ISSN:0975-9646

A Review on Cloud Data Storage in Virtual Perspective

B.Nagarajan[#1], Dr.J.Suguna[*2]

[#]*Ph.D., Research Scholar, Research and Development Centre,*
Bharathiar University,
Coimbatore-621046, India.
[*]*Associate Professor, Department of Computer Science,*
Vellalar College for Women, Erode, India

***Abstract*— CLOUD can be expanded as Common Location independent On-line Utility that is available on Demand. This review starts with study on cloud computing in virtual perspective. The cloud has many "on-demand" service models, of which, cloud storage is a model which is amorphous today and with neither a clearly defined set of capabilities nor any single architecture. This paper carefully studies on taxonomy of storage virtualization which forms the basis for cloud storage. The review on, CDMI and OCCI interface and their integration in cloud storage referential model helps researcher to implement Information Life-cycle Management (ILM) through tiered storage to store the backups. The cloud data can be retrieved through n-tier and those tiers are useful to allocate future storage requirement. This review concludes with challenges to be eschewed in cloud storage. To sum-up, this review will be a compendium for the researcher, to understand cloud data storage in virtual perspective.**

***Keywords*— Storage Virtualization, Cloud Data Management Interface (CDMI), Open Cloud Computing Interface (OCCI), Storage as a Service (StaaS).**

I. Introduction

Cloud computing is envisioned as next generation technology in IT. Instead of running program and data in an individual desktop computer, everything is hosted in the “CLOUD” a nebulous assemblage of computers and servers via Internet, where quality of service is provided on subscription basis.

Cloud storage is an "on-demand" service model in which data is maintained, managed and backed up remotely and made available to users over a network. It is a subscription based model. The cloud data can be stored and manipulated via a concept known as tiered storage [5].

The rest of this paper is organized as follows, Section 2 and its associate sections discusses about core of this review such as cloud computing and its "on demand" storage as a service model. Section 3 elaborates storage virtualization. Section 4 elucidates cloud storage referential model. Section 5 highlights challenges in cloud storage.

II. Key Technologies

A. *Cloud Computing*

National Institute of Standard Technology (NIST) [6] defines CLOUD as a model for enabling convenient, on demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction.

Cloud computing is about moving services, computation or data for cost and business advantage off-site to an internal or external, location-transparent, centralized facility or contractor. By making data available in the cloud, it can be more easily and ubiquitously accessed, often at much lower cost, increasing its value by enabling opportunities for enhanced collaboration, integration and analysis on a shared common virtual platform.

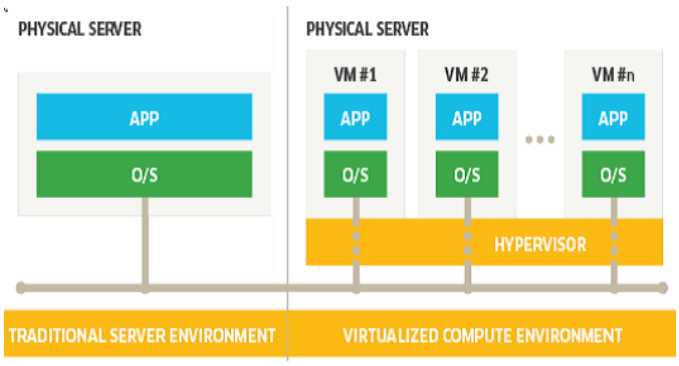

Fig. 1: Virtual Compute Environment

In the virtual cloud environment, each operating system appears to have the host's processor, memory and other resources all to itself. However, the hyper-visor is actually controlling the host processor and resources, allocating what are needed to each operating system in turn and making sure that the guest operating system (called virtual machines) cannot disrupt each other [11].

Cloud has five essential characters such as on demand self-service, broad network access, resource pooling, rapid elasticity and measured services and four deployment models such as public, private, community and hybrid [6]. Cloud has many "on-demand" service models on subscription basis over Internet. The agile feature of cloud computing has clinched Internet into an new age of “Everything as a Service” of which, StaaS is one such model which follows a storage virtualization in a "silo" approach to store and retrieve data from the remote tiers.

B. *STORAGE AS A SERVICE (STAAS)*

StaaS is a business model in which large company rents space in their storage infrastructure to a smaller company or individual.

