Handle a higher rate of Microsoft SQL Server transactions with Google Cloud C3 high CPU instances enabled by 4th Gen Intel Xeon Scalable processors

Compared to N2 high CPU instances with previous-gen processors, these instances delivered a higher SQL Server transaction rate

1.35x

the new orders per minute

Making customer experiences faster can improve satisfaction and lead to business growth. That's why, when you host your SQL Server databases in the cloud, selecting instances that can handle higher transaction rates can make a big difference.

To help organizations realize the benefits that newer technologies can have on database performance, we compared the SQL Server TPROC-C performance of two Google Cloud™ instance types: C3 high CPU instances with 4th Gen Intel® Xeon® Scalable processors and N2 high CPU instances with previous-gen processors. Both instance types used Google Cloud Hyperdisk Extreme storage, a new network block storage service. In our tests, the C3 high CPU instances were able to handle significantly more new orders per minute (NOPM) than similarly configured N2 high CPU instances with previous-gen processors.

The ability to handle more NOPM on C3 high CPU instances means that your organization could support more customers, reduce bottlenecks during peak times that cause customers to wait, or even save money by needing fewer instances to support your user base.

This project was commissioned by Intel.

How we tested

We compared two similarly configured high CPU instances with processor technology from different generations:

Table 1: Key configuration details for the two instances we tested. Source: Principled Technologies.

	N2 (previous gen)	C3 (current gen)
Instance	n2-highcpu-80	c3-highcpu-88
Region	us-east-1b	us-east-1b
CPU cores/ vCPU	80	88
Memory (GB)	80	176
Processor	Intel Xeon Platinum 8373C	Intel Xeon Platinum 8481C
Storage	1TB Hyperdisk Extreme, 200K IOPS (data) 60GB SSD persistent (log)	1TB Hyperdisk Extreme, 200K IOPS (data) 60GB SSD persistent (log)
DB size (WH/GB)	3,000/270	3,000/270

Google Cloud has many options for instance sizing, and performance varies significantly among them. At the time of our testing, the only C3 instance type available for general use was the C3 high CPU. While we would typically choose an instance with a higher memory-to-core ratio for a database workload, the standard and high-memory C3 instances were still in private preview at the time of testing. For comparison, we chose the N2 high CPU instance type and configured it as similarly as possible. Hyperdisk Extreme requires at least 64 vCPUs, so we configured the C3 instance with 88 vCPUs and matched the N2 instance as closely as we could, which was 80 vCPUs.

We sized the database to 3,000 warehouses, or roughly 300 GB, to ensure that it did not fit entirely into RAM and would engage the disks. We increased the HammerDB TPROC-C virtual users (VU) on each instance until the NOPM performance peaked and began dropping. We sized the disks with the goal of providing enough performance to avoid bottlenecks. Note: At the time of our testing, Google Cloud reported a known issue with Hyperdisks of limited performance on Windows instances. To avoid these known performance bottlenecks on the disks, we ran SQL Server on a Linux operating system.

To learn more about our testing, including step-by-step instructions for completing these tests, read the science behind the report.

About the HammerDB workload

We used the TPROC-C workload from the HammerDB benchmark to measure the performance of the instance types. According to HammerDB, this workload "is intentionally not fully optimized and not biased towards any particular database implementation or system hardware."[1] HammerDB developers derived their OLTP workload from the TPC-C benchmark specifications; however, as this is not a full implementation of the official TPC-C standards, the results in this paper are not directly comparable to published TPC-C results.

For more information, please visit https://www.hammerdb.com/docs/ch03s01.html.

Improving SQL Server performance by choosing Google Cloud C3 instances with 4th Gen Intel Xeon Scalable processors

Figure 1 shows the results of our HammerDB testing of new C3 instances with 4th Gen Intel Xeon Scalable processors vs. previous-gen N2 instances with 3rd Gen Intel Xeon Scalable processors. Both instance types used new Google Cloud Hyperdisk Extreme storage. The C3 high CPU instance outperformed the previous-gen instance, handling 1.35x the SQL Server new orders per minute using the HammerDB TPROC-C workload.

