A Highly Scalable Private Cloud Infrastructure for Educational Institutions based on Ubuntu Enterprise Cloud with Eucalyptus

E. Sivaraman
Ph.D Research Scholar
Department of Computer Science
NGM College, Pollachi, India

R. Manickachezian, Ph.D.
Associate Professor
Department of Computer Science (Aided)
NGM College, Pollachi, India

ABSTRACT

Every educational institution has its discrete goals, standards and regulations, security concerns, budget pressures, and technical legacies. No particular approach tackles all technology challenges, and solutions can hardly take a one-size-fits-all approach. The advancement toward cloud computing technologies facilitate organizations to access Internet-based information, applications, and services brings flexibility, increased choice, and capabilities to education IT departments. We provide a solution for academic institutions by building a highly scalable private cloud infrastructure based on Ubuntu Enterprise Cloud (UEC) with Eucalyptus. We discuss about deployment of minimal cloud infrastructure by considering fundamental network topology, with a single system serving as the "all-in-one controller", and one or more nodes attached to it. UEC infrastructure with Eucalyptus on different topologies such as implementation on a single server, two physical servers, three physical servers and more are discussed to identify how scalability is achieved. Finally we discuss the installation of education cloud with UEC based on Linux 10.04 and benefits of the private education cloud are analyzed.

General Terms

Education, Infrastructure, Network, Physical Machines, Servers

Keywords

Cloud, Eucalyptus, Private Cloud, Ubuntu Enterprise Cloud, Scalability, Virtualization

1. INTRODUCTION

Cloud computing is a kind of distributed computing that provides elastic, dynamic processing power and storage when needed. It basically gives the user the computing power when it needs it. When the college bell rings to study hours that last well into the night, students are demanding more technology services from their institutions. It's essential not only to keep pace with their evolving needs, but also to prepare them for the demands of the workplace tomorrow. At the same time, educational institutions are under increasing pressure to deliver more for less, and they need to discover ways to present rich, affordable services and tools. Those instructors who can deliver these sophisticated communications environments, including the desktop applications that employers use today, will be serving their students come across better jobs and greater opportunities in the future. Cloud computing can help afford those solutions. It's a network of computing resources located just about anywhere that can be shared. They provide to education a range of options not found in traditional IT models. In fact, the integration of software and assets organizations own with software and services in the cloud provides with new choices for balancing system management, cost, and security while helping to improve services.

2. DESIGNING AND IMPLEMENTING PRIVATE CLOUD INFRASTRUCTURE

2.1 Designing and Choosing the Right Solution

The primary step in the designing process is to involve many of institution's stakeholders as possible such as teachers, students, system administrators, etc., one person can't know everything. A solution that seems to be just right for the organization might have hidden issues, like security or privacy barriers, that one is not aware of. In that case, knowing about the existence of the complications, it would have impact on the design decision of choosing private or public cloud solution. Such a wrong decision in the early stages of the project could have a great economic impact later in the project, not to talk about the possible privacy and security breaches. A private cloud solution facilitates self-service provisioning, storage, pooling of networking, and compute resources to maximize the utilization of hardware by merging workloads from various applications and services in a managed, secure, and reliable way. It can help institutions allocate consumption of resources by different services or departments to better prioritize IT spend.

2.2 Ubuntu Enterprise Cloud with Eucalyptus

Ubuntu private cloud with Eucalyptus is preferred for the implementation for numerous reasons. First, it is Open Source Software (OSS), which is one of the important requirements for the project. In academic and educational environments, OSS solutions are primarily used, for the sake of interchangeability between different organization and because of socio-organizational reasons. Observations show that researchers and scientists in academic world usually do not prefer proprietary software solutions if they can find something equally good from open source community [1-2]. Ubuntu Enterprise Cloud is a heap of applications from Canonical included with Ubuntu Server Edition. UEC incorporates Eucalyptus along with a number of other open source software used to build cloud. UEC formulates it very easy to install and configure the cloud. Canonical also offers commercial technical support for UEC. Eucalyptus is infrastructure software that allows enterprises and government agencies to set up their own cloud computing environments. With Eucalyptus, educational institutions make more efficient use of their computing capacity, thus increasing productivity and innovation, deploying new applications faster, and protecting sensitive data, while reducing capital expenditure.

