ISSN 2250-0987
Megha Gupta et al, UNIASCIT, Vol 2 (1), 2012, 173-178

Improving scope of Cloud technology under Open Source Tools

Megha Gupta[1], Syed Imtiyaz Hassan[2]
[1]*Lecturer, Northern India Engineering College,New Delhi*
[2]*Assistant Prof., Jamia Hamdard University, New Delhi*

Abstract: Cloud Computing is a paradigm designed to tie together the power of networks of computers and communications in a more cost effective way. Clouds provide elastic capability such that it could serve a wide and constantly expanding range of information processing needs, including business, government, and education. The Cloud computing paradigm is growing rapidly and is being considered for adoption in many business platforms. Open source software mentions to software systems whose source code is available, allowing for immediate incorporation of improvements and adaptations of the system by its users. Cloud computing is having a growing impact on IT industries and business activities in many large organizations. Applications have on-demand access to resources for a fee which depends on amount of resources the application consumes. Different implementation strategies of the same application may result in different costs. Cloud computing has different open source software frameworks which have different functionalities and uses. This paper discovers and evaluates different type of existing open source tools available for Cloud Computing.
Keywords- Apache Hadoop, Cloud Computing, Eucalyptus, MapReduce, Nimbus, OpenNebulz, Open source

1.INTRODUCTION

Cloud computing is the next generation in computation. Certainly people can have the whole thing they need on the cloud. Cloud computing is the next expected step in the evolution of on-demand information technology services and products. Cloud Computing is an evolving computing technology that is speedily relating itself as the next large step in the development and deployment of an increasing number of distributed applications [1, 2]. Cloud computing delivers infrastructure, platform, and software's that is made available as Subscription-based services in a pay-as-you-go model to consumers. The significance of these services was emphasized in a recent report from the University of Berkeley as: "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" [12].

IT companies with innovative ideas for new application services are no longer required to make large capital expenses in the hardware and software infrastructures. By using clouds as the application platform, IT companies are freed from the trivial task of setting up basic hardware and software infrastructures. Thus they can emphasis more on revolution and establishment of business values for their application services [10]. Some of the modern and evolving Cloud-based application services include social networking, web hosting, content delivery, and real time instrumented data processing. Each of these application types has different composition, configuration, and deployment requirements.

Cloud computing also designates applications that are extended to be accessible through the Internet. These cloud applications use large data centre and powerful servers that host Web applications and Web services. Anyone with a suitable Internet connection and a standard browser can access a cloud application.

Open source tools denote to software system whose source code is made available for use or modification by third-party developers. Thus, unlike centralized, proprietary software development models, open source offers practical accessibility to the source code, allowing for instant and simultaneous incorporation of different approaches, and eventually, the separating of the system into modified variations. Here we describe and evaluate some of the existing open-source software frameworks such as Apache Hadoop, Nimbus, Cumulus project, OpenNebula, Eucalyptus, UEC and Cloudsim .

2. OPEN SOURCE CLOUD SOFTWARE

2.1 Apache Hadoop

The Apache Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using a

simple programming model. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Rather than rely on hardware to deliver high-availability, the library itself is designed to detect and handle failures at the application layer, so delivering a highly-available service on top of a cluster of computers, each of which may be prone to failures.[17]

The Hadoop framework provides applications with reliability. Hadoop implements a computational paradigm named Map/Reduce, where the application is divided into many small fragments of work, each of which may be executed or re-executed on any node in the cluster. In addition, it provides a distributed file system (HDFS) that stores data on the compute nodes, providing very high aggregate bandwidth across the cluster. Both Map/Reduce and the distributed file system are designed so that node failures are automatically handled by the framework. [17].

2.2 Nimbus

Nimbus is a set of open source tools that together provide an "Infrastructure-as-a-Service" (IaaS) cloud computing solution. Nimbus allows a client to lease remote resources by deploying virtual machines (VMs) on those resources and configuring them to represent an environment desired by the user. [12]

Nimbus also provides an implementation of Amazon's Elastic Compute Cloud (EC2) that allows user to use clients developed for the real EC2 system against Nimbus based clouds. Nimbus also allows clients to create auto-configuring clusters. Starting one VM is somewhat straightforward, but launching 10s or 100s together can become laborious without heavily customized scripts. Nimbus provides a general solution to this problem, one that allows user to even store VMs with no private credentials on-board. [11].

