ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645
Indian Journal of Science and Technology, *Vol 8(S8), 105–111, April 2015* DOI: 10.17485/ijst/2015/v8iS8/64230

A Performance Measurement Framework of Cloud Storage Services

Hangoo Jeon[1], Young-Gi Min[2] and Kwang-Kyu Seo[3]*

[1]Department of Management Engineering, Graduate School, Sangmyung University, Seoul, 110-743, South Korea
[2]Department of Convergence, Graduate School, Sangmyung University, Cheonan, 330-720, South Korea
[3]Department of Management Engineering, Sangmyung University, Cheonan, 330-720, South Korea; kwangkyu@smu.ac.kr

Abstract

Recently cloud storage services emerged as essential services with the spread of mobile devices have been developing into sharing contents with numerous users around the world by combining with personal cloud. In addition, they have been evolving into overcoming the limitations of mobile devices and supporting a diversity of mobile devices. Cloud storage services involve enormous investment expenses, but they are expected to be more profitable, so aggressive investments of them are being made and their service quality is being improved. Nevertheless, the performance of cloud storage services is not easy to evaluate due to the diversity of storages and their complex operating environments and so on. It is difficult to compare performance measurement results objectively and quantitatively because each service provider applies its own configuration methods and performance evaluation standards. This paper proposes a performance measurement framework for cloud storage services, which could provide objective and quantitative performance measurement information, to select a suitable cloud storage service by comparing and evaluating them. The proposed framework could help to manage cloud storage performance effectively and it is expected to contribute to market activation by providing comparable cloud storage service performance information to both providers and users.

Keywords: Cloud, Framework, Measurement, Performance, Storage

1. Introduction

Cloud computing is internet-based computing in which large groups of remote servers are networked to allow the centralized data storage, and online access to computer services or resources. Cloud computing virtualizes a wide range of IT resources such as servers, storages, networks and applications to provide services in the form of SaaS, PaaS or IaaS by users' demand. Cloud storage is a data storage service based on virtualization and cloud computing technologies, which has been used for efficiently sharing or synchronizing data between personal mobile devices such as smart phones and tablet PCs because individual possession of several mobile devices becomes common as they are popularized[1].

Recently cloud storage services emerged as essential services with the spread of mobile devices have been developing into a form of sharing contents with numerous users around the world by combining with personal clouds. They have been evolving into a form of overcoming the limitations of mobile devices and supporting a diversity of mobile devices. In this way, High Performance Computing (HPC) storage market is growing significantly in parallel with the development of related services, and the HPC storage market is expected to grow by an annual average of 8.9% during the period from 2011 to 2016 as

*Author for correspondence

Table 1. HPC storage market trend[3]

Year	2011	2012	2013	2014	2015	2016
Market Size(M$)	3,664	3,992	4,350	4,739	5,163	5,625

shown in Table 1. The global private cloud storage market is drastically increasing as shown in Figure 1. Typical cloud storage services currently provided are Apple's iCloud, Google's Google Drive, Microsoft's SkyDrive, Dropbox and SugarSync and so on[2].

Currently, cloud storage services are expected to be more profitable in spite of enormous investment expenses, so mobile device manufacturers, telecommunication companies and Internet service providers are making aggressive investments for dominating the storage market and improving their service quality. However, the performance of cloud storage services is not easy to evaluate due to the diversity of storages and their complex operating environments etc., and it is difficult to objectively and quantitatively compare performance measurement results because respective service providers apply their own configuration methods and performance evaluation standards.

Therefore, this study proposes a cloud storage performance measurement framework, which could provide objective and quantitative performance measurement information, to select a suitable cloud storage service by comparing and evaluating them. For this purpose, core technologies and performance characteristics of cloud storage services are analyzed to draw a performance metrics for cloud storage services, and to apply it to the proposed framework[5]. The research flow is shown in Figure 2. The developed framework could help to manage the performance of cloud storage services, and it is also expected to contribute to market activation by providing comparable cloud storage performance information to users.

