Thesis no: MSCS-2015-13

A framework to migrate and replicate VMware Virtual Machines to Amazon Elastic Compute Cloud

Performance comparison between on premise and the migrated Virtual Machine

Rajesh Bachu

Faculty of Computing
Blekinge Institute of Technology
SE–371 79 Karlskrona, Sweden

This thesis is submitted to the Faculty of Computing at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science in Computer Science. The thesis is equivalent to 20 weeks of full time studies.

Contact Information:
Author:
Rajesh Bachu
E-mail: rajeshbachu99@gmail.com

External advisor:
Ashok Ramu
Sr. Director Engineering,
Actifio Inc.

University advisor:
Sogand Shirinbab
Dept of Computer Science and Engineering

Faculty of Computing
Blekinge Institute of Technology
SE–371 79 Karlskrona, Sweden

Internet : www.bth.se
Phone : +46 455 38 50 00
Fax : +46 455 38 50 57

Abstract

Context Cloud Computing is the new trend in the IT industry. Traditionally obtaining servers was quiet time consuming for companies. The whole process of research on what kind of hardware to buy, get budget approval, purchase the hardware and get access to the servers could take weeks or months. In order to save time and reduce expenses, most companies are moving towards the cloud. One of the known cloud providers is Amazon Elastic Compute Cloud (EC2). Amazon EC2 makes it easy for companies to obtain virtual servers (known as computer instances) in a cloud quickly and inexpensively. Another advantage of using Amazon EC2 is the flexibility that they offer, so the companies can even import/export the Virtual Machines (VM) that they have built which meets the companies IT security, configuration, management and compliance requirements into Amazon EC2.
Objectives In this thesis, we investigate importing a VM running on VMware into Amazon EC2. In addition, we make a performance comparison between a VM running on VMware and the VM with same image running on Amazon EC2.
Methods A Case study research has been done to select a persistent method to migrate VMware VMs to Amazon EC2. In addition an experimental research is conducted to measure the performance of Virtual Machine running on VMware and compare it with same Virtual Machine running on EC2. We measure the performance in terms of CPU, memory utilization as well as disk read/write speed using well-known open-source benchmarks from Phoronix Test Suite (PTS).
Results Investigation on importing VM snapshots (VMDK, VHD and RAW format) to EC2 was done using three methods provided by AWS. Comparison of performance was done by running each benchmark for 25 times on each Virtual Machine.
Conclusions Importing VM to EC2 was successful only with RAW format and replication was not successful as AWS installs some software and drivers while importing the VM to EC2. Migrated EC2 VM performs better than on premise VMware VM in terms of CPU, memory utilization and disk read/write speed.

Keywords: Amazon Elastic Compute Cloud, Cloud Computing, Virtualization, Virtual Machine, VMware.

Acknowledgments

I thank my supervisor Sogand Shirinbab for providing advice and feedback on thesis and academic writing which no doubt helped in improving the thesis. I also extend my thanks to our examiner Martin Boldt for his efforts throughout the course.

List of Figures

List of Tables

Abbreviations

API	Application Programming Interface
AWS	Amazon Web Services
CLI	Command Line Interface
CPU	Central Processing Unit
EBS	Elastic Block Storage
EC2	Elastic Compute Cloud
ESX	Elastic Sky X
OS	Operating System
PTS	Phoronix Test Suite
RHEL	Red Hat Enterprise Linux
VHD	Virtual Hard Disk
VM	Virtual Machine
VMDK	Virtual Machine Disk

Contents

Chapter 1

Introduction

Cloud Computing is the new trend in the IT industry. Instead of spending many resources to maintain individual data center, companies can easily migrate to cloud by moving their data. The resources to maintain individual data center may include the process of thinking in which hardware to buy. Getting access after purchasing the servers may take some time. Companies also need to have technical staff to maintain these servers and it is very expensive to expand or change the configuration of the hardware. Due to these reasons the enterprises are moving to cloud. Moving to cloud provides many advantages as it is easy to access the infrastructure quickly and inexpensively. Amazon Elastic Compute Cloud (EC2) is an example for Infrastructure as a Service (IaaS) cloud provider. The intention of Cloud Computing is to achieve higher throughput and to tackle large scale computation problems by making better use of distributed resources putting them together [1]. The five essential characteristics of Cloud Computing are on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service. The three service models are Software as Service, Platform as a Service and Infrastructure as a Service. The four deployment models are private cloud, community cloud, public cloud and hybrid cloud. The objective of Cloud Computing is to bring software, information and shared resources as a service over a network which contributes benefits such as scalability, good performance, high availability, more mobility at low cost [2].

Virtualization forms the important block in the solutions of CC because it effectively uses the resources and reduces the cost. The main goal of virtualization is to manage the workload by transforming traditional computing to make it more scalable, efficient and economical. Virtualization can be applied at OS, hardware and server levels. Virtualization technology reduces hardware costs, energy and is rapidly transforming the fundamental way of computing [3]. Virtualization has quickly obtained dominant position in IT enterprises by delivering transformative cost savings and developing operational efficiency, flexibility and IT service levels. Virtualization is basically making a virtual image or version of a server, OS, storage devices or network resources so they can be used on multiple machines at same time. As virtualization technology is increasing day by day, companies

like Intel and AMD have seperately developed virtualization extension to the x86 hardware architecture [4].

Virtualization has enabled the migration of Virtual Machines by virtualizing the hardware layer. VM migration means moving a VM from one physical host to another. The three different types of VM migration are named as hot, cold and live migration. In live migration a VM can be migrated from source to destination when the VM's OS is running. In cold migration the VM is shutdown at the source host before migration. After migration, the VM is powered on at the target host. In Hot migration the hosted OS is suspended until the migration completes. Hot migration is efficient than cold migration because the run state of the OS is preserved and is less seamless than live migration [5].

1.1 Problem Statement

A lot of IT enterprises want to leverage the cloud for running their applications. The primary reason is that they want to reduce the capital expenditure incurred buying hardware and lab space etc. To facilitate this movement the enterprises are migrating to cloud. A case study was conducted in migrating an enterprise IT system to IaaS (Amazon EC2) and concluded that the system infrastructure in the case study would have cost 37% less over 5 years on EC2, and using cloud computing cloud have potentially eliminated 21% of the support calls for their system [6]. According to these results, companies would significantly benefit from migration of their systems to the cloud, however there are significant risks associated with this. These risks need to be identified and avoided. There are few studies focusing on migrating systems from non-virtualized environments to the cloud [6] but there are no research papers found in migrating and replicating VMs from VMware environment to Amazon EC2 except the guidelines provided by Amazon Web Services (AWS) for importing the VMs to EC2. However the hypervisor on EC2 is very different from hypervisor on VMware since EC2 uses Xen hypervisor and VMware uses ESXi. Companies can replicate VMs from VMware infrastructure to EC2 but due to different hypervisors VMware VMs are unable to run on EC2 [7]. There is no literature found in scheduled replication of VMware VMs to EC2 VM. The performance of the migrated VM may vary when compared to the on premise VM as the migrated VM runs on cloud platform.

This thesis is focused on finding a persistent method to migrate and replicate a backup of VMware Virtual Machine (VM) to Amazon EC2. VM migration is a process of moving VM from one physical server to another physical server. The advantages of VM migration includes Disaster Recovery, OS migration: as migration of VM is easier than migrating OS and applications independently, changing workloads in the server: instead of adding additional servers manually if workload

in a server is high. Then after finding a persistent method to migrate a backup of VMware VM to EC2, a trial on replicating on premise VM to migrated VM is done. Then comparison of the performance is made in terms of CPU, memory utilization and disk write/read speed between on premise VM to the migrated VM by using five different open source benchmarks Apache [8], John the Ripper [9], Stream [10], OpenSSL [11] and unpack-linux [12] which were selected from an open-source well known Phoronix Test Suite(PTS) [13].

1.2 Research Questions

1. How does the VMware VM perform when it is migrated to EC2 platform in terms of CPU, memory utilization and disk read/write speed?

   To answer this Research Question we need to find answer to these research questions:

   1.1. How to take a backup of VMware Virtual Machine and import it to EC2?

   1.2. What are the challenges in automatically importing and replicating the snapshots of VMware Virtual Machines to EC2?

   1.3. Which Virtual Machine has better performance, a Virtual Machine running on VMware or a Virtual Machine running on Amazon EC2 in terms of CPU, memory utilization and disk read/write speed?

1.3 Aim and Objectives

Aim: The aim of this research is to figure out how does the VMware Virtual Machine perform when it is migrated to EC2 platform. In order to reach this goal we need to figure out a method to migrate and replicate a snapshot of on premise VMware Virtual Machine to Amazon EC2 and run the snapshot on EC2. Then compare the performance of on premise VMware virtual machine to migrated EC2 VM in terms of CPU, memory utilization and disk read/write speed.

Objectives:

- Learn the state of practice method for taking a backup of VM.
- Find out the methods to import backups of VMs to EC2.
- Select a persistent method to migrate and run a backup of VM to EC2.
- Investigate the challenges in replicating VMware VM to EC2.

- Compare the performance of on premise VMware VM to the migrated EC2 VM.

1.4 Thesis Outline

The thesis is outlined as follows.
Chapter 2 covers background information regarding Cloud Computing, Virtualization, Hypervisors and the tools used in this thesis. Related works were mentioned in Chapter 3. In Chapter 4, case study and experiment methods were discussed. Results are reported and analyzed in Chapter 5. Discussions were mentioned in Chapter 6. Conclusions and Future work were mentioned in Chapter 7. References and appendix were mentioned after Chapter 7.

Chapter 2

Background

2.1 Cloud Computing

The National Institute of Standards and Technology (NIST) defines Cloud Computing as *"a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction"* [2].

The three service models of Cloud Computing are [2]:

- Software as a Service (SaaS): The capacity provided to the customer to use the provider's applications running on a cloud infrastructure. E.g., salesforce.com
- Platform as a Service (PaaS): The capacity provided to the customer to deploy consumer-created or acquired applications onto the cloud infrastructure. E.g., Google Apps.
- Infrastructure as a service (IaaS): The capacity provided to the customer is to provision processing, storage, networks and other fundamental computing resources where the customer is able to deploy and run arbitrary software like operating system and applications. E.g., Amazon EC2.

The four deployment models of Cloud Computing are [2]:

- Private Cloud: The cloud infrastructure is provisioned for exclusive use by single organization comprising multiple consumers (e.g., business units).
- Community Cloud: The cloud infrastructure is provisioned for exclusive use by specific community of customers from organization that have shared concerns (e.g., mission, security requirements).
- Public Cloud: The cloud infrastructure is provisioned for open use by general public. It exists on the premises of the cloud provider.
- Hybrid Cloud: The cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public).

2.1.1 Cloud Providers

As internet based Cloud Computing has gained tremendous change in recent times, the customers are outsourcing their storage and computations to public cloud providers as they pay for the service used on demand. Migrating to cloud provides new advantages such as cost and reliability as they don't need to invest lots of money on in-house infrastructure before launching their services [14]. Moreover the companies can save their costs as they don't need to maintain technical staff for managing their network infrastructure. As number of companies using cloud for storage has increased, the cloud providers such as AWS, Google and Microsoft are actively participating in providing services to make companies make use of their clouds [14].

AWS has been selected as the cloud provider in this thesis as it provides many services at lower costs. AWS has been named as the leader in the IaaS magic quadrant report for the fifth consecutive year. Gartner, an American IT research and advisory firm providing technology related insight has placed AWS in leaders quadrant and rated AWS high for having furthest completeness of vision and highest ability to execute [15].

2.1.2 Services provided by AWS

AWS is a collection of remote computing services which are also called as web services. AWS provides number of services such as compute, storage and content delivery, database, administration and security etc., but the most popular services are compute, storage and content delivery. By using AWS Elastic Cloud Compute (EC2), we can create virtual servers in the cloud and by using Simple Storage Service (S3), we can create scalable storage in the cloud [16].

Elastic Cloud Compute

Amazon EC2 provides scalable computing capacity in AWS. By using EC2 the companies can save their costs as they don't need to invest lots of money in the upfront hardware and using EC2 is very easier as it is easy to develop and deploy applications faster. It is very easy to launch as many virtual servers as required, configure security and networking, and it is also very easy to manage storage [16]. Some of the features provided by EC2 are:

- Elastic Block Storage (EBS) volumes which helps in storing persistent data. It provides persistent block level storage volumes for use with Amazon EC2 instances in the AWS Cloud. The EBS volumes are automatically replicated with its Availability Zone which helps in protecting from component failure and by offering high availability and durability.

