

**Cheng Yan**

**CLOUD STORAGE SERVICES**

**Thesis**

**CENTRIA UNIVERSITY OF APPLIED SCIENCES**

**Information Technology**

**June 2017**

**ABSTRACT**

<b>Centria University of Applied Sciences</b>	<b>Date</b> June 2017	<b>Author</b> Cheng Yan
<b>Degree programme</b> Information Technology		
<b>Name of thesis</b> CLOUD STORAGE SERVICES		
<b>Instructor</b> Kauko Kolehmainen	<b>Pages</b> 46	
<b>Supervisor</b> Kauko Kolehmainen		
<p>Cloud computing is a hot topic in recent research and applications. Because it is widely used in various fields. Up to now, Google, Microsoft, IBM, Amazon and other famous co partnership have proposed their cloud computing application. Look upon cloud computing as one of the most important strategy in the future. Cloud storage is the lower layer of cloud computing system which supports the service of the other layers above it. At the same time, it is an effective way to store and manage heavy data, so cloud storage attract some researchers. Therefore, the research of cloud storage will not only follow trends, but also has a high application value.</p> <p>Cloud storage is a distributed file system with complicated architecture. Firstly, it is implemented on top of the cloud computing infrastructure which is based on cheap, virtualized and unreliable physical hardware, Secondly, it supports large server scale, and has efficient heavy data storage. For all of these challenge, the key technologies of the system architecture and modules design is Cloud storage.</p> <p>This article discusses the background of the development of cloud storage, gives the definition of cloud storage, describes the characteristics of cloud storage, and proposes the mode of cloud storage architecture. Exposing the key technology of cloud storage systematically and comparing functions of different storage software.</p>		

<p><b>Key words</b> Cloud computing, cloud storage, storage virtualization, distributed extend, data protection.</p>
--

<b>1 INTRODUCTION.....</b>	<b>1</b>
<b>2 CLOUD COMPUTING.....</b>	<b>2</b>
<b>2.1 Cloud computing intorduction.....</b>	<b>4</b>
2.1.1 Deployment models .....	4
2.1.2 Service models .....	5
2.1.3 Essential characteristics.....	6
2.1.4 Common characteristics .....	6
<b>2.2 Cloud storage definition .....</b>	<b>8</b>
<b>2.3 Cloud storage features .....</b>	<b>10</b>
2.3.1 Comparison of cloud storage with traditional storage .....	10
2.3.2 Cloud storage technical characteristics.....	11
2.3.3 Imapact of cloud storage on the internet .....	15
<b>3 CLOUD STORAGE CORE TECHNOLOGY .....</b>	<b>16</b>
<b>3.1 Storage virtualization.....</b>	<b>16</b>
3.1.1 Multi-Tenancy .....	17
3.1.2 Virtualization implementation.....	19
<b>3.2 Distributed extend mode.....</b>	<b>22</b>
3.2.1 Distributed file system .....	22
3.2.2 Distributed database .....	25
<b>3.3 Information life cycle management .....</b>	<b>27</b>
<b>3.4 Data protection .....</b>	<b>29</b>
3.4.1 Fragmentation storage technology .....	29
3.4.2 Erasure coud.....	30
<b>3.5 Intelligent management .....</b>	<b>31</b>
3.5.1 Dynamic expansion of resources.....	31
3.5.2 Requested load balancing.....	32
3.5.3 Resource failure management .....	33
<b>4 TYPICAL CLOUD STORAGE APPLICATIONS .....</b>	<b>33</b>
<b>4.1 Public application.....</b>	<b>36</b>
4.1.1 Dropbox .....	36
4.1.2 Google S3 .....	37
4.1.3 iCloud.....	39
<b>4.2 Enterprice application .....</b>	<b>41</b>
<b>4.3 Telecom operators cloud storage application .....</b>	<b>42</b>
<b>4.4 Cloud storage prducts comparison .....</b>	<b>44</b>
<b>5 CONCLUSION .....</b>	<b>45</b>
<b>6 REFERENCES.....</b>	<b>46</b>

## 1 INTRODUCTION

Cloud computing has become one of the hottest terms of the 21st century emerging technology, and cloud computing relate to a variety of themes occupy the main status of the mainstream media, there are a variety of books on cloud computing has also been to shelves in recent years. Cloud storage is derived from the concept and practice of cloud computing. Through the personal application like Dropbox, iCloud, Google Drive, it coming into people's live, effectively changed the people's understanding of the storage, improved the way files are stored.

Internet makes the world accessible. Cloud storage is the key infrastructure to achieve seamless information sharing and service interaction experience from different users, different applications, different devices around the world. It will become a new public infrastructure service, like water, electricity, available in anytime, anywhere. The features of high performance, high flexibility, high capacity and high security will make cloud computing and cloud storage become the cornerstone of the future of internet innovation. At the same time it makes distributed database, mobile computing, search technology, Internet of Things and other technologies developed, and increase the user experience.

As an important part of cloud computing, the cloud storage is an important innovation in the storage of industry technology and service, it will solve the problems of many users on the storage like low price, high capacity, safety and stability.

In order to provide data storage services, cloud storage employs software to interconnect and facilitate collaboration between different types of storage devices. Compared to traditional storage methods, cloud storage poses new challenges in data security, reliability, and management. This paper introduces four layers of cloud storage architecture: data storage layer connecting multiple storage components, data management layer providing common supporting technology for multiple services, data service layer sustaining multiple storage applications, and user access layer. It then examines a typical cloud storage application—backup cloud (B-Cloud),and discusses its software architecture and characteristics.

## 2 CLOUD STORAGE

Cloud computing is a hot topic in recent research and applications. Up to now, Google, Microsoft, IBM, Amazon and some other famous companies have proposed their cloud computing application, and put cloud computing as one of the most important strategy in the future. Cloud storage is the lower layer of cloud computing system which supports the service of the other layers above it. In addition, it is an effective way to save and manage heavy data. So it focused even more attentions from some researchers. (Zhang 2008)

### 2.1 Cloud computing

The revolution of the personal computer took place in 1980s and 1990s. Then in the late 90s and the early 21st century, the internet set off the second IT industry revolution. In recent years, on a global scale, cloud computing once again set off a wave of information technology change, It was the third IT industry revolution. (Gollmann 2015.)

The definition of cloud computing can be traced back to 1961. The famous American computer scientist John McCarthy who is the winner of the Turing Award put forward the concept of the cloud computing. He thought that the computing power could be used by users like water, electric, gas or other public resource. However, for a long time, this idea has only stay in the dream stage. Until the beginning of the 21st century, this idea has been known with the development of technology and the maturity of the application and 2010 is the first year of real cloud era. (Gollmann 2015.)

In March 2006, Amazon introduced the Elastic Compute Cloud. The major equipment manufacturers and service providers have also launched on demand utility computing (utility computing is a service provisioning model in which a service provider makes computing resources and infrastructure management available to the customer as needed, and charges them for specific usage rather than a flat rate.) and other similar computing services. (Gollmann 2015.)

In August 9, 2006, Google CEO Eric Schmidt at the search engine conference (SES San Jose 2006) for the first time put forward the "cloud computing" concept: Cloud computing distributes computing and data to a large number of distributed computers, resulting in a strong scalability in computing and storage, and allows users through a variety of access methods, such as computers, mobile phones and other convenient access to the network to get the application and services. (Gollmann 2015.)

In October 2007, Google and IBM began in the United States University campus, including the Massachusetts Institute of Technology, Stanford University, University of California at Berkeley and the University of Maryland to promote cloud computing plan. The plan hopes to reduce the cost of distributed computing technology in academic research, and for these universities to provide the relevant hardware and software equipment and technical support, and the students can carry out various studies based on large-scale computing through the network. (Gollmann 2015.)

In November 2007, IBM officials announced a system called "Blue Cloud" that began commercializing cloud computing concepts. Subsequently, Yahoo, Hewlett-Packard and Intel all joined the cloud computing research, and put the concept of cloud computing applications in their own products. As a result, cloud computing gradually increase, and build the elasticity of calculation, on-demand calculation, utility calculation and other concepts together. (Gollmann 2015.)

2010, the world's major IT vendors stepped up the layout of cloud computing. Some companies even let the cloud computing as a strategic core, including General Electric, about 500,000 organizations around the world use the cloud computing platform to replace the original office system. Global cloud is moving into the high-speed development track. (Gollmann 2015.)

NIST (National Institute of Standards and Technology) defined cloud computing as a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. (NIST Workshop I 2010.)

### **2.1.1 Deployment models**

NIST believes that cloud computing should have five basic characteristics: on-demand services, a wide network, shared resource pool, fast expansion and pay on demand. In general cloud computing is a concept of innovation and technology re-creation for the IT infrastructure resources' deployment, management and use. NIST also proposed a model of cloud computing framework. Through this framework, the definition of cloud computing has been better explained, as shown in Figure 1. Cloud computing framework model is divided into four layers, from top to bottom are the Deployment Models, Service Models, Essential Characteristics and Common Characteristics (NIST Workshop I 2010.)

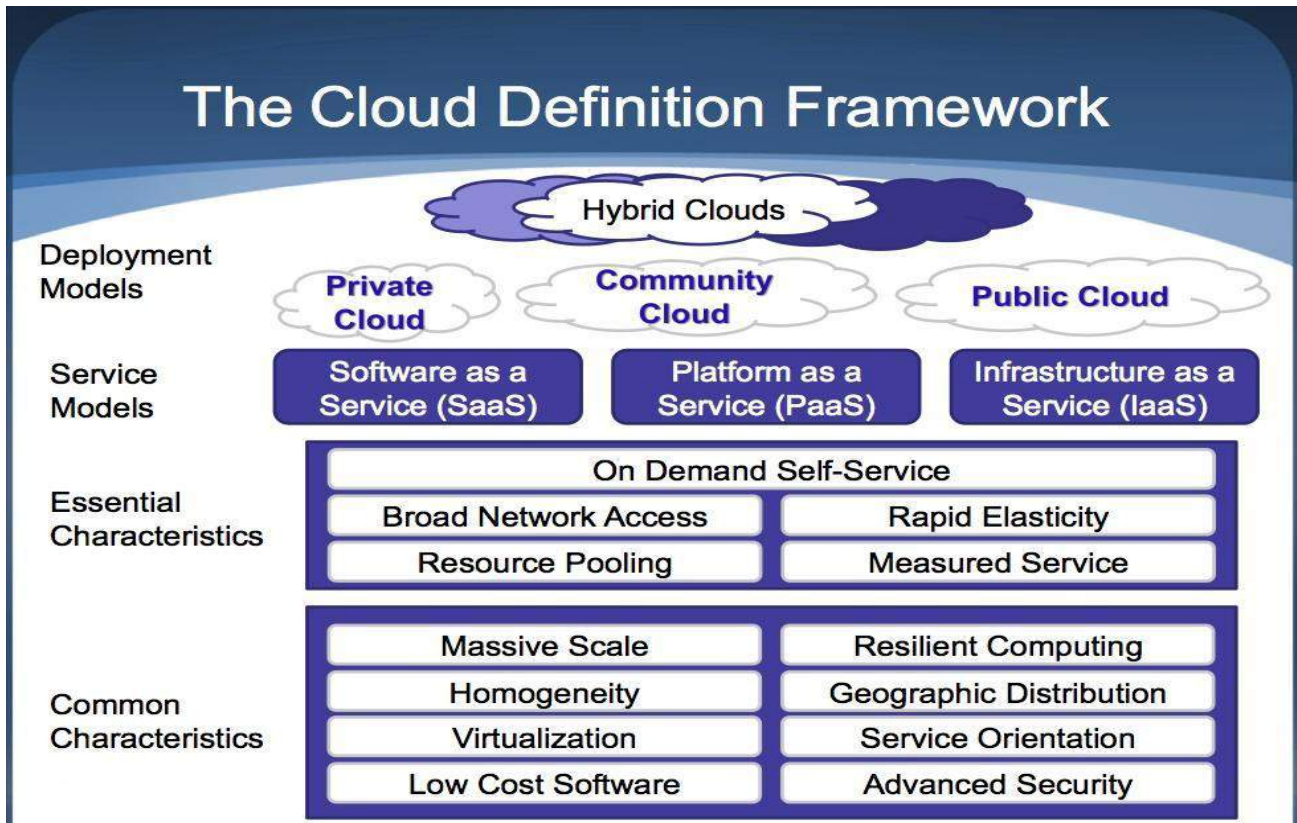


FIGURE 1. The Cloud Definition Framework (NIST SP 800-145, 2011)

NIST proposed deployment model is logically divided based on who owns and operates cloud components like data center. There are four models which are Private Cloud, Community Cloud, Public Cloud and Hybrid Clouds. (Zhang 2008.)

