Using cloudsim to model and simulate snapshot and D2D in cloud storage architecture

Hatem Taha[#1], Ali Manea[*2], Noura Aknin[#3], Kamal Addin El kadiri[#4]

[1, 3, 4] *Department of Informatics, Abdelmalek Essaâdi University, Tetouan, Morocco*
[2] *Department of Electrical Engineering, Sana'a University, Sana'a, Yemen*

Abstract *cloud storage service (STaaS) are increasingly gaining popularity and becoming one of the major service in cloud computing. Nowadays, although many cloud storage services available on the internet they aren't convenient and they have many challenges to apply it and benefit from it. In this paper, we will implement the model data backup technologies of our cloud storage architecture on a cloud computing simulation toolkit known as cloudsim using snapshot Amazon ec2 and disk-to-disk. We will present the new concept for cloud storage architecture to be able to store and backup data through model controller in layer backup technology that can be accessed through the APIs. The implementation aims to increase the availability of data, modernizing data protection and reduce in loss of data in the cloud storage environments. Also this solution could provide new information and gives addition that is useful for further researches and open new opportunity for the future cloud storage.*

Keywords — Architecture**,** *Cloud, Storage, Model, Backup, CloudSim*

I. INTRODUCTION

Cloud Computing is new technology generation that includes a collection branches of computer studies (security, programming, storage, Network…), and other sciences (physic, math…). It isn't just about accessing applications over the web. It can be almost anything, such as, it used for collection of data in the context of security or data management. So we consider the cloud computing as the primary source for all of the IT services, such as our first source of file storage (Images, files, documents), to deploy services, to share and to backup data as show in the figure 1.

Fig.1: Cloud Computing

As known, there is no specific description of this service, but it plays an important role to treat various services request of costumers using virtualization technology. More and more enterprises use the professional cloud storage service providers such as Amazon's S3[1], Google's GFS [2] and Microsoft's Live Services to store and manage their data. It is also one of the most important of the solutions associated with challenges of backup data. Cloud Storage integrates Grid Computing, Distributed Systems, Virtualization, Load Balancing and other traditional network technology. It has the properties of high reliability, high scalability, high efficiency and low cost etc.

Testing the model of the system with the real infrastructure is usually limited by specific infrastructure and takes a lot of money to set up the experiment environment and it is difficult to repeat the experiment. Therefore, a feasible way for developers for testing any system is simulation tools. Simulation tools enable developers to repeat it in a controllable and also can be downloaded and installed free of charge. In this paper, we propose CloudSim: an extensible simulation toolkit that enables modeling of Cloud storage environments. CloudSim is the simulation program for the above two purpose [3]. It enables seamless modeling, Implement, simulation, a creation of one or more virtual machines (VMs) and experimentation of cloud computing environment and the application services. It also allows simulation of multiple Data Centers.

This article explains the implementation and modeling of the fourth layer in the cloud storage architecture, modernizing data protection and backup and how users can benefit from it. Thus, backup is one important aspect of the cloud storage services, which provides the ability to restore data in the event of any disaster, without the need for specialized backup servers.

II. RELATED WORK

This section presents background information on various architectural elements for basis of Cloud Storage. It also presents requirements of various applications that need to implement this model as an alternate approach for achieving the acceptable solution.

A. Storage-as-a-Services (STaaS)

Storage-as-services: are systems which provide storage services to customers via the Internet that uses virtual technology (virtualized storage device). Integrates various storage devices of the network by a different application, which use a large amount of network bandwidth required to conduct storage. Cloud storage provides many different access interfaces for all application cloud storage architecture.

In general, users can access cloud storage to store and retrieve data using web services protocols, such as REST or SOAP. The services online backup is the most developed application in Cloud storage service (STaaS), but It faces challenges relate to the internet service, which we will study it in the next search, such as performance, scalability geographic, network security, availability and data encryption[4].

