Towards Cloud Computing Anatomy

Maria Salama and Mohamed Kouta

Abstract—Cloud Computing has recently emerged as a compelling paradigm for managing and delivering services over the internet. The rise of Cloud Computing is rapidly changing the landscape of information technology, and ultimately turning the long-held promise of utility computing into a reality. As the development of Cloud Computing paradigm is speedily progressing, concepts, and terminologies are becoming imprecise and ambiguous, as well as different technologies are interfering. Thus, it becomes crucial to clarify the key concepts and definitions. In this paper, we present the anatomy of Cloud Computing, covering its essential concepts, prominent characteristics, its affects, architectural design and key technologies. We differentiate various service and deployment models. Also, significant challenges and risks need are tackled in order to guarantee the long-term success of Cloud Computing. The aim of this paper is to provide a better understanding of the anatomy of Cloud Computing and pave the way for further research in this area.

Keywords—Anatomy, Aspects, Cloud Computing, Challenges.

I. INTRODUCTION

Cloud Computing has recently emerged as a compelling paradigm for managing and delivering services over the internet. The rise of Cloud Computing is rapidly changing the landscape of information technology, and ultimately turning the long-held promise of utility computing into a reality. Cloud computing is attractive to business owners as it eliminates the requirement for users to plan ahead for provisioning, and allows enterprises to start from the small and increase resources only when there is a rise in service demand.

As the development of Cloud Computing paradigm is speedily progressing, concepts, and terminologies are becoming imprecise and ambiguous, as well as different technologies are interfering. The hype around Cloud Computing further muddies the message. The overall confusion about the paradigm and its capacities turns the Cloud into an excessively general term that includes almost any solution that allows the out-sourcing of all kinds of hosting and computing resources. There are many definitions of Cloud Computing, but they all seem to focus on just certain aspects of the technology [1] – [6]. Moreover, the variety of technologies [1] in the Cloud, as well as the different notions of service models [7], [8], makes the over-all picture confusing. Thus, it becomes crucial to clarify the key concepts and definitions.

In this paper, we present the anatomy of Cloud Computing. We cover the essential definitions of Cloud Computing from different perspectives, prominent characteristics, and its organizational affects. The architectural design and key technologies on which clouds are underlying is also discussed to provide a better understanding of the clouds nature. More importantly, service and deployment models of clouds are clearly differentiated, to eradicate the concepts ambiguity. Also, significant challenges and risks need are tackled in order to guarantee the long-term success of Cloud Computing. The aim of this paper is to provide a better understanding of the anatomy of Cloud Computing and pave the way for further research in this area.

The rest of this paper is organized as follows. Section 2, introduces an overview about Cloud Computing, including definitions, characteristics and affects. In section 3, we differentiate different service and deployment models. In Section 4, we present the architecture of clouds and underlying technologies. Finally, we reflect the challenges of adopting Cloud Computing and our conclusion in Sections 5 and 6 respectively.

II. CLOUD COMPUTING OVERVIEW

This section presents a general overview of cloud computing, including its definitions, characteristics and organizational affects.

A. Defining Cloud Computing

In this section we gather integrative definitions of the Cloud proposed by many experts from different perspectives.

Cloud computing has been coined as an umbrella term to describe a category of sophisticated on-demand computing services initially offered by commercial providers. It denotes a model on which a computing infrastructure is viewed as a “cloud,” from which businesses and individuals access applications from anywhere in the world on demand [9]. The main principle behind this model is offering computing, storage, and software “as a service”. There are many definitions of Cloud Computing, but they all seem to focus on just certain aspects of the technology [10].

An early definition of Cloud computing has been proposed as follows: A computing Cloud is a set of network enabled services, providing scalable, QoS guaranteed, normally personalized, inexpensive computing platforms on demand, which could be accessed in a simple and pervasive way [1]. Markus Klems claims that immediate scalability and resources usage optimization are key elements for the Cloud. These are provided by increased monitoring, and automation of resources management in a dynamic environment.
Other authors disagree that this is a requirement for an infrastructure to be considered as a Cloud [2].

A Berkeley Report in February 2009 states “Cloud computing, the long-held dream of computing as a utility, has the potential to transform a large part of the IT industry, making software even more attractive as a service” [3].

From an economic perspective, Cloud Computing is defined as follows: “Building on compute and storage virtualization technologies, and leveraging the modern Web, Cloud Computing provides scalable and affordable compute utilities as on-demand services with variable pricing schemes, enabling a new consumer mass market” [4].