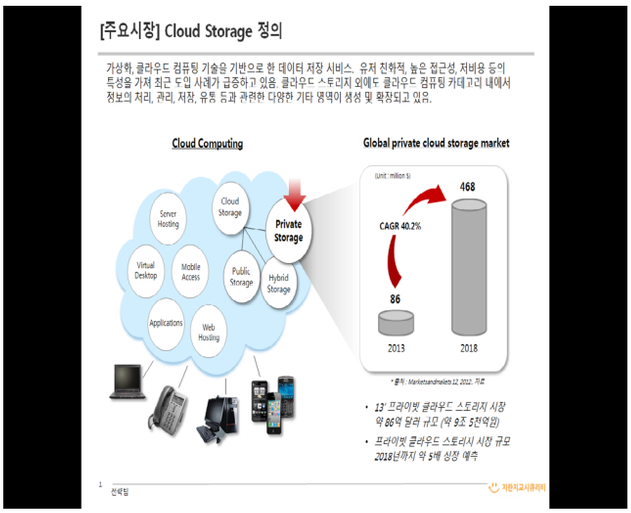

Figure 1. Global private cloud storage market[4].

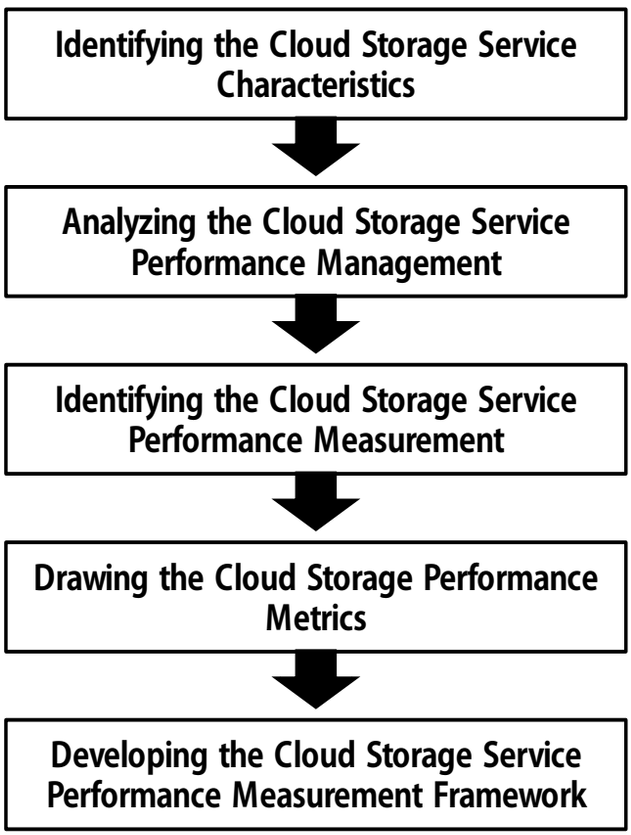

Figure 2. The research flow.

2. Cloud Storage Service

2.1 Storage Virtualization

Storage virtualization is a technology to integrate physical heterogeneous storage devices into a logical virtualized storage pool by software providing a virtualization function or separate hardware equipment and to manage them, which enables to allocate and use storages as necessary. This storage virtualization technology could raise the use rate for storage resources, and could lead to cost reduction. In addition, it could easily provide storage

scalability and availability. A virtualized storage does not really exist, but has the same characteristics as physical storage devices, so there is no need to change applications for accessing the virtualized storage, and it is used to improve characteristics of storage use rates, I/O performance, availability, cost reduction, manageability, file virtualization, file system virtualization, disk and block storage device virtualization, tape library virtualization, host- or server-based storage virtualization, storage device based storage virtualization, network based storage virtualization, object based storage, cluster distributed file system and so on[3].

2.2 Cloud Storage

Cloud storage means storage that could provide storage services via networks. The definition for such cloud storage might include beyond storage services depending on the interpretation. To build cloud storage, a storage virtualization technology is necessary. Therefore, it could be said that cloud storage provide services such as heterogeneous storage integration, data migration, backup, redundancy removal and fault recovery to users via networks.

Cloud storage has the following characteristics[3].

- Scalable on a large scale.
- Regardless of geographical locations.
- Based on commercial system.
- Price policy proportional to a capacity of the storage used or allocated.
- Easilt applicable to applications.

