Demystifying Cloud Computing

Paulo Neto

Faculdade de Engenharia da Universidade do Porto
Rua Dr. Roberto Frias, s/n 4200-465 PORTO, Portugal
pro10004@fe.up.pt

Abstract. The Cloud computing emerges as a new computing style which aims to provide on-demand network access to a shared pool of scalable and often virtualized resources (e.g., networks, servers, storage, applications, and services) that can be quickly provisioned and released. It became a mainstream in 2006 but so far there is still no consensus about its definition. This paper introduces and reviews the Cloud computing regarding to its definition, architecture, security and economical aspects. The purpose of this study is the creation of a baseline to start a PhD research within this subject.

Keywords: Cloud Computing, Grid Computing, Utility Computing

1 Introduction

With the fast improvement of computer processing and storage technologies in conjunction with the Internet success, the hardware resources have become cheaper and widely available. This technological trend has led to a new computing model called cloud computing, in which resources (e.g., CPU, storage and network) are provided as general utilities that can be leased and released by users through the Internet in an on-demand fashion.

Cloud computing can be seen as a platform that hosts applications and services and being driven by three significant trends. First, the wide shift to new Internet-based business models and Web 2.0 applications is driven by the growth in connected devices, real-time data streams, search operations, collaboration and social networking, and consumer-generated data. Second, global organizations are required to become integrated as they look to implement services-oriented architecture (SOA) applications and take advantage of software as a service (SaaS). Third, management, datacenter space and energy costs are driving to the requirement to improve the efficiency on the asset and human resources utilization.

Cloud computing provides computing as a utility. Just as electric companies provide electricity when and where needed, cloud computing vendors dynamically provision, configure, and de-provision IT (information technology) capability as needed, transparently and seamlessly. This allows IT consumers to focus on their specific problems and not on the computing resources they require.

There is an increasingly perceived vision that computing will be the 5th utility (after water, electricity, gas, and telephony) [5].

The remaining parts of this paper are organized as follows. Section 2 presents an overview and a definition of Cloud computing, Section 3 depicts a Cloud computing architecture, the underlying services and the deployment models and technologies. Section 4 compares Cloud with Grid computing emphasizing the role of the Web 2.0. The privacy and data security risks are discussed in Section 5. Section 6 describes some economical aspects and finally Section 7 presents some conclusions about this Cloud computing review.

2 Overview

In this section we present an overview of Cloud Computing including some definitions and a comparison with related concepts.

2.1 Definitions

The underlying concept of cloud computing is not a new one. John McCarthy [20] in 1961, was the first to publicly suggest (in a speech given to celebrate MITs centennial) that computer time-sharing technology might lead to a future in which computing power and even specific applications could be sold through the utility business model (like water or electricity). This idea of a computer or information utility was very popular in the late 1960s, but given up by the mid-1970s as it became clear that the hardware, software and telecommunications technologies of the time were not ready yet for that challenge. However, since 2000, the idea has resurfaced in new forms [20]. The first academic definition was provided by [8] in 1997 who called it a computing paradigm where the boundaries of computing will be determined by economic rationale rather than technical limits. When Eric Schmidt [19,18] explained his cloud computing view on the Search Engine Strategies Conference in 2006 and a couple of weeks later Amazon included the word cloud in the Elastic Cloud Computing (EC2), the term became a mainstream.

In October 2007 Google and IBM [9,14] announced a major research initiative to help students and researchers to address this new Internet-scale computing paradigm. There is still a little consensus how to define Cloud Computing [12]. The Berkeley researchers [16] define Cloud Computing as both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services.

The NIST also published a definition of cloud computing [15]:

> "Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."

From the hardware point of view, there is an illusion of infinite computing resources available on demand. The Cloud users don't need to concern themselves about the required resources for the future growth, and those resources will be paid in a pay-per-usage basis.

