

Cloud Storage Architecture

Gurudatt Kulkarni¹, Rani Waghmare², Rajnikant Palwe³
 Electronics & Tele. Dept
 1, 2, 3 Marathwada Mitra Mandal's
 Polytechnic. Pune India
gurudatt.kulkarni@gmail.com

Vidya Waykule⁴
 Electronics & Tele. Dept
 4 AISSM's College of Engineering,
 Pune, India

Hemant Bankar⁵, Kundlik Koli⁶
 Electronics & Tele. Dept
 5, 6 Vidya Pratishthan's Polytechnic,
 Indapur, India

Abstract— Designing storage architectures for emerging data-intensive applications presents several challenges and opportunities. Tackling these problems requires a combination of architectural optimizations to the storage devices and layers of the memory/storage hierarchy as well as hardware/software techniques to manage the flow of data between the cores and storage. As we move deeper into an era in which data is a first-class citizen in architecture design, optimizing the storage architecture will become more important. Cloud Storage Architecture is major topic in now a day because the data usage and the storage capacity are increased double year by year. So that some of the major companies are mainly concentrated on demand storage option like cloud storage. The existing cloud storage providers are mainly concentrated on performance, cost issues and multiple storage options.

Keywords- storage, private, CDMI, SLA

Introduction

Cloud computing data centers are modeled upon a simple design-for-failure infrastructure. They use low-cost, purpose-built, scalable solutions, including servers, storage systems and networking products, while still utilizing standard delivery models and massive economies of scale. Cloud computing data centers, however, do not purchase off-the-shelf systems designed for the traditional mass-IT market. These products are too expensive and include features that do not meet the cloud's unique data center environment and application requirements. These services are broadly divided into three categories:

- Infrastructure-as-a-Service (IaaS)
- Platform-as-a-Service (PaaS)
- Software-as-a-Service (SaaS).
- Storage as Service (SaaS)

1. *Infrastructure-As-A-Service*-Infrastructure-as-a-Service (IaaS) provides virtual servers with unique IP addresses and blocks of storage on demand. Customers can pay for exactly the amount of service they use, like for electricity or water, this service is also called utility computing.
2. *Platform-As-A-Service*- Platform-as-a-Service (PaaS) is a set of software and development tools hosted on the provider's servers. Google Apps is one of the most famous Platform-as-a-Service providers. This is the idea that someone can provide the hardware (as in IaaS) plus a certain amount of application software - such as integration into a common set of

programming functions or databases as a foundation upon which you can build your application. Platform as a Service (PaaS) is an application development and deployment platform delivered as a service to developers over the Web.

3. *Software-As-A-Service*-Software-as-a-Service (SaaS) is the broadest market. In this case the provider allows the customer only to use its applications. The software interacts with the user through a user interface. These applications can be anything from web based email, to applications like Twitter or Last.fm.

