IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

INTRODUCTION TO CLOUD STORAGE AND INFORMATION MANAGEMENT IN VIRTUALIZED ENVIRONMENT AND DISASTER RECOVERY CONSIDERATIONS

K. S. Patel[1], Dr. Amol B. kasture[2]

[1] Research Scholar - PhD, JJT University, Jhunjhunu, Rajasthan (India)

[2]Research Guide, JJT University, Jhunjhunu, Rajasthan (India)

ABSTRACT

This paper provides the analyses of the cloud computing environment along with the storage visualization used in the real time cloud environment. The cloud types are explained in this paper which deals with tier pros and cons respectively and oriented use cases for client to store their data on cloud. The paper also explains the clouds storage infrastructure service for the client usage as storage as a service to help selection the specific type of cloud storage based on the requirements. The paper explains disaster recovery mechanism and three levels of disaster recovery mechanism for clod storage to continue business applications in case of disaster recovery in the cloud environment. The combination of private cloud and inter-private cloud makes the system more robust in case of disaster recovery for cloud data storage and transparent recovery can be achieved.

Keywords: Keywords- Cloud storage, disaster recovery, reference model, inter-private cloud

I. INTRODUCTION

Cloud computing is a model where content and computation enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. Leveraging the current available work, this research synopsis is made to enhance the technology and to which can be served to computer engineering future.

A. Cloud computing:

"Cloud Computing refers to manipulating, configuring, and accessing the applications online. It offers online data storage, infrastructure and application." Cloud is a distinct Information technology environment that is designed for the purpose of remotely provisioning scalable and measured resources. This is a network of networks providing remote access to a set of decentralized IT resources. Much of the Internet is dedicated to the access of content-based IT resources published using the World Wide Web. IT resources provided by cloud environments are dedicated to supplying backend processing capabilities and user-based access to these capabilities. A cloud can be based on the use of any protocols that allow for the remote access to its IT resources. Protocols used by cloud refer to standards and methods which allow computers to communicate with each other in a pre-defined and structured manner.

Types of cloud

Cloud technology enables the sharing of resources in a way that dramatically simplifies infrastructure planning. With cloud computing technology, large pools of resources can be connected via private or public networks to provide dynamically scalable infrastructures for application, data and file storage. Additionally, the costs of computing, application hosting, content storage and delivery can be significantly reduced. Firms can choose to deploy applications on Public, Private or Hybrid clouds.

a. Public Clouds

Public clouds are owned and operated by third-party service providers. Benefit to the costumer from economies of scale because infrastructure costs are spread across all users, thus allowing each individual client to operate on a low-cost, “pay-as-you-go” model. Another advantage of public cloud infrastructures is that they are typically larger in scale than an in-house enterprise cloud, which provides clients with seamless, on-demand scalability. All customers on public clouds share the same infrastructure pool with limited configurations, security protections and availability variances, as these factors are wholly managed and supported by the service provider.

b. Private Cloud

Private clouds are those that are built exclusively for an individual enterprise. They allow the firm to host applications in the cloud, while addressing concerns regarding data security and control, which is often lacking in a public cloud environment. There are two variations of private clouds:

c. Hybrid Cloud

Hybrid clouds combine the advantages of both the public and private cloud models. In a hybrid cloud, a company can leverage third-party cloud providers in either a full or partial manner. This increases the flexibility of computing. The hybrid cloud environment is also capable of providing on-demand, externally-provisioned scalability. Augmenting a traditional private cloud with the resources of a public cloud can be used to manage any unexpected surges in workload.

B. Virtualization

In computing, virtualization refers to the act of creating a virtual (rather than actual) version of something, including virtual computer hardware platforms, operating systems, storage devices, and computer network resources. Network virtualization is a method of combining the available resources in a network by splitting up the available bandwidth into channels, each of which is independent from the others, and each of which can be assigned (or reassigned) to a particular server or device in real time.