The Figure 1 shows the main characteristics of a cloud service. It needs to be p
dem

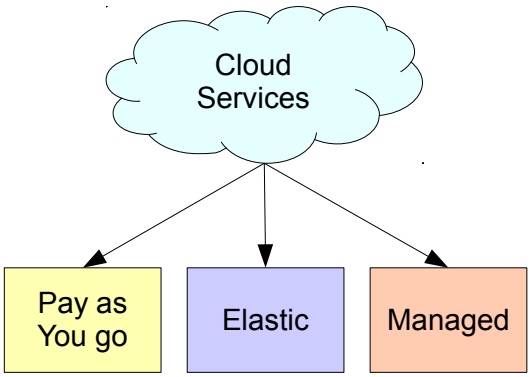

Fig. 1. Characteristics of a cloud service.

3 Cloud Computing Architecture

This section presents the cloud computing architecture, the typical deployment models as well as the underlying technologies.

3.1 Cloud Service Levels

Cloud services can be classified into four general types [6]:

Software as a Service (SaaS)
This is the most common type of cloud service and one that almost everyone has already used at some point. In the SaaS cloud model, the service provider supplies all the infrastructure along with the software product. Users interact with the service using a Web-based front end. These services cover a wide range, from Web-based e-mail like GMail to financial software like Mint [3].

Platform as a Service (PaaS)
Cloud service that provides software and product development tools hosted by the provider on the hardware infrastructure. The term PaaS is commonly used for cloud-based platforms to build and run custom applications. PaaS applications provide everything needed to build and deploy Web applications

and services accessible from anywhere on the Internet. The end users do not have to download, install, or maintain the system. Popular examples of this kind of a service are Google App Engine, Microsoft Windows Azure, Force.com, Morph and Bungee Connect [7].

Infrastructure as a Service (IaaS)
Cloud services that provide access to computing resources such as processing or storage which can be obtained as a service. The most popular examples of IaaS are Amazon Web Services (AWS) with its Elastic Compute Cloud (EC2) [1] for processing and Simple Storage Service (S3) [2] for storage , and and Rackspace [4].

data Storage as a Service (dSaaS)
Services that provide storage to be used by the consumer including bandwidth requirements [11].

Cloud computing can be viewed as a collection of services, which can be presented in layers services, as shown in Figure 2.

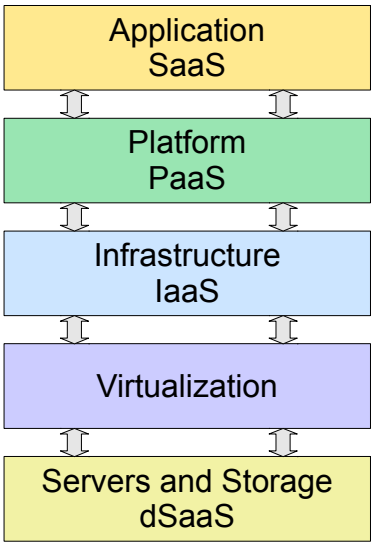

Fig. 2. Layered architecture of Cloud Computing.

3.2 Cloud Computing Deployment Models

There are four types of cloud computing deployment models [21]:

Public Cloud
A public cloud is the traditional concept where the resources are leased

through the Internet from an off-site third-party provider who bills in a pay per usage basis.

Community cloud
A community cloud is usually used by a set of organizations with similar requirements and interests in order to share the infrastructure but keeping a certain additional level of security and privacy. Examples of community cloud include Google's "Gov Cloud".

Hybrid cloud
Even without consensus with this term, a hybrid cloud is probably the use of physical hardware and virtualized cloud server instances together to provide a single common service. For instance, a hybrid storage cloud can be used in a tier fashion for archiving, backup and replication functionalities.

Private cloud
In private clouds the idea of leasing instead of buying and manage is lost, but for big organizations it can be acceptable with the IT department (or the hosting entity) playing the provider role. The storage and server hardware providers also allow some pay per usage policy, like having additional processors and storage capacity that will be charged just when temporarily activated (Capacity on Demand). Thus even in private clouds the backend overprovisioning can be avoided.

