An Overview of Cloud Computing Technology


Abstract—Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (eg networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. It has become a significant technology trend, and many experts expect that cloud computing will reshape information technology (IT) processes and the IT marketplace.

With the cloud computing technology, users use a variety of devices, including PCs, laptops, smart phones, and PDAs to access programs, storage, and application-development platforms over the Internet, via services offered by cloud computing providers. This paper presents an overview of cloud computing technology—deployment models, classes and characteristics.

Index Terms—Cloud computing, Infrastructure as a service, Platform as a service, Software as a service.

I. INTRODUCTION

Cloud Computing [1][2] is the term given to the use of multiple server computers via a digital network as if they were one computer. The ‘Cloud’ itself is a virtualization of resources—networks, servers, applications, data storage and services—which the end user has on-demand access to. These resources can be provided with minimal management or service provider interaction.

Cloud computing offers the end user resources without the requirement of having knowledge of the systems that deliver it. Additionally, the cloud can provide the user with a far greater range of applications and services. Therefore the cloud enables users and business scalable and tailored services.

Cloud Computing brings with it many benefits to the end user. These include:

- Access to a huge range of applications without having to download or install anything.
- Applications can be accessed from any computer, anywhere in the world.
- Users can avoid expenditure on hardware and software; only using what they need.
- Companies can share resources in one place.
- Consumption is billed as a utility with minimal upfront costs.
- Scalability via on-demand resources.

This paper presents an overview of the cloud computing technology. This paper is organized as follows: Section 2 presents deployment models of cloud computing. Section 3 presents different classes of cloud computing services. Section 4 presents cloud computing characteristics. And finally section 5 presents the conclusion.

II. CLOUD COMPUTING-DEPLOYMENT MODELS

Cloud computing is the next stage in evolution of the Internet. The cloud in cloud computing provides the means through which everything—from computing power to computing infrastructure, applications, business processes to personal collaboration—can be delivered to you as a service wherever and whenever you need. Cloud computing is offered in different forms[3][4]:

- Public clouds
- Private clouds
- Hybrid clouds
- Community clouds

A. Public clouds

Public cloud computing services are provided off-premise by third-party providers to the general public and the computing resources are shared with the provider’s other customers. This is pure cloud computing and there is no debate on this one.

B. Private clouds

Many large organizations prefer, or are legally obligated, to keep their servers, software and data within their own data centres; and private clouds enable them to achieve some of the efficiencies of cloud computing while taking responsibility for the security of their own data. By implementing cloud computing technologies behind their firewall, enterprises can enable pooling and sharing of computing resources across different applications, departments or business units. Unlike the pay-as-you-go model of public clouds, however, private clouds require significant up-front development costs, data centre costs, ongoing maintenance, hardware, software and internal expertise.

C. Hybrid clouds

Many enterprises take the ‘hybrid cloud’ approach by using public clouds for general computing while customer data is kept within a private cloud, community cloud or a more
An Overview Of Cloud Computing Technology

traditional IT infrastructure. The use of ‘virtual private cloud’ technology enables enterprises to connect their existing infrastructure to a set of isolated computing resources in public cloud infrastructure and to extend their existing internal IT management capabilities – such as security services, firewalls, and intrusion detection systems – to include their external virtual resources.

D. Community clouds

Community clouds are used by distinct groups (or ‘communities’) of organizations that have shared concerns such as compliance or security considerations, and the computing infrastructures may be provided by internal or third-party suppliers. The communities benefit from public cloud capabilities but they also know who their neighbors are so they have fewer fears about security and data protection.

III. CLASSES OF CLOUD COMPUTING SERVICES

Cloud computing services are divided into three classes, according to the abstraction level of the capability provided and the service model of providers, namely:

- Infrastructure as a Service (IaaS)
- Platform as a Service (PaaS)
- Software as a Service (SaaS)

Figure 1 depicts the layered organization of the cloud stack from physical infrastructure to applications.

![Cloud Stack Diagram]

**Figure 1: The cloud computing stack**

A. Infrastructure as a Service (IaaS)

Infrastructure as a Service (IaaS) is the delivery of computer hardware (servers, networking technology, storage, and data center space) as a service. It may also include the delivery of operating systems and virtualization technology to manage the resources. The IaaS customer rents computing resources instead of buying and installing them in their own data center. The service is typically paid for on a usage basis. The service may include dynamic scaling so that if the customer winds up needing more resources than expected, he can get them immediately (probably up to a given limit).

Dynamic scaling as applied to infrastructure means that the infrastructure can be automatically scaled up or down, based on the requirements of the application. Additionally, the arrangement involves an agreed-upon service level. The service level states what the provider has agreed to deliver in terms of availability and response to demand. It might, for example, specify that the resources will be available 99.999 percent of the time and that more resources will be provided dynamically if greater than 80 percent of any given resource is being used.