2.3 Major Technology of Cloud Storage

Cloud storage includes various technologies, however, the primary technologies are divided into 'connection and expansion method, storage virtualization, and storage stabilization'[3].

2.3.1 Connection and Expansion Method

In terms of storage configuration technologies, the most stable technology is the Storage Area Network (SAN) method. It connects servers and storage devices via a separate dedicated network unlike the existing method that storage devices are directly attached and accessed to servers, so multiple servers could share storage devices, furthermore, it could give faster and more stable access compared to the existing Direct Attached Storage (DAS) because it uses a dedicated network.

Another technology is the network Attached Storage (NAS), which has a structure of connecting with a network to enable file services, and it is scalable through different hardware without limit.

2.3.2 Storage Virtualization

There are three methods depending on where to place a virtualization engine for implementing storage virtualization. The first method is to directly install a virtualization engine on a storage array, which recognizes every disk in the storage array as a massive repository, and then allocates as much as necessary to users. This method should install a virtualization engine directly on a storage array, so it depends on hardware vendors in general.

The second one is that installs a virtualization engine on servers to provide integrated management of the connected storages. Because this method causes server overload, there could be more load as the number of storages is more increased.

The third is a method that installs a virtualization server between servers and storages, which is an 'out of band' scheme that uses different paths for data and control codes, respectively. To conceal complexity which is virtualization's original purpose, a virtualization engine processes too many tasks, consequently, there might be a heavy load on the engine itself, and virtualization with an in-band scheme imposes a heavy burden on the virtualization engine, therefore, an out-of-band scheme is favorable when connecting a lot of storages or storage servers.

2.3.3 Storage Stabilization

It is expected that storage products having functions for their stability in parallel with technologies making a less effect on the performance could help to enhance their competitiveness, and technologies such as redundancy, backup and snapshot are considered as being important for essential requisites of stability. In addition, encryption, compression and redundancy removal are considered to be technologies that could enhance the storage system to satisfy customers' demands.

2.4 Cloud Storage Service

As mentioned earlier, cloud storage is one of the cloud computing technologies, which could be said a technology that store and share users' data in data servers on Internet, and this cloud storage has an advantage of improving storage safety and reducing initial expenses

by enabling to use various devices as a terminal, to synchronously store data of every device and to store users' data on highly reliable servers. On the other hand, it has a disadvantage of being possible to drain personal information when the server is attacked and being difficult to store and use data according the communication environment. As mobile devices are popularized recently, users having various types of mobile devices such as mobile pads and smart phones in addition to notebooks could apply these devices to use cloud storage services. Cloud storage services enable users to access the same files from any device, and to share different types of contents such as photographs, movies, music and documents with others. Cloud storage service is a model to store data in the virtualized storage space connected via networks, which is supplied by third-party service providers in general, and they virtualize resources according to customers' requirements to offer storage spaces to customers. Customers store data by means of the storage space supplied by service providers, and the virtualized resources actually exist on several servers. It could be said a service that enables users to access the stored files through such a service anytime and anywhere if Internet is connected[6].

3. Cloud Storage Performance Characteristics

3.1 Storage Performance Measurement

To analyze characteristics for measurement the performance of cloud storage services, it is first examined how to measure the storage performance. In general, the storage performance measurement methods are defined as a characteristic on the transfer rate, and the performance measurement items are as follows.

3.1.1 Throughput

At first, the throughput measurement unit used to represent as the number of batch jobs that could be processed per day, and in the recent DB field, the system performance is measured by the number of transactions that could be processed per hour, or the number of table queries or updates. The throughput in the data communications field means the amount of data successfully transferred between two arbitrary positions over a period of time. In terms of performance, throughput could be represented as the sum of performance measure that elements organized into a system unit could make.

3.1.2 Bandwidth

It is directly proportional to the amount of data transferred or received over a period of time, and also proportional to the data complexity which represents the relevant system's performance level in the qualitative aspect. For example, a bandwidth required for downloading a page of text per second could not be equal to one required for downloading a photo file. In the analog system, the bandwidth measurement unit is Hz, which means the number of cycles changed per second.

