An Architectural and Technological overview of Cloud Computing

Neha Kulkarni
Dept. of Computer Science & Engineering, Dept of Biomedical Engineering
S.G.S.I.T.S Indore

VinayManurkar
Dept. of Computer Science & Engineering, Dept of Biomedical Engineering
S.G.S.I.T.S Indore

Niharika Machhiwal
Dept. of Computer Science & Engineering, Dept of Biomedical Engineering
S.G.S.I.T.S Indore

Abstract: Cloud computing is Internet-based computing, whereby shared resources, software and information are provided to computers .The cloud technologies such as Map Reduce has created new trends in parallel programming The support for handing large data sets, the concept of moving computation to data. Cloud computing presents a unique opportunity for batch processing and analytics jobs that analyze terabytes of data and can take hours to finish. If there is enough data parallelism in the application, users can take advantage of the cloud's new "cost associatively": using hundreds of computers for a short time costs the same as using a few computer for a long time. Programming abstractions such as Google's Map Reduce and its open-source counterpart Hadoop allow programmers to express such tasks while hiding the operational complexity of choreographing parallel execution across hundreds of cloud computing servers.

I. INTRODUCTION

Cloud computing is the latest and perhaps the most dramatic trend in advanced computing paradigms since the introduction of commodity clusters, which have dominated HPC for more than a decade. Clouds offer an amorphous distributed environment of computing resources and services to a dynamic distributed user base. Like clusters, cloud computing exploits economies of scale to deliver advanced capabilities. Unlike clusters ,cloud resources are nonspecific and provide basic capabilities but guarantee neither identical propertiesn from run to run nor high availability of specialized system types .At present, the use of cloud computing in computation science is still limited, but the first step towards this goal have been already done. Last year, the Department of Energy (DOE) National Laboratories started exploring the use of cloud services for scientific computing. On April 2009, Yahoo Inc. announced that it has extended its partnership with the major top universities in United States of America to advance cloud computing research and applications to computational science and engineering.

A. *The definitions of Cloud Computing (CC)*

Cloud computing is Internet-based computing, whereby shared resources, software and information are provided to computers and other devices on-demand, like a public utility. A technical definition [2] is "a computing capability that provides an abstraction between the computing resource and its underlying technical architecture (e.g., servers, storage, networks), enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction." This definition states that clouds have five essential characteristics: on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service.

Probably the most popular definition in business environment, NIST states that CC “is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. Definitions by CC or IT vendors reflect their specific viewpoints rather than a common understanding.

Amazon's definition is close to NIST, since their business is on CC services and not on hardware or software products: "Cloud Architectures are designs of software applications that use Internet-accessible on-demand services. Applications built on Cloud Architectures are such that the underlying computing infrastructure is used only when it is needed (for example to process a user request), draw the necessary resources on-demand (like compute servers or storage), perform a specific job, then relinquish the unneeded resources and often dispose themselves after the job is done. While in operation the application scales up or down elastically based on resource needs" .

B. Cloud Technologies

The cloud technologies such as Map Reduce and Dryad have created new trends in parallel programming [3]. The support for handing large data sets, the concept of moving computation to data, and the better quality of services provided by the cloud technologies make them favorable choice of technologies to solve large scale data/compute intensive problems. Cloud technologies such as Google Map Reduce, Google File System (GFS), Hadoop and Hadoop Distributed File System (HDFS), Microsoft Dryad, and CGL-Map Reduce adopt a more data-centered approach to parallel runtimes[4][5]. In these frameworks, the data is staged in data/compute nodes of clusters or large-scale data centers, such as in the case of Google. The computations move to the data in order to perform the data processing. Distributing file systems such as GFS and HDFS allow Google Map Reduce and Hadoop to access data via distributed storage systems built on heterogeneous compute nodes, while Dryad and CGL-MapReduce support reading data from local disks. The simplicity in the programming model enables better support for quality of services such as fault tolerance and monitoring.

II. COMMON CHARACTERISTICS

On-demand self-service provisioning: Computing resources can be acquired and used at anytime and anywhere allowing customizing the computing environments without the need for a direct service provider intervention. The user with administrative privileges has complete control of the configuration parameters and can set the desired characteristics of computing power, storage and network.

