A New Frontier in Power-Efficient OLTP

Micron® 5100 MAX SSD: More Users, Faster, Better Results for Demanding Databases

Solid state drives (SSDs) have migrated from specialized, high-cost, niche storage to IT's mainstream toolkit, with different SSD designs available to satisfy different workloads—from I/O-intensive to read-focused.

This migration to mainstream IT has many thinking differently about SSDs, their advantages and how to best use them and measure their effectiveness and value.

High IOPS performance is expected from SSDs, so now we tend to focus on other advantages like greater power efficiency and lower, more consistent response times. For example, we may explore database operations per watt (of system-level power consumed) or responsiveness (on average) or response consistency. We may be interested in how many users a given drive configuration or server node can reasonably support and the database throughput when the system is heavily loaded[2].

SSDs like the Micron® 5100 MAX can bring these benefits to I/O-intensive applications like PHP: Hypertext Preprocessor (PHP) with a database backend—a common use for MySQL®.

MySQL has its own distinct advantages in the relational database management system (RDBMS) space. Powering some of the largest, high-growth websites, MySQL remains free[3], broadly deployed[4] and very flexible with extensive community support.

In this paper, we compare MySQL Community Edition with OLTP workload test results (New Orders Per Minute [NOPM], NOPM/watt of system power, as well as average and 90th percentile transaction response times) to see how the high IOPS performance of the 5100 MAX with MySQL enables compelling OLTP results compared to legacy storage.

5100 MAX SSDs[1], MySQL and OLTP: Throughput, Power Efficiency, Low and Consistent Responses Supporting More Users[2]

5100 MAX vs. 8x HDDs

Metric	4x 5100 MAX	8x 5100 MAX
NOPM	31X	40X
Efficiency[2]	41X	43X
Average Response Time	95% lower	96% lower
90th Percentile Response Time	91% better	92% better

1. 4x 5100 MAX SSDs (RAID 10) and 8x 5100 MAX SSDs (RAID 10) compared to 8x 15K RPM HDDs (RAID 10).
2. Power efficiency in NOPM/watt. All comparisons made at user count saturation for each configuration – 'heavily loaded' defined as user count saturation. Details on user count saturation provided in "How We Tested" later in this paper. Ratios are NOPM/watt for each 5100 MAX configuration divided by NOPM/watt for the HDD configuration.
3. MySQL GPL licensing: https://www.mysql.com/about/legal/
4. Deployment data from https://www.mysql.com/why-mysql/

New Orders Per Minute and Number of Users By Configuration

We tested the HDD and 5100 storage configurations and assumed that anyone deploying MySQL would want to optimize their results, supporting as many users as possible until the platform is saturated[5].

We began testing with 32 users and increased the number of users until saturation. Once we reached each configuration's saturation point, we stopped adding users and recorded test results for comparison.

At saturation, the 4x 5100 MAX configuration shows 31X higher NOPM than the HDD configuration, while the 8x 5100 MAX configuration shows 40X higher than the HDD configuration.

Figure 1 shows NOPM (blue bar) and average response time (orange line) by user count for the 8x 15K RPM HDD configuration.

Figures 2 and 3 show the same results for the 4x 5100 MAX and 8x 5100 MAX configurations, respectively.

The saturation point (user count)[5] for each configuration is highlighted in blue.

Table 2 summarizes the user count, stop condition and measured NOPM for each configuration as well as the 5100 MAX configuration NOPM/15K RPM HDD NOPM ratio for each 5100 MAX configuration (NOPM ratio).

Configuration	User Count[6]	Stop Condition	Measured NOPM[7]	NOPM Ratio
8x 15K RPM HDD	32	Delivery 90% transaction time <5 seconds	1,821	---
4x 5100 MAX	256	NOPM plateaus	56,783	31x
8x 5100 MAX	320	NOPM plateaus	73,188	40x

