Chapter 1

Fundamentals of Cloud Computing

1.1. Introduction

In the introductory, we define the concept of cloud computing, describe the history, services, and types of cloud computing. The key technologies that enabled cloud computing are presented next. Then, we also discuss the issues about cloud computing features, standards, and security, and introduce the key cloud computing platforms, their vendors, and their offerings. We discuss cloud computing challenges and the future of cloud computing at last.

Cloud computing can be defined as an information technology (IT) paradigm that enables ubiquitous access to shared pools of configurable system resources and higher-level services that can be rapidly provisioned with minimal management effort, often over the Internet. Cloud computing relies on sharing of resources to achieve coherence and economies of scale, similar to a public utility. Advantages of the cloud computing technology include cost savings, easy scalability, and high availability.

1.1.1. History of Cloud Computing

Before we discuss the other topics of cloud computing, let's first see from a historical perspective where cloud computing fits in and whether it is really different or just the next fad.

Figure 1, adapted from Voas and Zhang (2009), shows six phases of computing paradigms, from dummy terminals/mainframes, to PCs, network computing, grid and cloud computing.

In Phase 1, dummy terminals are used to connect to powerful hosts shared by many users. At that time, the terminals were basically keyboards and monitors. In the second stage, independent personal computers (PCs) became powerful enough to satisfy users' daily work, which means that you didn't have to share a mainframe with anyone else. In the third phase a computer networks that

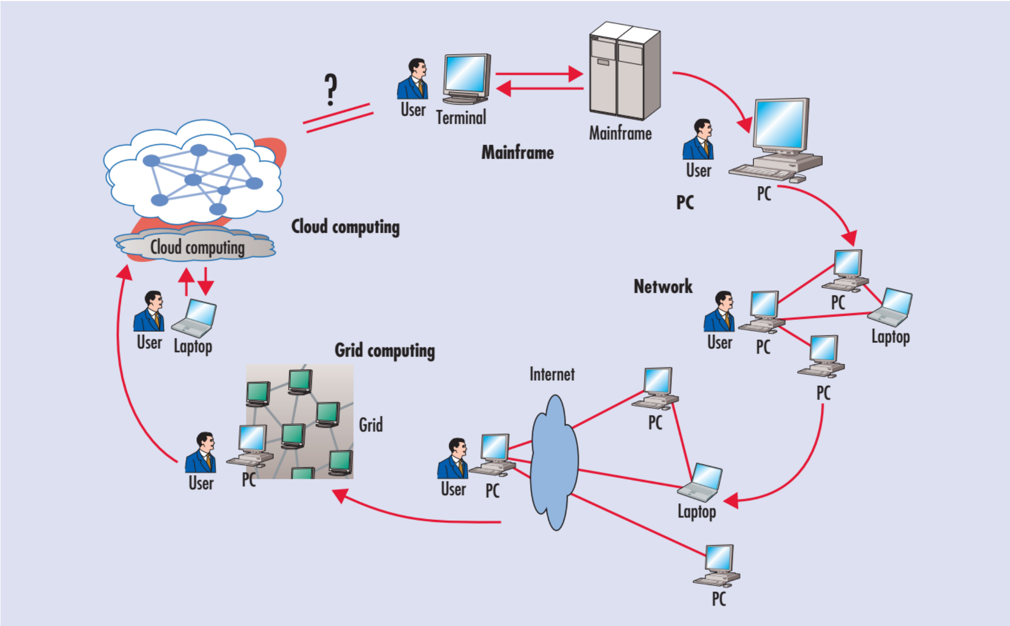

Figure 1. Computing Paradigm shift. Figure from Voas and Zhang (2009).

allowed multiple computers to connect to each other. You could work on a PC and connect to other computers through local networks (LAN) to share resources. In Phase 4, local networks were connected to other local networks to establish a more global network - users could now connect to the Internet to utilize remote applications and resources. In Phase 5 brought us the concept of an electronic grid to facilitate shared computing power and storage resources (distributed computing). People used PCs to access a grid of computers transparently. Now, in Phase 6, we can leverage all available resources on the Internet in an extremely scalable and simple way by cloud computing.

As **Figure 1** shows, a conceptual layer, a cloud on the Internet, hides all available resources (either hardware or software) and services, but it publishes a standard interface. As long as users can connect to the Internet, they have the entire Web as their power PC. Cloud computing thus refers to the techniques that enable and facilitate this scenario.

