SEARCH FOR A RELIABLE STORAGE ARCHITECTURE FOR RHIC*

S.Binello#, R.A.Katz, J.T.Morris, BNL, Upton, New York, USA

Abstract

Software used to operate the Relativistic Heavy Ion Collider (RHIC) resides on one operational RAID storage system. This storage system is also used to store data that reflects the status and recent history of accelerator operations. Failure of this system interrupts the operation of the accelerator as backup systems are brought online. In order to increase the reliability of this critical control system component, the storage system architecture has been upgraded to use Storage Area Network (SAN) technology and to introduce redundant components and redundant storage paths. This paper describes the evolution of the storage system, the contributions to reliability that each additional feature has provided, further improvements that are being considered, and real-life experience with the current system.

BACKGROUND

The RHIC operational RAID system contains executable versions of software used to operate the accelerator. It is also used to store accelerator equipment settings and accelerator performance data. In 2004, a project was initiated to upgrade the RHIC storage system in order to better meet the ever- increasing demands for space, performance and reliability.

The previous RAID system was an nStor 8LE that was directly attached to a SUN host via a SCSI bus. It had 320GB of storage with few redundancies.

Requirements for the new system were as follows:

1. It should be able to store all binary executables needed to operate RHIC, as well as all data produced over one year of RHIC operations, with some additional space to account for the growth in data rates over the next few years. Total storage requirements were estimated at about 5TB.
2. It should be highly reliable, with as little down time as possible.
3. It should be able to support data rates expected during the operation of RHIC. The upper limits on the previous system were 20MB for writes, and 40MB for reads. A safe estimate was to require at least double those rates.
4. It should cost under $100K

The second of these four requirements was the most difficult to measure. A high level of redundancy in system components, however, was considered likely to enhance system reliability since failures of single components could be tolerated. It was felt that redundancy paired with automatic failover should provide a satisfactory level of reliability.

*Work performed under the auspices of the U.S. Department of Energy
sev@bnl.gov

SYSTEM SELECTION

Architecture

Before purchasing a new RAID system, we investigated three storage architectures that were prominent in 2004:

1. Directly Attached Storage (DA) - where storage is directly attached to a small number of general purpose server systems (usually one or two) via a local bus such as SCSI or Fibre Channel.

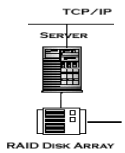

Figure 1: Diagram of typical DA configuration.

2. Network Attached Storage (NAS) - where storage is made available on a LAN via a dedicated file server appliance (NAS head).

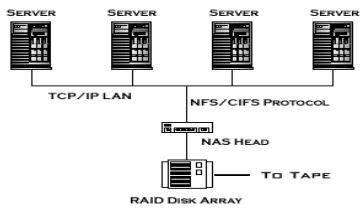

Figure 2: Diagram of typical NAS System.

3. Storage Area Network (SAN)[3] – where storage is located on a separate dedicated network. Every server is able to access every storage device.

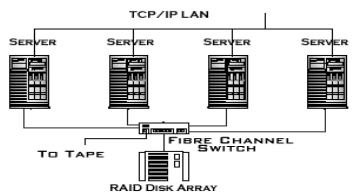

Figure 3: Diagram of typical SAN System.

Previous experience had been limited to storage directly attached to one or two hosts. The disadvantage with this approach was that it was not easy to replace hosts in the event of failure or add hosts when additional processing power was needed.

We decided to implement a small SAN. The SAN solution was somewhat more complicated than a NAS solution, but provided added flexibility that we considered important. In a SAN, servers may be added as needed to increase performance or provide an additional level of redundancy at the host level. We were also

concerned that the performance of a NAS box would be limited by the performance of the TCP/IP network. The SAN approach gave us the option of running some applications on servers that had direct access to storage, eliminating the need to go over the TCP/IP network.

