# Importance of redundant techniques used in hybrid and private cloud storage and its comparison

Abdul Ahad Afroz, Research Scholar, Shri JJT University
Dr. Prasadu Peddi, Research Guide, Shri JJT University

## ABSTRACT

*Report redundancy strategies were very useful mechanisms for providing fault tolerance and facts availability in any kind of storage. Cloud storage is not the exception. This paper gives an evaluation of classical file redundancy techniques carried out in cloud-storage deployment fashions, personal and hybrid. A small prototype of a private and hybrid cloud garage changed into carried out for this evaluation. The overall performance impact while report redundancy is best carried out in a non-public cloud as opposed to whilst redundancy is also dispensed in a public cloud (the hybrid model) is analyzed. Extra to classical report redundancy strategies, an modern method changed into evaluated for record redundancy based mostly on an statistics dispersal algorithm (IDA).The use of IDA represents an amazing choice for managing sensitive facts in hybrid cloud storage. On this approach, simplest elements of a file want to be despatched to the public cloud, fending off the complete document to be study from out of doors of the private region. In this context, there's a exchange-off between performance (for reconstructing the unique document, it's far first vital to gain all of its fragments) and the security stage that might determine the viability of the usage of IDA*

*Keywords: cloud computing, data management, virtualization.*

**1. Introduction :**To define the storage necessities for institutions or organizations of any length has emerge as a problem without a trivial answers. It's miles specially due to the very rapid generation of digital information whose conduct could be very dynamic [1]

In this context, it's far common for managers of storage assets, with the duty to make predictions approximately the resources as a way to be needed in the medium term, these ways are

A)        Predictions are beneath real needs. In this example, there may be a trouble of useful resource deficit.
B)        excessive expenditure on the acquisition of storage assets, that could produce a complex administration, probably with resources that will notbe used in the medium term.

In this case, the acquisition of storage offerings that put in force an elastic concept will become attractive, i.E., storage capability that can be increased or decreased on demand, with a value of acquisition and management relatively low.
These days, this service version is known as cloud computing. In this version, storage assets are provisioned on call for and are paid in keeping with consumption.
Services deployment in a cloud computing surroundings can be applied in basically three ways: non-public, public or hybrid. Inside the private choice, the infrastructure is operated entirely for a single organization; maximum of the time, it implies an preliminary sturdy investment because it's miles necessary for the organisation to buy a huge amount of storage sources and pay for the administration expenses.
The public cloud is the maximum traditional version of cloud computing. On this model, the infrastructure belongs to an external business enterprise, in which charges are a feature of the assets used.

These expenses include management. Finally, the hybrid model carries a mixture of each. A cloud computing environment is specifically supported through one of a kind technologies along with virtualization and carrier-orientated architectures.

A cloud environment offers   omnipresence and helps deployment to file storage offerings. It way that customers can access their documents from anywhere, even as there exists an internet connection and with out requiring the set up of a unique application (best a web browser is wanted).

Information availability, scalability, elastic carrier and pay simplest for consumption are very appealing characteristics found in the cloud carrier model. Virtualization plays a very essential role in cloud computing. With this technology, it's miles possible to have facilities which include a couple of execution environments, sandboxing, server consolidation, use

of multiple operating structures, software program migration, amongst others. Except virtualization technologies, rising tools that allow the creation of cloud computing environments, offering dynamic instantiation and launch of virtual machines and software migration also are

solution. Groups or users in wellknown can save sensitive or most frequently used statistics in the personal supporting the elastic service offered in this kind of computing model.

Presently, it's miles possible to locate numerous proposals for public cloud storage which includes Amazon S3 [2], RackSpace [3], or Google storage [4], which give high availability, fault tolerance and services and administration at low price. But, there nonetheless exist organizations that do not experience confident about storing their information in a third- celebration-owned surroundings. In these instances, such companies wanting to take blessings of the cloud computing centers would require to enforce a non-public cloud answer. Lamentably, this option is often past their budgets.

 In this case, a hybrid cloud version may be an low-cost infrastructure and less sensitive facts in the public cloud.