Comparing these six computing paradigms, it looks like that cloud computing is a return to the original mainframe computing paradigm. However, these two paradigms have several important differences. Unlike a mainframe, which is a physical machine that offers finite computing power, a cloud represents all possible resources on the Internet, suggesting infinite power and capacity. Meanwhile, unlike a simple terminal acting as a user interface to a mainframe, a PC in the cloud computing paradigm possesses significant power to provide a certain degree of local computing and caching support. In short, cloud computing has become a significant technology trend and could reshape the IT sector and the IT marketplace.

1.1.2. Layers of Cloud Computing

According to the abstraction level of the capability provided and the service model of providers, cloud computing services are divided into three classes: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) (Mell and T. Grance, 2009). **Figure 2** illustrates the layered organization, from the physical infrastructure to applications, of the cloud stack.

These abstraction levels can also be viewed as a layered architecture where services of a higher layer can be composed from services of the underlying layer

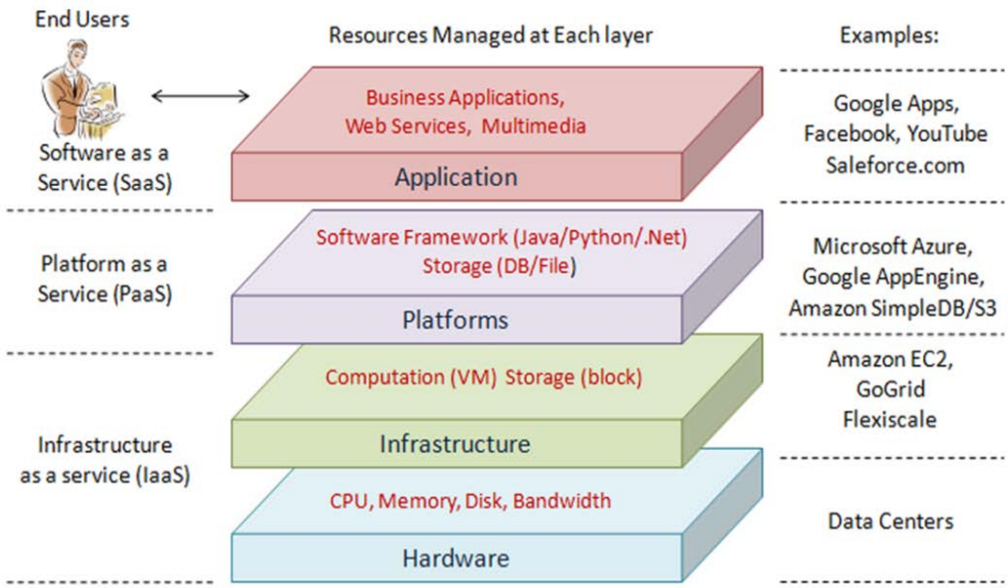

Figure 2. The cloud computing architecture. Figure from (Zhang et al, 2010).

(Buyya et al., 2011). The reference model of Buyya et al., (2011) explains the role of each layer in an integrated architecture. A core middleware manages physical resources and the VMs deployed on top of them; in addition, it provides the required features (e.g., accounting and billing) to offer multi-tenant pay-as-you-go services. Cloud development environments are built on top of infrastructure services to offer application development and deployment capabilities; in this level, various programming models, libraries, APIs, and Mashup editors enable the creation of a range of business, Web, and scientific applications. Once deployed in the cloud, these applications can be consumed by end users.

1) Infrastructure as a Service (IaaS)

Offering virtualized resources (computation, storage, and communication) on demand is known as Infrastructure as a Service (IaaS) (Sotomayor et al., 2009). The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications; and possibly limited control of select networking components (e.g., host firewalls).

Amazon Web Service mainly offers IaaS, which in the case of its EC2 service means offering VMs with a software stack that can be customized similar to how an ordinary physical server would be customized. Users are given privileges to perform numerous activities to the server, such as: starting and stopping it, customizing it by installing software packages, attaching virtual disks to it, and configuring access permissions and firewalls rules.

2) Platform as a Service (PaaS)

The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider (This capability does not necessarily preclude the use of compatible programming languages, libraries, services, and tools from other sources). The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment.

Google AppEngine, an example of Platform as a Service, offers a scalable environment for developing and hosting Web applications, which should be written in specific programming languages such as Python or Java, and use the services' own proprietary structured object data store. Building blocks include an in-memory object cache (memcache), mail service, instant messaging service (XMPP), an image manipulaton service, and integration with Google Accounts authentication service.