- Virtual computing environments, known as instances where we can launch, start, stop or terminate instances as per requirements. Number of EBS volumes can be attached/detached to/from the instance as required.
- Virtual Private Cloud (VPC) is a security feature provided by AWS for EC2 users. By using VPC we can launch AWS resources in a virtual network that we define. e.g., we can create a public-facing subnet for the webserver that has internet access and place the entire backend systems in a private facing subnet without internet access.
- Amazon Machine Images (AMI's) are pre-configured templates of the instances which consists of OS and some other additional software. Basic AMI's are provided by AWS and the users can create/sell/buy AMI's.
- Key pairs which helps in secure login to the instances.
- Instance store volumes which are also known as storage volumes and these volumes are deleted when instances are terminated.
- Regions and Availability Zones (AZs) in which resources can be stored in multiple physical locations.
- Security groups by which user can restrict access to their instances.
- Elastic IP address by which static IP address can be set for dynamic computing needs.

2.2 Virtualization

The Storage Networking Industry Association (SNIA) defines Virtualization as *"the act of abstracting, hiding, or isolating the internal functions of a storage (sub)system or service from applications, host computers, or general network resources, for the purpose of enabling application and network-independent management of storage or data"* [17].

VMware defines Virtualization as *"an abstraction layer that decouples the physical hardware from the operating system to deliver greater IT resource utilization and flexibility"* [18].
The three techniques of Virtualization are [5]:

(i) Full Virtualization: In full Virtualization, the OS is not aware that it is virtualized and thus it requires no modification. The aim or this technique is to create VMs which performs as the real machines.

(ii) Para Virtualization: The difference between para virtualization and full virtualization is that the unmodified OS is not aware of being virtualized and sensitive OS calls are trapped using binary translation.

(iii) Hardware assisted Virtualization: In this type of virtualization, the guests are allowed to run unmodified OS by using special features offered by hardware.

2.2.1 Virtual Machine

Virtualization is the enabler technology for cloud component, which uses a physical resource such as server and divides it into virtual resources called virtual machines [19]. Virtual Machine was invented by IBM in 1960's as a technique of tremendously expensive mainframe hardware [18]. IBM defined virtual machine as *"a fully protected and isolated copy of underlying physical machine hardware"*. While writing and testing any programs on VMs there is no fear of crashing the physical machine and it also doesn't cause any harm for other users [18].

VMware defines VM *"as a representation of a physical machine by software. It has its own set of virtual hardware upon which an operating system and applications are loaded"*. The traditional virtual machine system is diagrammatically represented in Figure 2.1 [18]. Virtual Machine Monitor (VMM) is a software layer that powers the hardware and it creates VMs which behaves like a physical machine that can run its own Operating System (OS). VMware Workstation brings such mainframe-class virtual machine technology to PC-based desktop workstation computers [18].

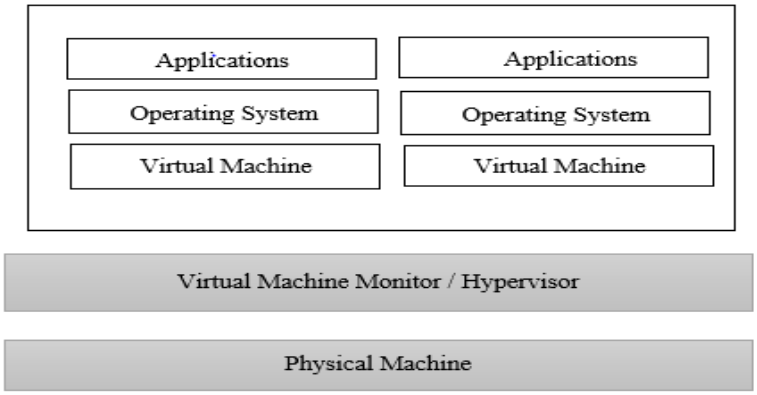

Figure 2.1: Traditional Virtual Machine System

2.2.2 Hypervisor

A hypervisor or Virtual Machine Monitor (VMM) is a software layer that powers the hardware and creates VMs which behaves like a physical machine that can run its own Operating System. Hypervisor will be on the top of the hardware and the VMs, applications and OS will be running on the top of the hypervisor[20]. Hypervisors are of two types. Type 1 and Type 2 hypervisor.

- Type 1 (native or bare metal) hypervisor run on host's hardware and manage the guest operating system [20]. E.g., Citrix Xen Server and VMware ESXi.
- Type 2 (hosted) hypervisor run on OS and OS runs on hardware. The other VMs will run on the top of the hypervisor [20]. E.g., VMware Workstation and Virtual Box

Different hypervisors are described in detailed below:

1. VMware ESXi (Elastic Sky X) is an enterprise-class Type-1 hypervisor developed by VMware for deploying and serving virtual computers. VMware ESXi is the next-generation hypervisor, providing a new foundation for virtual infrastructure [21].
2. Xen is the type 1 hypervisor which is available for open source and makes many instances of an OS or different OS run on a single host. Xen hypervisor is used for server virtualization, IaaS, desktop virtualization, security applications, embedded and hardware appliances. Amazon EC2 run their VMs on Xen Hypervisor [22].

2.2.3 VMware vSphere

VMware vSphere is VMware's Cloud Computing virtualization OS. It runs on ESXi (type 1) hypervisor. vSphere controls all the VMs running on the top of host and cluster through vCenter server. All these VMs will be running on ESXi hypervisor.
Workstation or virtual box were not selected as they run on Type 2 hypervisor. The advantage of Type 1 hypervisor is it provides high performance, availability and security compared to Type 2 hypervisor.
Each vCenter Server system manages multiple ESX hosts. We can run vSphere client and vSphere web access on multiple workstations [23]. The main components of VMware vSphere are:

- VMware ESX provides a virtualization layer that abstracts the processor, memory, storage and networking resources of the physical host into multiple VMs.

- vSphere client installs on windows machine and is the primary method of interaction with VMware vSphere. This acts as a console to operate VMs as an administration interface into the vCenter system and ESX hosts.
- VCenter server acts as a central administration point for ESX hosts connected on a network. The vCenter server is the working core of vCenter.
- The VMs will be running on Host and Cluster.

2.3 Snapshot

Snapshot is a context of storing the current disk state. Snapshot is defined as an *"ability to record the particular state of storage device at the given moment and preserve the snapshot as a guide for restoring the storage device in the event that it fails"* [24] A snapshot primarily creates point-in-time copy of the data. The snapshot copy can be taken instantly and made available for use by other applications for data protection, data analysis, data reporting, and data replication. As the original copy of the data is available as usually, snapshot copy is used to perform other functions of data. Snapshots helps in application availability, faster recovery, and easier backup management of large amounts of data and reduces exposure to data loss [24].

The tools for taking snapshots are:

- Snapshot in vSphere Client [23]: Snapshots can also be taken in vSphere Client. In vSphere, snapshots capture the entire state of virtual machine at the time we take the snapshot. Snapshot operation can be performed when a VM is powered on, powered off, or suspended. As vSphere client doesn't take snapshots in incremental blocks, it will be hard for replicating the VM by migrating the full snapshot to AWS. Thus vSphere snapshots are not used.
- IBM Spectrum Protect [25]: IBM provides Spectrum Protect Snapshot which enables advanced data backup and data recovery for virtual, physical and cloud software defined environments. It takes the backup in incremental blocks for the second backup of same VM, but it is not open source and is not available for free. Thus this type of snapshot was not selected.
- Actifio CDS [7]: Actifio CDS is an enterprise-class physical appliance that supports broad range of applications, host operating systems, third-party storage, and is designed to manage and protect nearly a petabyte of application data per system. CDS integrates seamlessly into any data center, leveraging either the IP network of Fibre-Channel SAN to capture data. Actifio CDS is used in this thesis to take snapshots of the VMs. Actifio

CDS is controlled from Actifio desktop (GUI).
The reason for selecting Actifio CDS is: for the first snapshot Actifio captures application data in its native format. After the first capture, Actifio ingests only unique changed blocks which helps in minimizing data movement, dramatically accelerating time to capture and reducing the burden on the production application. The only change blocks are applied to virtual copies via incremental merge which helps in delivering instantly accessible point-in-time data, without consuming the equivalent storage that would be required for physical copies [7]. Incremental blocks also helps in replication of VMs as replicating the incremental blocks is easier than replicating the entire snapshot. The amount of data to be replicated by incremental snapshot will be less when compared to the entire snapshot copy. The other reason is Actifio CDS was available to use as the case study in this thesis was conducted at Actifio.

Chapter 3

Related Work

Since the last few years companies like Twitter and Animoto have used clouds to construct highly scalable systems. Not only startups, even enterprise companies are attracted to cloud-based services as cloud providers market their services as being superior to in-house data centers in terms of financial and technical dimension as they are most cost effective, almost equally or more reliable and highly scalable [6].

Cloud computing has reached popularity in recent times and developed as a major trend in IT. While cloud research agenda has been pushed by industry in a high speed, academia has only recently joined as it can be seen through the sharp rise in workshops and conferences focusing on Cloud Computing. There were several white papers and general introduction to cloud computing, which provide an overview of the field, but no systematic review of the agenda has been taken by the academia yet [26].

Ali Khajeh Hosseini, D. Greenwood, and I. Somerville performed a case study of migrating an enterprise IT system to IaaS. This case study discusses about the implications of Cloud Computing from an enterprise of organizational perspective [6].

M. Jensen et al., S. Suashini et al, and B.R Knadukuri et al. have conducted surveys to find the security issues in the cloud [27][28][29] and A. Li et al. have compared the public cloud providers to migrate the storage [14], but no literature papers were found on how to migrate or replicate the VMs and make them run on EC2.

T. Mohammad and C.S Eati studied the performance of VM live migration over WAN but VM migration is limited to same hypervisor [5]. Y. Ashino and M. Nakae have discussed VM migration between different hypervisors, but their method succeeds only in VM migration between hypervisors such as VMware, Virtual Box and VirtualPC [30] but their study doesn't include migrating VMs from VMware ESXi hypervisor to EC2 Xen hypervisor.

S.Shirinbab, L.Lundberg and D.Ille have made a performance comparison of KVM, VMware and Xen Server using a Large Telecommunication Application [31] and similar to this Siavash Outadi and Jana Trchalikova have made a performance comparison of KVM and Xen hypervisors for telecommunication services and concluded that KVM demonstrated better performance in CPU utilization and response time. Xen has shown less down time in comparison to KVM which makes it suitable for large environments where maintenance, fault tolerance and manageability are essential [20].

P. Mehrotra et al. have made a performance evaluation of Amazon EC2 for NASA High Performance Computing (HPC) applications to evaluate the suitability of using Amazon's HPC cloud offering for NASA's HPC workload. They have concluded that Amazon EC2 have made great strides in HPC offering including reducing the overhead of its Virtualization layer and they also concluded that it cannot compete with HPC systems such as Pleiades where communication performance is important[32].
Sayaka Akioka and Yochi Muraoka have evaluated computational performance through some experiments with several HPC benchmarks and estimated the operational cost [33].
Keith R.Jackson et al., have made a performance analysis of HPC applications on the AWS cloud and concluded that EC2 is six times slower than a typical mmid-range Linux cluster, and twenty times slower than a modern HPC system [34].

Shiori Toyoshima et al., have made a storage access optimization with VM migration and basic performance analysis of Amazon EC2 and confirmed that execution time becomes shorter with the VM migration (cost of migration is also taken into account), compared with the case of accessing data over a network when a I/O-intensive application is executed [35]. But this thesis compare the performance between on premise and the migrated VMs.

Charles David Graziano had made a performance analysis of Xen and KVM hypervisors for hosting the Xen Worlds Project by running the open source benchmarks provided by Phoronix Test Suite [36]. Phoronix Test Suite has been also used by many other authors to benchmark performance in their their work [37, 38, 39, 40, 41], but this thesis compares the performance of the same Virtual Machines running on different hypervisors by running the same open source benchmarks using Phoronix Test Suite [13].

Chapter 4

Method

The main goal of this study is to figure out how does a VMware VM performs when it is migrated to Amazon EC2. In order to reach this goal, a case-study research has been done to select a persistent method to migrate VMware VMs to Amazon EC2. In this case-study we investigate different methods to import VMware's virtual machine into EC2. In addition to the case-study research, we have also conducted an experimental research to measure the performance of a virtual machine running on VMware and compare it with the same virtual machine running on EC2.

Motivation of the selected research methods

Case study type of research method is selected for investigating the challenges in migrating and replicating VMware VMs to Amazon EC2 as it is an empirical enquiry that investigates a contemporary phenomenon in depth and within its real-life context [42].

Experiments type of research method is selected for comparing the performance between the on premise and the migrated VM as the factors that are going to be examined such as CPU, memory utilization and disk read/write speed were already known before conducting the experiment.

Survey is not selected as they concentrate on opinions of the practitioners. The factors affecting the migration of VMs or the performance of the migrated VM may be unknown to the practitioners as there was no literature found in investigating the challenges in automatically replicating the VMs to EC2 and regarding the performance of the migrated VM.

4.1 Case Study

The main goal of this case study is to investigate the migration methods such as Management Portal for vCenter, EC2 import instance and EC2 import image provided by AWS to migrate VM backups to Amazon EC2, and figure out the

challenges faced while migrating and replicating on premise Virtual Machines to Amazon EC2. This case helps in finding the persistent method to migrate and replicate on premise Virtual Machines to Amazon EC2.

