How to Optimize Microsoft SQL Server Analytics Workloads With Micron® SSDs and DRAM

ACCELERATE MICROSOFT SQL SERVER ANALYTICS FROM MASSIVE DATA WAREHOUSES

Building efficient, optimized analytics platforms requires storage performance,[1] endurance and DRAM optimization. This document shows you how.

NVMe® SSD performance is typically much better than other storage choices.[2] This brief discusses testing performed on the Micron® 7450 SSD with NVMe, which is capable of up to 6.6 GB/s read and 3.5 GB/s sequential write performance. This is compared to a state-of-the-art SATA SSD, which is limited to just 540 MB/s read and 520 MB/s sequential writes.[2] This substantial difference in performance makes NVMe SSDs a much better fit for high-speed, high-volume analytics platforms. When combined with the right DRAM, the results can be even more impressive.

In this document, we show:

- A Microsoft SQL Server® configuration with different DRAM scaling options
- The effect of DRAM scaling on single-stream and multiple (multi)-stream performance
- SQL tuning options and their effects
- How to select SSDs with endurance based on the appropriate workload

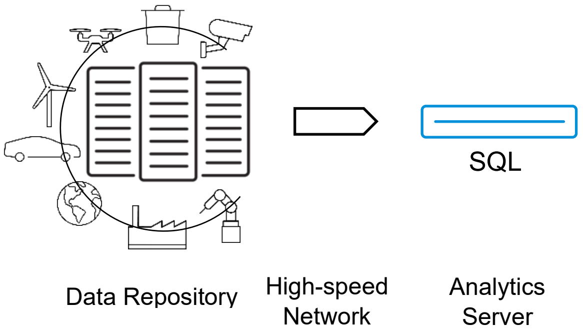

Figure 1: Testing layout overview[3]

Tests were performed on a Dell® EMC PowerEdge R7525 server with two AMD® EPYCTM 7713 (64-core, 128-thread, 2.0 GHz) CPUs, four 3.84TB Micron mainstream SSDs with NVMe, and 1TB of DRAM (16 dual-rank, 64GB Micron DIMMs in a one DIMM per channel configuration).

FOUR THINGS TO KNOW

Adding Memory Can Accelerate Query Processing

Expanding the memory capacity provides dramatic increases in overall system performance for single- and multi-stream analytics workloads. It is simple to improve an analytics server by adding additional memory. However, there is a point of diminishing returns.

For single stream workloads a 1:4 DRAM to dataset ratio proves optimal. For multi-stream we tested up to a 1:3 ratio and continued to see increasing performance.

Choosing the Right SSD Endurance

Increasing memory has the side effect of increasing the overall write rate to TempDB in the multi-stream use case. The increases mean that care must be taken when choosing SSD endurance.

Tuning SQL Degree of Parallelism

SQL Server allows direct control over its internal degree of parallel operations. The degree of parallelism (DoP) value dictates the number of worker threads used for a particular query. Different values for DoP can affect SQL performance and use models.[4]

Using Benchmarks Designed for Analytics Platforms

A decision-support benchmark can be used to run an online analytics processing (OLAP) workload based on the TPC-H specification.[5] When testing a design, it is important to use benchmarks for business-oriented, ad-hoc queries.

---

1. In this document, performance means IOPS, MB/s, latency or any combination thereof.
2. Additional information on the NVMe SSDs used is available here: www.micron.com/7450; additional information on the SATA SSDs used for comparison is available here: www.micron.com/5300.
3. Example deployment for the type of SQL analytics platform tested. Other deployments and configurations may produce different results.
4. Microsoft provides additional details here: https://docs.microsoft.com/en-us/sql/database-engine/configure-windows/configure-the-max-degree-of-parallelism-server-configuration-option?view=sql-server-ver15. Different values may give different results.
5. The TPC-H specification benchmark measures decision-support system performance. The TPC home page has additional details: http://www.tpc.org/tpch/#:~:text=The%20TPC%2DH%20is%20a,have%20broad%20industry%2Dwide%20relevance.

ADD MEMORY TO SPEED SQL QUERY EXECUTION

Scaling Memory, Single-Stream Operations

Single-stream testing measures the time taken to process a single query set (22 queries). Increased performance means each query takes less time to run (or, alternatively, that the platform can complete more queries for a given period — expressed as power test results).

Figure 4 shows how single-stream performance increases with memory scaling (lower y-axis values correspond to increased performance).

A baseline performance was established (1,187 seconds run time) using 256GB of memory. Table 2 shows that scaling memory reduces query run time (non-linearly).

