

Cloud Computing, Security and Rules

Zohre Bateni¹, Zahra Hajirezaei² and Behnaz Tamashoo³

¹ Prof assist, Engineering department, University of science and culture

² Graduate, Engineering department, University of science and culture

³ MS student, Engineering department, University of science and culture

Received: June 10 2013

Accepted: July 10 2013

ABSTRACT

Cloud computing is the basic model integrating both software and hardware sources with least amount of management problems. New services called “Cloud Services” have been published through this technology which in fact is provided by Cloud Providers for Cloud Users. In order to use it properly, before publishing such technology, the pros and cons of it must be analysed in details.

In this article first the definition of cloud computing and its advantages are mentioned following the method description of some service providers, and at the end the possible flaws and defects and the solutions to such problems are discussed.

KEYWORDS: Cloud computing, GfS, MapReduce, Cloud computing rules

1. INTRODUCTION

First computers were created in 1937[1]. Those computers were only capable of processing a restricted amount of commands on very limited data. By the advancement of technology in recent decades there was a dramatic increase in the ability of computers in terms of variety of actions, memory disk space, network and the power of computing and etc.

Network technology advancement and the ability to share massive amounts of data, gave users this chance that they don't feel the need for more free space on their electronic devices, on the other hand, they wouldn't be obliged to pay for all the software which rarely used, since in emergency times they can use the software online by paying much less price. Many practical softwares are designed and exercised for Cloud Computing Users, as their number grows, and these programs themselves emerge new needs and to come over those needs the designers and programmers of Cloud Computing are trying to empower the basic software and hardware of this technology. Powerful systems are required possessing enough space for all important software to install.

Two options are practical in order to reach such achievement. First is purchasing a very powerful and high tech system which will be too expensive in the beginning, although creating the software for such system would be much simpler, and second is to link a number of systems together which is probably less expensive, but creating software for such method would be dramatically more complicated and awkward. In addition when a number of systems are connected to each other the electrical energy must be provided, and on the other hand as too many systems beside each other would increase the heat, the room must be facilitated with air conditioning system[2].

Most ISPs nowadays have come to this technology and use powerful equipment and software to manage these systems. They provide people with internet and in turn they charge each person according to their monthly usage. Many large companies such as Google, Amazon, Yahoo, or some universities adopt Cloud Computing to use for research matters or providing internal or external users.

In addition to economical benefits, Cloud Computing has other advantages such as:

Cloud Computing is based on demands. In this system users can benefit from the services without communicating the service providers and this is done through the virtual spaces of websites or etc, and each user can do a different job individually.

They can use different standard mechanisms at hand, such as mobile phone, laptop or etc to connect to the system and meet their needs.

One feature of Cloud Computing is integrating two sources. Cloud Computing service providers use multi-tenant models to integrate the physical or virtual sources, and make them wiser according to the users demands.

Another feature is measurability of the services. Cloud based systems are able to measure services and also control and develop sources (saving, processing, network range, active user accounts). [3,4,5]

Another advantage of this service is that even small or newly-constructed companies can benefit from very powerful software and equipment without paying massive amounts of expenses.[6]

2. Cloud Computing in Google:

Google was a research project of two PHD students in 1996, and in 1997 the domain of www.google.com was officially identified[7]. It has provided services with billion files, such as Google maps, Google translator, and other popular and practical software like android system. Nowadays, it offers most of its services in Cloud realm. It uses very powerful and high-tech systems and adopts special technics to avoid processing problems of such massive amount of data.

For instance Google does not pay much for hardware and uses simpler ones, and pays much more for software instead and balances the lack of hardware with the use of powerful and complicated software. As Google does not purchase the software and creates them on its own, it can design and create them according to its own needs and boost the function of hardware noticeably.

The policy used to save document, data, and information in a computer or server is called File System. Google uses a particular type of file system called GFS to manage the files[8].

