International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

RESEARCH ARTICLE

CLOUD STORAGE GATEWAYS AND RELATED CHALLENGES

Mukesh Chandra Negi

Project Manager, TechMahindra Ltd, Noida, India
Mukesh_negi77@yahoo.com

Abstract

Storage has always been one of the biggest pain areas in the field of IT due to high cost, need of redundancy of data and consistent growth. Especially if your business deal with the heavy contents like audio and video files, multimedia files, animated graphics and images etc. Whatever technologies or applications you can use but storage is the only source of you data and information, whether it structure or unstructured data. Data is the base of any organization, so all organizations save and keep all structured and unstructured data in multiple copies in different forms like Disk, CD, Tapes etc. but the consistent growth and demand of storage makes it's more painful area for every organization. Cloud storage has emerged as a great innovation in the field of storage where all of your storage demands fulfilled by the third party cloud service provider remotely [1] [2]. You don't need to spend huge money on storage as well don't need to worried about consistent growth and future demand, however like other cloud service delivery models, there are still certain challenges exist like security, gateways and access model, encryption etc. which is bothering lots of organizations for acceptance of cloud storage. My research article will focus on all different cloud storage models, gateways and their related challenges.

I. INTRODUCTION

Cloud storage has dramatically decreases the cost of on premise storage models. In terms of architecture and access, it's completely different from the traditional storage media, and which makes it more challenging. It's completely managed by a third party cloud storage provider and you can access your data via internet using different protocols and methods. It's completely relies on internet where you data move between your device and cloud storage provider network via public network (Public Cloud Storage), and you remotely access it from anywhere. There is no specific standard as of now, and each cloud storage service provider has implemented their own different ways to reach and access data from their hosting cloud service storages [3] [4]. For security reasons, your data is transferred over the internet using secure https protocol, and all data remains in encrypted form during transfer over the internet. Further deduplication kind of feature is utilized to optimally utilize the network bandwidth where unnecessary duplicate data is filtered to travel over network again and again.

Fig 1 Cloud Storage High Level Model

II. CLOUD STORAGE MODELS

Like different cloud service delivery models, there are three cloud storage models -

A. Public Cloud Storage

Public cloud storage is just like public cloud model where your complete storage is stored over a public cloud. It's a multitenant shared model, where you can access it from anywhere, from any device using a proper internet connection. Since its deal with the public internet, so all risks and threats are same as exist with the public cloud model [4]. All security access models are in place, but still due to public network it's a big concern especially if you are dealing with some critical and sensitive data. Since you can access it from anywhere using valid credentials and role assign to you or by you, so you need more administrative control over public cloud storage model.

B. Private Cloud Storage

Private cloud as its names sounds, it's a private cloud of storage which is especially dedicated for your organization only. It's much safer then public cloud model and mostly used by organizations deal with critical and sensitive data like bank account information's, customers other private information's etc. it can be on premises or externally hosted. Due to private ownership, you have good administrative control over private cloud storage [6] [7].

See below figure to understand High Level Difference between Public and Private Cloud based storage
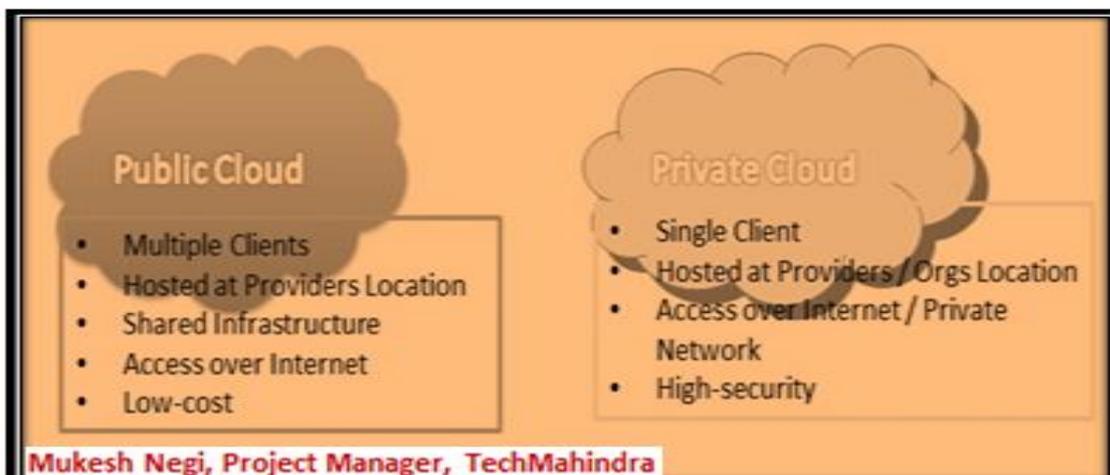
Fig 2 Public and Private Cloud Comparison

C. Hybrid Cloud Storage

Hybrid cloud storage is a combination of both, public and private cloud storage. Here you can distribute your data storage between both based on their criticality and functionality. Like critical data you can save on private cloud storage and other noncritical data you can save on public cloud storage, or based on growth and heavy requirements, you can offload heavy data to public cloud storage [8].