3.1.3 Response Time

It means the elapsed time required between the last inquiry (or write command) requested to another computer system (storage or another server) by a host (or server) and the response from the relevant computer system. For example, the time interval between the time when recognizing the end of inquiry and the time when indicating the first character responded from a user terminal could be said as the response time. However, measurement user input and system response only by a result indicate on the user terminal is not reasonable from mant perspectives. In the recent system environments where various device technologies are equally developed as well as servers, it could be said that the more practically applicable response time is the "perceived response time" that could be diversely measured from multiple device systems (storage, channel extender and router etc.). The measurement unit for the response time is ms in general.

3.2 Cloud Storage Performance Management

Basic metrics for the cloud storage performance are storage capacity, Input Output Operations Per Second (IOPS) representing the number of input and output operations processed per unit time, MB/s or Megabytes Per Second (MBPS) representing I/O data throughput per unit time, and response time representing the time to complete processing a user's request. Most of storage vendors do not open results of these performance metrics, and each vendor's own experimental conditions and methods are not also opened.

Factors influencing on cloud storage performance measurement are hardware conditions such as server performance, storage connection interfaces and whether

or not to use caches, software conditions such as file systems, operating systems and applications, and workload characteristics such as whether to use random access or sequential access, read/write ratio and I/O block size. Nevertheless, their methods representing the storage performance are different, and their performance measurement environments, measurement tools or systems are presented only through some white papers, and they publish the results of performance measurement which is difficult to evaluate fairly from a user perspective. In addition, even though the performance is significantly different depending on the types of workload, it is also difficult to estimate reasonable storage resource because their own applications' characteristics could not be known in detail[7].

3.3 Analysis of user Requirements

User requirements for measuring the performance of cloud storage services are as follows.

3.3.1 Real-Time Performance Measurement

It is possible to measure performance and provide information for cloud storage services in real time.

3.3.2 Same Performance Measurement Standard

The performance should be measured by the same performance measurement metrics, tools, procedure and configuration.

3.3.3 Provide Quantitative Performance Measurement

Objective and quantitatively comparable performance measurement results should be provided for cloud storage services.

3.3.4 Geographical Scalability

When measuring performance, it should be scalable to the distributed environment to reflect geographical characteristics of cloud storage services.

3.3.5 Flexibility of Expanding the Performance Measurement Tools

To secure reliability and objectivity of performance measurement information, it should be organized to easily add and remove the measurement tools for each metrics.

4. Cloud Storage Performance Measurement Framework

4.1 Cloud Storage Performance Measurement Metric

It is to extract the system's performance measurement metrics for measuring the cloud storage service performance, and typical performance metrics are throughput such as IOPS and MB/s, and response time in addition to CPU use rate and availability[8–12].

4.1.1 IOPS

PS is defined as the number of I/O operations processed per second, and the number of random read/write operations processed is measured between about 4KB and 16KB.

4.1.2 MB/s

MB/s is defined as the number of bytes processed per unit time by the storage, which is the transfer rate of data that could be processed per unit time.

4.1.3 Response Time

It is a time interval between a user's request and the time when completing a return to the user's request.

4.1.4 Reliability

It is a measure of representing stability on how long a storage system could be operated without a break.

4.2 Cloud Storage Service Performance Measurement Framework

The storage performance measurement metrics examined in the previous section could be used as the metrics for measuring the performance of cloud storage services. However, performance measurement results could be different depending on a reference position where measuring the performance because network performance has a great effect on cloud storage services. Therefore a cloud service performance measurement framework is proposed to measure the performance of storage services in various locations as shown in Figure 3.