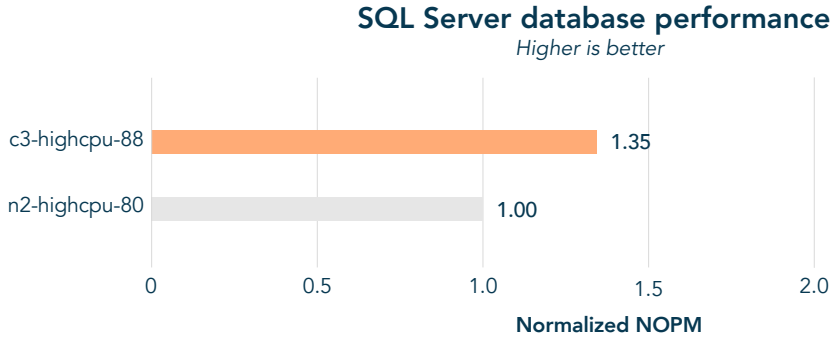

Figure 1: HammerDB results, in normalized new orders per minute, that the N2 and C3 instances achieved. Higher numbers are better. Source: Principled Technologies.

Handling that significant increase in database orders per instance shows that organizations hoping to minimize their cloud footprint could ultimately support fewer instances by selecting the C3 high CPU instance type over previous-generation hardware.

About Google Cloud Hyperdisk

Google Cloud now offers Hyperdisk, a network block storage service for NVMe or SCSI interfaces. According to Google Cloud, Hyperdisk "offers a scalable, high-performance storage service with a comprehensive suite of data persistence and management capabilities. With Hyperdisk you can easily provision, manage, and scale your Compute Engine workloads without the cost and complexity of a typical on-premises storage area network (SAN)."[2] There are two types of Hyperdisk: Hyperdisk Extreme, for maximum performance for the most demanding workloads, and Hyperdisk Throughput, for flexibility and efficiency of scale-out workloads. In our testing, we used 1TB Hyperdisk Extreme storage.

To learn more about Google Cloud Hyperdisk, visit https://cloud.google.com/compute/docs/disks/hyperdisks.

About 4th Generation Intel Xeon Scalable processors

The Intel strategy for 4th Gen Intel Xeon Scalable processors "aligns CPU cores with built-in accelerators optimized for specific workloads and delivers increased performance at higher efficiency for optimal total cost of ownership."[3] Intel claims the processors deliver "a range of features for managing power and performance, making the best use of CPU resources to achieve key sustainability goals. In addition, the Xeon CPU Max and the Max Series GPU add high-bandwidth memory and maximum compute density to solve the world's most challenging problems faster."[4]

To learn more about the latest processor technology from Intel, visit https://www.intel.com/content/www/us/en/products/docs/processors/xeon-accelerated/4th-gen-xeon-scalable-processors.html.

Conclusion

Cloud products are plentiful, and it can be easy for organizations to leave performance on the table by simply continuing to add more of the instances they are already using as their customer base or their demands grow. Taking this approach can needlessly waste an IT budget—but by choosing C3 high CPU instances with 4th Generation Intel Xeon Scalable processors, organizations can get significantly more performance per instance (compared to previous-gen instances), which could ultimately lead to savings. In our tests, where both instance types used Google Cloud Hyperdisk Extreme storage and were configured similarly, Google Cloud C3 high CPU instances with 4th Gen Intel Xeon Scalable processors delivered 1.35x the NOPM. If your organization seeks to maximize SQL Server performance per instance, selecting Google Cloud C3 high CPU instances with the latest processor technology could help you achieve those goals.

---

1. HammerDB, "About HammerDB," accessed July 25, 2023, https://www.hammerdb.com/about.html.
2. Google, "Google Cloud Hyperdisk," accessed July 27, 2023, https://cloud.google.com/compute/docs/disks/hyperdisks.
3. Intel, "4th Gen Intel Xeon Scalable Processors," accessed July 27, 2023, https://www.intel.com/content/www/us/en/newsroom/resources/press-kit-4th-gen-intel-xeon-scalable-processors.html.
4. Intel, "4th Gen Intel Xeon Scalable Processors."

Read the science behind this report at https://facts.pt/mJ83gQ6 ▶

Principled Technologies®

Facts matter.®

This project was commissioned by Intel.