Analysis of Storage Devices

Saurabh Suman[1], Mahesh Kokate[2], Rajni Kori[3]and Pravin Malviya[4]

[1,2,3,4]VJTI/Computer Department, Mumbai, India
[1]Email: Saurabh.mvjti@gmail.com., [2]Email: kokate_mahesh@yahoo.co.in
[3]Email: kori07rajni@gmail.com., [4]Email: malviyapravin2010@gmail.com

***Abstract* – In the recent growing IT sector it is important to store the data for future use and maintain those data. IT/storage managers and storage professionals across companies of all sizes face the different critical challenges such as Managing storage growth, Designing, deploying, and managing backup and recovery, Designing, deploying, and managing storage in a virtualized server environment, Designing, deploying, and managing disaster recovery solutions, Storage consolidation, Making informed strategic/big-picture decisions, Integrating storage in application environments (such as Oracle, Exchange, etc.), Designing and deploying multi-site environments, Lack of skilled storage professionals. In this paper we tell about different types of storage devices available now a days and their performance and limitations. We also discuss evolution of Storage Devices. How they differs from each other's and how effective they are in use and storage.**

Index Terms – CAS, DAS, NAS, SAN, RAID

I. INTRODUCTION

Information storage is the foundation of the information technology. As society increasingly relies on digitally stored and accessed information, supporting the availability, integrity, and confidentiality of this information is crucial. We need systems in which users can securely store critical information, ensuring that it persists, is continuously accessible, cannot be destroyed, and is kept confidential.

A large quantity of digital information is being created every moment by individual and corporate consumers of IT. This information needs to be stored, protected, optimized, and managed. Information is increasingly important in our daily lives. We have become information dependents of the twenty-first century, living in an on-command, on-demand world that means we need information when and where it is required. We access the Internet every day to perform searches, participate in social networking, send and receive e-mails, take pictures and videos through digital cameras, and satisfy many other personal and professional needs. Equipped with a growing number of content-generating devices, more information is being created by individuals than by businesses. Information created by individual's gains value when shared with others. When created, information resides locally on devices such as cell phones, cameras, and laptops. To share this information, it needs to be uploaded via networks to data centers.

The section II describes the evolution of the storage system, the section III describes about the different storage system and the section IV compares SAN,NAS and CAS.

II. EVOLUTION OF STORAGE SYSTEM

The evolution of open systems and the affordability and ease of deployment that they offer made it possible for business units/departments to have their own servers and storage. In earlier implementations of open systems, the storage was typically internal to the server.

The proliferation of departmental servers in an enterprise resulted in unprotected, unmanaged, fragmented islands of information and increased operating cost. The highlight of storage architecture is following

- *Redundant Array of Independent Disks (RAID):* This technology was developed to address the performance and availability requirements of data. It continues to evolve today and is used in all storage architectures such as DAS, SAN, and so on.
- *Direct-attached storage (DAS):* This type of storage connects directly to a server (host) or a group of servers in a cluster. Storage can be either internal or external to the server. External DAS alleviated the challenges of limited internal storage capacity.
- *Storage area network (SAN):* This is a dedicated, high-performance *FibreChannel (FC)* network to facilitate *block-level* communication between servers and storage. Storage is partitioned and assigned to a server for accessing its data. SAN offers scalability, availability, performance, and cost benefits compared to DAS.
- *Network-attached storage (NAS):* This is dedicated storage for *file serving* applications. Unlike a SAN, it connects to an existing communication network (LAN) and provides file access to heterogeneous clients. Because it is purposely built for providing

storage to file server applications, it offers higher scalability, availability, performance, and cost benefits compared to general purpose file servers.

- *Internet Protocol SAN (IP-SAN):* One of the latest evolutions in storage architecture, IP-SAN is a convergence of technologies used in SAN and NAS.

