

Performance Evaluation of Virtualization in a Private Cloud Test Bed

Subasish Mohapatra
Department of Computer
Science and Application
College of Engineering and Technology
Bhubaneswar, Odisha, India

Subhadarshini Mohanty
Department of Computer
Science and Engineering
College of Engineering and Technology
Bhubaneswar, Odisha, India

Arunima Hota
Department of Computer
Science and Engineering
College of Engineering and Technology
Bhubaneswar, Odisha, India

Shradha Pattanayak
Department of Computer
Science and Engineering
College of Engineering and Technology
Bhubaneswar, Odisha, India

ABSTRACT

Cloud computing is an interesting topic which is applied in the field of parallel and distributed computing. Cloud computing systems can access to large pools of data and computational resources through a variety of interfaces. Services provided by clouds are high performance computing, storage and application hosting. According to types of visibility clouds are divided into three types like: public cloud, private cloud and hybrid cloud. In case of public cloud resources are provided dynamically where as in case of private cloud resources are provided on demand. Reliability of private cloud is more. Here the paper, provide a comparative study between CPU performance of public cloud and private cloud (e.g. mycloud) .

Keywords

Cloud Computing, Virtualization, Private cloud, Eucalyptus, my cloud.

1. INTRODUCTION

The cloud is the next stage in the evolution of the Internet. It provides the means through which everything from computing power to business processes to personal collaboration is delivered to customers as a service wherever and whenever he needs it. Cloud means, an infrastructure that provides resources and services over the Internet. A storage cloud provides storage services (block or file based services); a data cloud provides data management services (record-based, column-based or object-based services); and a compute cloud provides computational services. Often these are layered (compute services over data services over storage service) to create a stack of cloud services that serves as a computing platform for developing cloud-based applications [1]. A cloud-based infrastructure is designed for data mining large distributed data sets over clusters connected with high performance wide area networks [2]. Cloud computing provides a convenient platform and also reduces the cost of the equipment required for processing huge amounts of data [3]. Cloud computing refers to the underlying infrastructure that makes it possible to scale services and resources exponentially in response to variable supply and demand. "Cloud computing is a style of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet." Users need not have knowledge of the technology infrastructure in the cloud that

supports them [4]. Different clouds offer different sorts of resources, e.g., processing, storage, management software, or application services. Clouds are typically constructed using large numbers of inexpensive machines. As a result, the cloud vendor can more easily add capacity and can more rapidly replace machines that fail, compared with having machines in multiple laboratories [1]. Different variants of cloud computing are Apache hadoop, Google Apps, Microsoft's Azure, Ubuntu enterprise cloud.

Main objectives of the paper are as follows:

- To analyze and study the performance of CPU of a virtual box using virtualization on a private cloud computing environment.
- To compare the CPU performance using cloud computing and without using cloud computing.
- To compare the Load on cloud web server and Load on non-cloud web server using Apache Benchmarking.

2. LITERATURE SURVEY

In this paper we have studied different cloud computing environments and analyzed their behavior as follows:

- Eucalyptus: Eucalyptus is an acronym for "Elastic Utility Computing Architecture for Linking Your Programs To Useful Systems". This cloud is a software platform for the implementation of private cloud computing on computer clusters. There is an open-core enterprise edition and an open-source edition. Currently, it exports a user-facing interface that is compatible with the Amazon EC2 and S3 services but the platform is modularized so that it can support a set of different interfaces simultaneously. The development of Eucalyptus software is sponsored by Eucalyptus Systems [5]. Eucalyptus works with most currently available Linux distributions including Ubuntu, Red Hat Enterprise Linux (RHEL), CentOS, SUSE Linux Enterprise Server (SLES), openSUSE, Debian and Fedora. It can also host Microsoft Windows images. Similarly Eucalyptus can use a variety of virtualization technologies including VMware, Xen and KVM hypervisors to implement the cloud abstractions it supports. Eucalyptus implements IaaS (Infrastructure as a Service) style private and hybrid clouds. The platform provides a single

interface that lets users access computing infrastructure resources (machines, network, and storage) available in private clouds-implemented by Eucalyptus inside an organization's existing data center and resources available externally in public cloud services [6]. The software is designed with a modular and extensible Web services-based architecture that enables Eucalyptus to export a variety of APIs towards users via client tools.