GFS is designed for optimization, saving, and recovering files and data and rest of Google tasks. Google is usually in need of massive file transferring, so in order to increase proficiency and efficiency greater units for transferring are selected. These units are called chunks. Each chunk has the capacity of about 64mb (In Ponix system the capacity of file transferring units which are called blocks is about 8kb). Google indicates that: "Selecting the massive capacity of chunks was the key element of increasing the efficiency"[9].

In designing GFS Google tried to optimize the efficiency of import-export equipment. It has also increased the speed of file reading to a level which is comparable to file reading from a disk. On the contrary the speed of writing is noticeably low which is of less importance, since unlike reading the action of writing needs to be done only once. Meanwhile GFS also supports the act of attachment (in a low speed)[10].

One of the Google's main defects is the increase of resistance of GFS error, which it overcame with the triple repetition of each chunk on different computers. The matter is redundancy which on the other hand increases the resistance of the error. It claims that Hosts or Disk drives merely cannot be trusted, although they can when they are combined [11]. Google uses Map-reduce to save and recover data and information[12]. Software program developer types Reduce, Map to do the functions[13]. This method was invented in Google to process massive amounts of rudimentary and raw data. Data must be received in thousands of machines and categorized in a reasonable time. Map-reduce allows Google engineers to present their procedures absent such worries like the publicity of data, and error resistance or etc[14].

The map function which is of Google library subordinates receives the entry data and presents a mild set of <keys , values>. Afterwards Map-reduce sends all entry data that share similar values to Reduce subordinate. In the next stage Reduce corresponds a set of data with similar values and this integration makes the entry data much smaller[15].

As it is seen the use of Map-reduce paved the way noticeably for Google to provide Cloud based services. Even though Map-reduce is only useful for data that are divisible, for Google it is much effective as Google softwares have such characteristics. This separation among the Google softwares has increased the error resistance. Since each part of the program is run separately in case an error occurs only that part of program which is running will be repeated and it does not affect the other parts [14].

In a study done in 2004 Google did two experiments on 1800 Cloud machines, and the results were published.

Each machine had a 2GHz Intel CPU, 4gb Ram, and two 160gb hard disks that were connected by IDE graphic link and Ethernet. In first experiment Google created and ran a program called Grep. Grep was designed for a 3 character search pattern among 10^{10} record of 100 bytes (1 terabyte data). The wanted pattern existed in record 92.337. Google claims that through Map-reduce it can find every item in 150 seconds.

The next experiment was the comparison between Map-reduce and Terasort in arrangement of one Terabyte data. Map-reduce did the arrangement in 891 seconds and Terasort did it in 1057 seconds.

The brevity of the software is also as important as its simplicity. Even though Google possesses numerous extensive servers, due to massive amount of various users of different services and also a variety of softwares, it is dealing with lack of space for saving data.

The other article issued by Google about Map-reduce revealed how much data is processed in their production system. Google claims that in September 2007 the cloud of Map-reduce had processed more than 400 petabytes (each petabyte equals 100 terabytes) in 2217000 turns. These amounts are incredibly huge and show that, Map-reduce has been really useful for Google calculations. [12]

Big Table which was invented based on GFS in 2006 as a database, was the third equipment Google adopted to achieve Cloud Computing. It could be said that Big Table is a management data base system. There are tables in this data in which lines and columns all possess names. The name of each column or row is considered a key. Just like any other data base the act of addition, omission, updating, and renewing are possible here. Since massive amount of data is going to be processed in Google data base it is not easy to manage them. In order to manage them more easily one can put some columns in a column family and devote them to a cloud machine to manage. Even though it is pretty difficult to create a column family in this system, since it is very easy to add or omit one column to this family, the use of Big Table is highly recommended. Another advantage of Big Table is the concept of Time Stamp. Through the presence of Time Stamp each slot in the table can save different versions of a data. These versions are numbered according to a 64bytes sentence of true numbers. By holding the Time Stamp constant and not changing it only one version will be accessible. Google claims that petabytes of data in thousand servers can be saved by Big Table[16].