Table 2: Saturation User Count, Stop Condition and Measured NOPM

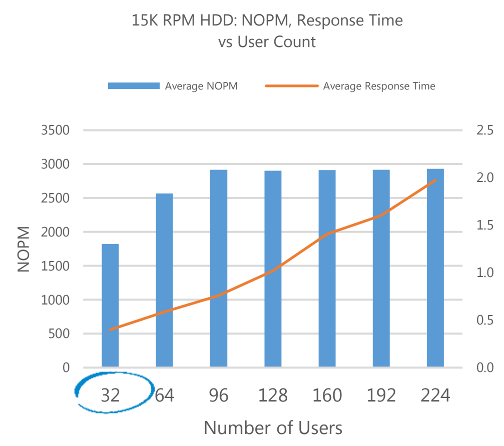

Figure 1: 8x HDD Configuration NOPM

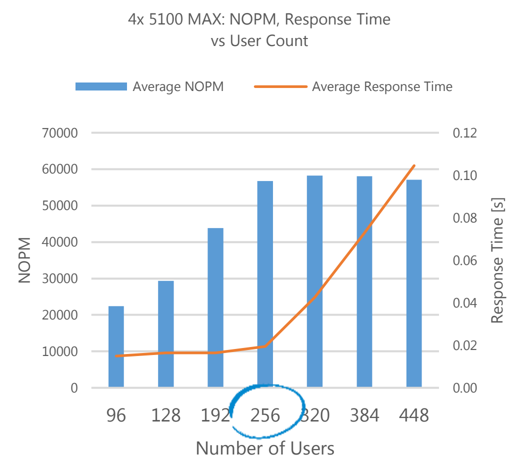

Figure 2: 4x 5100 MAX Configuration

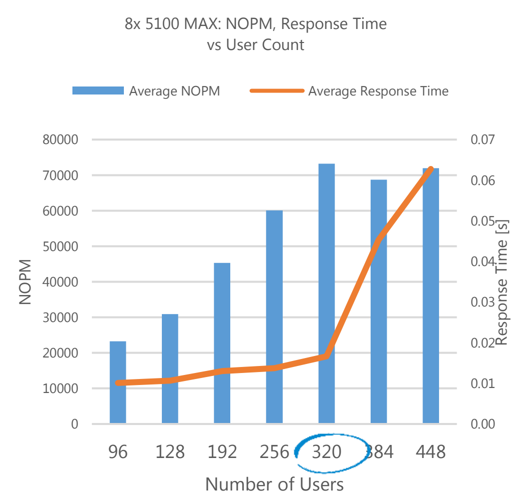

Figure 3: 8x 5100 MAX Configuration

5. Saturation based on OLTP test stop conditions suggested in the TPC-C specification. See How We Tested at the end of this document to see how we determined saturation for each configuration. For more details on TPC-C, see http://www.tpc.org/tpcc/
6. Once a user count met a stop condition, we did not add more users. For this and all subsequent testing, we fixed the user count to the value at which the stop condition was observed.
7. NOPM measured at the saturation user count

Calculating Power Efficiency (NOPM per Watt)

We measured system level power consumption (in watts) at each configuration's saturation user count during NOPM testing. We divided NOPM by power consumed to calculate a power-efficiency metric: NOPM per Watt for each.

This calculated metric shows measured NOPM for each configuration (again, at user count saturation) for each watt consumed. A larger number (more NOPM per Watt) indicated better power efficiency.

Figure 4 shows the results of these calculations—higher is better.

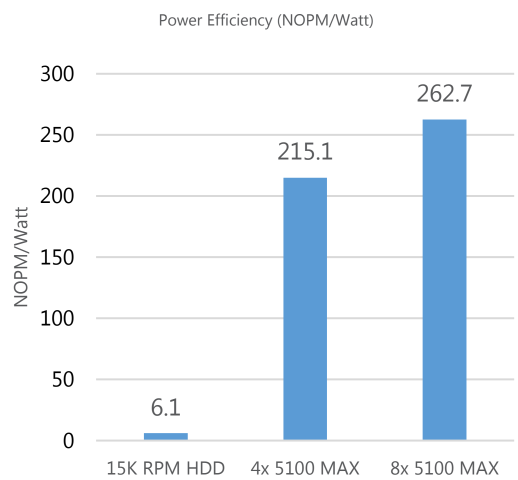

Figure 4: NOPM Per Watt (Calculated Power Efficiency)

The 15K HDD configuration generated an average of 6.1 NOPM per Watt while the 4x 5100 MAX generated an average of 215. The 8x 5100MAX generated an average of 262.

These differences are primarily due to:

- The 5100 MAX configurations draw slightly less average overall power during normal operation than the HDD configuration (other system components draw approximately the same power between configurations, negating their effect).
- The 5100 MAX configurations generate far more NOPM that the HDD configuration.

Table 3 summarizes these results and shows calculated NOPM/watt vs HDD ratios.

Configuration	User Count[8]	System Power	NOPM/Watt[9]	NOPM/Watt vs HDD
8x 15K RPM HDD	32	300 Watts	6.1	---
4x 5100 MAX	256	264 Watts	251.1	41x
8x 5100 MAX	320	279 Watts	262.7	43x

Table 3: Power Efficiency (NOPM/Watt)

8. Once a user count met a stop condition, we did not add more users. For this and all subsequent testing, we fixed the user count to the value at which the stop condition was observed.
9. NOPM measured at the saturation user count

Measuring Average Transaction Response Time (Latency)

For many OLTP deployments, fast database response (low average latency) is imperative. We measured average transaction response time for each configuration. Figure 5 shows these results.

The 8x 15K RPM HDD configuration showed an average transaction response time of 0.399 seconds, while the 4x 5100 MAX configuration showed 0.018 seconds and the 8x 5100 MAX showed 0.017 seconds.