Currently, Eucalyptus implements the industry-standard Amazon Web Services (AWS) API, which allows the interoperability of Eucalyptus with existing AWS services and tools. Eucalyptus provides its own set of command line tools called Euca2ools, which can be used internally to interact with Eucalyptus private cloud installations or externally to interact with public cloud offerings, including Amazon EC2 [7]. Eucalyptus includes these features:

- Compatibility with Amazon Web Services API
- Installation and deployment from source or DEB and RPM packages.
- Secure communication between internal processes via SOAP and WS-Security.
- Support for Linux and Windows virtual machines (VMs).
- Support for multiple clusters as a single cloud.
- Elastic IPs and Security Groups
- Users and Groups Management
- Accounting reports.
- Configurable scheduling policies and SLAs.

Components of Eucalyptus (Private Cloud)

Cloud Controller (CLC): The CLC is responsible for exposing and managing the underlying virtualized resources (machines (servers), network, and storage) via user-facing APIs. Currently, the CLC exports a well-defined industry standard API (Amazon EC2) and via a Web-based user interface.

Walrus: Walrus implements scalable “put-get bucket storage.” The current implementation of Walrus is interface compatible with Amazon's S3 (a get/put interface for buckets and objects), providing a mechanism for persistent storage and access control of virtual machine images and user data.

Cluster Controller (CC): The CC controls the execution of virtual machines (VMs) running on the nodes and manages the virtual networking between VMs and between VMs and external users.

Storage Controller (SC): The SC provides block-level network storage that can be dynamically attached by VMs. The current implementation of the SC supports the Amazon Elastic Block Storage (EBS) semantics.

Node Controller (NC): The NC (through the functionality of a hypervisor) controls VM activities, including the execution, inspection, and termination of VM instances.

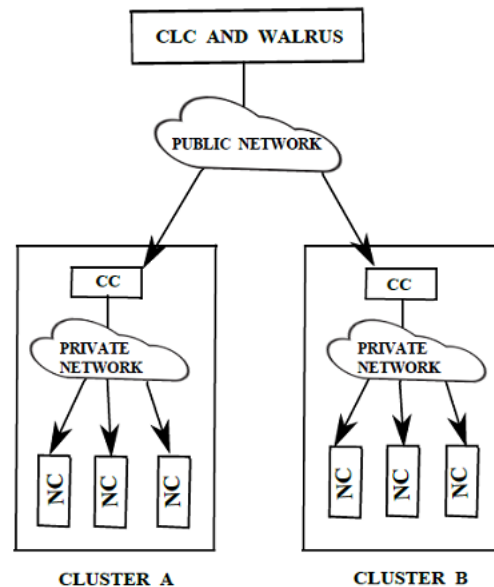


Fig 1: Eucalyptus Components [6]

- **Apache Hadoop:** The Apache Hadoop project develops open-source software for reliable, scalable, distributed computing. 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 [8]. Hadoop is a top-level Apache project being built and used by a global community of contributors, written in the Java programming language. Yahoo has been the largest contributor to the project, and uses Hadoop extensively across its businesses. The architecture of Hadoop consists of the Hadoop Common, which provides access to the file systems supported by Hadoop. The Hadoop Common package contains the necessary JAR files and scripts needed to start Hadoop. The package also provides source code, documentation, and a contribution section which includes projects from the Hadoop Community. For effective scheduling of work, every Hadoop-compatible file system should provide location awareness: the name of the rack (more precisely, of the network switch) where a worker node is. Hadoop applications can use this information to run work on the node where the data is, and, failing that, on the same rack/switch, so reducing backbone traffic. The Hadoop Distributed File System (HDFS) uses this when replicating data, to try to keep different copies of the data on different racks. The goal is to reduce the impact of a rack power outage or switch failure so that even if these events occur, the data may still be readable. A small Hadoop cluster will include a single master and multiple worker nodes [9]. The master node consists of a Job Tracker, Task Tracker, Name Node, and Data Node. A slave or worker node acts as both a Data Node and Task Tracker, though it is possible to have data-only worker nodes, and compute-only worker nodes; these are normally only used in non-standard applications. Hadoop

requires JRE 1.6 or higher. The standard startup and shutdown scripts require ssh to be set up between nodes in the cluster. In a larger cluster, the HDFS is managed through a dedicated Name Node server to host the file system index, and a secondary Name Node that can generate snapshots of the name node's memory structures, thus preventing file system corruption and reducing loss of data. Similarly, a standalone Job Tracker server can manage job scheduling. In clusters where the Hadoop Map Reduce engine is deployed against an alternate file system, the Name Node, secondary Name Node and Data Node architecture of HDFS is replaced by the file system specific equivalent [9].

