CrossMark

A Case Study on Effective Technique of Distributed Data Storage for Big Data Processing in the Wireless Internet Environment

Seong-Taek Park[1] · Yeong-Real Kim[1] · Seon-Phil Jeong[2] · Chang-Ick Hong[3] · Tae-Gu Kang[1]

Abstract Technology industry is experiencing its dramatic changes in the amount of data that requires management and the place that such an asset are stored. Despite the exponential growth of information, business leaders are expecting agility in order to complete a lot of works much faster and less expensively using data. With the generalized spread of smart devices under this rapidly changing digital environment and subsequent spread of wireless Internet users and communication expenditure, the amount of data via wireless Internet worldwide is increasing rapidly at a faster rate every year. However, the communication environment cannot catch up with the demands of consumers for the super high-speed wireless Internet. Accordingly, this paper aims to look at the establishment of the cloud storage-based file system that can provide services to meet the needs of users in the cloud computing environment via wireless Internet and the examples of the establishment of such a system.

✉ Tae-Gu Kang
  snookerk@hanmail.net

  Seong-Taek Park
  solpherd@cbnu.ac.kr

  Yeong-Real Kim
  yrkim@cbnu.ac.kr

  Seon-Phil Jeong
  spjeong@uic.edu.hk

  Chang-Ick Hong
  u2n@u2n.kr

[1] Department of Management Information System, Chungbuk National University, 1 Chungdae-ro, Seowon-Gu, Cheongju, Chungbuk, Korea

[2] Division of Science and Technology, BNU-HKBU United International College, 28 Jinfeng Road, Tangjiawan, Zhuhai, Guangdong Province, China

[3] U2N, Pyeongchon-daero, Dongan-gu, Anyang-si, Gyeonggi-do, Korea

1 Introduction

With the expansion of communication expenditure resulted from the spread of wireless Internet users, the amount of data as well as wireless traffic worldwide is exploding every year.

Market research agency Wireless Intelligence reported that the number of wireless Internet subscribers worldwide in the 1st quarter of 2013 are 1.6 billion people, 23 % of the total world population of 7 billion people and the number of wireless Internet users was increasing by 41 million people on average every month, which means a 43 % increase in a year [1]. With the development of LTE and 4G technology, private data such as document, image, and high-capacity image began to be stored into mobile devices, but the storage capacity of such devices is insufficient, compared to the capacity of contents that an individual wants to store. Accordingly, service providers are beginning to provide various types of cloud service.

Gartner, a research and advisory agency in the field of IT, selected cloud computing as one of the top 10 technologies to be used strategically by the majority of companies in 2011, in the Symposium/IT expo held in 2010 [2], and every country around the world is fighting a war without gunfire to dominate the next generation communication market. If not preparing for the next generation communication environment, Korea cannot avoid paying expensive loyalty to foreign companies like now. The current 4th generation communication LTE is implementing the speed of 10 Mb per second, providing high quality customized services using cloud service, and improving security and stability as well as implementing the integration/association of multiple clouds.

Cloud storage is highly expansive because it provides virtualized storage resources meeting the needs of users via network; it is not fixed at a particular geographical location; it is based on the commercial system; its price policy depends on the storage capacity; and it is flexible to applications [2–4]. Already, many companies are using this cloud system for cost-efficient data management in line with the flow of market, and given the current trends, it is expected that the cloud computing system can be more common and users' requirement will become more varied and thus establishing cloud storage environment accordingly will be important. But, foreign software companies are encroaching on the current Korean software market. For example, the market share of domestic software in public sector is 32 % and in particular, the market share of Operating System (OS) and Database Management System (DBMS), essential factors for the utilization of information system is no more than 4.7 %. Moreover, 96 % of the domestic outer mounted disc storage market is occupied by foreign software. Like this, domestic products are gradually disappearing because the domestic software market is severely encroached by foreign products. But purchasing managers are still not relying on the domestic products although the domestic software products hold a technical capacity no less than foreign software ones, because more than half of the domestic companies are small and medium-sized companies which can't afford to have an additional Quality Assurance (QA) team.

In this respect, this paper aims to look at the establishment of cloud storage file system that has more excellent price and performance compared to the foreign software products and the relevant cases of such an establishment. The established system is expected to be helpful for the companies which could not establish cloud storage even when they need high capacity processing due to expensive cost and play an outpost role for establishing a digital society.

In this paper comparatively analyzed the performance of local and overseas SAN systems by building "A" file system. Cost effectiveness was employed as the performance measure for analysis.