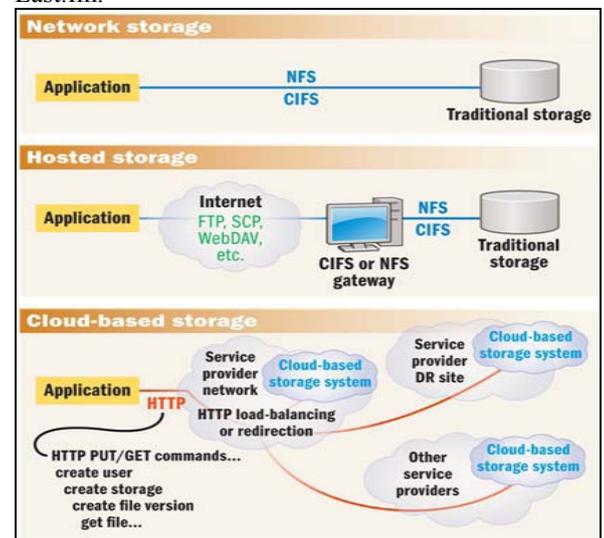


Figure 1 Evolution of Cloud Storage

4. *Storage-As-A-Service*-Storage as a Service is a business model in which a large company rents space in their storage infrastructure to a smaller company or individual. In the enterprise, SaaS vendors are targeting secondary storage applications by promoting SaaS as a convenient way to manage backups. The key advantage to SaaS in the enterprise is in cost savings -- in personnel, in hardware and in physical storage space. For instance, instead of maintaining a large tape library and arranging to vault (store) tapes offsite, a network administrator that used SaaS for backups could specify what data on the network should be backed up and how often it should be backed up. His company would sign a service level agreement (SLA) whereby the SaaS provider agreed to rent storage space on a cost-per-gigabyte-stored and cost-per-data-transfer basis and

the company's data would be automatically transferred at the specified time over the storage provider's proprietary wide area network (WAN) or the Internet. If the company's data ever became corrupt or got lost, the network administrator could contact the SaaS provider and request a copy of the data.

I. DEPLOYMENT MODELS [2]

Deploying cloud computing can differ depending on requirements, and the following four deployment models have been identified, each with specific characteristics that support the needs of the services and users of the clouds in particular ways

- A. *Private Cloud* - The cloud infrastructure has been deployed, and is maintained and operated for a specific organization. The operation may be in-house or with a third party on the premises.
- B. *Community Cloud* - the cloud infrastructure is shared among a number of organizations with similar interests and requirements. This may help limit the capital expenditure costs for its establishment as the costs are shared among the organizations. The operation may be in-house or with a third party on the premises.
- C. *Public Cloud* - The cloud infrastructure is available to the public on a commercial basis by a cloud service provider. This enables a consumer to develop and deploy a service in the cloud with very little financial outlay compared to the capital expenditure requirements normally associated with other deployment options.
- D. *Hybrid Cloud* - The cloud infrastructure consists of a number of clouds of any type, but the clouds have the ability through their interfaces to allow data and/or applications to be moved from one cloud to another. This can be a combination of private and public clouds that support the requirement to retain some data in an organization, and also the need to offer services in the cloud. [2]

II. STORAGE AS SERVICE CLOUD [3,4]

Cloud Storage as part of a storage infrastructure/concept offers companies an opportunity to meet today's demands of daily increasing data volumes. Savings in IT expenses stand in contrast to increasing expenditure for data protection and information security. Such conflict can be managed by storage management solutions that help handling large data volumes and simultaneously support numerous compliance demands. An intelligent storage management solution should be based on a multi-tier storage architecture which is a hierarchically structured manner of data storage. It also allows organizations to meet compliance demands if IT budgets are stagnating or even decreasing. Storage as a Service is generally seen as a good alternative for a small or mid-sized business that lacks the capital budget and/or technical personnel to implement and maintain their own storage infrastructure. SaaS is also being promoted as a way for all businesses to mitigate risks in disaster recovery, provide long-term retention for records and

enhance both business continuity and availability. Cloud storage is a service model in which data is maintained, managed and backed up remotely and made available to users over a network (typically the Internet). Cloud storage is amorphous today, with neither a clearly defined set of capabilities nor any single architecture. Choices abound, with many traditional hosted or managed service providers (MSP) offering block or file storage, usually alongside traditional remote access protocols or virtual or physical server hosting. Other solutions have emerged, typified by the Amazon S3 service, that resembles flat databases designed to store large objects. The Taneja Group defines cloud storage as a specific category within the larger field of —storage in the cloud solutions. Storage in the cloud encompasses traditional hosted storage, including offerings accessed by FTP, WebDAV, NFS/CIFS, or block protocols either remotely.

• *Types of Cloud Storage Systems*

There are many kinds of cloud storage solutions available today. Choosing the right kind of storage is of utmost importance. Each of these types has its own advantages and limitations. Selecting the correct underlying storage system can greatly impact the success or failure of implementing cloud storage.