Cloud storage can provide high reliability and security storage service at competitive price. Also, customers have the option to trade the privacy of their data for the convenience of storage services[5]. There are more than 60 CSPs (cloud storage providers) that deliver economies of scale by using different storage capacities to meet the needs of many organizations, passing the cost savings to their customer bases, such as Amazon, Google, Microsoft and DropBox. It can be an acceptable solution to keep up storage infrastructure and additionally to mitigate the risk of the data loss just in case the disaster. All customers know that there is a need for maintaining simplicity, scale back data storage prices, freedom from vendor lock-in, improve availability data, continuity and sensible performance.

From this groundwork, we see that may provide an opportunity for some emerging enterprise with growing their data. Also, this may be an effective solution which store resources in the cloud for government, enterprise and users. These users can at anytime, anywhere, through any device connected to the Internet cloud and easily access the data. In terms of work, there are a lot of problems and challenges exist.

B. Cloud Snapshot Technology

In this section, we would like to provide the reader a quick outline of some uses snapshotting generally, to understanding the opportunities and addressing a broad array of interconnected challenges for the structure of cloud storage. We will describe the snapshot cloud in architecture cloud storage and explain its underlying challenges at the core issue in backup and recovery. In any storage system, the protection of data is probably one of the most sensitive issues. More generally speaking, a snapshot is the state of a system at a particular point in time. High-availability systems may instead perform the backup on a snapshot that is a read-only copy of the data set frozen at a point in time and allow applications to continue writing to their data[6]. Snapshot could also provide another data access channel to accelerate I/O performance [6].

It can, therefore be said that the snapshots in the cloud storage system are used to avoid downtime, snapshots technology look and behave like complete backups[7]. There are an also large number of vendors offering different methods for exploiting snapshot to management, use, storage simpler, and consequently, more reliable. To give just two examples: Data Tiering and Remote Replication, these two are prime examples and crucial to any storage system [8][9].

C. How Does Snapshots Cloud Work?

Users must identify what they require of their cloud storage providers (CSPs), and to do this they must have a better understanding of how making use of cloud storage. In the context of this work, there are different methods adopted by vendors to create snapshots, each one having its own benefits and its drawbacks. Therefore, it is important to understand snapshot implementations. In a concise manner, we will present the primary ways in which snapshots can be incorporated into the cloud storage architecture to enhance data security.

Snapshots used in shadow-paging file systems. As NetApp WAFL and Oracle Solaris ZFS, Hadoop Distributed File System (HDFS). This also allows the administrator to create a new block device and to modify data, and to overwrite the old data of a logical volume. While there are a number of mechanisms that can protect the old data, for instance, ROW “Redirect-on-Write” Redirect-on-write is comparable to copy-on-write. This work focuses on the context in which copy-on-write is produced and modified. Likewise, this reduces the number of writes from two to one, which has to be written in new places and this is one of the most important challenges because the original data set can quickly become fragmented. However, with redirect-on-write, the original copy contains the

point-in-time data, that is, snapshot, and the changed data reside on the snapshot storage [7]. While COW "Copy-On-Write" Which allows the user makes changes to the same data "copy of the data", that they may occasionally need to modify the same resource, with the preservation of the original resource unchanged. Furthermore, make significant resource savings by sharing resources[7].

The main job of snapshot cloud is to backup or recover data. However, this presents some challenges, particularly in the working long-term and the numbers of the snapshot with these technology [10].

D. Cloud Disk-to-Disk

In the past few years, Solid State Drive (SSD) has started making its inroads into the enterprise IT environment and which contributed to reducing Hard Disk Drives (HDD) as a storage technology was widely used in the enterprises. For this reason, enterprises need higher input/output (I/O) performance.

Wherefore, we see that the cloud storage is hybrid approach combines new technology and traditional that used to reduce burdens for enterprises especially in the start-up phase, also to protecting the customer's data so that it could be extensively used for storing data and sharing. In response to these issues some innovators in this space are blending the capabilities of traditional onsite backup systems with cloud storage. This is even more important now, to ensure continued industry trend toward using disk-to-disk (D2D) backup for more frequent and timely data protection for the cloud storage system. In addition, the combination of all these innovation provides maximum protection for the customer data in cloud storage.

