Applied Mechanics and Materials Vols. 530-531 (2014) pp 650-661
Online available since 2014/Feb/27 at www.scientific.net

doi:10.4028/www.scientific.net/AMM.530-531.650

Cloud Computing Survey

Gang Lu[1,a*], Wenhua Zeng[2,b]

[1]Software School of Xiamen University,Xiamen,Fujian,361005

[2]Software School of Xiamen University,Xiamen,Fujian,361005

[a]372058396@qq.com,[b]whzeng@xmu.edu.cn

Keywords: cloud computing ,system framework, Parallel computing, Map/Reduce, Virtualization.

Index Terms—cloud computing, core techniques, general framework

Abstract. Since the concept of cloud computing was proposed in 2006, cloud computing has been considered as the technology that probably drives the next-generation Internet revolution and rapidly becomes the hottest topic in the field of IT. The paper synthetically introduces cloud computing techniques, including the currently non-uniform definition and the characteristics of cloud computing; The paper also introduces the core techniques of cloud computing, such as data management techniques, data storage techniques, programming model and virtualization techniques. Then the 4-tie overall technique framework of general cloud computing is talked about. Finally, the paper talks about the obstacles and opportunities.

Reference

The evolution of computer has experienced from mainframe to PC and the appearance of PC meets individual requirements along with convenience. However, from a global perspective researchers have gradually realized the problems resulted from its distributed existence. The biggest one of the problems is idle computer resource, which further leads to waste of electricity, waste of material and environmental pollution. Problems will create requirements and requirements will create motive force. Cloud computing came into the world in the way like that. In 2006, the concept of “Cloud Computing” was first proposed at SES (Search Engine Strategies Conference & Expo) by Eric Schmidt, the CEO of Google. Later, cloud computing has rapidly become the hottest topic of the whole internet industry, in which people see the possibility of the next generation of internet technological change.

The concept of cloud computing has come into being for 7 years, however, cloud computing has not got a united definition. Here, we refer to the definition of M. Armbrust et al[1]:

“Cloud computing refers to both the applications delivered as services over the Internet and the hardware and system software in the data centers that provide these services. The services themselves have long been referred to as Software as a Service (SaaS); the data center hardware and software is what we will call a cloud. When a cloud is made available in a pay-as-you-go manner to the general public, we call it a public cloud; the service being sold is utility computing; we use the term private cloud to refer to internal data centers of a business or other organization, not made available to the general public. Thus, cloud computing is the sum of SaaS and utility computing, but does not include small or medium-sized data centers, even if they rely on virtualization for management.”

Of course, there are several other definitions from different perspectives [2, 3, 4, 5, 6, 7, 8] . Some of them are given based on the development of cloud computing, some are given from the perspective of participants and some are given from the perspective of infrastructures. From these definitions we can conclude the characteristics of cloud computing:

a. Infrastructures of cloud computing are build on large-scale and cheap server cluster. The goal of making use of hardware resource with maximum efficiency is reached by building top applications cooperated with infrastructures.
b. Based on virtualization technology. The use of virtualization technology makes business logic separated from computation resource, which improve the access ability of services.
c. Commercial model of pay-as-you-go. Such model can make cloud computing service gain dynamic and high expandability and immensely improve utilization rate of computation resource; as a result, energy consumption per service can be reduced effectively.
d. What it provides is service. No matter what it provides (infrastructures, application platforms or applications), it is provided in the form of service.

The rest of this paper is organized as follows. Section 2 introduces the core technologies of cloud computing and form of its service. Section 3 talks about overall technical framework of cloud computing. Section 4 talks about main problems of current cloud computing. Section 5 concludes this paper.

Core technologies

Cloud computing is a kind of calculation method that is build on large-scale server clusters and operates on mass data. Cloud computing requires some unique technologies to implement the calculation method. The key technologies cloud computing depends on include data storage technology, data management technology, programming model and virtualization technology.

1.1 data storage technology

To ensure high availability, high reliability and economical efficiency, cloud computing adopts distributed data storage and ensure reliability with the help of redundant storage. In general, cloud computing systems have to satisfy requirements of a large number of users simultaneously and provide service for them parallelly, which means that data storage technologies used in cloud computing have to be high-throughput and high-transmit.

Until now, representative data storage technologies in the cloud computing include GFS (Google File System) and HDFS (Hadoop Distributed File System).

The design of GFS needs to ensure performance, reliability and scalability; besides, it is also affected by google's application load and technical environment[9].

