cost just 1.19 times as much as their M4 counterparts.[6] Because their performance outpaces their additional cost, M5n instances present a better value for your company's data warehouse analytics work, especially if you are interested in larger instance sizes.

At every data point, the newer 2nd Generation Intel Xeon Scalable (8259CL) Cascade Lake processor-powered M5n instances completed data analysis faster than the older Intel Xeon (E5-2686 v4) Broadwell processor-powered M4 instances.

The following hypothetical scenario may help you see a clearer picture of how our test results can relate to a real-world use case. Each night, a company has a two-hour window in which to analyze data on a 100GB database. The company uses this analysis to generate reports to open discussions on which products to improve and which geographic regions to focus their sales on, and to drive other business processes throughout the day.

Based on the results of our single-stream tests, we calculate that a new 64vCPU m5n.16xlarge instance would enable this hypothetical company to analyze 266 data streams within their two-hour analysis window. By contrast, the older m4.16xlarge instance would complete just 200 query sets (39 percent fewer) in the same time frame.

Put another way, the m5n.16xlarge instances would take just 1.5 hours to analyze the same number of data streams that the m4.16xlarge instances took 2 hours to complete, shrinking the required analysis window by 30 minutes (or 25 percent).

Over the course of a year, this hypothetical company would save 182.5 hours—or 7.6 full days' worth—of analysis time per year while maintaining the same rate of analysis they could get with the m4.16xlarge instance.

Scale up for improved performance gains

In our tests, large M5n instances outperformed the older M4 instances by a larger margin than the divide between medium instances. Whereas the medium M5n instances were 1.13 to 1.22 times as fast as the M4 instances, the large M5n instances were 1.27 to 1.39 times as fast.

1.13 – 1.22 times as fast
for medium instances

1.27 – 1.39 times as fast
for large instances

Conclusion

The more insights you can gain into your business, the more tools you'll have to handle challenges and plan for growth. Using OLAP workloads with your business intelligence applications is an effective way to boost the knowledge your organization has to work with.

In our tests, newer M5n series instances powered by 2nd Generation Intel Xeon Platinum (8259CL) Cascade Lake processors analyzed data warehouse query streams up to 1.39 times as fast as older M4 series instances powered by Intel Xeon (E5-2686 v4) Broadwell processors. This speed advantage could enable your business to reach critical decisions sooner and cut the time it takes to process data.

---

1 Transaction Processing Performance Council (TPC), "TPC Benchmark H," accessed November 3, 2020, http://www.tpc.org/tpc_documents_current_versions/pdf/tpc-h_v2.18.0.pdf.

2 Julien Simon, "New M5n and R5n EC2 Instances, with up to 100Gbps networking," accessed November 3, 2020, https://aws.amazon.com/blogs/aws/new-m5n-and-r5n-instances-with-up-to-100-gbps-networking/.

3 "Amazon EC2 Instance Types," accessed November 3, 2020, https://aws.amazon.com/ec2/instance-types/.

4 "AWS Nitro System," accessed November 16, 2020, https://aws.amazon.com/ec2/nitro/.

5 "Introduction to Intel Deep Learning Boost on Second Generation Intel Xeon Scalable Processors," accessed November 3, 2020, https://software.intel.com/content/www/us/en/develop/articles/introduction-to-intel-deep-learning-boost-on-second-generation-intel-xeon-scalable.html.

6 Cost data from "Amazon EC2 On-Demand Pricing," accessed November 3, 2020, https://aws.amazon.com/ec2/pricing/on-demand/.

Read the science behind this report at http://facts.pt/Bk4jM83 ▶

Principled Technologies®

Facts matter.®

This project was commissioned by Intel.