Status and Key Techniques in Cloud Computing

Juefu Liu
College of Information Engineering
East China Jiao Tong University
Nanchang ,China
JUEFULIU@yahoo.com.cn

Peng Liu
College of Information Engineering
East China Jiao Tong University
Nanchang ,China
liupengyjs@gmail.com

***Abstract*—Cloud computing is another buzzword and the latest effort after Web 2.0.It is a technology that access remote servers through Internet to maintain data and applications. Whilst cloud computing developes into a major trend in IT industry, academic research has recently joined.This paper compared existing definitions of cloud computing and gave personal opinion. The case of mainstream cloud computing platforms were introduced in helping to understand what's the cloud computing and then key techniques were analyzed.It also talked about the simulator-cloudsim and some aspects to be improved. The paper strives to describe the current status of cloud computing as well as its broad development prospects.**

Keywords-cloud computing; platform; key techniques;cloudsim

I. INTRODUCTION

With the advancement of the IT industry ,the emerging paradigm is that of Cloud computing which promises reliable services delivered through large-scale data centers that are based on compute and storage virtualization technologies.The consumers are supplyed with on-demand access to compute or storage resources at any time and they pay fees for their usage.Actually Leonard Kleinrock [1],in 1969,who has had the vision that "we will probably see the spread of 'computer utilities' which,like present electric and telephone utilities ,will service individual homes and offices across the country."

Industry observers say that cloud computing has great commercial potential.Market-research firm IDC expects IT cloud service spending to grow from about $16 billion in 2008 to about $42 billion by 2012 and to increase its share of overall IT spending from 4.2 to 8.5 percent.The cloud market as a whole is predicted to $160 bilion by 2011(source:Merrill Lynch).2010 will be the emergence of "enterprise-grade" cloud services [2].

Cloud computing has a bright future. Bioinformatics applications require heavy computation and management of huge amounts of data.M.Gaggero et al. [3] propose two relevant algorithms:BLAST and GSEA based on MapReduce and Hadoop.Qichang Chen et al. [4] use a high performance workflow system MRGIS to execute GIS applications efficiently.Aboulnaga,A. et al. [5] talk about how to deploy database appliances on the clouds.Andrew Newman et al. [6] present MapReduce-based RDF molecule store to distributed querying and reasoning.Kevin Beyer et al. [7] develop an enterprise content analytics platform that leverages the Hadoop mapreduce framework to support this class of analytics workloads.

II. DEFINITIONS

There are dozens of different definitions for cloud computing and there seems to be no consensus on what a cloud is.

Foster et al. [8] compare and contrast cloud computing with grid computing.They believe cloud computing is an evolved version of grid computing. Wikipedia [9] describes as "cloud computing is Internet-based computing, whereby shared resources, software and information are provided to computers and other devices on-demand, like the electricity grid. Michael Armbrust et al. [10] point out that cloud computing is the sum of SaaS and Utility Computing,but does not include private Clouds.

Luis M.Vaquero et al. [11] have studied More than 20 definitions. Chen G L. et al. [12] take the cloud computing as the new development of parallel computing.Buyya et al. [13] add that reach commercial it is necessary to strengthen the role of Service-Level Agreement(SLA).The NIST definition, depicted in Fig.1,the clearest and most comprehensive definitions of cloud computing , describes it as having five essential characteristics,three service models,and four deployment models [14].From my personal view,cloud computing is a model that dynamically supply real-time resources to users with different requirements by high-speed Internet in virtual resources pool and charge fees correspondingly.

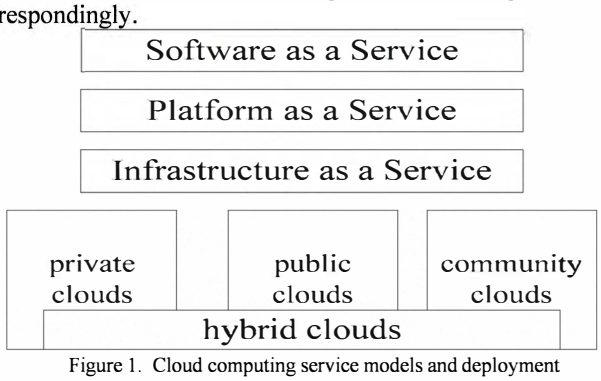

Figure 1. Cloud computing service models and deployment

The essential characteristics are summaryed as follows:(1).virtual resource pool;(2).QoS and SLA;(3). reliability and rapid elasticity;(4). measured service;(5). autonomy.

978-1-4244-6542-2/$26.00© 2010 IEEE

III. EMERGING CLOUD PLATFORMS

With the computing industry shifts toward providing PaaS and SaaS for consumers and enterprises,there will be an increase in the number of cloud platforms available to meet the needs.Several industrial organisations,academic and Open Source Initiative (OSI) have developed technologies and infrastructure.

