Several Public Commercial Clouds and Open Source Cloud Computing Software

Wenhua Zeng, Jianfeng Zhao, Min Liu
Cognitive Science Department. Xiamen University
Xiamen, Fujian, China
whzeng@xmu.edu.cn, zjfhi2008g@gmail.com

Jianfeng Zhao, Min Liu
Fujian Key Laboratory of the Brain-like Intelligent Systems(Xiamen University)
Xiamen, Fujian, China

***Abstract*—Cloud computing is usually used to describe the large-scale distributed infrastructure, platform and software services provided by third party, which is a hot topic in IT area in recent year. However, it is disputable in industrial and academic area that what is cloud computing and whether it is a novel technology. This paper attempts to draw a clear picture by the introduction of several public commercial clouds and open source cloud computing software, not to discuss the exactly concept of cloud computing. The commercial cloud computing Amazon EC2、 IBM smart cloud、Google App Engine and Windows Azure were introduced and the implementation of several open source cloud computing software such as Hadoop、Eucalyptus、openNebula and Nimbus were detailed. At last, several representative definitions were showed. The purpose of this paper is that a basic knowledge about cloud computing can be obtained by reading this paper.**

***Index Terms*—Cloud computing, Cloud services, Cloud computing software**

I. INTRODUCTION

The requirement for powerful computing drives people to spare no effects in developing new computing methods. The traditional concept of computing is broken up due to the appearance of cloud computing. Computing no longer means a mass of devices and complex management, but as a service sold on Internet. Computing become a kind of utility like power and water, users can purchase computing power like consuming power or water. At this time, there are many computing services named cloud computing.

However, it's vague that what cloud computing is. Some people think cloud computing is a revolution in IT, at the same time, others people argue that cloud computing is nothing but a new name. At present, no one definition about cloud computing is accepted widely. In this paper, we will not discuss the exactly mean of cloud computing, but try to introduce cloud computing by some famous cloud computing instances.

The main work of this paper is introducing some kinds of public cloud computing services and open-source cloud computing software with hope to illuminate the key problem of cloud computing by those instances.

II. SOME KINDS OF CLOUD COMPUTING SERVICES.

This section will introduce some kinds of cloud computing services provided by companies including Amazon EC2, IBM smart cloud, Google App Engine and Windows Azure.

A. Amazon EC2

Amazon Elastic Computer Cloud (EC2) is one of the most famous cloud computing center [1]. The framework of EC2 is showed in figure 1. Instance is the key of whole framework. From the view of users, an instance is a computer and a lot of instances compose a cluster. Users deploy their own tasks on the clusters. While for completing those, all needs are registering on the web of EC2 and paying for what you use.

In EC2, for initialing an instance, the configure of instance is needed, such as the CPU speeds, the RAM size, the OS type and other software, all those called the Amazon Machine Image (AMI). There are existed AMI can be used or the user can custom make their own AMI. AMI is stored in Amazon Simple Storage Services (S3) when it is not running. S3 is a distributed store system used to store user's data. Different with the traditional relationship database, there is no concept of the table and all the data are stored in object. Although there is a stored model in AMI, the model will be formatted when the instance restart. Therefore, if user needs to store data, an Elastic Block Store (EBS) should be applied, then mount the EBS on the instance as a store area.

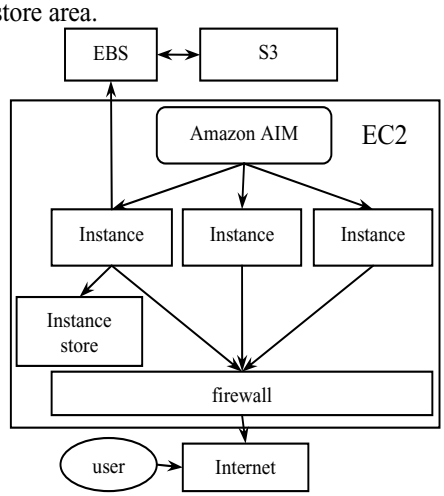

Figure 1. Amazon EC2

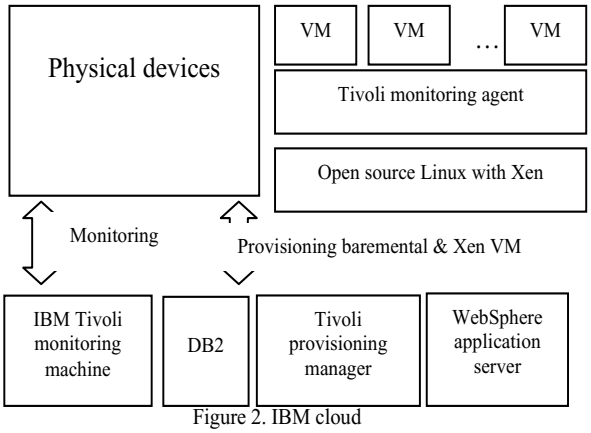

Figure 2. IBM cloud

B. IBM smart cloud

IBM introduced their own cloud computing in November 2001 named Blue Cloud. Recent a new open cloud computing platform called Smart Cloud [2] is established. Figure 2 displays the important part of IBM cloud platform. From the figure we know that Blue Cloud includes:

1) IBM Tivoli Monitoring: monitor the status of devices;

2) IBM DB2: a traditional database;

3) Tivoli Provisioning Manager: configure the devices automatically;

4) IBM WebSphere application service and some virtual components;

C. Google App Engine

The cloud service provided by Google is called Google App Engine (GAE) [3]. Similarly with EC2, user applies for computing resource by Internet. The difference with Amazon EC2 is that user doesn't obtain some instances, but a system. This system provides the services of storing files, database and bandwidth and so on. It's easy for users to transplant their Web service to GAE.

The advantages of using GAE are: relieve you from building servers and applying bandwidth from network operators, at the same time the capacity of service can be easily modified depending on the requirement. If users put web on their servers, users have to purchase new computers for adding the service capability. However, the resource will be wasted when the requirement is lower. Besides, the GAE provides some APIs implying the basic web functions, for example: URL catching, mail service, image processing, Memcache, account management and so on.

On the cloud area, cannot fail to mention is that Google disclosed their interior realization of their cloud service, which given a big boost to the researching of cloud computing. Google cloud is composed by four parts as shown in figure 3.

Google File System (GFS) [4]: a scalable distributed file system for large distributed data-intensive applications. Fault tolerance and high aggregate is two characters of GFS. GFS is usually deployed on commodity hardware and the data is stored in chunks. Each chunk is replicated on multiple chunk serves for reliability.

Bigtable [5]：a distributed storage system for structured data. A Bigtable is a sparse, distributed, persistent multi-dimensional sorted map. The map is indexed by a row key, column key, and a timestamp; each value in the map is an uninterpreted array of bytes.

MapReduce [6]: a programming model and an associated implementation for processing and generating large data sets. The user of the MapReduce library expresses the computation as two functions: Map and Reduce. The advantages of MapReduce are that it's easy to implement a distributed program and can be extended to very larger scale.

Chubby [7]: a lock service for loosely-coupled distributed systems. Chubby is intended to provide coarse-grained locking as well as reliable (though low-volume) storage for a loosely-coupled distributed system. The data consistency can be assured by Chubby.

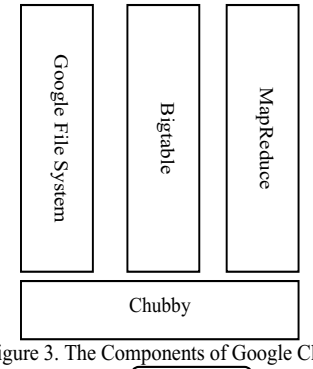

Figure 3. The Components of Google Cloud

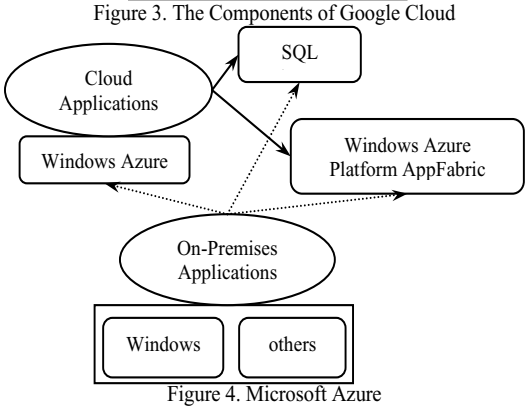

Figure 4. Microsoft Azure

D. Windows Azure

The commercial cloud computing system Azure is official launched by Microsoft in 2010 [8], from the using perspective, Azure is like GAE. Azure could be understood simply as Microsoft put their developing environment on the Internet. From the composition perspective, Azure consists of three parts as shown in figure 4 [9].

1) Windows Azure: an operating system as an online service;

2) Microsoft SQL Azure: fully relational cloud database solution;

3) Windows Azure platform AppFabric: makes it simpler to connect cloud services and on-premises applications.

The relationship of those three parts is: Windows Azure is the OS, Microsoft SQL Azure is database and Windows Azure platform AppFabric is used to connect the service in front and the program in end.

III. SEVERAL FAMOUS OPEN SOURCE CLOUD COMPUTING SOFTWARE.

In addition to the cloud computing services, there is a lot of open-source cloud computing software which can be used by users to build their own private cloud. We will introduce Hadoop, Eucalyptus, openNebula and Nimbus.