The private cloud is built for use by an enterprise customer alone, so it provides the most effective control of data, security and quality of service. The company has an infrastructure, and applications can be deployed on this infrastructure. Private Cloud can be deployed within the company data center firewall, It can also be deployed in a secure colocation facility. The private clouds may be within the enterprise, or on the outside. It may be managed by an company's IT administrator, or be managed by an outsourced service provider. It is also known as a private cloud because it only serves the company. (Loomis 2010.)

Public clouds typically manage and control resources by a third-party provider, it co-shared by multiple users. Hardware and technical assets such as physical addresses are outside the company. The characteristics are: high scale effect, low cost, single user data's reliability, security and controllability is poor. (Loomis 2010.)

The hybrid cloud is the product of the development of public clouds and private clouds. According to the different cloud objects, the cloud can be divided into private cloud and public cloud, in addition, the company may also have these two forms of clouds, constituting the hybrid cloud. (Loomis 2010.)

Community cloud refers to the certain area. Some of cloud computing service providers agree to provide computing resources, network resources, software and service capabilities. It consists the form of community cloud computing services, which is based on the advantages of networking within the community, easy integration of technology and other characteristics. To seek the common user needs within the community, to achieve the needs of regional users for cloud computing services. (Loomis 2010.)

### **2.1.2 Service models**

Cloud computing provides three services: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). IaaS infrastructure hardware resources provide cloud service vendors to users with cloud services. The most typical examples of applications in this area are Amazon Web Service (AWS), IBM, VMware. They are available as an infrastructure rental service. Elastic cloud in the field of IaaS propose to achieve on-demand characteristics, so users can choose their own needs of the basic resources, and only pay the corresponding costs. (Zhang 2008.)

PaaS provide a service as computing, development environment and other platforms. Users do not have to buy a server, use the platform to develop their own applications, it provided by the server and the Web for other users. The most typical application case in this area is Google's App Engine and Windows Azure. Both of them are served as a distributed platform. (Zhang 2008.)

SaaS (Software as a service) is different with PaaS, SaaS is more targeted that providing computing or storage resources service. And SaaS is also different with IaaS which provides a running environment for users to customize the application SaaS only provides some special services for application calls. (Zhang 2008.)

SaaS Provider deploys unified application software on the server according to their actual needs. Customers can order the necessary application software to the server through the Internet service provider. In this mode, the user can order the software without the need for maintenance of the software anytime anywhere. Service provider solely responses for the management and maintenance of the software, and also provides off-line operation and local data storage to users. For small and medium enterprises, SaaS



provides one of the best ways to use advanced technology. It completely eliminates the need of enterprise purchase, construction, maintenance of infrastructure and applications. SaaS orients multiple users simultaneously, but each user has an exclusive feel like the service. (Zhang 2008.)

### **2.1.3 Essential characteristics**

On Demand Self Service means that it according to user needs, the cloud give automatically to users with the appropriate performance. No service provider intervention is required. Broad Network Access means it use the broadband or mobile network provides the standard network and Internet access mechanisms to different kinds of terminals like mobile phones, Tablet and PC. (Davey 2009.)

Resource Pooling means the cloud computing service provider move the calculation, storage, network resources pooled into the resource pool, through multi-tenant mode to share to a number of consumers. The dynamic allocation according the needs of consumers on different physical resources and virtual resources. Rapid elasticity means the rapid expansion of resources and ability to shrink. From the user point of view, service providers have unlimited resources. Company will pay for the resources that they use. (Davey 2009.)

Measured service means that the resource optimization and control capabilities can be measured. At the same time, users can find relevant measurement data. It greatly improves the transparency of the service. (Davey 2009.)

### **2.1.4 Common characteristics**

Massive Scale is one of the cloud storage characteristics which means that the cloud computing systems use the large-scale expansion capabilities. For example, Google cloud computing has more than 100 million servers, Amazon, IBM, Microsoft, Yahoo and other cloud computing resource pools have hundreds of thousands of servers. Cloud computing can give users unprecedented scale and computing power. Resilient computing is one of the cloud storage characteristics which means in the cloud computing like the physical size of the resource and the logical scale can be dynamically scaled, it get near-linear performance expansion. It will meet the needs of applications and user growth. (Boss 2012.)

Homogeneity is one of the cloud storage characteristics which means that the scale of cloud computing makes the software and hardware used in cloud computing tend to homogenize. This feature can be

managed to achieve better security automation, such as configuration control, vulnerability detection, security audits but on the other hand, homogeneity also let the cloud computing services to appear in all users who use the service. Thus affecting large number of users and services. (Boss 2012.)

Geographic distribution is one of the cloud storage characteristics. In the cloud computing area, resources can be across the data center, cross-regional distribution. The scale of cloud computing breakthroughs in a single radio limit. It allows the user to roam the state to access the resource. (Boss 2012.)

Virtualization is one of the cloud storage characteristics which means cloud computing allows users to use different terminals at any location to enjoy the application services. The requested resource comes from "cloud". The application is running somewhere in the "cloud", but in fact the user does not need to understand the specific location of the application. People use a laptop or a mobile phone can achieve every storage they need through the network service. (Boss 2012.)

SOA (Service -oriented Architecture) is one of the cloud storage characteristics. It is the basic concept of cloud computing to achieve open architecture. SOA as a service-oriented architecture, is a model and methodology of a software architecture design. SOA using the existing company's software system to reorganize new software architecture. This software architecture will realize business changes, combined with existing services. It create a new software and service the entire enterprise business system. (Boss 2012.)

Low cost software is one of the cloud storage characteristics which means that in the existing cloud computing architecture, the service providers try to use the low-cost, high-reliability software to build cloud computing systems and the users can fully enjoy the low-cost advantage of the "cloud". It just a few hundred dollars a few days to complete works which needs tens of thousands of dollars and few months to complete. (Boss 2012.)

Advanced security is one of the cloud storage characteristics which means that in the cloud framework, there are four kinds of deployment models, in additional to the private cloud, the other three models public cloud, community cloud and hybrid cloud need to use the public network. Some information used storage into the PC or server, however we can put it into the cloud right now. Thus, how to ensure the safety of users account, if the application could move from one point to another point. Cloud service providers will see the details. All of these issues effect people's decision to use cloud computing, therefore, cloud computing service providers should start from both of technical and management aspects to

build a security system for services. When the users choose the cloud service, they have to confirm that the service provider has already received. (Boss 2012.)

## 2.2 Cloud storage definition

Up to 2016, the definition of cloud storage is still not very accurate, many people think that cloud storage is the network disk like Dropbox, Google Drive. But the network disk is just one of the forms of the cloud storage which is the closest expression to the public. It stores the user's file data to the network that achieve data storage and backup. It meet user's purpose of data storage, using, sharing and protection. Some people think that cloud storage is a kind of document network storage, such as Evernote notes storage services. Many people think the cloud storage is a system device to provide data storage and business access functions. The device is achieved through the cluster applications, network technology or distributed file systems and other functions. Cloud storage doesn't have the authoritative definition of the industry, but the industry has reached a basic consensus on the cloud storage. Cloud storage is not only a storage technology or equipment, but also a service innovation. (Colantoni 2010.)

The definition of cloud storage should have the following two parts. First one is that in the service-oriented aspects of the user, it provides on-demand service application model, users can connect to the cloud through the network, to save user's data in cloud storage anytime, anywhere. Secondly, in terms of build cloud storage service, it achieves massive, resilient, low-cost, low-power shared storage resources through the distributed, virtualized, intelligent configuration and other technologies. (Colantoni 2010.)

In the past 5 years, with the rapid development of cloud storage technology and applications, it has broken through the single point of the original IaaS layer definition. It formed a technical system, contains three layers of cloud computing infrastructure services (IaaS, PaaS, SaaS). At present, cloud computing services are mainly concentrated in the IaaS and SaaS layers. (Zhang 2008.)

The contents of IaaS and SaaS are not the same. From the perspective of IaaS, cloud storage provides a service for data storage, archiving, and backup. From the SaaS point of view, cloud storage service is very diverse the service has online backup document notes save network disk business photo preservation and sharing home video. Just like the IaaS business to provide cloud storage. Service providers like Amazon's S3, however, there are more cloud storage provider try to sell the SaaS business, such as Evernote and Google Docs. (Loomis 2010.)

Here is an example of using a cloud to store as IaaS services to help people understand the concept of cloud storage. When an enterprise builds a business platform, provider does not need to buy a large number of physical storage equipment. People try to order the IaaS service on the website. The cloud storage service has high level of the reliability and security. It running the order in 10 minutes after user paid. Enterprise and enterprise users can quickly access storage resources. Enterprises can also enjoy the purchase of storage services, such as multiple copies of data, hot data to speed up access and flexible policy configuration. At the same time, storage resources can be flexibly expanded based on how many spaces they have use. The enterprise pays the corresponding fees based on the actual use of storage space. By using cloud storage, the enterprise has the following benefits. Saving the cost of mining equipment. Reducing the human and resource costs to maintain storage equipment. (Loomis 2010.)

On the other hand cloud storage company get lots of benefits from the cloud management. After the integration of storage resources. The extra storage space lease to companies, not only save the resources, but also reduces operating costs. Let the user to deploy a remote storage of resources, changing the user's experience with storage device deployment. Cloud storage virtualization and intelligent management technology enables service providers to perform simple and efficient operation and maintenance for the cloud storage systems. (Loomis 2010.)

## **2.3 Cloud storage features**

Cloud storage is one of the best ways to keep business's important data safe and accessible. Many small and midsize businesses are turning to cloud storage services to keep files and documents backed up and available from any Internet-connected device — even when those files contain sensitive company and client data. (Loomis 2010.)

### **2.3.1 Comparison of cloud storage with traditional storage**

Because cloud storage is a new product in the field of storage, it will inevitably lead to a comparison with the traditional storage. There are many differences between them. The comparison of cloud storage and traditional storage in Table 1.

TABLE 1. The comparison of cloud storage and traditional storage. (Loomis 2010.)

Item	Cloud storage	Traditional storage
------	---------------	---------------------

Architec- ture	Not only a structure, but also a service. The underlying use of distributed architecture and virtualization technology, Easy to expand, single point of failure does not affect the overall service.	The architecture is a dedicated, specific hardware component that is used for a particular application
Service mode	On-demand use, according to the use of billing, service providers can quickly deliver and respond.	The user through the machine to buy or rent the storage capacity.
Capacity	Support for infinite expansion above the PB level	For a particular application storage, by the application requirements to define capacity, difficult to expand,
Data man- agement	Not only to provide the traditional way of access, but also to provide massive data management and external public service support, Using the strategy to protect data in the same time, such as chip storage, EC, ACL, certificate and other multiple protection strategies and technologies. Users can be configured flexibly.	The user data manager can see the information and is not safe enough. Usually use RALD to provide data protection, Users can not flexibly configure personalized storage policies and protection policies.

Cloud storage does not simply be seen as an architecture, it rather a service. The underlying use distributed architecture and virtualization technology. The service provided by hardware and software virtualization. Traditional cloud storage architecture is works in a particular area and it used for specialized and specific hardware components. This architecture includes the server, disk array, controller, system interface and other components, and it for a single service. (Loomis 2010.)

On-demand model means that according to the use of billing, it has some different compare cloud storage with traditional storage. Users can spend more time to enjoy more resources and services. The traditional storage business model is that the user needs to purchase the relevant package fees. The price is based on the specific rules set by the service provider or pay the hardware and software costs, even require additional software royalty fees and hardware maintenance related costs. (Foster 2010.)

Cloud storage with huge storage capabilities also has very good scalability performance. As a result, support is provided and can be extended to the PB level storage service as needed. The traditional storage through a dedicated array can also be achieved the capacity of the PB level, but its management and maintenance will be very difficult, and the cost is quite expensive. (Foster 2010.)

Cloud storage designers consider data management. They consider how to ensure data security and availability, therefore, a variety of protection strategies and techniques, such as erasure code (EC), Secure Sockets Layer (SSL), and access control lists (ACLs) are used to protect data security. Data in the cloud storage is distributed storage, but also take relevant backup technology and algorithms. Therefore, to ensure the reliability of data, data recoverability and scalability of system flexibility and other characteristics, while ensuring data availability and integrity in unpredictable hardware damage, data loss, the service will not stop in that time. Traditional storage uses more technical measures to ensure data availability, and the location of the disk to which the user data belongs is also known to the service provider that means there is a risk of information security. In addition, the general storage in the system upgrade, often users are told that their data is suspended. (Foster 2010.)

### **2.3.2 Cloud storage technical characteristics**

Reliability means that the cloud storage systems typically provide storage services for mass users. The reliability is very important significance for all users, and even the society functioning. Cloud storage system usually contains a large number of ordinary commercial hardware and software components that face various risk of failure. Therefore, under the premise of ensuring high reliability. It improve the overall efficiency of the cloud storage system being a problem that must be solved. (Foster 2010.)