A. *Object Storage Systems*

The motivation for object storage systems is simple - there is a need for having storage systems which can do more I/O computational work thereby relieving the hosts to do other processing work. Object storage mainly has two key characteristics: Individual objects and extended metadata. In such storage systems, data is stored and retrieved in the form of objects and these individual objects are accessed by a global handle. The handle may be a key, hash or a URL.

B. *Relational Database Storage Systems (RDS)*

Relational Database storage systems aims to move much of the operational burden of provisioning, configuration, scaling, performance tuning, backup, privacy, and access control from the database users to the service operator, offering lower overall costs to users. Due to this, the hardware costs and energy costs incurred by users are likely to be much lower because they are paying for a share of a service rather than running everything themselves.

C. *Distributed File Storage Systems*

It is a file system that allows access to files from multiple hosts sharing via a computer network and hence makes it possible for multiple users on multiple machines to share files and storage resources. The client nodes do not have direct access to the underlying block storage but interact over the network using a protocol. This makes it possible to restrict access to the file system depending on access lists or capabilities on both the servers and the clients, depending on how the protocol is designed.

III. CLOUD STORAGE ARCHITECTURES [3,4]

Cloud storage architectures are primarily about delivery of storage on demand in a highly scalable and multi-tenant way. Generically, cloud storage architectures consist of a front end that exports an API to access the storage. In traditional storage systems, this API is the SCSI protocol; but in the cloud, these protocols are evolving. There, you can find Web service front ends, file-based front ends, and even more traditional front ends (such as Internet SCSI, or iSCSI). Behind the front end is a layer of middleware that I call the storage logic. This layer implements a variety of features, such as replication and data reduction, over the traditional data-placement algorithms (with consideration for geographic placement). Finally, the back end implements the physical storage for data. This may be an internal protocol that implements specific features or a traditional back end to the physical disks.

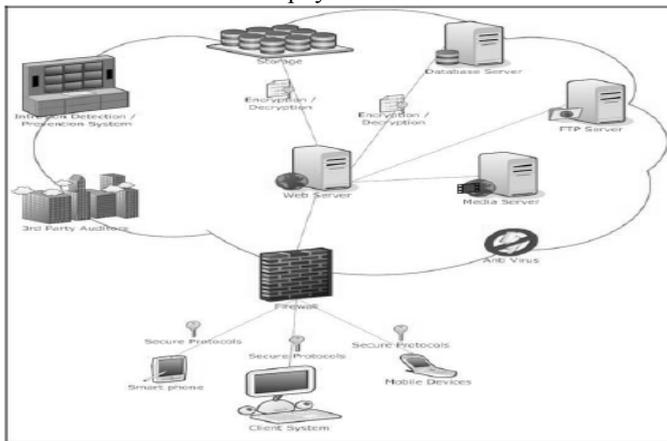


Figure 2.2 Cloud Storage Architecture

IV. STORAGE AS SERVICE CLOUD DESIGN VIEW(4,5)

A. Security

Security and virtualization are often viewed as opposing forces. After all, virtualization frees applications from physical hardware and network boundaries. Security, on the other hand, is all about establishing boundaries. Enterprises need to consider security during the initial architecture design of a virtualized environment. Data security in the mass market cloud, whether multi-tenant or private, is often based on trust. That trust is usually in the hypervisor. As multiple virtual machines share physical logical unit numbers (LUNs), CPUs, and memory, it is up to the hypervisor to ensure data is not corrupted or accessed by the wrong virtual machine. This is the same fundamental challenge that clustered server environments have faced for years. Any physical server that might need to take over processing needs to have access to the data/application/operating system. This type of configuration can be further complicated because of recent advances in backup technologies and processes.