This paper also examines the storage environment for implementing the cloud computing and describes how to establish cloud storage and how to utilize it. First, Chapter 2 understands wireless Internet and Cloud, Chapter 3 describes cloud storage environment and cloud-based storage, Chapter 4 looks at the examples of the establishment of storage-based file system, and finally Chapter 5 comes to a conclusion.

2 Understanding of Wireless Internet and Cloud

2.1 Cloud Computing

With the development of wireless devices and network technology, the time has come that anyone can obtain the information that they want by accessing to Internet anytime, anywhere. But, in most cases, users could simply obtain information and so for additional tasks or activities, they had to use their own PC. If you are a natural person not juridical person, you have to execute an operating system and document-preparing program and to do this, additional purchase is required. If you are a business entity, additional big costs will be incurred. A business entity has to assign terminal units and software products to every user for task efficiency, and if the number of users increases, the costs will also be incurred. Cloud computing, which appeared in such circumstances, is an attractive factor for both individuals and business entities. Cloud computing can be used anytime largely by users and so is called as aggregate of computer resources which are convenient for use [5]. All executive programs and hardware are connected by a large cloud, i.e. cloud network and so users just have to import each program at the terminal unit which provides only the function of network access. Your task contents are stored into web and can be used via the terminal unit anytime regardless of time and place.
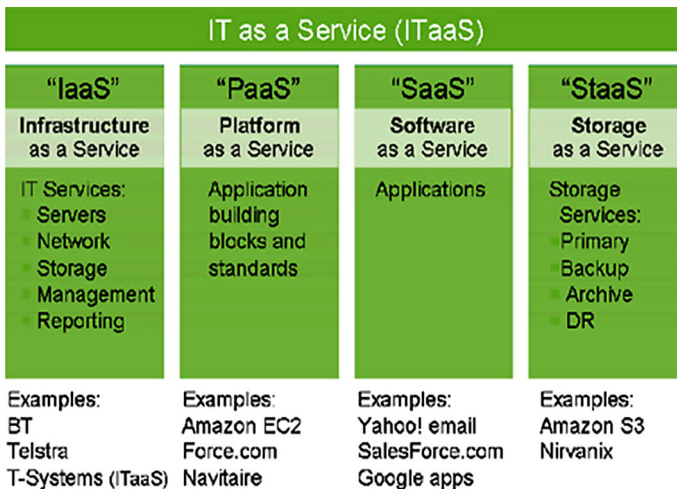

Fig. 1 Different types of cloud services that are available [9]

Chunlin et al. [6] proposes a market based hybrid cloud optimal scheduling optimization in hybrid cloud. Compared with other related algorithm, proposed MHCOS (Market based Hybrid Cloud Optimal Scheduling algorithms) achieve the better performance with acceptable overhead. Yang et al. [7] propose a proactive service handoff mechanism for MCC (Mobile Cloud Computing) using the Delaunay triangulation which is called as the DTSHM (Service Handoff Mechanism using the Delaunay Triangulation) mechanism.

The service provided by the cloud computing includes IaaS, PaaS, and SaaS, and also it includes everything required by the IT environments [8]. This is why the cloud service is also called as EaaS [9] (Fig. 1).

In SaaS, it provides software or applications for users to access the application that will be run in the cloud. In the PaaS layer, users can choose preferable operating systems and develop personal software by using supported resources in the cloud. In the IaaS layer, users can setup personal operating systems, configure computation environment, and develop software [10].

2.2 Wireless Internet Cloud Computing

With the appearance of smartphones whose accessibility to wireless Internet was improved, the private computing environment moved from the PC to the web-based mobile terminal unit and information use and communication method are also changing. As the wireless Internet was more common, the cloud computing service began to emerge, and in particular, with the use of wireless Internet through smartphones, cloud computing was combined to the existing mobile services and applied to the application and resources sharing-centered services. With the rapid spread of smartphones, there is an increasing demand for the corporate mobile services using wireless Internet in order to complete tasks such as access to corporate system, information search, file sharing, and project
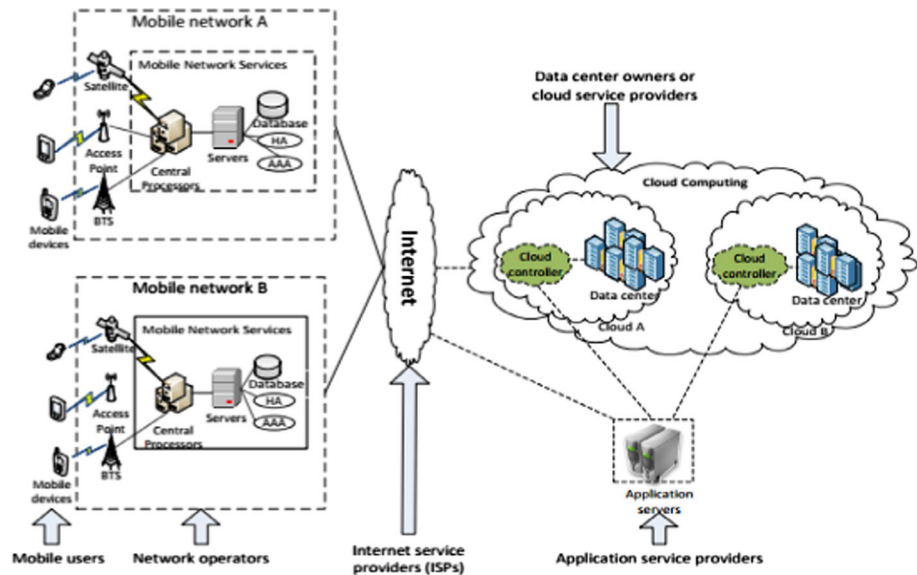

