Cloud computing: utility computing over the Internet

Taneli Korri
Helsinki University of Technology
tkorri@hut.fi

Abstract

Cloud computing has become a hot topic in the IT industry, as it allows people to buy computing resources in the same way as utility services like water and power. Cloud computing services are delivered over the Internet which makes them a viable alternative to building a own computing infrastructure. This paper examines what cloud computing is and what kind of advantages it brings. In addition, this paper looks at the disadvantages and compares cloud computing to it's near relative, grid computing.

KEYWORDS: utility computing, cloud computing, grid computing

1 Introduction

From the early days of the World Wide Web (WWW) to the present day, the number of people surfing on the Web has grown from a handful to hundreds of millions. The main reasons for this massive growth are the online services and the ease of connecting to the web. In the early 1990s, Internet connections were hard to come by. Home connections were almost non-existent and the lucky few who had a connection at home usually routed it through their universities. Today, the Internet is everywhere. Companies, universities, libraries and even coffee shops have Internet connections, which the employees, visitors and customers can use to surf the Web. Getting a connection to home is equally easy, Internet Service Providers (ISPs) sell connection services at affordable prices and the installation times vary from same day delivery to a few weeks.

Due to the ease of connecting, the Internet has expanded in every possible way. The user base has grown from a handful to over a billion, which has increased the business opportunities on it. This has in turn led to the situation where services offered on the Internet have grown rapidly, both in quality and quantity. As an example, the popular Internet search engines serve millions of search queries every day and popular Internet shops have tens of thousands of visitors browsing through their web sites daily. Serving these kinds of volumes requires serious investments in hardware, hardware which can cost a significant amount of money. But excess money is something people usually do not have. The current trend is to invest in the hardware infrastructure as the service becomes more popular. Or in some cases, investment occurs only after there's clear evidence that the current infrastructure cannot handle all of the users. But this is far from the optimal solution as infrastructure built this way cannot cope with a sudden boost of visitors. And in the opposite situation, during slow business days, much of the infrastructure will be unutilized [7]. To solve this never ending race for hardware investments, many are now looking at cloud computing.

Cloud computing doesn't have any precise definition [12], but when people talk about it they're talking about computation and storage service which can scale up and down as needed. And all this can be accessed over the Internet with the help of a credit card. Thus cloud computing can be broadly defined as hardware infrastructure as a service. Although this definition still generalizes a lot, as cloud computing can also be seen as platform and as a software service [10, 11]. As people pay for utility services like water and electricity, there shouldn't be any reason why they couldn't also be paying for computational power or storage which is available over the Internet [17].

Cloud computing is at the moment a hot topic in the IT industry. It has received a lot of time in the press in the recent years and since Amazon started its Amazon Web Services (AWS) [5] and Google its Google App Engine (GAE) [9] there's no sight of the interest in cloud computing decreasing. Cloud computing has even been mentioned as the key to Web 3.0 [7]. But during all this media interest, little or no time has been used to describe what cloud computing really is and where did it come from. Or how it will resolve the current problems in infrastructure. Equally, less time has been used to describe the advantages and the disadvantages of using it when building a service infrastructure.

Cloud computing has emerged from the same basic ideas and goals [8] as grid computing, which is mainly used in the academic world. However, cloud computing is more focused on solving the problems companies have faced when they have been working on their large computational problems and infrastructure. Although, the focus of cloud computing is a little bit different, it's still no wonder that both clouds and grids struggle with the same kind of problems [10].

But ideas and problems are where the similarity ends. The technical implementation and the user interface are usually somewhat different in clouds than in grids [10]. Clouds are usually built as clusters, which can individually scale up the calculation power and data storage as needed. These cluster are usually owned by one single company. Grids on the other hand are often built around clusters and connected to other similar grids which are not necessary owned by the same organization. On the technical side, clouds are usually built from cheap off-the-shelf hardware which anybody can buy. The computational power comes from the sheer quantity of these cheap servers and because they have been connected together efficiently in order that the computational problems

can be spread to many of these machines and solved in parallel. This has the advantage that the initial investment doesn't necessarily need to be big and the cluster can be expanded when more power is needed. Grids, by contrast, are traditionally built on big cluster computers which are powerful but expensive. Also, big cluster computers are seldom as upgradeable as a cluster of commodity computers.