The actual realization of data analysis and the establishment of statistical models to understand the hardware and software failure rules. According to the different services needs to design a variety of new coding mode, then try to set up the functional area in system. It has fault tolerance and refactoring performance and other characteristics. It will increase the overall data availability of the cloud storage system, while improving the overall operating efficiency of the system. (Foster 2010.)

Availability means that the businesses need to support users in different time zones and ensure 24/7 availability. Currently, I/O performance is considered first in most cloud storage platforms on the market. For the cloud storage platform, the redundant architecture part. Currently, cloud storage service providers to increase the availability of services while considering costs. However the current service level agreement on the market can not meet the needs of enterprise critical applications. Storage options in cloud storage, including multipath, controllers, different fiber networks, and improve cloud storage availability. (Foster 2010.)

In the past, security refers to the surrounding security, to make sure that no unauthorized access is allowed around. For the virtual world, a physical perimeter has disappeared. Therefore, the enterprise must assume that all transmitted data may be intercepted, there is no physical control on a system, and the execution of these rules must rely on other methods to restrict access to information. Encryption is an important way to restrict access to meaningful information. Therefore, when IT services are delivered over the cloud, encryption becomes an important issue of security, and it is very challenging because the data must be stored and saved in encrypted form. If the secret key itself is lost or damaged, the data itself will be lost or damaged. Therefore, data debris as an alternative, which enable the privacy of user data. (Foster 2010.)

In April 2010, SNIA announced the cloud storage standard - the CDMI specification. SINA said that CDMI is a direct specification, it provides a way for data centers to the cloud storage. Data center access to existing network storage resources should be able to switch to CDMI cloud storage resources fairly easily and transparently, however, the inadequacies of the CDMI specification does not provide reliability and quality to measure the quality of cloud storage service providers. It can not prevent the existence of such a risk of data loss. At the same time, the CDMI specification has not been widely used in the industry. The existing cloud storage service platform on the market, such as Amazon S3, Microsoft Azure and Google Drive, have adopted their own private interface specification. (Foster 2010.)

Cloud storage has the ability to reduce enterprise-class storage costs by using its capabilities and services. This includes the cost of purchasing storage, the cost of driving the storage, the cost of repairing the storage (when the drive fails), and managing the cost of storage. An example of a cloud storage solution is that a company called Backblaze started to store low-cost storage products. A Backblaze POD (storage rack) in a 4U chassis with 67TB of data packets, the price of less than 8,000 US dollars. This packet contains a 4U chassis, a motherboard, 4GB of DRAM, four SATA controllers, 45 1.5TB SATA hard drives and two power supplies, on the motherboard. Backblaze runs Linux (with JFS as the file system), using HTTPS and Apache Tomcat. Backblaze's software includes deduplication, encryption, and RAID6 for data protection. Therefore, cloud storage and services can significantly reduce the cost of enterprise-class storage, cloud storage has become a viable and cost-effective choice. (Foster 2010.)

The advantages of cloud storage is that the equipment of cloud storage has a large number of storage devices, distributed in different regions, cooperation between multiple devices. Many devices can provide the same service for an individual at the same time. Cloud storage as a platform service, cloud

storage vendors is based on user needs to develop a variety of platforms, such as IPTV application platform, video surveillance application platform and data backup application platform. As long as there is a standard public application interface, any authorized user can log on to the cloud storage system through a simple URL and use the cloud storage service. (Loomis 2010.)

Cloud storage capacity allocation function is not controlled by the physical hard disk. It can be in accordance with the needs of customers are expanded in an acceptable framework. Equipment failure and equipment upgrades will not affect the user's normal access. (Loomis 2010.)

Cloud storage technology take a different approach to data importance, and copy the file stored in a different server, therefore, when subjected to hardware damage, either a hard disk or server, the service does not always terminate. Because of the use of the index structure, system will automatically read and write instructions to other storage nodes. Read and write performance is completely unaffected. Managers only need to replace the hardware, data will not be lost. Put on a new hard disk server, system will automatically copy the file back, it always keep the backup file. In the expansion, first installed storage node is connected to the network, and new capabilities are automatically incorporated into the storage. (Loomis 2010.)

The cost of traditional storage mode, once the completion of the one-time investment of funds. system can not be used in the follow-up dynamic adjustment. With the upgrading of equipment, poor hardware platform is difficult to handle. As the business requirements change, the software needs to constantly updated and refactor. The cost of maintenance is high and ultimately uncontrollable. (Loomis 2010.)

For the cloud storage services, enterprises can dismantle the huge costs from the procurement of equipment and technical personnel. Maintenance work and system update upgrades are done by the cloud storage service provider. The cost of public cloud rental and private cloud construction costs will continue to decrease as cloud storage vendors competition. Cloud storage is the future trend of storage applications. The advantage of cloud storage is that if there is a standard public application interface, any authorized user can access the cloud storage system. View the relevant material, this convenient and fatal is damage to cloud storage. (Joseph 2010.)

Because each kind of equipment has a kind of attack method. If the user accesses the cloud storage system through a mobile port, the data may be leaked. Although many cloud storage using encryption or other security technology, but these security technologies can not fully protect cloud storage. Except



the second calibration or secondary encryption, but the second verification and encryption will undoubtedly increase the difficulty of the development for suppliers and enhance the difficulty of the user accessing the relevant data. (Joseph 2010.)

Slow access is the problem that cloud storage can not break at the moment. Many people criticize it. To the present, cloud storage can not handle frequent files. The database is not a storage object for cloud storage because it requires fast network connection. Tier1, Tier2 or block-based data storage is also beyond the storage capacity of cloud storage. (Joseph 2010.)

Only some of the huge archives and unstructured data are suitable for cloud storage, such as bank's account information, medical institutions, patient data and medical history data. The current slow access is due to the performance and bandwidth factors of the equipment provided by the cloud storage provider. People believe that these two points will be resolved in the near future. (Joseph 2010.)

The main function of cloud storage is to store the same data in different areas with large storage devices, information of data backup help users solve disaster problems. Although the user through the agreement with the supplier to avoid the loss of data, but on the other hand, intellectual property rights are not protected, data ownership is not guaranteed. (Joseph 2010.)

China's B2B research center says that with the demand for new technology, cloud storage services will be more and more complete. Security will get better, the portability of the data will be faster than the local storage. Through the optimization of the WAN and capacity optimization, the problem of data delay will be resolved, user's effective access rate will be improved. (Joseph 2010.)

### **2.3.3 The impact of cloud storage on the Internet**

The dependence of the enterprise on the search engine becomes weaker. At this stage, many people try to use Google when they encounter problems. With the development of cloud storage, especially the services of city cloud, industry cloud and corporate cloud. People will prefer to direct access to the cloud, complete the information search in the cloud. Cloud growth will also increase with instant communication tools, not just using search engines. Affected by this situation, the search engine profit model will change. (Boss 2012.)

With the establishment of enterprise cloud, industry cloud, city cloud, information is more systematic and structured, people will be more convenient to access information. Anyone who wants to find information about a particular aspect can find all the information about this in cloud. For example, many people are using computer and feel speed is very slow, so people want to find some ways to improve speed. In the search process is very east to enter a computer industry cloud, and this cloud not only includes the computer speed related information, but also includes the computer daily maintenance, computer equipment self-test related information. (Boss 2012.)

### **3 CLOUD STORAGE CORE TECHNOLOGY**

In Chapter 2, cloud storage technology might be characterized by high reliability, high availability, high security, standardization and low cost. The realization of these features mainly rely on distributed access, global access space, virtualization awareness, data flow and other technical support. This chapter will introduce and describe these key technologies. (Joseph 2014.)

#### **3.1 Storage virtualization**

The most popular understanding of storage virtualization is a performance of abstract storage hardware resources. It means that integrate one (or more) targets services or functions with other additional features integrating, unified comprehensive features to provide useful service. (Joseph 2014.)

Typical virtualization includes the following scenarios, Shielding the complexity systems, Adding or integrate new features, emulating, integrating or breaking down existing service functions. Virtualization is working on one or more entities, and these technology are used to provide storage resources and / or services. Storage virtualization is a technology that runs through the entire IT environment is used to simplify the underlying infrastructure that could be relatively complex. (Joseph 2014.)

The idea of storage virtualization is to separate the logical image of the resource and the physical store. Providing a simplified, seamless virtual view of resources for systems and administrators. For the user, the virtual storage resource is like a "storage pool". Users will not see the specific disk tape do not have to care about that data is from which path to specific storage device. From the management point of view, the virtual storage pool is taking centralized management. And then dynamically divided to each application according to the specific needs of the storage resources. (Joseph 2014.)

In other way, for the virtualization technology, people can use a disk array to simulate a tape library, provide storage resources to application that speed as fast as the disk, the capacity is similar as a large library. This is the VTL(Virtual Tape Library) which plays an increasingly important role in the company storage system (Boss 2012.)

The benefit of virtualization of the storage resources into a "storage pool" is storage pool integrate many scattered storage resources, thereby it also improving entirety utilization, while reducing system man-

agement costs. And for the virtualization technology, resource allocation function of storage virtualization has ability to divide and allocate resources, The integrated storage pool can be divided according to the requirements of "service level agreement", using the highest efficiency, the lowest cost to meet different applications in terms of performance and capacity requirements. (Joseph 2014.)

In addition to the benefits of time and cost, storage virtualization can also improve the performance and availability of storage environments. This is mainly due to "in a single control interface to dynamically manage and allocate storage resources." In recent enterprise operating environment, the data growth rate is very fast, and enterprise management data capacity to improve the speed is always far behind. Through virtualization, many works that takes time and repeats many times. Such as backup / restore, data archiving and storage resource allocation can be carried out by way of automation greatly reducing the manual work. Therefore, through the data management work into a single automated management system, storage virtualization can significantly reduce the gap between data growth and enterprise data management capabilities. (Loomis 2010.)

Only the network-level virtualization is the sense of the storage virtualization. That can be stored on the network of various brands of storage subsystem to one or more storage pool can be centrally managed, creating one or more virtual volumes of different sizes in the storage pool as needed. These virtual volume allocated to the server according to a certain read and write authorization. This achieves the goal of making full use of storage capacity, centrally managing storage, and reducing storage costs. (Loomis 2010.)

### **3.1.1 Multi-Tenancy**

For most people, multi-tenancy (Multi-Tenancy) can be regarded as a very new concept, however, this concept has been around for a long time. When the earliest, the concept of a multi-tenant refers to a separate instance of the software that can serve multiple organizations, and multiple organization services called multi-tenants. (Colantonio 2011.)

A multi-tenant software means that the designer can virtualize their data and configuration information, so that each organization using the software can use a separate virtual instance, and can be customized to the virtual instance. The tenant's data is both isolated and shared, thus solving the problem of data storage. In support of the multi-tenant model, the well-known cloud computing SaaS service provider

Salesforce.com is great, as of 2011, it has 73,000 customers. And these customers from 8 to 12 multi-tenant examples to support the company. (Colantonio 2011.)

The ratio is 1: 5000. That means each multi-tenant instance supports 5,000 users sharing the same data space. Analyzing from the architectural level, the Multi-Tenant model is an important difference between the SaaS and traditional technology. (Colantonio 2011.)

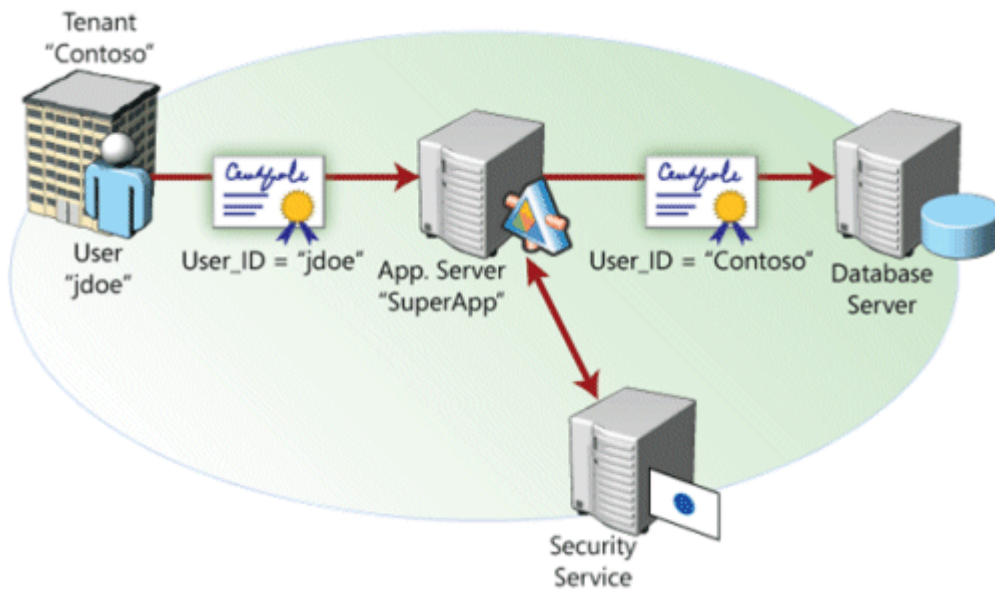


FIGURE 2. Multi-Tenant process (The Force.com Multitenant Architecture, 2008)

There are three main Options that SaaS multi-tenant safety the data.