See below figure to understand High Level Difference between Public, Private and Hybrid cloud based storage

Characteristic	Public cloud storage	Private cloud storage	Hybrid cloud storage
Scalability	Very high	Limited	Very high
Security	Good, but depends on the security measures of the service provider	Most secure, as all storage is on-premise	Very secure; integration options add an additional layer of security
Performance	Low to medium	Very good	Good, as active content is cached on-premise
Reliability	Medium; depends on Internet connectivity and service provider availability	High, as all equipment is on premise	Medium to high, as cached content is kept on-premise, but also depends on connectivity and service provider availability
Cost	Very good; pay-as-you-go model and no need for on-premise storage infrastructure	Good, but requires on-premise resources, such as data center space, electricity and cooling	Improved, since it allows moving some storage resources to a pay-as-you-go model

Fig 3 Public, Private and Hybrid Storage Comparison

III. CLOUD STORAGE GATEWAYS

Cloud based storage is totally on the cloud storage service provider premises and an end user, to access it you have to be some mechanism to access it remotely. Customer Storage Gateways are different hardware's of software's which you can use to access storage remotely. All respective hardware's or software's existing on the customer premises who wanted to access cloud storage, all such mediator hardware's or software's are called cloud storage gateway's. There are different ways to connect and access cloud storage and almost all cloud storage service providers has implemented their own way to provide access to clients. This is one of the panic areas due to lack of some universal standards in terms of security since all of your data travel across public network if it's public cloud storage. Gateway allows incompatible technologies to communicate with each other transparently by providing basic protocol translation [9]. There are different kinds of protocols used to access cloud storage. As of now, most of the public providers reply on internet protocols and prefer RESTful API over https instead of traditional NAS (Network Attached Storage) or SAN (Storage Area Network) protocols [9] [10]. Most of the cloud storage providers some great features like deduplication and compressions technologies which greatly reduce the cost as well increase the performance and efficiency. Most of the cloud service providers have replaced gateway word with controllers as it does more work than just a gateway. If you think the difference between on premises and cloud storage is only the location of storage, then it's not true. There is a great difference between on premises and cloud based storage connectivity which makes it more complicated and challenging.

Below are some standard ways implemented by different cloud service providers to provide access to cloud storage

A. API protocols

One of the great difference between traditional storage based access and cloud storage based access is the connectivity, how end user or clients can connect to storage and access the contents. Major differentiator is the API based connectivity which most of the cloud storage based providers providing as of now rather than using traditional NAS or SAN protocol connectivity. This API based connectivity you can say is a great innovation especially for developers and programmer but in parallel it's totally incompatible and painful wherever you have legacy applications [10].

Some of the Cloud Storage based API protocols implemented by various providers as of now are below –

SOAP (Simple Object Access Protocol)

SOAP is standard set of rules for Extensible Markup Language (XML)-based message exchange. It uses transport protocols like HTTP/HTTPS (Hypertext Transfer Protocol) and SMTP (Simple Mail Transfer Protocol). Since its web based and flow securely through secure protocols (HTTPS) and firewalls, so it can communicate with different programming languages and programs over the network [10].

REST (Representational State Transfer)

REST is flexible and much easier then SOAP. It's based on completely stateless architecture which read and parses XML based web pages. It's based on client and server program, where server programs used to serve data and client programs to request for data.

FASP (Fast and Secure Protocol)

FASP is proprietary of Aspera which is used for bulk migration or bulk moving of data. One of the great feature of FASP is it's re-sent the packets those are marked as dropped only. It's an alternate solution of TCP (Transmission Control Protocol) based transport methods [11].

B. File-based protocols supported by cloud storage gateways

It's another way or gateway to make cloud storage available for the local network. Cloud storage is completely object storage based and exposed to the outer world as web service [12]. You can access it programmatically using some SOAP kind of protocols using API's, so if you are going to access it using some file based protocols, then here has to be a physical or virtual device or appliance for translations between cloud storage API protocols and file based protocols to make it accessible.

IV. CLOUD STORAGE CHALLENGES

Despite of so many benefits, why most of the organizations still has not accepted the cloud storage solution? The reason is still open limitations and challenges. Major challenge is data integrity. Since your complete data is exist over third party cloud storage service provider, so it's big challenge for most of the organizations specially those deals with critical customer and information [12] [13]. Apart from security of data, there are few more open challenges exist where some universal standards need to be implemented to make this solution under proper governance and compliance so that any organizations can accept it without any hesitations.

Some of the open concerns and challenges are as below –

Data Transfer Rate & Performance

Data transfer rate and performance till an open issues since your complete data is exist over cloud provider premises storages which you need to access remotely from anywhere, depend on the distance between you and your service providers and bandwidth limitations and latency issues, it's still a big challenge.

Data Types

All kind of data are not suited for remote access and it's a cost constraint if you are going to buy high bandwidth for the same.

Bandwidth and Latency

High bandwidth do not work in case you have some latency issues due to certain more reasons like, chattiness of the network protocols, speed of light limitations which could lead to the bad end user experience.