3) Software as a Service (SaaS)

The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure (A cloud infrastructure is the collection of hardware and software that enables the five essential characteristics of cloud computing. The cloud infrastructure can be viewed as containing both a physical layer and an abstraction layer. The physical layer consists of the hardware resources that are necessary to support the cloud services being provided, and typically includes server, storage and network components. The abstraction layer consists of the software deployed across the physical layer, which manifests the essential cloud characteristics. Conceptually the abstraction layer sits above the physical layer.). The applications are accessible from various client devices through either a thin client interface, such as a web browser (e.g., web-based email), or a program interface. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user specific application configuration settings.

Applications reside on the top of the cloud stack. Services provided by this layer can be accessed by end users through Web portals. Therefore, consumers are increasingly shifting from locally installed computer programs to on-line software services that offer the same functionally. Traditional desktop applications such as word processing and spredsheet can now be accessed as a service in the Web. This model of delivering applications, known as Software as a Service (SaaS), alleviates the burden of software maintenance for customers and simplifies development and testing for providers.

Salesforce.com, which relies on the SaaS model, offers business productivity application (CRM) that reside completely on their servers, allowing costumers to customize and access applications on demand.

1.1.3. Types of Cloud Computing

Although cloud computing has emerged mainly from the appearance of public computing utilities, other deployment models, with variations in physical location and distribution, have been adopted. In this sense, regardless of its service class, a cloud can be classified as public, private, community, or hybrid (Mell and T. Grance, 2009), based on model of deployment as shown in **Figure 3**.

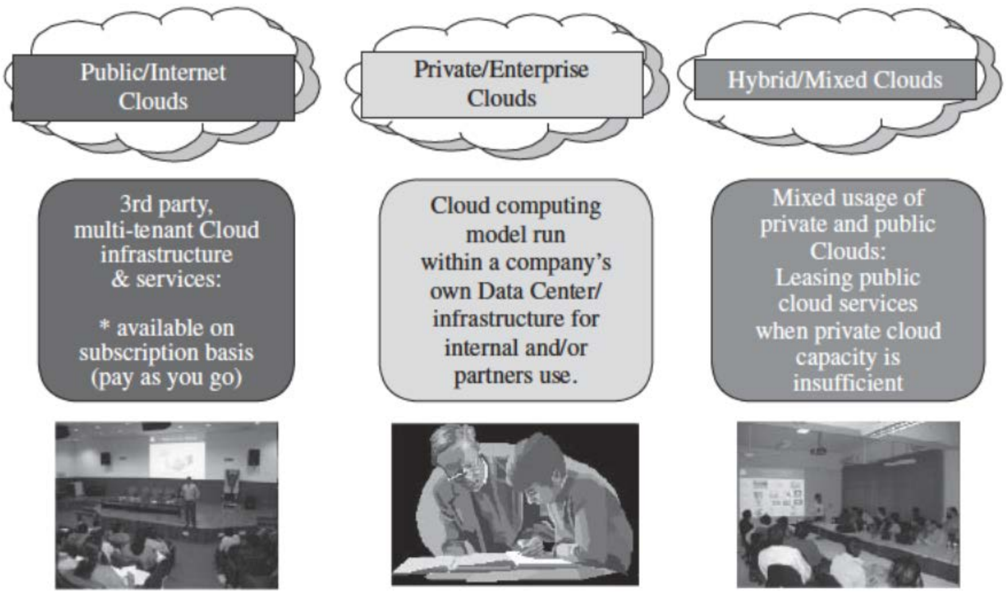

Figure 3. Types of cloud. Figure from (Buyya et al, 2009).

1.2. Key Technologies

This section introduces key technologies that support cloud computing. They include virtualization, Web service and service-oriented architecture, service flows and workflows, as well as Web 2.0 and Mashup.

1.2.1. Virtualization

The advantage of cloud computing is the ability to virtualize and share resources among different applications with the objective for better server utilization.

The underpinning for the majority of high-performing clouds is a virtualized infrastructure. Virtualization has been in data centers for several years as a successful IT strategy for consolidating servers. Used more broadly to pool infrastructure resources, virtualization can also provide the basic building blocks for your cloud environment to enhance agility and flexibility.