Independent database is the first option. A tenant use one database, which has the highest level of user data isolation, the best security, but the cost is high. Advantage of independent database is to provide independent databases for different tenants. It can help simplify the expansion of the data model design. To solve the unique needs of different tenants. If a failure occurs, recovering the data is relatively simple. (Colantonio 2011.)

Disadvantages are increasing the number of database installation, along with the increase in maintenance costs and acquisition costs. This program is similar with the traditional a customer a set of data and a set of deployments the difference is that the software is deployed in the carrier. If people are meeting with banks, hospitals and other need to be very high data isolation level tenants, you can choose this model to improve the rental pricing. If the price is lower, this solution is unacceptable to the operator. (Colantonio 2011.)

Share the database, isolate the data schema. This is the second option, multiple or all tenants share Database, but a Tenant and a Schema. Advantage is For tenants with higher security requirements provides a degree of logical data isolation, not completely isolated; Each database can support more tenants. Disadvantages is if a fault occurs, data recovery more difficult, Because the recovery database will involve data from other tenants; It is difficulties If there is a need for cross-tenant statistics. (Colantonio 2011.)

The third option is shared database, which means share data architecture. Tenants share the same Database and same Schema. But it can distinguish the tenant's data by using TenantID. This is the highest level of sharing. The disadvantages are that is the lowest level of isolation, the lowest security, the need for design and development to increase the amount of security development. Data backup and recovery is the most difficult thing, which need to be backed up and restored. If people want to serve the most tenants with the least number of servers and the tenant accepts a low isolation level to reduce costs, this is the best solution. (Colantonio 2011.)

### **3.1.2 Virtualization implementation**

Depending on the location of the virtualization implementation, storage virtualization can also be divided into host-based, storage-based, and network-based virtualization. Storage virtualization can be implemented at three different levels. Including the implementation of the file system software works on the host server, or the implementation of the array controller firmware (Firmware) works on the disk array, or the implementation of the Dedicated virtualization engine works on the Storage network. Cloud storage virtualization implementation plans shows as follows picture. (Hang 2004.)

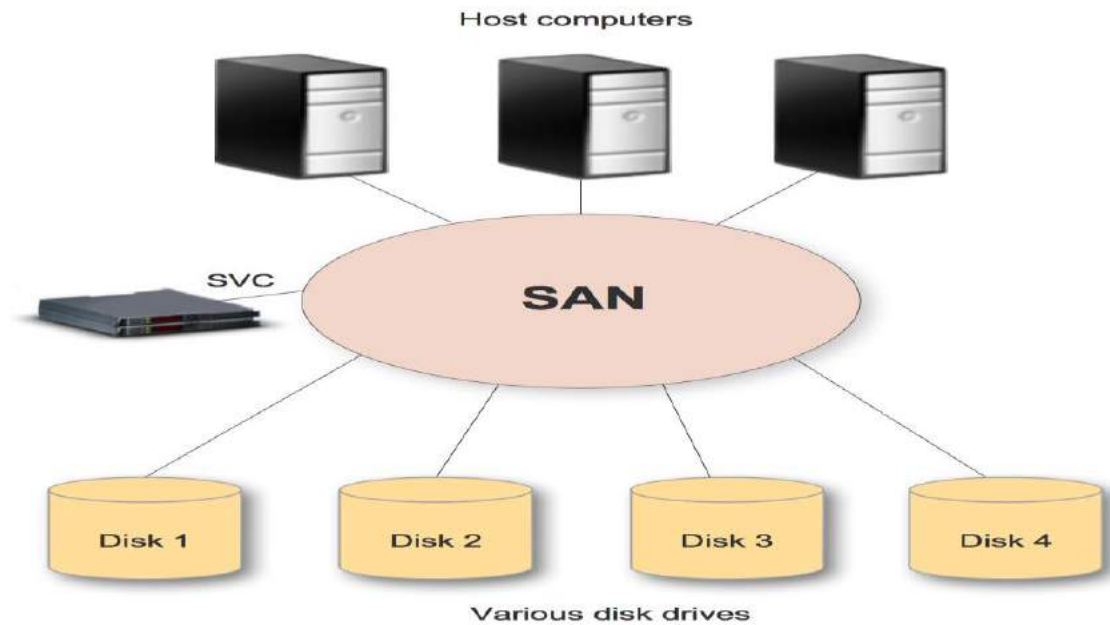


FIGURE 3. Multi-Tenant process (IBM Storage Virtualization, 2013)

Method 1 is the host-based virtual storage. Host-based virtual storage relies on proxy or software management. They are installed on one or more hosts, to achieve storage virtualization control and management. With the control software is running on the host, it will take a port from host processing time. Thus, this method has poor scalability, the actual performance of the operation is not very well. Host-based methods can also affect the stability and security of the system, because there may be unintentional access to protected data. This method requires to installed an appropriate control software on the host, a host failure may affect the integrity of the data in the entire SAN system. Software-controlled storage virtualization may also be due to different storage vendors hardware and software differences and bring unnecessary cost, the flexibility of this method is also relatively poor. (Hang 2004.)

However, because no additional hardware is required, host-based virtualization is the easiest one to implement. Its equipment cost is the lowest. Suppliers who use this approach tend to be software vendors for storage, and nowadays there already is a mature software products. These software can provide a graphical interface, easily management and virtualization SANs. In the host and small SAN structure has a good load balancing mechanism. In this sense, host-based storage virtualization is a cost-effective method. (Hang 2004.)

Method 2 is the virtualization based on storage devices. Storage - based storage virtualization approach, is dependent on the provision of related functions of the storage module. If there is no third party virtual

software. Storage-based virtualization often provides only one solution which is incomplete storage virtualization. For SAN storage systems that contain multi-vendor storage devices, the effect of this method is not very well. Functional module that depends on the storage vendor will exclude JBODS (Just-a-Bunch-of-Disks, one hard disk group) in the system. It affects the use of simple storage devices, because these devices do not provide storage virtualization capabilities. (Hang 2004.)

At the same time, using this method means that people will eventually lock a single storage vendor. Storage-based virtualization approaches also have some advantages. This approach is easier to implement in storage systems, it's easy to coordinate with a particular storage vendor's equipment and easier to manage. At the same time it is transparent to the user and the manager. However, they must note that because software to support virtualization can not be too much, the solution is more difficult to customize and monitor. (Hang 2004.)

Method 3 is the web-based virtual storage. The network-based virtualization approach means to implement storage virtualization in network devices. There are two ways as follow. First one is virtualization based on interconnected devices. If the method of the interconnection device is symmetrical, then the control information and data working on the same channel. If it is asymmetric, the control information and data works to the different paths. In a symmetrical way, the interconnected device may become an obstacle, but multiple device management and load balancing mechanisms can reduce the barriers to obstacles. In a multi-device management environment, when a device fails it is easier to find the server to replace the fallacious one. (Hang 2004.)

Virtualization based on interconnected devices can run on the dedicated servers, use standard operating systems such as Windows, Linux, or the operating system provided by the vendor. This method running in the standard operating system has the advantages under host method is easy to use, cheap equipment. Many device-based virtualization providers also provide additional functional modules to improve the overall performance of the system. That can have more performance and better functionality than standard operating systems, but requires higher hardware costs. (Hang 2004.)

However, some of the flaws in the device-based approach are the same as the host-based virtualization approach, because that still requires a proxy software running on the host or a host-based adapter, if any host failure or inappropriate host configuration may result in access to unprotected data. (Hang 2004.)



Second is Router-based virtualization. Router-based method is to achieve storage virtualization on the router firmware. Vendors typically also provide additional software running on the host to further enhance storage management capabilities. (Hang 2004.)

In this method, the router is placed on data channel of each host to the storage network. It used to intercept any command of the network from the host to the storage system. Because the router is potentially serving each host, most control modules exist in the firmware of the router. Compared to the host-based and most Internet-based devices, this method performs great. Because it does not depend on the proxy server running on each host, this method has better security than host-based or device-based methods. When the router connecting the host to the storage network fails, it still caused the data on the host could not be accessed. But only the host connected to the faulty router will be affected, other hosts can still access the storage system through other routers. The redundancy of the router can support dynamic multipath, which provides a solution to the above fault problem. (Hang 2004.)

### **3.2 Distributed extended mode**

A distributed operating system is a software over a collection of independent, networked, communicating, and physically separate computational nodes. Each individual node holds a specific software subset of the global aggregate operating system. Each subset is a composite of two distinct service provisioners. The first is a ubiquitous minimal kernel, or microkernel that directly controls that node's hardware. Second is a higher-level collection of system management components that coordination the node's individual and collaborative activities. These components abstracts microkernel functions and support user applications. (Tanenbaum, 2003.)

#### **3.2.1 Distributed File System**

Distributed File System refers to the file system management of physical storage resources are not necessarily directly connected to the local node. That connect by connecting the computer network with the node connected. The design of a distributed file system is based on the client / server model. A typical network may include multiple servers for multi-user access. In addition, peer-to-peer features allow some systems to play the dual roles of both the client and the server. For example, a user can publish a directory that allows other clients to access once accessed. This directory is like a local drive for a client. (Liu 2009.)

The computer manages the data through the file system, stores the data, and people can obtain the data into exponential growth. Simply by increasing the number of hard drives to expand the storage capacity of the computer file system the capacity size capacity growth rate data backup data security and other aspects of the performance is not well. Distributed file system can effectively solve the data storage and management problems will be fixed at a certain location of a file system, extended to any number of locations / multiple file systems, numerous nodes make up a file system network. Each node can be distributed in different locations through the network for communication between nodes and data transmission. When people use a distributed file system there is no need to care about where the data is stored on or from which node, they only need to manage and store the data in the file system as people would with a local file system. (Liu 2009.)

When the file system was originally designed, it was only for local data services within the LAN. However the distributed file system extends the service to the entire network. Not only change the data storage and management, but also has the advantage of a data backup, data security in the local file system. (Liu 2009.)

There is a large number of servers which form a collection. For the user is still a coherent system. First one is tanenbaum which is coordinated action in the components and a distributed network for computer. Second one is G. Coulouris which means when know there is a computer crash but your software runs never stops. Third one is leslie Lamport, definition is Distributed systems is the system that design to support the development of applications and services, can use the physical architecture by a number of autonomous processing elements do not share the main memory. But send asynchronous messaging through the network. Differences with hierarchical applications, layered applications (for example, Layer 3) are divided into application logic is a logical stratification not physical. The distributed system DS is a physical which layer which related to actual deployment. (Liu 2009.)

People can find out some distributed system features. For example, concurrency means sharing resources, taking ACID or Base principles see CAP theorem. The distributed system design follows the CAP theorem. CAP is consistency, availability, and partition tolerance. The CAP theorem holds that CAP can only satisfy two species. Scalability Scalable is an important feature that enables high-performance, high-throughput low latency latency through expansion. (Liu 2009.)

Reliability / availability means fault discovery and processing and recovery of fault-tolerant processing. There is a time proportional condition in a normal operating system. If a user can not access the system

the proportion increases. It is not available. Usability formula is  $\text{availability} = \text{uptime} / (\text{uptime} + \text{downtime})$  Fault-tolerant failover is a system in the case of an error everything is still working properly, indicating that the system is tolerant to the error. (Liu 2009.)

Message processing means that specific products are RabbitMQ ZeroMQ Netty and heterogeneity. Different operating system hardware program language developers middleware is a solution. The security is licensing authentication SSO and sign in Oauth. (Liu 2009.)

Transparency means the accessing transparency. Using the same operations for local and remote resources. Location transparent is accessing to resources without knowing its physical or network location. Concurrency Transparency has multiple processes can run concurrently using shared resources when they can not interfere with the processing of their processes and the copy. Transparency is multiple instances of a resource which can be used to replicate to improve reliability and performance without the need for a user to program a specialized application. (Liu 2009.)

Fault transparency is when a software hardware failure occurs, the user and application scenarios can continue to complete their tasks without being affected. Mobile transparency allows the system to move resources and customers. Performance transparency allows the system to reconfigure to improve performance load changes and zoom transparency is the extended the system on a scale without changing the application structure to improve throughput processing. (Liu 2009.)

