Seagate Nytro Flash Accelerator vs. the Competition

Technology Paper

The Seagate® Nytro™ Accelerator Card Outperforms the Competition in Performance Using an Oracle MySQL Database

This document is a guide to help increase MySQL database performance by implementing the Seagate Nytro flash accelerator card. By offering high performance with low latency and minimal CPU burden, the Nytro accelerator card can maximize transactional I/O performance for applications like online transaction processing (OLTP), data warehousing (DW) and data mining. The Nytro accelerator card is designed to improve user satisfaction and productivity by enabling applications that use the MySQL database to meet challenging service-level agreements (SLA) for response times and throughput.

This document also demonstrates how the Nytro accelerator card compares with a leading competitor's PCIe flash-based card running the same performance benchmark on identical hardware and database configurations. Finally, this document describes a reference architecture as a basis for planning and deploying PCIe-compatible flash storage, including a typical use case scenario.

Why Use Flash-based Storage with a MySQL Database?

As OLTP and DW applications have demanding requirements for short response times and high throughput, it makes it difficult for database administrators (DBAs) to maintain these systems as the number of users grows and the amount of data increases. During the application life cycle, performance bottlenecks might originate in one or more areas, including the network, processor or storage devices. However, correcting a bottleneck in one area may actually cause a bottleneck in another.

Seagate Nytro Flash Accelerator vs. the Competition

Flash-based storage provides performance levels that fall between the performance levels of HDDs and DDR3 memory. The recent innovation of mounting SSDs on a PCIe card can also alleviate throughput constraints that are caused by the storage interface. The Nytro accelerator card supports bandwidth up to 4GB/s, up to 300,000 IOPS and capacity up to 3.2TB.

Overview of the Seagate Nytro Accelerator Card

The Nytro accelerator card offers high performance with low latency and minimal CPU burden. This card is designed to maximize transactional I/O performance for MySQL databases and for other applications that require high-performance computing cycles. The Nytro accelerator card performs consistently across reads and writes, regardless of workload, by using industry-standard and widely deployed Seagate SAS software for easier system integration and management.

Configuring Linux and MySQL for Optimum I/O Performance

This section describes the configuration steps for both Linux and the MySQL database using the Nytro accelerator card.

To configure the Seagate Nytro accelerator card as a filesystem

You can tune the Nytro accelerator card for enhanced performance using Linux. In fact, the configuration steps that follow work in Linux as well as Oracle Linux with Unbreakable Enterprise Kernel (UEK). The following steps configure the Nytro accelerator card as a single filesystem for the MySQL database. Other options would be to use multiple cards and apply RAID to provide fault tolerance for extra data protection.

1. Align the Nytro card on a 1MB boundary in Linux. The following example shows how to partition the Nytro card as a single partition starting on a 1MB boundary:

```
echo "2048,," | sfdisk -uS /dev/sda

Disk /dev/sda: 24321 cylinders, 255 heads, 63 sectors/track
Units = sectors of 512 bytes, counting from 0
Device Boot       Start     End          #sectors     Id     System
/dev/sda1         2048      390721535    390719488    83     Linux
/dev/sda2         0         -            0            0      Empty
/dev/sda3         0         -            0            0      Empty
/dev/sda4         0         -            0            0      Empty
```

NOTE: The Nytro device can have any device address. In this case, the Nytro card has `/dev/sda`.

2. For creating an EXT-4 filesystem, use this line:

```
$ mkfs -t ext4 /dev/sda1 or mkfs.ext4 /dev/sda1
```

3. Mounting the new EXT-4 device (`noatime,nodiratime,max_batch_time=0,nobarrier`) options are described below:

```
$ mount -o noatime,nodiratime,max_batch_time=0,nobarrier/dev/sda1/mountpoint
```

4. Modify the kernel I/O scheduler. The I/O schedulers in the latest Linux releases have new I/O capabilities, including options to modify these settings at boot time. In the tests described in this document, the DEADLINE I/O scheduler was used with the file system mount options noted above. To invoke the DEADLINE I/O scheduler for the entire system, add this line to the `/etc/grub.conf` file and bounce the system.

   ```
   $ kernel /vmlinuz-2.6.39-300.28.1.el6uek.x86_64 ro root=/dev/mapper/vg_nytro-lv_root
   elevator=deadline
   ```

   An alternative way to invoke the DEADLINE I/O scheduler is to modify the `/etc/rc.local` file with the following command:
   `$ echo "deadline" > /sys/block/sda/queue/scheduler`

   To verify that the DEADLINE I/O scheduler is enabled, issue the following statement as root.