Resource pooling: Cloud service Providers share their resources with multiple users but they are configured to appear as a unique platform for individual end users. This multi-tenancy model allows to dynamically assign to several users physical and virtual resources according to their request without providing specific information on their position.

Rapid elasticity and scalability: The scalability is one of the most important feature of CC and allows to acquire more resources during a peak of demand and release them once they are no longer required. In this way the user has the illusion of infinite resources that can be used at the precise moment they become to be necessary for the proper service delivery

Broad network access: Resources are available over Internet through standard interfaces that allow the access from several devices such as PDA, laptops, desktops and mobile phones. For this reason, a key factor lies in the stability and availability of bandwidth that is guaranteed just from few years on a global scale.

Measured service: Cloud services are automatically managed and metered using appropriate monitoring and reporting systems for providing transparency for both the provider and end user. The use of specific metrics has allowed billing based on real usage of resources transforming, in this way, the IT in a commodity.

III. DELIVERY MODELS

Cloud providers deliver on three service layers: infrastructure, platform, and application.

Software as a Service (SaaS): It exists from long time and even anticipates CC. It can be considered as the basic idea behind the cloud and epitomized as "all-software in the end".A SaaS is an implementation of a business application that (a) is hosted in a cloud infrastructure and (b) is provided to customers through the Internet. This approach eliminates, for end users, the need to install and run the applications on their computers and reduces the cost of software by on-demand pricing.SaaS is on the top of the service stack and customers cannot control the underlying infrastructure and application platforms. In fact, the freedom of the end users is limited to the execution of the application and to certain user-specific configuration.

Platform as a Service (PaaS): A recent addition, it encloses everything for the whole software engineering lifecycle, from programming to deployment. The potential consumers of a PaaS service are software developers and testers, who are able to use infrastructure and do not need a tight control on it.This level includes two separate sections: (a) programming environment, with tools to support development and debugging, and (b) execution environment to deploy final application. Most PaaS vendors lock developers into their development platforms, and prevent direct communication with lower computing infrastructures or provide APIs with limited functionalities.

Infrastructure as a Service (IaaS): It is the lowest level of the service stack and contains all physical and virtual resources used to construct the cloud. Cloud infrastructure services are based on virtualized platforms, that are an evolution of the virtual private server, and offer to customers the possibility to buy on consumption whole range of physical resources, software and storage.In this way IaaS services avoid procurement, capacity planning, installation and configuration and reduce costs and time to deploy a system.The consumers are not able to control the underlying cloud infrastructure but have a complete control on operating systems, storage capacity and deployed applications.The high freedom and the wide possibilities of configuration led some researchers to divide this layer into sub layers related to the resources they depend on (a) computational resources, (b) data storage (c) communications.

summarizes the delivery and deployment models and explains how delivery models operate on deployment models. Within the delivery model we have shown the ways of nesting systems that ensure flexibility and efficiency.

IV. DEPLOYMENT MODELS

Each level of the cloud technology can be by the following models.

Private cloud: it is operated solely for one organization. A private cloud may be built and managed by the organization or external providers and may exist on or off premise [6]. Private clouds use virtualization and automated management technologies to ensure an high control on performance, reliability and security. Furthermore, a local data center reduces administrative and management tasks. This view has often been criticized for its similarity to the traditional proprietary server farm and the need to incur initial capital costs, a stark contrast to the basic philosophy of CC.

A. **Public cloud**: it provides computing services to public or industry group and it is owned by an external CSP .Public clouds do not require to invest initial capital on infrastructure and shift risks to infrastructure providers. However, it implies a lack of control on data, network and security settings, which hampers their effectiveness and spread in many business scenarios.
B. **Community cloud**: Multiple organizations with common requirements, as mission, policy and compliance, share cloud infrastructures across administrative domains. The infrastructure constituents of a community cloud can be managed by a partner organization and may exist on or off premise. An alternative is the Virtual Private Cloud (VPC) that is essentially a virtual platform running on top of public clouds, that is shared with other companies.
C. **Hybrid cloud**: It is a blended deployment model, with a composition of two or more clouds (private, community, or public) that are handled as a unique entity by a standardized technology that enables data and application portability. Generally, it addresses the limitations of the previous approaches. A hybrid cloud is commonly used to offload processes and data from the company private cloud to a reliable public cloud. This deployment model offers better flexibility, control and security over application data than public clouds but it requires a careful study to determine the best split of the components between public and private cloud.