The development of a prototype of a file garage provider carried out on a private and hybrid cloud environment using specially unfastened and open-source software (FOSS) enables us to research the conduct of various redundancy techniques, paying unique interest to the low value of the machine implementation, the machine efficiency, useful resource intake and the exclusive ranges of data privacy and availability that can be reached by means of a system like this

This paper is organized as follows: segment 2 introduces an offer for a cloud computing infrastructure based on unfastened open supply software (FOSS). It additionally describes the redundancy techniques that have been applied for this assessment. Segment three gives the assessment scenario and a performance analysis thinking about numerous aspects consisting of: the impact of getting an elastic garage carrier, the implementation of different redundancy techniques in each private and hybrid cloud computing environments. Phase four consists of the related paintings, and eventually phase 5 gives a few essential comments and conclusions.

## 2. Infrastructure description

In recent times, small and medium organizations (SMB) are going through within your means and technical challenges that get up when trying to reap the advantages of having their very own cloud computing surroundings

(private). The purpose of this concept is to help with the ones demanding situations with the aid of designing and enforcing a scalable and elastic distributed storage architecture based totally on unfastened and well-known open source equipment. This structure combines non-public and public clouds by means of developing a hybrid cloud surroundings. For this purpose, tools which includes KVM [5] and XEN [6] have been evaluated, that are beneficial for developing virtual machines (VM). OpenNebula [7], Eucalytus [8] and OpenStack

[9] are proper free alternatives for dealing with a cloud environment. OpenNebula was the choice for this prototype because there's enough to be had statistics on line that does not require a strong technical heritage

The difficult disks (HDs) integrated into the storage infrastructure are determined in business computer systems (commodities). The use of this type of HDs makes this architecture failure-prone. This case become the principle motivation to evaluate exceptional redundancy mechanisms, providing numerous tiers of information availability and fault tolerance. Discern 1 (a) indicates the center additives of our storage structure (the non-public cloud) and (b) a disbursed garage net software (DISOC) that is used as a proof of concept. It is able to also be observed that the personal cloud has an interface to get right of entry to a public cloud growing a hybrid environment.

The core components of the architecture are the following :

 virtual machine (VM). Extraordinary open supply gear were evaluated, inclusive of KVM [10] and XEN [6], for the introduction of digital machines. Some overall performance checks have been accomplished, it turned into discovered that KVM showed a slightly better performance than XEN (comparable effects can be found at [10]).

virtual system manager Module (VMMM). It has the function of dynamic instantiation and de- instantiation of digital gadget depending on the cutting-edge load on the infrastructure.

records get entry to Module (DAM). All the virtual disk area required by way of each VM was acquired

thru the usage of the data access Module Interface (DAM-I). DAM-I permits VMs to get access to disk space by means of calling the information get entry to Module (DAM), which affords obvious get entry to to the different disks which might be a part of the garage infrastructure. It allocates and retrieves man or woman files stored on exceptional record servers.

Load Balancer Module (LBM). It's far designed to distribute the load amongst special VMs instantiated on the bodily servers that make up the personal cloud.

Load manager (LM). It's far liable for monitoring the load that could arise in the non-public cloud

distributed storage at the Cloud (DISOC). It is a web-primarily based file storage machine this is used as a evidence of idea and become implemented primarily based on the proposed structure.
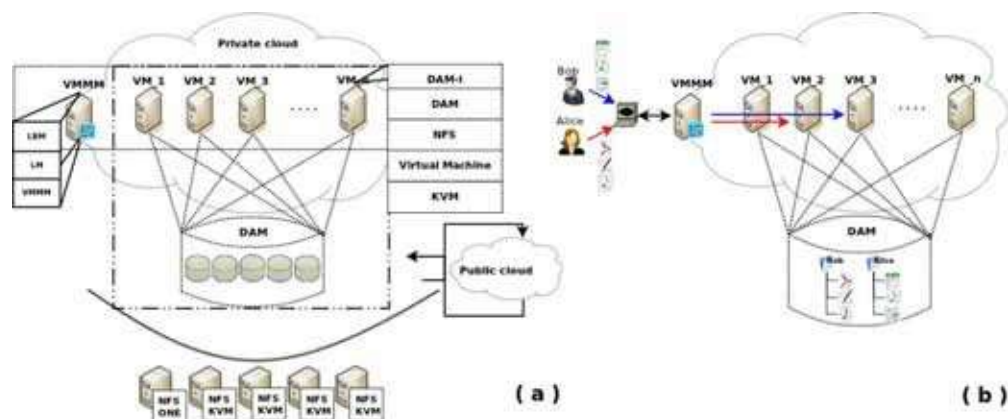


Figure 1. Main components of the cloud storage architecture.