Fig. 2 Mobile cloud computing architecture [12]

cooperation. In particular, Wi-Fi and other wireless Internet infrastructure are spreading and they are establishing an environment that one can access for 24 h a day, which brightens our view of mobile cloud market using wireless Internet [11].

2.3 Mobile Cloud Using Wireless Internet

The concept of mobile cloud computing using wireless Internet is very simple. It is a combination of the cloud computing indicating that the more you use, the more you pay and the mobile service. Its services are supported from the cloud via various mobile terminal units such as smartphone, tablet PC, smart pad, note book, and net book. The components of mobile cloud in wireless Internet environment can be divided into three sub-categories such as mobile terminal unit, application, and mobile cloud as shown in Fig. 2 [12].

2.4 Mobile Cloud Storage Using Wireless Internet

The amount of data generated by the use of PC, mobile, and notebook devices worldwide is 2.8 ZB (Zeta Byte) in 2012. The prospect has come that it will increase to 40 ZB until 2020. With the rapid increase of data scale, how to deal with big data has become a big issue. A large amount of storage space is necessary to store a massive amount of big data generated from the cloud storage [13]. To store and manage big data effectively, the cloud storage system uses distributed file system to connect several servers and nodes to network, which provides high availability and expandability [14].

3 Establishment of Cloud Storage

3.1 Cloud Storage Environment

Cloud system can be divided into private cloud and public cloud. Private cloud is limited to final users or partners who are directly associated with companies, whereas public cloud can be used by all Internet users without special restrictions. As shown in Fig. 3, the key of cloud storage service is in making an appropriate infrastructure so that it can provide each service as effectively as possible, regardless of public or private cloud environment [9].
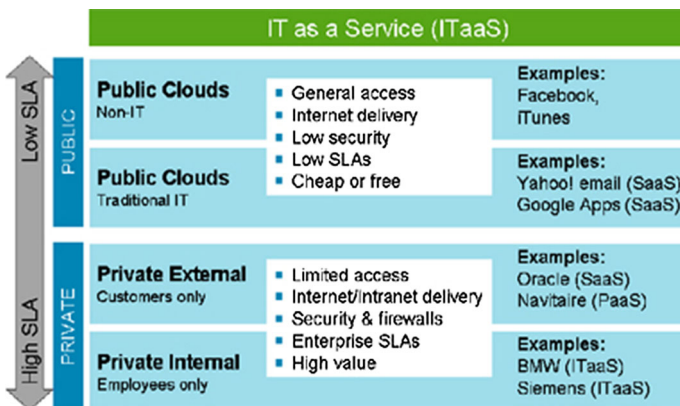

Fig. 3 Private versus public clouds [9]

Public cloud system and private cloud system have different fields and requirements for its use and every user may have different requirements for the storage service even in the same cloud environment. For example, user A may want high-performance storage service even if it has a small capacity, whereas user B may want a large-capacity storage service even if it has a low performance.

