High Performance MySQL Cluster – Cloud Reference Architecture using 16 Gbps Fibre Channel and Solid State Storage Technology

Evaluation report prepared under contract with Brocade

Executive Summary

As CIOs and datacenter managers consider increasing stresses on existing computing infrastructure and the seemingly insatiable demand for storage and rapid access to that storage, new infrastructure concepts must be investigated. The combination of recently announced 16 Gbps Fibre Channel infrastructure and Solid State Drive (SSD) technology offers a compelling way to significantly improve performance, lower costs and dramatically condense server and storage infrastructure footprints.

Many data centers today are struggling to maintain their SLAs for database performance because of conflicting workloads such as transactions and large data warehouse queries, un-optimized database code and a host of other reasons. One of these causes can be that large database server clusters, such as those running MySQL, are most often built with local disks and large amounts of system RAM in order to get maximum performance. This architecture, while somewhat simple to design, has huge risks with it that can impact performance, scalability and availability. For example, if these local disks fail, an instance of the database server may go down. In addition, this type of architecture can be scaled in only a linear fashion, which severely restricts the number of users per cabinet or rack, and the size of the database per server

Advances in technology are now available which can help overcome these limitations and dramatically improve performance, scalability and availability. With the advent of 16 Gbps Fibre Channel from Brocade and innovative SSD technology from GridIron Systems, the limitations of the existing database cluster architectures can be overcome, servicing significantly more users per cabinet, growing the maximum database size for each server, while significantly lowering capital expense and operating expense, including server costs, power and cooling. In some cases, the footprint savings can be significant, reducing many racks of equipment down to one, while dramatically increasing database transaction rates.

Brocade and GridIron Systems have jointly engineered a high-performance cloud reference architecture for high performance MySQL clusters. This architecture delivers:

- Unprecedented performance of over 1 million IOPS using equipment that fits in half of a datacenter rack.
- Datacenter infrastructure footprint consolidation of 10:1 that reduces density, power and cooling costs.
- CapEx and OpEx savings that are at least four times better than alternatives

Background and Application Environment

Conventional methods of deploying database cluster servers require configuring local disk storage with as much RAM as possible in each individual server so that the entire database can fit into RAM. While this makes for high-performing databases, it artificially limits the size of each database to typical RAM sizes of 64 GB, 96 GB, 128 GB, etc. In order to populate servers with the largest amounts of RAM available requires the purchase of the most expensive type of RAM available, significantly increasing server costs.

Limitations of Current Solutions

This type of local disk and large RAM configuration places limits on the number of users and the number of queries that can be serviced by each database server, and by extension, each server rack. It requires that hundreds or thousands of database servers be deployed to facilitate large environments such as large database clusters, massively parallel processing data warehouse environments and large Software-as-a-Service environments. As a result, the total database size per rack is constrained to a few terabytes and several racks of servers are typically deployed to handle the scalability needs of the environment.

Benefits of the High Performance MySQL Reference Architecture

In these large-scale environments, business revenue is determined, or limited, by the number of users that can be serviced concurrently, the number of database queries that can be run in a given time window, and the total amount of data that can be actively maintained. If a significantly larger number of users can be hosted in the same or smaller datacenter footprint and at lower cost using newer storage architectures, the business benefits are obvious.

This new reference architecture deploys the following technologies in order to significantly increase performance and scalability, and reduce footprint.

- New six-core servers that have one rack unit (1U) form factor
- 16 Gbps Fibre Channel switches and fabric adapters from Brocade
- OneAppliance™ FlashCube™ All-flash/SSD appliances from GridIron Systems

Brocade 6510 Switch

The Brocade 6510 Switch provides exceptional price/performance value, combining flexibility, simplicity, and enterprise-class functionality for virtualized data centers and private cloud architectures. Designed to enable maximum flexibility and investment protection, the Brocade 6510 is configurable in 24, 36, or 48 ports and supports 2, 4, 8, 10, or 16 Gbps speeds in an efficiently designed 1U package. It also provides a simplified deployment process and a point-and-click user interface — making it both powerful and easy to use.