- Google Apps: Google Apps is a service from Google providing independently customizable versions of several Google products under a custom domain name. It features several Web applications with similar functionality to traditional office suites, including: Gmail, Google Groups, Google Calendar, Talk, Docs and Sites. Google Apps is free and offers the same amount of storage as regular Gmail accounts. Google Apps for Business, which offers additional e-mail storage, is available for an annual fee per user account. Google Apps for Education, which is free, combines features from the Standard and Premier editions. App Engine provides a distributed data storage service that features a query engine and transactions [10]. Just as the distributed web server grows with your traffic, the distributed data store grows with your data. Users have the choice between two different data storage options differentiated by their availability and consistency guarantees. The App Engine data store is not like a traditional relational database. Data objects, or "entities", have a kind and a set of properties. Queries can retrieve entities of a given kind filtered and sorted by the values of the properties. Property values can be of any of the supported property value types. Data store entities are "schemaless". The structure of data entities is provided and enforced by user application code. The Java JDO/JPA interfaces and the Python data store interface include features for applying and enforcing structure within your app. This app can also access the data store directly to apply as much or as little structure as it needs. The data store is strongly consistent and uses optimistic concurrency control. An update of an entity occurs in a transaction that is retried a fixed number of times if other processes are trying to update the same entity simultaneously. Your application can execute multiple data store operations in a single transaction which either all succeed or all fail, ensuring the integrity of your data. The data store implements transactions across its distributed network using "entity groups." A transaction manipulates entities within a single group. Entities of the same group are stored together for efficient execution of transactions. User application can assign entities to groups when the entities are created.
- Microsoft's Azure: Microsoft's Azure Services Platform is a group of cloud technologies, each providing a specific set of services to application developers. The Azure Services Platform can be used both by applications running in the cloud and by applications running on local systems. The components of the Azure Services Platform can be used by local applications running on a variety of systems, including various flavors of Windows, mobile devices, and others. Those components include Windows Azure which Provides a

Windows-based environment for running applications and storing data on servers in Microsoft data centers. Microsoft .NET Services offers distributed infrastructure services to cloud based and local applications. Microsoft SQL Services provides data services in the cloud based on SQL Server. Live Services provides access to data from Microsoft's Live applications and others. The Live Framework also allows synchronizing this data across desktops and devices, finding and downloading applications, and more [11].

- Windows Azure: Windows Azure is simple to understand. It is a platform for running Windows applications and storing their data in the cloud. Windows Azure runs on a large number of machines, all located in Microsoft data centers and accessible via the Internet. A common Windows Azure fabric knits this plethora of processing power into a unified whole. Windows Azure compute and storage services are built on top of this fabric [12]. The Windows Azure compute service is based, of course, on Windows. For the initial availability of this service, a Community Technology Preview (CTP) made public in the fall of 2008, Microsoft allowed Windows Azure to run only applications built on the .NET Framework. The Windows Azure Platform is an application platform in the cloud that allows Microsoft datacenters to host and run applications. It provides a cloud operating system called Windows Azure that serves as a runtime for the applications and provides a set of services that allows development, management and hosting of applications off-premises. All Azure Services and applications built using them run on top of Windows Azure. Windows Azure has three core components: Compute, Storage and Fabric. As the names suggest, Compute provides a computation environment with Web Role, Worker Role, and VM Role while Storage focuses on providing scalable storage (Blobs, non-relational Tables, and Queues) for large-scale needs. Relational Database functionality is offered through SQL Azure, which is a scalable version of SQL Server that runs on the Azure platform. Fabric (Windows Azure Fabric) makes up the physical underpinnings of the Windows Azure platform as the network of interconnected nodes consisting of servers, high-speed connections, and switches. Conceptually, the repetitive pattern of nodes and connections suggests a woven or fabric-like nature. Compute and Storage components are part of the Fabric. Fabric resources and applications and services running on those resources are managed by the Windows Azure Fabric Controller service. It acts as the kernel of the Windows Azure distributed cloud operating system, providing scheduling, resource allocation, device management, and fault tolerance for the nodes in the Fabric. It also provides high-level application models for intelligently managing the complete application lifecycle, including deployment, health monitoring, upgrades, and de-activation. The Windows Azure Platform provides an API built on REST, HTTP and XML that allows a developer to interact with the services provided by Windows Azure. Microsoft also provides a client-side managed class library which encapsulates the functions of interacting with the services. It also integrates with Microsoft Visual Studio so that it can be used as the IDE to develop and publish Azure hosted applications. Windows Azure also offers Content Delivery (CDN) services as an option. The Azure CDN enables worldwide low-latency delivery of static content from

Azure Storage to end-users from 24 data centers worldwide [12].