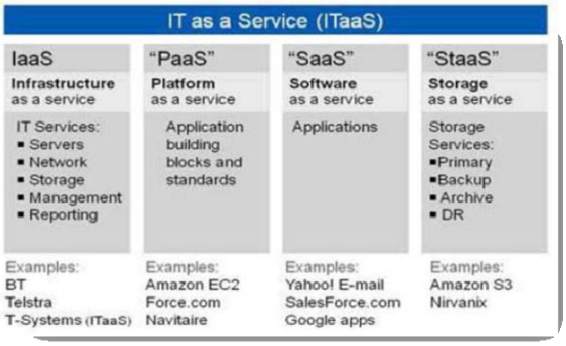

Fig. 2 : Cloud Services Redefined

Any individual or companies would sign a Service Level Agreement(SLA) whereby the StaaS provider agree to rent storage space on pay-per-use basis and then data would be automatically transferred at the specified time over the storage provider's proprietary wide network or through Internet. Popular Cloud Service Provider (CSP) providing StaaS include Google drive, Microsoft One Drive, Drop Box etc. They are targeting secondary storage applications by promoting Information Life-cycle Management (ILM) as a convenient way to manage backups [4].

ILM is a proactive strategy that enables an IT organization to effectively manage the data throughout its life cycle, based on predefined business policies [4]. This allows an IT organization to optimize the storage infrastructure for a maximum return on investment.

ILM products automate the processes involved in StaaS, typically organizing data into separate tiers according to specified policies and automating data migration from one tier to another based on defined criteria. *Tiered Storage is the assignment of different categories of data to different types of storage media in order to reduce total storage cost.* Categories may be based on levels of protection needed, performance requirements, frequency of use, and other considerations. A path management application, either as a component of ILM software or working in conjunction with it, makes it possible to retrieve any data stored by keeping track of storage cycle [5].

Automated Storage Tiering (AST) is a storage software management feature that dynamically moves information between different disk types and Redundant Array of Independent Disk (RAID) levels to meet space, performance and cost requirements. AST use policies that are set up by storage administrators [1].

StaaS is generally seen as a good alternative for a small or mid-sized business or an individual who lacks the capital budget and/or technical personnel to implement and maintain their own storage infrastructure.

III. STORAGE VIRTUALIZATION

Storage virtualization presents a logical view of the physical storage resources to host. This logical storage appears and behaves as physical storage directly connected to the host.

The Storage Networking Industry Association (SNIA) published a three level taxonomy that provides storage virtualization [2], [7] , [10].

The first level of storage virtualization should be in either Block level or in File level. Block level virtualization extends storage volumes on-line, resolving application growth requirement, consolidating heterogeneous storage array and enabling transparent volume access. It also provides the advantage of non-disruptive data migration through Storage Area Network (SAN). SAN carries data between servers (also called host) and storage devices through fibre channel switches. SAN enables storage consolidation and allow to share across multiple servers. It enables organization to connect geographically dispersed servers and storage. The two protocols that extend block-level access to application over Internet Protocol (IP) are namely internet Small Computer System Interface protocol (iSCSI) and Fibre Channel over IP (FCIP) [4].

iSCSI is an IP-based protocol that establishes and manages connection between storage, host and bridging. The data block is transported using TCP/IP. It has enabled IT organization to gain benefit of storage networking architecture at reasonable cost.

FCIP is a tunnelling protocol that enables distributed FC SAN islands to be transparently interconnected over existing IP-based local, metropolitan and wide-area networks. As a result, organizations now have a better way to protect, store and move their data while leveraging investments in existing technology. FCIP uses TCP/IP as its underlying protocol. In FCIP, the FC frames are encapsulated onto the IP payload. It does not manipulate FC frames (translating FC IDs for transmission). When SAN islands are connected using FCIP, each interconnection is called an FCIP link. A successful FCIP link between two SAN islands results in a fully merged FC fabric.

File-level virtualization addresses the Network Access Storage (NAS) challenges. It eliminates the dependencies between the data accessed at the file level and location, where the files are physically stored. It simplifies the file mobility. It provides user or application independence from location where the files are stored. It creates logical pool of storage, enabling users to use logical path, rather than a physical path to access files, thus enabling data sharing. NAS protocol enables CSPs to network together thousands of hard drives and manage them in one central server to provide the storage service. NAS enables CSPs to share data among users.