In this case, to user A, the storage volume that combines I/O interface having excellent performance such as InfiniBand and fiber channel, interface and DRAM-based SSD, and high-performance device such as RAID-based disc should be assigned, and to user B, the storage volume that combines iSCSI Interface using common Ethernet and general hard disc should be assigned. And user A and user B's access to each volume and authority differ and such related setting values should also be managed [15]. Like this, cloud storage service requires a method to manage storage effectively to meet the requirements of different users.
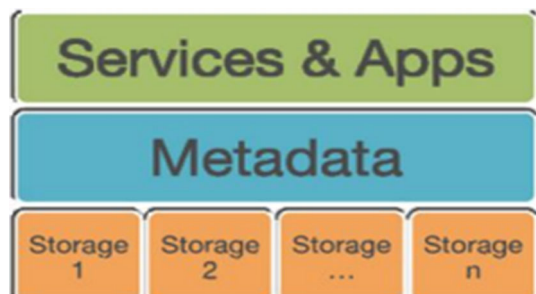
3.2 Cloud-Based Storage

Cloud-based storage is dispersed and stored and expanded according to the service requirements to improve the efficiency of the entire system. All data are copied until the final 1 bit and can be stored at a geographically separated place. The storage and maintenance of digital data is an essential part of digital library and must guarantee data consistency, integrity, and availability. To guarantee availability even when replacing the storage equipment for hardware obsolescence issues, meta data server should be separated from the dispersed storage for operation. Also, services should be separated from the storage to obtain data consistency, integrity, and availability from the logical aspects, regardless of storage disorder [16] (Fig. 4).

Data center planners should also fight with the amount of data that is increasing continuously. Entertainment industry is an important example that shows how the amount of digital assets is increasing exponentially. For example, according to Coughlin Associates 2012 report, the future demands for the digital storage in the entertainment sector for years will increase by 5.6 times and a total of 84 Exabyte storage space is expected to be necessary in 2017.

4 Examples of Establishment of Cloud-Based File System

One of the major advantages of cloud is to reduce Total Cost of Ownership (TCO) by optimizing the use of hardware as fully as possible. According to Symantec report, only 31 % of the entire capacity within the firewall is being utilized and cloud computing tries

Fig. 4 Cloud storage application (digital data maintenance) [16]

to solve such problems by integrating hardware resources, but TCO management is still not an easy task. Let's look at the cases of its establishment using cloud storage technology.

4.1 Cloud-Based File System

"A" file system is a large-scale cluster file system that can provide a large-capacity storage space of hundreds of petabyte by visualizing several to tens and thousands of servers into a single storage image. The existing enterprise-level storage is highly expensive because it uses a high-priced hardware, whereas "A" file system can be operated immediately if connecting common-use servers with Gigabit Ethernet and installing relevant software products. In other words, it has an advantage that it can provide a large-scale single value at a low cost. Capacity expansion can be easily supported without discontinuing the system already in operation. If purchasing hard disc and server as much as the capacity to add and installing "A" file system, storage capacity will increase, and as long as the network allows, you can add capacity infinitely and also expect the improvement of processing performance of the entire storage as well as the addition of capacity.

4.1.1 Composition of "A" File System

The composition of "A" file system can be expressed briefly as shown in Fig. 5. "A" file system is a system that one MGT server, multiple MDS servers, multiple DS servers, and multiple clients are connected via gigabit Ethernet, thus forming one big file system and providing services. To support the efficient dispersed storage of volume and storage pool for the purpose of supporting the expandability of storage capacity, the concept of chunk was introduced. Web monitoring server (AWMS) was provided for efficient management of network resources.

4.1.2 S/W Storage Solution and Coverage

"A" file system is an optimized software-based storage solution in small and medium-sized data utilization fields as well as wire/wireless Internet services that use large-scale

Fig. 5 "A" file system architecture

data such as cloud, User Created Contents (UCC), Internet Protocol Television (IPTV), and storage hosting. It has advantages in that users can virtualize the storage space of each server, ranging from small units to more than thousands of servers, into a single storage space if necessary; through such a virtualization, it is possible to expand the storage space up to more than the unexpectable data capacity; and resources are redistributed on the basis of a single space and the storage efficiency is maximized. Also, as shown in Fig. 6, the coverage of dispersed file storage is wider than that of SAN File System.

The development of technology acted as a pressure to force enterprises and data managers to establish a comprehensive storage plan for cost control. For example, Company "A" established high-performance architecture through a strategic investment on the hardware that formed the backbone of cloud computing infrastructure. It can also provide a higher capacity at lower costs than foreign systems by making use of the storage based on the established system and help reduce TCO by providing performance and reliability (http://www.seagate.com).

4.2 Company L's Introduction and Utilization Status

Company L, located in Korea, is a company which provides communication service. Company L established a cloud environment in line with this smart era. The type of introduced cases includes A Box, A Share, and A Home. First, A Box provides cloud-based video that can accept thousands of users, photo backup, smartphone backup/recovery, and various entertainment services as the number of the users who involve in mobile App/Web, large-capacity web hard, smartphone backup/recovery, and large-capacity sharing through wireless Internet is increasing and the amount and capacity of data is increasing.