UEC, which is Ubuntu Linux distribution with Eucalyptus cloud software integrated, is suitable since it is designed and suited to support private cloud solutions from a minimal starting point, and then scaling up or out as needed. This collection of the Ubuntu Server Edition with Eucalyptus and other applications makes it very easy to install and configure the cloud. Eucalyptus has been designed as modular set of five simple elements that can be easily scaled [3]. Eucalyptus offers both computing and storage, whereas other private cloud software like Open Nebula, Open Stack as for the time being, are offering only computational power. Ubuntu is the fastest budding Linux-based operating system with more than 10 million users and a huge community of developers [4-6]. Canonical, a corporation that supports and subsidizes Ubuntu, specializes in the free provision of software for common activities [7-8]. Canonical will also be contributing commercial support for internal clouds, along with technical support for installation. Thus, a prevalent community support as well as commercial support is available for Eucalyptus.

2.3 Components of UEC based Cloud with Eucalyptus

UEC with Eucalyptus is an on-premise private cloud platform, designed as a distributed system and is a modular set of five simple elements. Linux based processes that interrelate with each other form the cloud platform. The following components such as Node Controller (NC), Cluster Controller (CC), Walrus Storage Controller (WS3), Storage Controller (SC), Cloud Controller (CLC) comprise the Eucalyptus architecture (as shown in fig. 1) for UEC implementation [9]. Each component is acting as an autonomous web service that renders Web Service Description Language (WSDL) document defining the Application Programming Interface (API) to interact with it.

2.3.1 Node controller

Eucalyptus supports several hypervisors such as Xen and KVM in open source version, and VMware (ESX/ESXi) in Enterprise Edition [10]. Kernel-based Virtual Machine (KVM) is a virtualization key for Linux on hardware that contains virtualization expansions. A UEC node is a VT (Virtualization Technology) enabled server capable of running KVM as the hypervisor. UEC automatically installs KVM when the user prefers to install the UEC node. The VMs (Virtual Machines) running on the hypervisor and controlled by UEC are termed "instances". Even though Eucalyptus supports other hypervisors like Xen apart from KVM, but Canonical has chosen KVM as the preferred hypervisor for UEC. Node Controller runs on each one of the UEC node and controls the life cycle of instances running on the node. The NC interacts with the operating system 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 determine the node's physical resources such as the number of cores, the size of memory, the available disk space and also to discover about the state of VM instances running on the node and broadcasts this data to the CC.

2.3.2 Cluster Controller (CC)

A cluster is a collection of machines assembled together in the same network broadcast field. CC is the doorway point to a cluster. CC administers one or more Node Controllers and deploys instances on them. It manages the networking for the
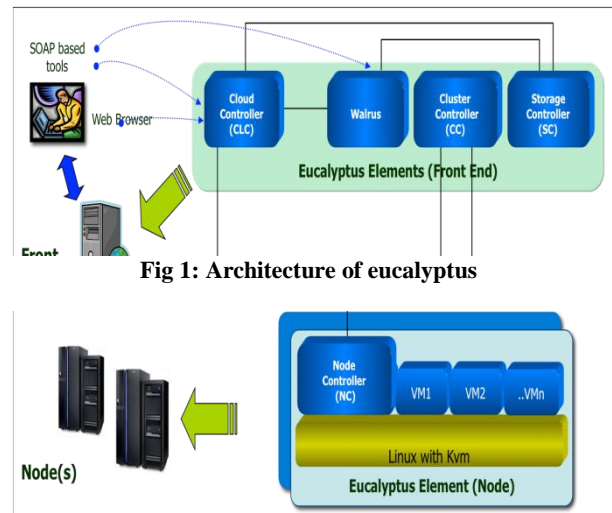

Fig 1: Architecture of eucalyptus

instances running on the Nodes under certain types of networking modes of Eucalyptus. It also manages the virtual network available to the instances, and collects information on NCs and reports them to Cloud Controller (CLC). In short we can say CC communicates with NCs on one side and CC on the other side.