A. Hadoop

Hadoop [10, 11] is a open-source cloud computing software contributed by Yahoo and maintained by Apache Software Foundation. Hadoop is inspired by Google cloud computing platform and can be seen as an open-source version of Google cloud computing. Similarly, Hadoop consists of Hadoop Distributed File System (HDFS), MapReduce and Hbase corresponding to Bigtable.

B. Eucalyptus

Eucalyptus （Elastic Utility Computing Architecture for Linking Your Programs to Useful Systems）[12] is an open source cloud computing project established by University of California, Santa Barbara. At present, Ubuntu Enterprise Cloud [13] is powered by Eucalyptus. As shown in figure 5, Eucalyptus has four components [14]. Node Controller (NC) controls the Virtual Machine (VM) instances running on the hosts. Cluster Controller (CC) monitors the cluster including the nodes information and the networking status. Storage Controller (Walrus) provides the interfaces like Amazon S3, which is used to store images and user's data. Cloud Controller (ClC) is the terminal interface of cloud users or the administrators. The figure illuminates that Eucalyptus is very modular. The reason is that a major goal of Eucalyptus is for academic research, therefore modular design is convenient for the researchers to replace a module.

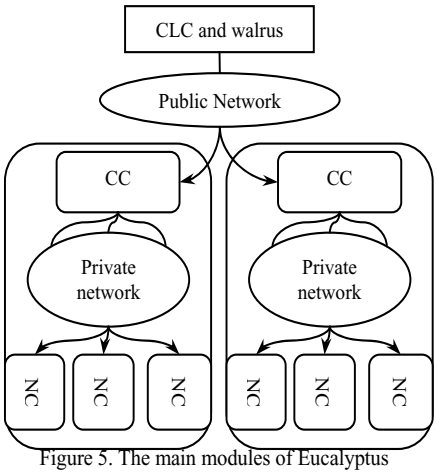

Figure 5. The main modules of Eucalyptus

C. openNebula

openNebula [15] is sponsored and maintained by Universidad Complutense de Madrid. openNebula can be used to built an infrastructure of cloud, the main aim of openNebula is to administrate the virtual data center. Figure 6 is the key components of openNebula. In figure, the Scheduler is used to schedule tasks. Command Line Interface (CLI) and client API provide the interface of programming. ONE core is a persistent back-end which is used to manage and control VM Life-cycle, Access Drivers is the of Interface different hypervisors and monitor physical resources. The greatest strength of openNebula is that openNebula does not only provide an open-source implementation of the most common cloud interfaces, but also unique, state-of-the-art features for the complete and comprehensive management of virtualized data centers and enterprise clouds.

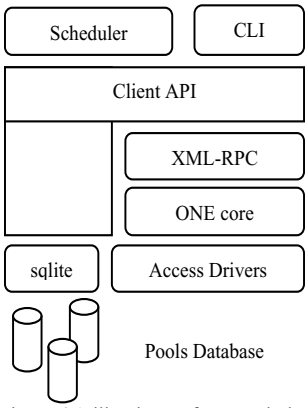

Figure 6 A illuminate of openNebula

D. Nimbus

Nimbus [16] is a open source software package for building a cloud computing system supported by University of Chicago and University of Florida, at the same time a science cloud platform [17] for science computing is built based on this software. The key components are shown in figure 7 [18]. In figure, the function of workspace control is equal to NC and workspace resource manager is equal to CC in openNebula. Workspace pilot can assist the users applied resource to access the cloud computing system. The context broker is used to explain the workflow. As a prime motive of Nimbus is science computing, the workflow scheduling is an important part of Nimbus.

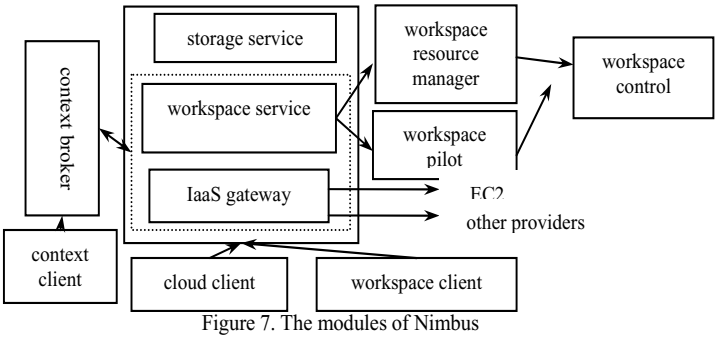

Figure 7. The modules of Nimbus

IV. DISCUSSION OF THE DEFINITION OF CLOUD COMPUTING.

Although there are many services and software called cloud computing, it doesn't exist a general accepted definition of cloud computing. It's a touchy issue to determine whether a product is cloud computing. Several influential definitions are listed as following: 1) Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services [19, 20] ; 2) A Cloud is a type of parallel and distributed system consisting of a collection of interconnected

and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements established through negotiation between the service provider and consumers [21] ; 3) Clouds are a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services) [22].