Brocade 1860 Fabric Adapter

Brocade 1860 Fabric Adapters are a new class of "stand-up" network adapter cards for servers with standard PCIe-based multi-function expansion slots. They feature Brocade's industry-unique AnyIO™ technology that allows any individual adapter port to be configured "on-demand" by software command as either a 16 Gbps Fibre Channel (FC) Host Bus Adapter (HBA), 10 Gigabit Ethernet (GbE) Converged Network Adapter (CNA) or 10 GbE Network Interchange Card (NIC).

With a dual-port Brocade1860 Fabric Adapter, FC and DCB FCoE, TCP/IP and iSCSI I/O protocols can all be running simultaneously on the same adapter card. The Brocade 1860 provides line-rate 16 Gbps FC performance data streaming bandwidth and over 1 million low-latency transaction IOPS per dual-port adapter. Most importantly, the Brocade 1860 can consolidate and dramatically reduce the number of network adapters required in a server while maintaining high-availability hardware redundancy such that cost/space-efficient 1U servers with as few as two PCIe slots can be deployed in reference architecture solutions.

Brocade VDX 6730 Switch

The Brocade VDX® 6730 Data Center Switch is a 10 Gigabit Ethernet (GbE) fixed port switch with LAN and native Fibre Channel ports. It supports multiple connectivity options, including classic ToR server deployments, Ethernet fabrics, and Ethernet storage connectivity for Fibre Channel over Ethernet (FCoE), iSCSI, NAS, and bridging Fibre Channel Storage Area Networks (SANs) and Ethernet fabrics.

GridIron OneAppliance FlashCube

GridIron's FlashCube™ silicon data storage appliances are applications/workload optimized to deliver the best performance out of solid state technology. FlashCubes provide:

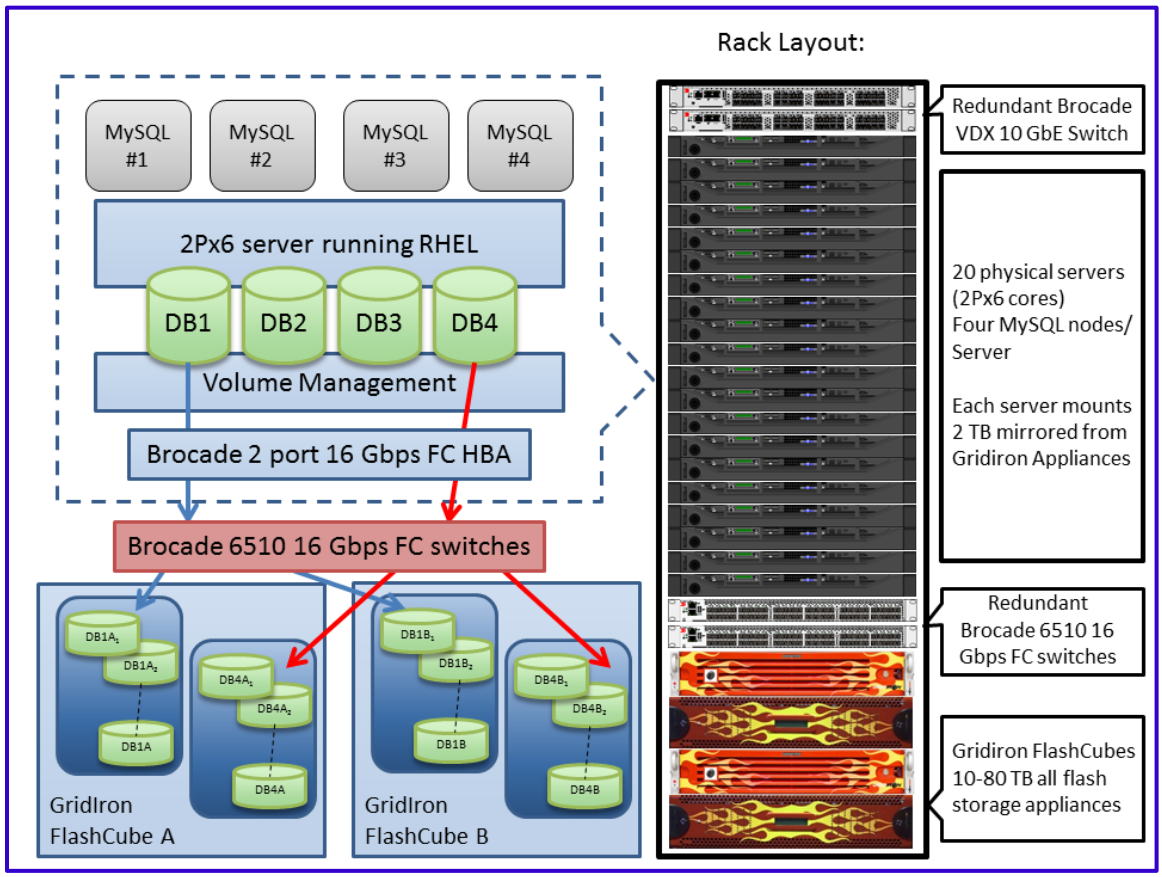
A. Between 10 – 80 TB of ultra-low latency (microseconds) solid state storage per appliance
B. Up to 12 GBps concurrent FC connectivity and over 1 million IOPS
C. The ability to use off-the-shelf form factor MLC SSDs for low acquisition cost

