2012 International Conference on Information and Network Technology (ICINT 2012)
IPCSIT vol. 37 (2012) © (2012) IACSIT Press, Singapore

Efficient Cloud Management for Parallel Data Processing In Private Cloud

Shenbaga Lakshmi and Senthamarai Vijai [+]

CSE Dept, Easwari Engineering College

Abstract. Cloud computing is gaining acceptance in many IT organizations, as an elastic, flexible, and variable-cost way to deploy their service platforms using outsourced resources. Many-task computing (MTC) paradigm embraces different types of high-performance applications involving many different tasks, and requiring large number of computational resources over short period of time. In this paper, we implement private cloud by using eucalyptus middleware. It basically used to implement infrastructure as a service (IaaS).Thus it helps for the organization to create their own cloud structure which eliminates renting from the public cloud providers like Amazon Web Services. It also offers flexible infrastructure services that can be easily utilized and managed by end users according to their needs. It enables enterprises and government agencies to establish their own cloud computing environments. An important issue in cloud computing is how resources can be allocated and managed in a cost-effective manner.

Keywords: many-task computing, eucalyptus, amazon web service, virtualization.

1. Introduction

Eucalyptus is a technology which allows anyone to set up their own cloud system on their own hardware. The name Eucalyptus is an acronym for “Elastic Utility Computing Architecture for Linking Your Programs to Useful Systems”. The framework essentially implements what is commonly referred to as "Infrastructure as a Service": a system with the ability to run and control collections of virtual machine instances deployed across a variety of physical servers. Focusing on the cost advantage for IT infrastructure and data center management, Eucalyptus provides a unique framework with a variety of interfaces to manage the resources. The hardware, network, and storage can easily be consolidated under the Eucalyptus cloud, hiding the heterogeneity in hardware, software stack, policies and configuration. Eucalyptus Systems develops enterprise grade technology solutions built on the Eucalyptus for private and hybrid cloud computing. Eucalyptus technology is quickly becoming the standard for on-premise cloud computing, delivering the cost efficiencies and scalability of cloud architecture with the security and control of deploying on an organization’s own IT infrastructure.

The rest of this paper is structured as follows. Section 2 presents background information on Eucalyptus, Section 3 reviews related work on other middleware, while Section 4 introduces the proposed approach. Section 5 describes the setting of the empirical study and Section 6 presents and discusses the results. Finally, section 7 concludes this work and points to future research directions.

2. Related Work

Middleware, in the context of distributed computing systems, was first described by Bernstein et al. [1] as a set of intermediaries for the components in a distributed computing system. This concept has been extensively utilized during the uprising of the Service-Oriented Architecture (SOA) where the services in question were in fact provided by middleware systems. Middleware in general is used to abstract the

[+] Corresponding author. Tel.: + 9444321089.
E-mail address: senthamarai.vijay@gmail.com.

differences between heterogeneous systems and expose a uniform interface. The different cloud middleware are

2.1. Open nebula

Borjasotomayor et al. [2] uses Open Nebula as a middleware that deploy and manage VMs, either individually or in groups that must be Co scheduled on local resources or external public clouds. It automates VM setup (preparing disk images, setting up networking, and so on) regardless of the underlying virtualization layer (Xen, KVM, or VMware are currently supported) or external cloud (EC2 or Elastic Hosts are currently supported).

2.2. Hadoop

Li, Jing-min et al [3] described about Apache Hadoop software library which 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.

2.3. Nimbus

Junjie Peng et al [4] presents Nimbus which is an open tool set, and also a cloud computing solution providing IaaS It permits users lease remote resources and build the required computing environment through the deployment of virtual machines. Nimbus have supported many nonscientific research domain applications

3. System Architecture

The system architecture of the proposed system is given in fig 3. A brief description of the components within the Eucalyptus system follows.

3.1. Components of eucalyptus cloud

The different components of Eucalyptus Cloud are

Node Controller (NC)
Cluster Controller (CC)
Walrus Storage Controller (WS3)
Storage Controller (SC)
Cloud Controller (CLC)

