Storage Virtualization: Towards an Efficient and Scalable Framework

Bilal Y. Siddiqui *bsiddiqui@islander.tamucc.edu*
School of Engineering and Computing Sciences
Texas A&M University-Corpus Christi
Corpus Christi, TX 78412, USA

Ahmed M. Mahdy *ahmed.mahdy@tamucc.edu*
School of Engineering and Computing Sciences
Texas A&M University-Corpus Christi
Corpus Christi, TX 78412, USA

Abstract

Enterprises in the corporate world demand high speed data protection for all kinds of data. Issues such as complex server environments with high administrative costs and low data protection have to be resolved. In addition to data protection, enterprises demand the ability to recover/restore critical information in various situations. Traditional storage management solutions such as direct-attached storage (DAS), network-attached storage (NAS) and storage area networks (SAN) have been devised to address such problems. Storage virtualization is the emerging technology that amends the underlying complications of physical storage by introducing the concept of cloud storage environments. This paper covers the DAS, NAS and SAN solutions of storage management and emphasizes the benefits of storage virtualization. The paper discusses a potential cloud storage structure based on which storage virtualization architecture will be proposed.

Keywords: Storage Virtualization, Cloud, Tiered Storage.

1. INTRODUCTION

To overcome critical storage issues faced by organizations, the need for a storage management solution arises. Storage architectures have been proposed which efficiently manage data in an environment. Intelligent storage systems distribute data over several devices and manage access to data [1]. Storage solutions are hence the top priority for organizations, considering integrity, availability and protection of data. With a plethora of options available, the most prevalent and traditional solutions have been direct-attached storage (DAS), network-attached storage (NAS) and storage area networks (SAN).

The storage components of the three traditional solutions namely DAS, NAS and SAN will be discussed. Various factors affect data storage selection such as performance, scalability, capacity, availability and reliability. Data protection operations such as backup and recovery are the backbone of any storage environment.

This paper focuses on a storage virtualization technology that creates logical abstractions of physical storage. Storage virtualization offers a number of benefits including reduction in management overheads and administration complexity [2].

Storage clouds on the Internet can be created using virtualization. The cloud storage system accommodates various storage sizes ranging from small storages to colossal devices. This paper discusses the cloud storage concept and a layered cloud storage structure. Virtual storage architecture will be advocated and analyzed.

Increasing the scalability and capacity of data is the inclination and goal of this work. Storage virtualization is the concept which encourages this purpose; improving data management in a storage environment.

2. INFORMATION STORAGE REVIEW

2.1 Direct Attached Storage (DAS)

DAS refers to a storage system attached directly to a server without the involvement of a network system. It is a non-networked storage solution, which does not provide access to other devices in the storage environment. This concept uses the server as the center for data transmission purposes. The main protocols used in DAS connections are ATA, SATA, SCSI, SAS and Fiber Channel [3].

The DAS solution for storage management, with its simple architecture is easy to understand and works well for small scale storage systems. It ensures security and efficiency of the highest order in such environments.

Deployment in large scale systems made the limitations of DAS clear. Since all data access transactions require transfer through a server, the performance of DAS storage method is affected in large scale storage systems such as an enterprise environment. When the performance of a system gets worse, it leads to a performance bottleneck. Moreover, disabling a server can make the whole system invalid, which is referred to as single point failure [4].

Other shortcomings of the DAS solution were issues pertaining to complex data sharing and high costs of data management. Large amounts of uncontrolled data meant wastage of system resources, which would lead to data inconsistency. The aforementioned are considered major disadvantages of the DAS architecture.

2.2 Network Attached Storage (NAS)

NAS is another primary solution for data storage. It is a file-level data storage which utilizes the built-in server as central data storage. NAS systems are normally networked and may be comprised of multiple logically or RAID arranged storage devices. Using file sharing protocols like AFP, NFS, and CIFS/SMB, file access is provided [4].

NAS appliances are storage units directly attached to a local area network (LAN). Where DAS architecture relied on link to a single server, NAS appliances can be accessed by any client on the network without server involvement. One solution to the problems suffered by DAS is to connect NAS appliances directly onto the network. These appliances only present/transfer data as files facilitating file sharing between users and servers across multiple operating systems.

Meeting the ever increasing demand of storage space requires adding extra capacity to NAS appliances. Eventually, when the limit is reached, the only solution is to deploy additional appliances. Servers need direct attached storage for high speed block based access. This creates a fragmented data environment. Data backup operations are also conducted on the LAN in the NAS storage method.

NAS provides solutions to a number of business requirements including solving the problems of DAS. However, as data requirements grow, NAS may become inefficient and/or inadequate. Increased dependency on the NAS controller means that all data will be funneled through the NAS appliance, which may result in a bottleneck. Limited backup functionality can also be a disadvantage of the NAS architecture.

