Journal of Biosphere, 5: 31-35, 2016 ISSN 2278 – 3342

Cloud Computing: A new beginning in Scientific Computing

Monika Mahajan

Department of Computer Application,
Govt. Degree College, Hiranagar, Kathua, J&K, India

Abstract: With the development of multicore processors, virtualization, distributed storage and automated management, a new type of computing mode named cloud computing is produced. It distributes computation task on the resource pool which consist of massive computers, so the application systems can obtain the computing power, the storage space and software service according to its demand. It can concentrate all the computing resources and manage them automatically by the software without human intervention. Cloud Computing is one contemporary technology in which research community has recently embarked. Scientific computing requires an ever increasing number of resources to deliver results for ever growing problem sizes in reasonable time frame. Cloud computing proposes an alternative in which resources are no longer hosted by researcher's computational facilities, but are leased from big data centres only when needed. The cloud computing paradigm holds great promise for the performance hungry scientific community.

Keywords: Cloud Computing, Virtualization, Distributed storage, scientific computing

Cloud computing is a new type of computing mode. It distributes computation task on the resource pool which consists of massive computers, so the application systems can obtain the computation power, the storage space and software service according to its demand. This kind of resource pool is called "cloud". The Clouds are some virtual computation resources which can be maintained and managed by them, usually they are some large-scale server clusters, including calculating server, storage server, the broadband resources and so on. Cloud computing can concentrate all the computing resources and manage them automatically through the software without intervene.

Characteristics of Cloud Computing

Five characteristics of cloud are:

- On-demand self-service: Computing capabilities can be provided to a customer according to the requirement of the user. Capabilities like storage and server time are allocated without human interaction.
- Broad Network access: Using standard mechanisms, the cloud can be accessed through network using thick or thin clients. Examples of the clients are tablets, laptops, mobile phones and workstations.
- Resource pooling: In the multitenant model, the computing resources are pooled to provide service to multiple consumers. The computing resources can be present anywhere geographically and the exact location of resources is not known to the user.
- Rapid Elasticity: Depending on the user requirement, the capabilities and resources in the cloud can be released and provided automatically.

- Measured Service: The services provided to the user are measured by the cloud system and are reported to the user and the provider. Based on the type of service, the cloud system optimizes and controls the resource use by a metering capability.

Service Forms of Cloud Computing

Cloud Software as a Service (SaaS):

SaaS is a software delivery model providing access to applications through the internet as a web-based service. Applications are built to be accessible to multiple users through a web browser.

Characteristics of SaaS:

- The software is made available through the Internet.
- The Software is maintained by the service provider.
- The license to the software is based on subscription or usage and billed on a recurring basis.
- Zero maintenance is required at the end-user side and hence SaaS applications are very cost effective.
- Software is available on demand and can be scaled up and down according to the demand.
- Software is upgraded and updated automatically and also supports multitenancy.

GoogleApps, Oracle on Demand, SalesForce.com and SQL Azure5 are some of the examples of SaaS.

Cloud Platform as a Service (PaaS):

PaaS solutions constitute the middleware on top of which applications are built and provide a development and deployment platform for running applications on the cloud.

Characteristics of Paas:

- Built-in security, scalability, and web service interfaces are provided by PaaS.
- Built-in tools for defining business rules and defining workflow and approval processes are provided by PaaS.
- Integration of applications with other applications on the same platform is easy.
- PaaS provides web services interfaces which enable us to connect the applications outside the platform.

Force.com, Google AppEngine, Windows Azure Platform, GoGrid Cloudcenter are some of the examples of PaaS.

Cloud Infrastructure as a Service (IaaS):

IaaS solutions are most popular and developed market segment of cloud computing. IaaS solutions bring all the benefits of hardware virtualization.

Characteristics of IaaS:

- IaaS provides virtual machines with pre-installed Operating Systems.
- Resources are available On-demand.
- IaaS allows storing copies of data in different locations.
- The computing resources in the cloud can be easily scaled up and down.

Examples of IaaS providers include Amazon ECC, Eucalyptus, GoGrid, Flexiscale, Linode, RackSpace Cloud, Terremark.