2.3.3 Walrus Storage Controller (WS3)

WS3 is correspondent to Amazon's S3. It implements the S3 like functionality – bucket based storage system with put or get storage representation (create and delete buckets, create objects, and put or get those objects from buckets). It affords a persistent simple storage service using SOAP and REST APIs compatible with S3 APIs for storing and serving files. It stores the machine images and snapshots and hence it is considered as a simple file storage system.

2.3.4 Storage Controller (SC)

SC provides persistent block storage for use by the instances on the cluster level. The instances are using Internet SCSI (iSCSI) or ATA over Ethernet (AoE) protocol to mount the virtual storage devices. SC permits the creation of block storage similar to the Elastic Block Storage (EBS) service from AWS. Its main functions are and creation of storage volumes, attaching, detaching and creation of snapshots.

2.3.5 Cloud Controller (CLC)

CLC is the entry point to UEC/Eucalyptus cloud, and is equal to Amazon's EC2. It monitors the running instances and collects information on the usage and availability of the resources in the cloud. It is the considered as frontend for managing the entire UEC/Eucalyptus cloud infrastructure. It provides an EC2/S3 compliant web services interface to the client tools on one side and for administrators to manage an infrastructure (such as images, users, groups, storage, network, and clusters) on the other side. Based on the information gathered about the infrastructure's load and resource availability, it judges the available resources and dispatches the load to the cluster

3. DEPLOYMENT OF MINIMAL CLOUD INFRASTRUCTURE

To build a private cloud, canonical suggests using Ubuntu Linux 10.04 as the operating system of choice [11]. For evaluation and experimental analysis, we assume a fundamental network topology, with a single system serving as the "all-in-one controller", and one or more nodes attached. We install, configure, register and perform several operations on a basic UEC setup that results in a cloud with a one controller "front-end" and one or several nodes for running Virtual Machine (VM) instances. To deploy a minimal cloud infrastructure, two dedicated systems are needed and they are front end and one or more nodes.

3.1 Front End Configuration

With reference to the hardware dimensioning, there are no fixed requirements, but rather recommendations [12]. The two dedicated systems – a "front end & controller" type (hosting CLC/WS3/CC/SC components), and "NC" type (hosting NC component). Table 1 shows the suggested configurations for the "frontend and controller" system that will run one or more of components like CLC/WS3/CC/SC.

Table 1. Recommendation for front end configuration

Hardware	Minimum	Suggested	Comments
CPU	1 GHz	2 x 2 GHz	for an all-in-one front end, at least dual core processor is recommended
Memory	2 GB	4 GB	depending on the number of components
Disk	5400rpm IDE	7200rpm SATA	slower/cheaper disk would work, but it is suggested to go for a faster versions (SATA, SCSI or SAN if the budget allows)
Disk Space	40 GB	200 GB	40 GB is only enough space for only a single image, cache, etc., Eucalyptus does not like to run out of disk space
Networking	100 Mbps	1000 Mbps	copying machine images (hundreds of MBs) is done over network; if the budget allows, consider 10 Gbps)

3.2 Other Node Systems (NC)

The other systems are nodes, which will run the NC and they further run instances. Table 2 shows the suggested configurations for the "NC" type system that will be used for

hosting NC component. Separation is done based on the different load that the different components inflict on the hardware – "front end & controller" type is not processor and memory demanding as the – "NC" type is. Besides, NC type should have VT extensions enabled before.