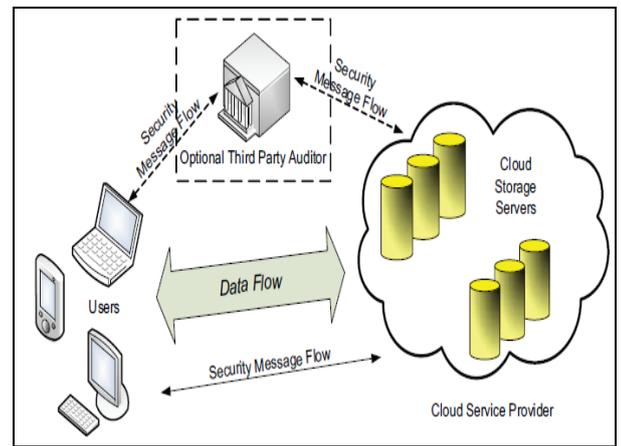


Figure 4.1 Cloud Security and Access Method

B. Automated ILM Storage

Information lifecycle management (ILM) has been at the heart of a very effective marketing campaign by vendors who sell multiple tiers of storage. Although the value proposition behind the ILM concept is simple — align the cost of storing data to the business value of the data — the real challenge comes in the actual execution of such an objective because most so-called ILM solutions are not granular enough to achieve this goal. To date, ILM has not been implemented in mass market clouds. The reason is twofold. First, the spinning media used in most clouds is usually found in the bottom tier of a typical ILM solution. With no lower tier to move data to, ILM can't be deployed. Second, the complexity and cost of implementing an ILM strategy that is granular enough to be effective has been incompatible with cloud economics. According to some industry reports, 70 percent of data is static. By storing the right data on the right media, enterprises can cut costs. Add the savings they can realize from deploying a cloud computing platform, and the financial benefits of implementing ILM in the cloud are significant. This should be possible without breaking applications or adding unnecessary complexity to operations.

C. Storage Access Method

As shown by the diagram below, there are three mainstream ways to access computing storage space. They are block-based (SAN or iSCSI), file-based (CIFS/NFS), and through Web services. Block and file-based access are most commonly found in enterprise application designs, and enable greater control of performance, availability, and security. At this point, most mass market clouds leverage Web services interfaces like SOAP and representational state transfer (REST) to access data. Although this is the most flexible method, it has performance implications. Ideally, an enterprise cloud provides all three access methods to storage to support different application architecture.

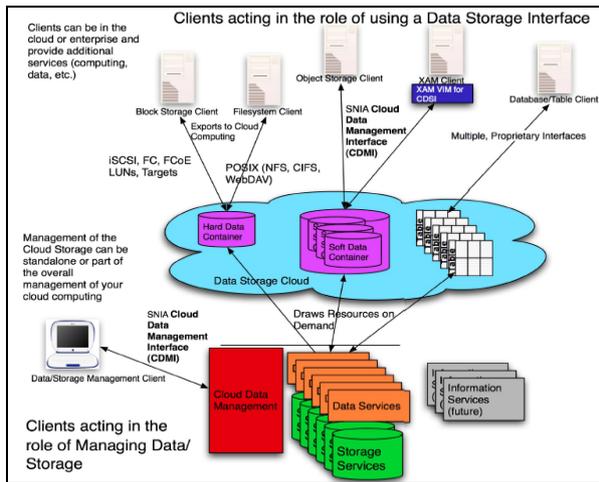


Figure 4.2. Cloud Storage reference model

D. Availability

IT infrastructure maintenance windows have largely been eliminated due to the necessity for enterprises to support users in multiple time zones with around-the-clock availability. Although SLAs are typically tied to availability, they can be difficult to measure, from a business perspective, due to the cascading effect of multiple infrastructure component SLAs. As mentioned earlier, I/O performance in the mass market cloud is often only best effort. If a cloud platform is dependent on parts of an infrastructure that are not managed by an internal IT group, then putting redundant infrastructure components and paths in place is the best way to mitigate the risk of downtime. Although cloud storage service providers continue to increase availability while watching costs, the SLAs in the current market do not meet the needs of enterprises' business-critical applications, as they often have caveats that exclude situations outside of the cloud provider's control.