### 2.1 Redundancy techniques

-Excessive availability is one of the vital features provided in a garage service deployed within the cloud. To accomplish it, the use of redundancy strategies has been the maximum useful concept [19][20]. DAM is the issue this is configured to provide special tiers of data availability. It currently consists of the subsequent redundancy policies: no-replication, general-replication, mirroring and IDA-based redundancy.

-No-Replication. This redundancy policy represents the information availability approach with the bottom stage of fault tolerance. In this method, best the unique version of a file is saved within the disk pool. It follows a round Robin allocation policy, which relies upon on disk availability. This policy prevents all files to be allocated in a unmarried server, offering a minimal fault tolerance

-Mirroring. This redundancy technique is a easy manner to ensure better availability, without high aid consumption. On this redundancy, whenever a report is stored in a disk, DAM creates a replica and places it on a specific disk following a spherical robin coverage

general-Replication. It represents the very best facts availability approach. In this technique, a copy of the document is saved in all of the report servers to be had. It's also the method that requires the highest intake of assets.

•       IDA-based totally redundancy. As a manner to offer better facts availability, with much less effect on the intake of sources, an opportunity method primarily based on data dispersal techniques may be used. The information Dispersal set of guidelines (IDA)

[12]  [11]    is an instance of this approach. While a report (of length is needed to be stored the usage of IDA, the record is damaged into n pieces of sizeFin which m<n. Those portions are disbursed in n distinctive disks. IDA only desires to gain m portions to reconstruct the precise report. On this context, although n-m disks failed, the record could though be recovered. It's far suitable that no greater than n-m document server fail

[13]   IDA affords better fault tolerance than mirroring without having to totally reflect the unique file. On this

prototype, IDA was evaluated with n = five and m = 3 (it means handiest a 60% of the original record can be replicated). IDA is very appealing for being used  a hybrid cloud environment, because it isn't vital to shop the complete report on a unmarried file server (disk). On this manner, it may be possible to send okay fragments of the file (wherein ok<m) to a public cloud storage with out revealing the entire content material of the authentic record.

Perfomance evaluation

A prototype of this architecture changed into carried out and used as the assessment state of affairs. It includes 5 business computers (commodities) whose characteristics are shown within the first segment of desk 1. The capabilities of the VMs that had been instantiated on the cited pcs are shown within the second segment of desk 1.

| Physical machines | | | | |
|---|---|---|---|---|
| PCs | Cores | Memory | Hard disk | Network |
| 1 pc | 4 | 4 Gb | 640 Gb | Etht 10/100 |
| 4 pc | 2 | 2 Gb | 250 Gb | Etht 10/100 |
| Virtual machines | | | | |
| 8 vm | 1 | 1 Gb | 1 Gb | Virtual Eth |
| 1 vm | 1 | 128 Mb | 1 Gb | Virtual Etht |

Table 1. Characteristics of the physical PCs and VMsused in the private cloud.

A total of 9 VMs had been created in a non-public cloud surroundings for this assessment. To be able to construct and check a hybrid cloud surroundings, it changed into vital to get entry to a public garage cloud (third-birthday party infrastructure). Two extraordinary public garage vendors were used in this test, Dropbox

[13]  and Phoenix (also referred to as TreeStore) [13]. Those 2 storage services were selected due to the fact they have got a loose service model and offer a simple software interfaces (API) for 0.33-birthday celebration builders. Our information get entry to Module (DAM) changed into additionally accountable for providing a transparent access to the outside garage infrastructure. It became required to send a legitimate consumer and password in both carriers. Moreover, for accessing Dropbox, it's also essential to obtain a key for developers. This secret's required via the Dropbox API. It's far vital to say that Dropbox is also capable of maintain files in the Amazon S3 garage infrastructure

Special workloads had been emulated, strolling concurrent customer applications that sent many parallel document upload and down load requests to our cloud storage prototype. The private garage cloud configuration turned into first examined via receiving 50, one hundred and one hundred fifty parallel requests. It's far well worth citing that after testing the hybrid cloud configuration, it was now not feasible to ship the same number of parallel requests used inside the private configuration. It become essential to lower this range due to the fact the general public cloud garage vendors (Dropbox and Treestore) could take it as an assault towards their servers and, thus, to dam the provider. Within the hybrid configuration take a look at, the numbers of parallel requests sent to the general public storage have been 10, 20 and 30. This personal and hybrid (non-public + public) cloud garage scenario changed into designed to assess the following: a) the impact of having an elastic carrier and, b) the behavior of the cloud storage infrastructure when applying different redundancy strategies for you to provide numerous ranges of information availability.