2.3 Storage Area Network (SAN)

SAN is the most popular solution, which has been rapidly spreading across organizations. This concept involves moving networked storage from its traditional position to a separate location, which is a network of its own. Storage devices can be connected directly to this network by a fabric which contains switches and hubs to connect any storage to the servers as if they were local. SAN is a dedicated high speed network, specifically created to facilitate data transfer not only between servers and storage devices but also between multiple storage devices. SANs depend heavily on a high-speed fiber channel technology to provide flexible connectivity requirements. Introduction of a storage fabric resolved some connectivity issues found in DAS architectures. This storage method helps achieve high speed data backup without any impact on LAN performance. SAN contributed to solve the scalability issues of DAS and NAS by allowing servers and storage devices to scale independently [4] [5].

The major disadvantages of SAN storage systems are that SAN upgrades may cost more than the initial installation and are not easy to merge/integrate with existing technologies or other vendor equipment. Users usually have to stick to a single vendor, which results in isolated pockets/islands of SANs. Storage virtualization offers a viable solution to the problems faced by SAN storage systems.

3. VIRTUALIZATION OF STORAGE

Virtualization involves abstracting or isolating internal functions of storage systems or general network resources to enable network independent data/storage management. According to [1], virtualization of storage is technically the pooling of physical storage from various storage devices, disguised to be a single storage device which can be administered centrally.

Storage virtualization increases performance of various tasks involved in storage administration, such as backup, recovery, archiving, etc. in a timely and cost effective manner. Provisioning and managing distributed storage as a consolidated resource promotes simplicity. Storage virtualization automates storage expansion, reducing manual provisioning and hence decreasing downtime.

Digital storage in a home storage environment has been growing steadily. The use of storage devices, including external hard drives, has become a common practice for the average consumer. These devices have a limited lifetime and therefore with age, they may fail. While the size and complexity are significantly less, the management of home digital storage can be improved by adopting techniques used in large scale datacenters [6].

Datacenters in an enterprise storage environment consist of a collaboration of storage devices and systems that are created by aggregating storage devices. These storage systems are networked together to improve data efficiency and movement in the network. Enterprises also deal with content stored on multiple types of storage systems, referred to as “storage tiering” [7]. This storage architecture is a common characteristic of large-scale storage systems such as those in a datacenter.

3.1 Cloud Storage

Clouds consist of numerous storage devices grouped by network, file systems and various storage components. Since clouds are presented as a service, they are referred to as a storage service system. Cloud storage architectures may be comprised of network devices, storage pools, file systems, service-level interfaces and common access interfaces, etc. It is made up of distributed resources, but acts as a single resource, hence providing high fault tolerance through data redundancy and distribution.

Cloud storage components can be categorized by their functions as physical and logical so as to provide more scope for interactions and compatibility. Figure 1 shows a potential cloud storage structure which is presented as a generalized and layered architecture.

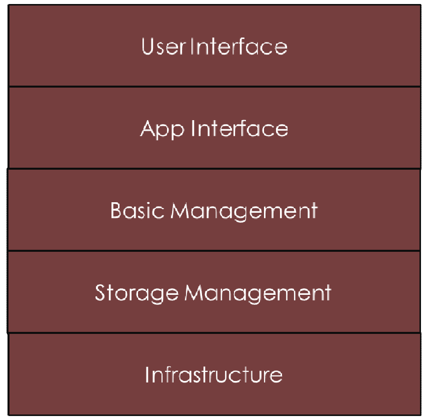

FIGURE 1: Cloud Storage Structure.

The structure from bottom to top consists of infrastructure, storage management, basic management, app-service interface and the user interface.

The infrastructure of a cloud storage system mainly consists of storage components such as storage devices, servers, and storage networks (NAS, SAN, FC, iSCSI, SCSI, SAS, etc.). It also comprises of wired/wireless networks.

In the storage management layer, there can be a storage manager or a management system which administers or advocates the management of storage in the infrastructure layer. Its most important role is data management. This layer collaborates between multiple domains to ensure data redundancy, fault tolerance, load balance and storage maintenance.

The basic management layer furthers the collaboration among storage devices by incorporating technologies such as cloud computing, distributed file system, etc., to enable data access. It is the central layer of a cloud storage structure which is the medium of interaction between the above and below layers.

App interface supplies a platform/link between applications and users for interaction. It supports various applications and platforms to help bundle applications under the cloud storage environment and be provided as a service to the users.

Finally, the user interface is the front end of the cloud storage structure. This layer is the point of interaction with the end users.

4. A PROPOSED VIRTUAL STORAGE ARCHITECTURE

For easy and dynamic data allocation, the end user has to have access to storage space without hindrance. To achieve this goal, this paper proposes a virtual storage structure in Figure 2.