- **Ubuntu Enterprise Cloud Architecture:** Ubuntu Enterprise Cloud (UEC) brings Amazon EC2-like infrastructure capabilities inside the firewall. The UEC is powered by Eucalyptus, an open source implementation for the emerging standard of the EC2 API. This solution is designed to simplify the process of building and managing an internal cloud for businesses of any size, thereby enabling companies to create their own self-service infrastructure [13]. Ubuntu is the only Linux distribution to position itself as a true cloud OS with three initial components of our cloud strategy already released. Two of these components are aimed at the infrastructure layer of the computing stack (commonly called IaaS or Infrastructure as a Service) while one component is aimed at the software layer (commonly called SaaS). Ubuntu is focused on building open-source software systems that will allow you and others to provide infrastructure, platform or application services. Ubuntu is not offering to host these services but instead we're defining an open-source stack (an open-source reference model) that any provider, whether internal to an organisation or external, can use to provide these services. By having multiple providers of the same service, risks to business continuity can be significantly reduced [14]. The architecture of Eucalyptus, which is the main component of Ubuntu Enterprise Cloud, has been designed as modular set of 5 simple elements that can be easily scaled such as Cloud Controller (CLC), Walrus Storage Controller (WS3), Elastic Block Storage Controller (EBS), Cluster Controller (CC), Node Controller (NC).

Issues and Challenges:

Privacy: Cloud computing poses privacy concerns basically, because the service provider at any point in time, may access the data that is on the cloud. They could accidentally or deliberately alter or even delete some info.

Compliance: In order to obtain compliance with regulations including FISMA, HIPAA, and SOX in the United States, the Data Protection Directive in the EU and the credit card industry's PCI DSS, users may have to adopt community or hybrid deployment modes that are typically more expensive and may offer restricted benefits.

Legal: As can be expected with any revolutionary change in the landscape of global computing, certain legal issues arise; everything from trademark infringement, security concerns to the sharing of propriety data resources.

Open source: Open-source software has provided the foundation for many cloud computing implementations, one prominent example being the Hadoop framework.

Open standard: Most cloud providers expose APIs that are typically well-documented (often under a Creative Commons license) but also unique to their implementation and thus not interoperable. Some vendors have adopted others' APIs and there are a number of open standards under development, with a view to delivering interoperability and portability.

Security: The relative security of cloud computing services is a contentious issue that may be delaying its adoption. Issues barring the adoption of cloud computing are due in large part to the private and public sectors unease surrounding the external management of security-based services. It is the very nature of cloud computing-based services, private or public,

that promote external management of provided services.

Sustainability: Although cloud computing is often assumed to be a form of "green computing", there is as of yet no published study to substantiate this assumption. Shifting the servers affects the environmental effects of cloud computing. In areas where climate favors natural cooling and renewable electricity is readily available, the environmental effects will be more moderate. (The same holds true for "traditional" data centers.)

Abuse: As with privately purchased hardware, crackers posing as legitimate customers can purchase the services of cloud computing for nefarious purposes. This includes password cracking and launching attacks using the purchased services.

Challenges

1. The main challenges include lack of trust in data security and privacy by users, organizational inertia, loss of governance, and uncertain provider's compliance. Trust is at the heart of the resistance that many customers have to the cloud. Concerns arise when their sensitive data and mission-critical applications move to a cloud computing paradigm where providers cannot guarantee the effectiveness of their security and privacy controls.
2. Several technical challenges related to the use of cloud computing include resource exhaustion, unpredictability of performance, data lock-in, data transfer bottlenecks, and bugs in large-scale distributed cloud systems. Low cost and computing resources available on demand are two key features of cloud computing. However, the market is becoming crowded with large providers. Because of high competition, many cloud providers over commit computing resources (eg, central processing unit [CPU] allocation, storage space, applications) to attract customers. In order to maintain the profit, they cut corners in the value-delivery system. For example, they may limit access to the cloud resources, or use out-of-date hardware or software or deploy older CPU technology. Unfortunately, most cloud customers are unable to govern the virtual architecture, and the providers usually do not permit an audit by the customers. The result is variable leading to unpredictable performance in the service. This difference between the customer's expectation and what the provider can really deliver presents a major technical challenge for the cloud customer to provide high-quality service to its own users.
3. There are many data security risks in the use of IT, such as hacker attacks, network breaks, natural disasters, separation failure, public management interface, poor encryption key management, and privilege abuse. Specific risks to cloud computing are separation failure, public management interface, poor encryption key management, and privilege abuse.