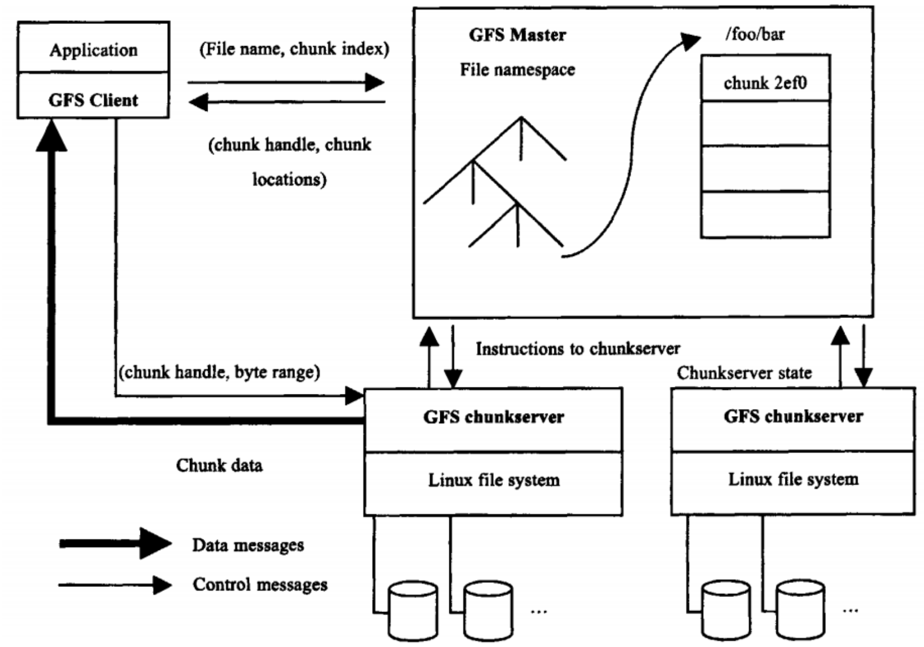

Figure 1. GFS system framework

The framework of GFS is shown as figure 1. A GFS cluster contains a master server (Master) and several chunk servers (Chunk). Master is responsible for storing all metadata, including namespace, access control, information about how to split file and information about where a file block can be accessed. Chunk is responsible for storing file blocks; in the process large file will be first separated into several fixed-size blocks (64M) and then each block will be stored in a chunk. When a client wants to access data, it will first request the Master to get information about the chunk(s) where the target data is stored; then, it will access the chunk(s) interactively based on the location information.

The method of redundant storage is adopted to ensure reliability of the data in GFS. Three other backups are saved for each file block by default. To ensure data consistency, each modification will be operated on all backups; a version number will be used to confirm whether all backups are consistent.

GFS separates the writing operation control signals from the data stream. When a client gains a writing authorization, the data to be updated will be transferred to all data duplicates; then, the client will send out writing request control signals after all duplicates receive the updated data. After all duplicates are updated, main duplicate sends writing operation completion control signals to the client.

HDFS is part of the Apache Hadoop Core Project, it can be seen as the open source implementation of GFS. HDFS is a highly fault-tolerant system, suitable to be deployed in cheap machines, it can create multiple copies of data blocks, and those copies will be distributed on the computer nodes. At the same time, it uses consistency models, scilicet ”write once, read many times”.

