IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 17, Issue 1, Ver. IV (Jan – Feb. 2015), PP 60-64
www.iosrjournals.org

Cloud Storage: Focusing On Back End Storage Architecture

Sarishma, Kartik Mishra
CSEwithspecialization inCloud Computing and Virtualization Technology (2012-2016)
UniversityofPetroleum&EnergyStudies, Dehradun,India

Abstract: *In the modern era of mobile and cloud computing, people are becoming more and more dependent on digital devices. In order to execute any application, a certain amount of storage space is mandatory which is to be used by the application as its own warehouse to store its data.While designing any storage architecture, we have data as the centre of attraction around which whole of our application design revolves.Cloud storage is a hot topic nowadays as the data storage capacity rates are increasing manifold's every year and has thus become a reality that all data centers and organizations should consider. This huge amount of data, thus poses a challenge for the construction of a good well defined, fault prone back end storage.This paper representsthe different available architectures that are used in storage technology foundation. Beginning with a conceptual overview of the SNIA reference model for cloud storage,the key concepts of cloud and other technologies which form a base for cloud storage are discussed. Followed by this, the three standard architectures related to cloud storage are discussed which are basicallyStorage Area Network (SAN), Direct Attached Storage (DAS) and Network Attached Storage(NAS). The paper concludes by pinpointing the future research and open challenges related to cloud storage.*

Keywords: *Mobile Computing, Storage architecture, Cloud Storage, SNIA.*

I. Introduction

Since a decade ago, the wide demand of increase in storage capabilities of devices aroused the need of well-formed storage architectures to meet the voracious storage demands. Earlier users used to buy their own hardware to store and carry their personal data. Slowly but unceasingly the amount of data which different users produced increased and so did the demand of storage mediums. Data can be considered as a pool of raw facts and figures which can be combined together to create a logical meaning. Depending upon the storage and managing medium of data, it can be classified into two categories, structured data and unstructured data. As the name suggest, structured data is stored in the form structures i.e. rows and columns which makes it much easier for retrieval as well as for access. On the other hand, unstructured data is the one which cannot be stored in the form of structures i.e. rows and columns and therefore, retrieval and access to unstructured data becomes pretty much difficult. Statistically speaking, about 80% of the data stored on cloud is unstructured which makes it difficult to use.

With the advent of cloud computing, users are facilitated with the view of unlimited storage space being available to them on pay per use basis. Individual users generate more data in the form of digital content such as audio, video, images, mp3, documents etc. as compared to the data generated by business enterprises. People leverage the availability of storage mediums to such an extent that they can move anywhere around the world with their own personal huge amount of data. They can communicate with people across the globe and share data with the geographically apart areas. This communication and sharing when viewed from an external perspective is quite simple but coming to the technicalities it becomes pretty hard to handle such a large amount of stored content. When it comes to transfer and sharing of data while deploying cloud services, it becomes important to manage every process efficiently. The segregation among data, security checks, latency, cost etc. are the factors which influence the cloud services related to storage. Uploading this data on cloud servers is done via network and it therefore becomes very important to focus on what kind of storage architecture a particular cloud service provider is using. There are a large number of available architectures for cloud storage but we are presenting the most viable, suitable and accepted architectures which are SAN, NAS and DAS. SAN is acronym for Storage Attached Network that provides access to consolidated, block level data storage. NAS stands for Network Attached Storage and is a file-level computer data storage server connected to a computer network providing data access to a heterogeneous group of clients. DAS is a directly attached storage in the form of hard disk drives connected through a Host Bus Adapter (HBA).

Rest of the paper is organized as: Section 2 explains the fundamentals or background related to cloud assisted storage mechanisms. Section 3 explains the SNIA reference model for cloud storage accompanied with a well-definedself-explanatoryreference diagram. Section 4 presents the selected architectures which lay the foundation of cloud storage which are followed bythree types of architectures i.e. SAN, NAS and DAS. The future research and open challenges related to cloud storage are presented in the section 5. The last section i.e. section 6 concludes our paper.

