Using the Tera-RamSan for High Bandwidth SAN File Sharing

The mechanical nature of disk drives becomes a weakness when multiple servers are accessing the same content. One server reading/writing data from a RAID results in sequential data access patterns. Multiple servers reading/writing data from the RAID results in random data access patterns. As the first user starts accessing data the disk capably spins and sequentially moves data to the server. When the second user starts accessing data from the same set of disks the read head becomes much less efficient. It must now stop reading from one part of the disk and begin reading from another part of the disk. As the two servers continue working, bandwidth for both servers drops. It would be bad enough if the drop in bandwidth was simply linear. Unfortunately, the drop in available bandwidth is more exponential. Compounding the situation is that as bandwidth decreases latency for data accesses increases. What started out as 5 millisecond response times for data accesses from the server become 20 millisecond and 100 millisecond delays as loads increase. To the users these delays are time-consuming and frustrating.

Defining the Solution

The ultimate storage solution would have four characteristics:

- Fast - capable of supporting many gigabytes per second of bandwidth;
- High capacity – capable of expanding to support several terabytes of storage;
- Shareable – content can be shared across multiple users without performance degradation
- Inexpensive – the solution must provide a return on investment that exceeds costs.

As is almost always the case, solutions in the market diverge from the requirements. Network attached storage would seem to be a great solution, except it is the slowest of the storage alternatives because they are slowed not only by the hard-disks that they use to actually store the data but they are also slowed by a very burdensome TCP/IP protocol. Fibre Channel attached RAID systems offer a higher level of performance, capacity and cost but still fail to meet the performance requirements of the most demanding shared environments.

Introducing the Tera-RamSan

Fortunately, Texas Memory Systems has a solution, the Tera-RamSan. The Tera-RamSan picks up where disk-based systems leave off:

- The Tera-RamSan is fast. In fact, it is the World's Fastest Storage®.
- The Tera-RamSan provides high capacities. The system can be deployed in multi-terabyte configurations.
- The Tera-RamSan, with the help of SAN file sharing software, is sharable.

The Tera-RamSan ***uses RAM*** (random access memory) as the storage media instead of hard disk drives. The natural benefit of RAM is that it is designed to handle random data accesses without slowing down. RAM is a solid state storage technology which means that there are no moving parts in the data path. Data is moved from the Tera-RamSan to the server in 20 microseconds. 20 microseconds is 250 times faster than the best case performance for disk based systems (5 milliseconds).

The speed of the Tera-RamSan storage media is complimented by high bandwidth connections to the editing servers. The Tera-RamSan ***uses Fibre Channel*** interfaces. The terabyte Tera-RamSan configuration supports a minimum of eight dual ported 2Gbit Fibre Channel controllers (or 16 ports). Each port supports close to 200MB/second of sustained random bandwidth meaning the entry-level terabyte configuration provides 3.2 Gbytes per second of bandwidth. The terabyte Tera-RamSan is scalable to 64 2Gbit ports which could provide as much as 12 Gbytes per second of bandwidth. A solution with this bandwidth is only available from Texas Memory Systems.

Technical Addendum

Tera-RamSan building blocks.
A terabyte configured Tera-RamSan is composed of eight independent (not interlinked) RamSan-325 solid state disks. The RamSan-325 has the following key characteristics:

- Non-volatile. Three batteries are used to keep the system powered in the event of external power failure. The system automatically uses this time to backup from memory to four RAID-3 protected disk drives.
- High capacity. Each unit provides up to 128GBytes of capacity.
- High bandwidth. Each unit includes from two to eight 2Gbit Fibre Channel ports for connectivity to the storage network or to servers.
- Easily managed. Each system has a web graphical user interface (GUI) for configuration and monitoring available through Ethernet. The systems can present from one to sixty-four LUNs. Each LUN can be mapped to one-to-many Fibre Channel ports.

Other Required Components
Because the Tera-RamSan involves a minimum of eight Fibre Channel links and each unit presents at least one LUN to the outside world it is desirable to plan ahead for implementation options. The minimum implementation is likely to require the following additional components:

- Fibre Channel switch(es). A Fibre Channel switch centralizes the connections between the Tera-RamSan and servers allowing any particular server access to the entire capacity of the Tera-RamSan without having every server directly connect to the array.
- Ethernet switch. As each unit has an Ethernet interface, it is desirable to have an Ethernet switch aggregate the Ethernet links out.
- Fibre Channel cables. Fibre optic cabling is required to connect the RamSan to Fibre Channel switches and to connect the switches to the hosts.
- SAN File Sharing software (discussed in detail below).