The idea is that virtualization disguises the true complexity of the network by separating it into manageable parts, much like your partitioned hard drive makes it easier to manage your files. Storage virtualization is the pooling of physical storage from multiple network storage devices into what appears to be a single storage device that is managed from a central console. Storage virtualization is commonly used in storage area networks (SANs). Server virtualization is the masking of server resources (including the number and identity of individual physical servers, processors, and operating systems) from server users. The intention is to spare the user from having to understand and manage complicated details of server resources while increasing resource sharing and utilization and maintaining the capacity to expand later.

C. Disaster Recovery

Disaster recovery (DR) involves a set of policies and procedures to enable the recovery or continuation of vital technology infrastructure and systems following a natural or human-induced disaster. Disaster recovery focuses on the IT or technology systems supporting critical business functions, as opposed to business continuity, which involves keeping all essential aspects of a business functioning despite significant disruptive events. Disaster recovery is therefore a subset of business continuity.

A disaster recovery plan (DRP) documents policies, procedures and actions to limit the disruption to an organization in the wake of a disaster. Just as a disaster is an event that makes the continuation of normal functions impossible, a disaster recovery plan consists of actions intended to minimize the negative effects of a disaster and allow the organization to maintain or quickly resume mission-critical functions. In information technology, disaster recovery steps may include restoring servers or mainframes with backups, re-establishing private branch exchanges (PBX) or provisioning local area networks (LANs) to meet immediate business needs. Business continuity describes the processes and procedures an organization must put in place to ensure that mission-critical business functions can continue during and after a disaster. The emphasis is more on maintaining business operations than IT infrastructure. Because business continuity and disaster recovery are so closely related, the two terms are sometimes combined as Business Continuity and Disaster Recovery.

II. CLOUD COMPUTING INFRASTRUCTURE SERVICE

There are two types of cloud as states in previous sections of this paper which are Private and Public clouds. These serve as the backbone for a different cloud computing service models in the current world. Currently the large scale and medium scale companies has been successfully adopting three common types of cloud computing service types. Infrastructure as a Service (IaaS), which means a service model around servers, capacity having large storage, and bandwidth for high speed network. Some of the examples of this mechanism includes Amazon EC2 and S3, etc. Platform-as-a-Service (PaaS) provides a managed platform which is eternally managed for deployment of the applications and services on top of it. This model typically provides tools for development such as databases and development studios for working with the APIs provided, as well as the infrastructure to host the built application and run them in real life world. it's examples may include Force.com, Microsoft Azure, etc. Software-as-a-Service (SaaS) is simply having a software system running on a computer that doesn't belong to the customer and isn't on the customer’s location. It is generally based on the concept of renting an application from a service provider rather than buying and installing and running software by the user themselves.

III. CLOUD STORAGE REFERENCE MODEL

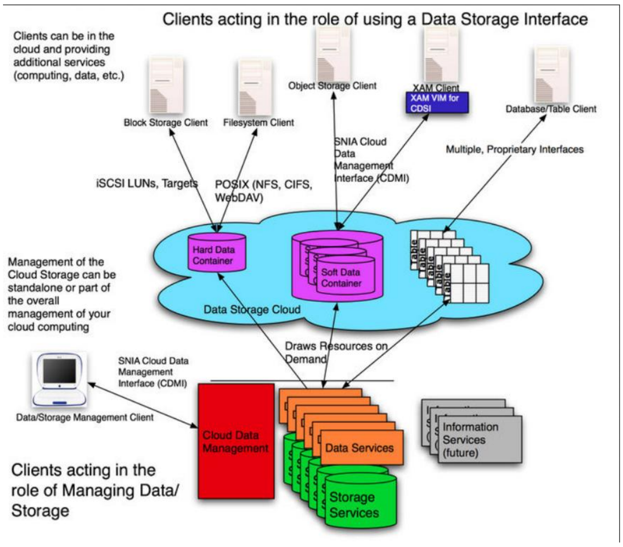
There are many facilities that the cloud storage provides are giving, some of them can be pay as you go, illusion of literally infinite capacity, and the simplicity of usage and management ease. It is important that any interface for cloud storage support these attributes, and at the time of allowing for a multitude of real world application cases and offerings, long into the future too. This model was originally brought by Storage Networking Industry Association (SNIA) which shows multiple types of cloud data storage interfaces supporting legacy and new applications for deployment and management. All of the interfaces allows storages to be provided to the users on demand, drawn from a pool of resources present at the storage layer of cloud. The data services can be applied to individual elements of data in the cloud and can be determined by the data system metadata which specifies the data requirements on the basis of individual elements or elements groups which are also be called as containers.