3.3 Technologies Behind a Cloud

Numerous underlying precursor technologies enabled cloud computing to emerge, like:

- Internet
- Virtualization
- Software-as-a-Service
- LAMP and WAMP stacks
- Web Hosting
- Database
- Inexpensive CPUs and storage
- SOA (service-oriented architectures)
- Sophisticated client algorithms, including HTML, CSS, AJAX, REST
- Client broadband
- SOA (service-oriented architectures)
- Large infrastructure implementations from Google, Yahoo, Amazon, and others

4 Comparing with Grid Computing

The definition of Cloud Computing overlaps with many existing technologies, such as Grid Computing, Utility Computing, Services Computing, and distributed computing in general. Cloud Computing is indeed evolved out of Grid

Computing and relies on Grid Computing as its backend and infrastructure support. The evolution has been a result of a shift in focus from an infrastructure that delivers storage and processing resources (such is the case in Grids) to one that is economy based in order to deliver more abstract resources and services (such is the case in Clouds) [10]. Cloud and grid systems share the same basic goal to reduce the computing costs. Grids are mostly designed to be general purpose, so they exhibit a complete set of available system capabilities and the resulting interface available for users and applications remains low level [13]. In contrast to the Grid system interfaces, cloud system interfaces are simpler and they do not expose internal system characteristics. Typically the exposed capability set is usually much more limited than the set of capabilities available in the Cloud system itself. Figure 3 adapted from [10] shows an overview of the relationship between Clouds and other domains that it overlaps with. Web 2.0 covers almost the whole spectrum of service-oriented applications, where Cloud Computing lies at the large-scale side. Supercomputing and Cluster Computing have been more focused on traditional non-service applications. Grid Computing overlaps with all these fields.

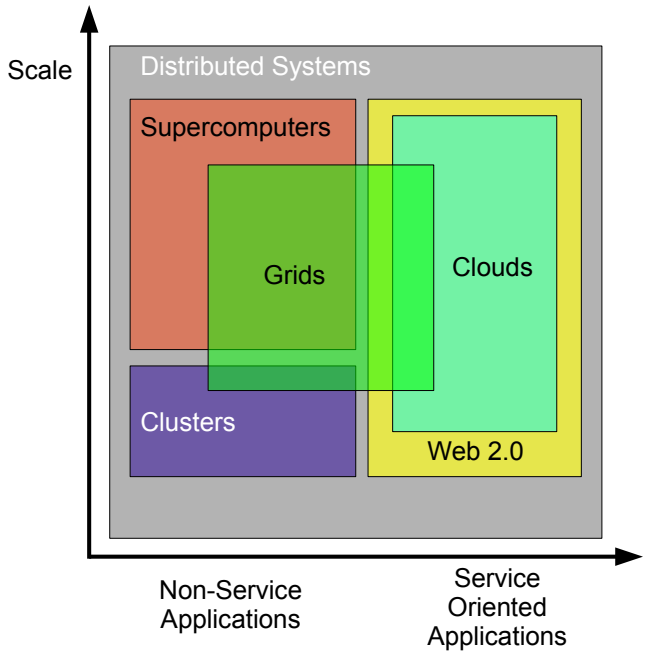

Fig. 3. Grids and Clouds interaction.

5 Privacy and Data Security Risks

Although the many advantages of cloud computing already presented, some security issues need to be carefully evaluated. Processing sensitive data outside the companies require additional cautions since the data bypass the physical, logical and personnel controls. The customers need to follow data retention policies enforced by the regulatory laws, thus the service provider need to be prepared and certified for those requirements. As long as the data is spread across multiple locations a special concern need to be taken regarding to the law on those specific locations (state/countries). In order to keep the data confidentiality, a widely tested encryption scheme should used because an encryption accident can make the data totally unusable. To protect against theft and or denial-of-service attacks by users, virtualization is the primary security mechanism in the today's clouds. It's a well known technology that protects against most attempts by users to attack one another or the underlying cloud infrastructure. Nevertheless neither all resources are virtualized nor the virtualization engines are bug free. Another important concern is the protection against the cloud provider. The virtualization technologies may allow the one who manages the lower layers to circumvent many security barriers. Although there is already strict legislation in US and EU to regulate and audit cloud computing providers [17], there is still some reluctance by several companies to adopt this new technology at least in a public cloud.