"A" Share, a service that one can share video or photo with friends during conversation, enables us to share one-air video and photo with friends.

Large-capacity sharing enables us to share video capacity without restriction and share up to 1000 photos at one time. Free contents function provides various free contents so that they can watch them with friends and also provides sports live, free movies, etc.
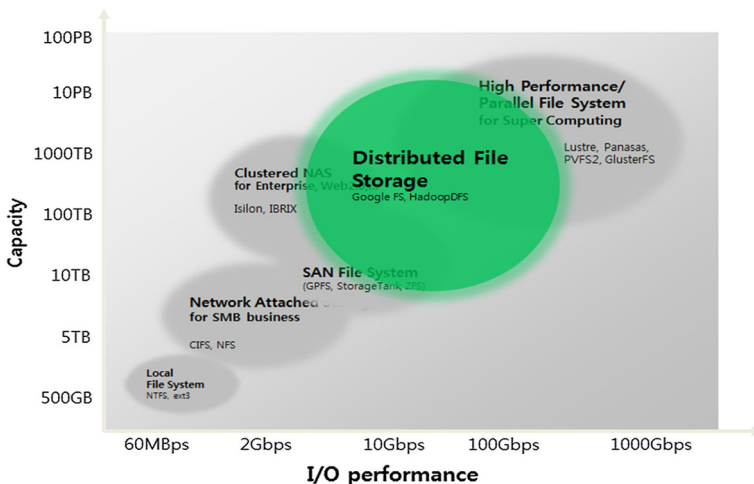

Fig. 6 Distributed file storage coverage ("A" File System)

A Home, cloud-based service that is a HD-level image quality and the real-time monitoring and invasion detection recorded videos can be stored, has such functions as real-time monitoring, invasion detection alarm, invasion recorded video, etc. Real-time Monitoring can provide real-time video monitoring at the HD level. In Invasion Detection Alarm, Closed-circuit television (CCTV) detects any movements automatically and transmits PUSH messages via smartphones. In case of invasion image records, invasion images can be automatically recorded before and after invasion detection, the cloud upload is possible, and the images can be checked immediately after it becomes cloud-based.

To provide such services, high-capacity storage needs to be introduced. Through the introduction of "A" file system for reducing the introduction/maintenance cost, stable customer service could be provided and the storage establishment cost could be reduced. Figure 7 represents the introduction and utilization of Internet and wireless Internet. It provides more excellent performance than the already-established system.

4.3 Company T's Introduction and Utilization Status

In the situations that smart devices were highly performed and varied and the applications that provided wireless multimedia services were increasing exponentially, and subsequently there was an increasing necessity of the equipment to distinguish, optimize, and offload them, Company T established an "A" file system to provide more stable service to users as wire/wireless communication solution agency.

After introducing the "A" file system, specifically the dispersed file storage solution in the mobile contents cache system, Company T used the cache system that consisted of more than 540 TB as main integrated storage, supported MDS Clustering for processing large-scale files, and brought about the performance effects such as metadata processing performance/data input and output processing performance if storage extension is necessary and scaling out of capacity.

As shown in Fig. 8, the dispersed file storage for mobile service has advantages in that if the number of file increases in large quantities, the metadata look-up load dispersion is possible and the look-up calculation support of 100,000 pieces of metadata per second is possible if the MDS Clustering (4EA) is constituted.
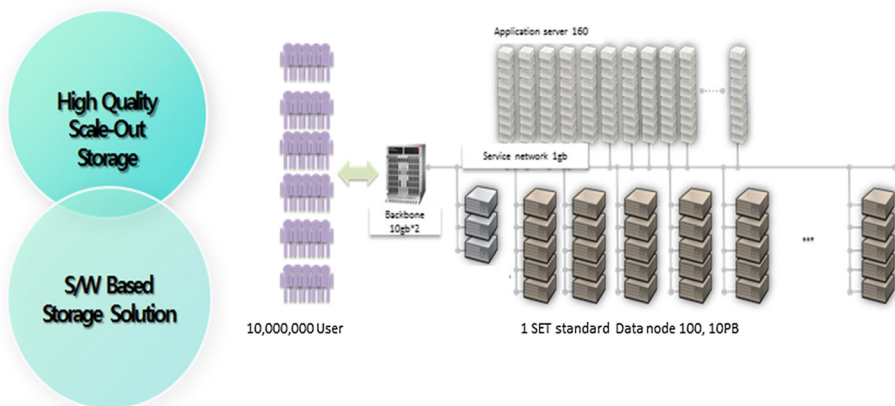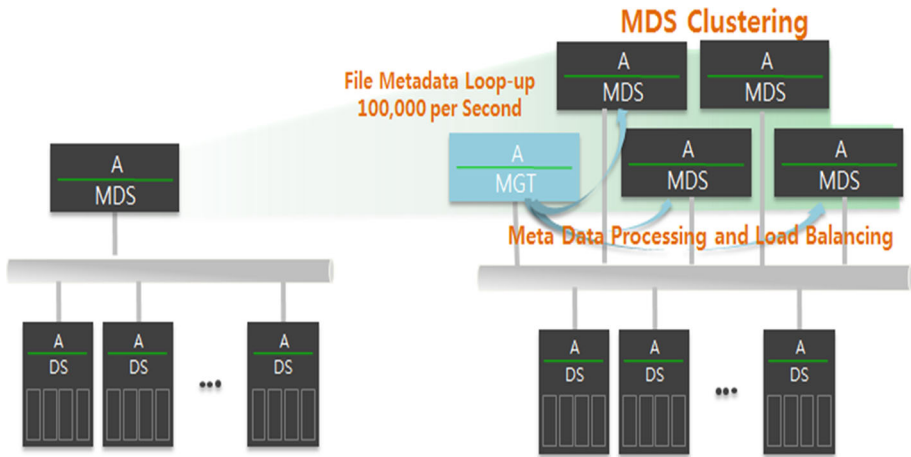