E. Primary Data Protection

Primary data is data that supports online processing. Primary data can be protected using a single technology, or by combining multiple technologies. Some common methods include the levels of RAID, multiple copies, replication, snap copies, and continuous data protection (CDP). Primary data protection within the mass market cloud is usually left up to the user. It is rare to find the methods listed above in mass market clouds today because of the complexity and cost of these technologies. A few cloud storage solutions protect primary data by maintaining multiple copies of the data within the cloud on non-RAID-protected storage in order to keep costs down. Primary data protection in the enterprise cloud should resemble an in-house enterprise solution. Robust technologies like snap copies and replication should be available when a business impact analysis (BIA) of the solution requires it. APIs for manipulating the environment

are critical in this area so that the data protection method can be tightly coupled with the application.

F. Storage Agility

Storage agility simply means being able to adjust storage needs as the business requires. Ultimately, this depends on the ability of the operating system to see storage as it changes, and the access method being used. Managed Operating System (OS images provided by the cloud provider) usually have the greatest agility when it comes to increasing disk space since the drive, mount point or logical volume manager naming standard are managed by the cloud provider. Custom images (OS images supplied by the customer) can still add space but the final configure items will be up to the customer since the exact configuration of the disk space is not known by the cloud provider. Mass market cloud offerings probably address this area the best of all nine areas discussed here. Most solutions have the ability to add incremental storage in some predefined amount. Removing space is also an option, but is usually done at the volume or mount-point level. As mentioned above, the ability of the operating system to react to these changes is usually the limiting factor.

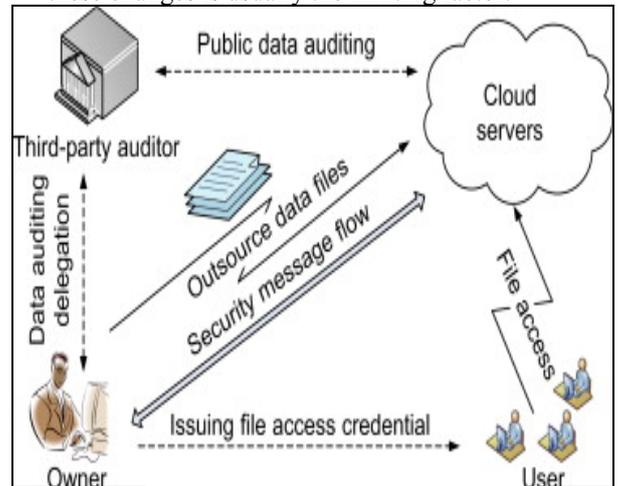


Figure 4.3 Architecture of cloud data service

G. Performance

Performance costs money. In a well-architected application, performance and cost are balanced. The key to achieving this is to match an enterprise's business performance requirements to the right technologies, which in turn requires that the enterprise translate its requirements from business language to IT metrics. Since this translation is difficult, enterprises often end up with static IT architectures that cannot meet the changing performance requirements of the business. Enterprise cloud computing provides a platform that is better suited to react to changes in performance requirements. Storage I/O in early cloud platforms typically possessed relatively high latency. That's because vendors have focused more on making the

data in the cloud readily accessible rather than improving SLAs related to performance, bandwidth guarantees, or I/Os per second (IOPs). There are two main reasons that latency remains relatively high: the type of access method and the type and configuration of the storage media being deployed. The access method consists of a combination of multiple layers of protocols (e.g., SOAP, NFS, TCP, IP, and FCP) over a physical layer of the OSI Model. Data access that includes a shared physical layer (like Ethernet) and several layers of protocols (like SOAP or NFS) generally introduces more latency than a dedicated physical layer (like Fiber Channel) running FCP. Most mass market clouds also include the Internet in the data access, which contributes to data access latency.

