Journal of Biosphere, 6: 21-26, 2017 ISSN 2278 – 3342

Research challenges in cloud computing technology

Vinay Arora and Monika Chopra

Abstract: In today's modern era Cloud Computing has been developed as a new standard for hosting and delivering services through the Internet. "Cloud Computing" has taken the attention of business owners as they don't require to plan in advance for provisioning, and can start with the small budgets and have to raise the resources only when there is demand of the service. However, in spite of huge opportunities offered by Cloud Computing to the IT industry, its development is still in the initial stages along with many challenges In this paper, we are going to provide the general view of Cloud Computing, its key points, architectural principles, and high-tech performance as well as research challenges. The main objective of this paper is to give a enhanced perceptive of the challenges included in Cloud Computing and discover significant research guidelines to surface the way for further exploration in this area.

Keywords: Cloud computing; cloud security; SaaS; Paas; IaaS; research challenges.

1. Introduction

Cloud Computing has achieved a fame and grown as a major development in IT. In spite of having many advantages of cloud computing, the organizations are not confident while accepting Cloud Computing services because of security issues and challenges related with organization of this technology [1]. While industry has been pushing the Cloud research agenda at high pace, academia has only recently joined, as can be seen through the sharp rise in workshops and conferences focusing on Cloud Computing. Lately, these have brought out many peer-reviewed papers on aspects of cloud computing, and made a systematic review necessary, which analyses the research done and explains the resulting research agenda. We performed such a systematic review of all peer- reviewed academic research on cloud computing, and explain the technical challenges facing in this paper [2].

Vinay Arora
Department of Mathematics,
UIET (PUSSGRC) Hoshiarpur,
Email: vinay2037@gmail.com

Monika Chopra (✉)
Department of Computer Science,
D.A.V. College Jalandhar.
Email: monicachopra@davjalandhar.com

With the rapid development of processing and storage technologies and the success of the Internet, computing resources have become cheaper, more powerful and more ubiquitously available than ever before. This technological trend has enabled the realization of a new computing model called cloud computing, in which resources (e.g., CPU and storage) are provided as general utilities that can be leased and released by users through the Internet in an on-demand fashion. In a cloud computing environment, the traditional role of service provider is divided into two: the infrastructure providers who manage cloud platforms and lease resources according to a usage-based pricing model, and service providers, who rent resources from one or many infrastructure providers to serve the end users. The emergence of cloud computing has made a tremendous impact on the Information Technology (IT) industry over the past few years, where large companies such as Google, Amazon and Microsoft strive to provide more powerful, reliable and cost-efficient cloud platforms, and business enterprises seek to reshape their business models to gain benefit from this new paradigm. Indeed, cloud computing provides several compelling features that make it attractive to business owners [3-9].

2. Overview of cloud computing

Cloud computing is the use of computing resources (hardware and software) that are delivered as a service over a network (typically the Internet). The name comes from the use of a cloud-shaped symbol as an abstraction for the complex infra-structure it contains in system

diagrams. Cloud computing entrusts remote services with a user's data, software and computation. Cloud computing is the use of computing resources (hardware and software) that are delivered as a service over a network (typically the Internet). The name comes from the use of a cloud-shaped symbol as an abstraction for the complex infrastructure it contains in system diagrams. Cloud computing entrusts remote services with a user's data, software and computation [10].

The main idea behind cloud computing is not a new one. John McCarthy in the 1960s already envisioned that computing facilities will be provided to the general public like a utility. The term "cloud" has also been used in various contexts such as describing large ATM networks in the 1990s. However, it was after Google's CEO Eric Schmidt used the word to describe the business model of providing services across the Internet in 2006, that the term really started to gain popularity. Since then, the term cloud computing has been used mainly as a marketing term in a variety of contexts to represent many different ideas. Certainly, the lack of a standard definition of cloud computing has generated not only market hypes, but also a fair amount of skepticism and confusion. For this reason, recently there has been work on standardizing the definition of cloud computing. [5]

a) NIST definition of cloud computing

Cloud computing is a model for enabling convenient on-demand network access to a shared pool of configurable computing resources networks(e.g. Servers, storage, applicationsand services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.[8]

The main reason for the existence of different perceptions of cloud computing is that cloud computing, unlike other technical terms, is not a new technology, but rather a new operations model that brings together a set of existing technologies to run business in a different way. Indeed, most of the technologies used by cloud computing, such as virtualization and utility-based pricing are not new. Instead, cloud computing has exploited these existing technologies to meet the technical and financial requirements of today's demand for IT.