Putting it All Together
Figure 1, below, shows the physical connectivity between servers, a Fibre Channel switch and two terabytes of RamSan units. The following types of connections are shown:

- The servers are connected to the Brocade Fibre Channel switch. In this case, a 32-port Fibre Channel switch is shown. At a minimum, each server should have one connection to the switch. It is also possible to have multiple HBAs in each server for redundancy. In some environments it is possible to use multiple HBAs for load-balancing.
- Each RamSan-325 unit is connected to the Brocade Fibre Channel switch. At a minimum, each RamSan-325 unit needs to have a connection into the switch. It is also possible to use multiple controllers in the RamSan for redundancy or to increment available bandwidth to that segment of the data.

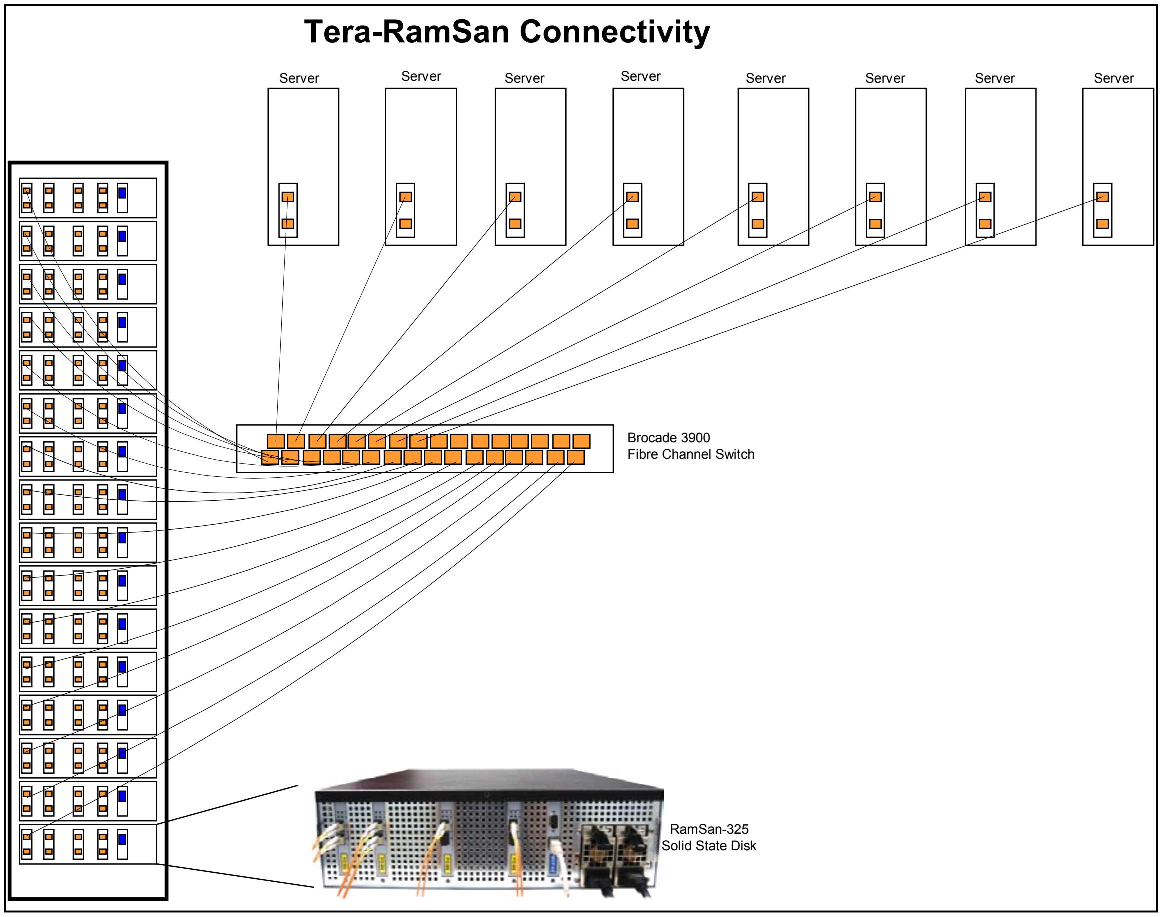

Figure 1: Tera-RamSan connectivity.