For example, increasing the memory from 256GB to 512GB resulted in a 28% run time reduction relative to the baseline, while increasing the memory from 512GB to 768GB showed an additional 23% reduction (relative to 512GB). And increasing from 768GB to 960GB resulted in an additional 7% run time reduction (relative to 768GB), suggesting a point of diminishing returns. The database size in this test is 3TB indicating a DRAM to dataset size of 1:4 is optimal in this case.

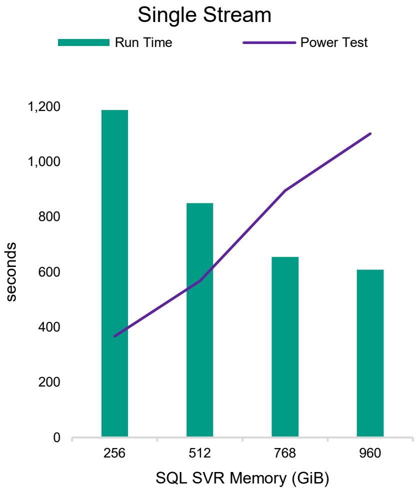

Figure 4: Single-stream performance

SQL SVR Memory (GB)	Run Time (sec)	Power Test	Reduction
256	1,187	305,728	Baseline
512	850	473,466	28%
768	655	745,662	23%
960	608	918,042	7%

Table 2: Memory increases and diminishing single-stream returns

Scaling Memory, Multi-Stream Operations

Multi-stream results reflect multiple queries being submitted by multiple, concurrent users; completion time for all queries is measured. Lower run time is better.

Figure 5 shows how multi-stream performance increases with memory scaling.

Table 3 shows that increasing memory from 256GB to 512GB showed a 37% run time reduction, while a memory increase from 512GB to 768GB resulted in a 36% run time reduction. Finally, as memory increased from 768GB to 960GB, we showed a 21% reduction, indicating a starting point for diminishing returns.

Multi stream testing accesses more of a given dataset, increasing the utilization of DRAM resulting in increased scaling. The DRAM to dataset ratio here is 1:3, with the expectation that more memory will continue to scale performance.

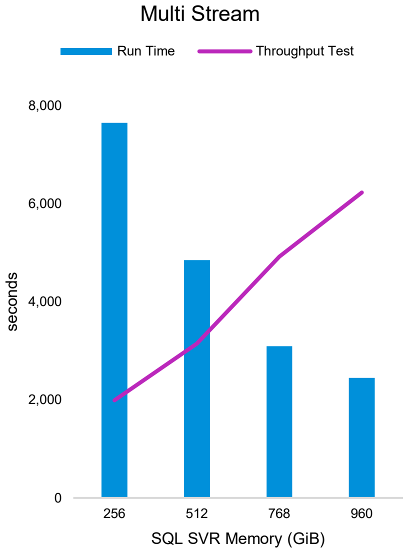

Figure 5: Multi-stream performance

SQL SVR Memory (GB)	Run Time (sec)	Throughput Test	Reduction
256	7,644	248,799	baseline
512	4,847	393,426	37%
768	3,095	614,723	36%
960	2,444	778,323	21%

Table 3: Memory increases and diminishing multi-stream returns

SELECT THE RIGHT SSD ENDURANCE

When selecting the optimal SSD configuration for an analytics workload, endurance may not be a primary consideration. This is because analytics (DSS/OLAP) workloads are typically thought of as read-intensive. Figure 6 shows that the endurance story is more complex.

SQL Server Memory (GB)	Streams	Write Rate (MB/s)	Platform TBW in 5 Years (PB)
256	1 (single)	108.6	16.0
512	1 (single)	82.4	12.1
768	1 (single)	102.2	15.0
960	1 (single)	107.7	15.8
256	8 (multi)	99.4	14.6
512	8 (multi)	124.1	18.2
768	8 (multi)	188.7	27.7
960	8 (multi)	235.8	34.7

Table 4: SSD total bytes written

As additional memory was inserted into the system (moving along the x-axis from left to right in Figure 6), the write rate in MB/s changed.

Table 4 shows the total bytes written (TBW) to storage over a planned platform lifecycle of five years (assuming 24x7 system operations). The four Micron 7450 3.84TB PRO SSDs used in this analysis each had a total bytes written (TBW) rating of 7,000TB for 4k random workloads. The total TBW of all four SSDs was 28,000TB or 28PB.