The impact of having an elastic service

As a first step, the effect of getting elasticity in the storage provider changed into evaluated compared to a static service (with out elasticity). Inside the elastic carrier, a new digital gadget is instantiated while a workload exceeds a defined threshold. The evaluation uses exceptional workloads generated by Autobench [14]. A physical machine with a single hard disk receiving an growing workload became as compared by way of making use of the same workload on a set of digital machines which might be incrementally instantiated within the equal bodily system. For this check, the workload essentially consisted of a hard and fast of requests of a dynamically generated Hypertext Preprocessor internet page. This web page emulates the processing time on a server by way

of going for walks a

sorting.Algorithm (bubble type). Seeking to emulate unique degrees of load at the server, it was defined a listing containing extraordinary portions of factors that had to be taken care of. The consequences shown in determine 2 constitute the common response time a patron acquired whilst the weight balancer simplest accessed to 1 physical device (constant line), and when the balancer accessed the equal physical machine the usage of from 1 to three VM instantiations (elastic line). It is able to be visible, when the workload is low, at the beginning of the test, how the response time presented through the static carrier (strolling most effective on one bodily machine) is higher, in some instances as much as 4 or five orders of importance, in comparison to that acquired inside the execution of the service having access to to 1 virtual machine.In this check, a maximum response time of 25 s turned into defined because the top threshold for a brand new VM instantiation. It approach that after the worldwide machine response time reaches 25 seconds, a new virtual machine may be instantiated and integrated into the garage provider. It additionally may be seen that the response time inside the elastic service has some huge falls throughout the take a look at. This behavior is not going on at the time of a new VM is instantiated, however at the time whilst the VM is included within the provider via the weight balancer. The instantiation and activation time of the brand new VM changed into among 60 and ninety seconds. When the workload increases, it's miles necessary to instantiate any other VM. For this take a look at, the elastic provider changed into capable to finish the workload supplying an acceptable reaction time, whilst the static service collapsed and couldn't finish all the requests despatched with the aid of the client. Likewise, whilst the reaction time is going beneath the threshold and continues for some time, a VM is launched. This descending hobby is monitored till a unmarried VM is running at the complete infrastructure.
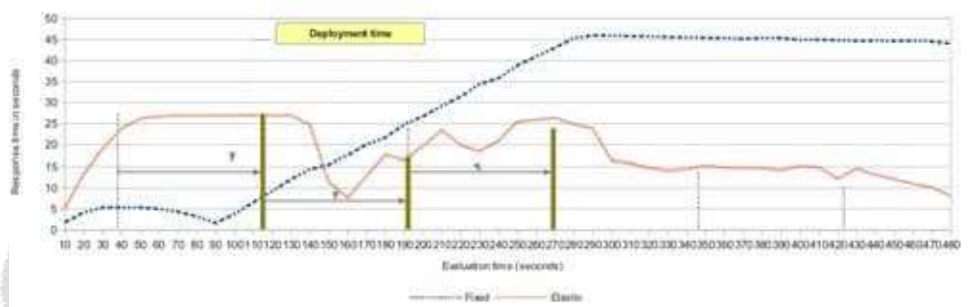


Figure 2. Performance comparison between a fixed and elastic storage service.

### *Evaluation of different redundancy techniques*

With the DAM component is possible to outline the level of information availability required inside the cloud storage prototype. It can be finished by means of applying extraordinary redundancy techniques. In this test, it become described a benchmark that allows seeing the blessings acquired of the usage of a disbursed garage machine as compared with a centralized version. Inside the first test, DAM became configured for getting access to a single disk the use of simplest one VM (emulating a centralized method) with a single document server (emulating centralized storage).

In the relaxation of the checks, it changed into continually considered a dispensed process (eight VMs) using a allotted garage gadget (5 disks that were allotted on exceptional storage servers encapsulated by way of DAM). Because the redundancy with the IDA approach is attractive for a hybrid cloud service, its conduct became in comparison in both cases, whilst it's miles most effective gaining access to a private storage cloud and while it is also having access to a public garage cloud (hybrid model).   Important metrics have been taken under consideration for those experiments: 1) reaction time: it considers the time from when the user clicks on the button to Upload or download a record till the point whilst the document loading or downloading has finished, in this test till the TCP connection is closed down. 2) service time: the time needed with the aid of DAM for finding a document (or a part of it) and getting the file geared up to be study through the gadget component this is inquiring for it.

