VIRIDENT

High Performance Oracle RAC with FlashMAX Connect Solutions Brief

Oracle Database's power and scalability make it very attractive for implementing business critical applications. Its integrated clustering feature, Real Application Clusters (RAC), allows it to provide uncompromising high availability and scale up from a single node to dozens as business needs grow.

This high availability and scalability come at a high cost, however. Oracle RAC clusters require a shared storage infrastructure, so that each node in the cluster can access the same storage as all other nodes. Affordable high performance, high availability clusters were compromised since adding another RAC node entailed not only the purchase, setup, and licensing of another server but a bump up in performance of the shared SAN infrastructure. This SAN infrastructure could dominate the costs associated with this scaling, and the SAN bottleneck could diminish the performance increases.

FlashMAX Connect vShare helps minimize the cost of Oracle RAC clusters while providing the highest available flash performance and reducing the need for expensive SAN storage solutions. It effectively converts the FlashMAX II into a shared storage infrastructure capable of full integration with Oracle RAC. The exceptional performance of a local PCI Express connected FlashMAX II device is now able to power clusters of up to 64 nodes and over 500TB of flash data.

Both Oracle RAC and Oracle RAC One Node can be used in this configuration.

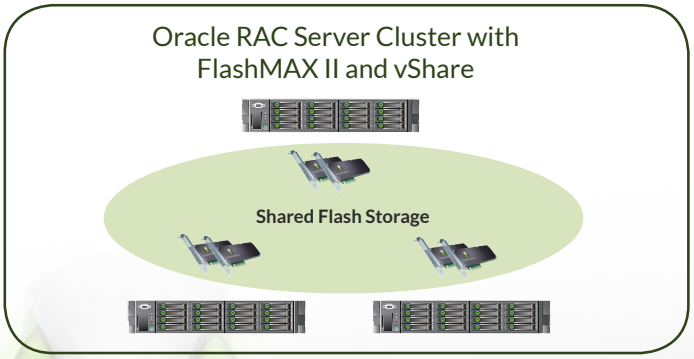

Figure 1: Oracle RAC cluster with FlashMAX Connect vShare architectural configuration

FlashMAX Connect vShare major benefits

- High availability RAC with performance of local attached PCIE flash
- Scale Oracle RAC clusters without having to scale backend SAN
- Ultra-low latency shared storage solution for Oracle RAC
- Simple to set up and manage, PCIE flash appears as shared storage to Oracle RAC

vShare compared to a traditional SAN RAC architecture

For Oracle RAC to function, it requires that all nodes of its cluster be able to access all storage LUNs. This precludes the cluster from using any per-node local storage (either rotating media or flash based) since per-node local storage is not globally visible. Because of this restriction, most implementations of RAC use a shared backend storage SAN, often sized very large to provide the IOPS required of a RAC cluster.

FlashMAX Connect vShare's revolutionary ability to make local flash storage visible to all nodes in a RAC cluster at microsecond latencies (versus millisecond latencies for SAN based storage), changes this architecture completely. Instead of a single expensive and slow SAN providing the RAC cluster with storage, each individual node of the cluster can contribute its own local, hyper-speed FlashMAX II storage to the RAC cluster.

Oracle RAC without a SAN

FlashMAX Connect vShares uses Oracle Automated Storage Manager's (ASM) built-in capabilities to provide seamless high availability to databases. The key to vShare's high availability lies in Oracle ASM's ability to group storage into different failure groups (FGs), and its ability to ensure that data is always replicated on at least two different failure groups. At a high level, a failure group is simply the set of volumes which are expected to have correlated failure events (in the case of vShare, all the cards in each server is identified as a separate failure group). With this information, Oracle ASM ensures that at least two copies of data are present on separate failure groups. This way, should a server go down (i.e. a failure group), there is no impact on data availability.

Oracle RAC Architecture with FlashMAX II and vShare

Below is an example of a 3 node Oracle RAC setup using ASM volumes providing either normal or high redundancy. Three identical nodes each have FlashMAX II devices installed for local flash storage. Each node makes its local FlashMAX II visible to the other nodes via an Infiniband fabric with Remote Direct Memory Access (RDMA) and microsecond latencies. Each node has all three ASM LUNs visible (one local and two remotely shared) thanks to FlashMAX Connect vShare.

When a server goes down, Oracle ASM automatically detects the error and continues service operation. When the failed server is brought back up, ASM will automatically reconnect and rebuild the data on the FlashMAX II. There is no additional management or setup required, once the vShare configuration is completed.