Cloud storage system has good scalability, fault tolerance, and internal implementation of the user transparent and other features. All of which are inseparable from the support of distributed file system. The existing cloud storage distributed file system includes GFS, HDFS, Luster, FastDF. Many of their design ideas are similar, but also have their own characteristics. (Liu 2009.)

Google File System (GFS) is an extensible distributed file system that is primarily used to handle large distributed data-intensive applications. GFS is a major feature of running on a large number of common low-cost hardware through the GFS file system to provide fault-tolerant features and to provide a large number of users can handle large amounts of high-performance data services. Compared with the traditional standard, GFS file size is large. It is mainly used to deal with large files. In addition, GFS mostly change the file by adding new data directly, rather than overwriting the existing data. Once the data is written the file only supports read operations. (Loomis 2010.)

The Luster file system is a typical distributed file system based on object storage technology. At present, the file system has been widely used in many high-performance computing institutions, such as the US Department of Energy, Sandia National Laboratory, Pacific Northwest National Laboratory. The Top500 machines is used in multiple units using Luster file system. (Loomis 2010.)

The large file properties of the Luster file system are great, the same data file is divided into a number of objects stored in different object storage devices by object-based data storage format. Large file I / O operations are assigned to parallel execution on different object storage devices, resulting in large aggregate bandwidth. In addition, because Luster combines the characteristics of the traditional distributed file system and the traditional shared storage file system design concept, so it has a more effective data management mechanism, global data sharing, object-based storage, storage intelligence, and rapid deployment and the other advantages. In spite of this, since Luster uses a distributed storage architecture to store metadata and data files separately access to the data before the need to access the metadata server. This process increases the network overhead so that Luster's small file I / O operation performance is poor (Loomis 2010.)

Parallel Virtual File System (PVFS) designed and successfully developed by Clemson University, is an open source parallel virtual file system built on top of the Linux operating system. The entire file system consists of management node, calculation node and I / O node composed of three parts. The management node is responsible for handling the file metadata. The compute node is used to perform various computational tasks. The I / O node is responsible for the storage and reading of the data file and is responsible for providing the required data to the compute nodes. (Loomis 2010.)

Throughout the cluster system, PVFS uses a consistent global namespace. In addition, PVFS applies object storage concepts to strip data files into multiple objects and store them across multiple storage nodes. Because in the network communication, PVFS only supports the TCP network communication agreement, which makes its lack of flexibility. In addition, due to the concept of PVFS application object storage data file storage which performance in dealing with small files is not ideal. (Loomis 2010.)

### **3.2.2 Distributed database**

Distributed database refers to the use of high-speed computer network would be physically dispersed multiple data storage unit connected to form a logically unified database. The basic idea is to store the data from the original centralized database to multiple data storage nodes connected through the network

to achieve greater storage capacity and higher concurrent traffic. In recent years, with the rapid growth of data volume, distributed database technology rapid development. The traditional relational database began to develop from the centralized model to the distributed architecture. The relational database based on the relational database relies on the data model and the basic characteristics of the traditional database start the process from centralized storage to distributed storage from centralized computing to distributed computing. (Vaidya 2010.)

On the other hand, with the increasing amount of data, relational database began to show some difficult to overcome the shortcomings. NoSQL as the representative of the non-relational database its high scalability high concurrency and have other advantages of the rapid development. A time on the market there have appeared a lot of key-value storage systems, document-based database NoSQL database products. NoSQL type database is becoming the main era of distributed database in the era of large data. (Vaidya 2010.)

In the face of amounts of data occurs growth and growing user demand. The distributed database must have the following characteristics in order to cope with the growing mass of data. For high scalability, distributed database must be highly scalable which can dynamically add storage nodes to achieve a linear expansion of storage capacity. High concurrency means distributed database must respond to large-scale users read / write requests in a timely manner can be randomly read / write data. High availability means it distributed database must provide fault-tolerant mechanism to achieve redundant data backup and to ensure high reliability of data and services. (Vaidya 2010.)

### **3.3 Information life cycle management**

Information lifecycle management (ILM) refers to strategies for administering storage systems on computing devices management that an information system data and its associated metadata are deleted from the generation and initial storage phases to the last. Different with the early data storage management methods, information lifecycle management technology is according to the user's operation from the full range of data management not just to automate the data storage process. The focus of the information lifecycle is naturally data. (Boss 2012.)

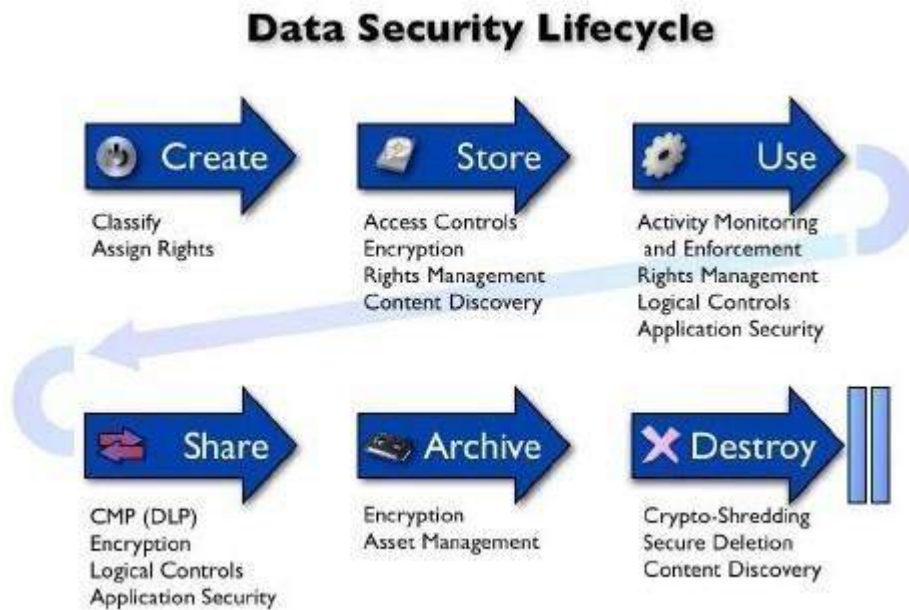


FIGURE 4. Information lifecycle management (webopedia ILM, 2013)

In the figure, the process of lifecycle is first create classify assign rights. Then is store, access controls encryption rights management content discovery. Using it activity monitoring and enforcement rights management logical controls application security. Sharing CMP(DLP) encryption logical controls application security. Archiving encryption asset management. At last, destroy Crypto-Shredding Secure deletion content discovery. (webopedia ILM, 2013.)

And then a new concept automatic tiered storage appeared. The basic principle is that the value of the data after the creation of the value of time will gradually reduce. The data is primarily accessed within 72 hours of its creation. After that, the number of visits will be reduced. The access frequency is getting lower, and the data will only be accessed occasionally after 30 days. At this time, the data becomes "passive data" or "cold data". As the value of the data decreases, the data should migrate to the low-speed, low-cost storage tier. If people want to do this manually, this repetitive operation is very tedious and difficult to meet the required workload. In other words, no one will do that. Automatic tiered technology is based on policies such as data creation time, access frequency, last access time, or response time. (Boss 2012.)

Automatic tiered technology is an important storage technology for data information life. The data has a life cycle access level storage devices are also different levels high reliability fast reading and writing equipment and the price is high. While the performance is low slow reading and writing equipment, the

price will be relatively low. Automatic tiered storage is making the data life cycle and storage devices match, which has reached the most economical and efficient storage resources. (Boss 2012.)

### **3.4 Data protection**

In cloud storage, the level of service provided can be defined according to the requirements of different users. For the high availability of data, the corresponding services can be provided by implementing different data protection means. Traditional storage usually uses copy protection for data protection. Data is usually stored across a copy of the node, the number of replicas depends on the user's overall measure of capacity utilization and data reliability. Replicas can provide higher data consistency. Data access performance and reliability are high protection. However, when the user's data is very large, a copy of the storage space may cause lower utilization rate, in particular, the absolute value of the storage space utilization will increase the cost of the user. In cloud storage, not only can copy protection, but also through the data of the fragment storage and rewritable code (Erasure Code) way to achieve data, these two technologies will cloud storage data protection pushed to a higher level. (Hang 2004.)

#### **3.4.1 Fragmentation storage technology**

Data fragmentation in cloud storage technology and the traditional centralized storage fragmentation mechanism is slightly different. The original data is also fragmented, but after the data fragmentation, in the cloud storage system requires the distribution of different fragments can be stored. After the data is fragmented, if the request is read, the algorithm that is restored to the original file after the slice is acquired is very expensive for the system. Therefore, the application of fragmented storage technology in cloud storage is selective, rather than building a cloud storage which must use the data fragmentation mechanism. (Hang 2004.)

In the cloud storage system, after the data sub-chapter, people can achieved distributed storage, each fragment will be stored on a different node up. However, when there are multiple fragments stored on the same node, to spread these fragments to a different hard disk up becomes a question. Because the purpose of data fragmentation are two, one is to improve the performance of parallel data write and read, and the other can be used with the copy and Erasure Code mechanism to achieve data protection. For these two purposes, the way data stored to the hard disk will directly affect whether the target can be achieved. (Hang 2004.)

At present, there are three ways to store data into a hard disk. First one is round-trip storage, that is, according to the number of hard disk to save. Second one is maximize the storage, to maximize the write a hard disk. When known the hard disk is already "full", and write to another hard disk. Last one is balanced write. Namely IO load very small hard disk in order to write. (Hang 2004.)

### 3.4.2 Erasure Code

The current mainstream cloud storage solutions, basically all of them have provide a Erasure Code way to protect data. Feature of the Erasure Code mode is no replication technology. At the same time, it can also improve the redundancy of data, while controlling the additional costs associated with data protection. In the Erasure Code data protection mechanism using  $M:K$  redundant fault tolerance mechanism for data protection. The most typical configuration is a 9: 3 or 10: 6 ratio. The data dividing and generating checksum to store and read. And the storage of data on whether to implement the Erasure Code can also be configured flexibly. (Zhilin Zhang, 2010.)

Implementation ratio base on data level. Users can define parameters themselves. Data continuity and integrity are certified. Erasure Code coding work is shown (  $M$  for data fragmentation,  $K$  for the test code,  $M:K = 9:3$  model to illustrated). (Zhilin Zhang, 2010)

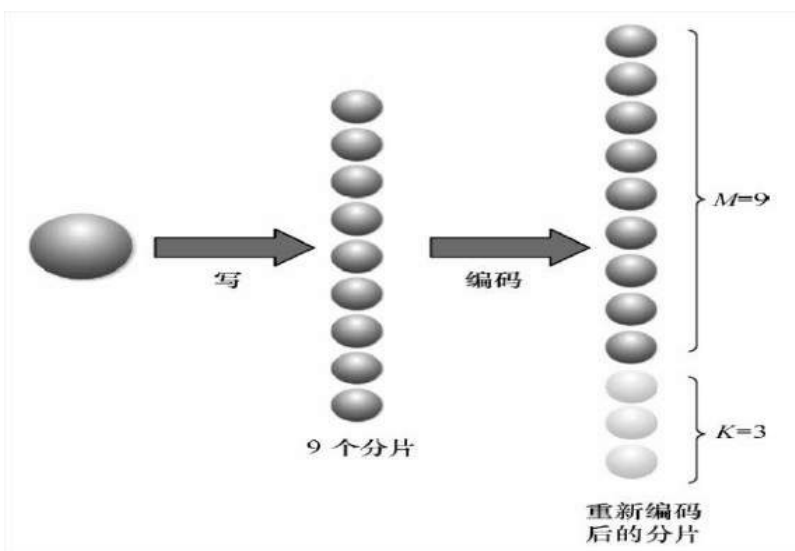


FIGURE 5. Ensure code work as  $M:K=9:3$  (Zhilin Zhang, 2010.)

Data is divided into  $M = 9$  data fragments and  $K = 3$  efficiency code. The check code for  $K = 3$  is calculated using Cauchy Reed-Solomon coding. Data slices are stored in separate fault tolerant domains, each slice is stored on a different hard disk. Storage cost ratio =  $K / (M)$ , where the storage cost ratio =  $3 / (9) = 33\%$ . For data fragmentation, there is consistency and integrity of the test. (Zhilin Zhang, 2010.)



The following figure shows the decoding of Erasure Code.