From the Quality of Service perspective, Clouds have been defined as a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services). These resources can be dynamically reconfigured to optimum resource utilization. This pool of resources is typically exploited by a pay-per-user model in which guarantees are offered by the Infrastructure Provider by means of customized SLAs” [5].

The US National Institute of Standards and Technology (NIST) has published a working definition [6] that seems to have captured the commonly agreed aspects of cloud computing.

This definition describes cloud computing using:
- Five characteristics: on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service.
- Four deployment models: private clouds, community clouds, public clouds, and hybrid clouds.
- Three service models: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS).

An encompassing definition of the Cloud taking into account cloud features has been proposed as follows. Clouds are a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services). These resources can be dynamically reconfigured to adjust to a variable load (scale), allowing also for an optimum resource utilization. This pool of resources is typically exploited by a pay-per-use model in which guarantees are offered by the Infrastructure Provider by means of customized SLAs. On the other hand, looking for the minimum common denominator would lead us to no definition as no single feature is proposed by all definitions. The set of features that most closely resemble this minimum definition would be scalability, pay-per-use utility model and virtualization [11].

B. Cloud Characteristics

The features of Cloud Computing are that it offers enormous amounts of power in terms of computing and storage while offering improved scalability and elasticity. Moreover, with efficiency and economics of scale, cloud computing services are becoming not only a cheaper solution but a much greener one to build and deploy IT services [8].

The Cloud computing distinguishes itself from other computing paradigms in the following aspects [1]:

- **On-demand service provisioning:** Computing Clouds provide resources and services for users on demand. Users can customize and personalize their computing environments later on, for example, software installation, network configuration, as users usually own administrative privileges.
- **QoS guaranteed offer:** The computing environments provided by computing Clouds can guarantee QoS for users, e.g., hardware performance like CPU speed, I/O bandwidth and memory size. The computing Cloud renders QoS in general by processing Service Level Agreement (SLA) with users.
- **Autonomous System:** The computing Cloud is an autonomous system and it is managed transparently to users. Hardware, software and data inside clouds can be automatically reconfigured, orchestrated and consolidated to present a single platform image, finally rendered to users.
- **Scalability and flexibility:** The scalability and flexibility are the most important features that drive the emergence of the Cloud computing. Cloud services and computing platforms offered by computing Clouds could be scaled across various concerns, such as geographical locations, hardware performance, and software configurations. The computing platforms should be flexible to adapt to various requirements of a potentially large number of users.

C. Affects

Recently, there has been a great deal of hype about cloud computing. Cloud computing is on the top of Gartner’s list of the ten most disruptive technologies of the next years. Since Cloud Computing makes several promises, in terms of time, management and operational costs, enterprises need to understand the affects of cloud computing which are focused on the following specific topics [12]:

- **The organizational change brought about with cloud computing:** The type of organizational change that cloud computing results in can be demonstrated by taking a look at, for example, IT procurement within enterprise. Simplistically, procurement is based on obtaining estimates for things, then getting those estimates signed-off by management to allow the procurement to proceed. Capital and operational budgets are kept separate in this process, and it can take several months between the decision to procure hardware and the hardware being delivered, setup and ready to use. The use of cloud computing can greatly reduce this time period, but there is more significant change related to the empowerment of users and the diffusion of the IT department’s authority. For example, a company’s training coordinator who requires a few servers to run a week-long web-based training course can bypass their IT department and run the training course in the cloud. They could pay their cloud usage-bill using their personal credit card and charge back the amount as expenses to their employee.
- **The economic and organizational implications of the utility billing model in cloud computing:** New cloud computing pricing models based on market mechanisms are starting to emerge but it is not yet clear how such models can be effectively used by enterprise. An example of such models is used by Amazon’s Spot Instances, which allows users to bid for unused capacity in Amazon’s data centers. Amazon
runs the user’s instances as long as the bid price is higher than the spot price, which is set by Amazon based on their data center utilization.

- The security, legal and privacy issues that cloud computing raises: Security, legal and privacy issues are widely acknowledged as being important in cloud computing. Most of the security and privacy issues in cloud computing are caused by users’ lack of control over the physical infrastructure. This leads to legal issues that are affected by a cloud’s physical location, which determines its jurisdiction. Furthermore, multi-tenancy brings the need for new solutions towards security and privacy. For example, denial of service (DoS) attacks were a common concern even before cloud computing became popular, but when an application is targeted by a DoS attack in the cloud, the user or owner could actually end-up paying for the attack through their increased resource usage. This could be significantly higher than the peak usage of that application in an in-house data-center with limited resources.