Today, the primary focus for virtualization continues to be on servers. However, virtualizing storage and networks is emerging as a general strategy. Results from a Gartner survey of 505 data center managers worldwide reports that planned or in-process virtualization of infrastructure workloads will increase from approximately 60 percent in 2012 to almost 90 percent in 2014. This continuing growth makes cloud computing an obvious next step for many organizations.

Virtualization is the foundation for an agile, scalable cloud—and the first practical step—for building cloud infrastructure. Virtualization abstracts and isolates the underlying hardware as virtual machines (VMs) in their own runtime environment and with multiple VMs for computing, storage, and networking resources in a single hosting environment. These virtualized resources are critical for managing data, moving it into and out of the cloud, and running applications with high utilization and high availability.

Virtualization is managed by a host server running a hypervisor—software, firmware, or hardware that creates and runs VMs. The VMs are referred to as guest machines. The hypervisor serves as a virtual operating platform that executes the guest operating system for an application. Host servers are designed to run multiple VMs sharing multiple instances of guest operating systems.

Virtualization also provides several key capabilities for cloud computing, including resource sharing, VM isolation, and load balancing. In a cloud environment, these capabilities enable scalability, high utilization of pooled resources, rapid provisioning, workload isolation, and increased uptime.

Today, the trend in virtualization has moved from reducing costs by consolidating data centers to increasing flexibility and agility through the pervasive use of virtualization for faster service deployment and dynamic placement of workloads. Pervasive virtualization is a strategic approach that provides a method for judiciously bringing legacy applications into your cloud to meet your strategic goals or as time and budget allow. Its benefits include better quality of service, improved availability and business continuity, faster resource deployment, and lower energy consumption. The role of virtualization is illustrated in **Figure 4**.

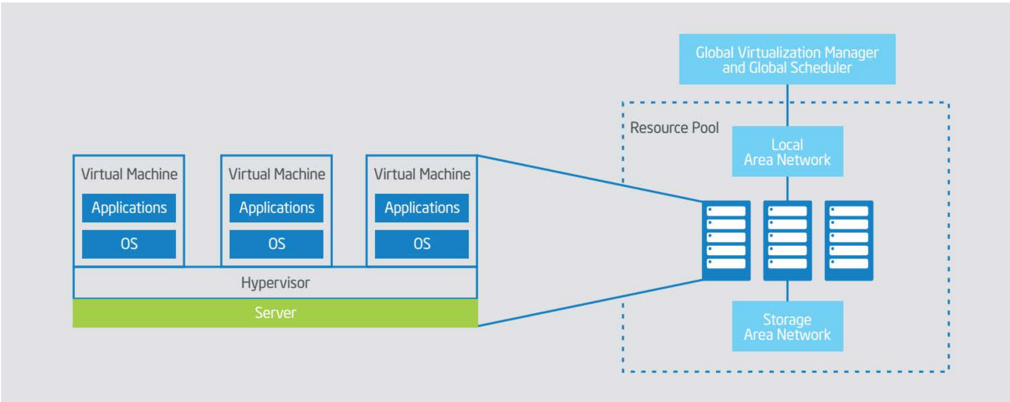

Figure 4. The role of virtualization.

Virtualization technologies include virtual machine techniques such as VMware and Xen, and virtual networks, such as VPN. Virtual machines provide virtualized IT-infrastructures on-demand, while virtual networks support users with a customized network environment to access cloud resources.

1.2.2. Web Service and Service Oriented Architecture

Web Services and Service Oriented Architecture (SOA) are not new concepts, but they represent the base technologies of cloud computing. Cloud services are usually designed as Web services, followed by industry standards including

WSDL, SOAP, and UDDI. A Service Oriented Architecture organizes and manages Web services inside clouds (Vouk, 2008). A SOA also includes a set of cloud services, which can be used on various distributed platforms.

1.2.3. Service Flow and Workflows

The concept of service flow and workflow refers to an integrated view of service-based activities in the cloud. Workflows have become one of the important research fields in the field of database and information systems (Vouk, 2008).

1.2.4. Web 2.0 and Mashup

Web 2.0 is a new concept that use of Web technology and Web design to enhance creativity, information sharing, and collaboration among users (Wang et al., 2008). Mashup, on the other hand, is a web application that combines data from multiple source into a single integrated storage tool. These two technologies are very beneficial for cloud computing.