   ```
   $ cat /sys/block/sda/queue/scheduler noop anticipatory [deadline] cfq
   ```

NOTE: In later Linux releases, the I/O scheduler defaults to "deadline".

5. In addition to changing the I/O scheduler, the `noatime`, `nodiratime`, `max_batch_time=0`, `nobarrier` file system mount options were invoked, which were added to the `/etc/fstab` file. These options eliminate some of the default behavior of the HDD that in turn increases the performance of the Nytro accelerator card. By applying these options, you will have some of the following benefits:
   - Eliminate the need for the system to create writes to the file system when objects are only being read. These options also enable faster access to the files, plus the benefit of less wear on the Nytro card.
   - Disable write barriers
   - Disable synchronous transaction batching
   - This example shows how the `/etc/fstab` entry invokes the `noatime`, `nodiratime`, `max_batch_time=0`, `nobarrier` options:
     `/dev/sda1 /u04 ext4 defaults,noatime,nodiratime,max_batch_time=0,nobarrier 1 2`

   An alternative to invoking the `noatime` option is to specify it when executing the `mount` command:
   `$ mount -o noatime,nodiratime,max_batch_time=0,nobarrier /dev/sda1 /u04`

6. Performance of the Nytro accelerator card can be improved by increasing queue depth (QD) from the default of 128 to 256 or higher, depending on the database workload. In order to modify QD, the `nr_requests` parameter will need to be modified to the same or larger value of the new QD value. Here are examples of modifying both the `nr_requests` and `queue_depth` parameters for `/dev/sda` which was used in our tests:

   ```
   $ echo "512" > /sys/block/sda/queue/nr_requests
   $ echo "512" > /sys/block/sda/device/queue_depth
   ```

7. Further performance benefits can be recognized by telling the operating system to complete an I/O request on the same CPU that initiated it. In some workloads, this can provide a nice jump in performance. To enable this feature on the Nytro accelerator card, use the following line:

   ```
   echo "2" > /sys/block/sda/queue/rq_affinity
   ```

8. Another operating system setting that can provide more performance is configuring the rotational setting to 0 (zero) which identifies the device as nonrotational, meaning it is a flash device.

   ```
   echo 0 > /sys/block/sda/queue/rotational
   ```

NOTE: To persist these settings across reboots, place all of these commands in the `/etc/rc.local` file.

9. For OLTP systems, using the following Deadline I/O Scheduler tuning settings will increase performance:

   ```
   echo 1 > /sys/block/sdX/queue/iosched/fifo_batch
   echo 0 > /sys/block/sdX/queue/iosched/front_merges
   echo 5 > /sys/block/sdX/queue/iosched/writes_starved
   ```

Seagate Nytro Flash Accelerator vs. the Competition

SEAGATE

MySQL Configuration

MySQL configuration was set up exactly the same on both servers. Since this test was to demonstrate the performance of the Nytro flash accelerator card vs. the competitor's card, the buffer memory settings were set on the low side to force more physical I/Os. The MySQL `my.cnf` configuration settings were:

```
[mysqld]
user=mysql
log_bin=/u04/binlog/mysql-bin
sync_binlog=0
binlog_order_commits=1
binlog_format='MIXED'
server_id = 10
datadir = '/u04/datadir'
default-storage-engine=INNODB
sql-mode="STRICT_TRANS_TABLES,NO_AUTO_CREATE_USER,NO_ENGINE_SUBSTITUTION"
max_connections = 600
query_cache_size=0
tmp_table_size=64M
thread_cache_size=100
myisam_sort_buffer_size=256m
key_buffer_size=16m
read_buffer_size=1m
read_rnd_buffer_size=1m
sort_buffer_size = 2m
innodb_additional_mem_pool_size = 128M
innodb_buffer_pool_size = 4G
innodb_adaptive_flushing=off
innodb_thread_concurrency=0
innodb_flush_log_at_trx_commit=2
innodb_read_ahead=0
innodb_max_dirty_pages_pct=60
innodb_monitor_enable='%'
performance_schema=ON
performance_schema_instrument='%=on'
slow_query_log=1
slow_query_log_file = "/u04/slow_query_logfile.log"
innodb_log_group_home_dir = '/u04/logs'
innodb_data_home_dir = '/u04/data'
innodb_file_per_table=1
innodb_log_buffer_size=4M
innodb_log_files_in_group=4
innodb_log_file_size=2G
innodb_flush_method = O_DIRECT
innodb_write_io_threads=64
innodb_read_io_threads=64
innodb_io_capacity=20000
innodb_data_file_path = "ibdata1:10G:autoextend"
expire-logs-days=1
innodb_flush_neighbors=0
innodb_lru_scan_depth=1024
```

NOTE: The default 16k pagesize was used for all benchmarks.