Hybrid Storage Requirements

If you are dealing with mix kind of data like critical and noncritical then you wouldn't be eliminate your data center storage network as you have to deal with your critical data on premises and noncritical data can be over cloud, but access and integrations in hybrid model may increase complexities and cost due to lots of other components like hardware, software etc. requirements.

Availability & Security

As I have mentioned earlier, security and data availability is still a big challenge as your complete data resides in a third party service provider premises, and which makes you insure feelings as well with respect to data availability. Certain questions in your mind would be, how secure is my data, what if my data will be corrupted, where the backup will be placed, what is the restore time, is there any chance of complete data loss etc.

Bandwidth Limitations

Bandwidth limitation over public cloud is still a limiting factor especially if you are dealing with very heavy data however vendors are still working and providing some features like deduplication and compression which decrease the required bandwidth as well as in parallel increase the performance [13].

Manageability

Manageability is also still a big concern since while you select a cloud vendor you may actually locked in yourself with a particular vendor, and there is no process or standards exist to move in from one cloud provider to another, and even you don't have any specific standards or ways to identify a cloud provider exactly according to your requirements

Protocol Translations and Interoperability

This is also one of the major concerns in cloud storage. All on premises techniques used FC(Fiber Cable), iSCSI to connect with the network storage while cloud storage protocols work only in file protocols like CIFS, NFS etc., and you can connect to both public and private cloud storages only via REST (HTTP based) or SOAP API's. So the access and interoperability issues between your existing applications and cloud make it complex in terms of adoptions and implementation [14] [15].

CONCLUSION

Cloud storage is one the great evolution in the filed IT. Large number of organizations has started using it to take benefits. But in parallel, lots of challenges have started surfacing as soon as more organizations from different sectors have started implementing the same. Major concern is data integrity and security, but still lots of implementation and performance related challenges are there as I have explained above in respective sections. Since all provider deals with their own standards and ways to implement the cloud solutions, thus it's not easy to address all challenges until and unless there is a single governance body to controls all risks, threats and concerns. There has to be some universal standard and governance body in place which should governs all the challenges with respect to the cloud storage.

REFERENCES

[1] Anthony T. Velte, Toby J. Velte, Robert Elsenpeter, Cloud Computing: A Practical Approach (McGraw Hill Publications, pp 135 – 144, 2010).

[2] C. Wang, Q. Wang, K. Ren, et al., "Ensuring Data Storage Security in Cloud Computing," Charleston,SC, United states, 2009.

[3] G. Ateniese, R. Bums, R. curtmola, et al., "Provable Data Possession at Untrusted Stores," in Ccs'07:Proceedings of the 14th ACM Conference on Computer and Communications Security, 2007, pp. 598-609.

[4] Ghemawat S., Gobioff H., Leung S., "The Google file system." Symposium on Operating Systems Principles 2003. New York: ACM.2003:29-43.

[5] Hendricks J., Sambasivan R. R., Sinnamohideen S.,et al. "Improving small file performance in object-based storage." [2006]. http://www.pdl.cmu.edu/PDL-FTP/Storage/CMU-PDL-06-104.pdf.

[6]James Broberg, Rajkumar Buyya, Zahir Tari, MetaCDN: Harnessing „Storage Clouds‟ for high performance content delivery, Journal of Network and Computer Applications, 1012–1022, 2009.

[7] Jonathan Hoppe, A Closer Look: Enterprise-Grade Cloud Storage. Online Available at: http://www.datacenterknowledge.com/archives/2011/03/17/a-closer-look-enterprise-grade-cloud-storage/

[8] Mackey G., Sehrish S., Jun W., "Improving metadata management for small files in HDFS." Cluster Computing and Workshops 2009. New York: IEEE Computer Society, 2009:1-4.

[9] Mike Hogan, Cloud Computing & Databases: How databases can meet the demands of cloud computing, ScaleDB Inc., November 14, 2008 Online Available at: www.scaledb.com/pdfs/CloudComputingDaaS.pdf

[10]Nicolas Bonvin, Thanasis G. Papaioannou and Karl Aberer, A Self-organized, Fault-tolerant and Scalable replication scheme for Cloud storage, SoCC '10 Proceedings of the 1st ACM symposium on Cloud computing, New York, USA, 2010, 205-216.

[11] Storage Networking Industry Association.Cloud Storage Reference Model,Jun.2009.

[12] S. J. Thomas Schwarz, E.L. Miller, "Store, Forget, and Check: Using Algebraic Signatures to Check Remotely Administered Storage," Lisboa, Portugal, 2006.

[13]searchcloudstorage.techtarget.com, http://searchcloudstorage.techtarget.com/definition/cloud-storage-gateway

[14]searchcloudstorage.techtarget.com, http://searchstorage.techtarget.com/magazineContent/Hybrid-cloud-storage

[15] Wenying Zeng, Yuelong Zhao, Kairi Ou, Wei Song,
Research on cloud storage architecture and key technologies, ICIS '09: Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human, ACM New York, NY, USA, 2009, 1044-1048.

176