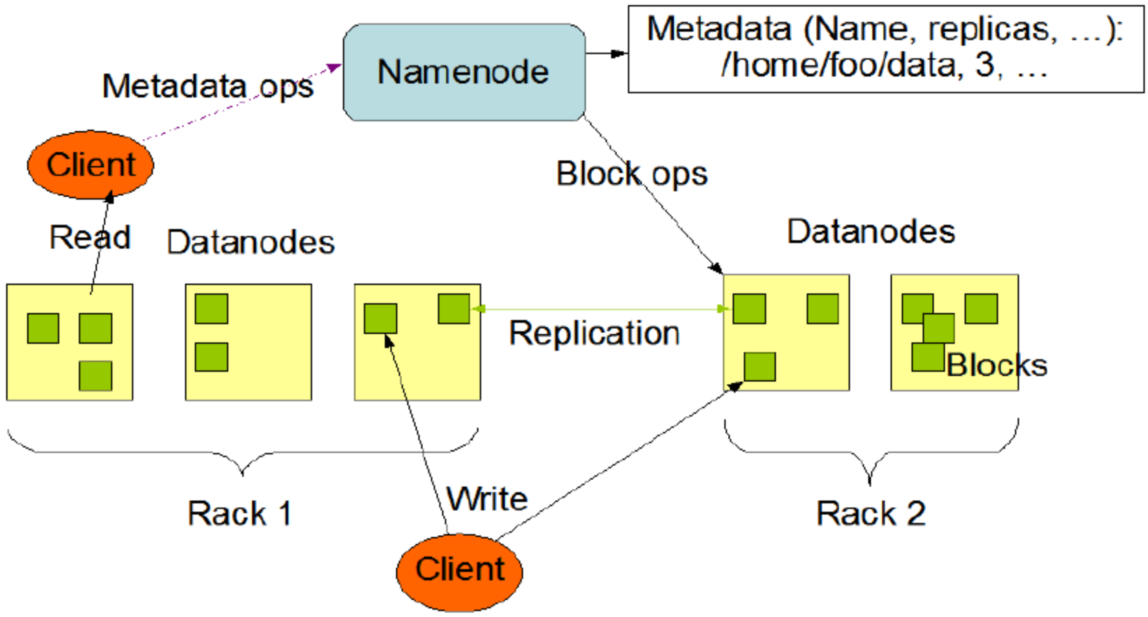

Figure 2. HDFS system framework

Figure 2 shows the HDFS system framework. HDFS uses the master/slave architecture, an HDFS cluster contains a Namenode and several Datanodes. A Namenode is a central server which is responsible for the namespace of file system, mapping data blocks to Datanodes and the access to files for the client. Generally, one Datanode is a single node , responsible for managing the storage of the node where it locates.

In addition to GFS and HDFS, Simple Storage Service (S3) of Amazon adopts the object-bucket model. Object is the basic unit of storage. It exists in the form of key-value pairs, and it's stored on the servers of the data center where the bucket is. The logical information of object is stored in the bucket and the bucket controls logical information using highly-reliable distributed hash table DHT.

1.2 Data management techniques:

The processing object of cloud computing system is large data sets, which requires corresponding data management techniques to manage the large data set efficiently. Secondly, how to find specific data in large data sets is also an issue which the cloud computing data management techniques must deal with.

Cloud computing is characterized by the mass data storage and a lot of analysis after reading. The frequency of reading data is much higher than the frequency of write-update, so the cloud data management have to be a kind of read-optimized data management.

A typical representative of cloud computing data management technology is Google's BigTable.

BigTable optimizes the data-read operation using the way of column storage. The storage structure which BigTable uses is <row: string, column: string, time: int64> ->string. BigTable's basic elements include line, column, record board and timestamp. Among these, the record board is a collection of rows. In the BigTable, data items are arrayed in accordance with the Lexicographical order of key words, each row is distributed to the recording board dynamically. The timestamp is a 64-bit integer which means different versions of data. Column family is a collection of several columns, the access permission of BigTable is controlled in the column family size. BigTable's logical structure is shown in figure 3.

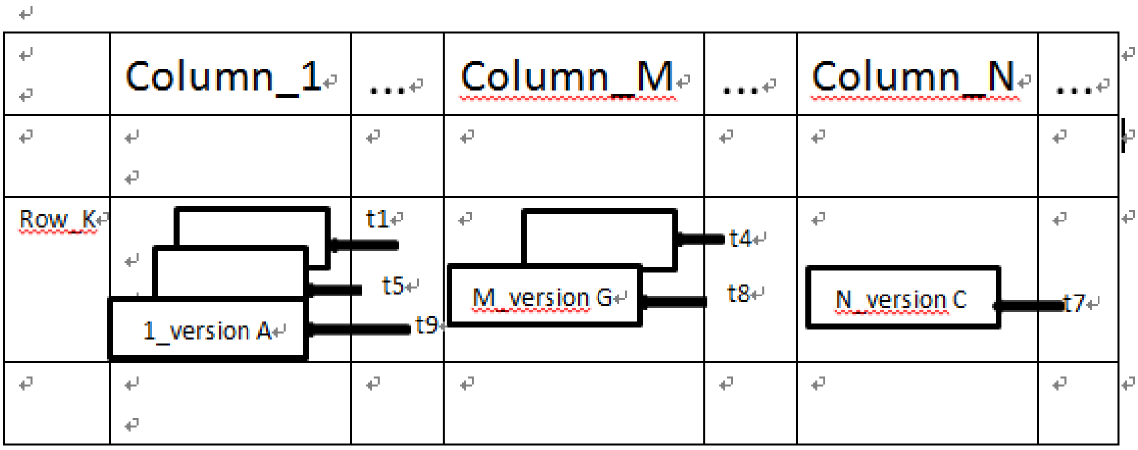

Figure 3. BigTable logical structure

In addition, in order to ensure that the data structure is of high scalability, BigTable adopts the three-tier form to store information of data location. When a client reads the data, firstly, it gets the location of Root Table from Chubby file, and reads the corresponding location information of METADATA tablet; then it reads the location of User Table which contains target data's location information from the METADATA tablet; afterwards, it reads the location information item of target data from the User Table; finally, according to the information, it reads the data from the server.