Table 2. Recommendation for NC configurations

Hardware	Minimum	Suggested	Comments
CPU	VT extensions	VT, 64-bit, Multicore	for an all-in-one front end, at least dual core processor is recommended
Memory	1 GB	4 GB	depending on the number of components
Disk	5400rpm IDE	7200rpm SATA or SCSI	since nodes are disk intensive, hard disks presents a performance bottleneck; consider faster versions (SATA, SCSI or SAN if the budget allows)
Disk Space	40 GB	200 GB	images are cached locally; eucalyptus does not like to run out of disk space
Networking	100 Mbps	1000 Mbps	copying machine images (hundreds of MBs) is done over network; if the budget allows, consider 10 Gbps)

4. UEC IMPLEMENTION ON DIFFERENT MACHINES

An UEC may consist of as few as two physical systems, up to potentially hundreds or possibly thousands of systems. Identifying a best configuration is a problem of the needs and the budget available, and several UEC topologies are presented for fitting the constraints [13]. In general, it is a problem of finding the number of physical servers that would host the various UEC web service components such as CLC, WS3, CC, SC, NC and figuring suitable types of distributions of those components across the servers. On the other hand, there will also be a need of another machine that can serve as a client machine and does not need to be a dedicated machine. Any OS that would include an internet browser such as Internet Explorer, Firefox, Chrome could be installed on the required client machine. That will be used only for accessing the web interface of UEC, and for making remote control sessions for viewing and interacting with computer desktops of instances, using Remote Desktop Protocol (RDP) or VNC [14-15]. The typical minimal eucalyptus setup for the components such as physical servers and client machine is shown in the fig. 2.
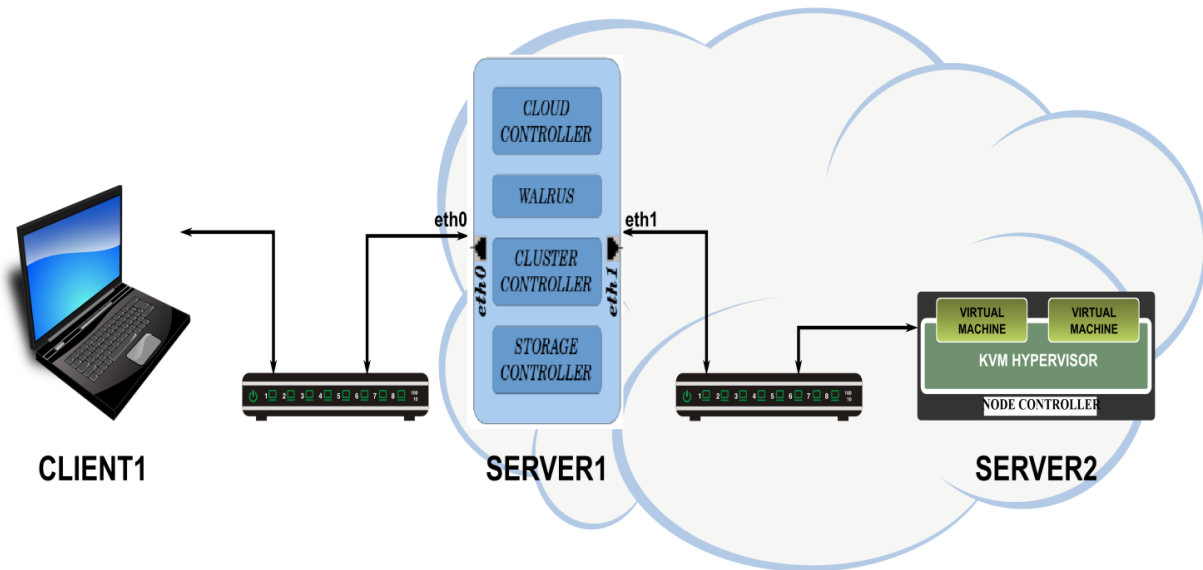
Fig 2. Typical minimal UEC with Eucalyptus set up for private education cloud

4.1 Implementation on Single Physical Machine

This is the most basic possible configuration, with all the Eucalyptus components such as CLC/Walrus/CC/SC and VM hosting (NC) implemented on the same machine. Implementing on one physical system is not a suggested topology for production use, but only for experiment purposes. But implementing Eucalyptus on single machine is achievable however with some limitations [16]. Implementing inside a virtual machine would generate an additional layer of virtualization, and KVM may not work properly, since VT enabled machines are required.