3. Yahoo Cloud Computing

Another Cloud Computing service provider is Yahoo. Just unlike Google that did each and every step of experiments and service providing programs in itself, Yahoo adopted a prepared method called Hadoop. Hadoop is an open text project to exercise something like Map-reduce, which contains distributed data like GFS, which is named HDFS. Unlike Google which is written by C++, Hadoop is written by JAVA. Eventually Hadoop developed and Yahoo was one of the greatest users and supporters of it.

Other Cloud Computing Service Providers are Amazon and Microsoft.

4. Some important problems of Cloud Computing:

- a) Safety: Cloud Computing can cause noticeable dangers in case of holding privacy. By using a Cloud Computing system important data are saved on Cloud servers, this will restrict the control of the data. If the Cloud Service Provider does not benefit from enough privacy or security systems, the safety of the data might be at great and fatal risk. Meanwhile it is not wise to sign a contract with a Cloud Computing Service Provider without knowing about its security system. It is necessary for both service provider and user to be aware of all security conditions of each other[17].
- b) Reliability: It is also important to cloud users whether the service is reliable or not. Due to publicity of Cloud Computing service, and also the long-time use of it by some users the reliability of the server becomes a vital matter as well, which will be achieved through error resistance technique and building hardwares.
- c) The possibility of abuse: when important personal data is stored in one data base a number of abusers will definitely be tempted to reach and publish them.
- d) The need of permanent access to a high speed internet: In case no internet would be available or a very poor service like dial-up would be used, the data in Cloud Service cannot be reached since documents or data saved on the web require a high range to be downloaded. In European or American countries such problem does not really count but in Iran providing internet system is not that easy that such problem can be ignored.
- e) Possible dangers in data protection: In Cloud Computing there are possible dangers in data protection. The most important possible dangers in protecting or processing the data can be divided into two categories:
 - e.1.) Lack of control on data and not enough information.

By placing information in a Cloud Computing Server individuals are not merely capable of controlling their own data, and cannot use techniques to ensure whether their data is inaccessible to public, safe and secure, clear, modifiable or etc. This can cause different matters such as:

 - e.1.1) Data Transfer between service providers can be problematic
 - e.1.2) A Cloud is created based on shared information and such characteristics. Service providers gather information from different sources and data lack similar roots.
 - e.1.3) Lack of privacy since a unified law is not legislated to direct Cloud Computing service providers in every country, so users' information might be abused even by some public or non-public organisations.
 - e.2) Lack of information and clarity about the method of processing: Having not enough information and clarity about the way of processing Cloud services can lead to potential dangers for controllers and data and since they are not aware of the dangers they do not know how to avoid them. One of the threats that controllers might not be aware of could be that personal data from different geographical parts of the world are processed, and protecting data follows different approaches and curfews. Meanwhile third world countries might suffer insufficient rules of

data protection or safe data transference, thus, personal data needs particular naming and this matter shall be identified to the controllers and processors.

Due to complexity of chain processing in the realm of Cloud Computing and in order to guarantee fair processing and enough protection for the data, specific rules need to be legislated.

5. Safety

Safety is one of the vital matters in real world and has always been an indispensable factor.

In virtual world too safety is counted as one of the vital and discursive matters. It is necessary to ensure proper safety of data, different channels and other things through different methods and means. For example it can be partly achieved through coding data or setting password, username, or other methods.

By publicity and increasing the number of Cloud service providers the importance of data security doubles. Unless users do not make sure of the method and place of storing their data, they will not feel safe to share their important documents or personal information. So it is necessary to ensure them of the permanence and security of their data, and if they realize that their data will be accessible whenever and wherever they desire, they will welcome such technology.

6. Security benefits of Cloud Services

Cloud Services have become highly popular due to some extensive security reasons.

Service providers use new methods which support backup and are not available at personal computers or organizational servers.

In softwares available on Cloud Services, security is always updated which in personal computers depend on the user and mostly does not happen in long intervals.

All users benefit equal security not concerning the amount and type of service the use.