Cloud Deployment Models

The four common deployment models are as follows:

Public cloud

The cloud infrastructure is owned by the cloud service provider. The cloud infrastructure exists in the premises of cloud provider. General public or a large industry group can access the cloud services for usage, on a pay according to usage method. The users are allocated the resources in the cloud on-demand. The resources are provided on a dynamic basis over the Internet. Small and medium enterprises (SMEs) benefit to great extent from using public clouds. Advantages of Public clouds are location independence, cost-effectiveness, reliability, flexibility, utility style costing and high scalability. Disadvantages are low security and less customizable.

Private cloud

The cloud infrastructure in a private cloud is operated solely for an organization. It can be managed by the organization itself or a third party. The private cloud can exist on premises or off premises. Advantages of private clouds are higher security and more privacy, more control, cost and energy efficiency. Disadvantages are limited scalability due to limited resources, inflexible pricing and private cloud is limited to a particular area.

Hybrid cloud

The cloud infrastructure in a Hybrid cloud is a composition of two or more clouds (private, community, or public). Each of them remain as unique entities but are linked together by standardized or proprietary technology. This technology enables data and application portability. Advantages of Hybrid clouds are scalability, flexibility, cost efficiency and security. Disadvantages are networking issues and security compliances.

Community cloud

The cloud infrastructure in a community cloud is shared by several organizations which have shared concerns (e.g. mission, security requirements, policy, and compliance considerations). It is generally managed by the organizations in the community or a third party and can be present either on-premises or off-premises. Advantages of Community clouds are that they are secure than public clouds and sharing of resources among several organizations. Disadvantages are that it is less secure than private cloud and requires governing policies for administration.

Key Technologies of Cloud Computing

Cloud computing systems use many technologies, of which the programming model, data management, data storage, virtualization are the key technologies.

Virtualization

Virtualization is a method of deploying computing resources. It separates the different levels of the application system including hardware, software, data, networking, storage and so on, breaks the division among the data center, servers, storage, networking, data and the physical devices, realize dynamic architecture, and achieves the goals of managing centralized and use dynamically the physical resources reducing the risk of management. In the cloud computing environment, all virtualization solutions are system integration solutions including servers, storage systems, network devices, software and service. They include multiple layers of virtualization technologies such as hardware virtualization, network infrastructure virtualization, application virtualization and desktop

virtualization, and combine several layers flexibly to realize the different models of virtualization solutions according to the application environment.

Mass Distributed Storage

In order to ensure high credibility and economy, cloud computing adopts distributed storage to save data, using redundancy storage to ensure the reliability of stored data and using high credible software to make up the incredibility of the hardware, therefore providing the cheap and credible mass distributed storage and computing system. The data storage system of cloud computing are Google File System (GFS) and Hadoop Distributed File System (HDFS) which is developed Hadoop team.

- GFS

  GFS is a distensible distributed file system. It is used in large and distributed applications which need to access mass data. The designing ideology of GFS is different from the traditional file system, which is designed for dealing large-scale data and the application property of Google. It runs on the cheap and common hardware, but it can provide fault tolerance function. It can provide high-performance service to a great deal of users.

- HDFS

  HDFS is a distributed file system which is applicable to running on commodity hardware. It is very similar to the existing distributed file system, but also with a significant difference, for example, HDFS is highly fault- tolerant and it can run on the cheap hardware; HDFS can provide data access with high throughput, so it applies to the application of large-scale dataset.

Parallel programming model

To enable users efficiently to use cloud computing resources and more easily enjoy services that cloud computing brings about; cloud computing programming model must make task scheduling and parallel execution transparent to users and programmers. Cloud computing adopts MapReduce programming model, which decomposes the task into multiple subtasks, and through two steps (Map and Reduce) to realize scheduling and allocation in the large-scale node. MapReduce is a parallel programming system developed by Google. It puts parallelism and fault tolerance, data distribution, and load balance in a database, and all the operations of data are summarized in two steps: Map and Reduce. MapReduce is mainly used in mass data processing. One of the features of the task scheduling strategy is scheduling priority the task the node which the data belong. This kind of scheduling scheme which is based on data position enables Map tasks to read and process data locally when the worker node which request task saves the data needing to process, thus reduces the network overhead and improve the performance of the system.