II. Background

Before going deep in cloud storage architecture and other technologies, we need to know about the basic background of some of the topics related to cloud computing. Proposal of storing data on cloud servers and use of virtualization techniques gave rise to the rapid emergence of cloud storage. Fundamentals of cloud computing concepts and their relation with the storage architecture are discussed below:

2.1 delivery Models:

1. Infrastructure as a Service (IaaS): This model provides access to virtualized infrastructure which acts as the basic hardware to store data on cloud. IaaS model provides virtual server space, memory, connections, bandwidth, load balancers etc. as a service to the end users. It can be considered as the base layer for storage space as all of the hardware needed for storing data is provided by IaaS i.e. it can be taken as a back end hardware providing model for cloud storage.
2. Platform as a Service (PaaS): IaaS provides the CSP's (cloud service providers) with virtualized bare hardware and platform is then needed to manage and perform other operations on the hardware. PaaS provides platform as a service which can be considered as a storage logic for the cloud storage. This model is responsible for managing, isolating, distributing and using data stored in virtualized hardware.
3. Software as a Service (SaaS): Infrastructure as well as platform on cloud is made available to end users by using SaaS whose basic functionality is to provide software as a service. The interfacing and other methods for user ease are presented in form of software under SaaS and henceSaaS can be considered as provider of front end for cloud storage.

2.2 Deployment models:

1. Public Cloud: The services available under public cloud are freely accessible by all on pay per usage basis. Cloud storage for all users is shared, although strict segregation is maintained between data of different users.
2. Private Cloud: The services available under private cloud are used by the users of an organization or of a closed network who own that private cloud. These services are not freely accessible by general public. The data stored on private cloud is thus more safe, secure and isolated. Moreover there is less chance of security breach when it comes to private cloud as the infrastructure is safe and is accessible to authorized users only.
3. Hybrid Cloud: Hybrid cloud can be considered as a combination of public and private cloud. The users can use their private cloud to store some secure data and can use public cloud at times when they need more functionality. For instance, when cloud storage limit is reached for some scenario for private cloud, at that time we can use public cloud to share the load. Critical data can be stored on private cloud and non critical data can be stored on public cloud.

III. SNIA Reference Model

The demand of cloud storage has increased drastically because of its astounding features like elasticity, pay per usage, management, view of unlimited storage, ease of use etc. Consequently, it becomes considerable to create a user interface for cloud storage which can support these qualities along with the providing an ability to stand up in future perspectives by competing with the latest trends. A reference model for cloud storage can be used to depict the different available interfaces for cloud storage which can support both the legacy as well as new futuristic applications. One such standard model proposed by Storage Networking Industry Association is discussed here, commonly referred as SNIA reference model. All of the interfaces interact with the end user and on the basis of user demand; it fetches the resources from the infrastructure pool. As depicted in the figure, CDMI (Cloud Data Management Interface) is an interface which will be used by different applications to manage, retrieve, create, remove and edit the user's data. The true potential of the hardware, storage logic and services can be determined by evaluating and observing such interfaces. In the centre, data storage cloud comprises of both the soft as well as hard data container which are used for storing the data. The cloud data management component, information services, data and storage services are used for managing the different types of demands which perform these functions by utilizing CDMI. The users can operate on cloud storage services by using a number of interfaces such as object storage client, XAM client, database or table client, file system client, block storage client etc. The rest of the figure is pretty much self explanatory.

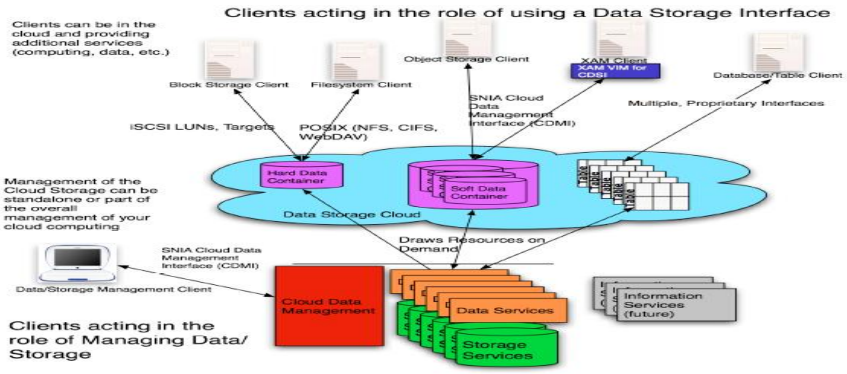

Figure 1: SNIA Reference Model [5]

IV. Architecture

The three basic types of storage are Storage Attached Network (SAN), Network Attached Storage (NAS) and Direct Attached Storage (DAS). All the three technologies have evolved over years and technological advancement in the field of storage lead to other technologies. In simple words DAS lead to NAS which in turn lead to SAN. All the three storage are described below.