This paper is organized as follows. Section 2 discusses the key elements of cloud computing. Section 3 looks into the advantages and Section 4 into to the disadvantages of cloud computing. Section 5 compares clouds to grids and discusses the main differences. Section 6 introduces the current biggest cloud computing platforms and looks at what they offer and how they fit in to the definition of cloud computing. Section 7 contains the conclusions where the papers findings are summed up.

2 Key elements of cloud computing

Cloud computing doesn't have a standardized definition. It seems that everyone closely related to cloud computing has a slightly different definition. Which has led to the fact that there are over 20 different definitions in all [12]. However, the definitions have much in common. There are many aspects that are repeated in numerous definitions like distributed computing, scalability, virtualization and service levels. So it seems that, although all of the definitions differ from each other, there is some kind of agreement of what cloud computing is all about.

2.1 Utility computing

When talking about cloud computing, distributed computing is understood as utility computing over the Internet. Which means computing services have moved from local data centers to hosted services which are offered over the Internet. From the customer perspective this means that he or she doesn't need to worry about the computing infrastructure anymore. Since it's bought as a service, the maintenance is the service provider's responsibility. Although the customer should be prepared for situations where the cloud isn't available, since nobody can guarantee a service level with 100% uptime.

2.2 Scalability

Scalability means that there's a way to add and remove resources to the services depending on the usage. In cloud computing this means offered services can be automatically scaled up and down depending on the customer's needs. This is actually one of the key elements of cloud computing. In traditional computing infrastructures computing power equals the amount of computers available in the data center, in cloud computing the computing power is almost unlimited. The customer can scale up the computing power as long as the service provider has free capacity and the customer is able to pay for it.

2.3 Virtualization

Cloud computing relies heavily on virtualization. The services are built on top of a virtualization layers which help the service providers to manage the service and offer standardized platform to the users. Virtualization is in fact another key element of cloud computing, it enables the service provider to offer the homogeneous service simultaneously to all customers, something that cannot be achieved, for example, in grid computing.

2.4 Service levels

Service levels have two distinct meanings in cloud computing. First meaning is the service level agreement (SLA) and the second one is the specified level the service works on the customers system.

Service level agreement SLA is the service provider's way of assuring the customer that the system works at least on the defined level. SLA sets the minimum computing resources the user has access to. For example, the SLA might state that the service provider offers computing resources which are capable of solving at least 25 billion floating point operations per second. Or that the offered hosting platform will be online for 99,9999% of the time.

Service level Today, cloud computing platforms have been used to offer services on three different levels: infrastructure, platform and software.

The lowest level available is infrastructure as a service (IaaS) [10, 8] or as sometimes called Hardware as a service (HaaS) [7]. At this level cloud computing works as the hardware behind the service. I.e. cloud computing service provides only the hardware which runs the customer's service, so the customer has to handle all the needed software installations and managing of resources.

The second level is platform as a service (PaaS) [10, 7]. This level offers the whole hardware and software platform to the customer, enabling the customer to focus fully on developing services. PaaS simplifies the development procedure as the customer doesn't need to worry about any hardware or software installations.

The last level available is software as a service (SaaS). In this level the customer is actually the end user. The customer buys access or a license to a software product, which is running on cloud computing technology. Usually the customer doesn't even need to install anything on his local machine as the software is operated with a web browser and available anytime from anywhere. There are some levels which are incorporated into the SaaS level, like database as a service (DaaS) [7]. DaaS offers specialized database software running on top of the cloud computing platform. One example is Amazon SimpleDB [18].

3 Advantages of cloud computing

There are some advantages in using cloud computing in the computing infrastructure. First of all the biggest advantage is initial cost savings. When companies or individuals launch

new products and services, they do not need to have a big initial investment in computer architecture. Cloud computing platform offers the hardware and makes sure the platforms hardware functions as supposed. Cost savings has actually two sides, the investments in hardware and the upkeep costs. Both of these are somewhat eliminated when building a service on cloud computing platform. Hardware is rented from the cloud computing service provider and the price is based on the usage [12, 10].