The NIST definition is one of the clearest and most comprehensive definitions of cloud computing and is widely referenced in US government documents and projects. [3] [4] This definition describes cloud computing as having five essential characteristics, three service models, and four deployment models. The essential characteristics are:

- *On-demand self-service*: We can acquire Computing resources and use them when needed without interacting with cloud service providers. Processing power, storage, virtual machines etc are included in these computing resources.
- *Broad network access:* We can access the computing resources over a network with various devices like laptops or mobile phones.
- *Resource pooling:* Multiple users share that resources pooled by Cloud service providers. This is called *multi-tenancy* in which e.g. A physical server may host a number of virtual machines that belong to different users.
- *Rapid elasticity***:** By scaling out a user can rapidly obtain more resources from the cloud . No longer required resources can be released and scaled back in.
- *Measured service:* Using appropriate metrics usage of resource is calculated such recording storage usage, CPU hours, usage of bandwidth etc.

Each cloud provides the different level of abstraction to the users, which is referred to as a service model as per the NIST definition. The three most widespread service models [8] are:

- *Software as a Service (SaaS* [3] [4]*)***:** In this users simply use a web-browser to access the software developed by others and offered as a service on the web.[2] Sales force's Customer Relationship Management software are accepted examples that make use of SaaS model of "Cloud Computing."
- *Platform as a Service (PaaS)* [3] [4]*:* Appliactions developed in thismake use of programming languages and tools compatible with the PaaS provider.[2] There is high level of abstraction provided by PaaS that helps them to focus on the development of their applications without worrying about the underlying infrastructure. Google App Engine is popular examples of PaaS.

- *Infrastructure as a Service (IaaS)* [3] [4]*:* In this computing resources are acquired from an IaaS provider and used for deploying and running their applications. [2] In opposite to the PaaS model, the IaaS model is having a low level of abstraction which allows the users to access the primary infrastructure by using virtual machines. IaaS gives users more flexibility as compared to PaaS because it allows the user for deploying any software stack on top of the operating system. Amazon Web Services' EC2 are well-liked examples of IaaS.

Different types of clouds depending on ownership and uses are deployed in NIST definition and are referred to as a cloud deployment model and the four popular models are:

- *Private cloud:* A cloud being used exclusively by single organization and may be operated by the organization itself or a third party. The "St Andrews Cloud Computing Co-laboratory" and "Concur Technologies" are example of the organizations having private clouds.
- *Public cloud:* A cloud that can be used (being paid) by the general public. Public clouds need important investment and are normally owned by big corporations like Microsoft, Google or Amazon.
- *Community cloud:* This is a cloud shared by a number of organizations as per their specific requirements. "The Open Cirrus cloud test bed" could be regarded as a community cloud whose objective is to support research in "Cloud computing".
- *Hybrid cloud:* This cloud is a mixture of other three deployment models. In a hybrid cloud each cloud could be separately managed but applications and data would be allowed to move across the hybrid cloud. Cloud bursting is allowed in Hybrid clouds in which a private cloud can burst-out to a public cloud when more resources are required.

Figure 1 is an overview of the common deployment and service models in "Cloud Computing", where the three service models could be deployed on top of any of the four deployment models [6].

3. Cloud computing technologies

In this section we provide a review of technologies being used in cloud computing environments [11].