A. Commercial Cloud Platform

Amazon Elastic Compute Cloud (Amazon EC2) [15] is a web service that provides resizable compute capacity for developers in the cloud. It works in conjunction with Amazon Simple Storage Service (Amazon S3), Amazon SimpleDB and Amazon Simple Queue Service (Amazon SQS) to provide a complete solution for computing, query processing and storage across a wide range of applications.The EC2 charges the user when the instance is alive.

Google Infrastructure for Cloud Computing consists of Google File System(GFS), BigTable, MapReduce and Chubby.

Blue Cloud ,proposed by IBM, adopts Xen,PowerVM and Hadoop to construct cloud computing environment for consumers.

Windows Azure Service Platform [16] includes the.NET Framework,SQL Services,Live Services and Windows Azure.

B. Academic Platform

High-level market-oriented cloud architecture [13] is proposed by Buyya.User/Brokers,SLA Resource Allocator,VMs and Physical Machines are four main entities.

Zhang Y X et al. [17] hava researched transparent computing system.It has three layers: transparent client, transparent network and transparent server.

Nuno Santos et al. [18] design a trusted cloud computing platform(TCCP) to attest the service is secure.

Luis M.Vaquero et al. [11] present a Cloud Actors to gain in flexibility and reduce costs.

C. Open Source Cloud Platform

Hadoop [19] is mature and widely used open-source software for reliable, scalable, distributed computing. A wide variety of companies and organizations use Hadoop for both research and production. Hadoop Distributed File System (HDFS) is the primary storage system used by Hadoop applications and creates multiple replicas of data blocks to enable reliable, extremely rapid computations.

There are other open-source cloud platforms ,for examples,Ecualyptus [20],ECP [21],Nimbus [22],etc.They are helpful to understand the cloud computing platform.

IV. KEY TECHNIQUES

In this section ,key technologies of cloud computing will be introduced:

A. Virtualisation technologies

Three commonly used virtualisation technologies are:Vmware Infrastructure,Xen and KVM.

VMware is the leading business virtualization infrastructure provider, offering the most trusted and reliable platform for building private clouds and federating to public clouds.Xen is an open source GPL project managed by XenSource. Xen open source community developed and maintained VMM and Hypervisor. KVM consists of a loadable kernel module, kvm.ko, that provides the core virtualization infrastructure and a processor specific module, kvm-intel.ko or kvm-amd.ko.

With increasing availability of VM technologies,the creation of customised environments on top of physical infrastructures is reality. The use of VMs brings several benefits such as: improved security, dynamic provision of VMs to services, server consolidation, performance isolation and so on. Systems based on virtual machines enable users to create virtual workspaces or virtual clusters atop theactual physical infrastructure to manage a cluster of computers [23].

B. Security Management

Security is of paramount importance to the flourish of cloud computing,yet it is still challenging today.It is the inhibitor to cloud adoption and security concerns are far beyond cloud to a certain extent. To tackle this area,researchers have obtained some achievements.

Siani Pearson [24] recommended privacy practices for cloud system designers,architects,developers and testers:

- Minimise personal information sent to and stored in the cloud
- Protect personal information in the cloud
- Maximise user control
- Allow user choice
- Specify and limit the purpose of data usage
- Provide feedback

ACM suggested that Integrity of the cloud infrastructure is ensured through the use of Trusted Computing [25]. Gartner Symposium presented that Offline VMs need to be protected form Tampering.Advanced security tools are needed.Cloud security alliance(CSA) helps with cloud security in security guidance and reference model.

C. Programming model

MapReduce is a programming model and an associated implementation for processing and generating large data sets [26]. The computation takes a set of input key/value pairs,and produces a set of output key/value pairs.The user expresses the computation as Map and Reduce.

Map takes an input pair and produces a set of intermediate key/value pairs. The MapReduce library groups together all intermediate values associated with the same intermediate key ,then passes them to the Reduce function.The Reduce accepts them and merges together these values to form a possibly smaller set of values.The model is described in Fig.2.

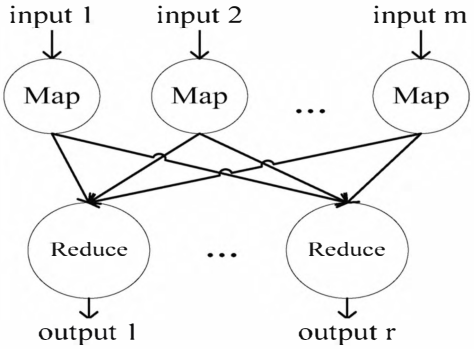

Figure 2.MapReduce model

Two main functions are written during programming:

$$Map:(in\_key,in\_value)\rightarrow\{(key_j),value_j \mid j=1...k\}$$

$$Reduce:(key,[value1,...,valuem])\rightarrow(key,final\_value)$$

The model is easy to use with parallel and distributed systems,since it hides the details of parallelization,fault-tolerance,locality optimization,and load balancing [26].A large variety of problems are easily solved with the ideas of MapReduce.It is a hot research direction to modify MapReduce.