4.2 Implementation on 2 Physical Machines

This is the typical minimal Eucalyptus setup for UEC. This configuration sets the entire user facing components (CLC/Walrus) and back-end control components (CC/SC) on a single system, and uses the second machine for VM hosting as shown in the fig. 3. This second physical machine hosts virtual instances, and therefore should be properly configured. In addition, this machine needs to be VT enabled. Scaling out option for this configuration is just adding more NCs to the existing set up.

4.3 Implementation on 3 Physical Machines

User facing components (CLC/WS3) are installed separately on the first (CLC) and second (WS3) physical server. Back-end controlling components (CC/SC) are installed on the third physical machine. The fourth physical machine is having the NC component installed. This fourth physical machine hosts virtual instances, and therefore should be properly configured. In addition, this machine needs to be VT enabled. Scaling out option for this configuration is just adding more NCs to the existing set up.

4.4 Implementation on 4 Physical machines

User facing components such as CLC/WS3 are installed separately on the first (CLC) and second (WS3) physical server. Back-end controlling components (CC/SC) are

installed on the third physical machine. The fourth physical machine is having the NC component installed. This fourth physical machine hosts virtual instances, and therefore should be properly configured. In addition, this machine needs to be VT enabled. Scaling out option for this configuration is just adding more NCs to the existing set up.

4.5 Implementation on 5 Physical Machines

Client facing components (CLC/WS3) are installed on the first physical server and back-end controlling components (CC/SC) for the first cluster are installed on the second physical machine, and the third physical machine is having the NC component belonging to that cluster installed. Back-end controlling components (CC/SC) for the second cluster are installed on the fourth physical machine, and the fifth physical machine is having the NC component belonging to the second cluster installed. These third and fifth physical machines are hosting virtual instances, and should be properly dimensioned. These machines needs to be VT enabled. Scaling out option for this configuration is just adding more NCs to the existing set up.
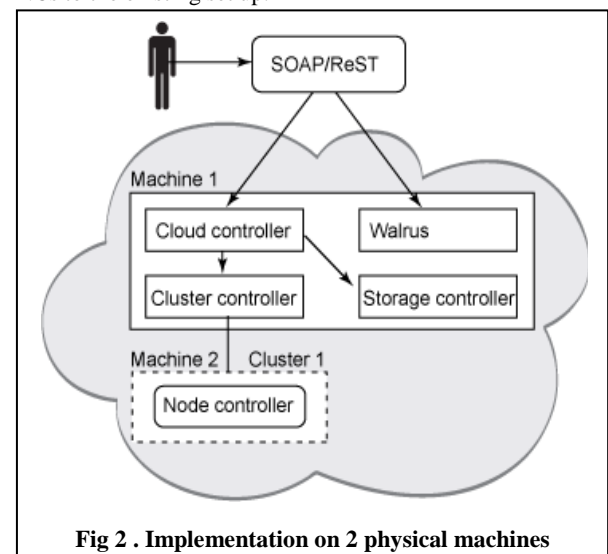

Fig 2 . Implementation on 2 physical machines

5. THE BENEFITS OF THE EUCALYPTUS EDUCATION CLOUD

The UEC Eucalyptus open source private cloud offers educational institutions the features necessary to improving the efficiency of infrastructure [17], which includes the following:

5.1 Data Center Optimization

Eucalyptus optimizes existing data center resources for an educational institution with consolidation through virtualization of all data center elements, including machines, storage, and network. Eucalyptus is compatible with most widely used virtualization technologies, including Xen and KVM hypervisors.

5.2 Automated Self-Service

Computer resource provisioning can be automated by allowing users to access their own flexible configurations of storage, machines, and networking devices as desirable through standardized web service protocols.

5.3 Web Services based

Universally accepted Web service protocols are used internally by the Eucalyptus, making its installation, operation, and maintenance similar to that of a high quality e-commerce website.