The next step after physically connecting all of the devices is to determine the best way to use the Tera-RamSan across the servers. As Figure 2 indicates, another key component in the equation is implementing a SAN shared file system, such as Red Hat's GFS (other SAN shared file systems are available from ADIC, Sistina, SGI and Tivoli):

> "Red Hat GFS allows Red Hat Enterprise Linux servers to simultaneously read and write to a single shared file system on the SAN, achieving high performance and reducing the complexity and overhead of managing redundant data copies. Red Hat GFS has no single point of failure, is incrementally scalable from one to hundreds of Red Hat Enterprise Linux servers, and works with all standard Linux applications."
> (from http://www.redhat.com/software/rha/gfs/).

GFS is installed on each server that has access to the SAN. GFS, or other SAN shared file systems, serve several crucial roles:

- GFS allows each of the servers to share access to each of the RamSan units by locking data that is being accessed by another server.
- GFS allows each of the servers to share access to each of the RamSan units by managing cache coherency, GFS cache management keeps a server from reading data from its own cache, which it believes to be the last update, when in fact the data has been updated on the central storage system by another server (unbeknownst to the reading server).
- GFS allows the eight unique LUNs presented by the RamSan to be flexibly presented to the servers by implementing logical volume management tools. At one extreme, this could mean that the entire terabyte of SSD capacity looks like a single volume to the servers.

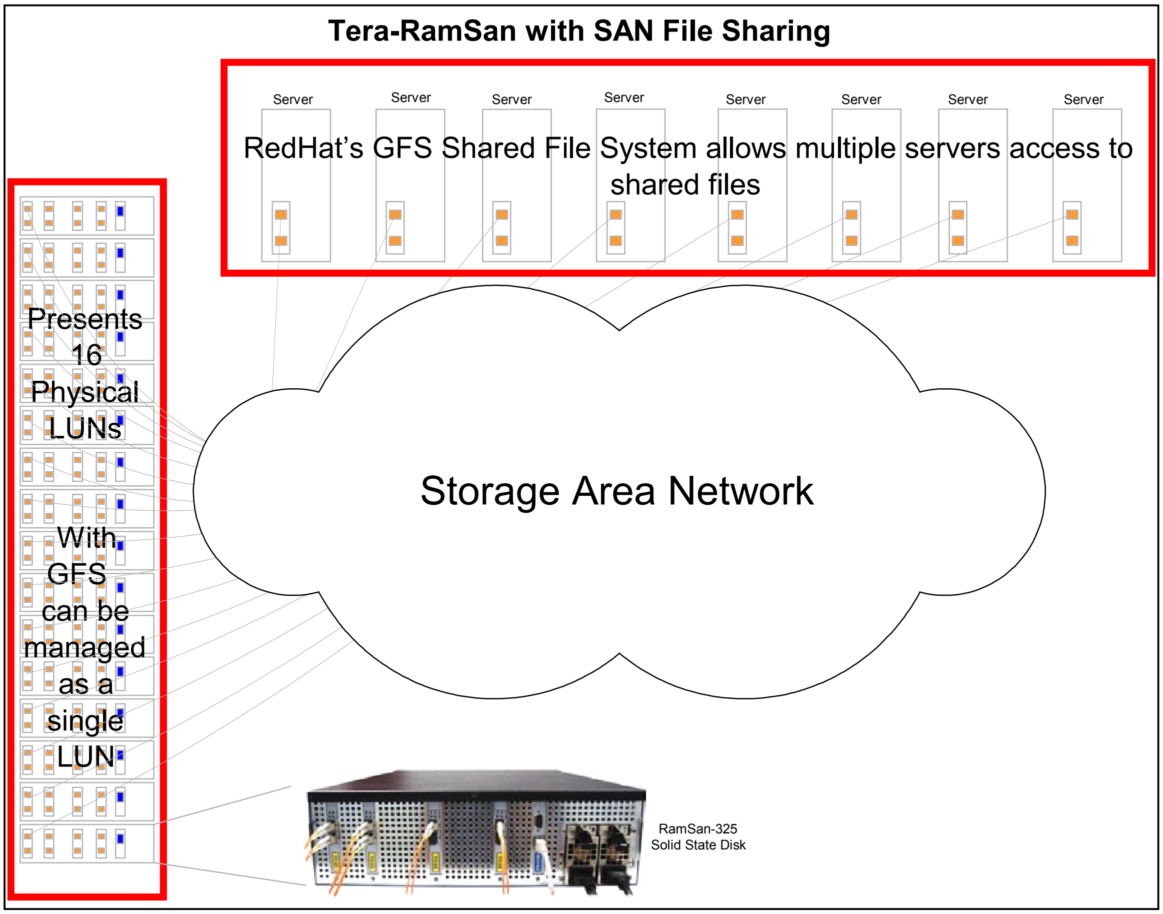

Figure 2: Tera-RamSan SAN File Sharing