Figure 5 shows a cloud computing architecture, adapted from (Hutchinson and Ward, 2009), in which an application reuses various components. The components in this architecture are dynamic in nature, run in a SaaS model, and use SOA. The components closer to the user are smaller in nature and more reusable. The components of the center include aggregate and extend services via Mashup servers and portals. Data from a service (such as addresses in a database) can be mashed up with mapping information (such as Yahoo or Google maps) to generate an aggregated views of the information.

1.3. Features of Cloud Computing

Compared with other computing paradigms, Cloud computing brings many new features (Wang et al., 2008; Grossman, 2009). There are briefly described in this section.

- Scalability and on-demand services

Cloud computing provides users with resources and services according to demand. Resources are scalable over multiple data centers.

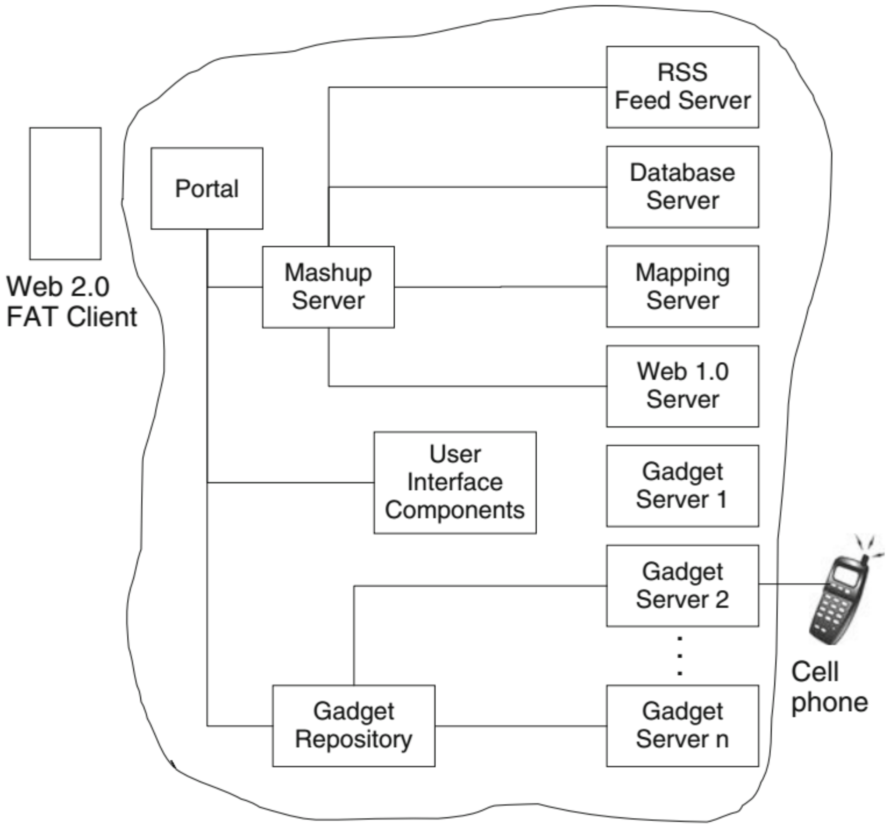

Figure 5. Cloud computing architecture uses various components at different levels (adapted from Hutchinson and Ward (2009)).

- User-centred interface

Cloud interfaces are location independent and can be accesses by well established interfaces such as Web services and Internet browsers.

- Guaranteed Quality of Service (QoS)

Cloud computed can guarantee users' QoS in terms of hardware/CPU performance, bandwidth, and memory capacity.

- Autonomous system

The cloud computing systems are autonomous systems managed transparently to users. However, the software and data in the clouds can be automatically reconfigured and integrated to a simple platform according to the needs of users.

- Pricing

Cloud computing does not need up-from investment. There is no need for capital expenditure. Users pay for services and capacity as they need them.

1.3.1. Cloud Computing Standards

Cloud computing standards have not been yet fully developed; however mangy of existing generally lightweight, open standards have promoted the development of cloud computing ("Cloud Computing," Wikipedia, http://en.wikipedia.org/wiki/Cloud_computing). **Table 1** illustrates some of these open standards, which are currently being used in cloud computing.

Table 1. Cloud computing standards ("Cloud Computing", Wikipedia, http://en.wikipedia.org/wiki/Cloud_computing).