Virtualization:

Virtualization is commonly defined as a technology that introduces a software abstraction layer between the hardware and the operating system and applications running on top of it [15]. This abstraction layer is called virtual machine monitor (VMM) or hypervisor and basically hides the physical resources of the computing system from the operating system (OS). Since the hardware resources are directly controlled by the VMM and not by the OS, it is possible to run multiple (possibly different) OSs in parallel on the same hardware. As a result, the hardware platform is partitioned into one or more logical units called virtual machines (VMs). "Virtuality"

differs from “reality” only in the formal world, while possessing a similar essence or effect. In the computer world, a virtual environment is perceived the same as that of a real environment by application programs and the rest of the world, though the underlying mechanisms are formally different. Two primary benefits offered by any virtualization technology are Resource sharing and Isolation. These are classified as full virtualization, para virtualization, os virtualization, hardware virtualization etc [16].

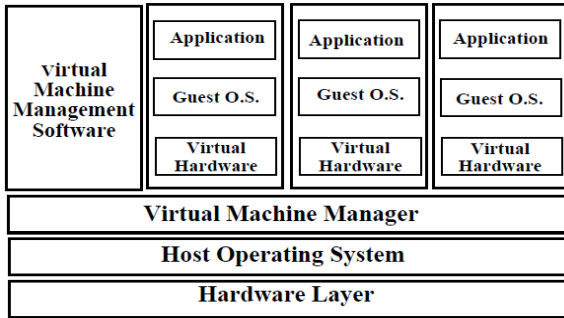


Fig 2: Diagram showing Full Virtualization

Flexibility, Availability, Scalability, Hardware utilization, Security, Reduced cost, Adaptability to Workload Variations, Load Balancing, and Legacy Applications are the major advantages of virtualization technology.

Hypervisor

Hypervisors are an important component of virtualized environments. Hypervisors are programs that allow multiple operating systems, known as guests, to run in virtual machines in an isolated fashion, and thus share a single physical machine, or host [17]. For management parallel OS systems on a machine, a virtualization layer is added between the hardware and operating system. This virtualization layer allows multiple operating system instances to run concurrently within virtual machines on a single computer, dynamically partitioning and sharing the available physical resources such as CPU, storage, memory and I/O devices [18].

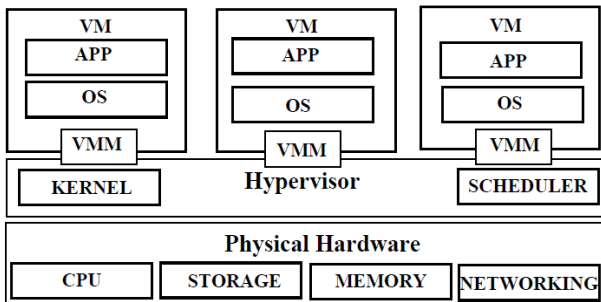


Fig 3: Virtualized Environment Representation

Observations:

From the above literature it is observed that in case of public cloud resources are provided dynamically where as in case of private cloud resources are provided on demand. Reliability of private cloud is more. However in this paper we have provided a comparative study between CPU performance of public cloud and private cloud by developing an cloud platform i.e. mycloud.

3. PROPOSED WORK

The underlying virtualized resources that comprise a Private cloud (i.e.mycloud) are exposed and managed by, the Cloud

Web Controller (CWC). The CWC is a collection of web services which are best grouped by their roles into three categories:

- Resource Services perform system-wide arbitration of resource allocations, let users manipulate properties of the virtual machines and networks, and monitor both system components and virtual resources.
- Controller Services govern persistent user and system data and provide for a configurable user environment for formulating resource allocation request properties.
- Interface Services present user-visible interfaces, handling authentication and protocol translation, and expose system management tools providing.

The Resource services process user virtual machine control requests and interact with the Cloud Web Controller to effect the allocation and de-allocation of physical resources. A simple representation of the system’s resource state (SRS) is maintained through communication with the Cloud Service Controller. The role of the SRS is executed in two stages: when user requests arrive, the information in the SRS is relied upon to make an admission control decision with respect to a user-specified service level expectation. VM creation, then, consists of reservation of the resources in the SRS, downstream request for VM creation, followed by commitment of the resources in the SRS on success, or rollback in case of errors. The middle tier of Controller Services handle the creation, modification, interrogation, and storage of stateful system and user data. Users can query these services to discover available resource information (Disk images and clusters) and manipulate abstract parameters (key pairs, security groups, and network definitions) applicable to virtual machine and network allocations.

