International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 11, November 2015
Available at http://internationaljournalofresearch.org

A Study of Data Storage and Management in Cloud Computing

Manoj Kumar Dhruw
Student, Bachelor of Engineering in Computer Science & Engineering Kirodimal Institute of Technology, Raigarh (C.G.), India
dhruwmanoj2010@gmail.com
Amarjit Ram
Student, Bachelor of Engineering in Computer Science & Engineering Kirodimal Institute of Technology, Raigarh (C.G.), India
amarjitram772@gmail.com
Rakesh Patel
Lecturer, Department of Information Technology Kirodimal Institute of Technology, Raigarh (C.G.), India
rakeshpatel.kit@gmail.com

Abstract

The Data Storage is a service of Cloud Computing which is provided by Cloud Computing provider. It allows cloud user to store data in any Cloud using Internet. In this paper we have discussed about Data Storage, challenges and about the Cloud Computing Management.

I. INTRODUCTION

A. What Is Cloud Computing ?

Cloud computing (‘cloud’) is an evolving term that describes the development of many existing technologies and approaches to computing into something different. Cloud separates application and information resources from the underlying infrastructure, and the mechanisms used to deliver them. Cloud enhances collaboration, agility, scaling, and availability, and provides the potential for cost reduction through optimized and efficient computing. More specifically, cloud describes the use of a collection of services, applications, information, and infrastructure comprised of pools of compute, network, information, and storage resources. These components can be rapidly orchestrated, provisioned, implemented and decommissioned, and scaled up or down; providing for an on-demand utility-like model of allocation and consumption. From an architectural perspective; there is much confusion surrounding how cloud is both similar to and different from existing models of computing; and how these similarities and differences impact the organizational, operational, and technological approaches to network and information security practices. [1]

II. CLOUD COMPUTING DATA STORAGE

A. What is Data Storage?

Cloud Storage is a service that allows to save data on offsite storage system managed by third party and is made accessible by a web services API. [2]

Storage Devices

Storage devices can be broadly classified into two categories:

- Block Storage Devices
- File Storage Devices [2]

Block Storage Devices

The block storage devices offer raw storage to the clients. These raw storage are partitioned to create volumes. [2]

File Storage Devices

The file Storage Devices offer storage to clients in the form of files, maintaining its own file system. This storage is in the form of Network Attached Storage (NAS). [2]

Cloud Storage Classes

Cloud storage can be broadly classified into two categories:

- Unmanaged Cloud Storage
- Managed Cloud Storage. [2]

(a) Unmanaged Cloud Storage

Unmanaged cloud storage means the storage is preconfigured for the customer. The customer can neither format, nor install his own file system or change drive properties.

(b) Managed Cloud Storage

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 11, November 2015
Available at http://internationaljournalofresearch.org

Managed cloud storage offers online storage space on-demand. The managed cloud storage system appears to the user to be a raw disk that the user can partition and format. [2]

Cloud Storage:

Cloud storage involves exactly what the name suggests storing your data with a cloud service provider rather than on a local system. As with other cloud services, you access the data stored on the cloud via an Internet link.

Even though data is stored and accessed remotely, you can maintain data both locally and on the cloud as a measure of safety and redundancy. [4]

Creating Cloud Storage System

The cloud storage system stores multiple copies of data on multiple servers, at multiple locations. If one system fails, then it is required only to change the pointer to the location, where the object is stored.

To aggregate the storage assets into cloud storage systems, the cloud provider can use storage virtualization software known as Storage GRID. It creates a virtualization layer that fetches storage from different storage devices into a single management system. It can also manage data from CIFS and NFS file systems over the Internet. The following diagram shows how Storage GRID virtualizes the storage into storage clouds: [2]

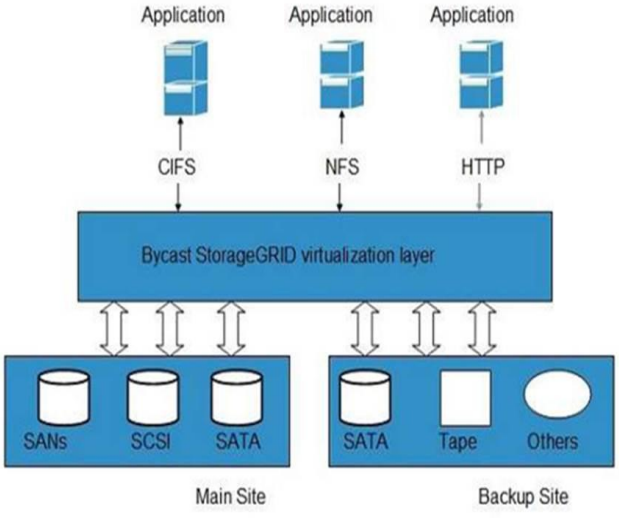

Fig. 1 Storage GRID virtualization the storage into storage clouds [2]

Virtual Storage Containers

The virtual storage containers offer high performance cloud storage systems. Logical Unit Number (LUN) of device, files and other objects are created in virtual storage containers.

