Title	An Integrated High Availability Computing Platform
Author	Yan Han Systems Librarian University of Arizona Library 1510 E. University Blvd. Tucson, AZ, 85721-0055 hany@u.library.arizona.edu Tel.: (520)626-2936 Fax: (520)621-8276
Created Date	2004-10-01
Updated Date	2004-01-05

An Integrated High Availability Computing Platform

1. Introduction

As users raise high expectations to access electronic contents and online services at anytime and anywhere, libraries are putting a lot of effort to provide electronic access to a variety of services and contents. A typical academic library provides information access to its own collections (e.g. catalog, archival collections) and external content (e.g. licensed databases, electronic journals). It also provides a wide range of services such as inter-library loan, document delivery, electronic reserve and virtual reference. In addition, it has to manage its internal operations such as human resources, financial department to facilitate its services to customers. The external services require multiple computing platforms to meet various needs of multiple commercial and open source systems, while the internal operations typically need a Microsoft Office environment including an email server and a file server, resulting in a wide range of computer hardware and software. Therefore, these diverse computing needs cause a complex computing environment, where a wide range of hardware, operating systems (OS), and applications co-exist.

The complexity results in many frustrated issues regarding all aspects of hardware and software environment management such as storage, backup, recovery, availability and security. Questions are often raised, including: what are some of strategies to reduce server downtimes due to any factors? How much storage should these servers have, respectively? What are backup schedules for different servers and how much should each server be backed up? How do we keep up with servers' hardware and software patches to enhance functionality and/or minimize security risks, while maintaining stability of systems at the same time? These routine tasks become problematic and time-consuming when multiple computing platforms are required to be managed.

Due to the complex external services and internal operations provided by the Library as well as its previous organizational structure, the Digital Library and Information Systems Team (DLIST) at the University of Arizona Library is managing more than 30 servers, which consist of a variety of hardware and software. The library's main web server and the ILS run on Sun's Solaris machines with Solaris 8. The inter-library loan (ILL) server, the mail server, the file server, the DNS server and the LDAP server are on Dell servers in Windows 2000/2003. Digital library systems such as contentDM are running on non-brand servers in Linux Red Hat OS. There are also multiple relational databases (e.g. MySQL, Oracle, PostgreSQL and mSQL) running on different servers and different OSs. To maintain this complex computing infrastructure, the team has three system administrators and one database administrator. Each system administrator manages servers in a specific OS environment (Windows, Linux, and Solaris), while the database administrator takes care of all database issues. Coordination on systems administration is necessary when multiple administrators are involved. The complexity and the difference among multiple servers, OSs, and databases result in extra coordination, increased turnover time, and

inefficient staff overhead.

2. Systems Analysis of Different Computing Needs

2.1 Problems and Requirements

Before taking any steps to solve the above problems, we have to fully understand and analyze business requirements, and design a system that fulfills our business requirements. A systems analysis approach was adapted to serve this critical purpose, which generally consists of preliminary investigation, problem identification, requirements analysis, decision analysis, system implementation and operation / support. (Whitten, Bentley & Dittman 2001) As the library decided to redesign its website to improve users' access to content and services, a new platform was planned to be purchased in order to fulfill the new computing needs. The new website consists of static and dynamic web pages. The static web pages are administered by Macromedia Contribute and the dynamic web pages such as subject guides will be handled by a Java/PHP/Database program. At the same time, to support the library's commitment to digital library and scholarly communication, another computing platform was planned to host a content management system (CMS) as well as storing archival files. One purpose of the CMS is to deliver journals, including the library's own journals (e.g. The Journal of Insect Science) and other organizations' journals (e.g. Rangelands). Due to the nature of the two systems, they have different computing needs regarding audience, systems requirements, and administration.

Since the website is the most important gateway to every information resources and services that the library provides, the platform must be stable and provide a true 24x7 High Availability (HA). The IT Industry usually measures a system's availability by its uptime. The table below provides an example of a range of availability. However, there is no official standard to define what a HA system is and what an accepted rate is. Different organizations have their own accepted rates. For example, EBay achieved 99.94% (June 2004) uptime, a 4% increase on site availability from 95.2% (June 1999) (eWeek, 2004).

Table I: Measuring Availability

Uptime	Downtime/Year	Downtime / week
98.00%	7.3 days	202 minutes
99.00%	3.65 days	101 minutes
99.50%	43.8 hours	50 minutes
99.80%	17.52 hours	20 minutes
99.90%	8.76 hours	10 minutes
99.99%	52.6 minutes	1 minute

According to statistics gathered, both software failure and hardware failure account for 50% of

unplanned downtime. Natural disasters, network failure and human errors are the other causes (Marcus, Stern 2003). The purpose of HA solutions is to minimize software and hardware failures, while other causes should be addressed by either training (for human errors), or unpreventable (for disasters), or higher level issues (for network failure).