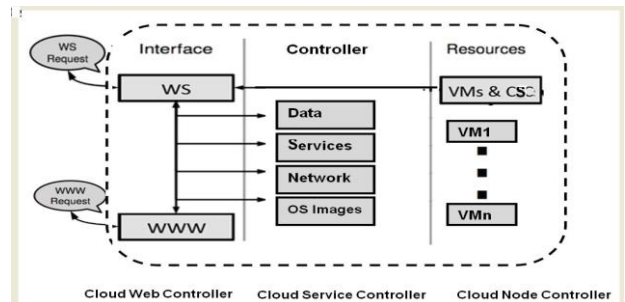


Fig 4: Service Level Architecture for Mycloud

The Resource Services interact with the Controller Services to resolve references to user provided parameters (e.g., keys associated with a VM instance to be created). However, these services are not static configuration parameters. As a result, the services which manage the networking and security group data persistence must also act as agents of change on behalf of a user request to modify the state of a running collection of virtual machines and their supporting virtual network. Lastly Cloud Node Controller contains number of virtual machines. These VMs act as resources for this service level architecture. These VMs can be accessed by the web services present in the interface [16].

Experimental Test bed Design

A private network is developed using three workstations to establish a heterogeneous cloud on the LAN. The proposed framework is used to establish a cloud with one Cloud Web Controller (CWC), one Cloud Service Controller (CC), and one Cloud Node Controller (NC).

Table 1 System Configuration with different parameters

Node Name	Type	CPU Frequency	RAM
Cloud Web Controller	Intel Pentium	2.66 GHz	4GB
Cloud Service Controller	Intel Core 2 Duo	2.66 GHz	2GB
Cloud Node Controller	Intel Pentium	2.20 GHz	2GB
Client Machine	Intel Core 2 Duo	2.4 GHz	2GB

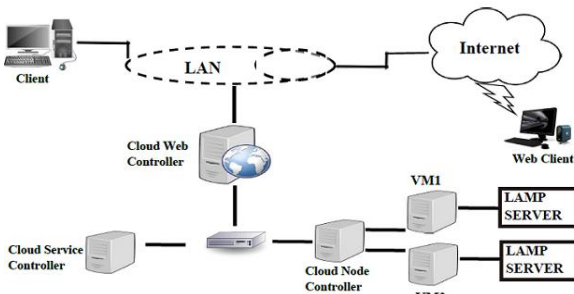


Fig 5: Experimental Testbed

System Design

The system design is divided into five sections: network design, My Cloud 1.0 installation, Software components, sample Web application, and workload generation, described below:

Network Design

The private network consists of four physical machines and one 100 Mbps Ethernet switch. Ubuntu 10.04 runs the front-end node, and Debian 5.0 (Lenny) runs installed on all other nodes. Figure 6.1 shows the basic network design. The frontend node is equipped with two Network Interface Cards (NICs). One NIC is used to connect the front-end with LAN, and another connects with the My Cloud 1.0 private network. HTTP and FTP servers are installed on the front-end node for the private network as required. Node1, Node2, and Node3 are members of the private network and are able to access the Internet using IP.

MyCloud 1.0 Deployment

My Cloud 1.0 is deployed in a on the private network as shown in Figure 5.

The design layout of my cloud is looked like:

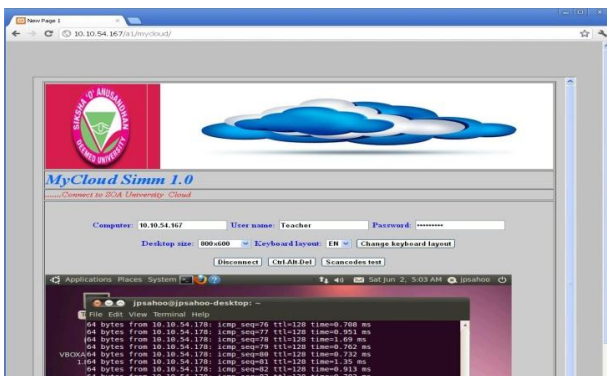


Fig 6: MyCloud Simm page layout

4. IMPLEMENTATION

The parameters which will be considered in this context are:

CPU utilization: CPU utilization is a key performance metric. It can be used to track CPU performance regressions or improvements, and is a useful data point for performance problem investigations. It is also fairly ubiquitous; it is reported in numerous places in the Windows family of operating systems, including Task Manager (taskmgr.exe), Resource Monitor (resmon.exe), and Performance Monitor (perfmon.exe). The concept of CPU utilization used to be simple. Assume you have a single core processor fixed at a frequency of 2.0 GHz. CPU utilization in this scenario is the percentage of time the processor spends doing work (as opposed to being idle).