In addition to Google's BigTable, other data management technologies such as Microsoft's DryadLINQ systems [12], Apache's Hadoop Database (HBase).

1.3 Programming Model/Task Scheduler:

Parallel computing is the essential characteristic of cloud computing. Therefore, cloud computing system must have a appropriate programming model so that the users can use it to write the distributed computing programs.

The most famous programming model in the cloud computing is Google's Map / Reduce model [13,14,15], most of other programming models of cloud computing systems are the variants of Map/Reduce.

Map/Reduce uses two simple concepts -"Map" and "Reduce" to form the basic unit of computing, users only need to provide their own Map function and Reduce function in order to handle massive amounts of data in parallel. Map function is responsible for defining the operations on each sub-block data; Reduce function is responsible for defining the merging operations on the intermediate computational results of each sub-block data. Other operations such as partition, allocation and scheduling of the input data, response to cluster nodes' failure, communication management among nodes and so on are in the charge of MapReduce framework. Figure 4 shows the process of execution of a Map/Reduce program, Map is responsible for handling each block and writing the results to the intermediate file, Reduce is responsible for merging the results in the intermediate file and outputting the final result.

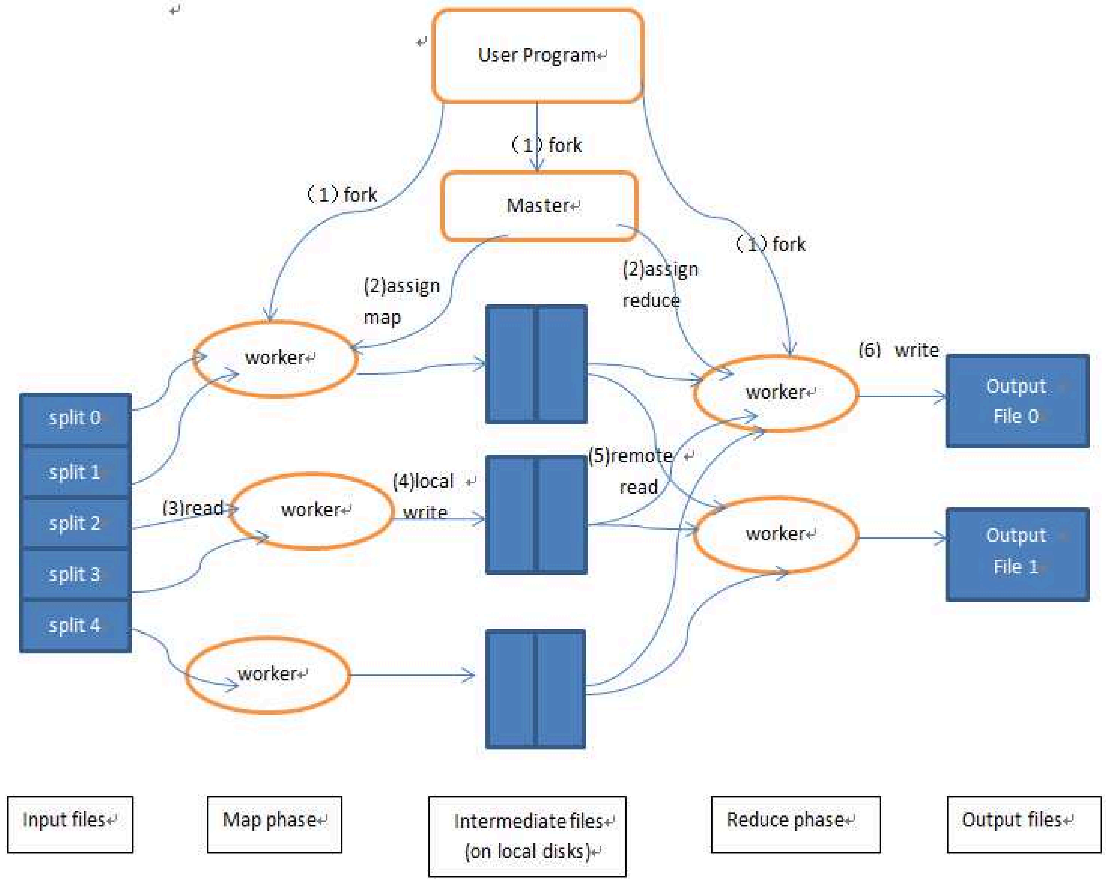

Figure 4. The process of execution of a Map/Reduce program

Hadoop's programming model is entirely based on the Map/Reduce framework, it consists of a single master JobTracker and slave TaskTracker each of each locates on a cluster node. Master is responsible for scheduling all tasks which constitute a job, these tasks are distributed in different slaves, master monitors their execution and re-execute the failed tasks. Slave is only responsible for executing the tasks assigned by the master.