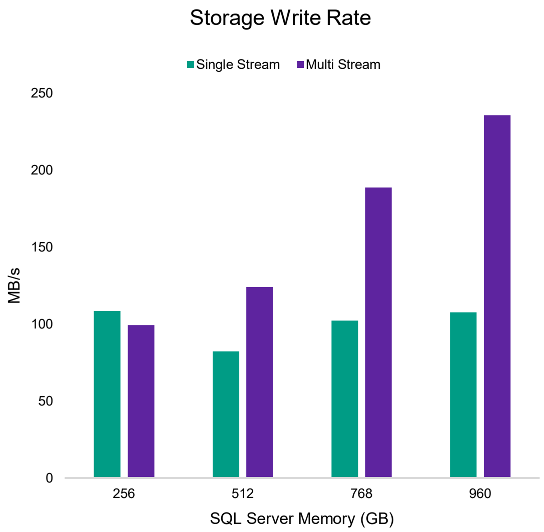

Figure 6: Single-stream and multi-stream write rates

Table 4 also shows that the 960GB memory configuration used for multi-stream (8 streams) testing exceeded this TBW value (34.7PB vs. 28PB). Since this is the only tested use case that exceeds the available TBW for the tested SSDs, the 3.84TB PRO SSD endurance is sufficient for all but this one use case.

Remedies are straightforward:

- **Spread TempDB across all the data SSDs.** This allows the system to use the endurance of all the SSDs (data files are written less infrequently than TempDB). Given the size of typical data warehouses, the number of SSDs required to reach the needed capacity should be sufficient to meet the endurance requirements.
- **Use SSDs with higher endurance for TempDB.** The Micron 7450 MAX 3.2TB SSD has a rated endurance of 17,500TB written for 4k random workloads. Sizing for the data warehouse is then a matter of measuring bytes written over time and adding additional devices for TempDB (or sizing to larger devices) to meet the endurance requirements.
- **Use additional SSDs.** Adding just one 3.84TB Micron 7450 PRO SSD adds 7,000TB of additional endurance (5 x 7,000TB for a total of 35,000TB). This combination provides the required endurance for 960GB SQL memory in the multi-stream use case shown in this document.

USE THE OPTIMAL DEGREE OF SQL PARALLELISM

SQL Server allows direct control over its internal DoP — a value that dictates the number of worker threads used for a particular query. Different values for DoP can affect SQL performance. Based upon empirical observations during single-stream and multi-stream tests an optimal DoP was established for final testing.

Optimal Single-Stream DoP for Analytics Workloads Is 192

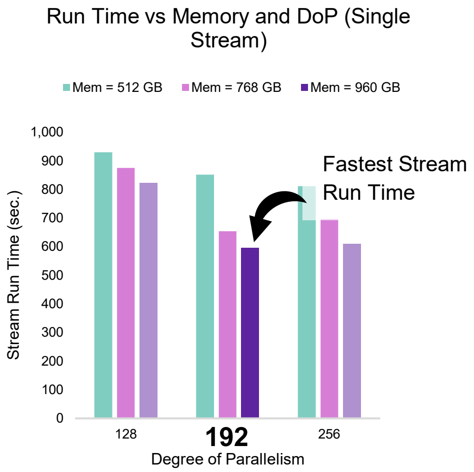

Figure 2: Single-stream DoP

A query stream is a sequential execution of a single query set submitted by a single emulated user. We used a power[6] test to measure single-stream queries executing DoP values from 128 up to 512. In Figure 2, lower Stream Run Time is better.

When all memory was available to the SQL Server (960GB), optimal DoP (or the DoP value that resulted in the shortest run time) was found empirically to be 192.

Optimal DoP values may vary based on configuration. Empirical determination helps ensure optimization.

Optimal Multi-Stream DoP Is 24

Multi-stream results reflect “batch” query processing, which means multiple queries submitted by multiple users, all completing as quickly as possible. This testing focused on measuring multiple query streams executing. In this case, lower Stream Run Time is also better.

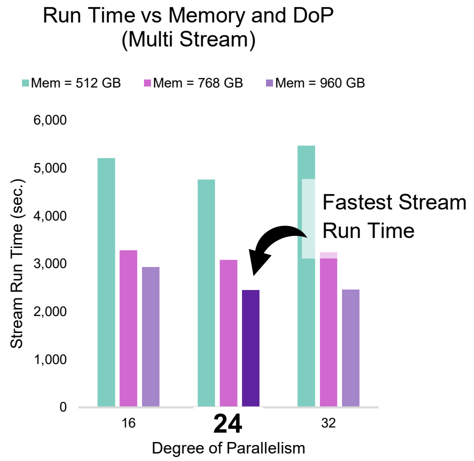

Figure 3: Multi-stream DoP

The minimum number of streams used was based on the scale factor of the data set.[7] We used a scale factor of 3,000, and hence 8 streams, as shown in Table 1.

When all memory was available to the SQL Server (960GB), optimal DoP was found empirically to be 24.

This value was used for all multi-stream testing.

Scale Factor	Streams
1	2
10	3
30	4
100	5
300	6
1,000	7
3,000	**8**
10,000	9
30,000	10
100,000	11

Table 1: Scaling factor and streams

---

[6] TPC-H documentation: http://tpc.org/tpc_documents_current_versions/pdf/tpc-h_v3.0.0.pdf. A power test is the measurement of a single-query stream executing with access to all available resources on the test system.