XAMMP: XAMPP is an easy to install Apache distribution containing MySQL and PHP. XAMPP is really very easy to install and to use - just download, extract and start.

Cloud Web Server Load Benchmarking test

In this experimental performance evaluation a tool for benchmarking the cloud server is used. It is designed to give an impression of how the cloud installation is performed. This especially shows how many requests per second certain cloud service is capable to serve. Benchmarking the service in cloud can provide insight into which sections of the application might need major optimizations, and helps in dealing with scalability issues. Apache Bench is a load testing utility that allows us to simulate load on a web server. To benchmark a web server the time it will take to give a page is not important: you don't care if a user can have his page in 0.1 ms or in 0.05 ms as nobody can have such delays on the Internet. Here are steps to carry out procedure along with an example.

Apache Benchmark Procedures

- You need to use same hardware configuration and kernel (OS) for all tests
- You need to use same network configuration. For example, use 100Mbps port for all tests
- First record server load using top or uptime command
- Take multiple readings and use the best result
- After each test reboot the server and carry out test on next configuration (web server)
- Again record server load using top or uptime command
- Carry on test using static html/php files and dynamic pages

5. EXPERIMENTAL RESULTS

CPU utilization: The cpu performance utilization is being monitored with ICMP ping flood as benchmark. The percentage of CPU utilization is recorded during the ping operation on host-host and host-guest connection.

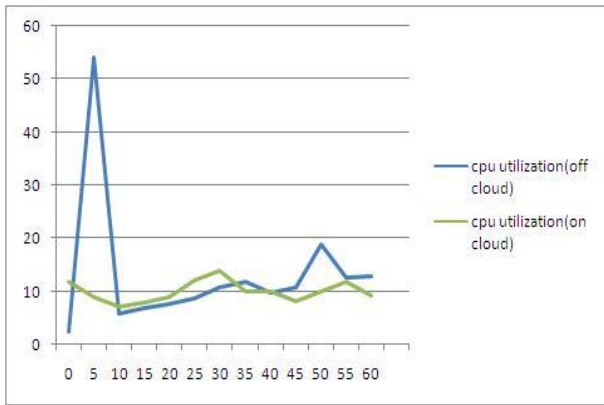


Fig 7: cpu performance using cloud service and without using cloud service

Apache Benchmarking

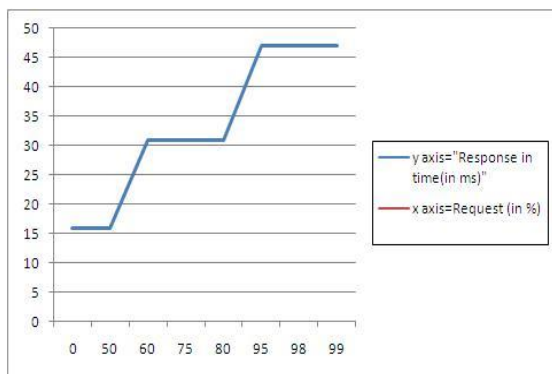


Fig 8: Load on Cloudweb server

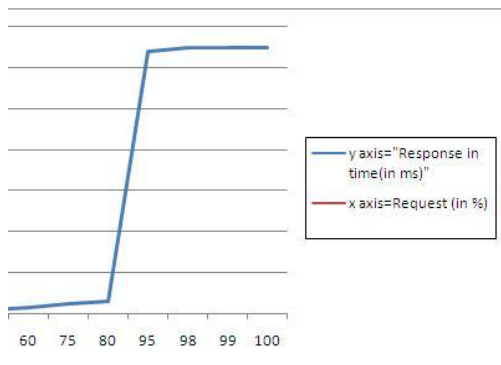


Fig 9: Load on noncloud webserver

6. CONCLUSION AND FUTURE WORK

MY CLOUDSIM system is built to allow administrators and researchers to deploy an infrastructure for user-controlled virtual machine creation and to control existing resources. Its hierarchical design targets resource commonly found within academic and laboratory settings, including but not limited to small- and medium sized Linux clusters, workstation pools, and servers. Here a virtual networking solution is used which provides VM isolation, high performance and a view of the network that is simple and flat. The performance evaluation shows the details of cpu utilization and load on cloud server using Apache benchmark. The evaluation of the virtualization with cloud computing has showed that virtualization performances are improving and tend to be as good as the native performances. Also the researches permit to outline the strengths and weaknesses of the different technologies used in

cloud computing. Due to that solutions are found to overcome the weaknesses. Also cloud computing and virtualization became more and more popular. After all the tests that have been performed on cloud computing it will be interesting to evaluate in more detail the cost of a private cloud for a company or a university. It would require evaluating the materiel needed to be able to virtualized a large number of machines and the price that it would cost. Also it would required to study the cost of solution that would be implemented such as the comparison of Open Source and commercial solution. Finally it would be necessary to evaluate the economy in terms of energy because the usage of cloud computing permit through the use of virtualization to reduce the energy consumption.