III. STORAGE DEVICES

A. Storage Area Network

The SAN (as shown in figure 1) as a network whose primary purpose is the transfer of data between computer systems and storage elements. The Storage Network Industry Association (SNIA) defines this. It can also be defined as a SAN can also be a storage system consisting of storage elements, storage devices, computer systems, and/or appliances, plus all control software, communicating over a network. It is sometimes referred to as "the network behind the servers." A SAN allows "any-to-any" connection across the network, using interconnects elements such as routers, gateways, hubs, switches and directors.

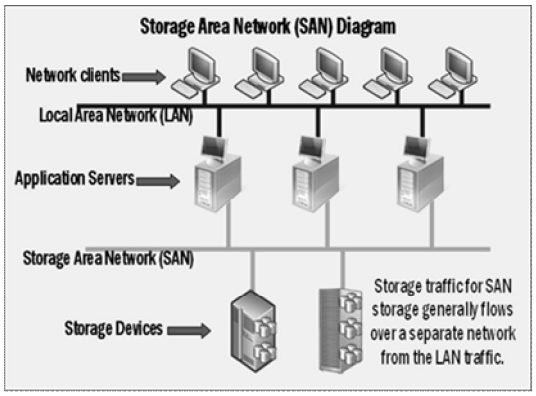

Figure 1: Storage area network

Components of SAN (as shown in figure 2)

A SAN consists of three basic components:

- Server
- Network infrastructure
- Storage

The component can be further broken down into some key elements:

- Node Ports
- Cablings
- Interconnecting Devices
- Storage Arrays
- SAN Management Software

Benefits of SAN

Sharing storage usually simplifies storage administration and adds flexibility since cables and storage devices do not have to be physically moved to shift storage from one server to another.

It has the ability to allow servers to boot from the SAN itself. This allows for a quick and easy replacement of faulty servers since the SAN can be reconfigured so that a replacement server can use the LUN of the faulty server.

By integrating storage devices, SAN increases the storage space usability and cost efficiency.

- SAN is the high-speed storage sharing system.
- SAN increases the network bandwidth and reliability of data I/O.
- SAN is separated from the regular network system, and has an ability to expand the storage capacity.
- SAN reduces the cost of the storage management since it simplifies the system fabric and devices management.

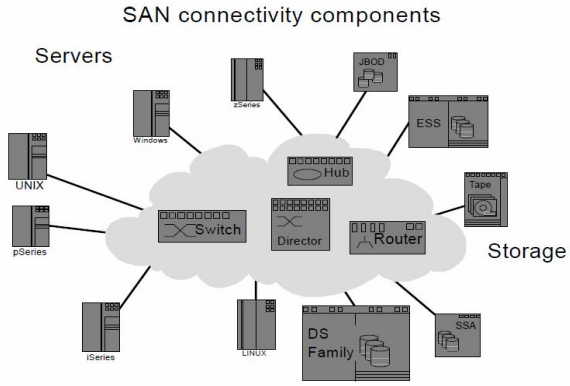

Figure 2: Component of Storage area network

Limitations of SAN

- Poor interoperability
- High Cost
- Difficult to manage
- Weak Security
- Scalability is a concern in complex deployments
- FC is not a routable protocol

A. Network Attached Storage

Network Attached Storage (NAS), as shown in figure 3, is basically a LAN-attached file server that serves files using a network protocol such as Network File System (NFS). NAS is a term used to refer to storage elements that connect to a network and provide file access services to computer systems. A NAS storage element consists of an engine that implements the file services (using access protocols such as NFS or CIFS), and one or more devices, on which data is stored. NAS elements may be attached to any type of network. Network Attached Storage separates the

application server from the storage. This increases overall system performance by allowing the servers to perform application request and the NAS to serve files or run applications.