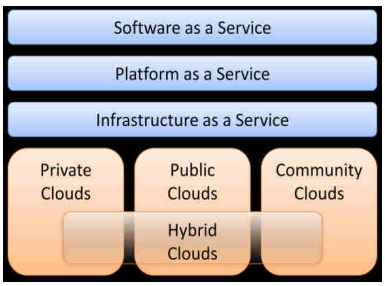

Fig. 1: Cloud computing deployment and service models.

a) Architectural design of data centers

A data center, a home to the computation power and storage, is core of "Cloud Computing" and contains a large number of devices like servers, switches and routers. Proper planning of this network architecture is decisive, as it will heavily rile applications performance and throughput in the distributed computing environment. Additionally, scalability and resiliency features are to be watchfully considered. Presently, a layered approach is the basic foundation of the network architecture design that has been tested in the biggest deployed data centers. The basic layers of a data center include the core, aggregation, and access layers, as shown in Fig. 2.

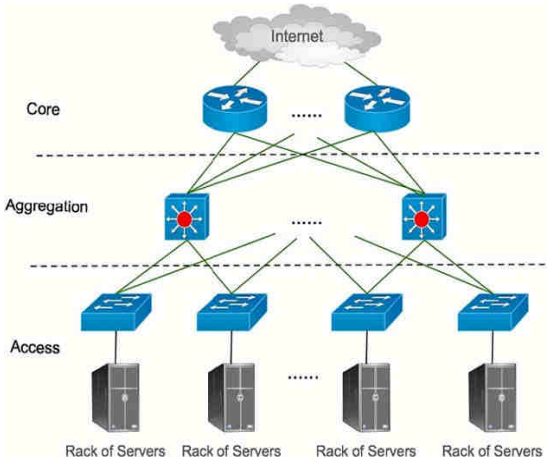

Fig. 2: Basic layered design of data center network infrastructure

In the access layer the servers in racks are physically connected to the network. There are normally 20 to 40 servers each rack and each connected to an access switch of 1 Gbps link. Access switches normally are connected to two aggregation switches for redundancy with links

of 10 Gbps. The aggregation layer generally provides significant functions, like domain service, location service, server load balancing, and many more. The central layer provides connectivity to many aggregation switches and provides a resilient routed fabric without single point of failure. The central routers handle traffic in and out of the data center. A famous practice is to control commodity Ethernet switches and routers to build the network infrastructure. The layered network infrastructure in different business solutions can be elaborated to meet specific business challenges [10].

Basically, the design of data center network architecture should meet the following objectives:

- *Uniform high capacity*: The available capacity on the network-interface cards of the sending and receiving servers should decide the maximum rate of a server-to-server traffic, and assigning the servers to a service should not be dependent on the network topology.
- *Free VM migration*: Virtualization permits the entire VM state for transmission across the network to move a VM from one physical machine to another. A cloud computing hosting service may move VMs for statistical multiplexing or dynamically changing communication patterns to attain high bandwidth for tightly coupled hosts or to attain variable heat distribution and power availability in the data center. We should design the communication topology to support rapid virtual machine migration.
- *Resiliency:* Failures will be common at scale. The network infrastructure must be fault-tolerant against various types of server failures, link outages, or server-rack failures. Existing unicast and multicast communications should not be affected to the extent allowed by the underlying physical connectivity.
- *Scalability*: The network infrastructure should be capable of scaling a large number of servers and allow for incremental growth.
- *Backward compatibility*: The network infrastructure should be backward compatible with switches and routers running Ethernet and IP. Because existing data centers have frequently leveraged commodity Ethernet and IP based devices that should also be used in the new architecture without much alteration. Highly interactive applications, that are sensitive to response time, are appropriate for geo-diverse MDC placed near the major populated areas.