1.4 Virtualization Technology:

Cloud computing requires the ability to dynamically cut and distribute computing resources, which is not an easy thing. Virtualization technology is the answer to this question. Virtualization in the cloud computing can be achieved in both hardware level and software level.

IBM's "Blue Cloud" computing platform is the typical representative of virtualization technology. "Blue Cloud" achieves virtualization in both hardware and software levels. Virtualization in the hardware level can be achieved with the use of IBM's p-series servers to get logical partition of hardware-LPAR, CPU resources in the logical partition can be managed by IBM Enterprise Workload Manager. In this way, plus the resource allocation strategies in the actual using process, it can legitimately distribute the resources to each logical partition. Virtualization in the software level achieves with Xen, with which it is can be a reality that multiple operating systems run on the same physical machine at the same time.

1.5 Service forms:

According to the service hierarchy of cloud computing, service forms of cloud computing are mainly as follows:

a. Software as a Service, SaaS

Software as a service's cloud computing deliver the services to the users through the browser. From the view of users, they can reduce expenses on setting up the server and buying software licenses; from the view of providers, they can reduce costs on the maintenance of their software. Salesforce.com is the most famous company in this field. Generally speaking, SaaS is often used in the field of human resources management and ERP.

b. Utility Computing

The idea of utility computing was put forward early, but it's not clearly until it is used in Amazon, Sun, IBM and other companies which provide virtual servers and storage service in recent years. Such cloud computing creates a virtual data center through IT industry so that it can put the memory, I/O resources, storage and computing resources together to construct a virtual resource pool to provide service for the entire network.

c. Platform as a Service, PaaS

PaaS provides the service in the abstraction levels on the cloud infrastructure, that is the software platform where the system runs, such as the development platform, commercial deployment platform, the application platform and so on. PaaS shields the users from the operation system, hardware and complexity of storage, so it requires that the providers should have good development capabilities and some resource management capabilities. In addition, cloud platforms provided by most providers is limited to particular programming languages and integrated development environment, for example, GAE only supports Python and the corresponding IDE.

d. Infrastructure as a Service, IaaS

IaaS provides users with computers (physical machines or virtual machines), storage space, network connectivity, load balance, firewall and other computing resources with the network; users deploy and run a variety of softwares on this basis.

Overall technical framework of cloud computing

The basic technical framework of cloud computing is divided to four layers [16]: the underlying infrastructure layer, the virtualization layer, the service resource layer and the service providing layer, as shown in Figure 5.