Fig. 7 Storage SET form

Fig. 8 Mobile service content cash system (Distributed File Storage)

4.4 Effects of Introduction and Comparison of "A" File System in Wireless Internet Environment

Figure 9 represents the supports for the web/command-based integration storage monitoring and management tool. First, it is possible to make an integrated monitoring of individual storage, entire storage load, and status information. And it has advantages in that it can drive, add, and delete cluster storage and systematic management and support for setting such options as storage function is possible.
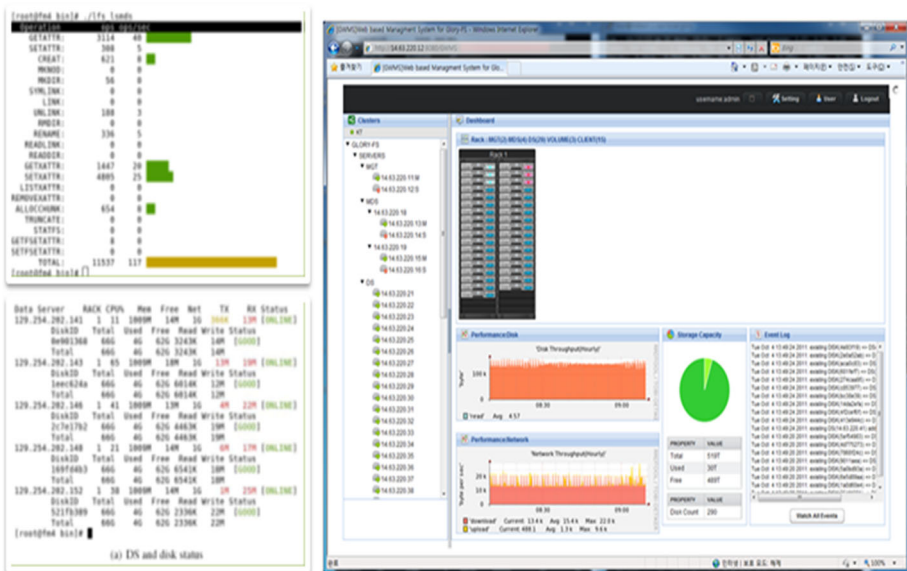

Fig. 9 "A" file system implementation effect (Mobile Environment)

The establishment of cloud storage file system through the introduction of "A" file system has advantages in that its storage establishment cost is less expensive than foreign systems, as a network storage solution that is not subject to H/W. In addition, it is applicable to the existing TCP/IP network and storage resources and common-use hardware and the solution introduction and maintenance cost is reasonable as individual/integrated license policy of "A" file system. We introduced cost effectiveness to measure performance of two systems and compared SAN system with "A" file system. In the first place, SAN system has 23,600 TCOs (Total cost of ownership) for establishing 33 TB, but "A" system has no more than 8840 TCOs for establishing 80 TB, and so it can reduce the cost by 63 % compared to the SAN system. Compared to the other parts, SNAN system has 2200 Network I/F FC HBA (11EA), but "A" system has 0, which is the part that shows the most difference (Fig. 10).