The following are detailed system requirements for the Library's web server:

1. The system must be stable to allow hundreds users simultaneous access.
2. The system must provide a true 24x7 HA, which minimize downtime due to any factors (e.g. hardware failure such as hardware failure, software failure).
3. The system must implement two backup schedules: incremental backup and full backup frequently.
4. The system requires at most 100GB storage for web pages.
5. The system must have redundancy such as Raid 5 to maintain data integrity.
6. The relational database, Oracle, must be in a stable and HA server with enough storage so that dynamic web pages can allow dozens of users' simultaneous access.
7. The system should be cost-effective with the possibility to expand.

The CMS system requirements are different from the web server system due to its nature. In essence, there are no real needs for HA and frequent backup, but it does require a lot of storage space for storing general PDF and big archival files. It is a primary archival server that will store all files related to the journals, including all files in the following formats: uncompressed TIFF, high quality and compressed PDF. The system requirements for the CMS system are:

1. The system must be stable to allow 10 users simultaneous access.
2. The system must have big storage space, currently at least 500 GB to store related files.
3. The relational database must be in a stable server with enough storage to host data from the CMS.
4. The system should have the potential to grow if more storage space needed.
5. The system must have redundancy such as Raid 5 to maintain data integrity.
6. It is desirable that the system provide 24x7 HA.
7. The system does not require frequent backup, because the contents will not be changed.
8. The system should be cost-effective.

Though their purposes and needs are different, they do have something in common. It is necessary to have both HA and data integrity for the two systems. In addition, both systems require frequent interactions with relational databases, which should be in a stable and HA environment. Multiple solutions were identified to fulfill these requirements to build an integrated HA system in order to fulfill different computing needs, including Linux clustering solutions and/or Windows Server 2003 Clustering server for achieving HA, and the Storage Area Networks (SAN) for accomplishing centralized data management.

2.2 Solutions for HA

There are two major Linux clustering solutions: commercial Linux such as Red Hat Enterprise Linux Advanced Server (AS) Cluster Suite (Red Hat, 2003) and other Linux products such as Debian and SuSE, associated with Linux-HA project < <http://www.linux-ha.org/> > and Linux Virtual Server (LVS) project < <http://www.linuxvirtualserver.org/> >. The Red Hat AS Cluster Suite consists of two features: the HA clustering, providing continued application operation in case of server failure, and IP load balancing providing network load balancing. For other Linux products, the Linux-HA project is very similar to the Red Hat's HA clustering, while the Linux Virtual Server provides load balancing. Currently these two projects are working together and are expected to collaborate in the future. The Linux HA project created the "heartbeat" package, which monitors hosts, informs the cluster when one of them failed, and then performs automatically failover. A sub-package "ldirectord" provides the integration with the LVS (Roberson, 2004).

While the Linux clustering solutions are getting mature, lots of sites are currently running this solution. Google made its own shortened version of Red Hat distribution and wrote its own code for load balancing (Harnur, 2000). It used 15,000 PCs to build the world largest Linux cluster and achieved nearly 100% uptime for processing over 150 million queries per day with <0.25 second response time (Holzle, 2002) (Barroso, Dean & Holzle 2003). The library has already used the Linux HA solution to support the "Tree of Life" project < <http://tolweb.org/> > and achieved near-100% uptime. The following diagrams illustrates the general concept behind the HA active-passive solution.

- (1) Figure 1: Node A and B are required to install the Linux HA "heartbeat" package with a crossover cable linking together. During the normal operation, Node A always processes any incoming requests. At the same time, Node B is standby and constantly sends signals to A to see if Node A is alive.
- (2) Figure 2: In case of Node A failure, the "heartbeat" package detects the failure and let Node B take over the IP. At this point, Node B is now responsible for handling incoming requests. In addition to the failover, there is also a need to synchronize files of Node A and Node B for a web server because of frequent changes of web contents. A module "rsync" continuously runs on the background to serve this purpose.

Figure 1: Linux HA Cluster Normal Operation

Figure 1: Linux HA Cluster Normal Operation

Figure 2: Linux HA Cluster Failover Operation

Figure 2: Linux HA Cluster Failover Operation

The Windows clustering technologies, consisting of server clusters and network load balancing, are Microsoft's counterpart to the Linux clustering technologies. Its concept is very similar to the Linux HA solutions. The Windows clustering technologies, a part of Windows Server 2003, provide better integration with the Windows environment.