III. SERVICE AND DEPLOYMENT MODELS

There are diverse dimensions to classify cloud computing, two commonly used categories are: service boundary (deployment models) and service type (service models).

- From the service boundary’s view, cloud computing can be classified as public cloud, private cloud and hybrid cloud. The public cloud refers to services provided to external parties. The enterprises build and operate private cloud for themselves. Hybrid cloud shares resources between public cloud and private cloud by a secure network. Virtual Private Cloud (VPC) services released by Google and Amazon are examples of Hybrid cloud.

- From the service type’s view, cloud computing can be classified as Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). SaaS providers to end users, while IaaS and PaaS provide services to ISV and developers - leaving a margin for 3rd-party application developers.

A. Deployment Models

The cloud model promotes four deployment models.

- Private cloud: The cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise.

- Community cloud: The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be managed by the organizations or a third party and may exist on premise or off premise.

- Public cloud: The cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.

- Hybrid cloud: The cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load-balancing between clouds).

- Infrastructure-as-a-Service (IaaS) or Hardware-as-a-Service (HaaS) solutions deliver IT infrastructure based on virtual or physical resources as a commodity to customers. These resources meet the end user requirements in terms of memory, CPU type and power, storage, and, in most of the cases, operating system as well. Users are billed on a pay-per-use basis. They have to set up their applications on top of these resources that are hosted and managed in data centers owned by the vendor. Amazon is one of the major players in providing IaaS solutions. Amazon Elastic Compute Cloud (EC2) provides a large computing infrastructure and a service based on hardware virtualization. By using Amazon Web Services, users can create Amazon Machine Images (AMIs) and save them as templates from which multiple instances can be run. It is possible to run either Windows or Linux virtual machines, for which the user is charged per hour for each of the instances running. Amazon also provides storage services with the Amazon Simple Storage Service (S3), users can use Amazon S3 to host large amount of data accessible from anywhere [8].

- Platform-as-a-Service solutions provide an application or development platform in which users can create their own application that will run on the Cloud. More precisely, they provide an application framework and a set of API that can be used by developers to program or compose applications for the Cloud. PaaS solutions often integrate an IT infrastructure on top of which applications will be executed. This is the case of

| TABLE I | SERVICE MODELS |
|---|---|---|
| Category | Characteristics | Product Type |
| SaaS | Customers are provided with applications that are accessible anytime and from anywhere. | Web applications and services (Web 2.0) | SalesForce.com (CRM) Clarizen.com (Project Management) Google Documents, Google Mail (Automation) |
| PaaS | Customers are provided with a platform for developing applications hosted in the Cloud. | Programming APIs and frameworks; Deployment system. | Google AppEngine Microsoft Azure Manjrasoft Aneka |
| IaaS | Customers are provided with virtualized hardware and storage on top of which they can build their infrastructure. | Virtual machines management infrastructure, Storage management | Amazon EC2 and S3; GoGrid, Nirvanix |
Google AppEngine and Microsoft Azure, while other solutions, such as Manjrasoft Aneka, are purely PaaS implementations. Google AppEngine is a platform for developing scalable web applications that run on top of data centers maintained by Google. It defines an application model and provides a set of APIs that allow developers to take advantage of additional services such as Mail, Datastore, Memcache, and others. AppEngine manages the execution of applications and automatically scales them up/down as required. Google provides a free but limited service, while utilizes daily and per minute quotas to meter and price applications requiring a professional service. Azure is a cloud service operating system that serves as the development, runtime, and control environment for the Azure Services Platform. By using the Microsoft Azure SDK, developers can create services that leverage the .NET Framework. These services have to be uploaded through the Microsoft Azure portal in order to be executed on top of Windows Azure. Additional services, such as workflow execution and management, web services orchestration, and access to SQL data stores, are provided to build enterprise applications. Aneka, commercialized by Manjrasoft, is a pure PaaS implementation and provides end users and developers with a platform for developing distributed applications for the Cloud by using .NET technology. The core value of Aneka is a service oriented runtime environment that is deployed on both physical and virtual infrastructures and allows the execution of applications developed by means of various programming models [8].