Second and probably the most wanted advantage is the scalability. Cloud computing platforms have the ability to scale computing power up automatically according to usage [8, 7], something that is impossible to achieve without a big investment in infrastructure. And investing in infrastructure will not guarantee the same benefits. When the usage increases the platform can be instructed to add more computing power to the service. In principle, there's no limit to the scaling, but the cloud computing service provider has some limits installed in order to keep the whole cloud stable and running. Since the service provider is offering the service the responsibility shifts from the user to the provider which in turn makes the user's life more easy.

In addition to scaling up, cloud computing platforms can of course scale the service down [8, 7]. Which means that there doesn't always need to be a massive infrastructure behind the service running it, only when the usage volume requires it. When compared to traditional infrastructure, a scaling cloud computing platform can utilize the given hardware more efficiently [7]. When there's a lot of traffic and users, scale the service up so that it won't choke. And when there's little or no volume, scale the service down to a bare minimum. Scaling enables the efficient usage of resources which leads to fully utilized hardware always. Traditional computing infrastructure is usually built according to the peak load, which means that when there's very little or no traffic, most of the hardware is not being utilized at all.

4 Disadvantages of cloud computing

Although cloud computing has many advantages which are hard or impossible to achieve with traditional infrastructure, it also has some disadvantages.

The pricing model differs from provider to provider and some of them are quite complex. For example, with Amazon Web Services the total price consists of used hours, used storage and the amount of data transferred [2]. The whole sum may be hard to predict when calculating the total costs.

The biggest disadvantage is the lack of open standards between cloud computing providers. Everybody has their own proprietary application programming interfaces (API) [17] which leads to vendor lock in. This is something everybody has to think about before starting to work on a cloud computing platform.

Other considerable things are the reliability of the company. Is it going to be around in five years? [10] How much will it cost to change the technology if the company discontinues the service? And what if the usage costs become too expensive? These are all questions the customer should think about before choosing cloud computing and a cloud computing provider. It might just be that cloud computing, although with many advantages, is a passing paradigm which will be superseded with something better. Or that there will be a lot of changes in the field, some providers may disappear and other may take their place. As there's no common API one has to be careful which provider they choose. It just might be that choosing a wrong provider causes more expenses than savings.

One disadvantage which affects mainly the government sector is the reliability and the security of the provider. There usually are privacy laws which the governments and companies have to obey. This means that they have to know exactly where the data is being stored and who can access it [17, 10]. In cloud computing all of the programs are running on the same hardware, although in different virtualized environments. So in theory it's possible that someone might be able to access the customer's private data through a vulnerability, since the applications are running on the same physical hardware.

5 Clouds compared to grids

Cloud computing and grid computing have evolved from the same principles and cloud computing can be described as subset of grid computing [7]. However, cloud computing and grid computing have a few key differences in the implementation and in the usage scenario.

First of all cloud computing and grid computing have different kind of usage scenarios. Grid computing is very much in use in the academic world and cloud computing is used much more in the corporate sector.

One reason for this difference is that grids are usually build from smaller grids which are operated by different organizations, usually by universities and institutions. By contrast, whole cloud computing environments are usually owned only by one company.

The second big difference is the cost of usage. Academic institutions usually grant computing time free of charge to their researchers [18]. And if these researchers use lot of computing time also on other grids, the debt is paid back by providing computing resources to these other grids some later time.

On the technical side both clouds and grids are somewhat alike. The hardware can be the same but there are differences in the services built on top of the hardware. Cloud computing platforms hide the underlying hardware by using virtualization technologies to be able to offer a standardized hardware platform to the customers [12].

Although there are lot of grid standards, the application which the user wishes to run on a grid has to be specially programmed to suit the chosen grid environment [8]. One key technical difference between grids and clouds is the fact that in cloud computing platforms allocating processes to resources is done automatically, if it has to be done by hand then the platform cannot be considered as cloud computing platform [7].