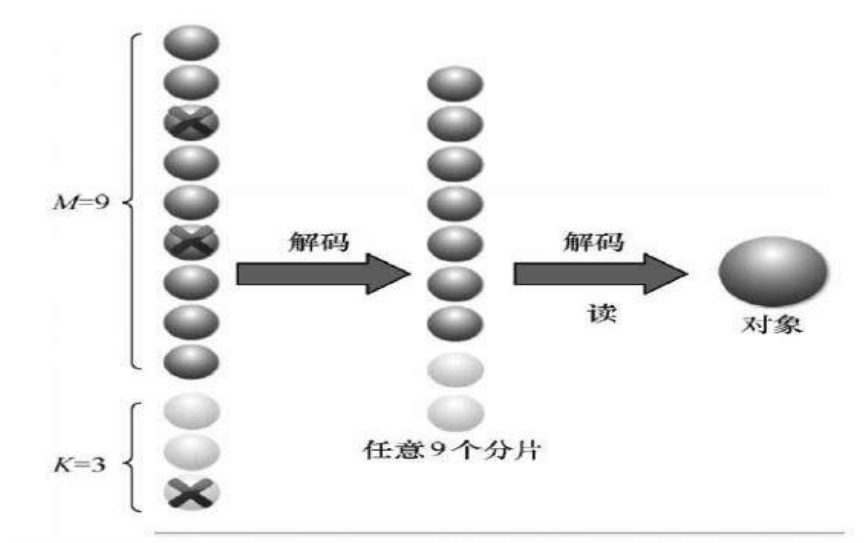


FIGURE 6. Ensure code work as M:K=9:3 (Zhilin Zhang, 2010.)

Data can be recovered from any slice. Data in a timely manner in the check code is partially lost, it can also restore. In the entire decoding process only need to sacrifice very little performance. (Zhang, 2010.)

### 3.5 Intelligent management

From the current cloud storage development trend, intelligent is becoming an indispensable feature in the cloud storage infrastructure. The existence of a large number of nodes in different geographical distribution of resources, the system supports millions of concurrent access and hundreds of millions of files or objects are cloud storage system is a significant feature. And these features require intelligent on the entire cloud storage system, Cloud storage is mainly reflected in the following aspects of intelligent. (Boss 2012.)

#### 3.5.1 Dynamic expansion of resources

Cloud storage service with the growth of business, data performance requirements are constantly changing. That is constantly asking for new requirements from the application of the stress requirements. In the configuration cloud storage, resource is to consider the dynamic expansion of resources. When building a cloud storage service, the cost is managed by demand balance. If the beginning of the peak demand in accordance with the allocation of cloud storage resources, then the cost will be greatly increasing. If the enterprise want to cut costs, the only way is maintaining the smallest storage resources, when the demand peak occurs, the lack of resources will bring disastrous consequences to the business system. In

particular, this scalability or the ability to dynamically increase resources is very important to many of today's Web-based applications. Web Service application has a large-scale features, Web Service application has a large-scale features, if people can dynamically developed, they can meet the needs of the scale of change as much as possible. (Joseph 2004.)

For cloud storage systems, the provision of storage resources and storage capacity is usually provided through the resources of the node. Therefore, the node can dynamically determine the direct expansion no matter the system has a high scalability. In a distributed cloud storage system, the storage area of the node also realizes the dynamic expansion for the storage domain because of the dynamic extension node function. Similarly, it is also possible to achieve this highly flexible scalability through the expansion of node resources in a storage domain. (Joseph 2004.)

### **3.5.2 Requested Load Balancing**

Mass storage will certainly support a large number of requests. In cloud storage, a typical feature is that these massive requests can be implemented within the system load balancing. The requests from the different addresses are sent to the respective bearer devices evenly and optimally according to the settings of the load balancing device. The cloud load in the load balance includes the edge of the network load balancing equipment to achieve dynamic balancing based on DNS analysis, also includes the cloud load internal load balancing mechanism. (Joseph 2004.)

The load balancing mechanism within cloud storage can be simply understood as load balancing between node resources. When a request occurs, the cloud storage intelligence has began to reflect its role. This request respond by which node, moreover it must be to the request of the data where the node is corresponding. If it is a large number of requests have been pressed on this node, the system have a better handling mechanism to handle this request or not become a question. This kind of request-based load balancing mechanism for cloud storage node resources is important to support a whole set of cloud storage systems. In general, the program will give an explanation for that. A common practice is through a large number of requests and judgments of system resource availability to carry on comprehensive coordination. (Joseph 2004.)

The equilibrium node at the same time bring us another important concept: the load balancing of the nodes can realize the dynamic expansion of the system. Because of different nodes corresponding to the

different requests, each node only needs to execute the request assigned to it, the performance has good protection. When a lot of large requests made at the same time, this mechanism under the cloud storage system has different nodes balancing the processing of requests for corresponding data. (Joseph 2004.)

For example, the system has eight orders, and there are 80,000 requests, the 80,000 requests will cloud balanced distribute to the node. If each node handles up to 10,000 requests, no matter how much the system resources , it will bring performance and user experience decline. When it becomes 800,000, people can expand the node to 80, because the internal load balancing mechanism does not require special treatment. When the 800,000 requests occurs, it must still evenly distributed up to 80 nodes , at this point each node is still dealing with 10,000 requests. Performance will not decline. Through this load balancing mechanism and the ability of nodes to expand, can realize the large-scale deployment of cloud storage needs. (Joseph 2004.)

### **3.5.3 Resource failure management**

Resource failure management also requires intelligence. The resources in the cloud storage usually refer to the hard disk. For the entire cabinet, and even to the entire storage domain, is distributed cloud storage data center. (Han 2006.)

The smallest unit of the hard disk failure will not bring data loss and performance degradation. This is a significant difference feature between cloud storage and traditional storage. Traditional hard disk data protection method is to use the RAID mechanism to carry out. However, in the cloud storage the concept of RAID has been greatly diluted. If the underlying is a traditional storage device, it must be to ensure the appropriate hard disk RAID protection. At first, this approach is to reduce the efficiency of the use of hard drives. Secondly, when there is a hard disk failure, to ensure the integrity and availability of data, people must go through the reconstruction of the RAID group. (Han 2006.)

In the real cloud storage, a single hard drive and multiple hard disk failure did not affect the use of the entire system and data. From the user to the administrator, they do not need to care about this level of hardware failure. When a hard disk failure the hard disk data will automatically generate a copy in the cloud. After the broken hard disk is replaced with a new disk, there is no data needed to write back to new hard disk, just join into the cloud storage system as a new storage resource. (Han 2006.)

## 4 Typical Cloud Storage Applications

Cloud storage applications are cloud storage providers through the open Internet to the user to pay for the use's storage services. According to the customer base and storage services provided in different ways, can be divided into public applications (for individuals, home users) and enterprise applications for business users. (Liu 2016.)

Typical cloud storage public applications are mostly served as SaaS layer services to users. There are many cloud storage applications synchronous backup document save offline download online editing and document sharing. Typical cloud storage enterprise applications mostly serve the user with the IaaS layer. Providing basic storage facilities and basic storage backup services for developers (mainly for enterprises) rental to facilitate their own office or according to business needs. Renter develops the business layer and ultimately provides a cloud of storage for business users. With the gradual maturity of cloud storage technology, more companies realized cloud storage prospects and begin to launch cloud storage applications. (Liu 2016.)

As a typical representative of the SaaS layer application Dropbox earlier appeared in front of the public user, provides synchronous backup function and change the user's office habits. Allowing users away from trouble that work outside the office need to carry the computer, and the file cloud sharing model is also closer to the distance of many Internet users. As a typical representative of the IaaS layer application, Amazon S3 has created the provision of basic storage facilities and interface for third-party companies to rent and provide cloud storage applications to public users. (Liu 2016.)

Because of knowing some of the successful transformation of Internet vendors telecom operators have also launched a cloud storage applications. Telecom operators cloud storage applications and cloud computing services are often inseparable. Cloud's character of integrated and high security are difficult to reach for Internet companies. These advantages may be the basis for the success of telecom operators in cloud storage applications. This chapter first introduced the typical cloud storage public application and enterprise application from the aspects of function, profit model and operating characteristics. Then introduced the telecom operators of cloud storage applications business model. (Liu 2016.)

### 4.1 Public application

Most of the cloud storage public application is the network disk business. The basic functions is upload download synchronous backup and online editing. While in the toll mode including free business and value-added charges business. Until now, many applications have their own features, like Dropbox is the feature of friendly interface convenience Google Driver has many third-party applications. And has its unique collaborative editing capabilities and image recognition search technology, SkyDrive relies on Microsoft's user to launch new products is iCloud put Apple's services and equipment together. (Liu 2016.)

#### **4.1.1 DropBox**

Dropbox introduced the online storage business in 2007. One of the largest and earliest implementations of cloud storage disk drives. Dropbox company has been set up 9 years, the number of users are more than 80 millions, the cumulative preservation of more than 100 billion files. Although a considerable number of users are free users, the operating income is still considerable. Dropbox products' device is under the Amazon storage device, easy to operate, with high service reliability and fast transmission speed. These advantages make Dropbox company gain more users. Dropbox's user covered business executives, retired people and students. (Liu 2016.)

First function is synchronous backup. It supports the user to set up a synchronous backup folder on the terminal to achieve synchronization with the backup on the network disk. Through a simple installation plug-in, people can put mail and other business generated documents automatically synchronized to the network disk. When the user opens and logs on to the computer, the synchronization folder in the new or modified files will automatically sync to the Dropbox. When the user logs in to Dropbox, the file is automatically synchronized from Dropbox to the computer. When the file in the folder changes, Dropbox only uploading changes to the file part, which has high efficiency. Second function is batch upload. Dropbox supports batch upload of files. Web page upload sa single file limit of 350MB. The client upload is unlimited, different packages have different upper limit traffic for upload and download. (Liu 2016.)

Third function is multi-version support. Dropbox supports multi-version save even if the file is deleted. It can also be restored to any historical version. So that even if the number of people involved in editing the file. Different users can find the historical version that he needs which will not be confused (Liu 2016.)

Operational characteristics include internet marketing model which is viral marketing and open interface. Dropbox uses viral marketing mechanism. The new registered users have only 2GB of space, If the user wants to increase the free space they can invite friends to register. Each successful registration of a friend will increase the storage space by 250MB. The current increase of up to 16GB of space. Inviting friends can be achieved through three ways like sending invitations to social networking sites. Sending an invitation link to the email contact by email and the dropbox provides an online mass invitation of link interface. (Liu 2016.)

Through opening interface to let third-party software vendors develop the applications for Dropbox, and get Dropbox profit sharing. More applications can increase the user's needing. Dropbox application bundle the account to maintain the relationship with the user. There are two types of third-party applications for Dropbox, bundle the account to achieve the contents of the Internet application and synchronization of Dropbox folder. Another is according to links jump after bundling to use the functionality of the application to handle Dropbox files. (Liu 2016.)

Dropbox's profit model is mainly using free trial and advanced service charges model. The first 2GB space is free, and then need to pay the monthly storage cost of the space size. 50GB space is \$ 99 per year, 100GB of space is \$ 199 per year. People can also get free space by inviting friends which is the account has bundled Facebook, Twitter account. It will slowly increase its package capacity, and let users still use Dropbox. At the same time, Dropbox charges a fee that is developing from the third-party software vendors, and divide profit to third-party vendors, improve the enthusiasm of developing applications. (Liu 2016.)

#### **4.1.2 Google Drive**

Google Drive as Google Inc. on April 25, 2012 launched of the cloud storage applications, replaced the previous Google Docs service. Users can log in to Google Drive via a unified Google account. Google Drive and Google's other business had closely linked and integrated Google's own business. (Liu 2016.)

Google Gmail users can send large attachments through Google Drive, and Google Drive users' video and pictures in Google Picasa can be shared to other Google Drive users, such as Gmail users. Google Drive covers the most important platforms today: Mac, Windows, Android and iOS, ranging from mainstream devices such as PCs, tablets. to smartphones such as Android phones and iPhones. The major

version has a local client version and a web version. Google Drive has fast speed of synchronization, and rely on the Chrome application store, it has lot of third-party application. In addition, Google in the industry has a good reputation and a huge amount of users, it provides Google Drive products a good development prospects. (Liu 2016.)

Users can store any files, such as photos, documents, and videos into Google Drive and access storage at any time through a variety of terminals. Google Drive not only upgrading Google Docs many features, but also increasing the sharing features through the interface integration for a large number of external applications, the number of applications is still increasing. (Liu 2016.)

Currently, Google Drive limits the capacity of a single upload file to 10GB. If people want to upload large files in the shortest possible time. Google Drive is the first choice. For free users, the system will automatically save up to 30 days up to 100 versions, if exceeded, it will delete the previous version, users can also save more versions by buying. (Liu 2016.)

Google Drive has the following two main functions except general cloud storage function, which are image recognition technology and share collaboration online editing. The Google Drive platform uses Google Goggles and Google OCR technology. Users can use image recognition technology to search information. If the user's keywords are included in the picture or scan file, then the pictures containing these words will appear in the scan results. (Liu 2016.)

