SystemFabricStor – RDMA Storage Target

Remote Direct Memory Access (RDMA) enabled storage from System Fabric Works provides higher rates, lower latency, and improved energy efficiency

SCALABLE ARCHITECTURES AND THEIR IMPACT ON STORAGE

Today's servers and storage systems consist of many interconnected and interoperating elements – processors, (with caches, memories, I/O bridge chips and channels), storage targets (rotating disks, tapes, and flash memory), and networks (Local, Wide, and Storage Area Networks - LAN, WAN, and SAN). Users and applications retrieve data by moving data between these elements. The speed, latency, and efficiency of the overall storage system are dependent on the architecture, balance, and compatibility between these elements.

The recent trend is that Individual processors have become more and more difficult to accelerate. Thus, increased parallelism of cores and servers is being exploited. Most storage no longer resides in the server and must be accessed via a SAN. System Fabric Works (SFW) is focused on improving the scalability, speed, and efficiency of remote storage to maximize performance - significantly reducing server CPU cycles that would otherwise be consumed by data movement and copies. As the number of systems and the size of storage grow, efficiency becomes increasingly critical in order to avoid diminishing returns. In this environment, the key capability to achieve efficiency is data access that reaches directly from storage to applications – i.e., Remote Direct Memory Access (RDMA).

DEPLOYMENT OF SAN APPLICATIONS USING SYSTEMFABRICSTOR

Traditional SANs use Fibre Channel fabrics and Host Bus Adapters (HBAs). The kernel level drivers and APIs that run over SANs utilize the legacy SCSI protocol to allow the operating system to communicate with "SCSI Targets". In RDMA fabrics, SRP extends traditional SAN functionality to transfer data directly between user applications and the storage device. Terminating the transfer at the application avoids an operating system data copy in each direction of transfer – dramatically cutting latency, increasing performance, and reducing CPU utilization. These benefits are the reason why SFW supplies and supports RDMA in its Storage Appliances.

SFW's SystemFabricStor Appliances leverage the scalability of open source software, RDMA standards, commodity processors, PCI Express, and SAS/SATA/NVRAM/Flash storage technologies. SystemFabricStor supports RDMA over Ethernet and InfiniBand and can be used by systems running Linux, Windows, or Oracle.

OPENFABRICS SOFTWARE AND APPLICATIONS INTERFACES FOR SYSTEMFABRICSTOR

SFW distributes and supports the OpenFabrics Enterprise Distribution (OFED) on RedHat and SUSE. The OFED stack contains the drivers, the operating systems' interfaces and the upper level protocols that enable Linux to be RDMA storage initiators. This software either allows legacy applications to reduce CPU usage and latency for the same level of work or it allows improving performance without adding compute resources. The latest OFED releases support Lustre protocols for clustered storage and the RDS protocols for Oracle Databases.

Support for RDMA is already built into Microsoft Windows Server 2012 (via SMB Direct).

IMPROVING APPLICATION, FILE SYSTEM, AND DATABASE PERFORMANCE WITH SYSTEMFABRICSTOR

RDMA interconnects are well documented for providing the most efficient, highest performance, and lowest latency file systems (e.g. Lustre). Databases (e.g. Oracle RAC 11g) and virtualization can be significantly accelerated when RDMA is enabled. Benchmarks show that RDMA can add significant performance to common application workloads.

- Oracle testing of their RDS protocol showed that RDMA enablement dropped CPU utilization during data ingest benchmarks more than 50% while performance improved ~5% over a non-RDMA enabled run (using identical network interfaces) [1]
- VMware testing of vMotion showed that RDMA was 36% faster while CPU utilization dropped 92% on the source and 84% on the destination server [2]
- InfiniBand technology provides interconnects for 226 of the Top500 Computer System, making it the most-used internal HPC system interconnect technology [3]

---

[1] Oracle's Next-Generation Interconnect Protocol: Reliable Datagram Sockets (RDS) and InfiniBand, presented by Paul Tsien, Oracle
[2] RDMA on vSphere: Update and Future Directions, Bhavesh Davda & Josh Simons, 3/26/2012
[3] November 2012 Top 500 list at http://www.top500.org/lists/2012/11