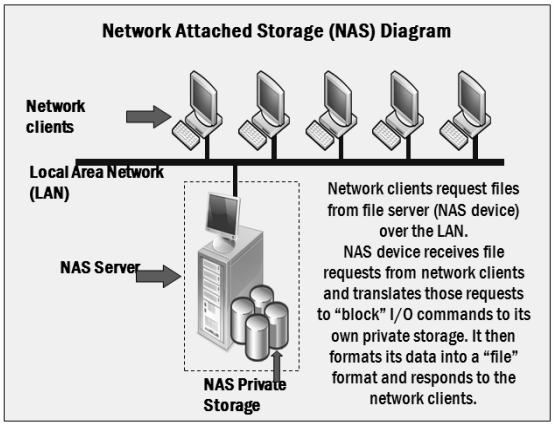

Figure 3: Network attached Storage

Components of NAS (as shown in figure 4)

- NAS head(CPU and Memory)
- Network infrastructure card (NICs):- It provides the connectivity to the network.
- An optimized operating system for managing NAS functionality.
- NFS and CIFS protocols for file sharing.
- Industry standard storage protocol to connect and manage physical disk resources, such as ATA, SCSI.

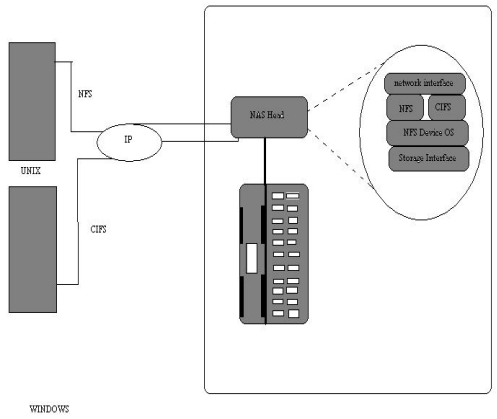

Figure 4: Component of network attached storage

Factors affecting NAS performance and availability

- Numbers of hopes
- Authentication with a directory service such as LDAP, Active directory, or NIS
- Retransmission
- Over utilized router and switches
- File or Directory lookup and metadata request
- Over utilized NAS devices

NAS Application and its Benefits

- File/Printer
- Application specific server
- Video imaging
- Graphical image store
- Centralized heterogeneous file sharing
- File system mirroring

- *Improved efficiency:* Eliminates bottlenecks that occur during file access from a general purpose file server because NAS uses a operating system specialized for file serving.

- *Improved flexibility:* Capability for clients on both UNIX and WINDOW platforms using industry standard protocol. NAS is flexible and can serve request from different types of clients from the same source.

- *Centralized storage: C*entralized data storage to minimize data duplication on client workstations, simplify data management, and ensures greater data protection.

- *Simplified management :* It provides a centralized console that makes it possible to manage file system efficiently.

- *Scalability:* Scale well in accordance with different utilization profiles and types of business applications because of the high performance and low latency.

- *High availabilit y :* Offers efficient replication and recovery option, enabling high data availability. NAS uses redundant network component that provide maximum connectivity option.

- *Security:* Ensures security, user authentication, file locking in conjunction with industry standard security schemas.

Limitations of NAS

There is no guarantee that the security information associated with the files and directories will be processed.

There will be performance degradation due to the fact that data is being accessed remotely.

An NAS server, using FTP or SMBs, uses a simple, clear text exchange of account name / password. This isn't as secure as Kerberos NAS devices which don't use NTFS may not support anything better than Simple File Sharing. Certain NAS devices fail to expose well-known services that are typical of a file server, or enable them in a way that is not efficient

If the NAS is occupied with too many users, too many I/O operations, or CPU processing power that is too demanding, the NAS reaches its limitations. A

server system is easily upgraded by adding one or more servers into a cluster, so CPU power can be upgraded, while the NAS is limited to its own hardware, which is in most cases not upgradeable.

B.Content-addressable storage

Content-addressable storage, also referred to as associative storage or abbreviated CAS, is a mechanism for storing information that can be retrieved based on its content, not its storage location. It is typically used for high-speed storage and retrieval of fixed content, such as documents stored for compliance with government regulations. Roughly speaking, content-addressable storage is the permanent-storage analogue to content-addressable memory .CAS is an object-based system that has been purposely built for storing fixed content data.

It is designed for secure online storage and retrieval of fixed content.CAS Architecture

A client accesses the CAS-Based storage over a LAN through the server that runs the CAS API.

The CAS API is responsible for performing functions that enable an application to store and retrieve the data.