Computing in grids and clouds has different concepts as cloud computing platform is one massive parallel computing system and grid system is a group of smaller grids. Applications are executed in parallel in clouds and sequentially in grids. Which means that if a user is trying to run a complex

computation, i.e. something requires hundred nodes, the user may have to wait in queue for some time in order for the hundred nodes to be free [10]. In cloud computing platforms the applications are run in parallel, which means the user doesn't have to wait for nodes to be freed. This implies that scalability is easier in clouds than in grids. Since the user doesn't have to wait for resources to be freed in order to scale, the user has the option to do his work in less time by using more nodes. Surprisingly it won't usually cost significantly more compared to running it on a less amount of nodes but for a longer time [18].

6 Available cloud computing platforms

Currently there are a few cloud computing providers which can be used by anybody. Probably the most know provider is Amazon with its Amazon Web Services (AWS) [5]. Google has taken a slightly different approach with its Google App Engine [9] by providing only a software platform. Microsoft has tried to catch up to Amazon and Google by introducing Live Mesh and a preview of Azure [14]. Table 1 contains information about the current cloud computing operators.

6.1 Amazon Web Services

Amazon is one the first companies which started to offer cloud computing services. Amazon offers a wide ranges of products, most of them being in the infrastructure level. The best known products are Amazon Elastic Compute Cloud (Amazon EC2) [2], Amazon Simple Storage Service (Amazon S3) [4] and Amazon Simple Queue Service (Amazon SQS) [3].

Amazon EC2 is Amazon's cloud computing platform. The service is based on Amazon Machine Images (AMI) which the user can create and share [17, 18]. Machine images contain the applications and libraries the user needs. The user can start instances of these images and the user is charged for the time the instances are running. One another thing affecting the bill is the amount of computing power. The user has to choose the virtualized hardware which runs the image. The hardware setups are predefined and they range from single core system with 1,7GB RAM and 160GB storage space to 8 core system with 7GB RAM and 1690GB storage space [18].

S3 is Amazon's reliable data storage with high availability. In S3 the user stores his or her files, or objects as they are called, in buckets. The system sets a limit of 100 buckets per user but the amount of objects is unlimited. Only the size of a single object is limited to 5GB. The user is charged for the amount of data in the storage and traffic from and to the S3 [18].

SQS is a simple reliable and scalable queue server, which can be used through web services interface. SQS can be used to implement reliable queue system between EC2 instances [18] and other distributed programs [3]. SQS's strong points are that it doesn't require any software installations and that it is accessible from anywhere.

6.2 Google App Engine

Google App Engine (GAE) [9] is Google's answer to Amazon Web Services. GAE is a platform service which runs applications written in the Python language. On top of the Python language standard libraries, GAE offers application programming interfaces to Google products like Google Accounts and email services [17]. Billing works a little different in GAE than in AWS. Google offers every user some resources for free and the user only has to pay if he exceeds these free resources. This means that testing GAE is easy and free, the only thing the user needs is a Google account [9].

6.3 Microsoft Live Mesh and Windows Azure

Microsoft Live Mesh [13] is Microsoft's first cloud computing product. Live Mesh is a service which enables the user to store files and applications to Microsoft's servers and access them anywhere anytime [17, 13]. In addition, the user can link computers to Live Mesh which enables the user to use these computers through Live Mesh like the user was sitting in front of it [13]. This of course requires that the target computer is turned on and connected to the Internet. Live Mesh files, applications and computers are access through a web interface or with a Live Desktop application.

Microsoft has another interesting cloud computing product, the Windows Azure [14]. However, Windows Azure hasn't been published yet and there's only some promotional material available. Windows Azure website depicts Azure as *"Windows Azure provides developers with on-demand compute and storage to host and manage web applications on the internet through Microsoft data centers."* [14] which makes it something to watch for in the future.

6.4 Others

Amazon and Google are the big players in cloud computing but there are also few other providers. 3tera and ElasticHosts [1, 6] are two which are also mentioned. Besides these companies, there are also Nimbus [15] and OpenNebula [16]. These are software projects which offer software for building a cloud computing system.

7 Conclusions

Cloud computing is new and hot topic in the IT industry. As cloud computing is a fairly new paradigm there isn't a precise definition yet for it. But many agree, that in order to a system to be a cloud, it has to support scaling and automatic issuing of resources. Cloud computing has evolved from the same principles as grid computing, but there are some key differences which distinguish them from each other. One of the biggest difference is that in grid computing the applications are run sequentially and in cloud computing they're run in parallel.