SYSTEMFABRICSTOR HARDWARE

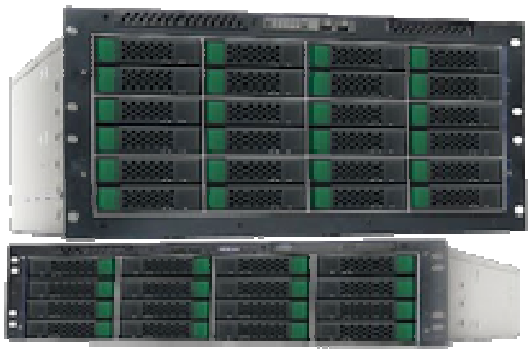
Standard hardware configuration for SystemFabricStor:

- x86 CPU (4-8 cores)
- PCI-E v3.0 High Performance RAID Controllers (1-4)
- 2+1 Hot Swap Power Supplies
- 3u rack mount (16 drive configuration-SSA16), ~70 lbs.
- 5u rack mount (24 drive configuration-SSA24), ~150 lbs.

SYSTEMFABRICSTOR CONFIGURATION OPTIONS

System Fabric Works sells and supports a range of SystemFabricStor appliances.

- Rotating disk - up to 24 SAS or SATA drives.
- NVRAM and Flash available for higher IOPS.
- Infiniband at speeds to 56 Gbps
- Ethernet to 10 Gb/s
- SFW allows selection of RAID level (0, 1, 5, 6 or JBOD) to match customer preferences for performance/data security.



SYSTEMFABRICSTOR PROTOCOL SUPPORT

Supported SystemFabricStor protocols:

- SCSI Remote Protocol (SRP) for RDMA –enables SCSI over Infiniband
- iSCSI Extensions for RDMA (iSER) – enables RDMA over Ethernet and Infiniband
- RDMA over Converged Ethernet (RoCE) – enables RDMA over Ethernet
- Internet Wide Area RDMA Protocol (iWARP) – enables RDMA over TCP
- NFS and CIFS can be supported by adding an external server that mounts SystemFabricStor as block storage along with software such as Nexenta or Chelsio's USS to provide the file system and NAS interface
- SystemFabricStor can act as an OST or MDT in Lustre implementations

CONTACT INFORMATION FOR SYSTEM FABRIC WORKS

Web: http://www.systemfabricworks.com
Sales: sales@systemfabricworks.com
CEO: Kevin Moran (512) 563-6674
kmoran@systemfabricworks.com

SYSTEMFABRICSTOR PEFORMANCE

Performance depends on the configuration selected by the customer. SystemFabricStor can be configured for high IOPS, high capacity, and/or high throughput:

- For high-IOPS applications, 15K RPM SAS drives are the starting point.
- SSDs and NVRAM drives or PCI-E based flash memory can be added for even higher IOPS performance
- For high-throughput applications, SAS drives provide up to 3.6 GB/sec of sustained performance.
- For high capacity applications, large SAS drives (to 4 TB) are available.
- Large capacity at a lower price point is available from SATA drives

SYSTEMFABRICSTOR CUSTOMERS

System Fabrics RDMA-enabled storage was developed for a proprietary customer in 2006. Other customers include Sandia National Labs, Patria, CSPI, and Padova Technologies.

CERTIFICATIONS AND ENVIRONMENTAL

ROHS	ROHS-5
CE	Completed
UL	Pending
CSA	Pending

Min temp.	10ºC/50ºF
Max temp.	35ºC/95ºF
Humidity	8%-90% (non condensing)

COMPLETE SOLUTIONS AVAILABLE FROM SYSTEM FABRIC WORKS

At System Fabric Works, RDMA is our DNA! We can assist you in architecting an end-to-end RDMA solution for your applications and infrastructure. In addition to our own products, we are a value added reseller of all of the major hardware and software vendors that provide components of a complete RDMA solution.