Redundancy techniques in a private cloud

This take a look at evaluated the response and service time perceived with the aid of customers that requested exclusive tiers of facts availability and fault tolerance to the garage prototype, while jogging on a personal cloud surroundings. Special redundancy strategies (see section 2.1) have been applied in MAD to perform this purpose.The left facet of figure 3 indicates the reaction and carrier time produced by extraordinary redundancy techniques in the course of the file uploading method. In this example, even though the worst provider time changed into produced with the aid of the full redundancy method, it may be visible that users perceived the worst overall performance when they have been getting access to a centralized storage carrier. It's miles thrilling to see how the no-replication method is showing the high-quality performance in the course of the importing system.
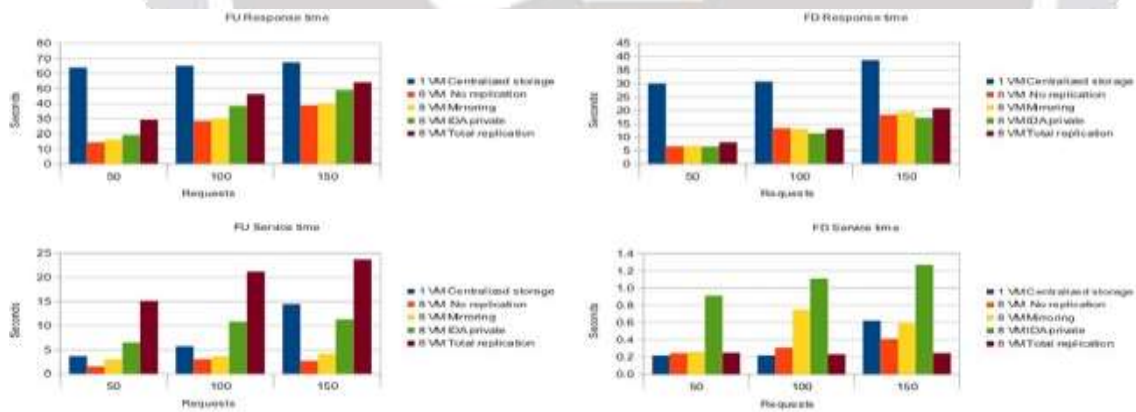


Figure 3. Average response and service time for file uploading (FU) and downloading (FD) using different redundancy techniques in a private cloud environment.

This conduct might be triggered due to the fact this technique does now not require any extra paintings for replicating a file and does not need to ship any extra statistics via the network. The right side of determine three indicates the response and carrier time perceived at some stage in the record downloading system using the private cloud. In this situation, despite the fact that the IDA method is generating the worst carrier time, the reaction time showed via the specific redundancy techniques was very similar.

IDA shows a completely aggressive reaction time and gives an appropriate level of fault tolerance. The entire redundancy method gives excessive records availability and fault tolerance, but it isn't always producing the pleasant response time. It can be caused by the way DAM is handling the dispensed disk pool. It's miles crucial

to observe that this redundancy method produces the best garage consumption.

Redundancy techniques in a hybrid cloud

The purpose of this test was to assess the conduct of the IDA redundancy technique implemented in a hybrid cloud (accessing the private and public cloud infrastructures). On this context, a reduced quantity of requests became generated due to Regulations given by way of the public storage companies.

It's far essential to word that the IDA technique could be appealing in hybrid cloud storage. IDA offers information availability, fault tolerance and a sure stage of privacy, because it does now not require a copy of a whole report to be sent to the general public cloud garage.

On this context, the response and service time perceived with the aid of users at some point of the document importing and downloading procedures have been compared. The overall performance of the version of IDA applied within the non-public cloud is taken as a reference factor.

The non-public model is in comparison to two IDA versions that get admission to every public cloud garage provider, Dropbox and Phoenix (TreeStore). The left side of figure four shows the response and provider time for the duration of the importing technique.