For that, We need a cloud storage is accessible to all, safe and scalable, with offsite backup for any data or any system anytime, anywhere, which means access to cheap store services for the customer. Furthermore, we see that these requirements may be realized in the technology Disk-to-Disk which is gaining in popularity in both large enterprises and small businesses. There are in fact a number of hybrid solutions with some hybrid cloud backup tools today, such as Acronis Inc., CommVault Systems Inc., and Symantec Corp.

E. Cloud Simulator

Definitions, tools, and methods should be clearly defined in the standards for cloud computing that can be used to ensure a search efficiency and effectiveness of results at the infrastructure and applications level and controlling them. As well as the use of real infrastructures that limits the experiments to the scale of the infrastructure and the complex applications. Also, makes the reproduction and repeatable of results an extremely difficult undertaking, The main cause for this being the conditions prevailing in the Internet-based environments are beyond the control of developers[3].

In order to overcome this challenge, the simulation such as CloudAnalyst, GreenCloud, Network CloudSim, EMUSIM and MDCSim enables us to accomplish goals of study and mechanisms that might not be implementable without a substantial investment, for example, accurate results, reduce the cost of analyzing, managing infrastructure and datacenter and monitoring applications. Cloud simulators are required for cloud system testing to decrease the complexity and separate quality concerns[11]. They enable performance analysts to analyze system behavior by focusing on quality issues of a specific component under different scenarios.

III.PROPOSED WORK

A. Modeling the Cloud Storage Architecture

In this section, we will describe how the model has implemented architecture [12] that we will propose to overcome the problems and challenges in the fourth layer (cloud backup technology) of architecture cloud storage and identify some points in it, and we will describe the existing CloudSim framework. One of the key aspects that make a Cloud storage architecture, data sharing, archiving, collaboration, and synchronization services different from a traditional storage is the massive deployment of virtualization technologies and tools.

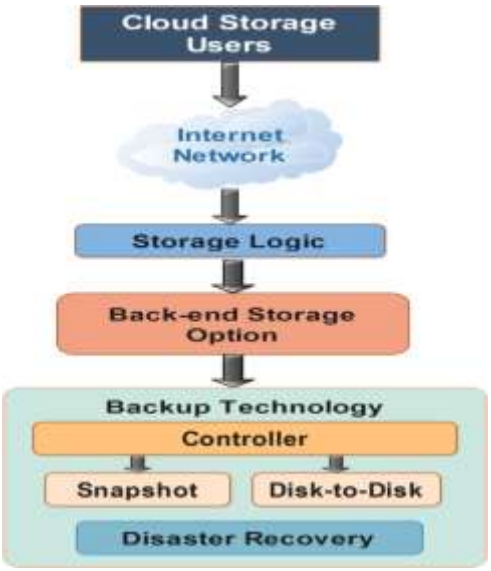

Fig.2 :Architecture Cloud Storage Overview [12]

Developers try to make further simulations to accommodate the needs, the Storage-level services (STaaS) related to the clouds can be simulated by extending the Datacenter entity of different cloud simulation.

On this face, we are seeking to implement hybrid solutions which could enable users cloud storage to select the method of keeping their data, management and raise the level of confidence they have in the proposed architecture of cloud storage. In addition, that has flexible and practical methods for keeping data and their services while reducing the costs associated with service delivery. Generally, this method is useful when evaluating architecture of cloud storage, and to improve storage efficiency, performance, reliability, and compatibility. Also, it's an unbiased method because it accurately reflects the controller between the snapshot and the D2D disk-to-disk for backup. On the other hand, it has particular advantage for modeling backup technology layer.

B. Modeling the Cloudsim

CloudSim is a framework simulation which enables to model, simulation, and experimentation of cloud computing and application services. CloudSim is a popular simulation environment, which offers capabilities to simulate computing Clouds that provision computing infrastructures as a service [3].

In this area, the Storage-level services (STaaS) related to the clouds can be simulated by extending the Datacenter entity of CloudSim [13]. The focus of CloudSim is on the modeling of so-called Cloudlets which are jobs, which can be scheduled on VMs (Virtual Machines). Models for hard disks, SAN and files are already included in CloudSim. However, capable interfaces to simulate object STaaS are missing, as well as a fine-grained model for the size of files. This type of cloud is a good alternative. Management is carried by a client that controls which applications should run. Servers, network, and storage devices are the property of the organization.