V. CONCLUSIONS

This paper introduces four public cloud computing services including Amazon EC2, IBM smart cloud, Google app engine, Windows Azure and four famous open source cloud computing software including Hadoop, Eucalyptus, openNebula and Nimbus. Users can purchase cloud computing services from the public cloud computing or built their own private cloud computing base on the open source cloud computing software. From the introduction of this paper, the reader can learn about the cloud computing and get a basic knowledge about cloud computing, which is useful for users to select cloud computing.

ACKNOWLEDGMENT

This work was supported by the Fujian Key Laboratory of the Brain-like Intelligent Systems (Xiamen University), P. R. China, the National Natural Science Foundation of China (Grant No. 60975076, 60903129, 60975084, 30900328).

REFERENCES

[1] Amazon Elastic Compute Cloud[Z]. City, 2011.
[2] K. Dowd. High performance computing[M]. Sebastopol, CA, USA: O'Reilly & Associates, Inc. , 1993.
[3] I. Foster, C. Kesselman. The Grid Blueprint for a New Computing Infrastructure (Second Edition)[M]. BeiJing: China Machine Press, 2005.
[4] S. Ghemawat, G. H, L. S.-T. The google file system[M]. Lake George, NY, United states: SOSP'03: Proceedings of the 19th ACM Symposium on Operating Systems Principles, October 19, 2003 - October 22, 2003, 2003.
[5] F. Chang, J. Dean, S. Ghemawat, W. C. Hsieh, D. A. Wallach, M. Burrows, T. Chandra, A. Fikes, R. E. Gruber. Bigtable: A distributed storage system for structured data[J]. ACM Transactions on Computer Systems, 2008, 26(Compendex).
[6] J. Dean, S. Ghemawat. MapReduce: Simplified data processing on large clusters[J]. Communications of the ACM, 2008, 51(1): 107-113.
[7] M. Burrows. The Chubby lock service for loosely-coupled distributed systems[M]. Berkeley, CA, USA: OSDI '06 Proceedings of the 7th symposium on Operating systems design and implementation, 2006.
[8] 徐光祐, 史元春, 谢伟凯. 普适计算[J]. 计算机学报, 2003, 26(9): 1042-1050.
[9] D. Chappell. Introducing windows azure[J]. Microsoft, Dec, 2009.
[10] A. J. K. van Soest, L. J. R. R. Casius. The merits of a parallel genetic algorithm in solving hard optimization problems[J]. Journal of biomechanical engineering, 2003, 125(141.
[11] A. Bialecki, M. Cafarella, D. Cutting, O. O'Malley. Hadoop: a framework for running applications on large clusters built of commodity hardware[J]. Wiki at http://lucene.apache.org/hadoop, 2005.
[12] E. Alba, J. M. Troya. A survey of parallel distributed genetic algorithms[J]. Complexity, 1999, 4(4): 31-52.
[13] E. Cantú-Paz. A summary of research on parallel genetic algorithms[R]. Urbana Champaign, IL, USA: Illinonos Genetic Algorithms Laboratory (IlliGAL), 1995.
[14] D. Nurmi, R. Wolski, C. Grzegorczyk, G. Obertelli, S. Soman, L. Youseff, D. Zagorodnov. The Eucalyptus Open-Source Cloud-Computing System[M]. IEEE Computer Society, 2009.
[15] G. Luque, E. Alba, B. Dorronsoro. Parallel genetic algorithms[J]. Parallel Metaheuristics, 2005, 105-125.
[16] M. Mejía-Olvera, E. Cantú-Paz. DGENESIS-software for the execution of distributed genetic algorithms[M]. Monterrey, México: Proceedings of the XX Conferencia Latinoamericana de Informática, 1994.
[17] 史蒂文斯. TCP/IP 详解: 协议 [M]. 北京: 北京大学出版社, 1999.
[18] K. Keahey, R. Figueiredo, J. Fortes, T. Freeman, M. Tsugawa. Science clouds: Early experiences in cloud computing for scientific applications[M]. Chicago, Illinois: Cloud computing and its applications, 2008.
[19] 刘鹏. 云计算(第二版)[M]. 北京: 电子工业出版社, 2011.
[20] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, M. Zaharia. A View of Cloud Computing[J]. Communications of the Acm, 2010, 53(4): 50-58.
[21] R. Buyya, C. S. Yeo, S. Venugopal, Ieee. Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities[J]. Hpcc 2008: 10th Ieee International Conference on High Performance Computing and Communications, Proceedings, 2008, 5-13.
[22] L. Vaquero, L. Rodero-Merino, J. Caceres, M. Lindner. A break in the clouds: towards a cloud definition[J]. ACM SIGCOMM Computer Communication Review, 2008, 39(1): 50-55