6 Cloud Computing Economics

One of the basic policies for any organization is the TCO (Total Cost of Ownership) reduction. For instance, provisioning a data center for the peak load it must sustain during the end of the month (batch jobs), leads to a resource underutilization during remaining time of the month. An organization in this situation may benefit from cloud computing paying by the hour for computing leading to cost savings even if the hourly rate to rent a machine from a cloud provider is higher than the rate to own one (including space, power, cooling and maintenance costs).

One of the basic policies for any organization is the reduction the TCO (Total Cost of Ownership). For instance, provisioning a data center for the peak load it must sustain during the end of the month leads to underutilization during the other days. In this situation an organization may benefit from cloud computing paying by the hour for computing leading to cost savings even if the hourly rate to rent a machine from a cloud provider is higher than the rate to own one (including space, power, cooling and maintenance costs).

Another example is a business startup that requires some computing resources at the beginning but will suffer an unexpected increase of the resources demand when it becomes popular. This type of organization doesn't need to invest upfront in computer resources that potentially will be used only months or years later.

Although the clouding computing costs are more expensive compared with owning a server for the same period, the cost of over-provisioning (having the server in almost idle status during many hours per day) or under-provisioning (increasing the system response time) is very high, driving the cloud computing solution very attractive from the business point of view.

In Figure 4 with a correctly antecipated peak load, there is a waste of resources due the lack of elasticity. The Figure 5 represents a typical underprovisioned environment where the demand overpass the capacity during the peak load and still presenting long periods of low resource usage. Figure 6 represents an on demand provisioning where the resources are dynamically adjusted based on the demand (elasticity) providing a great cost effective solution.

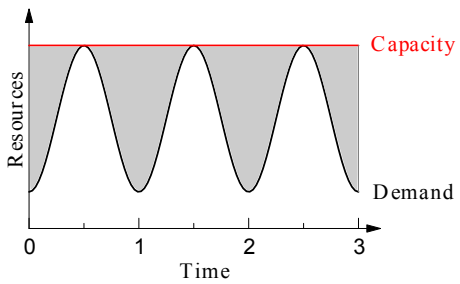

Fig. 4. Provisioning for peak load.

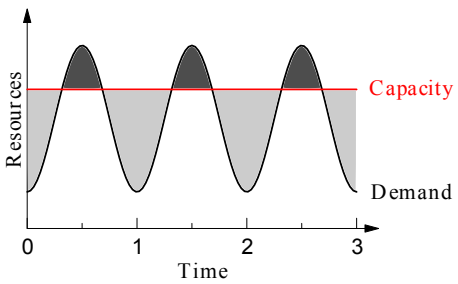

Fig. 5. Underprovisioning.

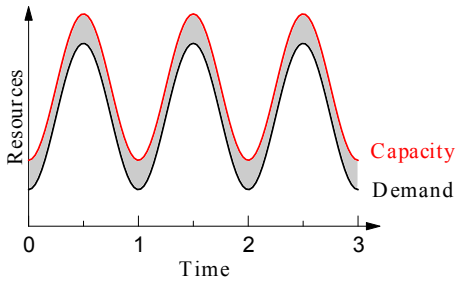

Fig. 6. On-Demand Provisioning.

In Figures 4,5 and 6, the grey area represents the waste of resources making the on-demand provision the most efficient.

Although the economical benefits explained before there are also some negative aspects using Cloud Computing. For instance transferring a large database over the network involves a significant cost (high bandwidth) and time. Some companies choose to ship the data in physical support (tape or disk) through a

courier in order to reduce the network costs and the time that takes the data import activity.