It may be seen how IDA suffers a high penalty whilst getting access to the outside garage (until 10 orders or importance). Even when the downloading technique (proper aspect of discern 4) showed a better overall performance,

the reaction time of IDA remains penalized whilst having access to the external garage in 6 or 7 orders of importance. This penalty at the IDA version on a hybrid surroundings is given specially by the negative net connection (it isn't a devoted hyperlink) used to send/acquire record fragments from the outside infrastructure:(garage:providers)
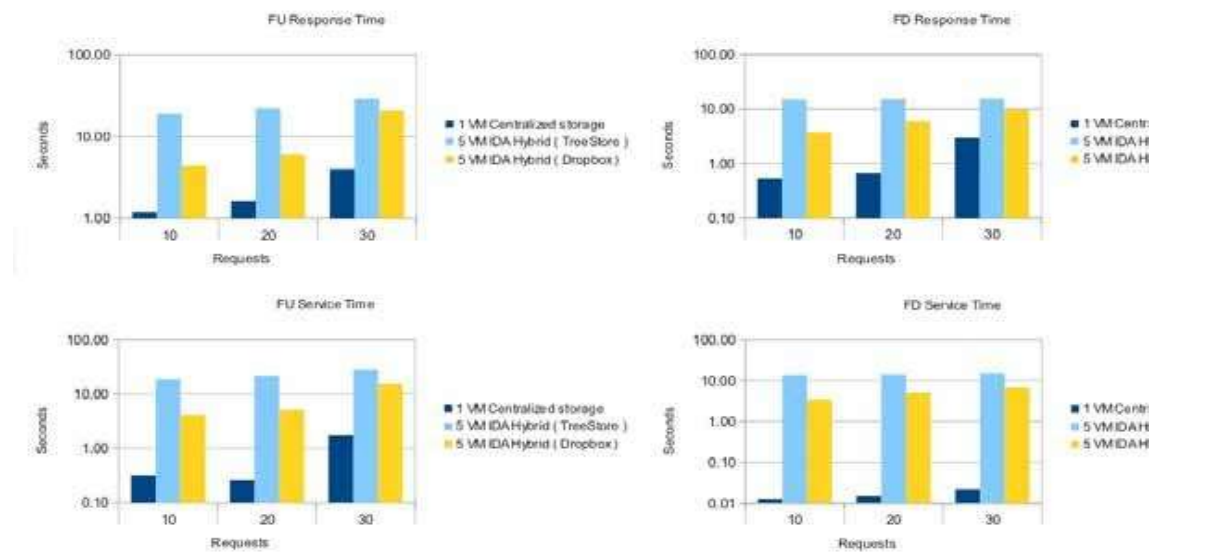


Figure 4. Average response and service time for file uploading (FU)and downloading (FD) in the evaluation of the hybrid cloud.

It's far really worth keeping in thoughts that one of the primary benefits of storing a few record fragments inside the outside infrastructure is the fact of having more storage space to be had within the non-public cloud.
It is also crucial to understand that, for protection reasons,
the range of fragments which can be despatched to the public infrastructure will never be greater than or identical to m, in which m is the range of portions required to construct the authentic file.
For checking out the behavior of this version of IDA, DAM changed into pressured to continually reap a fragment of a document from the public cloud (external vendors).
It must be cited that this is not the standard case, due to the fact in a real scenario, the hybrid model of IDA most effective could attain a fragment of a document from the public cloud inside the cases when it changed into now not able for DAM to obtain the m wished fragments from the private cloud, which means that that more than n-m disks had failed (worst case).

The two public storage vendors confirmed a comparable performance. However, the conduct of DropBox was barely higher than Treestore. It may be due to the maturity of the Dropbox API or a higher network connection to the Dropbox websites

### 3. Related work

Nowadays, Amazon S3 is considered a pioneer of cloud storage solutions. It offers to its users different storage rates, according to the amount of stored data. These rates vary depending on the data availability required by users. Data availability is related to the redundancy technique that will be used in the Amazon infrastructure [2].
There exist also solutions that take advantage of public cloud storage using redundancy techniques that were originated in RAID, for example RACS [15], which is a proxy that is located between multiple cloud storage providers and customers.

It is responsible for distributing data in a way that it provides an opportunity for clients to tolerate interruptions in a public cloud storage service or when the price for using the services is gettinghigh. It uses redundancy in order to support those possible situations. RACS offers to its users aninterface similar to Amazon S3, allowing operations such as PUT, GET, DELETE and LIST. Another proposal is HAIL [16], a cryptographic distributed system that allows file servers to provide a secure storage environment. HAIL supports the failure of any of the servers that make up the system, adding a degree of security to stored data using an approach based on the Reed Solomon error correction codes.