4.1 SAN:

A storage area network (SAN) provides access to consolidated, block level data storage that is accessible by the application running on any of the networked server. It carries data between servers (hosts) and storage devices through fibre channel switches. A SAN helps in aiding organizations to connect geographically isolated hosts and provide robust communication between hosts and storage devices. A SAN works on its own storage devices that are not reachable through the local area network by other devices and organization often choose SAN because of its features such as more flexibility, availability and performance than the other networked architectures.

4.2 Components of SAN:

A SAN is typically assembled using three principle components: cabling, host bus adapters (HBA) and switches. Cabling is the physical medium which is used to for establishing a link between every SAN device by using transmission mediums like copper or optical fibre based on distance requirement of the organization.HBA or Host Bus Adapter is an expansion card that fits into expansion slot in a server. HBA naturally offloads data storage and retrieval overhead from the local processer which results in improving server performance. Switch is used to handle and direct traffic between different network devices. It accepts traffic and then transmits the traffic to the desired endpoint device. In a SAN, each storage server and storage device is linked through a switch which includes SAN features like storage virtualization, quality of service, security and remote sensing etc.

4.1.2 Management of SAN:

Management is a vital part of SAN operation and is carried out by using out a tool referred as SRM. Storage resource management (SRM) applications are used to check and manage physical and logical SAN resources. Physical storage resources include the basic hardware like the RAID systems,storage arrays,magnetic tape libraries devices and FC switches, whereas logical storage structure involve the basic file systems and application-oriented storage fundamentals. Ideally, a centralized SRM tool should be able to detect storage resources, estimating their capacity and configuration, and computing the performance. The SRM tool should also be able to impact changes to the configuration and support reliable policies across the various storage technologies being managed.

4.3 NAS:

Network-attached storage or NAS is a file-level computer data storage server connected to a network and providing data accessibility to a diverse group of clients. NAS is specialized for the task assigned to it either by its hardware, software or by both and provides the advantage of server consolidation by removing the need of having multiple file servers.NAS also uses its own OS which works on its own peripheral devices. A NAS operating systems is optimized for file I/O and, therefore performs file I/O better than a primitive server. It also uses different protocols like TCP/IP, CIFS and NFS which are basically used for data transfer and for accessing remote file service.

4.3.1 Components of NAS:

A NAS device can be divided into components which are named as :

1. NAS head which is basically a CPU and a memory.
2. More than one Network Interface Cards (NIC's).
3. Optimized Operating System.
4. Protocols for file sharing (NFS or CIFS).
5. Protocols to connect and manage storage devices like ATA, SCSI, or FC.

4.2.2 Implementing NAS:

NAS can be implemented in two ways, either byintegrated implementation or by gateway implementation.

An integrated implementation is the one which has all of its component and storage system in a single enclosure.It has all components of NAS, like head and storage enclosed together making it a self-contained environment. In this a NAS head connects to the IP network, providing connectivity to clients and services like file I/O request. Storage management can consist of low range ATA to high-throughput FC disk drives and is managed by the management software.On the other hand, in gateway implementation NAS head shares its storage with SAN environment. It consist of independent NAS head and more than one or more storage arrays. Head performs the same function while storage is shared with other application requiring block-level I/O.Managing these is a more complex task as there are separate administrative tasks for head and storage. It also uses and utilizes the FC infrastructure like switches, directors or DAS.This type of NAS is most scalable as head and storage can be independently scaled up whenever it is required enabling high utilization of storage capacity by sharing it with SAN.

4.4 DAS:

DAS stands for Direct Attached Storage and as the name suggests, it is an architecture where storage connects directly to hosts. DAS is ideal for localized data access and sharing in environment where small server are located for instance, small businesses, departments etc. Block-level access protocols are used to access data through applications and it can also be used in combination with SAN and NAS. Based on the location of storage devices with respect to host, DAS can be classified as external or internal.In Internal DAS,the storage device is internally connected to the host by serial or parallel buses. Most internal buses have distance limitations and can only be used for short distance connectivity and can also connect only a limited number of devices.Moreover, they also hamper maintenance as they occupy large amount of space inside the server.Whereasin internal DAS the server connects directly to the external storage devices. SCSI or FC protocol are used to communicate between host and storage devices. It overcomes the limitation of internal DAS and overcome the distance and device count limitations and also provides central administration of storage devices.

4.3.2 . Why and why not to go for DAS?

There are many considerations which need to be focused upon while considering DAS. Whether to go for DAS or not is a challenging question. Following are some points which shortlist some factors and by considering these factors one can decide whether he should go for DAS or not.