Objectives:

- Find how the backups of Virtual machines are taken
- Figure out a persistent method to migrate and run a backup of VMware VM to EC2 by investigating all the methods provided by AWS
- Investigate the challenges in migrating and replicating the Virtual Machines.

4.1.1 Case and Unit of Analysis

This case study is conducted at Actifio [7], a Copy Data Virtualization company. Copy data comprises of multiple duplicates of same file. Copy data is obtained from multiple copies of volumes, backups, online copies for disaster recovery, etc. In a study IDC found that 75 % of storage is consumed by copy data [7]. The idea of copy data virtualization is to maintain one golden copy of the data and making it available to company without needing to store a separate copy for each use. In simple words, Actifio takes snapshot of a VM, later it only captures the incremental blocks of the VM and merges it to main snapshot. This snapshot copy is stored either at the same location or at the remote location [43]. But there is no method that migrates, runs and replicates the backup of golden copy to EC2. So the goal is to migrate this backup to EC2 and make the backup run on EC2. Then replicate the on premise running VM to the migrated VM. To make this possible, a VM is to be migrated and run on Cloud and replication is to be made.

In this case study the case is the Actifio's Copy Data Virtualization and unit of analysis is the VM migration and replication to EC2 using the snapshots taken from Actifio CDS. The unit of analysis is to investigate migration of on premise VM to EC2 by following the guidelines provided by AWS. To investigate VM migration, a backup of VM is taken using Actifio CDS. The procedure for taking backups of VMs was gained by conducting participant observations and informal discussions. Then backups are migrated to EC2 by following the guidelines provided by AWS.

Methods to migrate VM to EC2

VMware provides API to migrate VMs from one host to other host. By using vMotion, a VM can be migrated to specific ESX host only. MigrateVM operation helps in migrating VM's .vmx configuration file from one host to another without

moving its virtual disk file, but VMs can be imported to EC2 only by using the guidelines provided by AWS. Thus the API provided by VMware were not used in this thesis. [44, 16].

The tools available to import VM from existing environment to EC2 are

Management Portal for vCenter [16]
The AWS Management Portal for vCenter enables to manage AWS resources using VMware vCenter. Installing this plug-in inside the existing vCenter environment provides benefit of importing VMware VMs to EC2 instances. The instances can be imported to any AWS region. Following steps are followed to import VMware VM to EC2:

1. VM was created in vCenter.
2. The created VM was selected in vCenter inventory.
3. Guest OS was configured in a way that it can get its IP address from DHCP client.
4. VM is powered off.
5. Right click on the VM then Click migrate VM to EC2.
6. The guest OS running on the VM, AWS region, environment, subnet, private IP (optional), security group (they limit access to particular network ports) were selected.
7. Import task id was displayed by clicking on migrate to EC2.
8. After migration, newly created instance was shown by the port. The instance was started by clicking on start the instance button.

The other two methods to import VM to EC2 are EC2 Import Image and EC2 Import Instance. EC2 Command Line Interface tools and Java Development Kit are installed before using these methods. Before implementing these methods, the instructions to prepare the VM and the prerequisites were followed. The procedure for installing EC2 CLI tools and the prerequisites for import VM to EC2 are mentioned in Appendix A

EC2 Import Image [16] creates an import image task by importing the VM as an Amazon Machine Image (AMI) on the selected region and VM is launched from the AMI. To import VM using EC2 import image, AWS CLI was installed, the instructions to prepare the VM was followed, then the VMware VM was exported from its virtual environment by following the guidelines provided by VMware in exporting an OVF Template. Then the guidelines provided by AWS

were followed in importing the VM into Amazon EC2 [16].

EC2 Import Instance [16] creates an import instance task using the metadata from the specified disk image, and imports the image to Amazon EC2.

The VM migration process was started with EC2 import image then with EC2 import instance. To import VM using ec2-import-instance, a snapshot of VM is required. Hence participant observations are conducted to know the state of practice of taking snapshots of VMs at Actifio. Then the investigation is made by importing the VM using the mentioned methods.

Before migration, a VM is created on vCenter and windows 2008 R2 OS is installed in it. Then the VM was discovered on Actifio desktop and a template and profile were created, then the VM was protected using Actifio CDS. Then snapshot is taken and retrieved in a VMDK, VHD and RAW format. Now the snapshot files were stored on the physical machine and investigation was made. This procedure was gained from participant observations.

As EC2 accepts importing of VMs in three formats, the snapshots were taken from Actifio CDS in three formats i.e., VMDK, VHD and RAW formats. VMDK stands for Virtual Machine Disk. VMDK is a file format that describes containers for virtual hard disk drives to be used in virtual machines. VHD stands for Virtual Hard Disk. VHD is a file format which represents a virtual hard disk drive. It may contain what is found on a physical HDD, such as disk partitions and a file system, which in turn can contain files and folders. It is typically used as the hard disk of a VM. RAW is the raw format of the snapshot.

Replicating VM
Replicating VM feature helps in synchronizing and maintaining a scheduled replicated copy of the VMware VM on EC2. That means if a disaster occurs, the VM will be ready to use on remote location i.e, on EC2. This can be achieved by taking scheduled snapshots of VMware VM and rewriting it to the root volume of EC2 instance.

Replicating VM makes it easy when disaster strikes and we need to fail over to another datacenter or we need to test Disaster Recovery plans. By replicating VM, we can easily run a copy of the failed VM in a remote location which is already in a ready state. Further we can run disaster recovery tests as often as needed.
These steps were proposed and implemented to achieve Replicated VM functionality.

- Firstly EC2 Import Instance task will be made by migrating a snapshot of

VM in RAW format to EC2. Let us consider it as image 1. So now the same copy of VMware VM is on EC2. By doing this task, a VM from VMware environment will be imported to S3 bucket and turns into an instance on EC2.

- A dummy VM is launched on EC2. In this VM a folder named Rajesh in /dumps location is created.
- For the second time the snapshot of the VMware VM (consider as Image2) which contains extra files than Image 1 is taken and transferred it to the dummy VM using SCP.
  Syntax:

  ```
  scp -P 26 /rajesh/home/Desktop/image2.raw root@
  <Destination IP (Dummy VM)>:/dumps/rajesh
  ```

- Then the root volume of the EC2 VM was detached and attached to the dummy VM as a secondary volume and named it as /dev/xvdaa.
- Then the latest RAW format image (image2.raw) is copied to the secondary volume using dd and detached the volume and attached it back to the EC2 instance as the root volume.
  Syntax:

  ```
  dd if=/dumps/rajesh/image2.raw   of=/dev/xvdaa bs=1M
  ```

  Now the imported EC2 instance has the root volume with the second backup image i.e, image2.raw

This Procedure was followed to achieve replication of VM to EC2.

4.1.2 Case study protocol

Data collection protocol

Participant Observations (PO): Participant observation is when observing behaviors in natural setting, through active participation in the situation and manipulation of the environment [45]. Participant observation has been used in many practices as a tool for collecting qualitative data. This research method is one of the qualitative data collection methods.

Marshall and Rossman defined Observation as *"the systematic description of events, behaviors and artifacts in the social setting chosen for study"* [46]. Participant observation is the process of enabling researchers to learn about the activities

of the people under study in the natural setting through observing and participating in those activities.

Schensul and Lecompte defined participant observation as *"the process of learning through exposure to or involvement in the day-to-day or routine activities of participants in the research setting"* [46].

Participant observations has been selected as a data collection method as the validity of the study can be increased and the observations may help to have a better understanding of the context and phenomenon under the study.
By PO data collection method, observations have been made on how the backups are taken and stored either at on-premises or at remote location. While observations are made, informal discussions have been made and field notes is taken. Survey questionnaire and interviews were not selected as data collection method as we may not get hands on experience and clear knowledge on taking the backups of the VMs.

After POs are made, a clear idea has been gained on how the snapshots are taken and stored. Data has been collected by participant observations and informal discussions at Actifio. Field notes was taken while conducting observations and discussions. Then based on the observations and discussions, the collected data and the field notes is analyzed such that it makes readers easy to understand.

Analysis protocol

Qualitative data consist of words and observations, it doesn't contain numbers. To bring order and understanding with all data, analysis and interpretation was done. This process depends on questions the researcher want to answer, the resources and the needs of those who will use the information [47]. The outcome of the POs is in the form of text and narrative data which came from the field notes and from informal discussions with a team at Actifio.
According to [47], the analysis process consists of

- Get to know your data
- Focus on the analysis
- Categorize information
- Identify patterns and connections within and between categories
- Interpretation-bringing it all together

The same procedure has been followed in my thesis.

Get to know your data: The observations made, gathered data and field notes were understood. By reading the field notes time and again, some important impressions were made. As the goal of collecting data is clear, there are less chances of getting biased data.

Focus on the Analysis: As the purpose of evaluating the data was to find out the procedure for taking backups and how they were moved to target, some key questions are identified. By these questions the data analysis has started. Focus by case type of approach is selected. In focus by case, the analysis will be done on one case. In this thesis the analysis is done on a single case i.e., how the backup is taken and how it is stored.

Categorize information: As people refer to categorizing information as coding or indexing the data, the codes related to the data were gathered. To bring meaning to these codes the themes are identified by idea, concept, behavior, interaction, incidents, terminology and phrases used and are organized into coherent categories such that the codes summarize and bring meaning to the text. This is the important part of qualitative analysis as it involves reading and re-reading the text and identifying consistent categories. As categorizing the data was started, other codes that serves as subcategories were identified. Categorization was done until all the codes have been labeled. The question and the identified categories are mentioned below.

The initial question was: How the snapshot of Virtual Machine is taken and how it is stored.
From this question different categories have been made. The categories in this question are: snapshot, Virtual Machine, storage location.

There are two ways for categorizing the data. They are preset categories and emergent categories. In preset category a list of themes or categories will be made in advance and then data will be searched for these topics. The advantage of this category is that the researcher can start with the preselected categories. In emergent categories, rather than using preconceived themes or categories, the themes will be found emergently by reading the data. In this thesis the two methods were combined. The analysis started with preset categories and the other themes were added as they become important.

Identify patterns and connections within and between categories: After data is organized into categories by question, a look up for patterns and connections between and within the categories is done. Each category is summarized with data based on the observations made. Then the relation between these categories is found and relocated the categories in a way that the reading will be perfect and easily understood.

Interpretation-Bringing it all together: By using the themes and connections, the findings are explained. Analysis of the collected data was started by developing a list of key points and important findings and those resulted in categorizing and sorting the data.

4.2 Experiment

Conducting the case study in this thesis results in migrating on premise Virtual Machine to EC2. Phoronix Test Suite (PTS) [13] is used to benchmark the performance between the Virtual Machines.

The experimental setup consists of on premise VM running on vSphere which runs on ESXi hypervisor and the same VM migrated to EC2. After migrating, the VM runs on Xen hypervisor as EC2 uses Xen hypervisor. The hardware and software configuration of VMs are mentioned in Table 4.1.

	On premise VM (VMware)	Migrated VM (EC2)
Operating System	Ubuntu 14.04	Ubuntu 14.04
Architecture	X86_64	X86_64
File system	Ext4	Ext4
Hypervisor vendor	VMware	Xen
Virtualization type	Full	Full
Processor	Intel Xeon @2.4GHz	Intel Xeon @2.4GHz
Memory	1024MB	1024MB
Disk	8GB	8GB
L1i cache	32K	32K
L2 cache	256K	256K
L3 cache	30720K	30720K

Table 4.1: Hardware and software configuration of VMs

The experimental setup i.e., the VMs used in this experiment is gained from the case study. In the case study on premise VM is migrated to EC2. That means the same VM is also available on cloud platform. As the performance of the VM after migration is unknown, benchmarking is done between the VMs by selecting some open source benchmarks which compares the performance between the VMs.

Phoronix Test Suite [13] is an open source benchmark platform used in this thesis to benchmark and compare performance of the virtual machines. PTS is a benchmarking suite written in PHP designed for automating the process of benchmarking Linux systems. It provides a Command Line Interface to automatically download and install various commonly used open source benchmarks. The

main aim of running these benchmarks is to compare the performance between the migrated VM and on premise VM as the performance of the migrated VM is unknown. The performance of VM may vary after migration as it runs on different hypervisor compared to the on premise one.

In this thesis the on premise VM runs on VMware vSphere and vSphere VMs runs on ESXi hypervisor. As the VMs are migrated to EC2, the migrated VM runs on Xen hypervisor. Five benchmarks were selected out of many benchmarks to measure important attributes of a virtual machine. CPU, memory utilization and disk read/write speed type of attributes were calculated between the virtual machines.
The following benchmarks were selected and performance is calculated by running the same benchmark on different VMs with same configuration in different environments:

- Apache benchmark [8]: it calculates the sustainability on how many requests/second when carrying out 1,000,000 requests with 100 requests being carried out concurrently.
- John the Ripper [9]: this benchmark compares how many blowfish computations can be done per one second.
- Stream [10]: benchmarks system RAM read and write performance.
- OpenSSL [11]: measures the RSA 4096-bit encryption and decryption performance.
- Unpack [12]: it measures the amount of time the VM takes to extract content from a .tar.bz2 archive.