2.3 Solutions for Data Management

Because our data are distributed around multiple servers, there is no centralized storage place to store, backup and recover. A Storage Area Networks (SAN) solution was identified as a solution for centralizing data management so that we can provide availability, scalability and manageability to various data. The SAN is a back-end network connecting servers to storage devices and is implemented for environments that demand high data availability (Microsoft, 2002) (Freedman, 1999). It is a new technology often associated with Fiber Channel that appeared in the market in the late 1990s. Most SAN products can deliver data very fast (up to 2Gbps) and connect many

devices (up to 256). A SAN product is designed to eliminate a single point of failure from its system architecture view and therefore provides high data availability and redundancy. It usually contains at least two storage processing systems (i.e. CPU, cache, front-end and back-end ports are all at least dual to ensure redundancy and availability), multiple hard drives, and at least two independent loops connecting the processing systems and hard drives (EMC, 2002). One current problem with the SAN is that there are many vendors offering different products with a variety of hardware, firmware, and software. This problem might be solved soon, as vendors formed Storage Networking Industry Association (SNIA), whose purpose is coordinating works and generating standards (SNIA, 2004).

The SAN products were too expensive to fit in the budget of small / medium business. Recently lowered product price make it possible for the Library to use this solution. The SAN brings the Library the following benefits:

(a) Storage centralization

The SAN allow multiple servers to allocate and reallocate storage without moving storage around. Storage requirement is always a concern for servers and applications. The library 30 servers such as digital content, file sharing and email servers all have their own storage. System administrators sometimes have to move data around, but this can be a very difficult and time-consuming task due to complex settings of hardware and software. The centralization of storage with the potential to grow makes it easier to manage and eliminates the task of moving data around. This benefit meets our CMS system requirements for big storage and the possibility to grow in the future.

(b) Storage sharing

The SAN provides another possibility to share storage over multiple systems. Both systems build on the top of the SAN by sharing the storage and other advantages. Currently both systems are sharing the SAN's storage. In case of future needs for growth, there will be two ways to achieve it. We can add another multiple-disk array (with Terabyte space) to the current SAN. It is also possible to adjust the current setting and to re-allocate space for the desired systems.

In addition, the SAN can provide faster backup for servers. Our existing servers all have their own backup tape drives and software. This strategy requires a large number of tapes and tape drives, and even worse, the significant amount of time that system administrators spend on backup.

This solution has two pitfalls: cost and learning curve. With this size of implementation, we estimate that the cost of the current HA setting is about \$10,000 more than that of individual servers. However, the average cost per system will be down when multiple servers will be built on the top of the SAN. Furthermore, though the SAN is not difficult to manage, our systems administrators have to spend a few days to learn.

2.4 Evaluation

The systems analysis of the two systems allowed us to build an integrated HA platform that serves different computing needs of the web server and the CMS system. The following diagram is the system architecture view of our platform. Server A and B are using the Windows 2003 Clustering technologies with a "heartbeat" line linking together, providing a HA system to host the critical Library's web server. If there are software (e.g. OS patches) and hardware changes in Server A/B, the changes can be made at each server at one time. Server C is the CMS system that owns a lot of storage. All three servers are based on the SAN via two Fiber Channel Switches. The system eliminates a single point of failure except Server C. Another server is being planned to work with Server C to be another cluster. The platform has been running since July 2004 and achieved near-100% uptime.

Figure 3: The Library Integrated Platform System Architecture

3. Implications and Recommendations

3.1 Other Implications

The above example demonstrated one use of platform. It is also possible to consolidate other systems into this platform. Since this HA platform separates the implementation of HA and data management, the integrated HA platform can be extended and used for other systems. For example, file servers also require a HA environment in addition to huge storage, and this HA platform provides an excellent fit. Our next steps will focus on centralizing our database servers and providing HA clustering at the database level. The Oracle Real Application Clusters (RAC) is the proposed solution to consolidate multiple database software into a unified, centralized solution. In addition, we are considering improving our backup strategy with the SAN. The approach will use tape drives and libraries compatible with the SAN in order to reduce backup cost and staff time.

Our implementation demonstrates that different systems can share the same storage unit to host both open source and commercial systems. The use of new technologies enhances the functionality of current systems and brings new perspectives to integrate growing systems. Moreover, it provides a new way to consolidate and centralize routine tasks for managing libraries' technical infrastructure and at the same time fulfills different computing needs of multiple systems. Regarding its own functionality and sustainability, the system is both HA and flexible to minimize software and hardware failures.

3.2 Recommendations for Assessment and Implementation

For people who plan to adapt this approach, a systems analysis process is highly recommended to understand existing issues and desired requirements. The systems analysis generally consists of four phases: (1) understanding your business needs; (2) analyzing systems requirements; (3) analyzing and making decision; and (4) implementing the system.