Currently, the most high-profile IaaS operation is Amazon’s Elastic Compute Cloud (Amazon EC2). It provides a Web interface that allows customers to access virtual machines. EC2 offers scalability under the user’s control with the user paying for resources by the hour. The use of the term elastic in the naming of Amazon’s EC2 is significant. The elasticity refers to the ability that EC2 users have to easily increase or decrease the infrastructure resources assigned to meet their needs. The user needs to initiate a request, so this service provided isn’t dynamically scalable. Users of EC2 can request the use of any operating system as long as the developer does all the work. Amazon itself supports a more limited number of operating systems (Linux, Solaris, and Windows).

B. Platform as a Service (PaaS)

With Platform as a Service (PaaS), the provider delivers more than infrastructure. It delivers what you might call a solution stack — an integrated set of software that provides everything a developer needs to build an application — for both software development and runtime. PaaS can be viewed as an evolution of Web hosting. In recent years, Webhosting companies have provided fairly complete software stacks for developing Web sites. PaaS takes this idea a step farther by providing lifecycle management — capabilities to manage all software development stages from planning and design, to building and deployment, to testing and maintenance.

The primary benefit of PaaS is having software development and deployment capability based entirely in the cloud — hence, no management or maintenance efforts are required for the infrastructure. Every aspect of software development, from the design stage onward (including source-code management, testing, and deployment) lives in the cloud.

PaaS is inherently multi-tenant and naturally supports the whole set of Web services standards and is usually delivered with dynamic scaling. In reference to Platform as a Service, dynamic scaling means that the software can be automatically scaled up or down. Platform as a Service typically addresses the need to scale as well as the need to separate concerns of access and data security for its customers. Some examples of Platform as a Service include the Google App Engine, AppJet, Eteles, Qrimp, and Force.com, which is the official development environment for Salesforce.com.

C. Software as a Service (SaaS)

One of the first implementations of cloud services was Software as a Service (SaaS) — business applications that are hosted by the provider and delivered as a service. SaaS has its roots in an early kind of hosting operation carried out by Application Service Providers (ASPs). The ASP business grew up soon after the Internet began to mushroom, with some companies offering to securely, privately host applications. Hosting of supply chain applications and customer relationship management (CRM) applications was particularly prominent, although some ASPs simply
specialized in running email. Prior to the advent of this type of service, companies often spent huge amounts of money implementing and customizing these applications to satisfy internal business requirements. Many of these products weren’t only difficult to implement but hard to learn and use. However, the most successful vendors were those who recognized that an application delivered as a service with a monthly fee based on the number of users had to be easy to use and easy to stay with. CRM is one of the most common categories of Software as a Service; the most prominent vendor in this category is Salesforce.com

Buying Software as a Service offers a number of obvious advantages: the following provides some insight into why this approach to software delivery has gained so much traction with vendors and customers. The price of the software is on a per-use basis and involves no upfront costs from the service provider. (Of course, the reality is that your company may have some upfront work to do to get your data loaded into the Software as a Service application database and you may have to deal with ongoing data integration between your internal and cloud data stores.) Businesses get the immediate benefit of reducing capital expenditures. In addition, a business gains the flexibility to test new software on a rental basis and then can continue to use and adopt the software, if it proves suitable.

IV. CLOUD COMPUTING CHARACTERISTICS

Certain characteristics of a cloud are essential to enable services that truly represent the cloud computing model and satisfy expectations of consumers, and cloud offerings must be

- Self-service
- Per-usage metered and billed
- Elastic
- Customizable

A. Self-service

Consumers of cloud computing services expect on-demand, nearly instant access to resources. To support this expectation, clouds must allow self-service access so that customers can request, customize, pay, and use services without intervention of human operators.

B. Per-usage metered and billed

Cloud computing eliminates up-front commitment by users, allowing them to request and use only the necessary amount. Services must be priced on a short term basis (e.g., by the hour), allowing users to release (and not pay for) resources as soon as they are not needed. For these reasons, clouds must implement features to allow efficient trading of service such as pricing, accounting, and billing. Metering should be done accordingly for different types of service (e.g., storage, processing, and bandwidth) and usage promptly reported, thus providing greater transparency.

C. Elastic

Cloud computing gives the illusion of infinite computing resources available on demand. Therefore users expect clouds to rapidly provide resources in any quantity at any time. In particular, it is expected that the additional resources can be (a) provisioned, possibly automatically, when an application load increases and (b) released when load decreases (scale up and down).

D. Customizable

In a multi-tenant cloud a great disparity between user needs is often the case. Thus, resources rented from the cloud must be highly customizable. In the case of infrastructure services, customization means allowing users to deploy specialized virtual appliances and to be given privileged (root) access to the virtual servers. Other service classes (PaaS and SaaS) offer less flexibility and are not suitable for general-purpose computing, but still are expected to provide a certain level of customization.

V. CONCLUSION

Cloud computing refers to the delivery of computing and storage capacity as a service to a heterogeneous community of end- recipients. The cloud computing technology provides four deployment models: public cloud, private cloud, hybrid cloud and community cloud; Three service models: Infrastructure as a Service(IaaS), Platform as a Service(PaaS) and Software as a Service(SaaS); Four characteristics: self-service, per-usage metered and billed, elastic, and customizable.

REFERENCES