Cloud Data Management Interface (CDMI) is the interface that deals with functional aspects of the system and applications will use to create, update, retrieve, and delete elements of data from the cloud storage. Due to this interface the users will be able to view the capabilities of the cloud storage and can use the interface to manage groups of elements and the data placed there. Metadata can be written on containers and their data elements via CDMI interface. The interface will possibly be able to be implemented by the majority of existing cloud storage offerings and can be done with an adapter to the proprietary interface architecture, or direct implementation. The client libraries like eXtensible Access Method (XAM) can be used to CDMI interface

This interface can also be used by management and administrative of applications to manage accounts, containers, access security, information monitoring and even for storage that is accessible by other available protocols. The underlying capabilities and data services are exposed so that clients can understand by the upper layer of cloud. Conformant cloud offerings can be offered as a subset of either interface if they expose the limitations in the capabilities of the CDMI interface.

The cloud storage is not emerged branch of computer science field and hence there are lot many improvements are needed and have certain issues in todays implementation aspects which are described in this paragraph below. It gets the attention of IT people with its low cost comparatively and ability to adjust easily for capacity. The cloud storage offers reduction in the investment cost in total, but people has to suffer some of the technical, security, and integration, and organizational issues while dealing with it.

The data which is residing outside the normal enterprise's infrastructure, it is perceived that the it may loss the control over data if the cloud misbehaves. Although these concerns are hypothetical in nature and psychological rather than actual because of the features provided by cloud providers, due to some immaturity factors of cloud services, standards on the delivery of services and their evolving business model, and clients may have genuine concerns about the CSPs operational processes and availability factors. The complexity of cloud storage usage is something many people underestimate. Each vendor has different access methods, APIs to use which are generally implemented on non-standard basis that make integrating applications more difficult and costly if the application is big enough


Clients acting in the role of using a Data Storage Interface
Clients can be in the cloud and providing additional services (computing, data, etc.)
Block Storage Client
Filesystem Client
Object Storage Client
XAM Client
XAM VIM for CDSI
Database/Table Client
iSCSI LUNs, Targets
POSIX (NFS, CIFS, WebDAV)
SNIA Cloud Data Management Interface (CDMI)
Multiple, Proprietary Interfaces
Hard Data Container
Soft Data Container
Table
Data Storage Cloud
Management of the Cloud Storage can be standalone or part of the overall management of your cloud computing
Draws Resources on Demand
SNIA Cloud Data Management Interface (CDMI)
Data/Storage Management Client
Cloud Data Management
Data Services
Storage Services
Information Services (future)
Clients acting in the role of Managing Data/ Storage

.