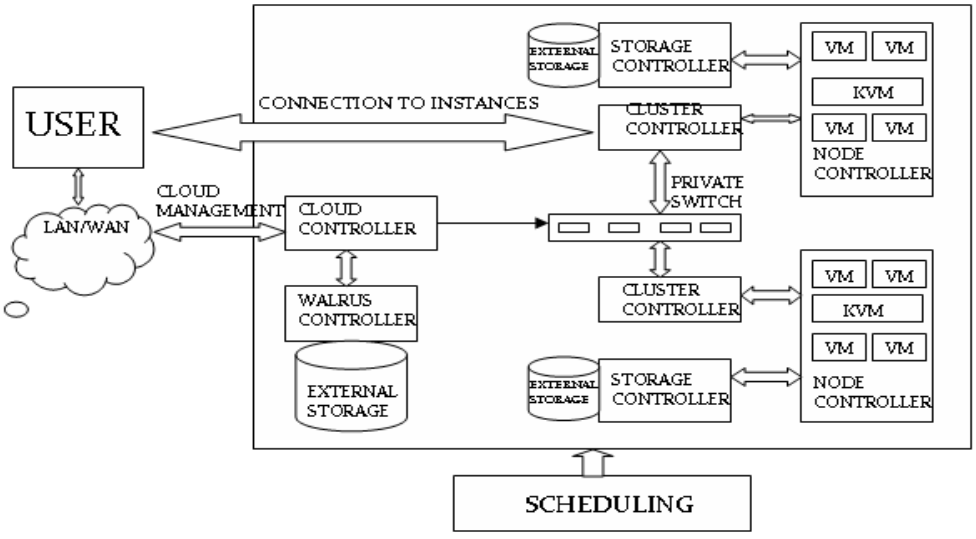

Fig. 1: System Architecture

Node Controller (NC). Node Controller runs on each node and controls the life cycle of instances running on the node. The NC interacts with the OS and the hypervisor running on the node on one side and the Cluster Controller (CC) on the other side. NC queries the Operating System running on the node to discover the node's physical resources - the number of cores, the size of memory, the available disk space and also to learn about the state of VM instances running on the node and propagates this data up to the CC.

Certain functions like Collection of data related to the resource availability and utilization on the node and reporting the data to CC, Instance life cycle management can be performed.

Cluster Controller (CC). CC manages one or more Node Controllers and deploys/manages instances on them. CC also manages the networking for the instances running on the Nodes under certain types of networking modes of Eucalyptus. CC communicates with Cloud Controller (CLC) on one side and NCs on the other side. Certain functions like receiving requests from CLC to deploy instances, deciding which NCs to use for deploying the instances, controlling the virtual network available to the instances, collecting information about the NCs registered with it and report it to the CLC

Walrus Storage Controller (WS3). WS3 provides a persistent simple storage service using REST and SOAP APIs compatible with S3 APIs. Certain functions like Storing the machine images, Storing snapshots, Storing and serving files using S3 API can be performed in this controller. WS3 should be considered as a simple file storage system.

Storage Controller (SC). SC provides persistent block storage for use by the instances. This is similar to the Elastic Block Storage (EBS) service from AWS. Certain functions like Creation of persistent EBS devices, providing the block storage over AoE or iSCSI protocol to the instances, allowing creation of snapshots of volumes

Cloud Controller (CC). The Cloud Controller (CLC) is the front end to the entire cloud infrastructure. CLC provides an EC2/S3 compliant web services interface to the client tools on one side and interacts with the rest of the components of the Eucalyptus infrastructure on the other side. CLC also provides a web interface to users for managing certain aspects of the UEC infrastructure. Certain functions like monitoring the availability of resources on various components of the cloud infrastructure, including hypervisor nodes that are used to actually provision the instances and the cluster controllers that manage the hypervisor nodes, Resource arbitration - Deciding which clusters will be used for provisioning the instances and monitoring the running instances.

3.2. Private cloud setup

The functional architecture of the private cloud setup is given in fig 3.1.Two servers (server 1 and server 2) will run a 64-bit server version and third server will run a Desktop 64-bit version (client 1). Then install the Desktop version on client 1 so that Firefox or other browsers can be used to access the web interface.

The following modifications are to be made for the private cloud setup.

The gateway for Server2 is set to the IP of the CC IP 192.168.20.1.This will enable the Server2 to connect to the enterprise network through Server1 (CC).

Server1 is a 64-bit server and Server2 is a 64-bit VT-enabled server.

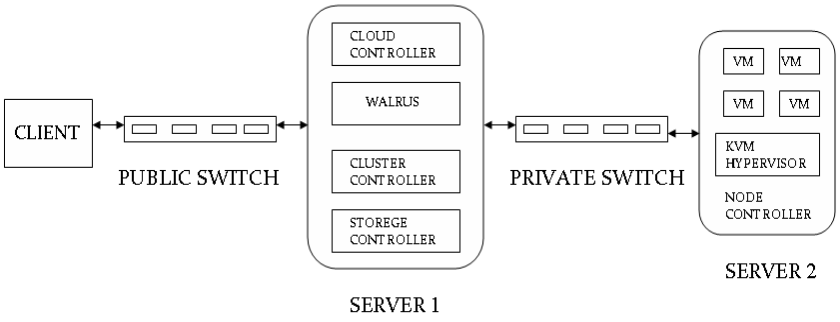

Fig. 2: Functional Architecture for Private Cloud Setup

The enterprise network runs on a classC private network 192.168.10.0 / 255.255.255.0