Following diagram shows a virtual storage container, defining a cloud storage domain:

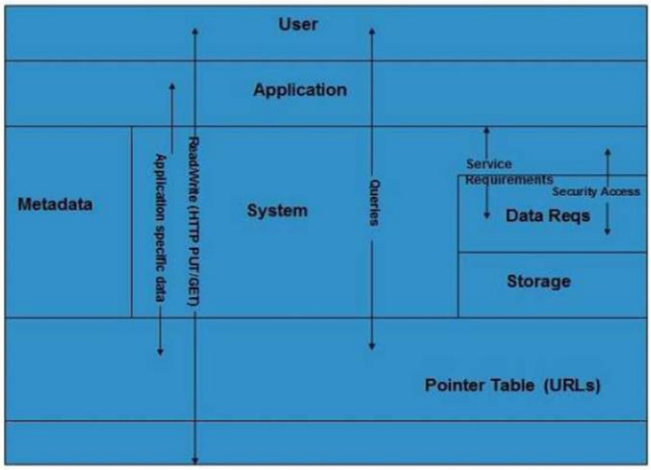

Fig. 2 Virtual storage container, defining a cloud storage domain [2]

Challenges

Storing the data in cloud is not that simple task. Apart from its flexibility and convenience, it also has several challenges faced by the customers. The customers must be able to:

- Get provision for additional storage on-demand.
- Know and restrict the physical location of the stored data.
- Verify how data was erased.
- Have access to a documented process for disposing of data storage hardware.
- Have administrator access control over data. [2]

III. CLOUD COMPUTING MANAGEMENT

(A) What is Cloud Computing Management ?

It is the responsibility of cloud provider to manage resources and their performance. Management of resources includes several aspects of cloud computing such as load balancing, performance, storage, backups, capacity, deployment, etc. The management is essential to access full functionality of resources in the cloud.

(B) Database Management

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 11, November 2015
Available at http://internationaljournalofresearch.org

The trickiest part of managing a cloud infrastructure is the management of your persistent data. Persistent data is essentially any data that needs to survive the destruction of your cloud environment. Because you can easily reconstruct your operating system, software, and simple configuration files, they do not qualify as persistent data. Only the data that cannot be reconstituted qualify.

The problem of maintaining database consistency is not unique to the cloud. The cloud simply brings a new challenge to an old problem of backing up your database, because your database server in the cloud will be much less reliable than your database server in a physical infrastructure. The virtual server running your database will fail completely and without warning. Count on it.

Whether physical or virtual, when a database server fails, there is the distinct possibility that the files that comprise the database state will get corrupted. The likelihood of that disaster depends on which database engine you are using, but it can happen with just about any engine out there. [3]

(C) Disaster Management

You are performing your backups and have an infrastructure in place with all of the appropriate redundancies. To complete the disaster recovery scenario, you need to recognize when a disaster has happened and have the tools and processes in place to execute your recovery plan. One of the coolest things about the cloud is that all of this can be automated. You can recover from the loss of Amazon's U.S. data centers while you sleep. [3]

Monitoring

Monitoring your cloud infrastructure is extremely important. You cannot replace a failing server or execute your disaster recovery plan if you don't know that there has been a failure. The trick, however, is that your monitoring systems cannot live in either your primary or secondary cloud provider's infrastructure. They must be independent of your clouds. [3]

Load Balancer Recovery

One of the reasons companies pay absurd amounts of money for physical load balancers is to greatly reduce the likelihood of load balancer failure. With cloud vendors such as GoGrid and in the future, Amazon you can realize the benefits of hardware load balancers without incurring the costs. Recovering a load balancer in the cloud, however, is lightning fast. As a result, the downside of a failure in your cloud-based load balancer is minor.

Recovering a load balancer is simply a matter of launching a new load balancer instance from the AMI and notifying it of the IP addresses of its application servers. You can further reduce any downtime by keeping a load balancer running in an alternative availability zone and then remapping your static IP address upon the failure of the main load balancer. [3]

Application Server Recovery

If you are operating multiple application servers in multiple availability zones, your system as a whole will survive the failure of any one instance—or even an entire availability zone. You will still need to recover that server so that future failures don't affect your infrastructure.