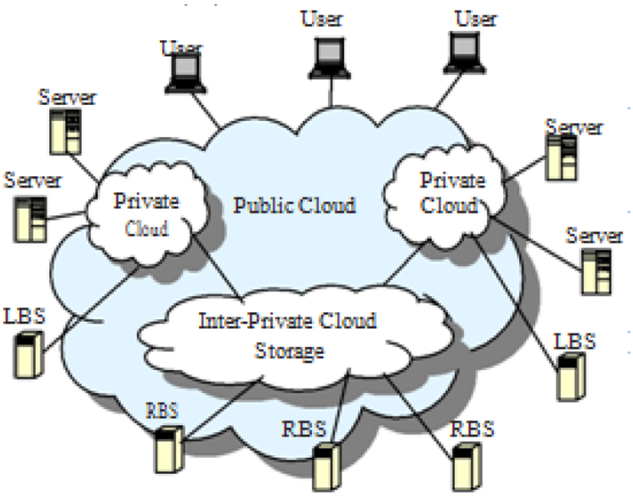
Few vendors provides software clients which implement common network file sharing protocols like NFS or CFIS but these are generally proprietary and cannot bridge between different cloud services across the heterogenous CSPs. The lack of standard protocols for accessing cloud storage means there is no interoperability between cloud storage providers, greatly complicating the data migration process which affects the online migration of the data. Another importance point to notice is access to cloud data is obviously limited by throughput of network and latency, and can also be dependent on Internet performance which is still poor in comparison to local network storage. Some vendors tries to enhance throughput with local caching and compression mechanisms, which don't improve latency of internet.

Another bigger issue can be data security with cloud storage for which most of the clients are worried about. If any possibility leakage, both in transfer and within a shared infrastructure, the experts agrees that encryption on all data stored can help in more secure data storage for the application. The archiving kind of data works good in the cloud because the data changes less frequently hence less computation will be needed. These data don't require high speed transactional access. Data in chunks can be easily compressed using data reduction technologies or it can be easily encrypted due to block based mechanisms. The applications with low I/O performance and tolerance for low downtime are suitable residents in cloud storage.

IV. DISASTER RECOVERY OF CLOUD STORAGE

The applications and data services are transparent in the cloud models where malicious attacks must be avoided. It is important and difficult to ensure data protection and security in cloud storage system. The necessity condition to establish fault tolerant function in less cost for cloud storage. Job of the fault tolerant function is to overcome single point failure which inherently avoids data loss across the cloud sites. At the same time, there must be fault-tolerant backup system present for the application data which can ensure that data consistency having a reliable backup if they are lost due to hazards or any disaster scenario.

Satisfaction of the continuity of application and the security of data, the structure of recovery mechanism is "distributed computing having centralized storage". Different requirements suggest that the disaster recovery has three levels viz. data-level disaster recovery, system-level disaster recovery, and application-level disaster recovery. First level, i.e. Data-level disaster recovery is the most basic type of recovery and it can ensure the security of the application data of user. System-level disaster recovery disaster has further requests for OS of application server, making DR time can be as short as possible to get uninterrupted application work. System-level disaster recovery requests real-time by which client could not feel that any disaster has occurred due to its less time. Most important feature of this system is to use the SAN - Storage Area Network to support access of the application data and backup. It is constituted by a number of high-performance routing devices having strong robustness. SAN provide a redundant fabric links for the application server to access the server and to backup data which can be used at the time of recovery. SAN as a large, stable data transmission network which requires a number of high-performance routing equipment to implement and the costly hardware.



As the SAN only provides application data access and backup, to achieve the response process cost only very little time, it is need to design their own associated routing protocols. It is need to use all of the authorized hardware and software devices to operate and all the enterprise data are stored. According to standards by SNIA for the network storage industry, Disaster recovery must be equipped with at least three different geographical locations which is also known as three way disaster recovery. If an enterprise private cloud storage is equipped with all of the DR and redundant backup systems then private cloud should not be seen as a cloud in the tight sense, because it is not thorough in resource rationalization and has the characters of heavy local configuration, lack of flexibility , high cost and other typical properties of non-cloud computing. Reasonably and more thoroughly approach is to private clouds to share a common cloud storage services for users. This shared public cloud storage services will be placed in the private clouds and it can be known as "inter-private cloud storage". Not only offerings for enterprise private cloud are provided with specialized storage services of disaster redundant backup but also for the cloud users with efficient and convenient mobile service.