The following diagram explains the configuration of this three node RAC cluster. This configuration can scale from 2 to 64 nodes with the current FlashMAX Connect vShare product. Each node can contain multiple FlashMAX II cards for higher capacities or performance.

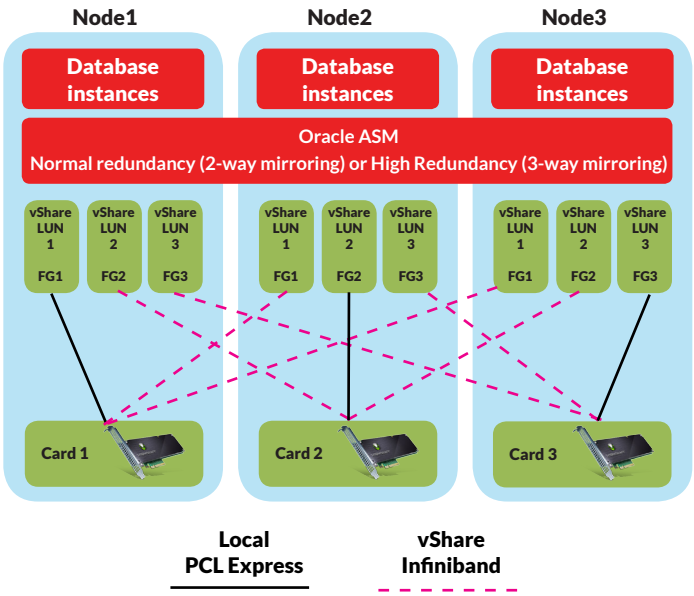

Figure 2: Oracle RAC cluster logical configuration, interconnectivity based

The physical configuration of this cluster is similar. An Infiniband backbone is used to connect multiple FlashMAX II cards using FlashMAX Connect vShare. All Oracle local data access is performed over this high speed, RDMA based network. Multi-pathing capability is embedded in FlashMAX Connect vShare and protects against both Infiniband port or switch failures. Oracle RAC's private network can be configured to use the same Infiniband ports as used by vShare, separate Infiniband ports, or private Ethernet ports.

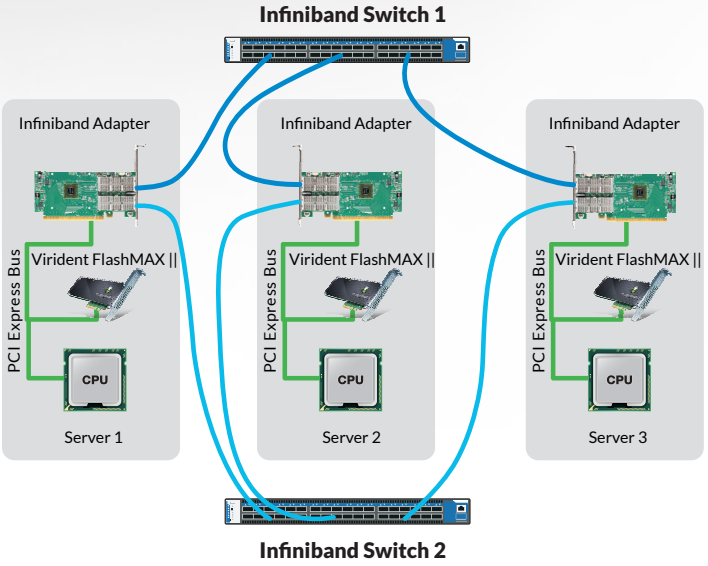

Figure 3: Oracle RAC cluster physical configuration

Performance of a 3-node Oracle RAC on Virident FlashMAX Connect vShare

A three-node Oracle RAC reference configuration was built using Oracle 11g Release 2 using six FlashMAX II 2.2TB devices, two per each node. The cluster was set up according to the architecture specified above, with each separate server's device set up as a separate failure group for high availability. Oracle Calibrate IO was run on a database with a 4KB block size to gauge performance levels.

Calibrate IO Results, 3-node RAC	
MAX_IOPS	1,434,268 IOPS
LATENCY	0
MAX_MBPS	12,310 MB/s

Table 1: Oracle Calibrate IO performance

Conclusion

Virident FlashMAX Connect vShare is a simple way to provide applications with the performance of the Virident FlashMAX II in a highly available Oracle RAC cluster without the need for traditional shared storage. It integrates seamlessly with Oracle ASM and RAC, to provide an easy to use solution. With its novel use of an a Remote Direct Memory Access fabric and integration with Virident's vFAS flash management, it can provide very high performance Oracle RAC solutions with a simple path for scaling up individual clusters or creating new ones.