2.3 Cumulus Project

The Cumulus project is an on-going Cloud computing project at the recently established Steinbuch Centre for Computing (SCC) at the Karlsruhe Institute of Technology (KIT). It intends to provide virtual machines, virtual applications and virtual computing platforms for scientific computing applications. The Cumulus project currently is running on high performance HP and IBM blade servers with Linux and the Xen hypervisor. Cumulus aims to build a test bed and infrastructure in the context mentioned above and integrates mainly already existing technology. [14]

2.4 OpenNebula

OpenNebula is an open and flexible tool that fits into existing data center environments to build any type of IaaS Cloud deployment. OpenNebula can be primarily used as a virtualization tool to manage your virtual infrastructure in the data-center or cluster, which is usually referred as Private Cloud. OpenNebula supports Hybrid Cloud to combine local infrastructure with public cloud-based infrastructure, enabling highly scalable hosting environments. OpenNebula also supports Public Clouds by providing Cloud interfaces to expose its functionality for virtual machine, storage and network management. [15]

2.5 Eucalyptus

Eucalyptus 2.0 is an open source Linux-based software architecture that implements scalable, efficiency-enhancing private and hybrid clouds within an organization's IT infrastructure. Eucalyptus provides Infrastructure as a Service (IaaS). This means that users can provision their own collections of resources (hardware, storage, and network) via Eucalyptus' self-service interface on an as-needed basis. A Eucalyptus cloud is deployed across an enterprise's "on-premise" data center and is accessed by users over enterprise intranet. Thus, with a Eucalyptus private cloud, sensitive data remains secure from external intrusion behind the enterprise firewall. Eucalyptus also interoperates seamlessly with Amazon's EC2 and S3 public cloud services and thus offers the enterprise a hybrid cloud capability. [16]

2.6 UEC

Ubuntu Enterprise Cloud (UEC) is included with Ubuntu Server Edition and integrates a number of open source projects . Ubuntu simplifies the complexities of cloud computing. Build within Amazon EC2 or create a private cloud on your own hardware. Ubuntu Server Edition helps user to create and control both. Ubuntu is the fastest route to a reliable cloud for business: it is the only distribution which includes all the components the user needs to build a cloud straight away. Private clouds give the power of cloud computing in one's own IT infrastructure. So one can get the benefits of cloud computing behind the security of his own firewall. Computing capacity can be grown or shrunk to meet user's application's needs. Amazon's Elastic Compute Cloud (EC2) allows user to build on-demand virtual systems on almost any scale without investing in hardware. Amazon EC2 fits perfectly with Ubuntu Enterprise Cloud's modularity, virtualisation capabilities, range of applications and optimised performance. [24]

2.7 Cloudsim

The CloudSim simulation layer provides support for modeling and simulation of virtualized Cloud-based data center environments including dedicated management interfaces for virtual machines (VMs),

memory, storage, and bandwidth. The fundamental issues such as provisioning of hosts to VMs, managing application execution, and monitoring dynamic system state are handled by this layer. A Cloud provider, who wants to study the efficiency of different policies in allocating its hosts to VMs (VM provisioning), would need to implement their strategies at this layer. Such implementation can be done by programmatically extending the core VM provisioning functionality. There is a clear distinction at this layer related to provisioning of hosts to VMs. A Cloud host can be concurrently allocated to a set of VMs that execute applications based on SaaS provider's defined QoS levels.Design of the CloudSim software framework and its architectural components. Initial releases of CloudSim used SimJava as discrete event simulation engine [5] that supports several core functionalities, such as queuing and processing of events, creation of Cloud system entities (services, host, data center, broker, virtual machines), communication between components, and management of the simulation clock. [13]

3. REVIEW STUDY

Cloud Computing is a technology that will definitely revolutionize the way we deploy and run our applications, operating various resources in much efficient and economical manner. But as of now, the 'Cloud' technology doesn't suit every kind of application that we use today.

Hence, below are the 3 basic characteristics of an application that suits 'Cloud computing' rather than conventional methods of app deployment and resource consumption.

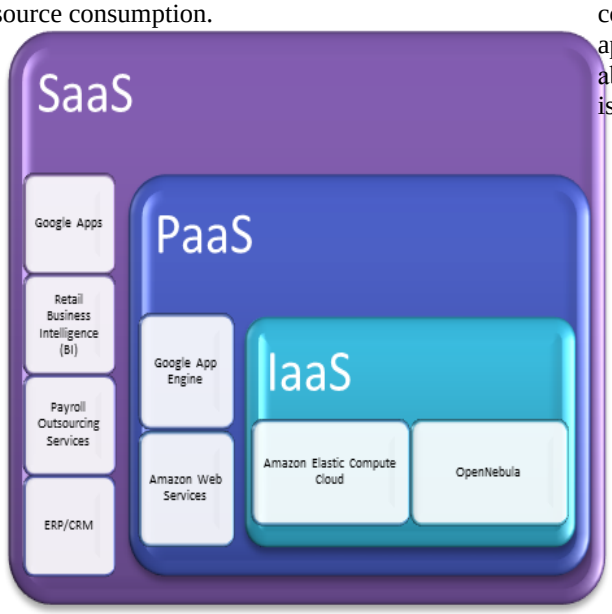

Figure 1: Cloud Services

3.1Applications which needs Flexibility:

The best part about Cloud computing is that it provides flexibility, shared and automated management that distributes resources over the cloud network. So, if the application's resource utilization is inconsistent rather spiky then "Cloud" will suit the most. For instance, travelling related sites hit peak traffic in December and quite average traffic otherwise so they can be considered appropriate for 'Cloud'.

3.2 Application which is growing exponentially or Demanding Scalability:

An app that's estimated to hit the rooftop and might continue its exponential nature of growth for quite some time is well suited for 'Cloud computing'. It's not a smart decision to buy out limited capacity hardware to take care of an application that will start overburdening the current infrastructure in lesser time than expected.

3.3Application which wants to run economically:

Cloud computing (specially 'Utility computing') lets its user to pay for what he actually uses and not for what he has reserved. Hence, apps (most of them are highly uncertain about their early growth),are better to be deployed on 'Cloud' to save out critical bucks.

4. MAJOR ISSUES OF USING CLOUD

More companies are starting to recognize and realize the benefits and advantages of cloud computing. However, as with any developing approach, there is some fear, uncertainty, and concern about the technology's maturity. Some of the major issues of using cloud computing are as follows

System / Property	Amazon Elastic Compute Cloud (EC2)	Google App Engine	Microsoft Live Mesh	Sun Network.com (Sun Grid)	GRIDS Lab Aneka
Focus	Infrastructure	Platform	Infrastructure	Infrastructure	Software Platform for enterprise Clouds
Service Type	Compute, Storage (Amazon S3)	Web Application	Storage	Compute	Compute
Virtualization	OS Level running on a Xen Hypervisor	Application Container	OS Level	Job management system (Sun Grid Engine)	Resource Manager and Scheduler
Dynamic Negotiation of QoS Parameter	None	None	None	None	SLA based Resource Reservation on Aneka side
User Access Interface	Amazon EC2 Command - Line Tools	Web-based Administration Console	Web-based Live Desktop and any devices with Live Mesh installed	Job submission scripts, Sun Grid Web Portal	Workbench, Web-based portal
Web APIs	Yes	Yes	Unknown	Yes	Yes
Value-added Service Providers	Yes	No	No	Yes	No
Programming Framework	Customizable Linux - based Amazon Machine Image (AMI)	Python	Not applicable	Solaris OS, Java, C, C++, FORTRAN	APIs Supporting different programming models in C# and other .Net supported Languages

Figure 2: Vendors and its representative cloud platform

. 4.1 Performance, latency

Performance concerns may stop some companies from using cloud computing for transaction oriented and other data-intensive applications. Customers who are a long distance from cloud providers could experience latency, particularly if there is a lot of

traffic and their code is not optimized for efficient transmission

4.2 Security and privacy

With the businesses' information and critical IT resources outside the firewall, customers worry about their vulnerability to attack. This could limit cloud computing desirability Users want to be confident that their cloud-computing provider is following standard security practices, which requires disclosure and inspection. For example, users don't necessarily want providers to have multiple customers sharing the same virtual hardware and network resources. Data stored in the cloud might be used anywhere in the world and thus might be subject to state or national data-storage laws related to privacy or record keeping.

4.3 Related bandwidth costs

While companies can save money on equipment and software with cloud computing, they could incur higher network-bandwidth charges from their service providers. Bandwidth costs might be low for smaller Internet-based applications that aren't data-intensive but could be very high when, for example, a company makes a multi terabyte database available via cloud computing.

4.4 Vendor lock-in and standards

There are no cloud-computing standards for elements and processes such as APIs, the storage of server images for disaster recovery, and data import and export. This is currently hampering adoption by limiting the portability of data and applications between Systems. If a company is dissatisfied with one cloud-computing service—or if the vendor goes

out of business—the firm cannot necessarily easily and inexpensively transfer service to another provider or bring it back in-house. Instead, the company would have to reformat its data and applications, and transfer them to a new provider, a potentially complex process. And if the company brings the service in-house, it would have to hire employees with the skill necessary to work with the technology.