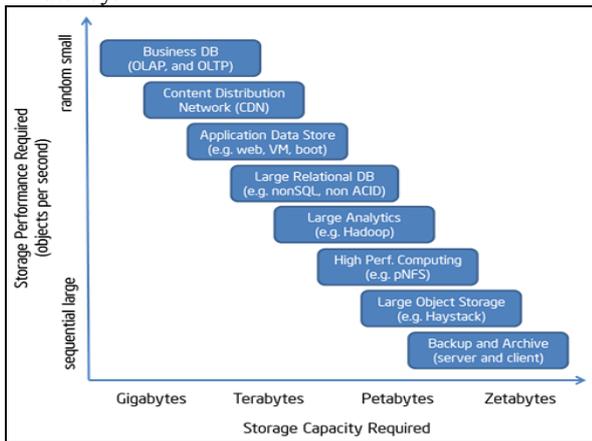


Figure 4.4 Cloud Storage Space Vs Performance

H. Scalability

Scalability must be provided not only for the storage itself (functionality scaling) but also the bandwidth to the storage (load scaling). Another key feature of cloud storage is geographic distribution of data (geographic scalability), allowing the data to be nearest the users over a set of cloud storage data centers (via migration). For read-only data, replication and distribution are also possible (as is done using content delivery networks).

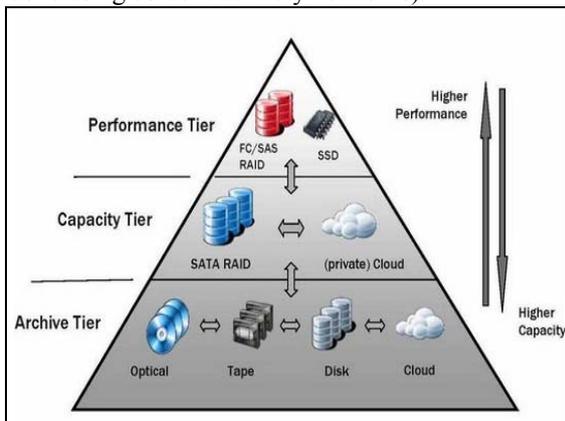


Figure 4.5 Cloud Storage variations at Layers

V. CLOUD STORAGE STANDARDS [5]

Businesses, governments, non-profit organizations and individual consumers are all facing growing challenges in storing, managing, protecting and mining the explosion of data being generated in an increasingly digital world. Cloud storage standards can help these groups address the accessibility, security, and portability and cost issues associated with the relentlessly growing pools of data. Cloud storage standards can also help define roles and responsibilities for data ownership, archiving, discovery, retrieval and shredding/retirement. Service level agreements (SLAs) around data storage assessments, assurance and auditing also must be defined in a consistent manner. Four key groups can benefit from the CDMI standard:

A. Cloud storage subscribers (users):

Service-level expectations for cloud storage security, portability, protection, performance and other criteria among different cloud storage services are best queried and compared over a standard interface. CDMI provides cloud storage subscribers with a simple, common interface to help them discover the appropriate set of compatible cloud storage service providers for their specific requirements.

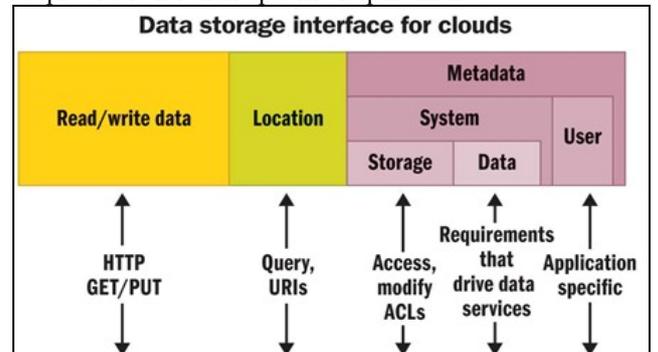


Figure 5.0 Storage Interface at Cloud

B. Cloud storage service providers:

Publishing cloud storage service capabilities via a standard interface helps ensure broad market coverage for service providers. CDMI provides a common interface for cloud storage service providers to advertise their specific capabilities and help subscribers discover them. CDMI helps service providers advertise as many or as few capabilities as required matching their targeted subscriber bases. CDMI also provides unique, non-standard extensions for service providers that want to differentiate without sacrificing broad market addressability.