Comparing the HDD configuration response times to those of the 4x 5100 MAX shows that the 4x 5100 MAX configuration is 95% lower while 8x 5100 MAX is 96% lower than the HDD configuration.

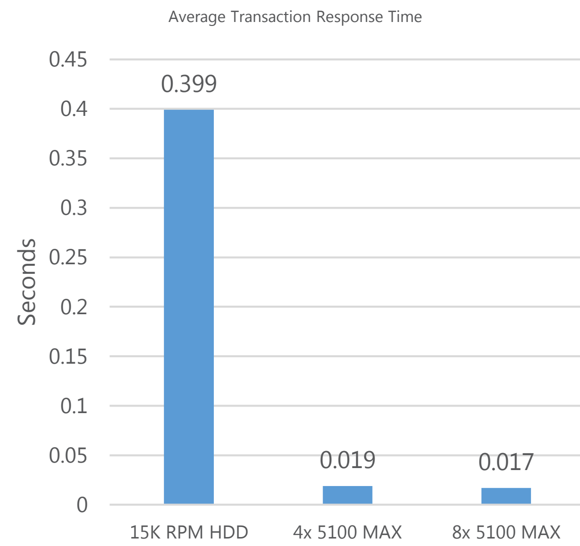

Figure 5: Average Response Time

It is important to note that these average transaction response times were measured while the 5100 MAX configuration was generating far greater NOPM than the 15K RPM HDD configuration.

Table 4 summarizes these results and shows each 5100 MAX configuration's average response time reduction (compared to the HDD average response time) as a percentage.

Configuration	User Count	Average Response Time	Average Response Time Reduction
8x 15K RPM HDD	32	0.399 sec.	---
4x 5100 MAX	256	0.019 sec.	95%
8x 5100 MAX	320	0.017 sec.	96%

Table 4: Average Response Time

Measuring Average Transaction Response Time Consistency

Low average transaction response time is highly desirable in high-throughput OLTP systems—a database that can respond quickly can offer better results. For some implementations, latency consistency (measured as 90th percentile transaction response time) may be more important. The 90th percentile transaction response time metric shows the time in which 90% of the transactions complete.

If an OLTP system can't tolerate an occasional outlier (during which the database response takes much longer than average), 90th percentile transaction response time (latency consistency) may be a primary driver. Figure 6 shows these results.

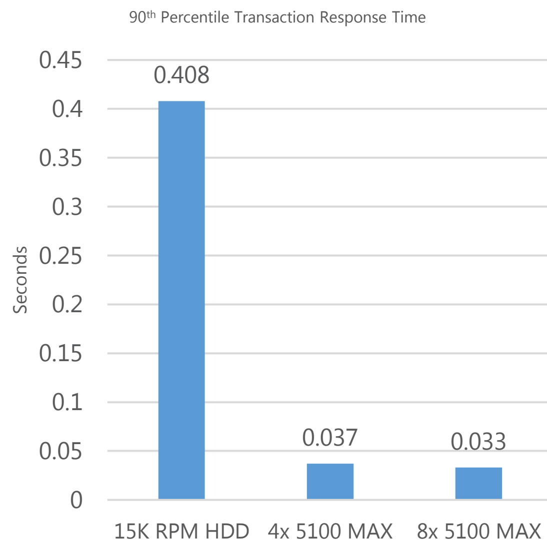

Figure 6: 90th Percentile Transaction Response Time

We measured the 90th percentile transaction response time for each configuration (again, at their saturation user counts) and compared the results (a lower 90th percentile transaction response time means the database responses are more consistent). As with average transaction response times, more consistent transaction response time is better. These results are summarized in Table 5.

Configuration	User Count[6]	90th Percentile Response Time	90th Percentile Response Time Reduction
8x 15K RPM HDD	32	0.408 seconds	---
4x 5100 MAX	256	0.037 seconds	91%
8x 5100 MAX	320	0.033 seconds	92%

Table 5: 90th Percentile Transaction Response Time

Summary

In this technical brief, we showed how MySQL Community Edition with an OLTP workload excels on two different 5100 MAX SSD arrays (4x and 8x 5100 MAX, RAID 10) when compared to legacy storage. We demonstrated that combining the openness of MySQL with the low power and high IOPS performance of the 5100 MAX SSD can enable compelling results.

We've come to expect great results from SSDs, and IT's focus is turning to the real advantages SSDs can offer in the data center—metrics like power efficiency and lower, more consistent response times.

When we look at database throughput from open platforms like MySQL, metrics like NOPM per watt and the number of users we can really support before our platform is saturated matter and help drive deployment choice. Choose wisely. Choose the 5100 MAX SSD for MySQL/OLTP.

Learn more at micron.com.