- **Software-as-a-Service** solutions are at the top end of the Cloud computing stack and they provide end users with an integrated service comprising hardware, development platforms, and applications. Users are not allowed to customize the service but get access to a specific application hosted in the Cloud. Examples of SaaS implementations are the services provided by Google for office automation, such as Google Mail, Google Documents, and Google Calendar, which are delivered for free to the Internet users and charged per user for professional quality services. Examples of commercial solutions are SalesForce.com and Clarizen.com, which provide online CRM (Customer Relationship Management) and project management services, respectively [8].

IV. ARCHITECTURE AND UNDERLYING TECHNOLOGIES

This section describes the architectural model and the key technologies currently used for Cloud Computing.

A. Architecture

Many organizations and researchers have defined the architecture for Cloud Computing. Basically the whole system can be divided into the core stack and the management. In the core stack, there are three layers: (1) Resource (2) Platform and (3) Application.

The resource layer is the infrastructure layer which is composed of physical and virtualized computing, storage and networking resources.

The platform layer is the most complex part which could be divided into many sub-layers; e.g. a computing framework manages the transaction dispatching and/or task scheduling. A storage sub-layer provides unlimited storage and caching capability.

The application server and other components support the same general application logic as before with either on-demand capability or flexible management, such that no components will be the bottle neck of the whole system [13].

Based on the underlying resource and components, the application could support large and distributed transactions and management of huge volume of data. All the layers provide external service through web service or other open interfaces. Cloud Architecture is depicted in Figure 1.

![Cloud Architecture](image)

**B. Technologies behind Cloud Computing**

A number of enabling technologies contribute to Cloud computing, several state-of-the-art techniques are identified [1]:

- **Virtualization technology**: Virtualization technologies partition hardware and thus provide flexible and scalable computing platforms. Virtual machine technologies, such as VMware and Xen, offer virtualized IT infrastructures on demand. Virtual network advances, such as VPN, support users with a customized network environment to access Cloud resources. Virtualization techniques are the bases of the Cloud computing since they render flexible and scalable hardware services.

- **Orchestration of service flow and workflow**: Computing Clouds offer a complete set of service templates on demand, which could be composed by services inside the computing Cloud. Computing Clouds therefore should be able to automatically orchestrate services from different sources and of different types to form a service flow or a workflow transparently and dynamically for users.

- **Web service and SOA**: Computing Cloud services are normally exposed as Web services, which follow the industry standards such as WSDL, SOAP and UDDI. The services organization and orchestration inside Clouds could be managed in a Service Oriented Architecture (SOA). A set of Cloud services furthermore could be used in a SOA application environment, thus making them available on various distributed platforms and could be further accessed across the Internet.

- **Web 2.0**: Web 2.0 is an emerging technology describing the innovative trends of using World Wide Web technology and Web design that aims to enhance creativity, information
sharing, collaboration and functionality of the Web. Web 2.0 applications typically include some of the following features/techniques:
- CSS to separate presentation and content
- Folksonomies (collaborative tagging, social classification, indexing & social tagging).
- SemanticWeb technologies
- REST, XML and JSON-based APIs.
- Innovative Web development techniques such as Ajax.
- XHTML and HTML markup.
- Syndication, aggregation and notification of Web data with RSS or Atom feeds.
- Mashups, merging content from different sources, client- and server-side.
- Weblog publishing tools.
- Wiki to support user-generated content.
- Tools to manage users’ privacy on the Internet.

The essential idea behind Web 2.0 is to improve the interconnectivity and interactivity of Web applications. The natural technical evolution that the Cloud computing adopts is Web applications in nature are Web applications which render desirable computing services on demand. It is thus a network storage system, which is backed by distributed storage providers (e.g., data centers), offers storage capacity for users to lease. The data storage could be migrated, merged, and managed transparently to end users for whatever data formats. Examples are Google File System and Amazon S3. A Mashup is a Web application that combines data from more than one source into a single integrated storage tool. The SmugMug is an example of Mashup, which is a digital photo sharing Web site, allowing the upload of an unlimited number of photos for all account types, providing a published API which allows programmers to create new functionality, and supporting XML-based RSS and Atom feeds.