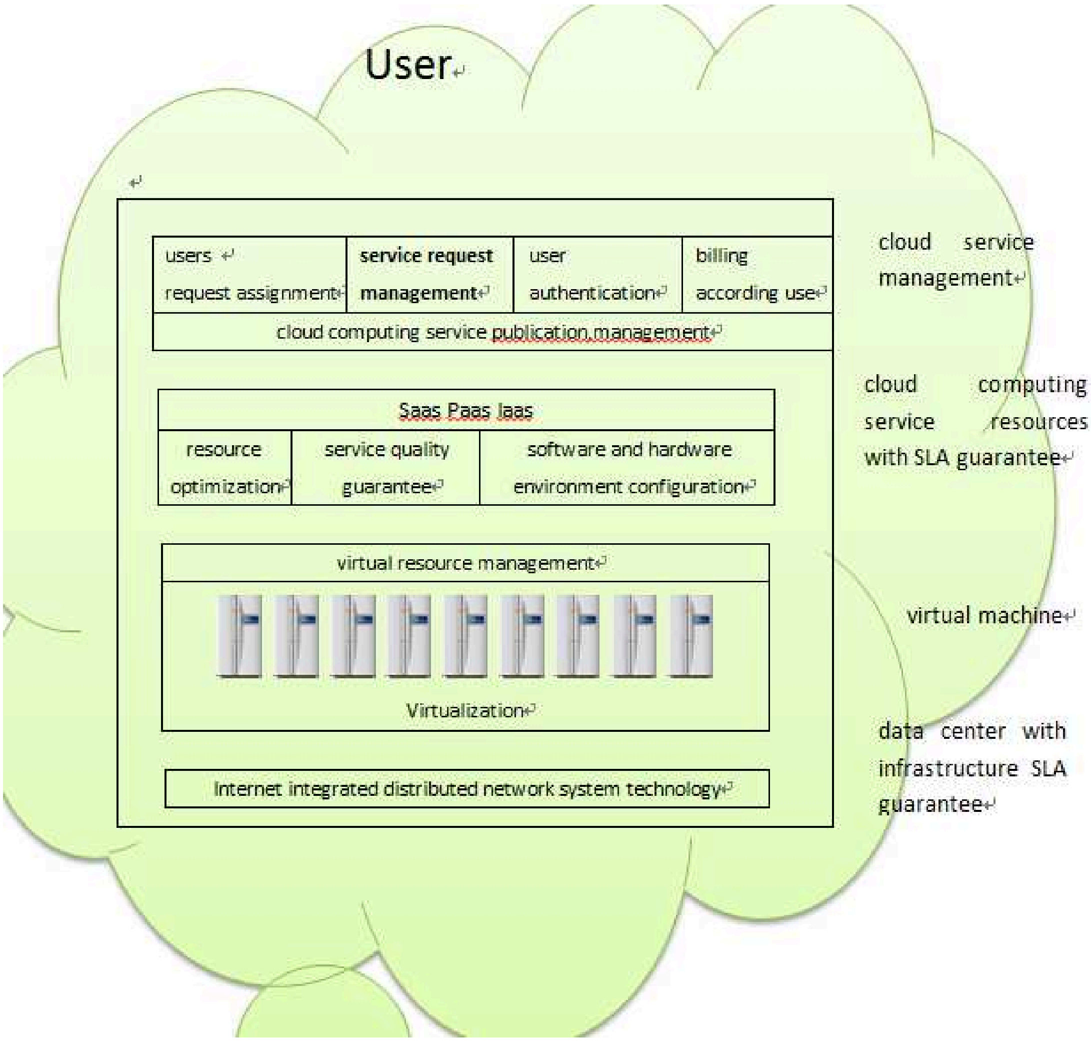

Figure 5. Overall technical framework of cloud computing

a. Tier 1: the underlying infrastructure layer, the physical machines compose the large data center with SLA guarantee through the network and the service encapsulation of the cluster technology[17].
b. Tier 2: the virtualization layer, based on the infrastructure data center or large-scale distributed systems, with the use of the virtualization technology, this layer creates a virtual layer to provide the virtual machines which are logically independent. This mechanism enables the cloud computing framework to dynamically separate and manage computing resources. Michael Armbrust et al[1] think that in order to support cloud service on the upper layer, the virtualization mechanism need to meet three conditions, which are as follows:
   i. The illusion of infinite computing resources available on demand, thereby eliminating the need for Cloud Computing users to plan far ahead for provisioning.
   ii. The elimination of an up-front commitment by Cloud users, thereby allowing companies to start small and increase hardware resources only when there is an increase in their needs.
   iii. The ability to pay for use of computing resources on a short-term basis as needed (e.g., processors by the hour and storage by the day) and release them as needed, thereby rewarding conservation by letting machines and storage go when they are no longer useful.

c. Tier 3: service resource layer. It manages the resource directly based on the virtual resource platform and provides the cloud computing service resources with SLA guarantee. There exists a certain distance from the resource management on the virtualization layer to the cloud computing service with SLA guarantee, in addition of this, the united management of the resources on the virtualization layer is essential.
d. Tier 4: the service providing layer, in addition to related management technologies of the platform, cloud computing also need to provide management measures related to the business, and provide users with service's publishing and management based on the service resources, including users' authentication, request management, request distribution and other functions, among of which the most important is pay-as-you-go.

Existing successful examples of cloud computing systems almost meet the 4-tier partition on the whole, such as Google AppEngine, Microsoft Azure, Amazon WebService. The introduction to these three instances has appeared on many research articles, so it doesn't need to be described again. Here, we describe only the comparison between the three resource models of the cloud computing system, as shown in Table 1.