Seagate Nytro Flash Accelerator vs. the Competition

Benchmark Results

Before adding the Nytro accelerator card or the competitor's flash PCIe-based card to the test configuration, Linux and the MySQL database were configured using techniques described in the above sections. OLTP benchmarks were used for these tests. Quest Benchmark Factory software was used for these benchmarks with transactions per second (TPS) and average response time results (Figure 1).

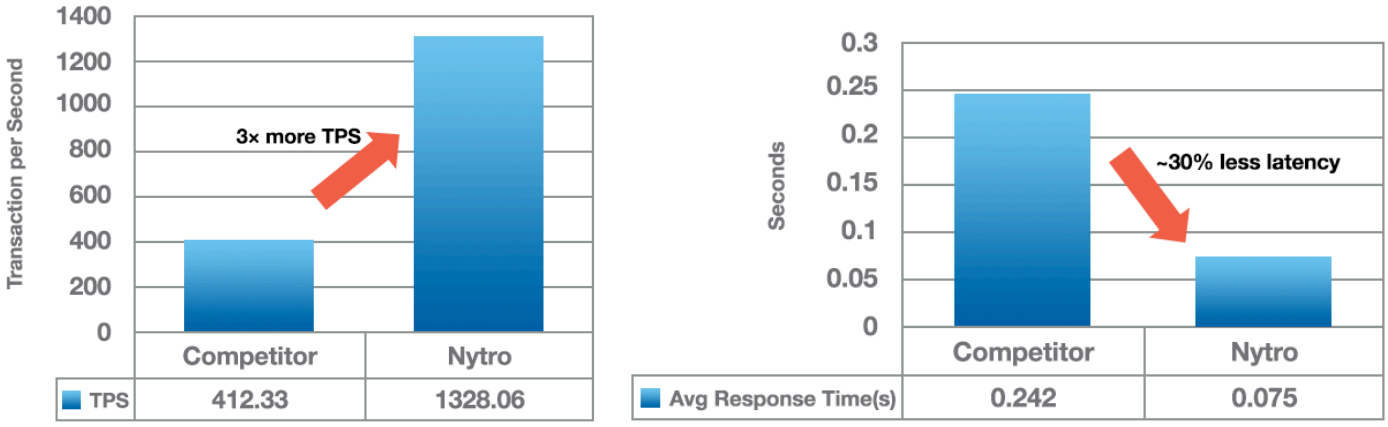

Figure 1. Transactions per Second and Average Response Times Compared

Competitor's Results

The results from the benchmark using the competitor's PCIe flash accelerator card were:

- Average TPS: 412.33
- Average Response Time: 0.242 seconds

Nytro Accelerator Card Results

The results from the benchmark using the Nytro accelerator card were:

- Average TPS: 1328.06
- Average response time: 0.075 seconds

In summary, the Seagate Nytro accelerator card had more than 3× the TPS and ~30% less latency than the competitor.

NOTE: None of the results from in-house testing were audited or published by the Transaction Performance Council (TPC), the copyright owner of TPC-C. The TPC-C transactions are defined according to the TPC-C standard specification.

Seagate Nytro Flash Accelerator vs. the Competition

Conclusion

Based on this benchmark testing with the Nytro accelerator card and a MySQL database, large performance gains were realized over the competition using the configuration presented in this guide for OLTP type of workloads.

The Seagate Nytro accelerator card has demonstrated to be a superior-performing PCIe-based flash solution on the market which will work well in these database benchmark tests. Not only can the Nytro accelerator card be faster than the competition, but the installation, tuning and implementation is markedly more user-friendly due to the collaborative work between Seagate and the Oracle Linux and MySQL groups.

seagate.com

AMERICAS Seagate Technology LLC 10200 South De Anza Boulevard, Cupertino, California 95014, United States, 408-658-1000
ASIA/PACIFIC Seagate Singapore International Headquarters Pte. Ltd. 7000 Ang Mo Kio Avenue 5, Singapore 569877, 65-6485-3888
EUROPE, MIDDLE EAST AND AFRICA Seagate Technology SAS 16–18, rue du Dôme, 92100 Boulogne-Billancourt, France, 33 1-4186 10 00