4.5 Cloud Storage Selection Method

The cloud storage selection methods are as follows [17]. First, you should not focus on the price only even if you introduce a small or medium-sized cloud system. Second, you should contact storage supplier or solution consultant to identify where walls are present as well as node and storage scale-out, size, availability, and management function. Third, you should understand that there are such factors as expandability, performance level, availability, and carbon emission in the intersecting points of enterprise-level or modular storage architecture. Fourth, you should identify the entire necessary costs for IT department beyond the cost at hand necessary for introducing it. Fifth, you should apply all costs to cloud design from various perspectives. Finally, you should approach in the way of making a modelling to the other problematic areas to be considered after modelling the hardware cost first.

Referring to the above cloud storage selection methods well is paramount, and if you want to introduce a system for your own company, the most important thing is to identify
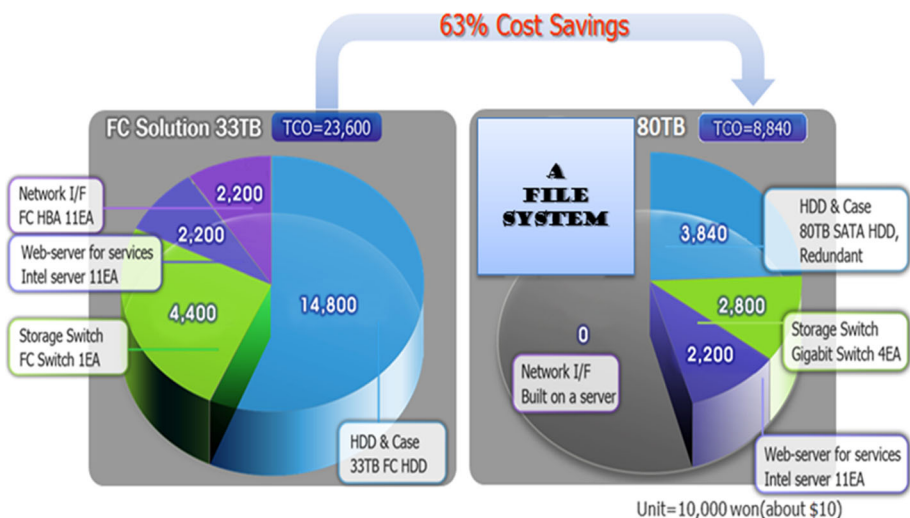

Fig. 10 Cost effectiveness (SAN system VS "A" file system)

the current status of your company, and additionally, collecting information on the cloud storage-related technology and market regularly and reflecting it into the decision making is also important.

5 Discussion and Conclusion

In the midst that the amount of data as well as wireless traffic worldwide is increasing exponentially every year with the spread of wireless Internet users in this rapidly-changing wireless Internet environment and the expansion of communication expenditure, there are continuous corporate efforts to satisfy customer complaints against wireless Internet environment. In particular, with the use of wireless Internet via smartphone, cloud computing is combined with the existing mobile services and applied to the application and resources sharing services. With the rapid spread of smartphone, there is an increasing demand for corporate mobile services via wireless Internet through which one can perform tasks such as information search, file sharing, and project cooperation by accessing to the corporate system and thus enterprises are putting more interest in more stable wireless Internet services at a lower cost.

This paper focuses on the cloud storage system utilizing the "A" file system that provides cloud storage environment in such a wireless Internet environment. This paper finds that the companies who provide service at wireless Internet environment are applying the "A" file system to provide more stable wireless Internet services at a lower cost and satisfy user's requirements and improving the user's satisfaction by providing cloud computing environment and though such efforts, they can establish a more efficient and easy-to-manage cloud storage environment.

Recently, a numerous number of world IT companies are providing cloud computing storage services or to provide these, working faster, and the prospects are coming that cloud storage will be used by individuals, corporate bodies, or various organizations. But there are dark realities that many companies in Korea are buying the cloud storage computing system from most of the foreign IT companies at an expensive cost. But as there is a domestic technology, "A" file system that can provide the cloud computing environment that evolves various information system environments to low-cost, high-efficient, and high-available next-generation computing environment, it's expected that the establishment of the cloud storage computing environment made in Korea will increase and more researches and developments will continue to be made.

The present paper proposed "A" File System for the benefit of significant cost reduction in building cloud storage over the legacy systems. The proposed system can save cost by no less than 63 % compared to existing systems. The proposed "A" File System's cost efficiency for cloud storage environment was substantiated here.

The present findings will give implications and serve as a useful guide for corporate personnel planning to build or introduce some cloud storage computing environment. Given the diversity of cloud service, the present findings could provide reference data for future studies on interoperability between services other than cloud storage.

Also, in that the proposed system is limited to cloud storage, which is one of multiple cloud services, further studies need to delve into other approaches to apply the system to various services.