How We Tested

Storage Configurations

We used a database server (2U rack-mount with two Intel® Xeon™ 14-core processors, 256GB of memory and an LSI®/Avago 3108 RAID on chip (ROC)-based hardware controller with default options set and a load generator (VM residing on a similar server as part of a VM cluster, its CPU and memory remained consistently with the VM sized to ensure it was not a bottleneck) connected via a 10 Gb/E switched network. We used an Avago-based hardware RAID controller configured as 'pass through' and used host-based RAID (MDADM) to manage drive RAID.

Drive Type	Number of Drives	RAID
5100 MAX	4	10
5100 MAX	8	10
15K RPM HDD	8	10

Table 6: Storage Array Configuration

Software

- Database server: CentOS® 7.3 with MySQL Community Edition 5.7
- VM Load Generator Server: VMware® ESXi™ 6.0.0, VM OS: CentOS 7.3

Schema Sizing

Our test creates nine interrelated tables, of which the basic building block is a warehouse. For each warehouse, there are 10 districts and each district serves 3,000 customers. Each warehouse has 100,000 items it stocks, and these items compose new orders. When the schema is created, there are at least 30,000 'historic' orders per warehouse—one per customer, minimum.

For this paper, we used a database with 10,000 warehouses resulting in the population of records for each of the tables as shown in Table 7.

With indexes, the total database size is roughly 1TB when testing begins. This resulted in a dataset that was larger than the memory available to the database.

Table	Rows
WAREHOUSE	10,000
DISTRICT	100,000
ITEM	100,000
NEW_ORDER	90,000,000
CUSTOMER	300,000,000
HISTORY	300,000,000
ORDER	300,000,000
STOCK	1,000,000,000
ORDER_LINE	3,000,000,000

Table 7: Schema Sizing

Configuring Users and Warehouses

We used a test process that ensures when a user is created, that user is assigned a list of warehouses, such that all warehouses are assigned to a user, and each warehouse is assigned to only one user. This helps ensure better consistency (because warehouses are not randomly assigned) and lets us easily change the database size without changing the number of database connections (as might happen if there were fewer users than warehouses). For example, if we ran the test with 20 warehouses and five users, the assignments would be as shown in Table 8.

User	Assigned Warehouses
1	1, 6, 11, 16
2	2, 7, 12, 17
3	3, 8, 13, 18
4	4, 9, 14, 19
5	5, 10, 15, 20

Table 8: Schema Sizing

This results in more consistent performance from one run to the next. Additionally, the entire dataset is accessed equally instead of as a hot dataset size that is dependent on the number of users. (Note: It's important to realize that these changes mean that comparisons between these numbers and other published data may lead to inconsistent and/or inconclusive results.)

Keying and Think Time

The TPC-C specification describes two delays between transactions: Keying Time and Think Time. To address this in our testing, we use a parameter for 'static delay' that sets a delay between every transaction, allowing us to slow the transaction rate per user by a consistent amount. We tuned this parameter along with the number of 'users' to scale the load until we find the maximum performance and user count saturation point. For this testing, the static delay is set to 100ms for the 5100 configurations, and 250ms for the 15K HDD configuration.

Stop Conditions

We increased system load (user count) until we reached one or more of the following limits:

- 80% CPU utilization (not observed during testing)
- Five-second 90% average transaction response times
- NOPM plateaus
- Sharp increase in response time (a.k.a. "kink")

Table 9 shows the number of users for all tests for each configuration. This table should be carefully examined and understood to put measured results in context.

Configuration	User Count	Stop Condition
8x 15K RPM HDD	32	DELIVERY 90%h percent transaction time <5 seconds
4x 5100 MAX	256	NOPM plateaus
8x 5100 MAX	320	NOPM plateaus

Table 9: Saturation user count, stop condition and measured NOPM

All measurements and calculated values are based on the stated user counts. For the 8x 15K RPM HDD configuration, we found that if we tested with more than 32 users, the DELIVERY 90% response time exceeds five seconds. For the 4x 5100 MAX and 8x 5100 MAX configurations, we used 256 and 320 users, respectively, for all tests. At 256 users, the 4x 5100 MAX configuration reached a NOPM plateau; at 320 users, the 8x 5100 MAX configuration did as well. These three stop conditions set the tested user counts for all metrics and all calculated values.

90% Transaction Response Time

We used 90% response time stop conditions for each of the five transactions as noted in Table 10 (see section 5.2.5.7 of the TPC-C specification from Transaction Processing Council, 2016; for more details see http://www.tpc.org/tpcc/).

For our testing, we stopped increasing load if any of the individual transaction 90% response times exceed these limits.

Transaction	90% Transaction Response Time Stop Condition
New Order	5 seconds
Payment	5 seconds
Order Status	5 seconds
Delivery	5 seconds
Stock Level	20 seconds

Table 10: 90% Transaction Response Time Test Stop Conditions