Motivation: The PTS benchmarks are valid tests to determine how VMs performance and throughput varies as the VMs performing the same computationally or I/O intensive task changes. The main aim of using the expect scripts provided by PTS is to compare how well a VM is performing when it is migrated from on premise to cloud.
The process of comparing one's business process and performance metrics to industry best or best practices from other companies is called benchmarking. In this thesis benchmarking is done by comparing the performance between the on premise VM and the migrated VM.
Phoronix Test Suite is installed in both the Virtual Machines. The other required dependencies were downloaded by PTS automatically.

4.2.1 Scenario 1

This experiment is related to the performance of the Virtual Machine using the Apache Benchmark. This benchmark is designed to test the performance of a

given server. It calculates the performance by stressing a given server to determine how many requests/second it is able to handle. The version of the benchmark that is included with PTS attempts to execute 1,000,000 requests to the server 100 requests at a time. It then measures how many requests/second the VM has made [8][13]. Executing Apache benchmark on PTS runs the benchmark for three times and then the average of those values is given as the final result for that run.
Motivation: This benchmark was chosen as it gives a good idea how the VM is able to handle increasing I/O stress in terms of CPU and memory usage. Running this benchmark on both the virtual machine gives a clear idea on the performance in terms of CPU and memory usage.

4.2.2 Scenario 2

This experiment is related to the performance of the Virtual Machine using the John the Ripper benchmark. Out of the available Traditional DES, MD5 and Blowfish, blowfish test was selected as it used to test how the various VMs perform under very calculation intensive operations. Blowfish is a symmetric block cipher that can be used as a drop-in replacement for DES. Blowfish is a variable-length key, 64-bit block cipher [48]. Executing John the Ripper benchmark on PTS runs the benchmark for three times and then the average of those values is given as the final result for that run [9][13].
Motivation: This benchmark was chosen as it gives a good idea of how the VM performs that intensive operations. This benchmark calculates the number of blowfish calculations were made per one second on each VM. The VM with more Blowfish computations has good performance compared to the other one.

4.2.3 Scenario 3

This experiment is related to the performance of the Virtual Machine using the Stream benchmark. This benchmark is used to measure the sustainable memory bandwidth rather than the burst or peak performance of a machine. Out of four operating modes of the stream benchmark i.e. copy, scale, sum, triad, copy mode was used as it measures transfer rates by copying a large array from one location to another, instead of doing any additional arithmetic operations. The array used in this benchmark is larger than the cache of the machine and structured such that data re-use is not possible [10][13]. Executing Stream benchmark on PTS runs the benchmark for ten times and then the average of those values is given as the final result for that run.
Motivation: This benchmark calculates the memory bandwidth for a copy task by copying a large array from one location to another. The virtual machine with high bandwidth has good performance when compared to the other one.

4.2.4 Scenario 4

This experiment is related to the performance of the Virtual Machine using the OpenSSL Benchmark. OpenSSL is an open-source toolkit that implements Secure Sockets Layer and Transport Layer Security protocols. This benchmark uses the open source SSL toolkit to measure the performance of signing and verifying using 4096 bit RSA operations [11][13]. Executing OpenSSL benchmark on PTS runs the benchmark for three times and then the average of those values is given as the final result for that run.
Motivation: This benchmark is used to calculate the number of RSA calculations per second. The VM with more RSA calculations per second performs better than the compared VM.

4.2.5 Scenario 5

This experiment is related to the performance of the Virtual Machine using the Unpack Benchmark. The Unpack benchmark from PTS measures the time requires to extract a Linux kernel from a tar.bz2 archive. This benchmark's performance is mainly a factor of disk read/write speed but the CPU is also a factor to uncompress the bz2 archive [12][13]. Executing Unpack-Linux benchmark on PTS runs the benchmark for four times and then the average of those values is given as the final result for that run.
Motivation: As this benchmark's performance is mainly a factor of disk read-/write speed and as CPU is also a factor to uncompress the bz2 archive [13], running this benchmark gives a clear idea on how the VMs perform when disk and CPU factors are considered. The VM extracting Linux kernel from a tar.bz2 archive in less time is considered to be better performance compared to the other VM.

Chapter 5

Results and Analysis

5.1 Case Study Results

After conducting participant observations, clear idea has been gained on creating the VM on vCenter, installation of OS on the VM, protecting and taking the backup of the VM. The detailed instructions are as follows.

Creation of VM: Firstly a VM is to be created so that the backup of that VM can be taken. For creating the VM, VMware vSphere client is used. After opening vSphere, IP address, username and password were entered and logged in. Under vCenter different hosts can be added. Under the host, VMs were created then the process for creating a new VM was started. Two types of configurations were available on vSphere. They are typical and custom. By selecting typical, a new VM will be created with the most common devices and configurations. By selecting custom, a new VM with additional devices or specific configuration options will be created. After selecting configuration, host was selected by naming the VM. Then the guest OS and OS version was selected. Network was selected then selection of the data store was done from where the VM should allot the virtual disk and size of the disk was mentioned. Here three types of disks can be selected. They are thick provision lazy zeroed, thick provision eager zeroed and thin provision. Then the VM was created.

Installation of OS on VM: OS in the VM was installed by clicking on Virtual Machine Properties, selecting CD/DVD and entering the location of the ISO file of the OS. Then by opening the console, the VM with OS was ready.

Protecting VM (taking Snapshot): Now this VM is discovered Using Actifio CDS. After logging into Actifio desktop, click on Discover VM. Then select the protected VM by using a template and a profile. A template was created by scheduling the snapshot and a profile was created by selecting the Disk pool to store the backup. Then the backup copy was retrieved form the storage. Then the guidelines provided by AWS were used to migrate the VM to EC2. The instructions for installing EC2 CLI tools are mentioned in Appendix A. The backup

image of VM was taken in VMDK, VHD and RAW formats.

5.1.1 VM migration using vSphere plugin

The VM migration started with vCenter plugin and all the steps were followed as mentioned in the method chapter. After migration task is completed, a newly created instance was shown by the port. The instance was started by clicking on start the instance button.
The migration method by using this method was successful and it is not used for migration as the goal is to migrate snapshot of the VM to EC2 and not the running VM.
Then investigation was done on the other two methods provided by AWS.

5.1.2 VM import using EC2 Import Image

As mentioned in the method chapter, after exporting the Virtual Machine from its Virtual Environment, AWS EC2 import-image command was used.
The Virtual Machine in OVA format was uploaded to Amazon S3 using the upload tool provided by AWS [16]. Then AWS CLI was used to import the OVA image. The syntax used is as follows:

```
$ aws ec2 import-image --cli-input-json "{  \"Description\":
\"Windows 2008 OVA\", \"DiskContainers\": [ { \"Description\":
\"First CLI task\", \"UserBucket\": { \"S3Bucket\":
\"my-import-bucket\", \"S3Key\" : \"my-windows-2008-vm.ova\" } } ]}"
```

After running this command, import task was started and the VM was available in the Amazon Machine Image's list. The particular AMI was selected and the VM was launched.
Importing the VM using this method was successful for Windows, Ubuntu and RHEL VMs but this method is not used to migrate the VMs as the goal is to migrate the snapshot of the VM to EC2 and run the snapshot on EC2 but not the running VM on vCenter.

5.1.3 VM import using EC2 Import Instance

Importing a VMware VM image of VMDK format to EC2: The script used to import VMDK image to EC2 was:

```
ec2-import-instance rjsh_vm.vmdk -f vmdk -p Windows -o access key
-w secret key -t m1.small -a i386 -z us-east-1a -b bucket name
```

Importing this VMDK image to EC2 was not successful as it throws an error saying this does not appear to be a Stream Optimized VMDK. It contains invalid

magic number -795951053.
The same migration method was tried using snapshot containing Ubuntu and RHEL OS and importing was not successful as it throws the same error. Then migration method was tried using VHD format.

Importing VMware VM image of VHD format to EC2: The script used to import was:

```
ec2-import-instance rjsh_vm.vhd -f vhd -p Windows -o access key
-w secret key -t m1.small -a i386 -z us-east-1a -b bucket name
```

Importing the VHD format was not successful as it throws an error saying this does not appear to be a dynamic VHD image. Image starts with *****. As importing the VHD format was also not supported, the next investigation was made using RAW format.
The same migration method was tried using snapshot containing Ubuntu and RHEL OS and importing was not successful as it throws the same error. Then migration was tried using RAW format.

Importing VMware VM image of RAW format to EC2: After trying with the VMDK and the VHD format, an attempt to import RAW format of the backup image to EC2 was done. The script used to import was:

```
ec2-import-instance rjsh_vm -f raw -p Windows -o access key
-w secret key -t m1.small -a i386 -z us-east-1a -b bucket name
```

Finally, importing the RAW format was successful. To verify that VM is migrating to EC2, three real time VM snapshots were taken from Actifio's datastore in the RAW format containing Windows, Ubuntu and RHEL OS and the same migration method was used to migrate the VMs. As this method runs without any errors, this method was selected as persistent method to migrate VMs to EC2.

After running the ec2-import-instance command, it clearly mentions the details like Disk image format as RAW, Volume Size, Availability Zone, Status, etc. then a new manifest file was created with a name snapshotname.vmdkmanifest.xml. Then the manifest file was uploaded first, then the backup image was uploaded in number of parts. RAW file of 8GB is sent in 820 parts and stored in S3 bucket (in this case). Now after the upload is complete, the uploaded file is converted into an EC2 instance. The migrated VM was accessed by logging into AWS management portal. After entering console, EC2 was selected. After clicking on the running instances, the number of running instances were displayed. For connecting to the instance, rdesktop software was used which helps in connecting to the remote machines. rdesktop was installed and the command to connect to instance is:

```
\#rdesktop -u <username> -fP <IP of the virtual machine>
```

Then the console will be displayed.

5.1.4 Importing image directly from Datastore

Until now the investigation has been made by importing RAW image to the physical machine, later image was migrated to EC2. Now a trial has been made by importing an image directly from Datastore. Actually, Actifio stores the backups in Disk format. But to make this investigation possible, Actifio have implemented File system in User space (FUSE). FUSE is an OS mechanism for computer OS like UNIX and allows non-privileged users to create the file systems without editing kernel code [49]. This investigation has been made by importing the RAW format of the image directly from the data store and by implementing this method, a copy of data is saved when compared to the above migration methods.

5.1.5 Replicating the VM

Replication of VM was implemented by following the proposed method mentioned in the method chapter. After implementing the proposed method, when booting is done to EC2 instance which has the latest RAW format image in its root volume, the status checks stops at $\frac{1}{2}$ and the system log says that OS is not found. So, to investigate what is happening while importing a VMware VM to EC2, comparison is done by comparing the C drive files in both the VMs i.e., VMware VM and EC2 VM. Based on these observations, it is found that AWS installs some software and drivers when a backup is imported using ec2-import-instance. The screen shot of the differences are shown in Figure 5.1.

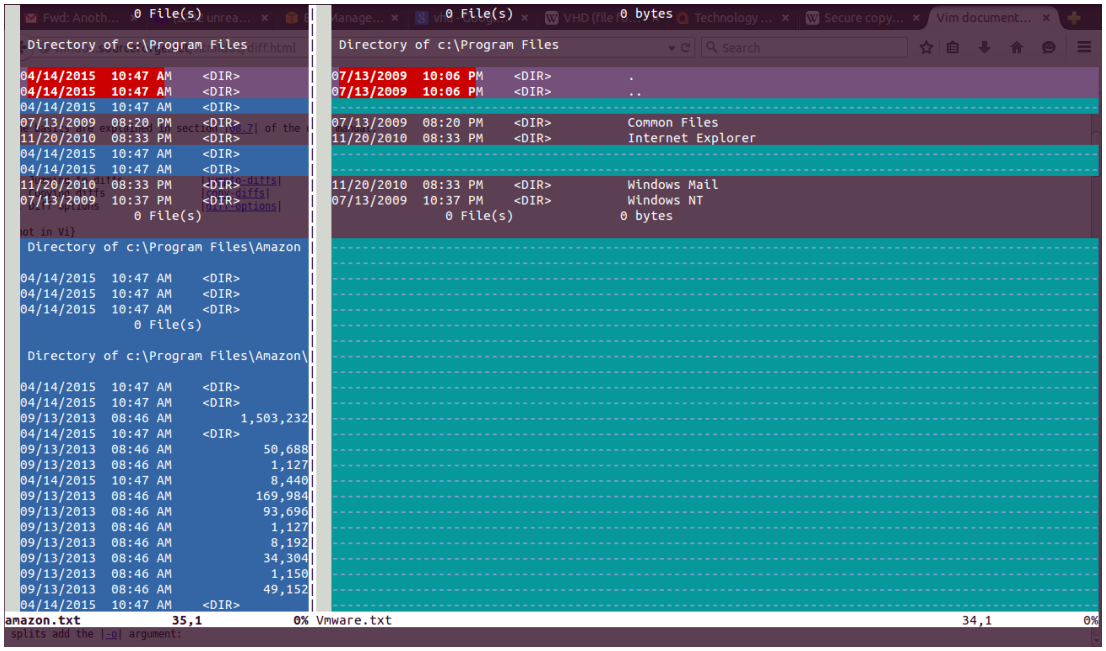

Figure 5.1: Comparison of C drive files between EC2 VM and VMware VM

The same experiment is made for RHEL VMs and the same happens in RHEL machines.
A trial was made by injecting AWS software and Xen tools on VM before importing, but it was not successful. Some discussions were made with AWS technical support team and concluded that it is not possible to inject Xen drivers or software on the VM before importing because AWS automatically installs the drivers and software while ec2-import task is made.