The recovery of a failed application server is only slightly more complex than the recovery of a failed load balancer. Like the failed load balancer, you start up a new instance from the application server machine image. Once the server is operational, you must notify the load balancer of the existence of the new server (as well as deactivate its knowledge of the old one) so that the new server enters the load-balancing rotation. [3]

(D) Database Recovery

Database recovery is the hardest part of disaster recovery in the cloud. Your disaster recovery algorithm has to identify where an uncorrupted copy of the database exists. This process may involve promoting slaves into masters, rearranging your backup management, and reconfiguring application servers.

The best solution is a clustered database that can survive the loss of an individual database server without the need to execute a complex recovery procedure. When an instance goes down, however, any number of related issues may also have an impact on that strategy:

- The database could be irreparably corrupted by whatever caused the instance to crash.
- The volume could have gone down with the instance.
- The instance's availability zone (and thus the volume as well) could be unavailable.
- You could find yourself unable to launch new instances in the volume's availability zone. [3]

(E) Cloud Management Task

The cloud provider performs a number of tasks to ensure efficient use of cloud resources. Here, we will discuss some of them: [2]

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 11, November 2015
Available at http://internationaljournalofresearch.org

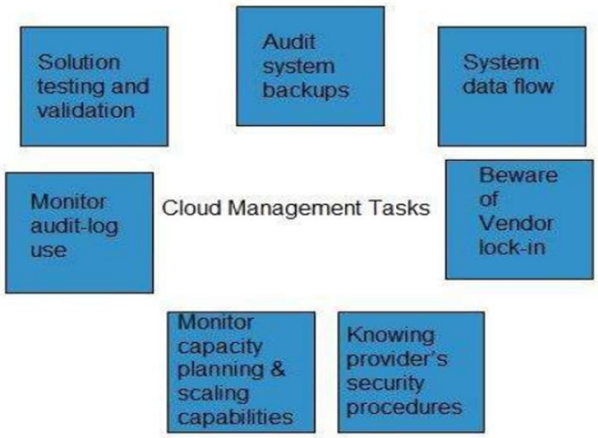

Fig. 3 The Cloud Management Tasks [2]

Audit System Backups

It is required to audit the backups timely to ensure restoring of randomly selected files of different users. Backups can be performed in following ways:

- Backing up files by the company, from on-site computers to the disks that reside within the cloud.
- Backing up files by the cloud provider.

It is necessary to know if cloud provider has encrypted the data, who has access to that data and if the backup is taken at different locations then the user must know the details of those locations. [2]

Data Flow of the System

The managers are responsible to develop a diagram describing a detailed process flow. This process flow describes the movement of data belonging to an organization throughout the cloud solution. [2]

Vendor Lock-In Awareness and Solutions

The managers must know the procedure to exit from services of a particular cloud provider. The procedures must be defined to enable the cloud managers to export data of an organization from their system to another cloud provider.[2]

Knowing Provider's Security Procedures

The managers should know the security plans of the provider for the following services:

- Multitenant use
- E-commerce processing
- Employee screening
- Encryption policy. [2]

Monitoring Capacity Planning and Scaling Capabilities

The managers must know the capacity planning in order to ensure whether the cloud provider is meeting the future capacity requirement for his business or not.

The managers must manage the scaling capabilities in order to ensure services can be scaled up or down as per the user need. [2]

Monitor Audit Log Use

In order to identify errors in the system, managers must audit the logs on a regular basis. [2]

Solution Testing and Validation

When the cloud provider offers a solution, it is essential to test it in order to ensure that it gives the correct result and it is error-free. This is necessary for a system to be robust and reliable. [2]

Conclusion

The study of Data Storage we can kwon that how data are Store in any Cloud and how data are maintained by Cloud Provider. The data Storage and Management is major issue for Cloud computing Provider. The main challenges in Cloud Computing are Database Recovery, Application Server Recovery and Load Balancer Recovery.

References

[1] An Study of Security Issues & Challenges in Cloud Computing by Mili Patel, Rakesh Patel, Department of Information Technology of Kirodimal Institute of Technology Raigarh (C.G.), India.

[2] http://www.tutorialspoint.com/cloud_computing/cloud_computing_overview.

[3] Cloud Application Architectures by George Reese, April 2009 First Edition, Published by O'Reilly Media, Inc.

[4] Anthony T.Velte, Toby J.Velte, Robert Elsenpeter, Cloud Computing a Practical Approach, TATA McGraw-Hill Edition 2010.