D. Data Management

Bigtable is the most famous of data management in cloud computing.It is a distributed storage system for managing structured data to give clients dynamic control over data layout and format. Fig.3 is a organization of BigTable system [27].

Bigtable is combined of client LIBRARY,master server and tablet server,which are depicted in the Fig3. correspondingly.The client access to file directory after executing a open(),that is opening the lock.Then client communicates with tablet server.The Bigtable master performs metadata ops and load balancing between tablet servers.

Although varied demands,Bigtable has successfully provied a flexible,high-performance solution for all of the google products,which are web indexing,google earth and google finance [28].

Cloud monitoring,energy management and resource scheduling are also key technologies.

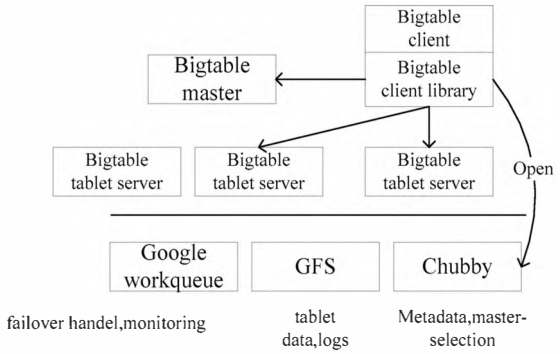

Figure 3. organization of BigTable system

V. CLOUDSIM

Cloudsim [29] ,leaded by Buyya, allows cloud customers to test their services in repeatable and controllable environment free of cost, and to tune the performance bottlenecks before deploying on real clouds.It provides a generalized, and extensible simulation framework that enables seamless modeling, simulation, and experimentation of emerging cloud computing infrastructures and application services.

Its functionalities:

- modeling and simulation of large scale cloud computing data centers
- modeling and simulation of virtualized server hosts
- modeling and simulation of energy-aware computational resources
- modeling and simulation of federated clouds
- dynamic insertion ,stop and resume of simulation
- user-defined policies for allocation of hosts to virtual machines and policies for allocation of host resources to virtual machines

The latest version is cloudsim 2.0 released on May 17, 2010.CloudSim has been tested and ran on Sun's Java version 1.6 or newer and Ant is also needed to compile Cloudsim.The simulate process using cloudsim is summarized as Fig.4.

There are examples and implementations of different policies for VM and Cloudlet schedulers provided with within the CloudSim package.CloudSim 2.0 contains support for model and simulation of hot topics in cloud computing, including green/power-aware cloud computing and federated cloud computing.We can implement our scheduling algorithms in subclasses of VmScheduler (for scheduling of VMs) and in subclasses of CloudletScheduler (for scheduling of Cloudlets). We also can implement our policies of cloudlets submission in a subclass of DatacenterBroker.

With the simulator ,we can schedule tasks in our scenarios.However some aspects of the simulator need to improve in personal view:

(1). Unfriendly interface.When the user want to compile and run the simulation,they must face the Java codes.

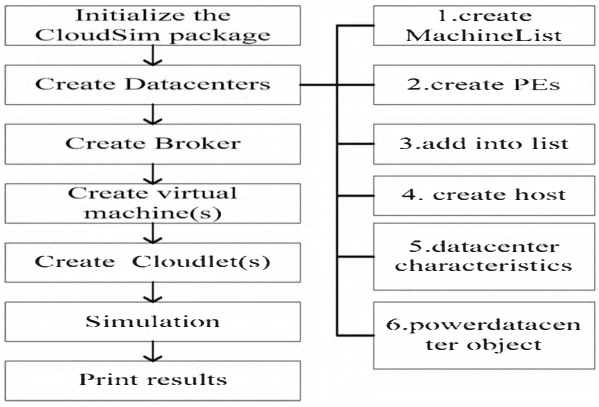

Figure 4. simulate process using cloudsim

The visualmodeler in gridsim is a good example .May be we could modify the simulator ,then let it have friendly interface to be used practically.

(2). Efficiency.When we develop our scenario,it is not so easy to complete new codes.It is complex and excrescent during amending the codes.

(3).To study the cloudsim is difficult ,for a lot source codes need to learn.