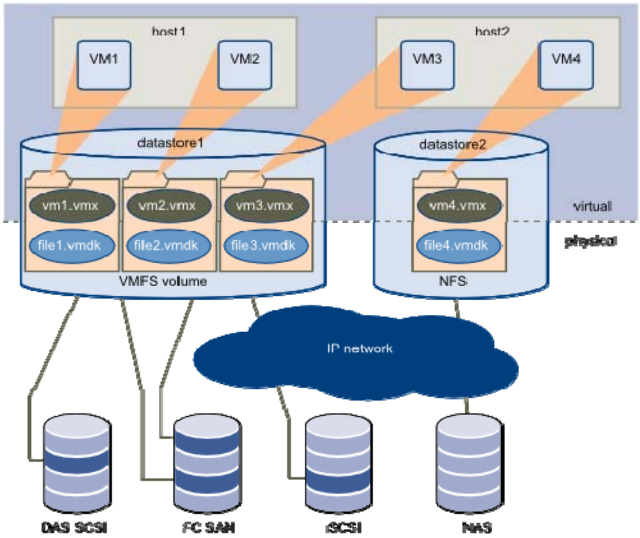

Figure 3: Storage Virtualization

In second level, it should be implemented in server, storage network and storage. The server virtualization includes path management, volume management and replication. In storage network virtualization, path redirection, access control and load balancing are managed. In storage virtualization, volume management with Logical Unit Number (LUN), replication and RAID are managed. LUN help in disk utilization whereas RAID is an enabling technology that leverages multiple disks as part of a set, which provides data protection against Hard Disk Drive (HDD) failures. In general, RAID implementation also improves the I/O performance of storage system by storing data across multiple HDDs.

In third level, the storage virtualization can be implemented either as In-band or out-of-band. In In-band virtualization, data-path is within virtualization function. It is useful for static application. In out-of-band virtualization, data path is external and is useful for virtual storage applications such as tiered storage [4], [5].

The key benefit of storage virtualization is to increase storage utilization through tiers, adding or deleting storage without affecting the application's availability and non-disruptive data migration.

IV. Cloud Storage Referential Model

SNIA is based on Cloud Data Management Interface(CDMI) which has four standards such as cloud storage subscriber(users), cloud storage service provider, cloud storage service developer and cloud storage service broker [2], [10]. These standards provides functional interface,

Used to create, retrieve, update and delete data elements from the cloud. As part of these interfaces, the client will be able to manage container and the data that is placed in these container enables inter-operable cloud storage and data management.

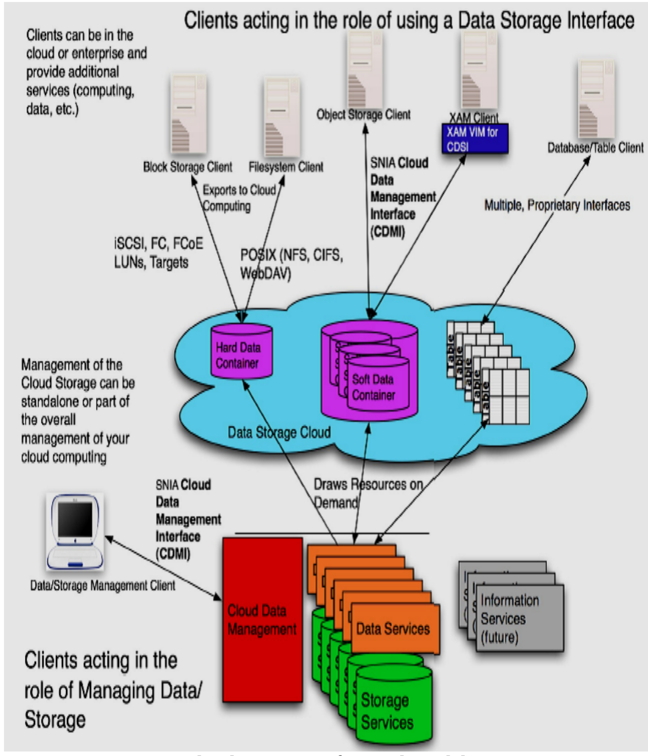

Figure 4: Cloud Storage Referential Model

A container is not only a useful for abstraction of storage space, but also serves as a grouping of the virtual data stored in it and a point of control for applying data services in the aggregate. CDMI can be used to manage container exported for use by cloud computing infrastructures [2], [7]. CDMI containers are accessible not only via CDMI as a data path but also through other protocols such as OCCI.