Sharing collaboration and online editing by setting up authorization for Google Drive, people can specify that the file becomes a shared data for a group and the group member can synchronize to the collaborative office team. Compared to the feature of Dropbox support multi-user decentralize edit the same file. Google Drive is undoubtedly more suitable for team office and more efficient. The Google Docs page supports 15 different types of files to online viewing and editing, including documents such as offices adobe CAD and PostScript. (Liu 2016.)

Operational characteristics of Google Drive are open interface and different kinds of third-party applications. Google's SDK Toolkit enhances the enthusiasm of third-party application providers to driving them to create more applications and increase the number of users. At the same time, Google Drive allows users to share content across multiple platforms. Users can add applications in Google Chrome to Google Drive to complete the application bindings when loading. Then users can use the application to edit the files in the network disk without installation. Google Drive has become Chrome OS theory

"local drive." Different kinds of third-party applications shows that until now the applications are include the video playback applications such as Pixorial Video document editing applications such as Aviary and Loupe e-mail applications such as Gmail Attachments To Drive office applications such as Doc Sign, Hello Fax, 8reader and task management classes such as Smart Sheet, Ganttter. (Liu 2016.)

Profit mode of Google Drive is provides users with 5GB of free storage. Paid packages are 25GB, 100GB and 1TB and several other. To meet customer demand for large-capacity storage space. In addition, Google Drive's integrated applications included free apps and fee-based apps, charging fees for users of paid apps and divided profits to developers. (Liu 2016.)

#### **4.1.3 iCloud**

iCloud is the service for the use of Apple's cloud equipment launched in Apple's June 7, 2011. Users can get 5GB of free space. The other optional payment model currently has 15GB, 20 US dollars / year and 25GB 40 US dollars / year 55GB for 100 US dollars / year. iCloud putted Apple music services, system backup and other products together. The iCloud platform can also store personal data to Apple's servers and automatically push them to each Apple device for the user over a wireless network, including iPhone, iPod, Touch, iPad, Apple TV, and Mac. (Liu 2016.)

Apple company use iCloud products through Wi-Fi and other wireless networks to achieve users' terminal information daily backup. Backups include contacts, calendars, documents, photos, music and books. The size of synchronization unit is limited. The current free account is 25MB, paying is 250MB. (Liu 2016.)

Cloud document synchronization can create documents to all types of Apple devices, people can automatically sync to the cloud, modify the record can also be synchronized. For Cloud photo synchronization, any Apple device photos can be automatically saved to the cloud, and to achieve and other Apple devices are synchronized with each other. iCloud can save up to 30 days of photos, iOS devices can save the latest 1000 photos, and Mac and PC can save all the photos. (Liu 2016.)

For Cloud synchronization in iTunes, users purchase the music that can be downloaded multiple times in different devices with the same Apple ID. The free space in the cloud is 5GB, which can save messages, documents and backup data. iTunes music does not occupant space. Compared to Google and Amazon's cloud music service, Apple has the powerful iTunes Match, which includes 18 million songs



for 256kbit / s, AAC format songs, and the cost is 24.99 US dollars / year. iTunes Match by scanning the user's collection of music library find out which music is not using Apple platform to match the music. Then Apple will provide users the same quality cloud-based version. (Liu 2016.)

Because iCloud does not support large-capacity single-file synchronization. The space is small in general, they are not concerned about its profit model, but rather its business model features. Careful analysis of iCloud's specific business they can find out. iCloud is enhanced version of Apple "Mobile Me" they used to provide. This service has been there before cloud computing which is a client / server application model. But iCloud has a great impact on Apple's business model. Free service replace the original fee service, then rely on value-added services charges. (Liu 2016.)

Enhancing user viscosity means that put iCloud server as the center unified users of different Apple devices can synchronize all kinds of application data including contacts calendars documents photos music and books. iCloud connect all the Apple's service together on the cloud and close contact effective combination of all Apple's software and hardware devices. Once the Apple user experienced iCloud, then they are going to buy the Apple's other hardware and software products. (Liu 2016.)

Attracting users to change hardware devices is also one of aim for Apple. Apple's iPhone series, iPad series in hardware and software have been to be perfect, only to join more attractive applications to attract more users to replace the hardware device. The reason they do this is that apple's profit is coming from the hardware business. (Liu 2016.)

## **4.2 Enterprise application**

Cloud storage enterprise applications expressed as providing the underlying storage facilities, and basic storage, backup service developers rent (mainly for enterprises), to facilitate the office of its own employees, such as online document editing, collaboration between employees, file sharing and a higher level of security encryption. And re-development business layer by the tenant, providing cloud storage services to the users, such as file storage, online sharing. Amazon S3, Amazon Simple Storage Service which is the Amazon online provider of online storage services. Users (mainly for small and medium business users) save the file to Amazon's server through REST, SOAP and BitTorrent in the WEB service interface. Function of Amazon S3 provides a simple Web service interface, allowing business users to store and retrieve data from any location on the network and any time. It also allows any developer to access this scalable, reliable, secure, fast and inexpensive infrastructure. (Liu 2016.)

Operational characteristics have regional storage, unlimited storage, tiered storage and disaster recovery. Users can choose the nearest site to store data. It used to improve the upload and download efficiency. At present, Amazon has five storage locations in the world. Users can choose from the United States, the western United States, the European Union, Asia and Asia Pacific. For unlimited storage and tiered storage, Amazon provides two storage tiers the standard layer and the reduced redundant storage tier. The standard layer is used to store critical data. Redundant storage layer is used to store less important data, such as pictures, caching. And the price is relatively cheap. Requiring of the Redundant storage layer is high, it up to 99.99% durability and 99.99% availability. The standard layer can also restore an old version of the storage object, with the SLA protocol. It can keep running when the devices occur some problem sometime. The purpose of this design is to prevent data loss when the data center is in a disaster or accident. For disaster recovery, people can use S3 to disaster recovery data on a single or multiple site in enterprise. S3 provides a DR final solution that combines durability, scalability, availability and security. It has functions of unlimited storage version control and Amazon AWS import and export functions. (Liu 2016.)

Amazon S3 charges fees for tenants which is including data space rental fees and requests. Charge according to the user selected storage mode, the use of storage space size. Larger storage space has the smaller average cost. Amazon S3 space rental fee table below.

存储空间	每月单 GB 的价格
0 ~ 1TB	0.125 美元
1 ~ 50TB	0.110 美元
51 ~ 500 TB	0.095 美元
501 ~ 1000TB	0.090 美元
1001 ~ 5000TB	0.080 美元
5001 ~ 10000TB	0.055 美元

FIGURE 7. Amazon Z3 pay plan (Amazon.com, 2013.)

Amazon S3 network services introduce a new business model, known as the "requestor to pay." Service providers using S3 services can choose their data to join this new model. The cost of data transfer borne by the data requester (end user), not the owner of the data. In the past, the owner of the S3 folder paid

for the storage and transmission of the data. It inhibits the opening of the data, and the new model changes the rules of the past. (Liu 2016.)

Facebook is a public service site, which uses the device under the Amazon S3 storage device. The Web service interface with store photos, multimedia files through REST, SOAP, and store it on the Amazon server. Amazon S3 is scalable, reliable, secure, fast and inexpensive infrastructure to achieve the storage needs of the business generated by the large user data. At present, Facebook has the largest number of college students on all major business sites, 34 million active users. Facebook also is the largest photo sharing application in the United States, the number of photos is 8.5 million per day. Obviously, Amazon S3 contributed lots of values. (Liu 2016.)

### **4.3 Telecom operators cloud storage application**

Telecom operators cloud storage business mainly provide services for enterprise customers, high-end business and home customers, among them, corporate customers focusing on medium-sized enterprises. Key enterprise users include transportation, manufacturing and public utilities industries, to provide enterprise space rental services, enterprise-class remote data backup and disaster recovery services, video surveillance systems and other services. Public market focus on business people, young people and home users to provide online disk online storage services. Users can use these services to download files from the Internet online or offline to the storage space, and dynamically share to friends, friends can be indirectly shared. Compared with the original single-threaded, it improves the speed of the file spread and expand the transmission chain. (Liu 2016.)

AT & T is one of the largest telecom operators in North America. Currently. It offers cloud hosting services, cloud storage services and cloud storage services specifically for medical imaging. In the early North American market, cloud storage or cloud computing platform basically provided by the Amazon. In North America's user base, Amazon's EC2 and S3 are the synonymous with cloud computing. From personal to business, they all use the Web or third-party client to connect to Amazon's cloud services. (Liu 2016.)

Similarly, with the development of mobile Internet, users rely on the operators to provide access to the network more and more. Therefore, the most direct choice for users in North American is using cloud services which is provide by operators directly. Since 2008, more users emerge the demand of AT & T services, but they don't have the concept of AT & T's cloud service platform better cloud host cloud

storage Platform and a solution of personal and enterprise data at that time. Among of users, there is a very important customer: the US medical industry users. In the medical profession, the patient's medical records start using electronically for a long time. The massive accumulation of medical imaging, electronic medical records and other data, making the storage needs of various hospitals is huge. Data sharing between hospitals and hospitals, hospitals and medical research institutions have a high demand for the distribution storage of medical image data. (Zhu Liu 2016.)

With the AT & T's first cloud storage service which is for medical imaging officially launched, AT & T has built a service for personal cloud storage applications based on the same approach until the end of 2011. At the same time, cloud storage solutions and AT&T cloud host has been connected very well. (Liu 2016.)

AT & T's cloud storage service provides a resilient, scalable storage architecture. This service can meet the user on-demand application space, according to the use of space to pay. This kind of basic cloud storage service provides similarly to the Amazon cloud storage services, and both of their underlying use is based on the object of the storage system. (Zhu Liu 2016.)

From this cloud storage service combined with the carrier's mode of view, it provides customers some benefits. For example, extensible requirements unlimited access on demand on-demand storage. And application interface (API) can access online in anytime and anywhere. They also support the management of the network and multi-level security methods. (Liu 2016.)

Predefined storage policy selection, providing third-party program integrated cloud storage services. Flexible pricing model, so that customers choose to pay only the use of storage space or get the specific amount of storage for the AT & T Committed Storage Volume (CSV) option. (Liu 2016.)

#### 4.4 Cloud storage products comparison

TABLE 2. Three cloud storage products comparison. (Weibo 2016.)

	OneDrive	Dropbox	Google Drive
File size limit	2GB	No limit for the application	5TB
Free space	15GB	2GB	15GB
get extra space?	Yes	Yes	No
Pay plan	2 dollar/month 100GB 4 dollar/month 200GB	10 dollar/month 1TB	2 dollar/month 100GB 10 dollar/month 1TB

	7 dollar/month 1TB		
Support platform	Windows, Mac, Android, iOS	Windows, Mac, Linux, Android, iOS, Blackberry, Kindle Fire	Windows, Mac, Android, iOS,

The table shows the comparison with three companies. At first, compare different field of three companies. Then shows the best company for every field.

One Drive is the product from Microsoft. The user who using Windows8 and above this version will find this function in basic system. They can find out the file path in the browser interface. But any user can download this service from the OneDrive application store. The previous versions users of Mac and Windows8 can download and installed the desktop software. Android, iOS, WP, and Xbox users can also download applications on their platforms to use the OneDrive service. Figure 8 will show the Google drive file path. (Weibo 2016.)