Storage Components

We purchased the RAID system from nStor (now Xyratex). A similar, smaller, directly attached system was already in place at RHIC to support our software development efforts. The new RAID system had the following features:

- Redundant components: The system has dual power supplies, fans, and cards.
- Fibre drives: Fibre drives are highly reliable. They have an MTBF of above 1M hours under heavy usage[1] as opposed to 500K hours to under 1M hours for ATA/SATA, and less than 500K hours for the IDE drives available at the time. Indicative of the reliability of Fibre drives is the 5 year warranty period, as opposed to 1 to 3 years for ATA drives. Even today with the wide availability RAID systems with SATA drives, Fibre drives are still the preferred choice for critical systems.
- Dual Active/Active controllers: Dual RAID controllers with automatic failover provide insurance in the event of a single controller failure. The additional active/active capability provides increased performance. Both controllers are used to service requests. Just one controller is used in an active/spare configuration.
- Dual-loop: Each drive can be accessed by one of two independent paths. If one of the loops fails, the drive can still be accessed from the other loop.

Storage Configuration

The storage was configured in the following manner:

- RAID 5: To handle the eventuality of a drive failure the RAID system was configured to implement RAID 5[2]. This provides a cost effective solution, in that it does not require twice as many disks as in RAID 1 (Mirroring), where all data is duplicated on a separate disk. RAID 5 arrays only require the equivalent of one additional drive to contain the parity data that is necessary to rebuild data from a failed drive.
- Spare disk: We reserved one disk as a spare to be used to rebuild data from RAID 5 parity data in the event of a drive failure.
- RAID arrays: In order to reduce the risks of multiple drive failures, we divided the storage into three RAID arrays. Since a second drive failure within a RAID 5 array in the process of rebuilding itself would be catastrophic, dividing the storage into multiple RAID arrays decreases the likelihood of a second failure within the same array. However, each additional RAID array does consume the equivalent of one disk drive for parity data.

Network Configuration

In addition to ensuring redundancies in the storage device, it was also essential that redundancy exist in the network connecting the hosts to the data storage device.
All the network components were purchased from Qlogic.

- Switches: SAN switches allow multiple hosts to connect to a storage device. Two switches were used to connect the storage to servers to create a small SAN. If one switch fails, all traffic is automatically re-routed through the surviving switch.
- Host Bus Adapters (HBA's): HBA's are used to connect hosts to the switches. In order to create redundant paths to storage, each host had two HBA's. Each HBA on a host was connected to a different switch. Load balancing was enabled on each host so that both links are utilized in the transfer of data.
- Host Failover driver: A break in communication in a path leading from a server to storage is detected by the HBA driver on the host. This driver is responsible for automatically failing over all communication from a failed path to the surviving path.

Storage Software

Rapid notification of failures is important in a storage system with automatic failover. The failover to redundant components provides an immediate solution to a failure, but intervention is necessary to re-establish full redundancy in the system. It is imperative that system administrators obtain prompt notification of a failure so that corrective action can be taken.

Vendors for both the storage system (nStor) and the network components (Qlogic) provide software to configure and monitor their respective systems. Each package also supports e-mail notification in the event of a failure. The nStor software is Web based and detects the failure of any storage components. The Qlogic software resides on the hosts and detects a failure with any of the links to the RAID system.

EXPERIENCE

The operational storage system was trouble free for the first year. An identical directly attached system purchased earlier for our development efforts had also run well for two years.

The second year proved problematic. The main issue encountered was that one of the controllers would hang but would not be failed over for a time period that varied from minutes to hours. This disrupted operations and also led to corruption of the file system on four occasions. After some investigation we discovered that the EXT3[5] file system on our Linux servers was susceptible to corruption if communication with storage was totally lost. Under these conditions it was not enough just to restore the RAID system. It was also necessary to reconstruct the file system.