C. Modeling of Controller Details

In the classical cloud storage the user must subscribe for each service to benefit from it. For example, D2D, combined with compression and de-duplication disk-based solutions, is used for local, daily and recurring backups that have shorter retention but that have more granularities or Snapshots which provide a way of creating virtual time to get essential data protection completed while minimizing impacts to applications and boosting productivity. Then each service has its own limits and conditions that make the task more complex and not reliable. Disk-to-disk; snapshots; backups, and replication have become popular options for near-term and real-time data protection to meet RTO and RPO requirements. The important point is that some applications require RTOs and RPOs of zero or close to zero and need some technologies such as synchronous replication, data mirroring combined with snapshots or continuous data protection across multiple sites. Due these mentioned reasons we try to create new model to exploit from two technologies and get advantages of each one using one service, one count.

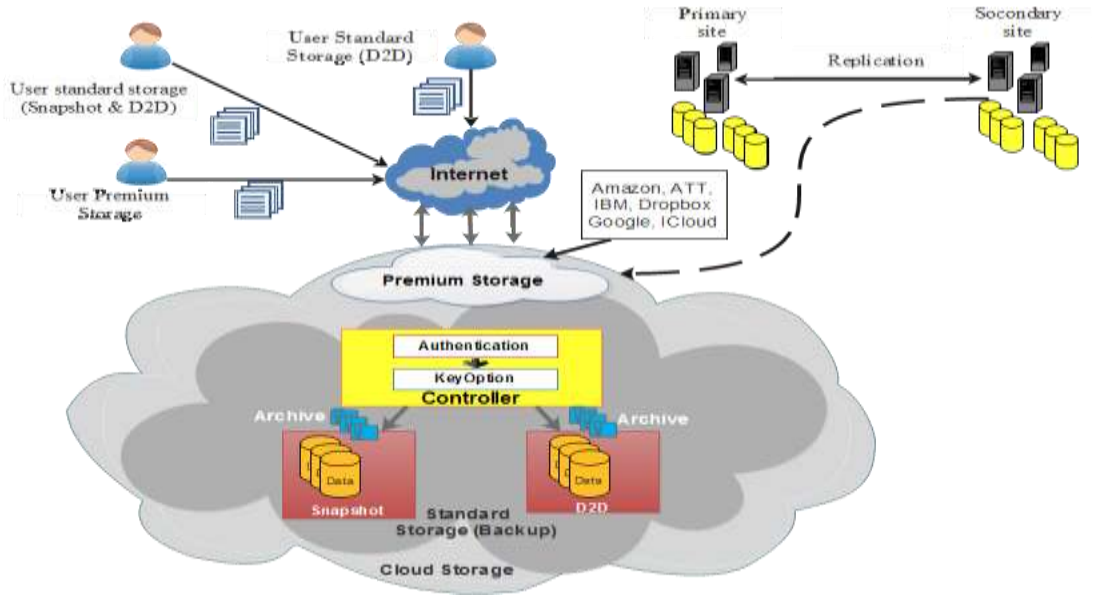

Fig 3: Evolution Architecture Cloud Storage

Figure 3 shows how the above scenario works. Starting at the right, normal operations with data are replicated to the secondary or cloud storage along with routine snapshots, journal, or log file D2D copies made to multiple locations. Should a failure occur at the primary site, such as failure of a disk drive or other component, it can be isolated and contained using various techniques. Should a more serious incident occur, failover can occur to the secondary or premium storage in clouds (Amazon, IBM, Dropbox, Microsoft Azure,...), where access can continue or resume or where recovery can quickly occur to a given point of time. Should an even more serious incident occur that results in the primary; secondary site or premium storage and their resources not being available, the disk-to-disk and snapshot data at a Cloud Backup Technology layer can be leveraged?