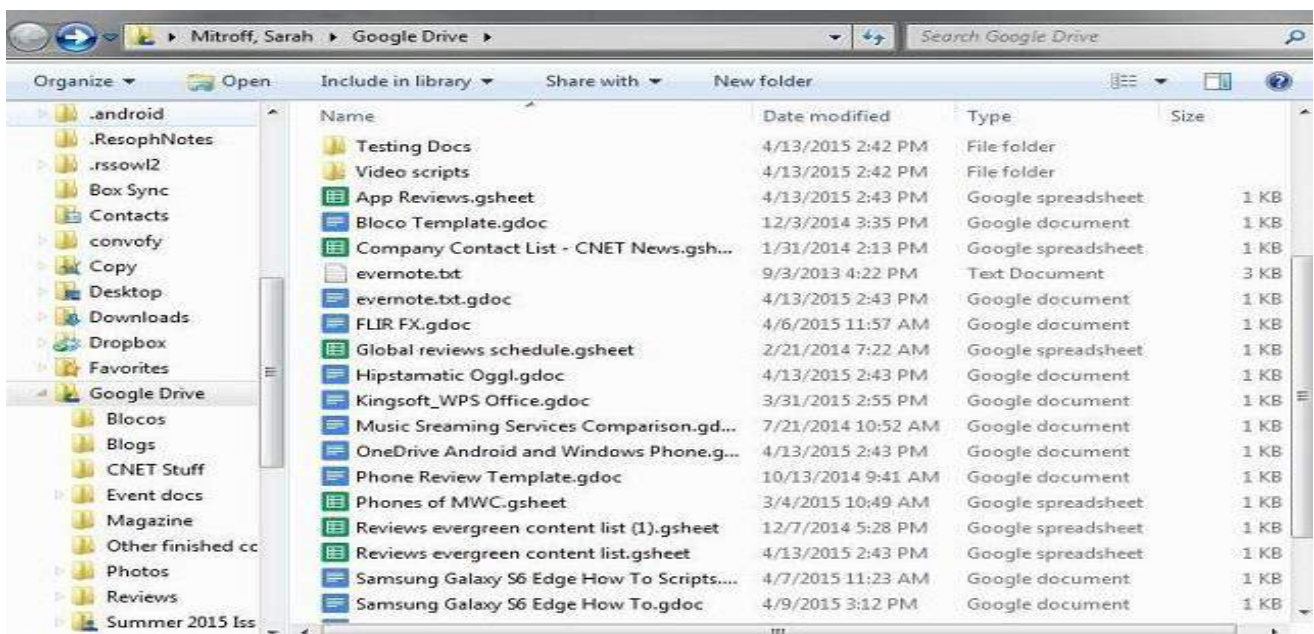


FIGURE 8. Google Drive file path

Advantage of OneDrive is that can connects to Windows devices directly, because it is built on the computer, tablet and mobile phone equipped with Windows operating system. It is easy to open and edit files on OneDrive on other Microsoft applications. Because OneDrive is closely connect with Office software. OneDrive is a good choice for those people who use Office software frequently. (Weibo 2016.) Disadvantages is if people do not have all the Windows devices OneDrive's appeal is not huge enough. Although other devices have a corresponding application, OneDrive is born for Windows. In order to

use OneDrive, people must register a Microsoft account and then you can use Outlook, Xbox Live and other Microsoft services. The need for these services up to users. Currently OneDrive has a limit of twenty thousand files. Microsoft say that it was solving the restrictions, but the solution time is still unknown. OneDrive is suitable for the people who use Windows computers, tablets and mobile phones, and people have to access files easily from these various devices. (Weibo 2016.)

The picture shows the people who has large number of different devices to share in different drive. Dropbox is popular in cloud storage because it is reliable, easy to use and easy to set up. After the file is uploaded to the cloud. People can access it through the Dropbox website, Mac, Windows and Linux desktop applications, or iOS, Android, Blackberry and Kindle Fire mobile applications. (Weibo, 2016.)

The biggest advantage of Dropbox is that all the equipment are very good. Application is simple to use. It has elegant design. Everyone can get started. Desktop applications connect very well with computer file system. Disadvantages is that Dropbox's website design is the most primitive of all cloud storage service applications. Simple and clean, but people can not choose the file display mode. However, users can get sharing options on the Dropbox website, which almost cover the lack of terrible design. Dropbox is suitable for the people who share their files using a large number of different devices. Figure 9 will show the Dropbox user interface. (Weibo 2016.)

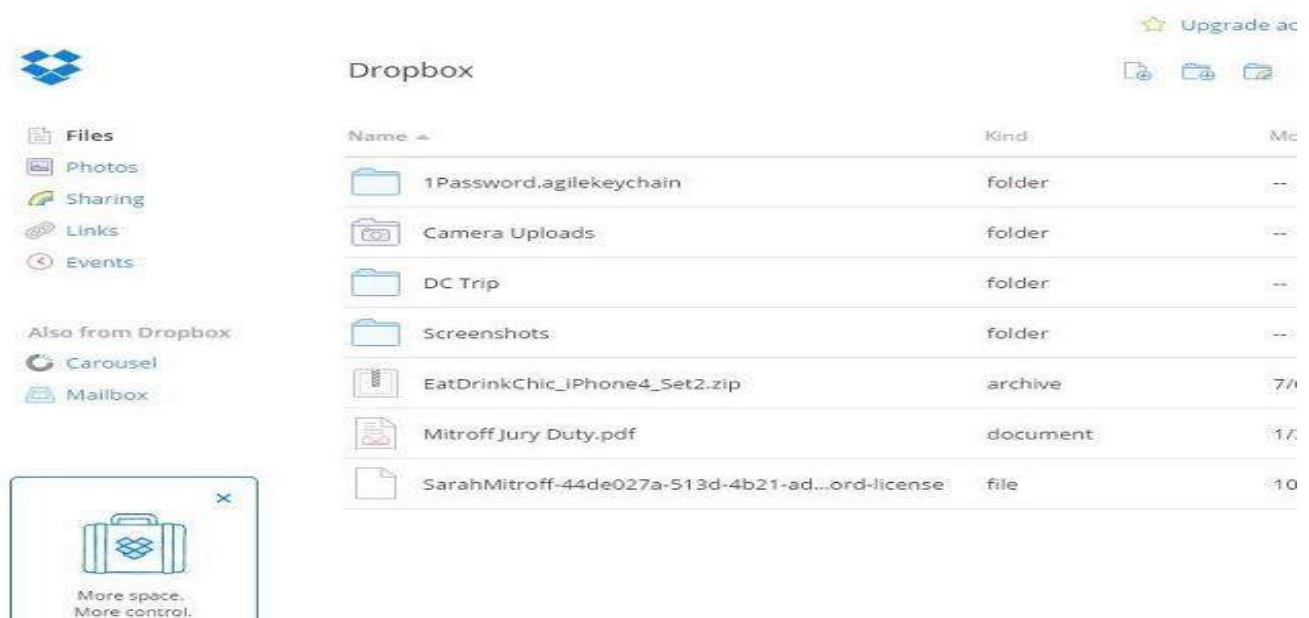


FIGURE 9. Dropbox user interface

4 years ago, Google Drive was a series of online office tools Google Docs, and then it evolved into Google Drive. With Google Drive people can get text editors, form tools and slideshow editing tools, plus 15GB of free cloud storage. Advantage of Google Drive is if users have a Google account, they do

not need any setup. More importantly, if they use Gmail, only need to click the next button then it can be stored in the cloud. Figure 10 will show the Google drive application screenshots. (Weibo 2016.)

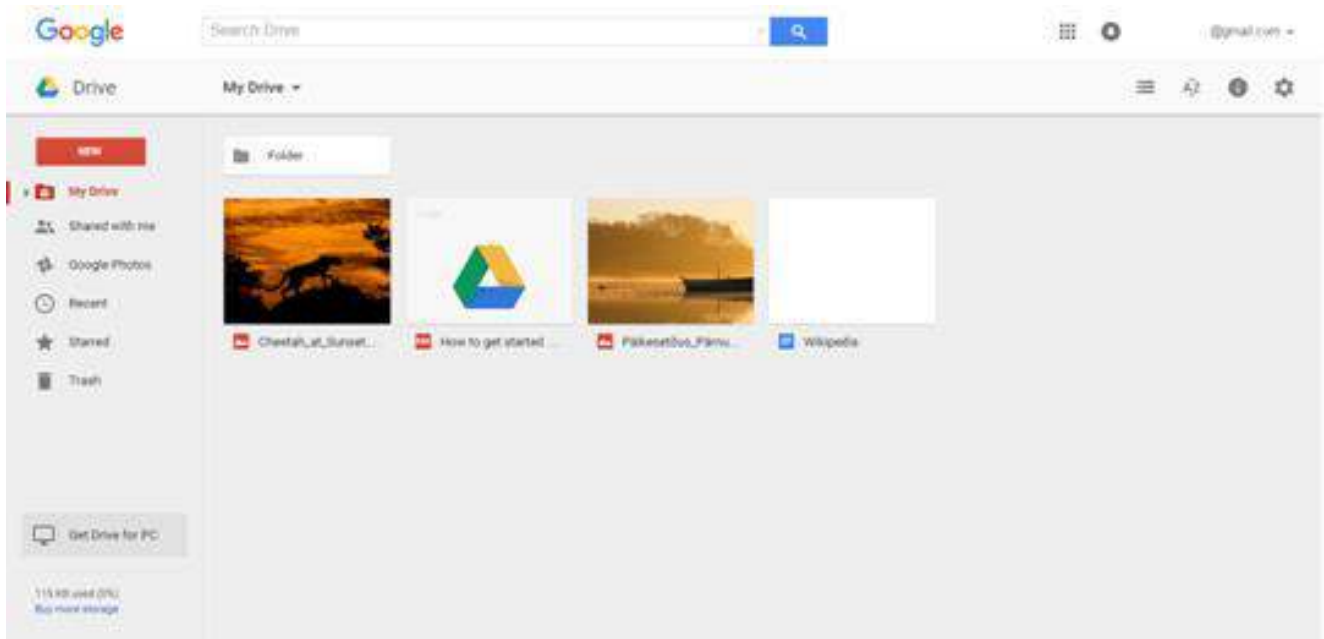


FIGURE 10. Google Drive application screenshots.

Disadvantages is that although people can manage files and photos on Google Drive, they can not automatically upload mobile photos to the cloud. But Google's Google+ app has an automatic backup feature, and photos will be sent to Google+. People hope that Google will let them like the image editing feature of Google+ and the file editing service of Google Drive which is showing as follow. Google Drive is suitable for Google fans, or the people who want to use cloud storage with office tools for users. (Weibo 2016.)

## 5 CONCLUSION

Clouds create unique requirements for data on security performance data protection availability and manageability. In order to dispel a lot of worries. Potential cloud adopters' requirements must be systematically addressed, and the concept of virtual storage containers provides a useful structure for thinking about how to meet these requirements. By meeting the needs of storage providers or cloud service providers will be able to create multi-tenant storage infrastructure security flexibility high functionality and interoperability. With the rapid development of communication and internet technology, mobile communications and the internet become the largest potential market and most attractive areas. Using distributed cloud storage technology to store terminal data, while ensuring the integrity and security of data. The overall architecture design this system follows terminal-oriented C/S model.

In this article, giving the research situation of cloud storage system. Analyzing the status of the development of cloud storage system. On this basis, the article investigates the key technology involved in-depth study, including two ways of storage that are virtualization and distributed storage technology. Researching the cloud terminal technology including architecture and features of the cloud platform cloud storage application components and lifecycle and cloud storage application development environment. Researching distributed storage technology includes distributed systems network storage and distributed file system technology. This article focus on file life cycle literacy principles and file transfer principle. On the basis of in-depth study of these key technology, chapter 3 have analyzed the technology people need and the appropriate method people should do like that. Then find out the system feasibility analysis and demand analysis. Based on the technical feasibility and economic viability of the system, verify the feasibility and design the business process. Chapter 4 introduce and compare four popular cloud applications, which find out the advantages and disadvantages and suitable ways for all kind of people.



## REFERENCES

- Amazon, Inc., “Amazon EC2 Service Level Agreement.” <http://aws.amazon.com/ec2/sla/>. Accessed: 2016-04-04.
- Boss, M. Security & Privacy, 2012.
- Barnett C. & Barnett O. 2004. Embedded C Programming and the Atmel AVR. Clifton Park, USA Delmar Cengage Learning.
- Davey, B. A. and Priestley, H. A. Introduction to lattices and order Cambridge: Cambridge University Press, 2009
- Foster F. Kuhn, D. R. and Chandramouli, R. “Role-Based Access Control, Artech House,” 2010.
- Gollmann, D. 2005 Computer Security 2nd Edition: John Wiley and Sons.
- Han, J. Kamber, M. and Pei, J. 2006. Data mining: concepts and techniques.
- INCITS, A. American national standard for information technology, role based access control, ANSI, 2004.
- Jeremy, B. 4th edition 2004. Exploring Arduino: Tools and Techniques for Engineering Wizardry.
- L. M., Rodero-Merino, L., Caceres, J. & Lindner, M. (2009). A Break in the Clouds. Towards a Cloud Definition. Computer Communications Review, Vol. 39, No. 1, pp. 50—55.
- Liu, J. and Qian, Z. 2016 “Mining Boolean and General Fuzzy Weighted Association Rules database,”.
- Lu, J. and Qian, Z. “Mining Boolean and General Fuzzy Weighted Association Rules database,” 1999.
- R. Van Renesse, K. P. Birman, and W. Vogels, May 2003 Astrolabe: A robust and scalable technology for distributed system monitoring, management, and data mining. ACM Trans. Comput. Syst., 21(2):164–206.
- SZALAY A S, KUNSZT P, THAKAR A, 2004. et al. Designing and mining multi—terabyte astronomy archives: The sloan digital sky survey.
- VARIA J. 2012. Cloud architectures—Amazon Web services [EB/OL]. [2009 — 03—01]. <http://acmbangalore.org/events/monthly-talk/may-2008-cloud—architectures—amazon-web-service.html>
- Weber, H. A. “Role-Based Access Control: The NIST Solution,” , 1997.