CAS architecture is a Redundant Array of Independents Nodes (RAIN).IT contains the storage node and access nodes networked as a cluster by using a private LAN that is internal to it.

Clients access the CAS on a separate LAN, which is used for interconnecting clients and servers to the CAS.

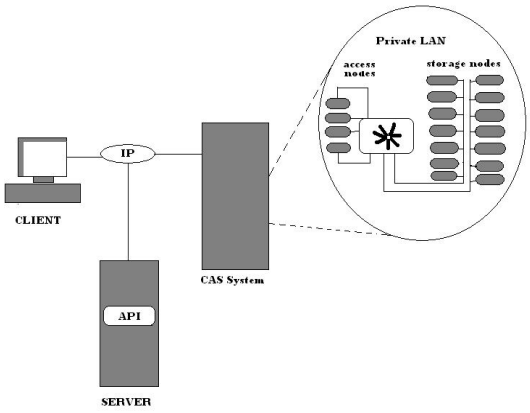

Figure 5 : CAS Architecture

Features and Benefits of CAS

- *Content authenticity*: It assures the genuineness of stored content. Content authenticity is assured because the address assigned to each piece of fixed content is unique as a fingerprint.
- *Content integrity*: Refers to the assurance that the stored content has not been altered .Use of hashing algorithm ensures content integrity.
- *Location independence*: CAS uses a unique identifier that application can leverage to retrieve data rather than a centralized directory, path names, URLs . This yields complete content mobility to application across locations.
- *Single instance storage***:** The unique signature is used to guarantee the storage of only a single instance of an object. This signature is derived from the binary representation of the object.
- *Retention enforcement*: Protecting and retaining data object is a core requirement of an archive system.
- *Record level protection and disposition*: All fixed content is stored in CAS once and is backed up with a protection scheme.
- *Technology independence*: The CAS system interface is impervious to technology changes.
- *Fast record retrieval*: CAS maintain all content on disks that provide "sub second "time of first byte" (200ms-400ms) in a single cluster.

Limitations

For data that changes frequently, CAS is not as efficient as location-based addressing. In these cases, the CAS device would need to continually recompute the address of data as it was changed, and the client systems would be forced to continually update information regarding where a given document exists. For random access systems, a CAS would also need to handle the possibility of two initially identical documents diverging, requiring a copy of one document to be created on demand.

IV. COMPARASIONS OF SAN, NAS AND CAS

Points	Storage area network (SAN)	Network Attached storage (NAS)	Content addressed storage (CAS)
Connectivity	Fiber Channel ,ISCSI	IP	IP
Data access method	Block	File	File, object
Key requirement	Deterministic performance, support for high-transaction	Sharing, collaboration	Long-term data retention, data integrity
Types of application	OLTP ,Data warehousing ,ERP	File and print server consolidation, product design	Archiving reference data store
Typical market segmentation	Large business	All	Middle and large business
Limitation	High cost, difficult to manage, less security	Performance degradation due to remotely accessing of data	For data that are changes frequently its not efficient for that.

V. CONCLUSION AND FUTURE WORK

In the survey and analysis we discuss about different types of storage devices, application and limitation. We have studied the architecture of the SAN, NAS and CAS. How these storage devices are useful in today world .We uses these storage devices as our requirement and based on their performance.

In future we combine the SAN and NAS technology over an IP for better performance and for bigger enterprises.

REFERENCES

[1] Bilski T., A Formal Model for Data Security Evaluation, [In:] *Selected Papers of the Fifth International Conference on Computational Science and Applications*, (Eds. M. Gavril ova, O. Gervasi), Kuala Lumpur, IEEE, 2007, s. 253-257.

[2] Bilski T., Data storage and transmission convergence Concept, E. Kozan (ed.) *Proceedings of the Fifth Asia Pacific Industrial Engineering and Management Systems Conference*, Queensland University of Technology, 2004, 14.8.1-14.8.16.

[3] http://www.snia.org

[4] http://www.nas-san.com/

[5] http://www.storagesearch.com/

[6] http://www.brocadekorea.com

[7] Information Storage and Management by EMC^2