C. Cloud storage service developers

Operating systems such as Windows, Solaris, Linux and Apple's iPhone have proven the value of standard interfaces for application developers. The success of the cloud will therefore depend on standard interfaces for computing, networking and storage. CDMI provides the only multi-vendor, industry-standard development interface for application developers that want to store data in the cloud. CDMI also ensures a broad infrastructure of compatible service providers for application developers, thereby creating

the broadest possible market of potential subscribers to cloud application developers.

D. Cloud storage service brokers

As subscribers entrust more important data to cloud storage providers, the need to "de-risk" the relationship between subscribers and providers becomes paramount. Enterprises or government entities may also have complex cloud storage requirements that exceed the capabilities of any individual cloud storage provider. In that case, a suite of federated cloud storage services may be required. Cloud storage service brokers can step in and offer "middle-man" services to subscribers. For example, brokers could offer "cloud insurance" via CDMI by combining a primary and secondary set of cloud storage providers to the broker's customers (subscribers). If the primary cloud storage service provider has an outage or terminates the service altogether, the broker-assigned secondary cloud storage service can take over according to the SLAs. Similarly, cloud storage brokers can use the discovery interfaces of CDMI to assemble a custom suite of services. That custom "cloud suite" would be a federation of several distinct cloud storage service providers, presented as a single cloud storage service by the broker to the subscriber.

CONCLUSION

Cloud Storage with a great deal of promise, aren't designed to be high performing file systems but rather extremely scalable, easy to manage storage systems. They use a different approach to data resiliency, Redundant array of inexpensive nodes, coupled with object based or object-like file systems and data replication (multiple copies of the data), to create a very scalable storage system. Designing storage architectures for emerging data-intensive applications presents several challenges and opportunities. Tackling these problems requires a combination of architectural optimizations to the storage devices and layers of the memory/storage hierarchy as well as hardware/software techniques to manage the flow of data between the cores and storage. While there are issues of non-uniformity across cloud vendors there is a requirement to provide uniform user interfaces and seamless integration with the mainstream desktop and server computing. Moreover, since a cloud infrastructure is a distributed system, storage facilities may be designed like the distributed file system.

ACKNOWLEDGMENT

Mr.Gurudatt Kulkarni one of the authors is indebted to Principal Prof. Mrs. Rujuta Desai for giving permission for sending the paper to the conference. Mrs. Rani Waghmare is also thankful to the Secretary Principal B.G. Jadhav, Marathwada Mitra Mandal for giving permission to send the paper for publication. We would also like to thanks our colleagues such as Lecturer Mrs. Geeta Joshi and Jayant Gambhir for supporting us.

REFERENCES

1. <http://searchsmbstorage.techtarget.com/feature/Understanding-cloud-storage-services-A-guide-for-beginners>
2. Jiyi WU^{1,2}, Lingdi PING¹, Xiaoping GE³, Ya Wang⁴, Jianqing FU¹, 2010 International Conference on Intelligent Computing and

Cognitive Informatics, —Cloud Storage as the Infrastructure of Cloud Computing□

3. Storage Networking Industry Association. Cloud Storage for Cloud Computing, Jun.2009.
4. Curino, Jones, Popa, Malviya, Wu, Balakrishnan and Zeldovich, Relational Cloud : A Database-as-a-Service for the Cloud, 2010
5. http://www.infostor.com/index/articles/display/0442659564/articles/infostor/backup-and_recovery/cloud-storage/2010/march-2010/snia-develops_standards.html
6. Storage Networking Industry Association. Cloud Storage Reference Model, Jun.2009