Virident FlashMAX Connect
vHA MySQL Solutions Brief

MySQL instances power many of the web's most demanding applications. Its ease of use and stability make it very simple to implement web applications. However making those applications highly available is not nearly so simple and can often involve serious administrative and performance issues.
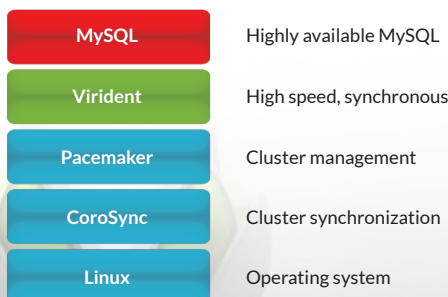
FlashMAX Connect vHA makes rolling out highly available, highly performant, applications as easy as deploying a single instance. By leveraging the exceptional stability and performance of the FlashMAX II device for storage, with a RDMA back-to-back connection, it handles synchronous replication of data sets with minimal effort.

FlashMAX Connect vHA features

- Synchronous replication at microsecond latencies over an RDMA network
- Integration with standard cluster management tools
- Performance indistinguishable from local FlashMAX II device
- Automatic failover for service continuity
- Support for multiple FlashMAX II devices and replicated stores in a single server

Where does vHA fit in a high availability stack?

Virident FlashMAX Connect vHA can integrate with multiple cluster management services to provide a high availability storage subsystem for database applications. For a highly available MySQL instance, it integrates with the widely available, open source Pacemaker and Corosync stack. Since it works directly at Virident's vFAS level, it is able to provide the highest performance replication.

MySQL	Highly available MySQL
Virident	High speed, synchronous
Pacemaker	Cluster management
CoroSync	Cluster synchronization
Linux	Operating system

Why use FlashMAX Connect vHA over DRBD or application replication

While there are other existing alternatives for providing synchronously replicated, highly available storage, none can provide both the performance and simplicity of FlashMAX Connect vHA.

DRBD, the Distributed Replicated Block Device, can be configured to provide synchronous replication of a filesystem. But since it is unaware of the underlying storage device, and has not been optimized for the exceedingly high IOPS and bandwidth of the Virident FlashMAX II, it is simply unable to provide the same performance as vHA, even when coupled with a Virident FlashMAX II

Application managed replication is often an asynchronous process and somewhat complicated to set up and manage. In this configuration, should a server fail there can simply be no guarantee that all transactions would have been replicated to the backup server. It is also can impact application performance due to the overhead of this replication.

## A simple, high performing, highly available MySQL instance using vHA

The cluster shown below consists of a pair of matched servers and operating systems, both equipped with Virident FlashMAX II devices and a single, back-to-back connected pair of QDR Infiniband cards for replication traffic. Note that while Infiniband is used for this cluster, no Infiniband switches or other infrastructure are required, only the single cable linking the two servers. Of course, if so desired, the Infiniband network may include additional switches.

MySQL serves application queries on the 1G Ethernet network using a virtual IP (VIP) that is redirected by the cluster manager to the currently active node of the cluster. MySQL's databases are stored on the vHA FlashMAX II device to provide MySQL with a bottleneck-free IO subsystem. The logs are also stored on the replicated vHA storage. All writes are synchronously replicated between the current master and warm standby with delays on the order of 5 to 10 microseconds (compared to milliseconds on some other replication technologies).

When a server goes down, automatic failover time is on the order of seconds. While the cluster manager used can determine a cluster failure in seconds, the actually startup time of MySQL on the secondary node will depend on the amount of write log to be replayed. Once MySQL is fully operational the client applications merely need to reconnect and continue their operations.
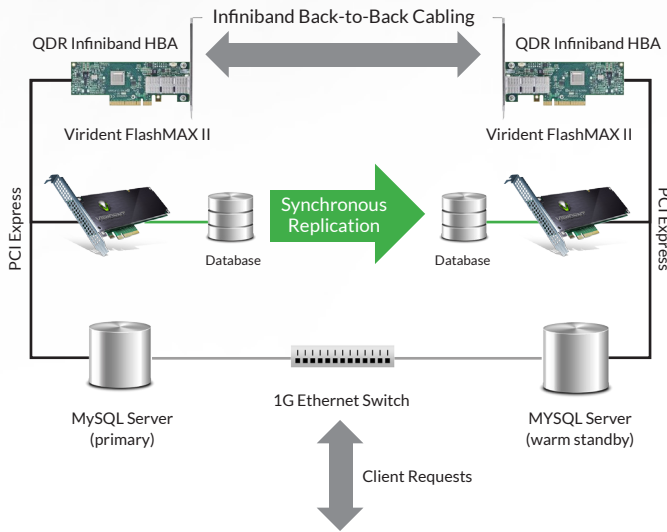
MySQL vHA cluster requirements

The MySQL solution runs on the following hardware and software:

- ▶ CentOS 6.2, 64 bit server edition
- ▶ MySQL 5.x database
- ▶ Pacemaker, Corosync cluster manager
- ▶ 2 x Virident FlashMAX II devices
- ▶ 2 x Mellanox ConnectX-3 QDR/FDR Infiniband adapters
- ▶ Infiniband copper cabling
- ▶ Gigabit switch for incoming client connections

MySQL vHA cluster configuration

A full implementation guide is available from Virident that walks through all the steps required to build a high availability MySQL cluster. Please contact a Virident representative to acquire that document for more details.

Comparison of HA cluster performance using DRBD and vHA

A test was run using a TPCC-like OLTP benchmark on both DRBD and FlashMAX Connect vHA. The same machines were used to run both clusters. All replication was performed over the Infiniband connection between the two nodes. Dataset size was 2500W, and MySQL's buffer pool was set to 65GB. The results are clear, with vHA's tight integration into the FlashMAX vFAS stack it is capable of giving real application performance boosts to MySQL, as it nearly doubled the TPMC-like performance of the DRBD based replication.

Transactions-per-min for TPCC-like workload

vHA TPmC = 18,708

DRBD TPmC = 11,901

vHA DRBD
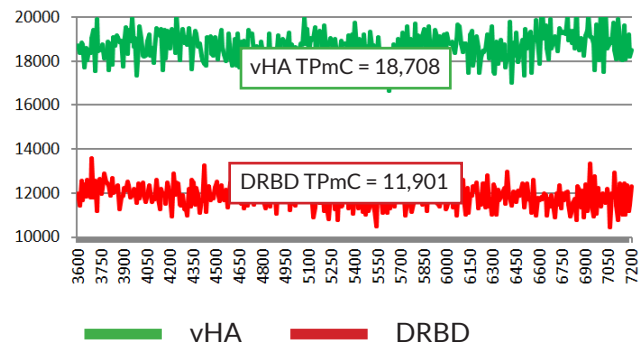
Conclusion

Virident FlashMAX Connect is a simple way to provide applications with the performance of the Virident FlashMAX II in a highly available cluster. It integrates seamlessly with standard cluster managers to provide an easy to use solution. With its novel use of back-to-back Infiniband connections, and integration with Virident's vFAS flash management, it provides synchronous replication nearly twice as fast as the nearest competitor over the same fabric.