7 Conclusion

This paper reviewed the Cloud Computing technology from the architecture and deployment models point of view and presented some economic advantages and disadvantages and potential security issues. It has been produced as baseline to start a PhD research within this subject and aims to contribute to the Cloud computing evolution. Since cloud computing has embedded a blend of technologies, with so many opened issues like security, it may drive thousands of research projects that will change the computing paradigms during the current decade.

References

1. Amazon Elastic Compute Cloud (Amazon EC2). `http://aws.amazon.com/ec2/`, [Accessed December 2010]
2. Amazon Simple Storage Service (Amazon S3). `http://aws.amazon.com/s3/`, [Accessed December 2010]
3. Mint - Free Personal Finance Software. `http://www.mint.com/`, [Accessed January 2011]
4. The Rackspace Cloud. `http://www.rackspacecloud.com/`, [Accessed December 2010]
5. Buyya, R., Yeo, C., Venugopal, S., Broberg, J., Brandic, I.: Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. Future Generation Computer Systems 25(6), 599–616 (2009)
6. Chaganti, P.: Cloud services for your virtual infrastructure, Part 1: Infrastructure-as-a-Service (IaaS) and Eucalyptus. IBM developerWorks (December 2009), [Accessed December 2010]
7. Chaganti, P.: Cloud services for your virtual infrastructure, Part 2: Platform as a Service (PaaS) and AppScale. IBM developerWorks (January 2010), [Accessed December 2010]
8. Chellappa, R.: Cloud computing: emerging paradigm for computing. INFORMS 1997 (1997)
9. Chiu, W.: Google and IBM Initiative for Computer Students. `http://www.ibm.com/ibm/ideasfromibm/us/google/images/podcast_google.pdf` (October 2007), [Accessed December 2010]
10. Foster, I., Zhao, Y., Raicu, I., Lu, S.: Cloud computing and grid computing 360-degree compared. In: Grid Computing Environments Workshop, 2008. GCE'08. pp. 1–10. Ieee (2009)
11. Furht, B., Escalante, A.: Handbook of Cloud Computing. Springer (2010)
12. Geelan, J.: Twenty-One Experts Define Cloud Computing. `http://cloudcomputing.sys-con.com/node/612375` (January 2009), [Accessed December 2010]
13. Jha, S., Merzky, A., Fox, G.: Using clouds to provide grids with higher levels of abstraction and explicit support for usage modes. Concurrency and Computation: Practice and Experience 21(8), 1087–1108 (2009)

14. Lohr, S.: Google and I.B.M. Join in Cloud Computing Research. The New York Times (October 2007), [Accessed December 2010]
15. Mell, P., Grance, T.: The NIST Definition of Cloud Computing. National Institute of Standards and Technology (2009)
16. Michael, A., Armando, F., Rean, G., Anthony, D., Randy, K., Andy, K., Gunho, L., David, P., Ariel, R., Ion, S., et al.: Above the clouds: A berkeley view of cloud computing. EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2009-28 (2009)
17. Sotto, L., Treacy, B., McLellan, M.: Privacy and Data Security Risks in Cloud Computing. Electronic Commerce & Law Report (2010)
18. Sullivan, D: Conversation with Eric Schmidt hosted by Danny Sullivan. (http://www.google.com/press/podium/ses2006.html) (August 2006), [Accessed December 2010]
19. Wikipedia, the free encyclopedia: Eric Schmidt. `http://en.wikipedia.org/wiki/Eric_Schmidt` (2010), [Accessed December 2010]
20. Wikipedia, the free encyclopedia: John McCarthy (computer scientist). `http://en.wikipedia.org/wiki/John_McCarthy_(computer_scientist)` (2010), [Accessed December 2010]
21. Wikipedia, the free encyclopedia: Cloud computing. `http://en.wikipedia.org/wiki/Cloud_computing` (2011), [Accessed January 2011]