5.2 Experiment Results

The results are reported and analyzed in this chapter.

5.2.1 Scenario 1

As mentioned in the method chapter, Apache benchmark was ran on both the VMs. The performance of a given VM was calculated by the Apache benchmark. The performance was calculated by stressing the server to determine how many requests it can handle per one second. This version of benchmark attempts to execute 1,000,000 requests to the server 100 requests at a time. Then it measures how many requests/second the VM has made. The results of PTS are mentioned in Appendix B Table B.1 and B.2. The comparisons are mentioned in Table 5.1.

Test No	**EC2** (Requests/sec)	**VMware** (Requests/sec)
1	7480	5746.09
2	7425.55	5519.78
3	7480.41	5505.14
4	7478.4	5635.58
5	7556.84	5421.78
6	7504.09	5465.24
7	7339.14	5496.83
8	7503.19	5504.7
9	7504.08	5494.32
10	7392.52	5502.53
11	7495.98	5663.12
12	7438.1	5422.16
13	7423.92	5393.07
14	7464.51	5588.17
15	7432.56	5485.91
16	7518.83	5485.1
17	7473.36	5422.35
18	7487.88	5562.69
19	7448.21	5553.29
20	7459.38	5595.89
21	7427.62	5494.48
22	7408.57	5466.74
23	7451.42	5610.52
24	7485.62	5670.82
25	7384.62	5584.41

Table 5.1: Apache Benchmark

By running this benchmark for 25 times, Mean was calculated for the obtained results. The formula which was used for calculating mean is as follows

$$\overline{X} = \frac{\Sigma X_i}{N}$$

where
$\overline{X}$ is mean,
ΣX_i is sum of all data values,
N is number of all data values.

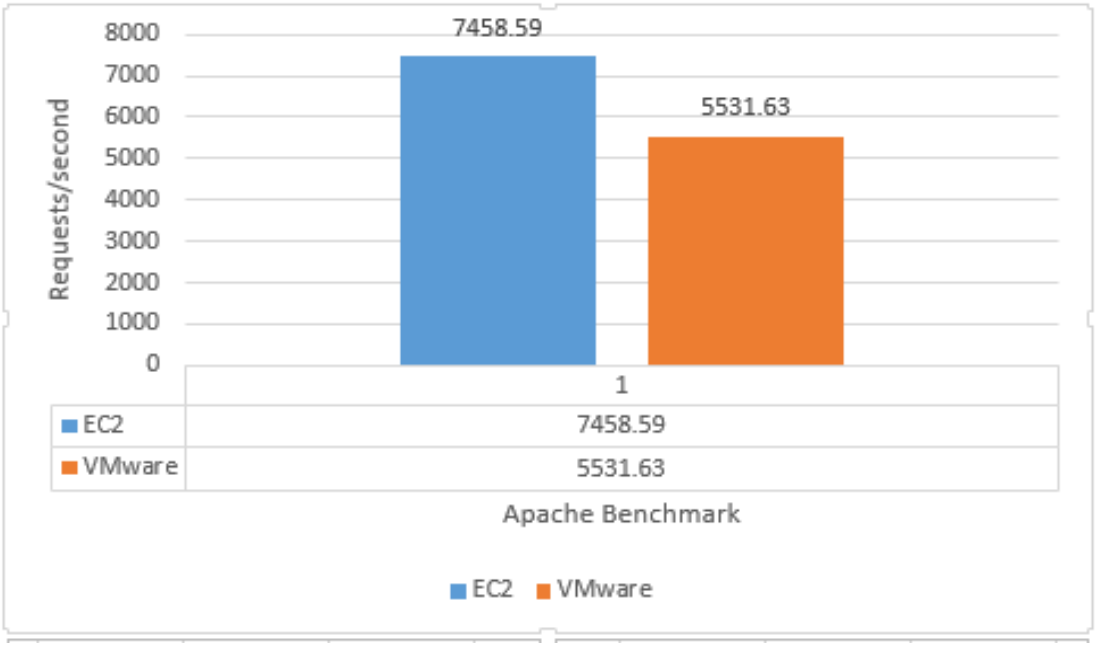

Figure 5.2: VMs performance-Apache Benchmark

The analysis has been done by taking independent variable on the x-axis and dependent variable on the y-axis as shown in Figure 5.2. Here the benchmark running on both VMs is considered as an independent variable and the number of requests/second of each VM is considered as the dependent variable. Running Apache benchmark on migrated EC2 VM made an average of 7459 Requests/second and on premise VMware VM made an average of 5532 Requests/second. From this experiment it is clear that EC2 VM handle the I/O stress better than VMware VM in terms of CPU usage and memory usage. The performance difference between the VMs may be due to the different hypervisors or the different virtualization technology used by the hypervisors.

5.2.2 Scenario 2

This experiment has been conducted to measure the performance of the VM using John the Ripper benchmark. This test is made to know how the VM perform under very intensive operations. The results of PTS are mentioned in Appendix B Table B.3 and B.4. The comparisons are mentioned in Table 5.2.

Test No	**EC2** (BF comp/sec)	**Vmware** (BF comp/sec)
1	806	692
2	809	668
3	807	670
4	785	636
5	800	667
6	811	666
7	813	685
8	801	681
9	802	687
10	769	656
11	787	675
12	774	675
13	783	671
14	804	642
15	787	628
16	795	645
17	800	660
18	786	646
19	807	674
20	806	670
21	784	679
22	794	658
23	770	700
24	783	701
25	789	640

Table 5.2: John the Ripper Benchmark

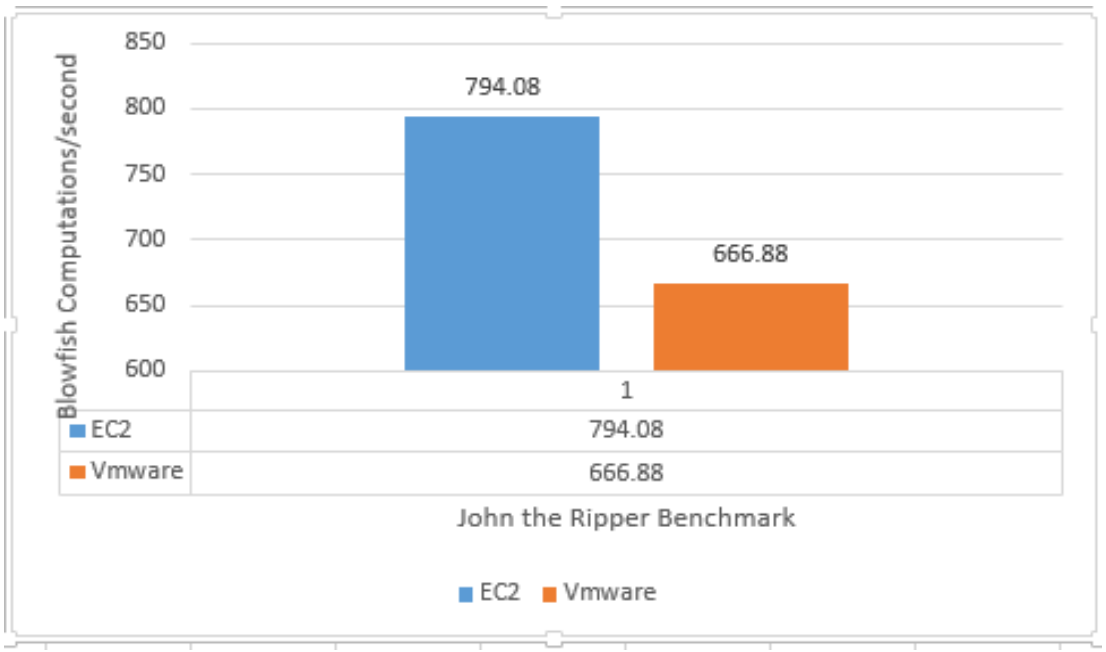

Figure 5.3: VMs performance-John the Ripper Benchmark

The analysis has been done by taking independent variable on the x-axis and dependent variable on the y-axis as shown in Figure 5.3. Here the benchmark running on both VMs is considered as an independent variable and the number of Blowfish computations/second of each VM is considered as the dependent variable. Running John the Ripper benchmark on migrated EC2 VM made an average of 794 Blowfish Computations/second and on premise VMware VM made an average of 667 Blowfish Computations/second. As the VM with more Blowfish computations/second performs better, it is clear that EC2 VM performs better than VMware VM under very calculation intensive operations. The performance difference between the VMs may be due to the different hypervisors or the different virtualization technology used by the hypervisors.

5.2.3 Scenario 3

This experiment is conducted to compare the performance of the VM using Stream benchmark. This benchmark is used to measure the sustainable memory bandwidth rather than the burst or peak performance. The array used in this benchmark is larger than the cache of the machine and is structured such that data re-use is not possible. The results of PTS are mentioned in Appendix B Table B.5 and B.6. The comparisons are mentioned in Table 5.3.

Test No	**EC2** (MB/sec)	**VMware** (MB/sec)
1	11302.23	10887.27
2	11300.72	10907.72
3	11420.84	10908.5
4	11322.92	10932.38
5	11439.77	10924.06
6	11330.28	10868.17
7	11390.32	10922.41
8	11454.84	10921.48
9	11409.07	10919.33
10	11588.8	10913.98
11	11481.66	10896.31
12	11401.69	10897.99
13	11286.89	10897.8
14	11406.52	10889.21
15	11323.37	10867
16	11399.71	10927.25
17	11247.03	10891.02
18	11280.52	10922.09
19	11282.75	10917.82
20	11307.6	10918.04
21	11290.9	10930.23
22	11175.62	10908.44
23	11304.59	10900.27
24	11244.47	10928.13
25	11288.25	10907.38

Table 5.3: Stream Benchmark

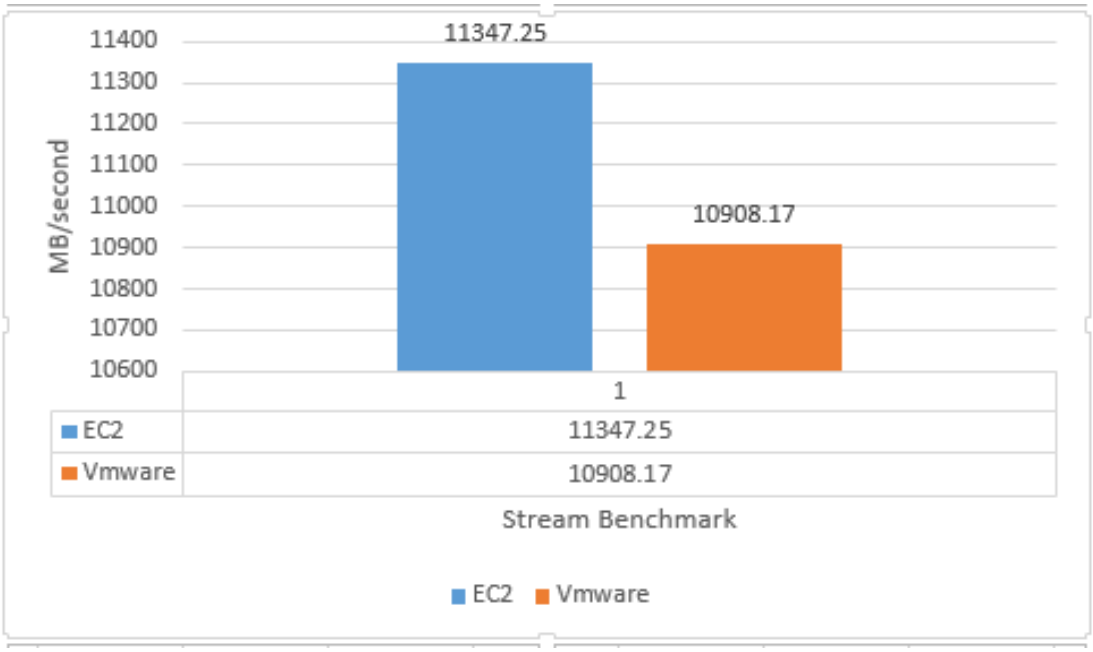

Figure 5.4: VMs performance-Stream Benchmark

The analysis has been done by taking independent variable on the x-axis and dependent variable on the y-axis as shown in Figure 5.4. Here the benchmark running on both VMs is considered as an independent variable and the bandwidth i.e, MB/second of each VM is considered as the dependent variable. Running Stream Benchmark for copy operation on migrated EC2 VM made an average of 11347 MB/second and on premise VMware VM made an average of 10908 MB/second. As the VM with more MB/second in copying a large array from one location to another performs better, it is clear that EC2 VM performs better than VMware VM. The performance difference between the VMs may be due to the different hypervisors or the different virtualization technology used by the hypervisors.