The hardware and software intellectual properties were developed for demanding use-cases such as accelerating Write/Log/ETL and Flash Cache operations for databases and big data applications.

FlashCubes allow you to concentrate large amounts of tier-0 storage into a small amount of rack space, saving on power costs and conserving valuable rack space in the data center, without the failure and maintenance issues of disk-based storage.

The high internal hardware bandwidth (12 GBps full-duplex) of the FlashCube makes it suitable for shared applications – while most PCI-e attached arrays or SSD/PCI-e cards have bandwidth for only ONE server and cannot be shared over a SAN.

A typical rack layout for this new type of server and storage architecture is shown in the diagram below.



Scalability

In this new configuration, scalability is achieved several ways. Because the newer servers have more cores and processing power per server, more database instances can be hosted on the same server, extending the number of databases and number of users that can be serviced per server. The consolidated I/O from these additional servers can overwhelm traditional storage architectures thus preventing the realization of the full potential of server consolidation. The MySQL reference architecture, by virtue of using high-performance all-flash storage appliances for hosting the databases, provides very high concurrent bandwidth and IOPS while at the same time providing very low I/O latencies (sub-millisecond latencies). Additionally, since the storage is not built into the server, the database capacity can grow as needs grow without requiring hardware changes to the servers. The all-flash storage appliances allow the reference architecture to fully utilize the compute capabilities of the database servers to do real work by eliminating wait times experienced with spinning disks.

Total I/O Bandwidth

The high-performance MySQL reference architecture delivers significant improvements in total I/O bandwidth that eliminate bottlenecks typically encountered in conventional architectures. Using the 16 Gbps Brocade 1860 Fabric Adapter, each port provides 1.6 GBps of raw bandwidth. With a typical configuration of two dual-port adapters, each server now has 6.4 GBps of raw bandwidth to the storage, compared to the lower storage bandwidth capabilities found in today's typical deployments. The improved bandwidth allows the GridIron FlashCube appliances to fully

support the higher bandwidth and IOPS requirements of the consolidated servers thus eliminating the I/O bottleneck found in other storage architectures.

High Availability

The high-performance MySQL reference architecture provides for redundant Ethernet networks and redundant Fibre Channel networks to deliver high availability. Each server mounts mirrored 2 TB LUNs from the all-flash storage appliances via two separate, redundant paths.

Replication

Replication of server databases is easily possible within the rack shown above or to another similarly configured rack at another location. The high-speed Ethernet and Fibre Channel networks allow for much faster replication and restoration of data which reduces recovery time.