Cloud computing offers a lot of possibilities for those who need a lot of computing power and storage. It's something to look into especially for people who cannot invest in own hardware. It's also something people who are building web

	Amazon Web Services	Google App Engine	Microsoft Live Mesh	3tera	ElasticHosts
Service level	Infrastructure	Platform	Infrastructure	Infrastructure	Infrastructure
Pricing	By usage	Free, By usage	Free	Unknown	Subscription, By usage
Data centers	USA, Europe	Unknown	Unknown	USA	UK

Table 1: Current cloud computing operators

services should look into, as cloud computing offers scalability in both computing power and storage. But cloud computing is not a silver bullet, which will magically solve all problems. Lack of standards in the field of cloud computing has led to the current situation where every cloud computing provider has their own API and their own way of doing this. This has led to the fact that cloud computing provider has to be chosen in very early stage of a software project. If the choice is not made, there might be a significant amount of work to be done in order to have the software work on a different cloud computing platform.

References

[1] 3tera. Cloud computing for web applications. Accessed 16.3.2009 http://www.3tera.com/.

[2] Amazon. Amazon elastic compute cloud (amazon ec2). Accessed 16.3.2009 http://aws.amazon.com/ec2/.

[3] Amazon. Amazon simple queue service (amazon sqs). Accessed 10.4.2009 http://aws.amazon.com/sqs/.

[4] Amazon. Amazon simple storage service (amazon s3). Accessed 10.4.2009 http://aws.amazon.com/s3/.

[5] Amazon. Amazon web services. Accessed 18.3.2009 http://aws.amazon.com/.

[6] ElasticHosts. Elastichosts - flexible servers in the cloud. Accessed 18.3.2009 http://www.elastichosts.com/.

[7] Francesco Maria Aymerich, Gianni Fenu, Simone Surcis. An Approach to a Cloud Computing Network. In *First International Conference on the Applications of Digital Information and Web Technologies*, pages 113–118, August 2008.

[8] Giacomo V. Mc Evoy, Bruno Schulze. Using clouds to address grid limitations. In *MGC '08: Proceedings of the 6th international workshop on Middleware for grid computing*, pages 1–6, New York, NY, USA, 2008. ACM.

[9] Google. Google app engine. Accessed 16.3.2009 http://code.google.com/appengine/.

[10] Ian Foster, Yong Zhao, Ioan Raicu, Shiyong Lu. Cloud Computing and Grid Computing 360-Degree Compared. In *Grid Computing Environments Workshop*, pages 1–10, November 2008.

[11] Lizhe Wang, Jie Tao, Marcel Kunze, Alvaro Canales Castellanos, David Kramer, Wolfgang Karl. Scientific Cloud Computing: Early Definition and Experience. In *10th IEEE International Conference on High Performance Computing and Communications*, pages 825–830, September 2008.

[12] Luis M. Vaquero, Luis Rodero-Merino, Juan Caceres, Maik Lindner. A Break in the Clouds: Towards a Cloud Definition. *SIGCOMM Computer Communication Review*, 39(1), December 2008.

[13] Microsoft. Live mesh beta. Accessed 16.3.2009 https://www.mesh.com/Welcome/default.aspx.

[14] Microsoft. What is the azure services platform? Accessed 16.3.2009 http://www.microsoft.com/azure/whatisazure.msp

[15] Nimbus. Nimbus - nimbus open source iaas cloud computing software. Accessed 18.3.2009 http://workspace.globus.org/.

[16] OpenNebula. Opennebula :: start:. Accessed 18.3.2009 http://www.opennebula.org/.

[17] Rajkumar Buyya, Chee Shin Yeo, Srikumar Venugopal. Market-Oriented Cloud Computing:Vision, Hype, and Reality for Delivering IT Services as Computing Utilities. In *10th IEEE International Conference on High Performance Computing and Communications*, pages 5–13, September 2008.

[18] Scott Hazelhurst. Scientific computing using virtual high-performance computing: a case study using the amazon elastic computing cloud. In *SAICSIT '08: Proceedings of the 2008 annual research conference of the South African Institute of Computer Scientists and Information Technologists on IT research in developing countries*, pages 94–103, New York, NY, USA, 2008. ACM.