The Controller Sub-layer ensures an access to resources securely by setting many configurations. It is responsible for managing data and selecting the best of technology to exploit the Cloud Backup Technology layer, As Figure 3 shows. In addition to making copies of data that are stored in different locations (a local copy on disk, another copy on a fileserver, another stored at an off-site cloud or managed service provider site), retention is also important. Precisely by devising the controller layer to two new other sub-layers: authentication and KeyOption based on CloudSim as the simulator.

A controller can be seen as a manager that ensures that all resources needed for system backup and recovery are available. This is based on a cloud Storage Model within our article [12]. In the fourth layer are used two technologies for backup (snapshot and Disk-to-Disk) in backup technology layer, Snapshot and Disk-to-Disk that are composed of the rules of control and applications for storage data.

In other terms we add a particular work that one controller controls the whole Cloud Backup Technology, it includes list of users and list of its data as an instance of IOStreams class in JAVA. The user authenticate and choose the technology recommended by our model automatically or manually by user himself to ensure this procedure perfectly we add new model as sub-layer called controller which is seen from it name it control all operations of users in the Cloud Storage.

1) ***Authentication Model:*** The authentication model serves to access securely in cloud storage. Although different technologies storage: disk-to-disk (D2D) and snapshot. Each service it has an own profile, but our work based on one account to facility access to the cloud storage so we create a new model called "USER", the profile can be a client of Amazon services or profile of disk-to-disk (D2D) technology implemented by us. To benefits services from Amazon, we need an Access Key and SecretKey. With two profiles make the task complex. The proposed model combines profiles to one profile account in order to improve scalability, flexibility.

2) ***Keyoption Model:*** KEYOPTION is set of constraints and conditions for using Cloud Backup Technology including the size of data, number of copies, duration data on the disk and the cost. It indicates which technology can and should the user use it: SNAPSHOT OR disk-to-disk (D2D). To deal with these constraints we define keys as attributes on a model class called "Parameter" that its values can be configured depending on provider service. The model proposed save data automatically into volume using snapshot when size of data in volume is less than limited size and number of copies less than one hundred copies into the disk, in other cases "KeyOption" consider disk-to-disk (D2D) as main technology storage. The limited size is the maximal value of data permitted by the provider, so when the total size data in the volume is exceed this value the model automatically change the technology to D2D. The attribute "duration" means how long data still into the disk when user use snapshot as main technology service, but when data exceed sixty days the data will be removed automatically or can the user refer to disk-to-disk (D2D) technology. And finally "limited copies" is the last attributes that from it name seen is number of copies. In the cloud storage system, Snapshots and disk-to-disk (D2D) provide significantly easier, reliable, effective-cost, automated, and faster backups than any traditional backup system.

D. Modeling Snapshot Technologies

Snapshot technology are commonly used to enhance data protection, recovery time objectives and recovery point objectives as mentioned in this article[7].Amazon service offers the third party API for developing and writing code by Java programming and you can work with snapshots using the AWS Storage Gateway console, one of the AWS Software Development Kits (SDKs), or the AWS Storage Gateway REST API. We based on it to develop our model called "MYSNAPSHOT"; we redefine two procedures in the model: CREATSNAPHOT and SREACHSNAPSHOT.

As in the site web of Amazon explain that procedure CREATSNAPSHOT serve to create new image for volume EBS, then the second procedure SEACHSNAPSHOT aim to looking for snapshot images which created before.

E. Modeling Disk-to-Disk Technologies

According to the article[14], disk-to-disk (D2D) enhance the server levels for important data types in the organization by increasing reliability of backup and data protection.

The advantage of this, when user subscribe to a cloud storage service, user can lease one or two kinds of technologies backup from the backup cloud storage, which can access via the API. This model to

improve functionality of cloud backup layer, it choose any type of technology, or both according to the data storage management method or the moving of our interests in the Cloud Storage.

Disk-to-disk (D2D) requires both software and hardware to replicate data to disk on another. In this task we try to implement disk-to-disk (D2D) technologies by creating the model Disk written by JAVA language. The model has set of disks which each disk has a D2D identity to describe where data will be transferred.