All servers have some backup servers to store second copy of data. The backup servers and master server can be distributed in different sites. The backup server may include the local backup server and the remote backup server. Once the data are stored in LBS, data integration can be down in a leisure time including redundant data delete, disk compression and finishing work. Remote backup stage can be limited by the network connection ability of the cloud then upload of GB, PB magnitude of data across the enterprises and is a great challenge to store this much data on cloud. Due to this, in the initial remote backup, the mechanism of physical transfer can be used and the temporary line from cloud storage node for computation to business clients can also be provided to move a large amount of data into it. After the initial backup the data can backup daily via incremental approach which is used to reduce pressure on network bandwidth and improve speed of backup and replication. Reduction of the traffic is also an important means to enhance the service experience in the cloud environment. In the case where a dedicated client is used compression and encryption can be done before the data backup to replace the SSL based encryption during data transmission across the sites and reduce transmission costs as less bits are transferred. Backup service can not only support the application integration, but they also have ability to satisfy the requirements of concurrent, magnanimity and common application interface.

In the hardware architecture, inter-private cloud storage does not require high-speed data processing in real time, so it is not need very high-end storage technologies (such as the Fibre Channel attached disk array), and the Iscsi channel disk array is enough. However, in software architecture, the inter-private cloud storage requires a high intelligence, policy-based SaaS service standards and information management functions. In particular, it is important to achieve a strong separation and protection technology of security data.

REFERENCES

[1] K. Keeton, D. Beyer, E. Brau, A. Merchant. "On the Road to Recovery: Restoring Data after Disaster". Proceedings of the 1st ACM SIGOPS/EuroSys 06, European Conference on Computer Systems, ISBN:1-59593-322-0, 2006.

[2] R. Chow, P. Golle, M. Jakobsoon "Controlling data in cloud: outsourcing computation without outsourcing control" Fujitsu Laboratories of America ISBN: 978-1-60558-784-4 doi>10.1145/1655008.1655020

[3] QU Ming-cheng, WU Xiang-hu,; LIAO Ming-hong, et al. "A Disaster-Tolerant Storage Model and a Low Data Failure Model for Data Grid". Acta Electronica Sinica. 2010.38(2),pp 315-320.

[4] T. Wood, E. Cecchet, K.K. Ramakrishna "Disaster Recovery as a cloud service: Economic Benefits &deployment challenges" university of Massachusetts AT&T LABS – 2nd USENIX conference on Hot topics in cloud computing Boston Year of Publication: 2010

[5] G. Zhang, L. Chiu, Ling Liu "Adoptive Data Migration in Multi-tiered Storage Based Cloud Environment" CLOUD '10 IEEE 3rd International Conference in Cloud Computing ISBN: 978-0-769541303 doi>10.1109/ CLOUD. 2010.60; Year-2010.

[6] ARMBRUST M, FOX A, GRIFFITH R, et al. "Above the Clouds: A Berkeley View of Cloud Computing" [R]. Berkeley, CA, USA: University of California, 2009.

[7] Dai Yuanshun. "The Brief Review of Cloud Computing Technologies". Information and Communications Technologies. 2010.2,pp 29-35.

[8] G. Zang, L. Chiu "Adaptive Data Migration in Multi-tiered storage based cloud envoirnment"Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference On page(s): 148 - 155 Print ISBN: 978-1-4244-8207-8, Date: 5-10 July 2010

[9] ZHOU Ke, WANG Hua, and LI Chunhua. "Cloud Storage Technology and Its Application", ZTE Communications, 2010.16(4),pp 24-27.

[10] K. Keeton, C. Santos, D. Beyer, J. Chase, and J. Wilkes. Designing for Disasters. Conference On File And Storage Technologies,2004.

[11] W. Hoe, I - Len Yen "Dynamic service and data migration in clouds" Computer Software and Applications Conference, 2009. COMPSAC '09. 33rd Annual IEEE International ISBN: 978-0-7695-3726-9 doi> 10.1109/COMPSAC.2009.127