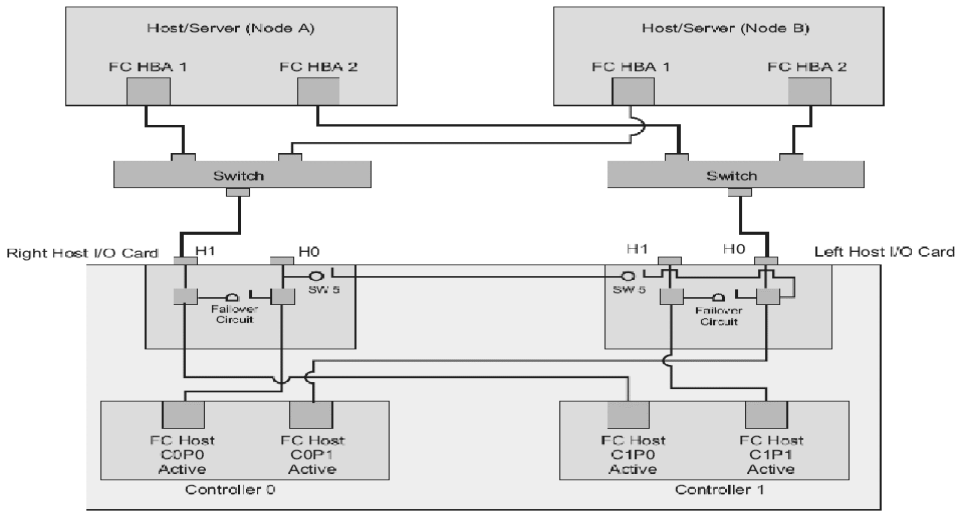

Figure 4: Diagram showing system with redundant paths.

We were unable to intentionally recreate this problem for diagnostic purposes. Eventually, all components were replaced. This included a chassis replacement and an upgrade of the controllers to a later model. We have not experienced the problem since. Our suspicion is that a firmware upgrade in 2005 may have been responsible. This was reinforced by the fact that our development system, which had run problem free for the previous two years, began exhibiting the same problem at around the same time. The firmware in both systems has since been upgraded again. Unfortunately, we have experienced two instances of a different type of failure in which faulty drives are not failed over correctly.

A hot backup system had been in place at RHIC for many years. All critical executables, data and scripts needed to operate RHIC are backed up daily to this system. As a result of these failures, we placed increased emphasis on developing written procedures and establishing annual drills to switch the entire operational system over to the hot backup system.

PLANS

In an attempt to increase reliability, we are currently planning to replace the nStor/Xyratex RAID system, while maintaining the current SAN architecture. One system under consideration is an EMC system sold by DELL (CX-3).

Two additional capabilities that could prove useful are:

1. Clustering: to automatically failover a host in case of a failure, or simply for host maintenance or upgrades.
2. Mirroring: switching to a completely separate hot backup system is a time consuming (approximately 4 hours if all goes well) manual process, it would be useful to have a secondary storage device on the SAN dedicated to mirroring data from the primary storage device. An automated process could be put in place to switch to the mirror in the event of a serious RAID failure. Note that this capability would probably not prove useful if corrupted file systems were simply duplicated to the mirror.

CONCLUSION

As our current plans imply, we have been satisfied with many of our underlying choices. Aside from the problems with the RAID controllers, we experienced few problems with the system. The Fibre drives have proved reliable, only three failures out of 36 drives over three years. The redundant links have also proved useful. We experienced only one failure, and that was handled automatically and seamlessly. We have not had any need to utilize the flexibility provided by the SAN network, but it is still reassuring to know that it exists.

REFERENCES

[1] Greg Schulz, “Balancing act – Match your date to the correct disk”, searchstorage.com February 2004,
[2] William George, “RAID explained”, White Paper Puget Systems, April 2006
[3] “What is a SAN”, OSSI. Jan 2000
[4] “SAN Fundamentals” SUN April 2007
[5] Stephen Tweedie, EXT3, Journaling Filesystem, July, 2000 Ottawa Linux Symposium