5.2.4 Scenario 4

This experiment has been conducted to measure the performance of the VM using OpenSSL benchmark. This benchmark uses the open source SSL toolkit to measure the performance of signing and verifying using 4096 bit RSA operations. The results of PTS are mentioned in Appendix B Table B.7 and B.8. The comparisons are mentioned in Table 5.4.

Test No	**EC2** (RSA/sec)	**VMware** (RSA/sec)
1	100.43	96.73
2	100.8	95.09
3	100.97	96.56
4	101	93.51
5	101.3	95.09
6	100.37	94.8
7	100.57	99.9
8	100.73	96.9
9	101.07	95.34
10	100.87	94.19
11	100.33	94.77
12	100.93	93.5
13	100.9	97.93
14	101.1	93.9
15	100.43	97.87
16	100.7	91.21
17	101.47	99.97
18	101.07	96.84
19	100.2	96.13
20	100.07	99.73
21	100.43	101.51
22	101.43	95.36
23	100.83	100.63
24	100.9	98.01
25	101.67	96.35

Table 5.4: OpenSSL Benchmark

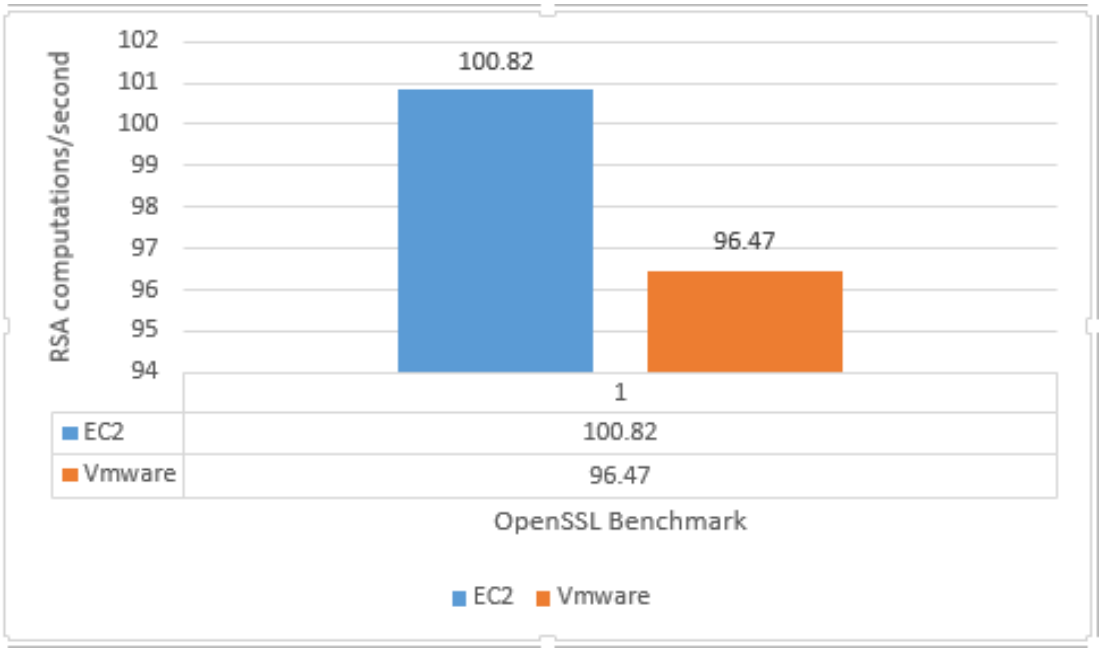

Figure 5.5: VMs performance-OpenSSL Benchmark

The analysis has been done by taking independent variable on the x-axis and dependent variable on the y-axis as shown in Figure 5.5. Here the benchmark running on both VMs is considered as an independent variable and the number of RSA computations/second of each VM is considered as the dependent variable. Running OpenSSL Benchmark on migrated EC2 VM made an average of 101 RSA Computations/second and on premise VMware VM made an average of 96 RSA Computations/second. As this benchmark uses open source SSL toolkit to measure the performance of signing and verifying using 4096 bit RSA operations, the VM with more RSA Computations/second performs better. It is clear that EC2 VM performs better than VMware VM. The performance difference between the VMs may be due to the different hypervisors or the different virtualization technology used by the hypervisors.

5.2.5 Scenario 5

This benchmark from PTS measures the time required to extract a Linux kernel from tar.bz2 archive. The results of PTS are mentioned in Appendix B Table B.9 and B.10. The comparisons are mentioned in Table 5.5.

Test No	**EC2** (seconds)	**Vmware** (seconds)
1	11.56	12.41
2	11.4	12.41
3	11.83	12.62
4	11.46	12.45
5	11.79	13
6	11.76	12.3
7	11.71	12.93
8	11.64	12.47
9	11.81	13
10	11.46	13.14
11	11.65	12.85
12	11.8	12.85
13	11.36	12.67
14	11.35	12.77
15	11.56	12.6
16	11.32	12.94
17	11.67	13.03
18	11.72	12.85
19	11.53	12.99
20	11.98	12.9
21	11.22	12.73
22	11.58	12.89
23	11.47	13.01
24	11.98	13.06
25	12.09	12.96

Table 5.5: Unpack Benchmark

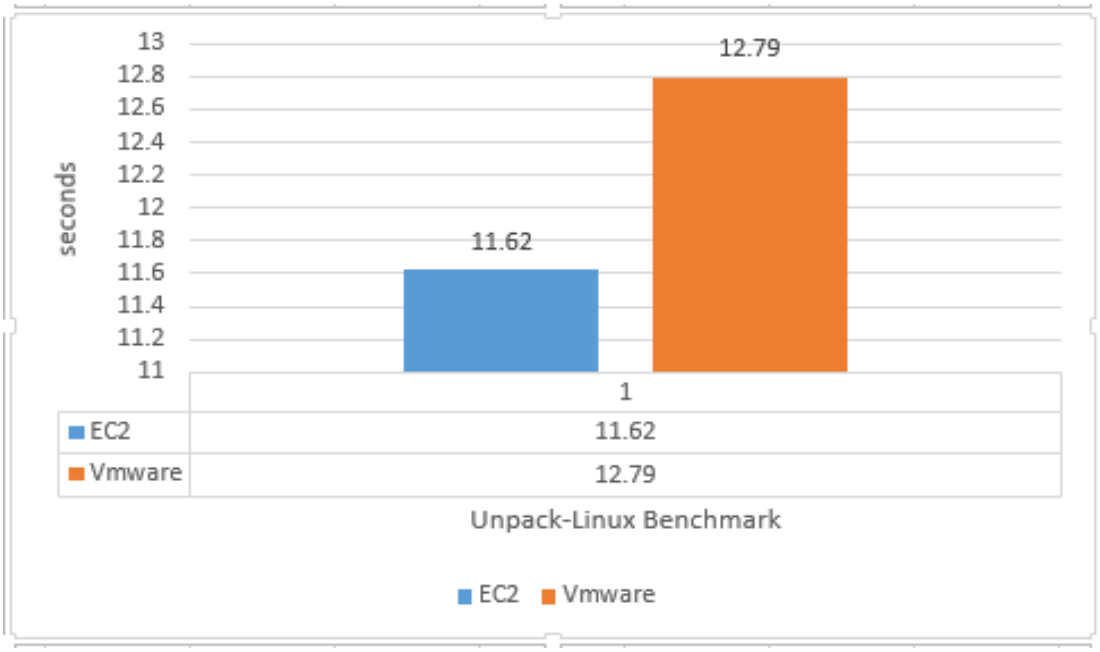

Figure 5.6: VMs performance-Unpack Benchmark

The analysis has been done by taking independent variable on the x-axis and dependent variable on the y-axis as shown in Figure 5.6. Here the benchmark running on both VMs is considered as an independent variable and the time i.e., number of seconds to unpack the linux for each VM is considered as the dependent variable. Running Unpack-Linux Benchmark on EC2 VM took an average of 11.62 seconds and VMware VM took an average of 12.79 seconds to extract Linux kernel from tar.bz2 archive. As the VM extracting Linux kernel from a tar.bz2 archive in less time considered to have better performance, EC2 VM performs better than VMware VM. The performance difference between the VMs may be due to the different hypervisors or the different virtualization technology used by the hypervisors.

Chapter 6

Discussion

The thesis was started by conducting an investigative case study for migrating and replicating VMware VMs to EC2. Participant Observations were conducted as a data collection method to know the state of practice methods for taking snapshots of VMs. Migration of VMs to EC2 using vCenter plugin and EC2 import image was successful for Ubuntu, RHEL and Windows VMs. But this method was not selected as the goal is to migrate the snapshot of the Virtual Machine. Migration of VMs to EC2 using EC2 Import Instance was successful only for RAW format of images and was not successful for VMDK and VHD images. The reason for not importing VMDK and VHD snapshot was clearly shown as an error saying that the VMDK snapshot was not a stream optimized VMDK and the VHD snapshot does not appear to be a dynamic VHD. The investigation started with Windows VM then with Ubuntu and RHEL. The error was same for all three machines. VM import was successful using RAW format for EC2 Import Instance. Hence in this case RAW format was selected as a persistent method for importing VM to EC2. Replication was not successful as AWS installs some software and drivers while importing the VM to EC2.
In the experiments, the migrated EC2 VM performed better when compared to the on premise VMware VM in all the five benchmarks. The selected benchmarks have calculated the CPU, memory utilization and disk read/write speed. Based on the comparison, the migrated VM is said to have better performance compared to the on premise VM. The reason for better performance because the virtualization technology used by the hypervisors on which VMs were running may be different as the hypervisors used by VMware and AWS are different.

6.1 Independent and Dependent Variables

An Independent variable is a variable that stands alone and is not changed by other variables that are measured. Dependent Variable is a variable that depends on other factors. Independent variable causes a change in dependent variable and it is not possible that dependent variable could cause a change in independent variable.
The choice of the hardware and the environment where the VM is running i.e.,

VMware/Amazon EC2 can be considered as the independent variables and the parameters of the implemented benchmarks such as CPU, memory utilization and disk write/read speed are the dependent variables in this work.

6.2 Threats to Validity

- **Internal Validity** refers to how well an experiment is done, especially whether it avoids confounding i.e., more than one possible independent variable acting at same time [50]. Taking a VM snapshot containing more data takes more time for investigation. To overcome this threat, snapshots of very less size (below 4 GB) were considered. Reliability of the selected benchmarks is a threat to the comparison process. To overcome this threat five benchmarks were considered which calculates CPU, memory utilization and disk read/write speed.

- **External Validity** concerns whether the outcome of the experiment can be generalized to population [50]. Investigation on importing on premise VMware VM to different cloud providers with different hypervisor is a potential threat. This can be mitigated by using the VM migration guidelines provided by other cloud providers.

- **Construct Validity** concerns the relation between theory and observation [50]. To overcome this improper interpretation of theory, regular meetings with external and internal supervisor were made. Experimental setup has been made in accordance with the recommendations of the external supervisor.

- **Conclusion Validity** concerns the relationship between the treatment and the outcome [50]. To overcome this threat in performance comparison, all the tests were conducted for 25 times and all the benchmarks between on premise VM and migrated VM were ran concurrently.

Chapter 7

Conclusions and Future Work

This thesis concludes by answering all the Research Questions. From the investigation conducted in the case study, a method is selected by taking the snapshot in RAW format and import the VM to Amazon EC2. Snapshots of VMDK and VHD formats were not selected as the VMDK format was not stream optimized VMDK and the VHD format was not dynamic. Replicating the VM was not successful as AWS installs some software while a VM is being imported to EC2. Based on the experiments, migrated VM running on Amazon EC2 performs better than on premise VMware VM in all the selected benchmarks in terms of CPU, memory utilization and disk read/write speed. Thus EC2 VM is said to have better performance than VMware VM in terms of CPU, memory utilization and disk read/write speed.

7.1 Answering Research Questions

1. How does the VMware VM perform when it is migrated to EC2 platform in terms of CPU, memory utilization and disk read/write speed?
   Answer: The author has followed a mixed area of approach in this thesis. Case study and Experiments were conducted to answer this research question. Based on the conducted experiments, the migrated EC2 VM performed better in all the benchmarks than the on premise VMware VM in terms of CPU, memory utilization and disk read/write speed.