VI. CONCLUSION

This paper has presented the cloud computing's status,and the state-of-the-art techniques.It introduced the mainstream cloud platforms and talked about cloudsim..

In the future,the Internet of things combined cloud computing will promote the success of cloud computing.It is must to make the standards of the interface in cloud computing.For market-oriented cloud computing ,QoS and SLA should be enforced.Interoperability between clouds needs extended interaction protocols.Maybe Gloud(Grid+Cloud) [27] is coming after some years.Anyhow,we should believe that Cloud computing is promising.

REFERENCES

[1] L.Kleinrock, "A vision for the Internet," ST Journal for Research, Vol.2, No.1, Nov.2005,pp.4-5.

[2] http://www.idc.com [03 Dec 2009].

[3] Massimo Gaggero,Simone Leo,Simone Manca, "Parallelizing bioinformatics applications with MapReduce," http://www.cca08.org/papers/Poster10-Simone-Leo.pdf.

[4] Qichang Chen,Liqiang Wang,Zongbo Shang, "MRGIS:A MapReduce-Enabled High Performance Workflow System for GIS,"2008,pp.646-651,doi:10.1109/eScience.2008.169.

[5] Aboulnaga,A., et al., "Deploying Database Appliances in the Cloud,"Data Engineering, Vol.32, No.1,2009,pp.14-21.

[6] Andrew Newman,Yuan-Fang Li,Jane Hunter, "Scalable Semantics-The Silver Lining of Cloud Computing," 2008,pp.111-118.

[7] K.Beyer,V.Ercegovac,R.Krishnamurthy, "Towards a Scalable Enterprise Content Analytics Platform," Data Engineering, Vol.32, No.1,2009,pp.29-36.

[8] Foster,I.,Zhao,Y.,Raicu,I.andLu,S., "Cloud Computing and Grid Computing 360-Degree Compared," In Grid Computing Environments Workshop(GCE'08),Nov.2008,pp.1-10.

[9] http://en.wikipedia.org/wiki/Cloud_computing.

[10] Michael Armrust et al., "Above the Clouds:A Berkely view of Cloud Computing," Technical Report No.UCB/EECS-2009-28,Feb.2009.

[11] Luis M.Vaquero et al., "A Break in the Clouds:Towards a Cloud Definition,"ACM SIGCOMM,Vol.39,No.1,Jan.2009,pp.50-55.

[12] Chen G L,Sun G Z,Xu Y,et al., "Integrated research of parallel computing:Status and future,"Chinese Sci Bull,2009,54(11):1845—1853,doi:10.1007/s11434-009-0261-9.

[13] Buyya Rajkumar,Chee Shin Yeo,Srikumar Venugopal, "Market-oriented cloud computing:vision,hype and reality for delivering IT services as computing utilities," Technical Report, Aug. 2008.

[14] I.Sriram,A.Khajeh-Hosseini, "Research Agenda in Cloud Technologies,"ACM Symposium on Cloud Computing(SOCC2010),2010.

[15] Amazon Elastic Compute Cloud,http://aws.amazon.com/ec2/.

[16] D.Chappell, "Introducing the Azure services platform,"White paper,Oct.2008.

[17] Zhang Y X,Zhou Y Z, "Transparent computing:a new paradigm for pervasive computing," Ubiquitous Intelligence and Computing(UIC 2006) ,Berlin,Heidelberg:Springer-Verlag,2006.

[18] Nuno Santos,Krishna P.Gummadi,Rodrigo Rodrigues, "Towards trusted cloud computing,"In Proc.HotCloud,Jun.2009.

[19] http://hadoop.apache.org/.

[20] http://www.eucalyptus.com/.

[21] http://www.enomaly.com/.

[22] http://workspace.globus.org/.

[23] Marcos Dias de A., Alexandre di Costanzo, Rajkumar Buyya, "A cost-benefit analysis of using cloud computing to extend the capacity of clusters," Cluster Comput,Apr.2010, doi:10.1007/s10586-010-0131-x.

[24] Siani Pearson, "Taking account of privacy when designing cloud computing services,"ICSE'09 Workshop,2009,pp.44-52.

[25] Richard Chow et al., "Controlling data in the cloud: outsourcing computation without outsourcing control," Conference on Computer and Communications Security,2009,pp.85-90.

[26] JeffreyDean,SanjayGhemawat,"MapReduce: Simplified Data Process ing on Large Clusters," OSDI'04,Dec.2004.

[27] Liu Peng, "Cloud Computing, "Beijing:Publishing House of Electronics Industry,Mar.2010,pp.27-28.

[28] Chang F,et al., "BigTable:A distributed storage system for structured data," ACM Transactions on Computer Systems,26(2),2008,pp.1-26.

[29] http://gridbus.org/cloudsim/.