Applications	Communications: HTTP, XMPP Security: OAuth, OpenID, SSL/TLS Syndication: Atom
Client	Browsers: AJAX Offlines: HTML5
Implementations	Virtualization: OVF
Platform	Solution stacks: LAMP
Service	Data: XML, JSON Web services: REST

1.3.2. Cloud Computing Security

One of the key issues in implementing cloud computing is bringing virtual machines that contain critical applications and sensitive data, to public and shared cloud environments. As a result, potential cloud computing users worry about the following security issues ("Cloud Computing Security," Third Brigade, http://www.cloudreadysecurity.com/).

- Will the users still have the same security policy control over their applications and services?

- Can it be proved to the organization that the system is still safe and meets SLA?
- Is the system complaint and can it be verified to auditors of the company?

In traditional data centers, the common security methods include peripheral firewall, non military zones, network segmentation, intrusion detection and prevention systems, and network monitoring tools.

The security requirements for cloud computing providers start with the same technologies and tools as for the traditional data centers, including the application of the strong network security boundaries. However, physical segmentation and hardware based security cannot prevent attacks between virtual machines on the same server. Cloud computing servers use the same operating systems, enterprise and Web applications as localized virtual machines and physical servers. Therefore, attackers can remotely exploit loopholes in these systems and applications. In addition, co-location of multiple virtual machines will increase the attack area and risk to MV-to-VM compromise. Intrusion detection and defense systems need to be able to detect malicious activity in the VM level, regardless of the location of the VM within the virtualized cloud environment ("Cloud Computing Security," Third Brigade, http://www.cloudreadysecurity.com).

In summary, the virtual environments that deploy the security mechanisms on virtual machines including firewalls, intrusion detection and prevention, integrity monitoring, and log inspection, will effectively secure the VM cloud and prepare for deployment.

1.4. Cloud Computing Platforms

Cloud computing has great commercial potential. According to market research firm IDC, IT cloud services spending will grow from about $16B in 2008 to about $42B in 2012 and to increase its share of overall IT spending from 4.2% to 8.5%.

Table 2 presents key players in cloud computing platforms and their key offerings.

Table 2. Key Players in Cloud Computing Platforms (adapted from Lakshmanan (2009)).

Company	Cloud computing platform	Year of launch	Key offerings
Amazon.com	AWS (Amazon Web Services)	2006	Infrastructure as a service (Storage, Computing, Message queues, Datasets, Content distribution)
Microsoft	Azure	2009	Application platform as a service (.Net, SQL data services)
Google	Google App.Enging	2008	Web Application Platform as a service (Python run time environment)
IBM	Blue Cloud	2008	Virtualized Blue cloud data center
Salesforce.com	Force.com	2008	Proprietary 4GL Web application framework as an on Demand platform
Baidu	Baidu Cloud	2015	Infrastructure as a service (Storage, Computing, Message queues, Datasets, Content distribution)
Alibaba	Ali Cloud	2009	Infrastructure as a service (Storage, Computing, Message queues, Datasets, Content distribution)
Tencent	QCloud	2010	Infrastructure as a service (Storage, Computing, Message queues, Datasets, Content distribution)
Huawei	Huawei Cloud	2011	Infrastructure as a service (Storage, Computing, Message queues, Datasets, Content distribution)

Table 3 compares three cloud computing platforms, Amazon, Google, and Microsoft, in terms of their capabilities to map to different development models and scenarios ("Which Cloud Platform is Right for You?," http://www.cumulux.com/).

1.4.1. Pricing

Pricing for cloud platforms and services is based on three key dimensions: (i) compute, (ii) storage, and (iii) bandwidth.

Compute is based on the time unit required to process requests by running instances, applications or computers. **Table 6** compares the pricing of three major cloud computing platforms.

Table 3. Cloud Computing Platforms and Different Scenarios (adapted from "Which Cloud Platform is Right for You?," http://www.cumulux.com/).