7. REFERENCES

- [1] Lal Kanhaiya and Mahanti,N.C, A Novel Data Mining Algorithm for Semantic Web based Data Cloud, International Journal of Computer Science and Security (IJCSS), Vol. 4, Issue 2, pp. 160, 2010.
- [2] Grossman Robert and Yunhong Gu, Data Mining Using High Performance Data Clouds: Experimental Studies Using Sector and Sphere, Proceedings of The 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008), ACM, pp. 920-927, 2008.
- [3] Chao-Tung Yang, Wen-Chung Shih, Guan-Han Chen and Shih-Chi Yu, Implementation of a Cloud Computing Environment for Hiding Huge Amounts of Data, International Symposium on Parallel and Distributed Processing with Applications(ISPA), 2010.
- [4] Erl Thomas, Ricardo Puttini, and Zaigham Mahmood. Cloud computing: concepts, technology, and architecture. Pearson Education, 2013.
- [5] Araujo, Jean, et al. "Software aging in the eucalyptus cloud computing infrastructure: characterization and rejuvenation." ACM Journal on Emerging Technologies in Computing Systems (JETC) 10.1 (2014): 11.
- [6] Lonea, Alina Madalina, Daniela E. Popescu, and Octavian Prostean. "A survey of management interfaces for eucalyptus cloud." Applied Computational Intelligence and Informatics (SACI), 2012 7th IEEE International Symposium on. IEEE, 2012.
- [7] Yadav, Sonali. "Comparative study on open source software for cloud computing platform: Eucalyptus, OpenStack and OpenNebula." International Journal Of Engineering And Science 3.10 (2013): 51-54.
- [8] Vavilapalli, Vinod Kumar, et al. "Apache hadoop yarn: Yet another resource negotiator." Proceedings of the 4th annual Symposium on Cloud Computing. ACM, 2013.
- [9] Murthy, Arun C., et al. Apache Hadoop YARN: Moving Beyond MapReduce and Batch Processing with Apache Hadoop 2. Pearson Education, 2013.
- [10] Patidar, Shyam, Dheeraj Rane, and Pritesh Jain. "A survey paper on cloud computing." 2012 Second International Conference on Advanced Computing & Communication Technologies. IEEE, 2012.
- [11] Buyya, Rajkumar, and Diana Barreto. "Multi-cloud resource provisioning with Aneka: A unified and integrated utilisation of microsoft azure and amazon EC2

- instances." 2015 International Conference on Computing and Network Communications (CoCoNet). IEEE, 2015.
- [12] Nodehi, Tahereh, Sudeep Ghimire, and Ricardo Jardim-Gonçalves. "Toward a unified intercloud interoperability conceptual model for IaaS cloud service." *Model-Driven Engineering and Software Development (MODELSWARD)*, 2014 2nd International Conference on. IEEE, 2014.
- [13] West, Andrew. "Ubuntu and business ethics: Problems, perspectives and prospects." *Journal of Business Ethics* 121.1 (2014): 47-61.
- [14] West, Andrew. "Ubuntu and business ethics: Problems, perspectives and prospects." *Journal of Business Ethics* 121.1 (2014): 47-61.
- [15] Castrucci, Marco, et al. "Key concepts for the Future Internet architecture." *Future Network & Mobile Summit (FutureNetw)*, 2011. IEEE, 2011
- [16] Sahoo, Jyotiprakash, Subasish Mohapatra, and Radha Lath. "Virtualization: A survey on concepts, taxonomy and associated security issues." *Computer and Network Technology (ICCNT)*, 2010 Second International Conference on. IEEE, 2010.
- [17] Desai, Ankita, et al. "Hypervisor: A survey on concepts and taxonomy." *International Journal of Innovative Technology and Exploring Engineering* 2.3 (2013): 222-225.
- [18] Khan, Ashiq, et al. "Network virtualization: a hypervisor for the Internet?." *IEEE Communications Magazine* 50.1 (2012): 136-143.