IP addresses allocated for the cloud instances as public IP addresses:

192.168.10.200-192.168.10.220 (Enterprise range)

Table. 1: Hardware Setup for Eucalyptus Cloud Infrastructure

Hardware	Server 1		Server 2		Client 1	
	Minimum	Suggested	Minimum	Suggested	Minimum	Suggested
CPU	1GHz	2 x 2GHz	VT extensions	VT,64Bit,Multi core	VT extensions	VT,64Bit, Multi core
Memory	1GiB	2GiB	1 GiB	4 GiB	1 GiB	2GiB
Disk	5400 rpm IDE	7200rpm SATA	5400rpm IDE	7200rpm SATA	5400rpm IDE	7200rpm SATA
Disk Space	40 GiB	200 GiB	40GiB	100 GiB	40GiB	100 GiB
Networking	100 Mbps	1000 Mbps	100 Mbps	1000 Mbps	100 Mbps	100 Mbps

3.3. Steps for setting up eucalyptus cloud

Step1: preparation of the requirement

Step2: install the cloud /cluster /storage /walrus front end server.

Step3: install the node controller.

Step4: register the node

Step5: installing cloud administrative credentials through command line.

Step6: installing cloud images.

Step7: running an instances

Verifying resources

Checking images

Checking security group

Installing a keypair

Running of the instances

Step8: monitoring and accessing the instances

4. Experiments and Results

A private cloud has been setup based on Ubuntu's 10.04 Server edition, that consists of two Servers – Server A and Server B. Server A acts as the cloud, cluster, warehouse and storage controller and Server B acts as node controller. We configured Machine A on a Core2duoX6800 processor based machine with 2GB DDR 2 RAM and 80 GB Hard disk. Machine B is running on an AMD PhenomeII X4 965 processor with 4 GB DDR 3 RAM and 250 GB Hard disk. The nodes communicate through a fast local area network.