OCCI is a free, open, community consensus driven API, targeting cloud infrastructure services [2]. With OCCI, cloud computing clients can invoke a new application stack, manage its life cycle and also manages the resource that it uses. The OCCI interface can also be used to assign storage to a virtual machine in order to run the application stack such as that exported by SNIA's CDMI interface. Figure well explains CDMI and OCCI in an integrated cloud computing environment. The exported CDMI containers can be used by the Virtual Machines in the cloud computing environment as virtual disks on each guest.

With the internal knowledge of the network and virtual machine, Cloud Infrastructure Management Application (CIMA) can attach exported CDMI containers to the virtual machines. CDMI provides a type of export that contains information obtained via the OCCI interface. In addition, OCCI provides a type of storage that corresponds to exported CDMI containers. OCCI and CDMI can achieve interoperability initiating storage export configuration from either OCCI or CDMI interfaces as starting points. Although the outcome is the same, there are differences between the procedures using CDMI's interface over the OCCI's as a starting point.

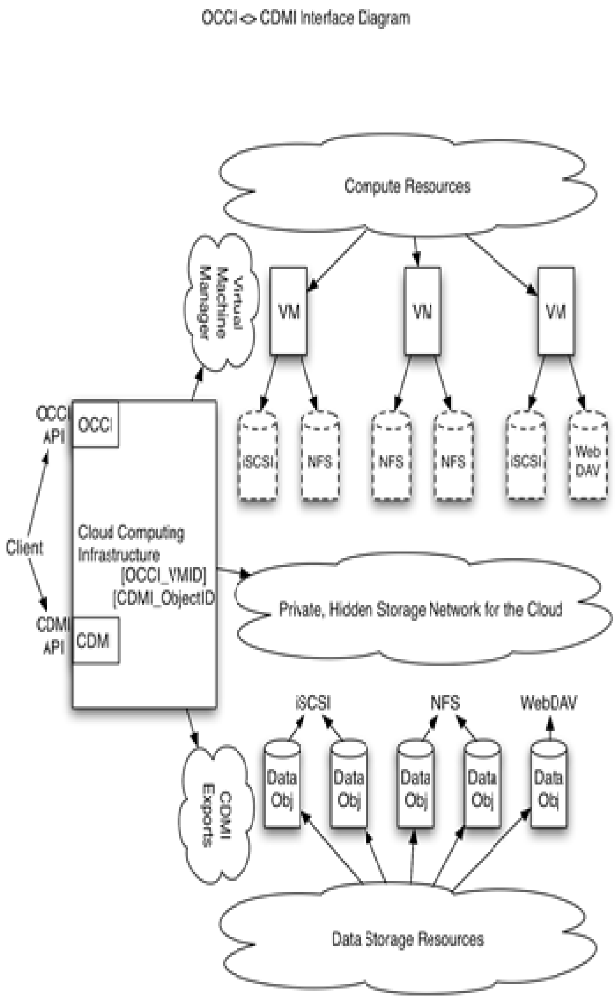

Figure 5: OCCI – CDMI Interface Diagram

The interoperability between CDMI and OCCI starts with client creating a CDMI container through the CDMI interface and export it as an OCCI export type. The CDMI container object ID is returned as a result. The client then creates a virtual machine through the OCCI interface and attaches a storage volume of type CDMI using the Object ID. The OCCI virtual machine ID is returned as a result. The client then updates the CDMI container object export information with the OCCI virtual machine ID to allow virtual machine access to the container. The client then starts the virtual machine through the OCCI interface [2]. This interoperability helps to achieve StaaS using ILM in tiered storage.

V. CHALLENGES IN CLOUD STORAGES

The challenges in cloud storage includes storage up-gradation cost, data security, path management with Cloud, multi-site data sharing, disaster recovery and handling storage growth. Of these challenges, researchers' are working especially on disaster recovery and handing storage growth [1].