TABLE 1. THE COMPARISON OF RESOURCE MODELS AMONG THREE CLOUD COMPUTING SYSTEMS

	Google AppEngine	Microsoft Azure	Amazon WebService
Virtual machine's computing model	Have defined application's framework and architecture in advance, providing the programmers with the processing service based on Python; all the storage of persistent states is put in the MegaStore; fully automatic changes in the computing scale and the stored resource scale.	Based on the Microsoft Common Language (CLR) virtual machine technology; provide automatic load balance; computing models are supervised and managed; it can't completely execute all commands, limited to CLR; some basis functions including load balance don't need developers' concerns.	Provide the elastic computing resources which are easily monitored; developers can make use of instances to carry out the computing tasks in any form; developers need to burden the tasks to build their own instances or let the third party provide supports.
Storage model	（1）Distributed large-scale database management system-BigTable （2）The scalable storage model which faces users' application and is based on BigTable Google storage model is specially designed for Google applications, not for others.	（1）SQL data service（2）Azure storage service The API defined by Microsoft is needed and users hardly use their own custom storage service.	According to the service chosen by the users, the platform uses the storage model-EBS without automatic adjustment or the storage model-S3 with the adjustment, different models provide different guarantees.
Network model	Fixed topology based on the three-tier network architecture, whose scale change is invisible to programmers.	Automatically assign it based on the programmers' description to application components (ROLE).	Based on the open standards on IP layer, internal details are hidden. The security group provides the limits to communication on each node.

Challenges and opportunities

a. Security issues

All the time, the security of data and application in cloud computing is one of the most contentious focuses, and is also one of the biggest factors that prevent the companies from moving to the cloud platform. The security issues related to cloud computing can be divided into technology issues and non-technology issues.

The security problems about technology include threats from outside the cloud and threats from inside the cloud. The outside threats are similar to the ones that large data centers faces now. Compared with the external threats, the internal threats need more concerns, such as steal behaviors or denial-of-service attacks that the cloud registered users do, and the users' isolation in the resource-sharing condition is a big concern. Under the circumstances where the cloud computing technology is immature, multiple users unconsciously share a physical resource, which may bring some security risks, for example, only Hadoop's latest version supports the cluster access control and authentication on users' level, which had been the Hadoop's potential risks.

The security problems about non-technology mainly refer to the policy and legal issues due to the across localization that are brought by the distributed nature of cloud computing. For instance, the distributed storage strategy of cloud computing may break through the monitoring range of the local government, some sensitive data's loss or leakage may bring the security risks on economic, political and other aspects. So a nation should establish unified regulatory policies to ensure that the cloud computing develops and is applied safely and efficiently.

b. The lack of standardization issues

The lack of the standardization of cloud computing platforms prevents the interoperability between different cloud computing platforms and the migration of the users' data and applications. It leads to such situation that a service can be obtained only from one provider, which precisely becomes a hampered factor when enterprises consider migrating to the cloud. After the selecting a service, the availability of this service is entirely ensured by this cloud provider, the companies often have not enough trust on it, especially for the data storage service, either cloud providers' operation errors or companies' collapse may seriously affect the availability of the data. However, if there is a unified standard, the situation will be different, because at this time the availability of the service can be jointly guaranteed by multiple providers. About this, Michael Armbrust et al [1] mentioned it when they discussed 10 obstacles and opportunities of cloud computing.

Currently, the standardization research of cloud computing is in its infancy, and some cloud computing organizations or vendors try to define the standards of cloud computing. For example, Open Grid Forum set up Open Cloud Computing Interface Group, which is responsible for studying and defining the remote infrastructure management APIs[19]. Open Cbud Onsortium is studying the standards which aim to improve the interoperability of cloud computing, and some cloud computing providers also begin to take the interoperability issues into account. For instance, Eucalyptus can provide the interfaces compatible with Amazon's AWS, that is to say, users can use Eucalyptus's cloud computing service just like using EC2 or S3 service on Amazon, and it's easy to switch the service between Eucalyptus and Amazon AWS [21]. In early 2010, the International Telecommunication Union –ITU set up the "special working group of cloud computing", who is responsible for assessing the current standards and creating the common standards of cloud computing according to the conditions, in order to change the current situation where cloud computing doesn't have standards.

c. Regulatory and legal issues

The virtual and the transnational nature of cloud computing leads to many problems on regulatory and legal aspects. The first is boundary issue of data and applications of cloud computing, cloud computing makes it possible for the data storage to exceed the supervision range of the local government, or the situation where it is not compatible with the local policy of data storage to happen; the second is the qualification examination issues of the cloud computing providers, people's concerns requires corresponding examine on the cloud computing providers to ensure the security of the data and the quality of the service.

d. Other issues

There are many other factors which prevent cloud computing from developing, such as the transmission problems of large data sets; the capability can't be predicted under the resource-sharing condition, especially for the capability of the network, the disk I / O and the parallel tasks' dispatch; the ability to dynamically and quickly adjust the scale according to requirements is lacking, especially the dynamic adjustment of the storage scale; existing software licensing model isn't suited to the pay-as-you-go model of cloud computing.

Obstacles and opportunities coexist. Each of the problems above is also a direction of the cloud computing future research. The solution to each problem may take cloud computing one step forward.