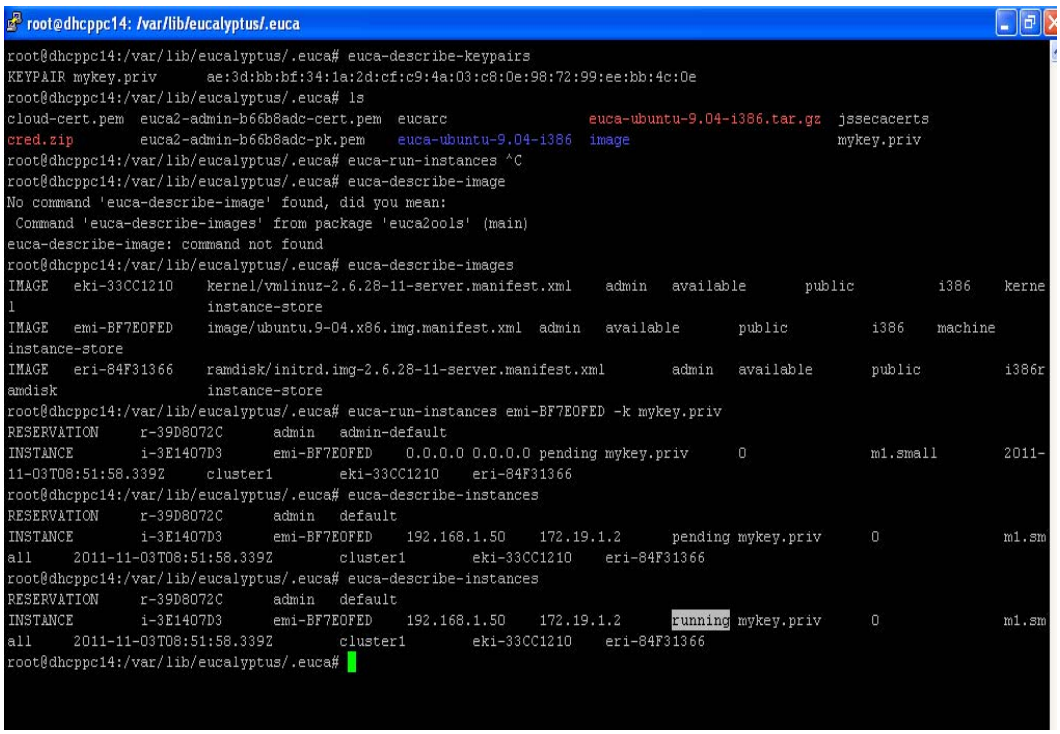

Fig. 3: Running of Instances

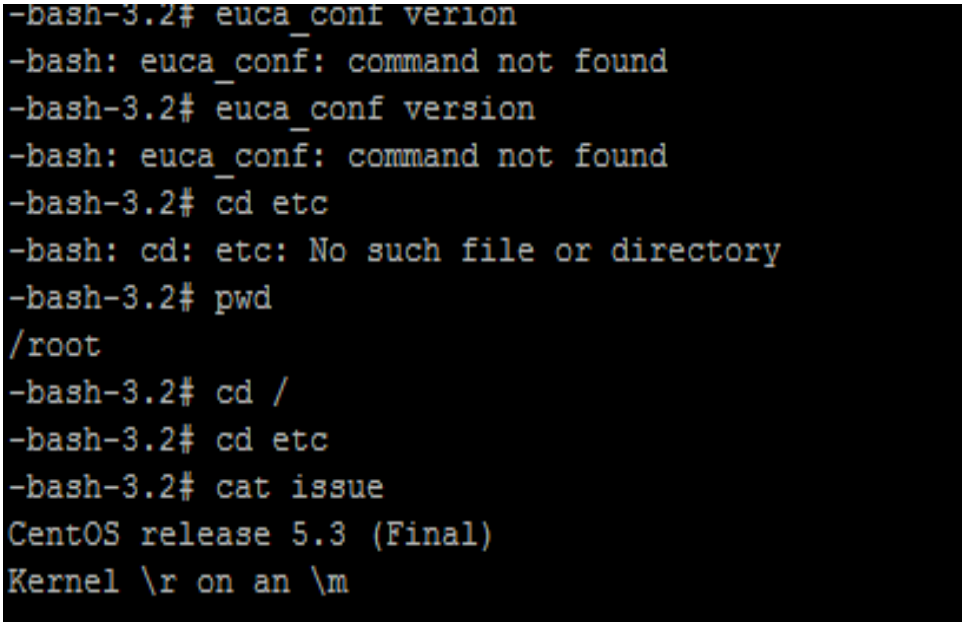

Fig. 4: Working In a Instances

5. Conclusion

In this paper, a clear view of how the private cloud can be setup and how an instance i.e.) a virtual machine can be created and thus new Operating System can be boot from the virtual machine. It basically used to implement infrastructure as a service (IaaS). Thus it helps for the organization to create their own cloud structure which eliminates renting from the public cloud providers like Amazon Web Services. It also offers flexible infrastructure services that can be easily utilized and managed by end users according to their needs. It enables enterprises and government agencies to establish their own cloud computing environments. In the future, a proper CPU Scheduling and security management can also be incorporated in the private cloud setup.

6. References

[1] Bernstein, P.A. Middleware: A Model for Distributed System Services Communication.

[2] Borja Sotomayor, Ruben S.Montero, Ian Foster Virtual Infrastructure Management In Private And Hybrid Cloud published by the IEEE computer security, (2010) by, pp.14-22.

[3] Li, Jing-min; He, Guo-hui, Research Of Distributed Database System Based On Hadoop Published by International Conference On Information Science And Engineering (2010) by pp.1417-1420

[4] Junjie Peng, Xuejun Zhang, Zhou Lei, Bofeng Zhang, Wu Zhang, Qing Li.Comparison Of Several Cloud Computing Platforms Published by International Symposium On Information And Science Engineering (2009)

[5] Balachandra Reddy Kandukuri, Ramakrishna Paturi V, Dr.Atanu Rakshit, Cloud Security Issues, An International Ieee Conferences On Service Computing On (2009) Pp.517-520.

[6] Paul Marshall and Kate Keahey and Tim Frrman, Improving Utilization of Infrastructure Clouds, IEEE journal on May (2011), pp.205-214.

[7] Sohan Singh Yadav and Zeng Wen Hua, Cloud: A Computing Infrastructure On Demand published by an IEEE International Conferences on Computer Engineering and Technology on April (2010), pp.v1-423 - v1- 426.

[8] Daniel Warneke and Odej Kao, Exploiting dynamic resource allocation for efficient parallel data processing in the cloud, proc. parallel and distributed system (2011), vol.22 pp.985-997.

[9] Grabauer B.,Walloshek T. and Stocker E., Understanding Cloud Computing vulnerabilities, proc. Computer Engineering and Technology (2011),vol.99 pp.149-167.

[10] Shuai Zhang, Xuebin Chen, Shufen Zhang and Xiuzhen Huo 'Comparison between cloud computing and grid computing', conf. computer application and system modeling, vol.11 pp.v11-72 –v11-75.