Since data are distributed separately throughout the Cloud server they are not readable by human.

In case any problem occurs, the recovery and restoring data will be much cheaper.

Cloud providers use realtime system to determine any influx and security check for each demand[18].

7. Potential problems in Cloud Services security

There is too much information about different people in Cloud Computing service providers, so these centers are threatened by different institutions. The first threat is the service provider itself that in specific times may abuse the information or under some possible pressures might share the information with courtiers. The other threat is hackers or other invaders of virtual world. The existence of a site abundant with such information might attract them and another possible threat might be data loss due to system breakdown.

In 2008 a bit of corrupted data in the servers used by S3 (Simple Storage Service) Amazon, caused system to stop working for some hours. As nowadays users are given online storages in gigabytes the possibility of such matters is doubled [19].

In another case in the early 2009 a hacker managed to reach all personal google-apps documents of an individual by reaching out his/her email on twitter and send them to the media[20]. Although the staff of twitter insisted that this matter was hacking the google, it made no difference in the results as the information had been revealed.

In October 2009 and following a breakdown in one of the servers of Danger in Microsoft Corp which provides storing spaces, one million information of T-Mobile Sidekick was lost, though much of the information was recovered later[21].

And here is the answer of Cloud Computing service providers to the problems mentioned above: <<Now Clouds are safer than anything else you use>>. eran feigenbaum the manager at Google Security Centre of practical software says: "Cloud Computing service providers can assure your safety in a much more practical way than millions of users of data centres in independent corporations"[22]. Think of the threats in former approaches. About two third of the people who participated in survey forgot their USBs in different places, one laptop is stolen in every 53 seconds in the US! We can never increase safety to 100% but at least we can try to be like this." [23] In a conference of Computer Security in April 2009, john chambers, the junior manager of Cisco regarded Cloud Computing as Security Nightmare"" and remarked that it cannot be dealt with by common methods"[24]. In the same conference Ronald Rivest a computer science technician in MIT who invented Regular Security Algorithm, RSM said that it was better to replace Cloud Computing with Swamp Computing. He indicated after that he meant to encourage the users to analyse the apparently simple problems of this technology[25].

NIST organization aims to provide inter-Cloud services which will bring ease and efficiency to the users. It can clearly be observed that Cloud Computing Technology have flaws as well and even great companies have had

massive defeats in their job. One of the top computer science technicians Larry Patterson who is in charge of an internet exam platform in the university of Parinstone says: “Even if we say Clouds are secure against external invasions, how can we make sure they are not abused by the providers themselves, so they have to prove worthy over long periods of time. It is true that the provider possesses high security mechanisms but can it be trusted against different people and make sure that this service will last more than 5 or 10 years. There are some security matters that must be considered. Mere benefiting from the technology is not enough as it might be difficult and follow a lot of security problems[26].

8. Legal restrictions of Cloud Computing

In advanced countries in order to protect the data, particular legal restrictions are made for processing personal data and accessibility through network systems such as Telecomoperator, that operators must obey them.

There are also rules that include controllers whether inside or outside of EEA. One of these regulations is the position of the company and its range of activities[27]. Cloud Computing also includes different roles and it must be mentioned who are going to take them and what their duties are. In 2010 in a negotiation the actors were divided to two categories of controllers and processors and the main responsibility of controllers was known to choose a person to compile the curfews of data protection in the language of computer. These people included Cloud customers, Cloud service providers, and sub-contractors.

The need to protect data for the customers and Cloud Service Providers:

- 1) Compiling the basic laws
- 2) The clarity of the function
- 3) Identifying goal and restrictions
- 4) The possibility of recovering or omitting the data when they are not needed
- 5) Protecting curfews between controller and processor

9. Conclusion and offers

Technology is advancing every day and if we do not pace ourselves with the technology we'll be isolated in the modern world. Each technology has its own pros and cons and tries to compensate the preceding ones. By analysing the advantages and disadvantages it will be regarded as whether harmful or advantageous to advance and to be continued in the future.