(1) Scenario	On-premise application unchanged in the cloud
Characteristics	Multiple red legacy, java or .NET based application
Amazon	Threat the machine as another server in the data center and do the necessary changes to configuration
Google	Needs significant refactoring of application and data logic for existing Java application
Microsoft	If existing app is ASP.NET application, then re-factor data, otherwise refactoring effort can be quite significant depending on the complexity
(2) Scenario	Scalable Web application
Characteristics	Moderate to high Web application with a back-end store and load balancing
Amazon	Threat the machine instance as another server in the data center and do the necessary changes to configuration. But scalability and elasticity is manual configuration
Google	Use dynamically scalable features of App Engine and scripting technologies to build rich applications
Microsoft	Build scalable Web applications using familiar .NET technologies. Scalingup/down purely driven by configuration.
(3) Scenario	Parallel processing computational application
Characteristics	Automated long running processing with little to no user interaction.
Amazon	Need to configure multiple machine instances depending on the scale needed and manage the environments.
Google	Platform has minimal built-in support for building compute heavy applications. Certain application scenarios, such as image manipulation, are easier to develop with built-in platform features.
Microsoft	With worker roles and storage features like Queues and blobs, it is easy to build a compute heavy application that can be managed and controlled for scalability and elasticity.
(4) Scenario	Application in the cloud interacts with on-premise data
Characteristics	Cloud based applications interacting with on-premise apps for managing transactions of data
Amazon	Applications in EC2 server cloud can easily be configured to interact with applications running on premise.
Google	No support from the platform to enable this scenario. Possible through each app using intermediary store to communicate.
Microsoft	From features like Service Bus to Sync platform components it is possible to build compelling integration between the two environments.

Continued

(5) Scenario Characteristics	Application in the cloud interacts with on-premise application
Amazon	On-premise applications
	Applications in EC2 server cloud can easily be configured to interact with applications running on premise.
Google	No support from the platform to enable this scenario. Possible through each app using intermediary store to communicate.
Microsoft	From features like Service Bus to Sync platform components it is possible to build compelling integration between the two environments.

Storage is usually measured by the amount of data stored in GB per month per month.

Bandwidth is measured by calculating the total data of incoming and outgoing platform services through transactions and batches. In general, data transmission between services on the same platform is free on many platforms.

In summary, by analyzing the cost of cloud computing, the cost of deploying an application may vary according to the characteristics of the application. As can be seen from **Table 4**, the unit price of the three platforms is very similar. In addition to unit pricing, it is also important to transform it into monthly application development, deployment and maintenance costs.

Table 4. Pricing comparison for major cloud computing platforms (adapted from "Which Cloud Platform is Right for You?," http://www.cumulux.com/.)

Resource	UNIT	Amazon	Google	Microsoft
Stored data	GB per month	$0.10	$0.15	$0.15
Storage transaction	Per 10 K requests	$0.10	--	$0.10
Outgoing bandwidth	GB	$0.10 - $0.17	$0.12	$0.15
Incoming Bandwidth	GB	$0.10	$0.10	$0.10
Compute time	Instance Hours	$0.10 - $1.20	$0.10	$0.12

1.4.2. Cloud Computing Components

The main elements comprising cloud computing platforms include computer hardware, storage, infrastructure, computer software, operating systems, and platform virtualization. The leading vendors providing cloud computing components are shown in **Table 5** ("Cloud Computing," Wikipedia, http://en.wikipedia.org/wiki/Cloud_computing).

Table 5. The leading vendors of cloud computing components.

Cloud computing components	Vendors
Computer hardware	Dell, HP, IBM, Sun
Storage	Sun, EMC, IBM
Infrastructure	Cisco, Juniper Networks, Brocade Communication
Computer systems	3tera, Eucalyptus, G-Eclipse, Hadoop
Operating systems	Solaris, AIX, Linux (RedHat, Ubuntu)
Platform virtualization	Citrix, VMWare, IBM, Xen, Linux KVM, Microsoft, Sun xVM

1.5. Challenges of Cloud Computing

In summary, the new paradigm of cloud computing offers many of benefits and advantages over the previous computing paradigms and many organizations are adopting it. However, there are still many challenges that researchers and practitioners in the field are addressing (Leavitt, 2009). They are briefly introduced as follows.

1.5.1. Performance

The main performance problem may be for some intensive transaction-oriented and other data-intensive applications, where cloud computing may lack sufficient performance. In addition, users who are far away from cloud providers may encounter high latency and delay.

1.5.2. Security and Privacy

When using cloud computing, the company is still worried about security. When information and time, the customer is worried about the vulnerability of the attack. The key IT resources are outside the firewall. The security solution assumes that cloud computing providers follow standard security practices, as mentioned earlier.

1.5.3. Control

Some IT departments are worried because cloud computing providers have a full control of the platforms. Cloud computing providers typically do not design platforms for specific companies and their business practices.