Total Cost and Cost per Rack

Because the databases are stored in the all-flash storage appliances, server memory can be configured for the needs of the database engine, without having to attempt to store all the data in-memory. This typically lowers the memory requirement for the servers, or allows the servers to run more instances of the databases per physical server.

With one rack unit (1U) servers, 20 servers can be placed within half of a full-height rack, along with all of the networking and dense storage and infrastructure required to support an increased number of users and total database capacity. This new server and storage configuration provides the equivalent of many racks of a more traditional local storage and large-memory server configuration. The cost of one rack of this configuration is far less than the cost of many (10 – 30) racks of the equivalent traditional configuration.

Power Consumption

Several factors contribute to reduced power consumption. Fewer servers are required to support the database users, reducing power consumption. The all-flash storage consumes consider ably less power than traditional storage. For example, a 20 TB FlashCube from GridIron Systems consumes approximately 600 watts while 20 TB of traditional disk storage can consume up to 1500 watts or more, depending on the type of disk drives deployed.

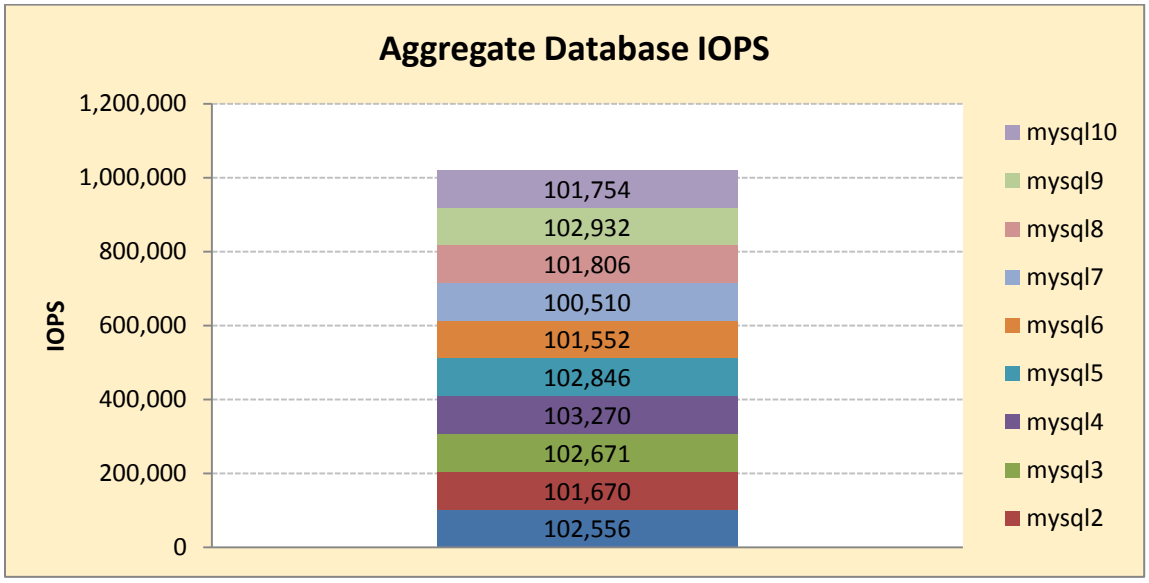
Given that this new one-rack configuration provides the equivalent of 10 – 30 racks of traditional server and storage infrastructure, the electrical power savings are enormous. In addition, the reduced equipment power requirements also translate into reduced cooling requirements.

Architecture Comparison Summary

	High-performance MySQL Architecture	Conventional Architecture
Scalability	Larger	Smaller
Total I/O Bandwidth	Higher	Lower
High Availability	Fully redundant	Partly redundant
Replication	Faster	Slower
Total Cost	Lower	Higher
Power Consumption	Lower	Higher

Performance Results

Demartek audited the test results of a specific configuration of this type of architecture. We ran forty instances of MySQL across 10 physical servers (even though the reference architecture accommodates up to 20 physical servers) concurrently addressing eight mirrored volumes configured in the GridIron FlashCube appliances. We measured the rate at which MySQL transactions were completed in the test period. The charts below describe the transaction rates observed. Databases were 256 GB each.