Data management

- Cloud computing needs to process and analyze mass and distributed data, therefore, data management technology must be able to efficiently manage large data sets. Cloud data management offers potential benefits such as speeding up technology deployments and reducing both capital expenditures and system maintenance costs

Few Cloud Projects in Scientific Computing

1. **LHC project:** Using the world's largest, most complex scientific instruments, CERN continues its study of the most basic element of matter i.e higgs boson. The raw data per event is around one million bytes (1 Mb), produced at a rate of about 600 million events per second.

That computational power at CERN is delivered through four cloud environments the Organization's IT team created using OpenStack , the suite of open source cloud software.

2. **Cloud Enabled Space Weather Platform (CESWP):** The purpose is to bring the power and flexibility of cloud computing to space weather physicists. The goal is to lower the barrier for the physicists to conduct their science i.e, to make it easier to collaborate with other scientists, develop space weather models, run simulations, produce visualizations and enable provenance.
3. **Rapyuta:** The RoboEarth Cloud Engine is an open source Platform-as-a-Service (Paas) framework designed specifically for robotics applications. It helps robots to offload heavy computation by providing secure customizable computing environments in the cloud. The RoboEarth Cloud Engine's nick name is "Rapyuta", after the castle in the sky inhabited by robots featured in the Japanese movie Rapyuta's computing environments allow robots to easily access the RoboEarth knowledge repository. Furthermore, they are tightly interconnected, paving the way for deployment of robot teams.

Cloud computing is a new kind of commercial computing model developed on the basis of grid computing, public computing and SaaS. It can distribute computing tasks to the resources pool consisting of massive computers, enabling different application systems to acquire computing power, storage space and various software services according to needs. The ultimate goal of cloud computing is to provide calculation, services and applications as a public facility

References

[1] Shufen Zhang, Hongcan Yan, Xuebin Chen,"Research on Key Technologies of cloud computing ", Physics Procedia 33(2012)1791-1797.

[2] Haibo Yang, Mary Tate, "A Descriptive Literature Review and Classification of cloud Computing Research", Communications of the Associations for Information Systems: Vol 31, Article 2.

[3] By: John W. Rittinghouse and James F. Ransome , cloud computing: Implementation, Management, & Security **,**CRC Press, Taylor & Francis Group, Boca Raton, FL, 2010, ISBN: 978-1-4398-0680-7.

[4] Alexandru Iosup, Simon Ostermann,Nezih Yigitbasi,Radu Prodan, , Thomas Fahringer, , and Dick Epema, ," Performance Analysis of Cloud Computing Services for Many-Tasks Scientific Computing", IEEE TPDS, Many-Task Computing, November 2010.

[5] Rani, B. Kezia, B. Padmaja Rani, A.Vinaya Babu," Cloud Computing and Inter-Clouds - Types, Topologies and Research Issues", Procedia Computer Science 50 (2015) 24 – 29.

[6] Aymerich, F.M. Fenu, G. Surcis, S. An approach to a Cloud Computing network. Applications of Digital Information and Web Technologies, 2008.

[7] Zhang Weimin, Tang Jianfeng. Cloud Computing. Science Press, 2009.

[8] M. Armbrust et al., "Above the Clouds, A Berkeley View of Cloud Computing", Technical Report, University of California, Feb, 2009.

[9] S. N. Srirama, "Scientific Computing on the Cloud (SciCloud)", http://ds.cs.ut.ee/research/scicloud.

[10] Hadoop: http://hadoop.apache.org/.

[11] K. Keahey, T. Freeman. "Science Clouds: Early Experiences in Cloud computing for Scientific Applications," Cloud Computing and Its Applications 2008 (CCA-08), Chicago, IL. October 2008.

[12] https://www.openstack.org/user-stories/cern/

[13] http://www.computerworld.com/article/2960642/cloud-storage/cerns-data-stores-soar-to-530m-gigabytes.html

[14] http://www.datacenterknowledge.com/archives/2012/03/02/european-research-forms-the-science-cloud

[15] http://roboearth.org/software-components