b) Distributed file system over clouds

Google File System (GFS) is the proprietary distributed file system that is developed by Google and particularly designed to give efficient, reliable access to data with large clusters of commodity servers. Files are divided into the chunks of 64 megabytes, and are normally appended to or read and only extremely rarely overwritten or shrunk. Comparing with traditional file systems, GFS is designed and developed to run on data centers to generate very high data throughputs, low latency and tolerate individual server failures. Then there is the open source Hadoop Distributed File System (HDFS) that stores large files across a number of machines. It attains reliability by substituting the data across the number of servers. The file system is built from a group of data nodes, each serving blocks of data on the network using a block protocol exact to HDFS. Data is also available on HTTP that allows access to all contents through web browser or other types of clients. Data nodes can communicate to rebalance data distribution, to move copies around, and to provide the replication of data high.

c) Distributed application framework over clouds

HTTP-based applications generally obey the rules to some web application structure such as Java EE. In the modern data center environment, group of servers are used for calculation and data-intensive jobs like financial trend analysis, or for film animation. “MapReduce” is a software structure launched by Google to maintain distributed computing on the large data sets on groups of computers. “MapReduce” consists of one Master, to which client applications request the MapReduce jobs. The Master pushes work out to existing task nodes in the data center, determined to keep the tasks as close to the data as possible. The Master has the information of node containing the data, and other hosts that are nearby. If the task is not possible to be hosted on the node where the data is stored, precedence is given to nodes having same rack. This reduces the network traffic on

the main backbone that also helps to improve the throughput, as the backbone is usually the blockage. If a task fails or times out, it is rescheduled. If the Master gets failed, then all the ongoing tasks are lost. The open source Hadoop MapReduce project is inspired by Google's work. Currently, many organizations are using Hadoop MapReduce to run large data-intensive computations.

4. Research challenges

In spite of widely being adopted by the industry, the research on "Cloud Computing" is still at its infancy. Security is one of the major concerns that hold back the growth of "Cloud Computing" service model. The idea of possession of confidential data to third party is dangerous and the consumers need to be more thoughtful in understanding the risks of data breaches in this new environment [1]. In this section, we summarize some of the challenges in research of cloud computing.

1. Automated Service Provisioning

One key feature of "Cloud Computing" is the ability of acquisition and release of resources on-demand. The aim of a service provider is to allocate and de-allocate the resources from the cloud to fulfill its service level objectives (SLOs), with minimum operational cost. However, it is unclear how a service provider can attain this objective. Particularly it is difficult to determine the mapping of SLOs such as QoS requirements to low-level resource requirement such as CPU and storage requirements. Furthermore, to achieve high quickness and respond to rapid demand situations such as in ash crowd effect, the resource providing decision making must be made on-line.

Automated service provisioning is an old problem. Dynamic resource provisioning for the Internet applications has been studied broadly in the past. This approach typically involves: (1) Developing an application performance model which anticipates the number of application instances needed to handle demand at every particular level for satisfying QoS requirements; (2) Periodically anticipating the demand in future and finding resource requirements through the performance model; and (3) Automatic distribution of resources with predicted resource requirements. Application performance model can be constructed using various techniques, including Queuing theory, Control theory and Statistical Machine Learning.

2. Virtual Machine Migration

Virtualization can offer important benefits in "Cloud Computing" by enabling virtual machine migration to balance the load across the data center. Additionally virtual machine migration enables the strong and highly responsive provisioning in data centers. The major benefit of VM migration is to avoid hotspots; moreover the data should be transferred consistently and efficiently, with integrated consideration of resources for applications and physical servers.

3. Server consolidation

Server consolidation is an effective approach to maximize resource utilization with minimum energy consumption in a "Cloud Computing" environment. Live VM migration technology is generally used to secure VMs residing on a number of under-utilized servers over a single server, so as to set the remaining servers to an energy-saving state. Moreover, dependency in VMs, such as communication requirements, has also been considered in recent times.

However, server consolidation activities should not affect the performance of application. It is known that the usage of resource (also known as the footprint) of individual VMs may vary over time. Hence, it is sometimes important to observe the actuations of VM footprints and use this information for effective server consolidation. At last, the system must rapidly respond to resource congestions as they occur.

4. Traffic management and analysis

Analyzing data traffic is significant for today's data centers. For example, many web applications depend on the analysis of traffic data for optimization of customer's experiences. Network operators also need to be aware of traffic over the network to make many of the decisions regarding management and planning.