As can be seen in the chart, the aggregate IOPS from all the database servers exceeded one million IOPS. The GridIron FlashCubes were able to meet the consolidated IOPS and bandwidth requirements of the servers, where each individual physical server was pushing upwards of 100,000 IOPS. While this chart shows the IOPS using 4K block size, larger block sizes were used to saturate the 16 Gbps FC links from a bandwidth standpoint. The GridIron FlashCubes were able to handle mixed-size workloads to accommodate the needs of the database servers.

Conclusion

With the advent of Brocade 16 Gbps Fibre Channel infrastructure and GridIron Solid State Drive (SSD) technology based acceleration solutions, environments with large database server clusters, such as those running MySQL, have a compelling alternative architecture that provides significant improvements in performance, cost, manageability availability and footprint.

Test results for the high-performance reference architecture running MySQL demonstrated significant advantages in performance, including breaking the 1 million transactions-per-second barrier, which is a significant achievement, especially with the reduced amount of equipment needed to achieve this.

The reference architecture from Brocade and GridIron Systems can be used for any data-intensive environment to deliver:

- Extreme performance
- Dramatic footprint consolidation
- Compelling CapEx and OpEx savings
- Fast, easy, and cost-effective scaling

Enterprise IT teams should consider the high-performance MySQL reference architecture when the need to service increasing numbers of users and database queries per cabinet goes beyond the capabilities of traditional deployments while reducing power consumption in the process.

Links for additional information:

- Brocade Communications Systems, Inc.: www.brocade.com
- GridIron Systems, Inc.: www.gridironsystems.com

Appendix – Test Environment

The test environment was similar to the configuration diagram shown above, except that ten physical servers were used with each physical server running four instances of the MySQL database for a total of 40 instances:

- Server #1: MySQL instances 1 – 4
- Server #2: MySQL instances 5 – 8
- …
- …
- Server #10: MySQL instances 37 – 40

Each server:

- 2x Intel Xeon X5675, 3.06 GHz, 12 total cores, 24 logical processors
- 192 GB RAM **
- Linux kernel 2.6.18-238.el5

** Even though each of the physical servers had 192 GB of RAM, each of the MySQL instances was configured to use only 64MB of RAM for caching (the mysql innodb_buffer_pool_size parameter). The cache size was intentionally kept low to drive as much I/O traffic as possible for the workload.

MySQL information:

- Version: MySQL Ver 14.14 Distrib 5.5.15, for Linux (x86_64) using readline 5.1
- Percona build 5.5.15-21.0

The following storage devices were carved out of the GridIron All-Flash storage arrays. All of the devices are used for MySQL. The total flash capacity of the storage arrays was 24 TB.

/dev/sdab /dev/sdai /dev/sdap /dev/sdaw /dev/sdf /dev/sdm /dev/sdt
/dev/sdac /dev/sdaj /dev/sdaq /dev/sdax /dev/sdg /dev/sdn /dev/sdu
/dev/sdad /dev/sdak /dev/sdar /dev/sdh /dev/sdo /dev/sdv /dev/sdz
/dev/sdae /dev/sdal /dev/sdas /dev/sdi /dev/sdp /dev/sdw
/dev/sdaf /dev/sdam /dev/sdat /dev/sdc /dev/sdj /dev/sdq /dev/sdx
/dev/sdag /dev/sdan /dev/sdau /dev/sdd /dev/sdk /dev/sdr /dev/sdy
/dev/sdaa /dev/sdah /dev/sdao /dev/sdav /dev/sde /dev/sdl /dev/sds

---

The original version of this report is located at
http://www.demartek.com/Demartek_Brocade_GridIron_16GFC_MySQL_Evaluation_2012-04.html.