IV. Conclusion and Future Work

Cloud storage is a service model of cloud computing in which data is maintained, managed, backed up remotely and make available to users over a network. in this area, simulation tools are becoming more and more important in the evaluation of the Cloud storage model, its approaches helps us to evaluate architecture of cloud storage that are available with various specifications. Some of them we have used to implement backup technology such as CloudSim. It enable for us to test our proposal and solution to this new technique with reduced costs and time-to-implement. In future work we try to integrate our proposal model as new layer or sub-layer for CloudSim Simulator, so we will develop a new API that can be used it for users and developers to manage data in storage disk. There for we will add more features as monitoring backup disk to show details of each data center in cloud storage and also application web based on HTML5 to get simulation more rich. Finally, this technique is being developed in order to give users the experience of managing a private cloud environment that facilitates the backup and data storage, using technologies already exists.

Reference

[1] D. Bermbach et S. Tai, « Eventual consistency: How soon is eventual? An evaluation of Amazon S3's consistency behavior », in *Proceedings of the 6th Workshop on Middleware for Service Oriented Computing*, 2011, p. 1.

[2] S. Ghemawat, H. Gobioff, et S.-T. Leung, « The Google File System », in *Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles*, New York, NY, USA, 2003, p. 29–43.

[3] R. Buyya, R. Ranjan, et R. N. Calheiros, « Modeling and Simulation of Scalable Cloud Computing Environments and the CloudSim Toolkit: Challenges and Opportunities », *ArXiv09074878 Cs*, juill. 2009.

[4] « What is Storage as a Service (SaaS)? - Definition from WhatIs.com », *SearchStorage*. [En ligne]. Disponible sur: http://searchstorage.techtarget.com/definition/Storage-as-a-Service-SaaS. [Consulté le: 01-déc-2015].

[5] S. Kamara et K. Lauter, « Cryptographic Cloud Storage », in *Financial Cryptography and Data Security*, R. Sion, R. Curtmola, S. Dietrich, A. Kiayias, J. M. Miret, K. Sako, et F. Sebé, Éd. Springer Berlin Heidelberg, 2010, p. 136‑149.

[6] C. Lijun, L. Zhaoyuan, et J. Weiwei, « A snapshot system based on cloud storage Log-Structured Block System », in *2014 IEEE 11th International Conference on Networking, Sensing and Control (ICNSC)*, 2014, p. 392‑398.

[7] « Understanding and exploiting snapshot technology for data protection, Part 1: Snapshot technology overview », 26-avr-2006. [En ligne]. Disponible sur: http://www.ibm.com/developerworks/tivoli/library/t-snaptsm1/. [Consulté le: 09-mars-2016].

[8] A. P. Amarendran, T. Chatterjee, et P. Varadharajan, « Automated, tiered data retention », US8805806 B2, 12-août-2014.

[9] B Lakshmi, , C Chandrasekhara Reddy, et , SVSRK Kishore, « ARCHIVING AND MANAGING REMOTE SENSING DATA USING STATE OF THE ART STORAGE TECHNOLOGIES », *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Hyderabad, India, p. 1153‑2014, 09-déc-2014.

[10] « B | SNIA ». [En ligne]. Disponible sur: http://www.snia.org/education/dictionary/b#backup_copy. [Consulté le: 31-déc-2015].

[11] A. Oujani et R. Jain, *A Survey on Cloud Computing Simulations and Cloud Testing*. .

[12] Taha HATEM., M. Ali., E. K. Kamal Eddine, et A. NOURA., « EXPANSION OF ARCHITECTURE CLOUD STORAGE WITH SNAPSHOT AND DISK-ToDISK IN CLOUD COMPUTING », *GLOBALJOURNAL OFENGINEERINGSCIENCE ANDRESEARCHES*, india, avr-2016.

[13] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. De Rose, et R. Buyya, « CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms », *Softw. Pract. Exp.*, vol. 41, n° 1, p. 23‑50, janv. 2011.

[14] « How to implement disk-to-disk backups ». [En ligne]. Disponible sur: http://searchdatabackup.techtarget.com/feature/How-to-implement-disk-to-disk-backups. [Consulté le: 22-sept-2016].