Though, there are various challenges for existing traffic measurement and analysis methods in Internet Service Providers (ISPs) networks and enterprise to widen to data centers. Initially, the thickness of links is much higher than that in ISPs or enterprise networks, which makes the worst case situation for existing methods. Secondly, most recent methods can calculate traffic matrices between a small numbers of end hosts, but even a modular data

center can have a large number of servers. Finally, existing methods usually assume somehow patterns that are reasonable in Internet and enterprises networks, but the applications deployed on data centers, such as MapReduce jobs, considerably change the traffic pattern. Additionally, there is tighter coupling in application's use of network, computing, and storage resources, than what is seen in other settings. Currently, there is not much work on measurement and analysis of data center traffic.

5. Data Security

Data security is another significant research topic in the area of "Cloud Computing." As service providers normally do not have access to the physical security system of the data centers, they must depend on the infrastructure provider to attain the full data security. Yet for a virtual private cloud, the service provider can only identify the security setting distantly, without knowing its full implementation. The infrastructure provider must attain the following objectives: (1) *confidentiality*, for the secure data access and migration, and (2) *audit ability*, for verifying whether there is tempering of security setting of applications or not. Confidentiality is generally attained by using cryptographic protocols, and audit ability can be attained using remote verification techniques. Remote verifications usually require a trustworthy platform module (TPM) to produce non-forgeable system summary (i.e. system state encrypted with the help of TPM's private key) as the proof of system security. In this case, it is dangerous to build trustworthy mechanisms at every architectural layer of the cloud. Initially, the hardware layer must be reliable using hardware TPM. Secondly, the virtualization platform should be reliable with secure virtual machine monitors. VM migration must only be permitted if both source and destination servers are reliable. New work has been dedicated to design the efficient protocols for establishment and management of trust.

5. Conclusions

"Cloud Computing" has been developed as a new standard for hosting and delivering services through the Internet. The growth of "Cloud Computing" is fast changing the scenery of information technology, and ultimately spinning the long-held promise of utility computing into a truth. In spite of the important benefits offered by "Cloud Computing", the latest technologies are not grown enough to identify its full potential. Many important challenges in this area, including automatic resource providing, power management and security management, have only started to attain attention from the research community. So, there is still big prospect for researchers to make ground breaking contributions in this field, and bring important effect to their growth in the industry.

References

1. Rajan, S. & Jairath, A. "Cloud Computing: The Fifth generation of Computing, International Conference on Communication Systems and Network Technologies (2011) 15(4) Publisher: Ieee, Pages: 665-667
2. Darko Shuleski et al. "Impact of Cloud Computing Technology Implementation in Public Sector", Proceedings of 10th International Conference "Challenges of Modern Management" Bucharest (Romania) Nov 2016.
3. Sean Carlin et al. "Cloud computing Technologies" International Journal of Cloud Computing and Services Science (IJ-CLOSER) Vol. 1, June 2012.
4. Mell, P. & Grance, T. (2011) "The NIST Definition of Cloud Computing (Draft)", Publisher: U.S. Department of Commerce.
5. Cloud Computing, en.wikipedia.org/wiki/ Cloudcomputing, Dec. 2009.
6. Cloud Hosting, CLoud Computing and Hybrid Infrastructure from GoGrid, http://www.gogrid.com.
7. S. Kumar et al, vManage: loosely coupled platform and virtualization management in data centers. In: Proc of international conference on cloud computing, 2009.
8. NIST Definition of Cloud Computing v15, csrc.nist.gov/groups/SNS/cloud-computing/cloud-def v15.doc.
9. Sun Microsystems, Introduction to Cloud Computing Architecture, 2009.
10. M. Al-Fares et al., A scalable, commodity data center network architecture. In: Proc SIGCOMM, 2008.
11. M. Armbrust et al., Above the clouds: a Berkeley view of cloud computing. UC Berkeley Technical Report, 2009.