Technological Issues

CC uses several current technologies and therefore perpetuate their issues. On the other side, the innovative service characteristics of CC pushed researchers to analyze the quality of service. By looking into cloud computing as a white box, researchers have identified potential critical areas: network configuration and data management.

A. Networking

In cloud computing two different networks coexist: (a) a logical network, made of virtual machines, that are deployed on physical servers interconnected by logical links, and (b) a physical network, made of physical servers and physical networking devices. This dual network requires to map optimized flows on the physical network and to allocate bandwidth to meet communication requirements. One approach is based on a “1-cuts” algorithm for node assignment. However, dynamics of resources looks too complex to be managed by predictions based on simulations. Therefore, a system of voluntary collaboration between resources is necessary, as suggested by the principle Catallaxy. The agents dynamically allocate resources by routing traffic demands in an optimal way, lowering cost and optimizing performance other solutions for agility and flexibility in cloud computing propose a virtualized optical network or an implementation of a reliable multicast service to reduce bandwidth on links while ensuring required reliability and flexibility.

B. Data Management

CC uses a layered storage area to run computing cycles and redundancy replication to ensure high availability. Moreover, the flexible architecture of CC provides, ideally, an infinite amount of resources, that allow data-intensive flows otherwise unapproachable. Invisible to cloud users, this structure ensures a stable, reliable and massively parallel computing engine. Some researchers have directed their efforts to create a distributed programming framework that supports simplified data-parallel applications. The implementation of this framework ensures long-term storage for large amounts of data and provides a natural data-parallel computing framework that operates the distributed datasets by assigning a group of operators to each element in datasets. The structure of CC guarantees optimal management of data, but raises critical issues in data mining. In mining methods and algorithms should exploit available resources and avoid unnecessary redundant operations. Some researchers have identified customizable templates, assembled recursively and capable to handle a wide class of SQL data mining querie. In addition to traditional mining, CC has fostered new capabilities in automatic data management and content interpretation. Most popular search engines use CC to parse file contents in the attempt of creating a semantic web.

REFERENCES

[1] M. Egger, G. D. Smith, and K. O’Rourke, “Rationale, potentials, and promise of systematic reviews”, Systematic reviews in health care: Meta-analysis in context, pp. 3-19, 2001.
[2] H. Yang and M. Tate, Where are we at with cloud computing?: a descriptive literature review, 2009.
[3] I. Sriram and A. Khajeh-Hosseini, “Research agenda in cloud technologies”, arXiv:1001.3259,. 2010.
[4] A. Khajeh-Hosseini, I. Sommerville, and I. Sriram, “Research challenges for enterprise cloud computing”, arXiv:1001.3257, 2010.
[5] G. Motta and N. Sfondrini, “Research studies on cloud computing: a systematic literature review”, 17th International Business Information Management Association Conference (IBIMA), 2011.
[6] P. Mell and T. Grance, The NIST definition of cloud computing. National Institute of Standards and Technology, 53, 2009.
[7] J. Varia, Cloud architectures, White Paper of Amazon, jineshvaria. s3, 2008.
[8] D. Chappell, Introducing the Windows Azure platform. Retrieved May, 30, 2010.
[9] J. Carolan, S. Gaede, J. Baty, G. Brunette, A. Licht, J. Remmell, L. Tucker, and J. Weise, Introduction to Cloud Computing architecture- White Paper, 2009.
[10] L. Wang, J. Tao, M. Kunze, A. C. Castellanos, D. Kramer, and W. Karl, “Scientific cloud computing: Early definition and experience”, IEEE, pp. 825-830, 2008.