   1.1. How to take a backup of VMware Virtual Machine and import it to EC2?
   Answer: The backup of VM is taken using Actifio CDS. The procedure for taking backups is gained by conducting Participant Observations and Informal Discussions at Actifio.

   1.2. What are the challenges in automatically importing and replicating the snapshots of VMware Virtual Machines to EC2?
   Answer: The challenges in automatically importing and replicating the snapshots of VMware VMs to EC2 are as follows. As the VMDK and VHD snapshot used in this case were not stream optimized and

not dynamic,automatic migration of this format was not successful. Automatic replication was not successful as AWS installs some software while importing VMs to Amazon EC2. These are the challenges in automatically importing and replicating VMware VMs to EC2.

1.3. Which Virtual Machine has better performance, a Virtual Machine running on VMware or a Virtual Machine running on Amazon EC2 in terms of CPU, memory utilization and disk read/write speed?
Answer: For the considered metrics, migrated Amazon EC2 VM performed better than on premise VMware VM in terms of CPU, memory utilization and disk read/write speed. The open source benchmarks used for comparison are Apache [8], John the Ripper [9], Stream [10], OpenSSL [11] and Unpack [12].

7.2 Future Work

In this thesis, migrating a VMware Virtual Machine to Amazon EC2 was successful, but scheduled replication was not successful. As the challenges in replicating VM were mentioned, future works can be done by implementing a framework that replicates the on premise VMware VM to Amazon EC2. The problems in migrating Actifio's VMDK and VHD snapshots can be investigated. Performance comparison is made only on Ubuntu VMs. This can be done on Windows and RHEL VMs also.

References

[1] B. P. Rimal, E. Choi, and I. Lumb, "A taxonomy and survey of cloud computing systems," in *INC, IMS and IDC, 2009. NCM'09. Fifth International Joint Conference on*, pp. 44–51, IEEE, 2009.

[2] M. Peter and G. Timothy, "The nist definition of cloud computing," *NIST special publication*, vol. 9, 2011.

[3] L. MALHOTRA, D. AGARWAL, and A. JAISWAL, "Virtualization in cloud computing," *J Inform Tech Softw Eng*, vol. 4, no. 136, p. 2, 2014.

[4] "A Performance Comparison of Hypervisors." `http://www.vmware.com/pdf/hypervisor_performance.pdf`. [Online; accessed 25-Sept-2015].

[5] T. MOHAMMAD and C. S. EATI, "A performance study of vm live migration over the wan," *Electrical Engineering*, 2015.

[6] A. Khajeh-Hosseini, D. Greenwood, and I. Sommerville, "Cloud migration: A case study of migrating an enterprise it system to iaas," in *Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on*, pp. 450–457, IEEE, 2010.

[7] "Actifio-Copy Data Virtualization for the Enterprise and Cloud." `http://www.actifio.com/`. [Online; accessed 24-Sept-2015].

[8] "Welcome to The Apache Software Foundation!." `http://apache.org/`. [Online; accessed 28-Sept-2015].

[9] "John the Ripper password cracker." `http://openwall.com/john/`. [Online; accessed 28-Sept-2015].

[10] "STREAM: Sustainable Memory Bandwidth in High Performance Computers." `https://www.cs.virginia.edu/stream/`. [Online; accessed 28-Sept-2015].

[11] "OpenSSL:Cryptography and SSL/TLS Toolkit." `http://www.openssl.org/`. [Online; accessed 28-Sept-2015].

[12] "The Linux Kernel Archives." `https://www.kernel.org/category/about.html`. [Online; accessed 28-Sept-2015].

[13] M. Larabel and M. Tippett, "Phoronix test suite." `http://www.phoronix-test-suite.com/`. [Online; accessed 27-Sept-2015].

[14] A. Li, X. Yang, S. Kandula, and M. Zhang, "Cloudcmp: comparing public cloud providers," in *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, pp. 1–14, ACM, 2010.

[15] "Magic Quadrant for Cloud Infrastructure as a Service, Worldwide.." `https://aws.amazon.com/resources/gartner-2015-mq-learn-more/`. [Online; accessed 24-Sept-2015].

[16] "Amazon Web Services." `http://www.aws.amazon.com/`. [Online; accessed 25-Sept-2015].

[17] "Storage Networking Industry Association." `http://www.snia.org/`. [Online; accessed 24-Sept-2015].

[18] J. Sugerman, G. Venkitachalam, and B.-H. Lim, "Virtualizing i/o devices on vmware workstation's hosted virtual machine monitor.," in *USENIX Annual Technical Conference, General Track*, pp. 1–14, 2001.

[19] U. Gurav and R. Shaikh, "Virtualization: a key feature of cloud computing," in *Proceedings of the International Conference and Workshop on Emerging Trends in Technology*, pp. 227–229, ACM, 2010.

[20] Y. Ashino and M. Nakae, "Virtual machine migration method between different hypervisor implementations and its evaluation," in *Advanced Information Networking and Applications Workshops (WAINA), 2012 26th International Conference on*, pp. 1089–1094, IEEE, 2012.

[21] C. Chaubal, "The architecture of vmware esxi," *VMware White Paper*, vol. 1, p. 7, 2008.

[22] D. Chisnall, *The definitive guide to the xen hypervisor*. Pearson Education, 2008.

[23] S. Lowe, *Mastering VMware vSphere 5*. John Wiley & Sons, 2011.

[24] N. Garimella, "Understanding and exploiting snapshot technology for data protection, part 1: Snapshot technology overview," *internet accessed on Dec*, vol. 22, p. 8, 2009.

[25] "IBM Spectrum Protect." `http://www-01.ibm.com/software/tivoli/csi/backup-recovery/`. [Online; accessed 28-Sept-2015].

[26] I. Sriram and A. Khajeh-Hosseini, "Research agenda in cloud technologies," *arXiv preprint arXiv:1001.3259*, 2010.

[27] M. Jensen, J. Schwenk, N. Gruschka, and L. L. Iacono, "On technical security issues in cloud computing," in *Cloud Computing, 2009. CLOUD'09. IEEE International Conference on*, pp. 109–116, IEEE, 2009.

[28] B. R. Kandukuri, V. R. Paturi, and A. Rakshit, "Cloud security issues," in *Services Computing, 2009. SCC'09. IEEE International Conference on*, pp. 517–520, IEEE, 2009.

[29] S. Subashini and V. Kavitha, "A survey on security issues in service delivery models of cloud computing," *Journal of network and computer applications*, vol. 34, no. 1, pp. 1–11, 2011.

[30] Y. Ashino and M. Nakae, "Virtual machine migration method between different hypervisor implementations and its evaluation," in *Advanced Information Networking and Applications Workshops (WAINA), 2012 26th International Conference on*, pp. 1089–1094, IEEE, 2012.

[31] S. Shirinbab, L. Lundberg, and D. Ilie, "Performance comparison of kvm, vmware and xenserver using a large telecommunication application," *connections*, vol. 1, no. 15, p. 30, 2014.

[32] P. Mehrotra, J. Djomehri, S. Heistand, R. Hood, H. Jin, A. Lazanoff, S. Saini, and R. Biswas, "Performance evaluation of amazon ec2 for nasa hpc applications," in *Proceedings of the 3rd workshop on Scientific Cloud Computing Date*, pp. 41–50, ACM, 2012.

[33] S. Akioka and Y. Muraoka, "Hpc benchmarks on amazon ec2," in *Advanced Information Networking and Applications Workshops (WAINA), 2010 IEEE 24th International Conference on*, pp. 1029–1034, IEEE, 2010.

[34] K. R. Jackson, L. Ramakrishnan, K. Muriki, S. Canon, S. Cholia, J. Shalf, H. J. Wasserman, and N. J. Wright, "Performance analysis of high performance computing applications on the amazon web services cloud," in *Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on*, pp. 159–168, IEEE, 2010.

[35] S. Toyoshima, S. Yamaguchi, and M. Oguchi, "Storage access optimization with virtual machine migration and basic performance analysis of amazon ec2," in *Advanced Information Networking and Applications Workshops (WAINA), 2010 IEEE 24th International Conference on*, pp. 905–910, IEEE, 2010.

[36] C. D. Graziano, "A performance analysis of xen and kvm hypervisors for hosting the xen worlds project," 2011.

[37] T. Deshane, Z. Shepherd, J. Matthews, M. Ben-Yehuda, A. Shah, and B. Rao, "Quantitative comparison of xen and kvm," *Xen Summit, Boston, MA, USA*, pp. 1–2, 2008.

[38] A. Lenk, M. Menzel, J. Lipsky, S. Tai, and P. Offermann, "What are you paying for? performance benchmarking for infrastructure-as-a-service offerings," in *Cloud Computing (CLOUD), 2011 IEEE International Conference on*, pp. 484–491, IEEE, 2011.

[39] S. Wang, Y. Chen, W. Jiang, P. Li, T. Dai, and Y. Cui, "Fairness and interactivity of three cpu schedulers in linux," in *Embedded and Real-Time Computing Systems and Applications, 2009. RTCSA'09. 15th IEEE International Conference on*, pp. 172–177, IEEE, 2009.

[40] A. Basu *et al.*, "Performance of eucalyptus and openstack clouds on futuregrid," *International Journal of Computer Applications*, vol. 80, no. 13, pp. 31–37, 2013.

[41] M. Jarus, S. Varrette, A. Oleksiak, and P. Bouvry, "Performance evaluation and energy efficiency of high-density hpc platforms based on intel, amd and arm processors," in *Energy Efficiency in Large Scale Distributed Systems*, pp. 182–200, Springer, 2013.

[42] R. K. Yin, *Case study research: Design and methods*. Sage publications, 2013.

[43] R. Ashok, M. L. Woodward, P. J. Abercrombie, M. Mutalik, T. Taylor, C. A. Provenzano, and D. F. Chang, "Virtual data backup," Feb. 13 2015. US Patent App. 14/622,479.

[44] "Virtual Infrastructure SDK." `https://www.vmware.com/pdf/SDKProgrammingGuide.pdf`. [Online; accessed 03-Oct-2015].

[45] G. Guest, E. E. Namey, and M. L. Mitchell, *Collecting qualitative data: A field manual for applied research*. Sage, 2012.

[46] B. B. Kawulich, "Participant observation as a data collection method," 2005.

[47] E. Taylor-Powell and M. Renner, "Analyzing qualitative data. university of wisconsin-extension program development & evaluation," 2003.

[48] B. Schneier, "Description of a new variable-length key, 64-bit block cipher (blowfish)," in *Fast Software Encryption*, pp. 191–204, Springer, 1994.

[49] "FUSE for OS X." `http://osxfuse.github.io/`. [Online; accessed 04-Oct-2015].

[50] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in software engineering*. Springer Science & Business Media, 2012.

Appendices

Appendix A

Appendix A

A.1 Procedure for Installing EC2 CLI tools

1. Install Java Development Kit on the machine.
2. Amazon EC2 API tools were installed. The version of installed EC2 tools was ec2-api-tools-1.6.12.0
3. AWS Certificates were downloaded. The certificates consists of the private key and the access key.
4. Environmental Variables were configured.

```
export EC2_HOME=<Path of EC2 tools>
export PATH=EC2_HOME/bin
export EC2_PRIVATE_KEY='ls $EC2_HOME/pk-*.pem'
export EC2_CERT='ls $EC2_HOME/cert-*.pem'
export JAVA_HOME=<Java Home Path>
```

   (a) Here EC2_HOME is pointed to the location where amazon tools were installed.
   (b) .pem files (security key and access key) were kept in EC2_HOME directory.
   (c) JAVA_HOME was pointed appropriately.
   (d) All these contents were pasted in **bashrc** home.

Verification of the installed EC2 tools was done by typing the command **ec2-describe-regions**. This command displays the available AWS regions.

Pre-requisites for importing VM to AWS

The pre-requisites for importing VMs to AWS [16] are mentioned in Table A.1.