1.5.4. Bandwidth Costs

Through cloud computing, companies can save cost on hardware and software; however they could incur higher network bandwidth costs. The bandwidth cost of smaller Internet-based applications may be low, which are not data intensive, but they may grow significantly for data-intensive applications.

1.5.5. Reliability

Cloud computing still can not always offer round-the-clock reliability. In some case, cloud computing services suffered a few-hours outages.

In the future, we can expect more cloud computing providers, richer services, established standards, and best practices.

In the field of research, HP Labs, Intel, and Yahoo have launched the distributed Cloud Research Test Bad, with facilities in Asia, Europe, and North America, with the aim of developing innovations including cloud computing specific chips. IBM has launched the Research Computing Cloud, which is an on-demand, globally accessible set of computing resources that support business processes.

1.6. Future of Cloud Computing

In summary, cloud computing is definitely a computing paradigm/framework that will remain in the future for a long time. In the near future, cloud computing can occur in all directions. One possible solution in the future is that an enterprise may use a distributed hybrid cloud as illustrated in **Figure 6**.

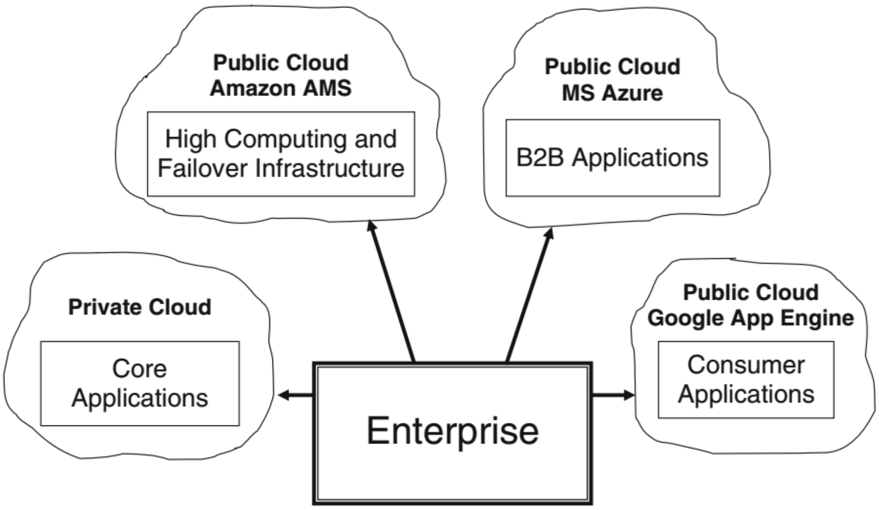

Figure 6. Distributed hybrid cloud architecture (adapted from Lakshmanan (2009)).

In this case, the enterprise will use the core applications on its private cloud, while some other applications will be distributed on multiple private clouds, which are optimized for specific applications.

References

Grossman, R. L., 2009, The case for cloud computing. IEEE ITPro, 23-27.

Hutchinson, C., and Ward, J., 2009, Navigation the next-generation application architecture. IEEE ITPro, 18–22.

Leavitt, N., 2009, Is cloud computing really ready for prime time? IEEE Computer, 15-20.

Mell, P., & Grance, T., 2009, The NIST Definition of Cloud Computing, National Institute of Standards and Technology, Information Technology Laboratory, Technical Report Version 15.

Rajkumar Buyya, James Broberg, and Andrzej Goscisk, 2011, Cloud Computing: Principles and Paradigms, Wiley, A John Wiley & Sons, Inc., Publication.

Sotomayor, B., Montero, R. S., Llorente, I. M., & Foster, I., 2009, Virtual infrastructure management in private and hybrid clouds, IEEE Internet Computing, 13(5):14-22.

Voas, J., & Zhang, J., 2009, Cloud computing: New wine or just a new bottle? IEEE ITPro, 15-17.

Vouk, M. A., 2008, Cloud computing—Issues, research and implementations. Proceedings of the ITI 30th International Conference on Information Technology Interfaces, Cavtat, Croatia, 31-40.

Wang, L., Tao, J., and Kunze, M., 2008, Scientific cloud computing: Early definition and experience. Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications, Austin, TX, 825-830.

Wikipedia, https://en.wikipedia.org/wiki/Cloud_computing, last accessed on May, 1, 2018.

Zhang, Q., Cheng, L., & Raouf Boutaba, 2010, Cloud computing:state-of-the-art and research challenges, J. Int. Serv. Appl., 2010(1):7-18.