- In investigation phase (phase 1), you may ask high-level questions related to current issues and business needs. For example, how much is my budget? What is the implementation timeline? Who will be involved? What services will the system(s) provide? Does the system(s) need HA only or both HA environment and huge storage? For example, if your system needs HA only, it would be easier to skip the SAN unit.
- In phase 2, you should ask questions at the system level and research functional requirements to meet your needs. The functional requirements at the hardware level include: OSs (Windows and/or Linux), system uptime (e.g. 99.5%), storage requirement (e.g. 500GB), database interactions, and backup. A checklist of functional requirements and constraints should be generated.
- During the decision making phase (phase 3), you can evaluate different clustering technologies and different SAN products against the checklist. When considering clustering technologies,

it is better to focus on your team skills and your budget because the clustering technologies are mature. When choosing SAN products, you should consider critical information about vendors such as reputation, products specifications, customer services, OS support, and backward compatibility. For example, if the vendor is not a SNIA member, its future is questionable. Reviews and listservs are excellent resources to find out more information.

- Once a consensus is reached to implement this approach, an implementation plan should be drafted, including: timeline, people involved, and system specifications (servers/network: host names, IP addresses, Network Interface Cards, OS version, HA and/or SAN drivers, backup strategy). A typical implementation takes 2 to 7 days to complete. A testing plan is recommended to test all the components of the system. For example, testing the "Heartbeat" line and recording the time of system recovery; testing connections between the SAN and the servers. All documents from these phases are critical in implementation and administration and should be fully archived.

4. Conclusion

This paper describes requirements and problems in the Library's current system architecture, and its approach to build an integrated hardware platform to serve different computing needs. It discusses an approach to build a reliable HA platform to integrate different systems. The critical components of the HA platform are the Linux/Windows Clustering solutions for achieving systems' availability, and the SAN for managing data by centralizing and sharing storage. Benefits of this approach include: (a) reliable, available, and serviceable systems with minimal downtime; (b) storage centralization, sharing and growth possibility for easier data management; (c) centralized management for easier server management such as backup. We integrated the Library's critical web server and the CMS based on the HA and flexible storage platform, which achieved nearly 100% uptime. The platform provides tested and indisputable answers to all the questions raised in the beginning of this paper, including unplanned downtime, storage issues, backup difficulty, and system patches. This platform is a major system improvement for the Library to better serve users' needs. Our future plans will focus on centralizing backup strategy, clustering databases and integrating other servers using this approach.

Biography

Yan Han is a Systems Librarian at the University of Arizona Library. He received MS in Computer Science and MLIS from the University of Western Ontario, Canada. His current work involves the integration and development of systems and tools for digital libraries and library systems, including the Geotechnical, Rock and Water Resources (GROW) Digital Library <<http://www.grow.arizona.edu>>, digital journal delivery such as the Journal of Range Management Archives <<http://jrm.library.arizona.edu>> and the Rangelands. He can be contacted at: 1510 E. University Blvd, Tucson, AZ, 85721-0055, USA. Email: hany@u.library.arizona.edu.

References

- Barroso, L.A., Dean, J. & Holzle, U. (2003), " Web Search for a Planet: The Google Cluster Architecture", *IEEE Micro*, vol. 23, no. 2, pp. 22-28.
- EMC, (2002), *Introduction to the CX600: the next-generation CLARiiON storage system*, EMC.
Available: http://www.emc.com/pdf/products/clarion/c955_cx600.pdf [2004, 10/27].
- eWeek, (2004), "Marketplace to the world", *eWeek*, vol. 21, no. 35, pp. 22-24.
- Freedman, A. (1999), *The computer desktop encyclopedia*, 2nd edn, Amacom, New York.
- Harnur, S. (2000), *Google Relies Exclusively on Linux Platform to Chug Along*. Available:
<http://www.hpworld.com/hpworldnews/hpw009/02nt.html> [2004, 10/26].
- Holzle, U. (2002), *The Google Linux Cluster*, University of Washington, Seattle. Available:
http://www.cs.washington.edu/info/videos/asx/colloq/UHoelzle_2002_11_05.asx. [2004, 10/26].
- Marcus, E. & Stern, H. (2003), *Blueprints for high availability*, 2nd edn, Wiley, New York :
Chichester.
- Microsoft, (2002), *Microsoft computer dictionary*, 5th edn, Microsoft Press, Redmond, Wash.
- Red Hat, (2003), *Delivering High Availability Solutions with Red Hat Cluster Suite*, Red Hat.
Available: http://www.redhat.com/whitepapers/rha/RHA_ClusterSuiteWPPDF.pdf [2004, 10/22]
- Roberson, A., (2004), *Linux-HA downloads*. Available: <http://www.linux-ha.org/download/> [2004, 10/27].
- SNIA, (2004), *About SNIA*, Available: <http://www.snia.org/about/> [2004, 10/26]
- Whitten, J.L., Bentley, L.D. & Dittman, K.C. (2001), *Systems analysis and design methods*, 5th edn, McGraw-Hill, Boston.