4.3.2.1 Why to go for DAS:

1. It requires low investment than other networking architectures.
2. Less hardware and software are needed to setup and operate DAS.
3. Configuration is simple and can be deployed easily.
4. Managing DAS is easy as host based tools such as host OS are used.

4.3.2.2 Why not to go for DAS:

1. Major limitation of DAS is that it doesn't scale up well and it restricts the number of hosts that can be directly connected to the storage.
2. Limited bandwidth in DAS hampers the available I/O processing capability and when capability is reached, service availability may be compromised.
3. It doesn't make use of optimal use of resources due to its lack of ability to share front end ports.

V. Future Work

1. Isolation: Isolation is maintained between data of different users in cloud storage but despite of many efforts, there are cases when this isolation is compromised leading to personal loss of the user. Creating a mechanism for achieving zero isolation among different user data is a challenge for researchers.

2. Security breaches: Providing ids and passwords, cross checks on login of user accounts, authentication etc. is done nowadays to seek proper identification of users. After all this, cloud storage is still an open platform for online security breaches where information like bank account number, bills and other data can be easily compromised. Creating tools which can reduce such breaches is still a challenge.
3. Back-up and disaster recovery: Large scale catastrophic loss caused by events such as server failure etc. can lead to interference or complete loss of the user's personal data and considering this on large scale, a very high number of users can get affected in absence of backup data centres. Maintaining back up for disaster recovery is something which many cloud providers don't prefer. This puts a risk of loss or interference of user data.
4. Malicious insider: Any malicious user can easily implant his malicious code in other user's VM thereby providing him with access to data of other users. In cloud, virtualized infrastructure is used which makes this process more comforting for malicious insiders.
5. Some more issues like control over data, interoperability, increase in performance, decrease in cost, anywhere access etc. are still open for research and future developments.

VI. Conclusion

From the time of evolution of cloud storage, it is designed in such a way so as to deliver functionalities like high scalability, low cost, easy management etc. Cloud storage does not merely focus on delivering high performance output. The performance and other factors of cloud storage depend largely on the infrastructure used and it lays the foundation for any type of storage. Our paper focuses on the back end cloud storage architectures (when it comes to cloud storage considering networks). It covers the fundamentals concepts related to cloud computing and their relation with cloud storage, after which SNIA reference model is discussed. Followed by this we have given a conceptual overview of SAN, NAS and DAS with concept implementation. The paper concludes by pinpointing the future research challenges.

Acknowledgment

The author would love to acknowledge our professors and friends for imparting helpful comments. The standard disclaimer applies.

References

[1]. EMC, and EMC Education Services. Information Storage and Management: Storing, Managing, and Protecting Digital Information. LibreDigital, 2010.
[2]. Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., ...&Zaharia, M. (2010). A view of cloud computing. Communications of the ACM,53(4), 50-58.
[3]. Chunhua, ZHOU Ke WANG Hua LI. "Cloud Storage Technology and Its Application [J]." ZTE Communications 4 (2010): 013.
[4]. Rimal, B. P., Choi, E., &Lumb, I. (2009, August). A taxonomy and survey of cloud computing systems. In INC, IMS and IDC, 2009. NCM'09. Fifth International Joint Conference on (pp. 44-51). Ieee.
[5]. Wu, Jiyi, et al. "Recent Advances in Cloud Storage." Proceedings of the Third International Symposium on Computer Science and Computational Technology (ISCSCT'10). 2010.
[6]. Meyer, Dutch T., et al. "Fast and cautious evolution of cloud storage."Proceedings of the 2nd USENIX conference on Hot topics in storage and file systems. USENIX Association, 2010.
[7]. Jadeja, Yashpalsinh, and KiritModi. "Cloud computing-concepts, architecture and challenges." Computing, Electronics and Electrical Technologies (ICCEET), 2012 International Conference on. IEEE, 2012.
[8]. ZHANG, Hu, and Ming-dong LI. "Cloud Storage Technology and Its Applications." Journal of Yibin University 12 (2012): 022.
[9]. Goda, Kazuo. "Direct Attached Storage." Encyclopedia of Database Systems. Springer US, 2009. 847-847.
[10]. Gibson, Garth A., and Rodney Van Meter. "Network attached storage architecture." Communications of the ACM 43.11 (2000): 37-45.
[11]. Clifford, Mark, Norm Miles, and Bruce R. Rabe. "Storage area network (SAN) management system for discovering SAN components using a SAN management server." U.S. Patent No. 7,194,538. 20 Mar. 2007.
[12]. Zeng, Wenying, et al. "Research on cloud storage architecture and key technologies." Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human. ACM, 2009.