In the proposed framework, gauge nodes were deployed as a scalable form to measure performance in a geographically distributed environment because the per-

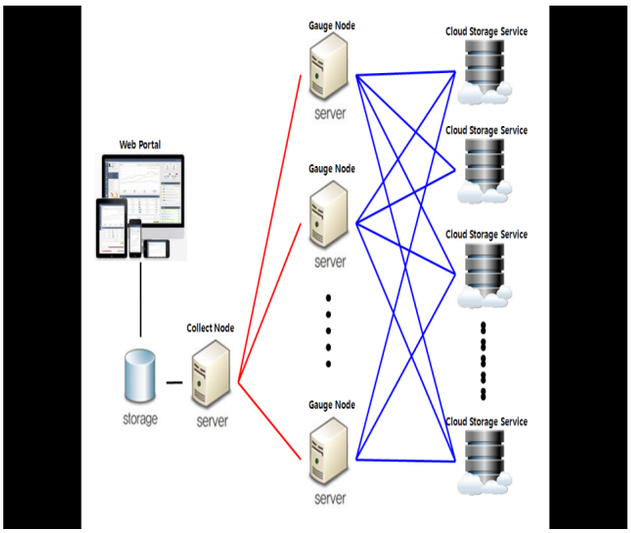

Figure 3. Cloud storage performance measurement system framework.

formance measurement results might be different by its geographical characteristics. In addition, because gauge nodes could measure performance by various benchmark tools (IOPS, COS Bench, Crystal Disk Mark and so on), they are designed as a structure of being able to easily add and remove performance measurement tools so that could eliminate dependency on tools for each performance measurement metrics, and implemented as a form that could actively respond to the performance metrics change caused by the technology of cloud storage services. The performance values measured by gauge nodes are gathered by collect nodes to analyze and process into an objective and performance-comparable form to provide it to users.

5. Conclusion

In general, cloud storage services provide a various advantages such as high scalability, availability and cost reduction based on the storage virtualization technology, and the use of cloud storage services is growing as mobile devices are recently popularized. Accordingly many IT companies offer storage services in their cloud computing, or quickly prepare for offering their services. At this point in time, this paper identified and defined the performance measurement items and metrics for cloud storage services. Finally a performance measurement framework of Cloud storage services framework was proposed to compare the performance measurement results objectively and quantitatively and it could help service providers could manage their performance, and presented a criterion of selecting services to users.

However, a result of measuring these performance metrics is one that uses a benchmark tool to artificially measure the performance of a storage stage, so it could be used for reference to the storage performance rather than evaluating the whole storage performance. In the future, an additional study is needed to develop a framework considering various characteristics of cloud storage services such as their prices and functions, and it is also expected to provide a more specific criterion of selecting cloud storage services to users by analyzing a correlation between various factors such as performance, prices and functions and so on.

6. Acknowledgement

This research was supported by the ICT Standardization program of MISP (The Ministry of Science, ICT & Future Planning) (2014-PM10-05). This research was also supported by Cloud Quality Evaluation System Operation program of MISP (The Ministry of Science, ICT & Future Planning) (KISA 2014-004).

7. References

1. Shin JB, Kim YG, Park WR, Park CI. A method for data access control and key management in mobile cloud storage service, Journal of IEMEK, Embedded Technology Dominates the Universe. 2013; 8(6):303–9.
2. Kim JJ, Han KJ. Cloud storage trends for big data, ICT weekly technical trends. National IT Industry Promotion Agency; 2014.
3. Kim YC, Cha MH, Lee SM, Kim YK. Trends of storage virtualization technologies on cloud computing. Analysis of Telecommunications Research Institute; 2009.
4. Rohan. Advanced Energy Storage Technologies Market by Type, Applications and Strategic Opportunities (Forecast to 2016). Marketsandmarkets.com; 2012.
5. Yang JW. Design of storage cloud monitoring for Qos monitoring. Journal of KEIA, Korea Environmental Industry Association. 2010; 4(3):57–61.
6. Jun CJ, Lee JH, Jeon IS. Research about factor affecting the continuous use of cloud storage service: user factor, system factor, psychological swiching cost factor. Journal of CALSEC, Society for E-Business Studies. 2014; 19(1):15–18.
7. Cloud Storage. 2013 Industry Report; 2013.

8. Ko DS. A Study on development of the evaluation guideline for the storage performance. Journal of KONI, the Korea Navigation Institute. 2010; 14(2):266–71.
9. Wang, Yaguang. COS bench adaptor development guide-Version 2.5.3. Intel; 2013.
10. Available from: https://www.storageperformance.org
11. Available from: https://www.cloud-qos.or.kr
12. Available from: https://www.kindstorage.net