In addition, public cloud garage infrastructures including Amazon S3 [2], Rackspace [3], Google storage
[4] are being utilized by disbursed record structures such as Dropbox [12], Wala [17], and ADrive [18] that permit users to shop and share documents via web applications.

### 4. Conclusions

This proposal presented a contrast of various redundancy techniques that have been applied in a personal and hybrid cloud storage infrastructure. A description of the additives of this infrastructure turned into made.

It turned into validated that it's miles viable to improve the time of device deployment and overall performance when an elastic services (virtualized) is carried out on bodily machines.
The usage of the bodily machines resources can be optimized, mainly when they are strolling systems (just like the garage service) with an unpredictable workload.

The redundancy techniques evaluated on this paper had been implemented in a facts get right of entry to module named (DAM). It's miles a simple mechanism for storage consolidation on a private and hybrid cloud environment, which is capable of offer different stages of statistics availability primarily based on person necessities. DAM makes use of a lightweight algorithm for record allocation, decreasing the amount of metadata wished with low resources consumption. It's miles shown how hybrid cloud surroundings, implemented with
free

Available software program tools, may be a good answer for the ones establishments that aren't confident of storing realistic statistics in public storage clouds, and have economic and technical boundaries for building their very own large personal cloud. The prototype defined in this paper display how viable is to construct a modest personal cloud and combine it with a consolidated public cloud. On this context, this paper showed how using a redundancy method primarily based on an data dispersal algorithm (IDA) allows obtaining the advantages of the general public cloud storage with out exposing the whole content in their documents in a 3rd-party infrastructure.

### *References*

[1] John F. Gantz et al, "The Expanding Digital Universe:A Forecast of Worldwide Information Growth Through 2010", An IDC White Paper - sponsored by EMC (online). Available from: http://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf

[2] M. R. Palankar et al, "Amazon S3 for science grids: a viable solution?" DADC08, Proceedings of the 2008 international workshop on Data-aware distributedcomputing, pp. 55-64. June 2008.

[3] Rackspace Cloud Files (online). Available from: http://www.rackspace.com/cloud/cloudhostingproducts/files

[4] Jeffrey Dean, "Evolution and future directions of large-scale storage and computation systems at Google", SoCC10, Proceedings of the 1st ACM symposium on Cloud computing, pp. 1-1 2010.

[5] Irfan Habib, "Virtualization with KVM", Linux Journal. Volume 2008 Issue 166, Article No. 8. Feb. 2008.

[6] P. Barham et al, "Xen and the art of virtualization", SOSP 2003 Proceedings of the 19th ACM symposium on Operating systems principles, pp. 164-177. 2003.

[7]   B. Sotomayor et al, "Virtual Infrastructure Management in Private and Hybrid Clouds", IEEE Internet Computing, vol. 13, no. 5, pp. 14-22, Sep./Oct. 2009.
[8]
[9] Bill Childers, "Build your own cloud with Eucalyptus", Linux Journal, volume 2010 issue 195, Article No. 1. July 2010.

[10] Openstack homepage (online). Available from: http://www.openstack.org/Comparative of XEN and KVM (online). Available from: http://virt.kernelnewbies.org/XenVsKVM

[11] M. O. Rabin, "Efficient dispersal of information for security, load balancing, and fault tolerance", J. ACM 36,2 (April 1989), pp 335-348.
[12]      Dropbox (online). Available from: http://www.dropbox.com/features
[13] J. L. González and R. Marcelín Jiménez. "Phoenix: Fault Tolerant Distributed Web Storage". FTRA, Journal of Convergence of the Future Technology Research Association, vol 2, no.1. 2011 pp. 79-86.
[14]      Autobench (online). Available from: http://www.xenoclast.org/autobench

[15] Abu-Libdeh, H et al., RACS: a case for cloud storagediversity, Proceedings of the 1st ACM Symposium on Cloud Computing. Pp. 229-240.  June 2010.
[16] Bowers K. D. et al., "HAIL: a high-availability and integrity layer for cloud storage", Proceedings of the 16thACM Conference on Computer and Communications Security. pp. 187-198. November 2009.

[17] Wala, Secure online storage (online). Available from:http://www.wuala.com

[18]  ADrive, Web storage (online). Available from: http://www.adrive.com