References

1. Jung, G. (2013). *World population one out of every four wireless internet*, *electronics information center day trend*. http://203.253.128.6:8088/servlet/eic.wism.EICWeb
2. Gartner. (2010). *Gartner identifies the top 10 strategic technologies for 2011*. http://www.gartner.com
3. Kim, Y. C., Cha, M. H., Lee, S. M., & Kim, Y. G. (2009). Trends of storage virtualization technologies on cloud computing. *Electronics and Telecommunications Research Institute, 24*(4), 69–78. **(in Korean)**.
4. Carol, S. (2009). *Storage explained: Cloud storage defined*. http://SearchStorage.com.
5. Kim, H. Y., Min, O. G., & Nam, G. H. (2010) The technology trend of mobile cloud. *Electronics and Telecommunications Research Institute, 25*(3), 43–44 **(in Korean)**.
6. Chunlin, L., & LaYuan, L. (2015). Efficient market strategy based optimal scheduling in hybrid cloud environments. *Wireless Personal Communications, 83*(1), 581–602.
7. Yang, Z., Liu, X., Hu, Z., & Yuan, C. (2014). Seamless service handoff based on delaunay triangulation for mobile cloud computing. *Wireless Personal Communications*. doi:10.1007/s11277-014-2229-6.
8. NIA. (2009). ICT new technology paradigm: Cloud computing strategy. *CIO Report, 17*, 1–40. **(in Korean)**.
9. O'Neal, J. (2009). *NetApp, storage infrastructure for the cloud*. http://www.netapp.com/us/communities/tech-ontap/tot-cloud-storage-0509.aspx
10. Wang, Y., Chen, T., & Wang, D. (2015). A Survey of mobile cloud computing applications: Perspectives and challenges. *Wireless Personal Communications, 80*(4), 1607–1623.
11. Yang, H. D., & Hwang, S. W. (2010). A study on the improvement of mobile cloud computing services. *DIGIECO focus*. http://www.digieco.co.kr.
12. Dinh, H. T., Lee, C., Niyato, D., & Wang, P. (2013). A Survey of mobile cloud computing: Architecture, applications, and approaches. *Wireless Communications and Mobile Computing, 13*(18), 1587–1611.
13. Kim, J. K., & Kim, D. H. (2014). Data access frequency replication method using Erasure code in cloud stroage system. *Journal of the Institute of Electronics and Information Engineers, 51*(2), 325–331. **(in Korean)**.
14. Abadi, D. J. (2009). Data management in cloud: Limitation and opportunities. In *Proceedings of IEEE conference on data engineering* (pp. 1–10).
15. Choi, J. W., Eom, H. S., & Yeom, H. Y. (2011). Constructing a cloud storage system with SCST. *Korean Institute of Information Scientists and Engineers Conference, 38*(2), 127–130. **(in Korean)**.
16. Lim, T. Y. (2011). Survey on cloud computing based digital library: Focus on foreign university library use cases. *Digital Library, 64*, 60–69. **(in Korean)**.
17. Merrill, D. (2009). *Choose the method of storage TCO perspective*. Special Feature, http://www.his21.co.kr.

Seong-Taek Park received an M.B.A. degree in 2003, and the Ph.D. degree in Management Information Systems in 2010, both from Chungbuk National Univeristy (CBNU), South Korea. He is currently a Professor at The Korea Association of Software Manpower (KASOM). His present research interests include: Big Data, Patent Valuation, Patent Strategy, cloud computing, and Patent Management.

Yoeng-Real Kim received Ph.D. Degrees in Management Information Systems from University of Nebraska. He is the professor at Chungbuk National University. His research interests include ERP, MIS, Bigdata, cloud computing.

Seon-Phil Jeong received an M.B.A. degree and the Ph.D. degree in Management Information Systems in 2008, both from Chungbuk National Univeristy (CBNU), South Korea. He is a Professor at BNU-HKBU United Internation College. His present research interests include: Big Data, Computer science, ERP, cloud computing.

Chang-Ick Hong graduated from the Korea Advanced Institue of Science and Technology (KAIST) Industrial Engineering. He is currently a director at U2N Co., Ltd. His present research interests include: Cloud Service, Big Data, Archive, HPC (High Performance Computing), Enterprise Contents Management, CDN (Contents Delivery Network), IOT(Internet Of Things).

Tae-Gu Kang is getting a the Ph.D. degree in Management Information Systems at Chungbuk National University (CBNU). He received master degree in 2007 from the same university. He is currently a director at SOMA Consulting and Solution. His present research interests include: Big Data, ERP, Cloud Computing, and Wireless Internet.