[7] More details about TPC-H benchmark and scaling factors are available here: https://docs.deistercloud.com/content/Databases.30/TPCH%20Benchmark.90/index.xml. Different values may give different results.

CONCLUSION

Data analytics opportunities are as massive as data growth itself. In May of 2021, Globe Newswire noted that the 2021 global data analytics market was estimated at $25 billion (USD) in 2021, with a revenue compound annual growth rate (CAGR) of 25% during the forecast period to 2030.[8]

When designing SQL analytics platforms, these four elements are essential to consider:

1. **Adding Memory:** Additional memory can speed analytics queries. While this is one of the simplest ways to improve performance, there is a point of diminishing returns.
2. **SSD Endurance:** Increasing memory has a side effect — it also increases the overall write rate to storage. It is necessary to plan for SSD endurance appropriately.
3. **DoP:** This directly-tunable SQL Server parameter dictates the number of worker threads used for a particular query. Different values for DoP can affect SQL performance and use models. Higher DoP is not always better.
4. **Test With Analytics Benchmarks:** When testing a design, it is important to use benchmarks for business-oriented, ad-hoc queries.

Sizing the analytics platforms correctly, understanding the resultant workload changes, and identifying points of diminishing returns can help ensure that a system is built right the first time.[9]

---

[8] Global Newswire estimates for the global data analytics market, not just SQL analytics: https://www.globenewswire.com/en/news-release/2021/05/18/2231324/0/en/Global-Data-analytics-Market-Size-to-Grow-with-a-CAGR-of-25-from-2021-to-2030.html

[9] This paper uses the Micron 7450 SSD with NVMe. Additional details on this and other Micron data center SSDs are available here: https://www.micron.com/products/ssd/usage/data-center-ssd. Micron DRAM product information is available here: https://www.micron.com/products/dram

TESTING DETAILS

Hardware

Database Server	Qty.	Description
Server	1	DellEMC PowerEdge R7525
CPU	2	AMD EPYC 7713 (64-core, 128 threads) @ 2.0 GHz
Memory	1	TB Micron DRAM
Storage: Database	4	Micron 7450 SSD with NVMe, 3.84 TB
Storage: OS	1	Dell Express Flash CD5, 960GB

Table 5: Hardware configuration

Software

Database Server	Ver.	Description
Ubuntu	20.04	Operating System
Microsoft SQL Server	2019	Enterprise Core Edition (x64)

Load Gen. Server	Ver.	Description
TPC QGEN	2.17.3	

Table 6: Database and load-generation software

Test Process

The following steps were completed prior to testing:

1. Restored SSDs to FoB[10] state (via NVMe format)
2. Created partitions or namespaces for:
    - Data
    - TempDB

   Note: During testing, the refresh functions were not used. Consequently, there was no activity to the SQL logs, so they remained on the boot device with the SQL Server application. Your configuration may differ.

3. Restored database from a previously created backup
4. Created columnstore indexes
5. Gathered statistics

For each test run, the following steps were completed:

1. Set database parameters
2. Restarted database
3. Started data captures
4. Began query execution
5. Completed queries
6. Stopped data captures

Measuring Performance: Total Queries per Hour

Note: This paper does not contemplate official TPC-H submissions. This section is intended for background information only.

When submitting official TPC-H results, the following calculation is used to produce a single output from a test run:

$$Power@Size = \frac{3600 \times ScaleFactor}{\sqrt[24]{\prod_{i=1}^{i=22} QI(i,0) \times \prod_{j=1}^{j=2} RI(j,0)}}$$

$$Throughput@Size = \frac{NumStreams \times 22 \times 3600}{Full\ Stream\ Time} \times ScaleFactor$$

$$QppH@Size = \sqrt{Power@Size \times Throughput@Size}$$

In short, the Power@Size value is effectively the geometric mean of the query and refresh function times, scaled with the ScaleFactor value. The Throughput@Size value is simply the naïve queries per hour scaled by the ScaleFactor. The final output for an official TPC-H run is the geometric mean of the power and throughput performance metrics.

For this report, we modified the Power@Size equation slightly:

$$Power@Size = \frac{3600 \times ScaleFactor}{\sqrt[22]{\prod_{i=1}^{i=22} QI(i,0)}}$$

Because we did not perform refresh functions, we used the geometric mean of the 22 queries and scale-by-scale factor.

---

10. For a better understanding of SSD performance states, see “Understanding SSD Performance Using the SNIA SSS Performance Test Specification,” Section 2: https://www.snia.org/sites/default/files/UnderstandingSSDPerformance.Jan12.web_.pdf