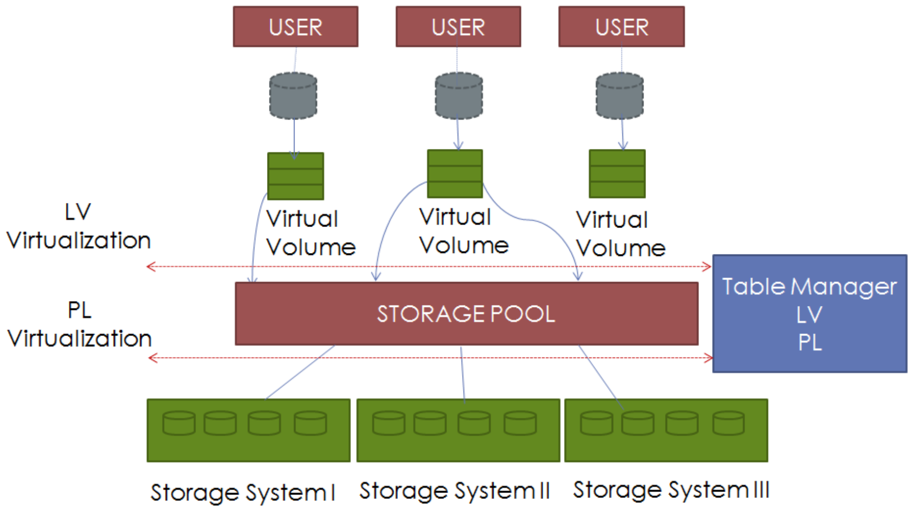

FIGURE 2: Proposed Virtual Storage Architecture.

As seen, storage systems are virtualized by storage virtualization techniques. The end users are provided with a logical volume of storage. This volume of storage will have a size as specified by the end user. There is a mapping table between the logical and virtual storages for each user. These tables record/preserve data on mapping relationships among users and storage pools. The user will be able to see enough space to meet its storage needs.

To manage mapping tables, a table manager is devised. This table manager manages logical-virtual virtualization and physical-logical virtualization. Separate managers can also be setup for independent logical-virtual/physical-virtual virtualizations.

The most severe problem virtualization can help solve is management of storage, considering the heterogeneous nature of a storage environment consisting of file systems, operating systems, management consoles, storage systems, management software, etc. Coping with the exponentially increasing demand for storage space both physical and virtual environments poses a major challenge. Data protection in scenarios such as server overload, server recovery, failed backups and service interruption is essential since critical data is present in a SAN.

In the proposed architecture of virtual storage, since the logical volume has been made virtual, users can be allocated virtual spaces of any size. Even if the allocated size exceeds the physical storage space, the user is unable to see the physical storage space. This helps in increasing simplification of storage administration, storage capacity and data utilization [6].

Since the data is virtualized into storage pools, storage space can be expanded as users need more storage space. Both the physical and logical spaces can be expanded dynamically according to the users' needs. Users can have huge virtual storage spaces.

However, since the proposed architecture involves the introduction of mapping tables and mapping managers, there is a possibility of increased complexity of the storage environment [6].
To increase productivity and meet consumer demands, valuable data has to be secured both internally and externally [9]. Virtualizing the physical space to logical volumes and the logical volumes to virtual volumes can make the system prone to malicious attacks. This architecture will be investigated further to ensure efficiency and scalability. Future work will include incorporation of encryption/decryption protocols in the proposed framework for assurance against malicious attacks.

5. CONCLUSION

This paper discussed various components of the three traditional storage topologies, namely direct-attached storage (DAS), network-attached storage (NAS) and storage area networks (SAN). The concepts of cloud storage and storage virtualization have been focused upon, following which a layered cloud storage structure has been discussed. Finally, considering the efficiency and scalability of data utilization as focal points, a virtual storage architecture has been proposed for future consideration.

6. REFERENCES

[1] Somasundaram, G., Shrivastava, A. “Information Storage and Management: Storing, Managing, and Protecting Digital Information,” John Wiley & Sons, 2009.

[2] Clark, T. “Storage Virtualization: Technologies for Simplifying Data Storage and Management,” Addison-Wesley Professional, 2005.

[3] Peng, Y., Zhu, Y., and Luo, J. “A Network Storage Framework Based on SAN,” The 7th International Conference on Computer Science & Education (ICCSE), 2012.

[4] Han, D. “SNIS: A Storage Networking Integrating NAS and SAN,” Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, 2005.

[5] “Storage Technology Foundations Networked Storage Introduction,” EMC Education Services, 2007.

[6] Coughlin, T. “Virtualization of Consumer Storage,” IEEE 14th Symposium on Consumer Electronics, 2010.

[7] Farley, M. “Building Storage Area Networks,” Osborne Press, 2001.

[8] Mahalingam, P., Jayaprakash, N., and Karthikeyan, S. “Enhanced Data Security Framework for Storage Area Networks,” Second International Conference on Environmental and Computer Science, 2009.