Conclusion

As the technology which probably drives the next Internet revolution, cloud computing is worth researches' further study and discussion. This paper introduces the definition and characteristics of cloud computing; then introduces the core technologies which cloud computing relies on, including the large-scale data storage technology, data management technology, programming models and virtualization technology; then introduces the overall technique framework of general cloud computing, and compare the resource models of three cloud computing platforms. Finally, it talks about the main challenges about the current development of cloud computing.

References

[1] Michael Armbrust, Armando Fox et al. Above the Clouds: A Berkeley View of Cloud Computing. Tech. Rep. UCB/EECS-2009-28, EECS Department, U.C. Berkeley, 2009

[2] Foster I, Zhao Yong, RaicuI, et al. Cloud computing and grid computing 360-degree compared [EB/OL], http://arxiv.org/abs/arxiv09010131 , 2010.

[3] Buyyaa R, Chee S Y, Srikumar V, et al. Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility[J]. Future Generation Computer Systems, 25 (6) : 599-616, 2009.

[4] Luis M V, Luis R M, Caceres J, et al. A break in the clouds: towards a cloud definition [J] . ACM SIGCOMM Computer Communication Review, 39 (1) : 50-55, 2009.

[5] Boss G，Malladi P，Quan D，Legregni L，Hall H．Cloud computing．IBM White Paper, 2007. http://download.boulder.ibm.corn/ibmdl/pub/soRware/dw/wes/hipods/Cloud_computing_wp_final_80ct.pdf

[6] Cloud computing. http://en.wikipedia.org/wiki/Cloud_computing

[7] Liu Peng, Cloud computing[M], Beijing: Electronic Industry Press, 2010.

[8] Chen Kang, Zheng Weimin. Cloud computing: System instance and Research condition [J]. , Journal of Software ,2009

[9] Ghemawat S，Gobioff H，Leung ST．The Google file system．In：Proc．of the 1 9th ACM Symp on Operating Systems Principles. New York：ACM Press．2003．29—43．
[10] YU Y, ISARD M, FETTERLY D et al. DryadLINQ: A system for general purpose distributed data-parallel computing using a high-level language. Proceedings of the 8th USENIX Symposium on Operating System Design and Implementation. New York: ACM Press, 2008: 1-14.
[11] CHANG F, DEAN J, GHEMAWAT S, et al. BigTable：A distributed storage system for structured data [J]．ACM Transactions on Computer Systems，2008，26(2)：1—26．
[12] YAN B Q, RHODES P J. Toward automatic parallelization of spatial computation for computing clusters. Proceedings of the 17th International Symposium on High Performance Distributed Computing. New York: ACM Press, 2008: 45-54.
[13] Dean J，Ghemawat S．MapReduce：Simplified data processing on large clusters．In：Proe. of the 6th Symp．on Operating System Design and Implementation．Berkeley：USENIX Association，2004．137-150．
[14] Dean J, Ghemawat S. Distributed programming with Mapreduce. In: Oram A. Wilson G, ads. Beautiful Code. Sebastopol: O'Reilly Media, Inc．，2007．371—384．
[15] Dean J, Ghemawat S. MapReduce: Simplified data processing on large. Clusters. Communications of the ACM, 2005, 51 (1): 107-113
[16] Wang Jiajuan, Lv Zhihui, Wu Jie, Zhong Yiping. Cloud computing technology development analysis and applications discussion. Computer Engineering and Design, 2010, 31 (20).
[17] Francesco Maria Aymerich, Gianni Fenu,Simone Surcis. An approach to a cloud computing network [C]. First International Conference on the Applications of Digital Information and Web Technologies, 2008:113-118.
[18] Narsimha C, Siddalinga K M, Rahman. Security implementation in Hadoop [EB/OL]. [2010-7-2]. http://search.iiit.ac.in/cloud/presentations/28.pdf
[19] Open Grid Forum. Open cloud computing interface WG (OCCI-WG) [EB/OL]. http://www.gridforum.org/News/news.php?id=145 .
[20] Open Cloud Consortium. Open cloud consortium [EB/OL]. http://opencloudconsortium.org/
[21] Eucalyptus Systems Inc. Amazon web services compatibility [EB/OL]. http://www.eucalyptus.com/
[22] International Telecommunication Union. Focus group on cloud computing (FG cloud) [EB/OL]. http://www.itu.int

Advances in Measurements and Information Technologies
10.4028/www.scientific.net/AMM.530-531

Cloud Computing Survey
10.4028/www.scientific.net/AMM.530-531.650

DOI References

[15] Dean J, Ghemawat S. MapReduce: Simplified data processing on large. Clusters. Communications of the ACM, 2005, 51 (1): 107-113.
http://dx.doi.org/10.1145/1327452.1327492