5.4 Scalable Infrastructure

UEC Eucalyptus based clouds are highly scalable [17-18], which enables an organization to efficiently scale-up or scale-down data center resources according to the needs of the it.

5.5 Elastic Resource Provisioning

This benefit allows users to flexibly reconfigure computing resources as requirements transform. This helps the organization workforce remain adaptable to sudden changes in business needs.

5.6 Open Source Innovation

The open source development community provides highly transparent and extensible, Eucalyptus open source core architecture that supports value-adding customizations and innovations [20].

6. CONCLUSION

Cloud computing is a budding business model that delivers computing capabilities over the Internet in a highly flexible and elastic manner. Education IT will persist to run applications within its own environment whilst adding new applications and services that run in the education cloud. The focus is on making solutions for the real world of hybrid IT environments by providing cost-effective software and services that support efforts to promote learning, create opportunities, and address challenges facing education institutions. Ubuntu Enterprise Cloud, a Linux distribution built around Eucalyptus, is supposed as a complete solution for a private education cloud. UEC is feasible as platform for a private cloud suitable for educational institution. In this paper, we have presented Eucalyptus, an open-source implementation of an IaaS system. This paper describes a deployment to be performed in order to compare the performance of UEC with Eucalyptus, and their suitability for meeting the challenges facing users of cloud computing. Constant rapid advances in cloud computing are leading to more complex and dependable systems for creating private clouds in academic environments. UEC, the first complete open source offering is a definite contender, which with further improvement in database performance could go beyond commercial alternatives within the near future.

7. REFERENCES

[1] Daniel Nurmi et al., "The Eucalyptus Open-source Cloud-computing System", Cluster Computing and the Grid, 2009. CCGRID '09. 9th IEEE/ACM International Symposium

[2] Jonathan S Ward, "A Performance Comparison of Clouds – Amazon EC2 and Ubuntu Enterprise Cloud", StACC: St Andrews Cloud Computing Co-laboratory, 2009

[3] S. Wardley, et al., "Ubuntu Enterprise Cloud Architecture", Canonical, USA, Technical Report, August 2009

[4] "Ubuntu - An introduction to cloud computing", 2011

[5] J. Jackson, "Dell Releases Ubuntu-powered Cloud Servers", PCWorld, 2011.

[6] "Deployment Guide - Ubuntu Enterprise Cloud on Dell Servers SE", DELL, 2011

[7] Canonical, "Ubuntu Case Study – Knowledge Tree Moves to the Cloud", 2011

[8] S. Wardley, et al., "Ubuntu Enterprise Cloud Architecture", Canonical, USA, Technical Report, August 2009.

[9] Johnson. D, et al., "Eucalyptus Beginner's Guide – UEC Edition", 2010

[10] "What software environments are supported?", https://www.eucalyptus.com/faq#q7

[11] Review: Ubuntu Enterprise Cloud (2010), http://www.networkworld.com/reviews/2010/051010-ubuntu-cloud-test.html

[12] "UEC CDInstall - Community Ubuntu Documentation", https://help.ubuntu.com/community/UEC/CDInstall

[13] "UEC topologies", https://help.ubuntu.com/community/UEC/Topologies

[14] Understanding the Remote Desktop Protocol (RDP), http://support.microsoft.com/kb/186607

[15] "VNC remote access software for desktop and mobile platforms", http://www.realvnc.com/

[16] Zoran Pantić and Muhammad Ali Babar, "Guidelines for Building a Private Cloud Infrastructure", A Technical Report, 2012

[17] "Five Steps to Enterprise Cloud Computing (2013)", A Eucalyptus white paper

[18] "Clouds provide infinite scale", https://www.eucalyptus.com/resources/cloud-overview/cloud-myths-dispelled

[19] Xiaoyu Chen, Gary B. Wills, Lester Gilbert, David Bacigalupo, " Using Cloud For Research: A Technical Review", A Technical Report, 2010

[20] Marvin Rambhadjan and Arthur Schutijser, "SURFnet cloud computing solutions", A Technical Report, 2010