In terms of Cloud Computing as a new technology many advantages can be remarked such as: reliability, accessibility and cheapness of data, and standardizing the processes; and disadvantages like: the need to permanent internet access which in Iran is a bit problematic, security problems, lack of static laws and regulations, lack of privacy, lack of trust, etc. If such problems are solved this technology will be much more popular and too many good services will be available, although, even now there are too many services such as email and networking web sites with extensive data bases.

Thus it is suggested that Cloud service providers try to centralize the aim of legislating international curfews and rules in order to gain public trust.

REFERENCES

-
1. <http://www.computerhope.com/issues/ch000984.htm>
 2. http://www.nsa.gov/research/_files/publications/cloud_computing_overview
 3. <http://www.lians.ca/documents/0001636>
 4. <http://www.casselsbrock.com/CBArticle/Data Security The Case Against Cloud Computing>
 5. Data Security The Case Against Cloud Computing, *Bernice Karn Partner, Cassels Brock & Blackwell LLP*, march 2011
 6. <http://www.smallbusinesscomputing.com/biztools/articlephp/3939301/Cloud Computing Tips for Small Business.html>

-
7. Barroso LA, Dean J, Hölzle U. Web search for a planet: the Google cluster architecture. *IEEE Micro*.2003 March-April; 23(2).
 8. Ghemawat S, Gobioff H, Leung S. The Google file system. *Proceedings ACM Symposium on Operating Systems Principles*. 2003 October.
 9. http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/en//archive/gfs-sosp2003
 10. [http://en.wikipedia.org/wiki/Google File System# CITEREF Ghemawat GobioffLeung 2003](http://en.wikipedia.org/wiki/Google_File_System#CITEREF_Ghemawat_GbioffLeung_2003)
 11. [http://en.wikipedia.org/wiki/Google File System# CITEREF Ghemawat GobioffLeung 2003](http://en.wikipedia.org/wiki/Google_File_System#CITEREF_Ghemawat_GbioffLeung_2003)
 12. Dean J and Ghemaway S. MapReduce: Simplified data processing on large clusters. *Proceedings Operating Systems Design and Implementation*.2004 December.
 13. Dean J and Ghemaway S. MapReduce: Simplified data processing on large clusters. *Communications of the ACM*.2008 January; 51(1).
 14. [http://code.google.com/edu/parallel/mapreduce tutorial.html](http://code.google.com/edu/parallel/mapreduce_tutorial.html)
 15. Dean, Jeff andghemawat, Sanjay. MapReduce: Simplified Data Processing on Large Clusters http://static.googleusercontent.com/externalcontent/untrusted_dlcp/research.google.com/en/us/archive/mapreduce-osdi04
 16. <http://lab.google.com/papers/bigtable.html>
 17. <http://www.informationweek.com/cloud-computing/the-clouds-five-biggest-weaknesses/229202319>
 18. Barry Reingold & Ryan Mrazik, “Cloud Computing: The Intersection of Massive Scalability, Data Security and Privacy (Part 1)” (2009) 14:5 *Cyberspace Law* 1 at 2-3, online: Perkins Coie http://www.perkinscoie.com/files/upload/PS_09-06_Cloud_Computing_Article.pdf
 19. Amazon S3 Availability Event: July 20, 2008
 20. http://www.computerworld.com/s/article/9135591/Hacker_break_in_of_Twitter_e_mail_yields_secret_docs
 21. <http://www.engadget.com/2009/10/11/sidekick-failure-rumors-point-fingers-at-outsourcing-lack-of-ba/>
 22. <http://www.bcs.org/content/conWebDoc/48493>
 23. <http://blog.eio.com/2012/05/01/how-to-prevent-laptop-theft/>
 24. http://www.computerworld.com/s/article/9131998/Cloud_computing_a_security_nightmare_says_Cisco_CEO
 25. <http://cloudcomputing.syscon.com/node/1231725>
 26. <https://www.technologyreview.in/computing/24284/page2/>
 27. http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp179_en