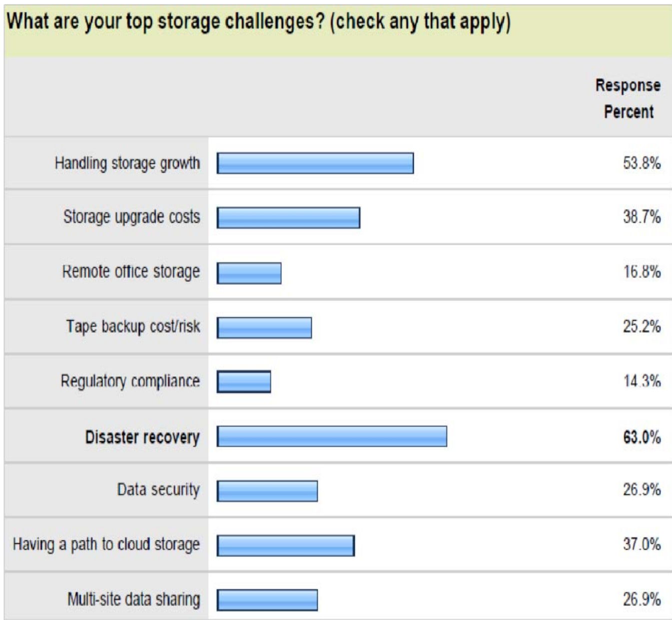

Figure 6: Challenges in Cloud Storages

A. *DISASTER RECOVERY*

This is the foremost challenge in cloud storage. It is the replication and hosting of physical or virtual servers by a third-party to provide fail-over in the event of a man-made or natural catastrophe.

Generally disaster can be caused by various factors such as hardware failure, natural disaster, closure of cloud services, cloud related malware, inadequate infrastructure design and planning and others. Data can be lost partially or fully due to these disaster which has be recovered. For such reasons, a suitable Disaster Recovery Plan (DSP) can be especially useful for small to mid-size business or individual that lacks the necessary expertise to invest in and maintain their own off-site Disaster Recovery environment. It should be flexible as the business need changes. Researcher has scope to develop a suitable Disaster Recovery as a Service (DRaaS).

B. *HANDLING STORAGE GROWTH*

Data storage administrator will assess and forecast future requirements, so that, planning and implementation of storage growth would ascertain organization to save money and budget on storage. Factors to consider include tier(s) of data storage. Automated Storage Tiering (AST) allows less frequently accessed items to be moved to a lower (and less costly) tier, freeing faster (and more expensive) tiers for those applications that need it. Techniques like data de-duplication and file compression can help reduce the amount of storage capacity needed by eliminating extra copies of the same file and reducing the amount of space needed to store files and images. Researcher has sufficient scope to develop policies and programs for AST to handle future storage growth.

VI. Conclusion And Future Work

The key technologies reviewed in this paper provides in-depth knowledge to understand cloud computing and its "on-demand" cloud service models. The CDMI Integration with OCCI in virtual perspective helps to build a reliable storage virtualization. StaaS discussed in this review will be a compendium to understand the underlying concept of cloud storage. The challenges discussed at the end of this review provide direction and scope of research in cloud storage. The future work include careful study on storage virtualization and Cloud Storage Referential model, through which a reliable and secured n-tiered storage is achieved for StaaS, which will benefit all walks of people.

References

[1] Buyya, J.Broberg and A.Goscinski, "Cloud Computing Principles and Paradigms," John Wiley and Sons, 2011.
[2] SNIA Manual for Cloud Storage, the Storage Networking Industry Association, SNIA 2011.
[3] A white paper by Andaman, EMA Senior Analyst, "Virtualization 101: Technologies, Benefits and Challenges", August 2006.
.
[4] G.Somasundaram and A.Srivastava, "Information storage and Management – Storage, Managing and Protecting Digital Information, Wiley and Sons 2009.
[5] A.D.Luca and M.Bhide," Storage Virtualization for Dummies", John Wiley and Sons 2010.
[6] P.Mell and T.Grance, "National Institute of Standards and Technologies (NIST) for Cloud Computing" 2011.
[7] F.Bunn, N.Simpson and R.Peglar "SNIA Technical Tutorial for Storage Virtualization ", March 2013.
[8] J.Y.Luo, "Secure Cloud Storage ", a thesis submitted to University of Waterloo – Canada 2014.
[9] J.Wu and L.Ping , "Cloud Storage as an Infrastructure of Cloud Computing" , 2010 International conference of Intelligent Computing and Cognitive Informatics , pp. 380-383 IEEE November 2010.
[10] G.Kulkarani, V.Waykule and H.Bankar, "Cloud Storage Architecture", 7th International Conference on Telecommunication, Systems, Services and Application, pp.76-81, IEEE 2012.
[11] S.M.Bansode, "Mitigating Cloud Virtualization Vulnerabilities", International Journal of Recent Advances in Engineering and Technology, ISSN (online) 2347-2812, Vol.1, Issue. 3, pp.9-11, 2013