4.5 Transparency

Companies cannot pass reviews of their capabilities by potential clients if they can't demonstrate who has access to their data and how they keep unauthorized personnel from retrieving information. Cloud-computing vendors are addressing this concern by having third parties audit their systems in advance and by documenting procedures designed to address customers' data-security needs

4.6 Reliability

Cloud computing does not always provided round-the-clock reliability.

5. CONCLUSIONS

With the expansion of Cloud technologies, there is a new need for applications to study and analyse the benefits of the technology and how best to apply the technology to large-scaled applications. The latest struggles to design and develop Cloud technologies focus on important innovative methods, plans and mechanisms for efficiently managing Cloud infrastructures. Cloud platforms are yet not at the focus of most people's attention. The charms of cloud-based computing, including scalability and lesser prices, are very factual. The open source tools growth will improve the scope of the cloud computing in deferent's fields like business, telecommunication and online Gaming.

REFERENCES

[1] Message Passing Interface (MPI), www.mcs.anl.gov/mpi/,

[2] P. Mell and T. Grance, "The NIST Definition of Cloud Computing; http://csrc.nist.gov/groups/SNS/cloudcomputing/

[3] B. Hayes, "Cloud computing", Communications of the ACM, (7):9–11, 2008.

[4] G. Gruman and E. Knorr, "What cloud computing really means", InfoWorld, 2008.

[5] J. Geelan, "Twenty one experts define cloud computing", Electronic Magazine, 2008.

[6] L. M. Vaquero, L. Rodero-Merino, J. Caceres and M.Lindner, "A Break in the Clouds: Towards a Cloud Definition", ACM SIGCOMM Computer Communication Review, Volume 39, Number 1, 2009.

[7] A. Lenk, Markus Klems, J. Nimis, S. Tai and T. Sandholm"What's Inside the Cloud? An Architectural Map of the Cloud Landscape", CLOUD'09, 2009.

[8] B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, I. Pratt, A.Warfield, P. Barham, and R. Neugebauer, "Xen and the Art of Virtualization", Proceedings of the ACM Symposium on Operating Systems Principles, 2003.

[9] M. Hibler, R. Ricci, L. Stoller, J. Duerig, S. Guruprasad, T. Stack, K. Webb, and J. Lepreau, "Large-scale Virtualization in the Emulab Network Testbed", Proceedings of the 2008 USENIX Annual Technical Conference, 2008.

[10] Hewlett-Packard, "HP Integrated Lights-Out 2 User Guide", Technical report, HP, 2009.

[11] Amazon EC2, http://aws.amazon.com/,

[12] Nimbus, http://www.nimbusproject.org/,

[13] Cloudsim, http://www.cloudbus.org/cloudsim/

[14] Lizhe Wang, Jie Tao, Marcel Kunze, Dharminder Rattu and Alvaro Canales Castellanos, "The Cumulus Project: Build a Scientific Cloud for a Data Center", CCA08, 2008.

[15] OpenNebula, http://www.opennebula.org,

[16] Eucalyptus, http://www.eucalyptus.com/,

[17] Apache Hadoop, http://hadoop.apache.org/,

[18] S.Ghemawat, H. Gobioff and S.T. Leung, "The Google File System", ACM Symposium on Operating SystemsPrinciples, 2003.

[19] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "Openflow: Enabling in novation in college networks", 2008.

[20] G. DeCandia, D. Hastorun, M. Jampani, G. Kakulapati, A. Lakshman, A. Pilchin, S. Sivasubramanian, P. Vosshall and W. Vogels, "Dynamo: Amazon's Highly Available Key- Value Store", ACM Symposium on Operating Systems Principles, 2007.

[21] F. Chang, J. Dean, S. Ghemawat, W. C. Hsieh, D. A. Wallach, M. Burrows, T.r Chandra, A. Fikes and R. E. Gruber, "Bigtable: A Distributed Storage System for Structured Data", Symposium on Operating System Design and Implementation, 2006.

[22] Caroline, http://labs.oracle.com/projects/caroline/,

[23] Django web framework, http://www.djangoproject.com,

[24] Uec, http://www.ubuntu.com/cloud

[25] GoogleAppEngine, http://code.google.com/appengine/,

[26] MicrosoftAzure, http://www.microsoft.com/windowsazure OpenId, http://openid.net/,

[27] http://my.opera.com/blu3c4t/blog/2008/11/01/a-glance-to-cloud-computingE. Deelman, G. Singh, M. Livny, B. Berriman and J. Good, "The Cost of Doing Science on the Cloud: The Montage Example", Proceedings of Super Computing 2008.

[28] J. Napper and P. Bientinesi, "Can Cloud Computing Reach The TOP500?", Conference On Computing Frontiers, 2009.