World-wide distributed storage system: A Cloud storage model should foresee:
- A network storage system, which is backed by distributed storage providers (e.g., data centers), offers storage capacity for users to lease. The data storage could be migrated, merged, and managed transparently to end users for whatever data formats. Examples are Google File System and Amazon S3. A Mashup is a Web application that combines data from more than one source into a single integrated storage tool. The SmugMug is an example of Mashup, which is a digital photo sharing Web site, allowing the upload of an unlimited number of photos for all account types, providing a published API which allows programmers to create new functionality, and supporting XML-based RSS and Atom feeds.
- A distributed data system which provides data sources accessed in a semantic way. Users could locate data sources in a large distributed environment by the logical name instead of physical locations. Virtual Data System (VDS) is good reference.

Programming model: Users drive into the computing Cloud with data and applications. Some Cloud programming models should be proposed for users to adapt to the Cloud infrastructure. For the simplicity and easy access of Cloud services, the Cloud programming model, however, should not be too complex or too innovative for end users. The MapReduce is a programming model and an associated implementation for processing and generating large data sets across the Google worldwide infrastructures. The MapReduce model firstly involves applying a map operation to some data records – a set of key/value pairs, and then processes a reduce operation to all the values that shared the same key. The MapReduce-Merge method evolves the MapReduce paradigm by adding a merge operation. Hadoop is a framework for running applications on large clusters built of commodity hardware. It implements the MapReduce paradigm and provides a distributed file system – the Hadoop Distributed File System. The MapReduce and the Hadoop are adopted by recently created international Cloud computing project of Yahoo!, Intel and HP.

V. CHALLENGES

Despite the initial success and popularity of the cloud computing paradigm and the extensive availability of providers and tools, a significant number of challenges and risks are inherent to this new model of computing. Providers, developers, and end users must consider these challenges and risks to take good advantage of cloud computing. Issues to be faced include user privacy, data security, data lock-in, availability, disaster recovery, performance, scalability, energy-efficiency, and programmability.

- Security, Privacy, and Trust. Security and privacy affect the entire cloud computing stack, since there is a massive use of third-party services and infrastructures that are used to host important data or to perform critical operations. In this scenario, the trust toward providers is fundamental to ensure the desired level of privacy for applications hosted in the cloud [8]. Legal and regulatory issues also need attention. When data are moved into the Cloud, providers may choose to locate them anywhere on the planet. The physical location of data centers determines the set of laws that can be applied to the management of data. For example, specific cryptography techniques could not be used because they are not allowed in some countries. Similarly, country laws can impose that sensitive data, such as patient health records, are to be stored within national borders [10].

- Data Lock-In and Standardization: A major concern of cloud computing users is about having their data locked-in by a certain provider. Users may want to move data and applications out from a provider that does not meet their requirements. However, in their current form, cloud computing infrastructures and platforms do not employ standard methods of storing user data and applications. Consequently, they do not interoperate and user data are not portable. The answer to this concern is standardization.

- Availability, Fault-Tolerance, and Disaster Recovery: It is expected that users will have certain expectations about the service level to be provided once their applications are moved to the cloud. These expectations include availability of the service, its overall performance, and what measures are to be taken when something goes wrong in the system or its components. Generally, users seek for a warranty before they can comfortably move their business to the cloud, that is Service Level Agreement (SLA). SLAs, which include QoS requirements, must be ideally set up between customers and cloud computing providers to act as warranty.

- Resource Management and Energy-Efficiency: One important challenge faced by providers of cloud computing services is the efficient management of virtualized resource pools. Physical resources such as CPU cores, disk space, and network bandwidth must be sliced and shared among virtual machines running potentially heterogeneous workloads [14]. In addition to optimize application performance, dynamic resource management can also improve utilization and...
consequently minimize energy consumption in data centers. This can be done by judiciously consolidating workload onto smaller number of servers and turning off idle resources.

VI. CONCLUSION

In this paper, we have presented the anatomy of Cloud Computing, covering its essential concepts, prominent characteristics, its affects, architectural designs and key technologies. Also, service and deployment models of clouds were clearly differentiated. Although, a great popularity and apparent success have been visible in this area, significant challenges and risks need to be tackled by industry and academia in order to guarantee the long-term success of Cloud Computing. Visible trends in this sphere include the emergence of standards; the creation of value-added services by augmenting, combining, and brokering existing compute, storage, and software services; and the availability of more providers in all levels, thus increasing competitiveness and innovation. As the development of Cloud Computing paradigm is speedily progressing, we hope our work will provide a better understanding of the anatomy of Cloud Computing, and pave the way for further research in this area.

REFERENCES