Accepted Operating System by Amazon Web Services	Windows: 2012, 2008, 2003 and 7 can be imported
	Linux: 5.1-5.11, 6.1-6.6, and 7.0-7.1 can be imported
	CentOS:,5.1-5.11, 6.1-6.6, and 7.0-7.1
	Ubuntu:,12.04, 12.10, 13.04, 13.10, 14.04, 14.10
	Debian:,6.0.0-6.0.8, 7.0.0-7.2.0
Accepted Image types	VMDK, RAW and VHD formats.
Instance types	General purpose, compute optimized, memory optimized, storage optimized and GPU.
Windows (32 and 64 bit)	VM Import supports MBR-partitioned volumes that are formatted using the NTFS file system and doesn't support GUID Partition Table (GPT)
For Linux (64 bit only)	VM Import supports MBR-partitioned volumes that should be formatted using ext2, ext3, ext4, Btrfs, JFS or XFS file system and doesn't support GPT partitioned volumes

Table A.1: Pre-requisites for importing VM to AWS

Appendix B

PTS Results

Test no	Average	Result 1	Result 2	Result 3
1	**7480**	7522.34	7451.18	7466.48
2	**7425.55**	7452.89	7320	7503.77
3	**7480.41**	7477.92	7485.93	7477.39
4	**7478.4**	7493.56	7494.06	7447.58
5	**7556.84**	7534.58	7539.84	7596.09
6	**7504.09**	7516.47	7499.41	7496.39
7	**7339.14**	7331.45	7345.2	7340.77
8	**7503.19**	7501.01	7553.89	7454.68
9	**7504.08**	7531.61	7481.28	7499.35
10	**7392.52**	7443.53	7377.68	7356.36
11	**7495.98**	7438.19	7508.53	7541.23
12	**7438.1**	7454.35	7411.97	7447.98
13	**7423.92**	7433.34	7392.3	7446.13
14	**7464.51**	7486.7	7476.36	7430.48
15	**7432.56**	7427.55	7440.69	7429.43
16	**7518.83**	7541.09	7500.79	7514.62
17	**7473.36**	7491.42	7368.55	7560.1
18	**7487.88**	7566.54	7408.25	7488.86
19	**7448.21**	7486.23	7416.29	7442.13
20	**7459.38**	7436.12	7439.84	7502.17
21	**7427.62**	7317.12	7369.56	7596.19
22	**7408.57**	7453.19	7417.93	7354.58
23	**7451.42**	7393.53	7391.5	7569.23
24	**7485.62**	7522.2	7459.81	7474.85
25	**7384.62**	7468.19	7337.82	7347.85

Table B.1: Apache Benchmark-Amazon EC2
Values in Requests/second

Test no	**Average**	**Result 1**	**Result 2**	**Result 3**
1	**5746.09**	5794.29	5728.68	5715.29
2	**5519.78**	5471.24	5603.5	5484.6
3	**5505.14**	5459.44	5597.42	5458.56
4	**5635.58**	5631.55	5532.63	5742.57
5	**5421.78**	5464.26	5286.68	5514.39
6	**5465.24**	5224.69	5764.35	5406.67
7	**5496.83**	5695.29	5243.58	5551.61
8	**5504.7**	5378.4	5563.29	5572.4
9	**5494.32**	5438.36	5277.37	5767.24
10	**5502.53**	5402.27	5722.75	5382.58
11	**5663.12**	5706.69	5651.23	5631.44
12	**5422.16**	5437.48	5620.69	5208.32
13	**5393.07**	5240.42	5410.58	5528.21
14	**5588.17**	5796.41	5653.62	5314.49
15	**5485.91**	5309.78	5740.59	5407.36
16	**5485.1**	5471.57	5532.45	5451.27
17	**5422.35**	5696.72	5239.68	5330.65
18	**5562.69**	5428.57	5719.76	5539.73
19	**5553.29**	5594.67	5370.65	5694.55
20	**5595.89**	5799.7	5433.34	5554.64
21	**5494.48**	5351.69	5448.38	5683.359
22	**5466.74**	5334.512	5729.22	5336.49
23	**5610.52**	5535.59	5781.31	5514.67
24	**5670.82**	5701.48	5749.73	5561.26
25	**5584.41**	5650.21	5438.56	5664.45

Table B.2: Apache Benchmark-VMware
Values in Requests/second

Test no	**Average**	**Result 1**	**Result 2**	**Result 3**
1	**806**	786	814	818
2	**809**	822	814	792
3	**807**	816	787	819
4	**785**	778	788	790
5	**800**	810	771	820
6	**811**	819	794	821
7	**813**	820	819	800
8	**801**	775	820	808
9	**802**	776	811	820
10	**769**	769	769	769
11	**787**	810	771	781
12	**774**	768	767	787
13	**783**	770	794	786
14	**804**	777	819	815
15	**787**	782	798	781
16	**795**	796	817	773
17	**800**	791	794	815
18	**786**	774	783	800
19	**807**	800	816	805
20	**806**	800	816	801
21	**784**	792	783	776
22	**794**	789	814	780
23	**770**	780	767	764
24	**783**	790	770	789
25	**789**	789	789	788

Table B.3: John the Ripper-Amazon EC2
Values in Blowfish computations/second

Test no	**Average**	**Result 1**	**Result 2**	**Result 3**
1	**692**	685	721	670
2	**668**	655	723	627
3	**670**	677	671	661
4	**636**	636	620	653
5	**667**	644	632	724
6	**666**	642	713	642
7	**685**	721	714	620
8	**681**	670	692	680
9	**687**	677	672	711
10	**656**	665	665	638
11	**675**	642	699	683
12	**675**	679	715	631
13	**671**	658	664	690
14	**642**	663	632	632
15	**628**	630	626	627
16	**645**	638	649	649
17	**660**	620	670	689
18	**646**	621	694	622
19	**674**	630	696	696
20	**670**	655	633	722
21	**679**	682	693	663
22	**658**	639	623	711
23	**700**	696	694	711
24	**701**	719	686	698
25	**640**	630	626	664

Table B.4: John the Ripper-VMware
Values in Blowfish computations/second

Test no	Average	Result 1	Res 2	Res 3	Res 4	Res 5	Res 6	Res 7	Res 8	Res 9	Res 10
1	**11302.23**	11336.3	11361.3	11224.1	11189.7	11094.2	11261.2	11259.7	11419.7	11474.5	11401.6
2	**11300.72**	11141.4	11078.1	11357.3	11108.9	11286.6	11618.6	11472	11334.6	11146.7	11463
3	**11420.84**	11491	11464.5	11276.3	11392.7	11323.5	11549.8	11526.6	11727.6	11201.4	11255
4	**11322.92**	11143.8	11325.8	11091.1	11284.3	11304.3	11463	11239.9	11381.3	11435	11560.7
5	**11439.77**	11279.6	11487.7	11308.3	11549.8	11567.3	11514.9	11536.5	11507.6	11564.1	11081.9
6	**11330.28**	11220.2	11128.2	11073.3	11066.6	11384.6	11483.6	11435	11226.3	11590.9	11694.1
7	**11390.32**	11294.7	11585.9	11333.1	11404.2	11646.6	11208.5	11366.1	11334	11282.8	11447.3
8	**11454.84**	11450.7	11554	11559.9	11586.7	11304.3	11388.7	11405.7	11245.3	11552.2	11500.9
9	**11409.07**	11225.6	11212.3	11152.8	11281.8	11695.1	11396.8	11480.2	11496	11496.8	11653.3
10	**11588.8**	11639.7	11594.1	11346.7	11638.9	11487.7	11723.3	11658.3	11788.2	11680.5	11330.6
11	**11481.66**	11410.5	11680.9	11793.8	11733.8	11436	11319.3	11221.7	11399.3	11396	11425.3
12	**11401.69**	11513.3	11595.9	11351.7	11393.7	11646.6	11250.1	11254.2	11149.9	11245.3	11616.2
13	**11286.89**	11257.4	11418.9	11344.4	11087.2	11381.3	11180.2	11244.6	11380.7	11266.9	11307.3
14	**11406.52**	11536.5	11516.7	11277.1	11596.7	11415.4	11744	11661	11132.7	11180.2	11004.9
15	**11323.37**	11081.8	11083.4	11269.1	11266.1	11277.3	11436.8	11497.5	11660.2	11302.7	11358.8
16	**11399.71**	11454.8	11575.7	11458.9	11421.2	11444.2	11157.7	11088	11481	11431.2	11484.4
17	**11247.03**	11175.5	11215.5	11235.9	11177	11200.5	11224.1	11485.9	11230.5	11422	11103.4
18	**11280.52**	11155.2	11225.8	11061.1	11207.6	11314	11232.7	11400.1	11606.9	11468.9	11132.9
19	**11282.75**	11248.7	11233.5	11115	11111.1	11317	11351.5	11422	11495	11299.5	11234.2
20	**11307.6**	11125.7	11025.3	11403.2	11445	11549	11645.6	11478.5	11202.9	11098.1	11102.7
21	**11290.9**	11191.2	11179.4	11375.7	11455.5	11480.2	11404.9	11427.8	11145.8	11150.6	11097.9
22	**11175.62**	11065.9	11265.9	11251.8	11142.1	11182.6	11192.1	11044.4	11305.8	11213.8	11091.8
23	**11304.59**	11473.6	11385.4	11383.1	11196.8	11190.4	11249.3	11205.9	11420.5	11239.9	11301
24	**11244.47**	11138	11247	11217.7	11159.1	11111.1	11372.6	11264.6	11474.3	11246.3	11214
25	**11288.25**	11205.9	11181.8	11158.4	11341.1	11409	11345.9	11279.6	11278.6	11405.9	11276.3

Table B.5: Stream (Copy)-Amazon EC2
Values in MB/second

Test No	Average	Res 1	Res 2	Res 3	Res 4	Res 5	Res 6	Res 7	Res 8	Res 9	Res 10
1	**10887.27**	10859.2	10789.7	10981.1	10915.7	10795.2	10792.1	11003	10831.4	10878.9	11026.4
2	**10907.72**	10959.6	10807.5	10902.3	10863.7	10863.8	10901.3	10960.4	11025.1	10930.9	10862.6
3	**10908.5**	11024.2	10853.8	10865.9	10831.4	10984.2	10872.9	10872.6	10960.1	10978.4	10841.5
4	**10932.38**	11015.2	10904.9	10907.3	10924.6	10915.3	10900	11005.4	10992.6	10865.6	10892.9
5	**10924.06**	10944.4	10982.6	10820.8	10982.6	10988.4	10830.5	11014.6	10846.2	10938.6	10891.9
6	**10868.17**	10867.6	10802.1	10924.4	10941.6	10813.3	10866.8	10895.8	10891.7	10812.1	10866.3
7	**10922.41**	11018.4	10860.3	10873.4	10990.3	11010	10804.5	10944.5	10952.8	10872.4	10897.5
8	**10921.48**	10909.2	11008.3	10817.6	10833.2	11004.6	10919.8	11000.1	10861.4	10843.2	11017.4
9	**10919.33**	10881.3	10910.4	10789.2	11002.4	11008.6	10917.4	10855.3	10853.6	10959.7	11015.4
10	**10913.98**	10896.2	11019.3	10791.2	10979.6	10985.3	10830.7	10969.9	11011.6	10820.4	10835.6
11	**10896.31**	10985.3	10974.7	10976.7	10899.4	10837.2	10792.1	10806.4	10899.3	10871.6	10920.4
12	**10897.99**	10911.3	10875.4	10838.9	10985.1	10950.5	10844.7	10816.6	10882.4	10881.4	10993.6
13	**10897.8**	10813.4	11006.3	10860.4	10829.6	10864.4	10956.4	10979.1	10892.1	10904.7	10871.6
14	**10889.21**	10896.4	10826.6	11027.1	10789.2	10834.2	10917.1	10825.5	10963.9	10997.4	10814.7
15	**10867**	10949.1	10814.2	10795.6	10851.8	10800	10823.1	10935.4	10976.9	10821.3	10902.6
16	**10927.25**	10975.3	10843.6	10885.7	10877.2	10970.4	11002.3	10984.1	11028.6	10789.1	10916.2
17	**10891.02**	10847.1	10945.8	11024.7	10899.7	10832.5	10867.5	10821.3	10794.9	10848.6	11028.1
18	**10922.09**	10923.8	10928.4	10945.7	10977.1	10882.8	10995.4	10823.2	10972.6	10855.9	10916
19	**10917.82**	10841.2	10979.6	10814.2	11010.4	10828.6	10807.9	11009.1	10916.4	10971.2	10999.6
20	**10918.04**	10945.2	10901.8	10955.4	10998.9	10864.3	10883.9	10942.7	10931.1	10868.9	10888.2
21	**10930.23**	10975.4	10908.3	11004.7	10794.8	10942.6	10795.9	10961.9	10964.8	10993.4	10960.5
22	**10908.44**	10989.8	10807.4	10925.9	10915.1	10888.1	10917.5	10963.8	10909.2	10861.7	10905.9
23	**10900.27**	10807.2	10867.9	10977.1	10852.7	10939.2	10925.8	10846.1	10890	10872.6	11024.1
24	**10928.13**	10844.3	10988.9	11025.3	10914.8	10830.1	10951.6	10852.9	11021.6	11000.5	10851.3
25	**10907.38**	10889.4	10845.6	10803.9	10975.8	10889.2	10981.6	10887.5	10909.2	10910	10981.6

Table B.6: Stream (Copy)-VMware
Values in MB/second

Test No	Average	Result 1	Result 2	Result 3
1	**100.43**	100.2	100.2	100.9
2	**100.8**	101.2	100.6	100.6
3	**100.97**	100.3	100.8	101.8
4	**101**	100.9	101.1	101
5	**101.3**	100.4	101.3	102.2
6	**100.37**	100.1	100.9	100.1
7	**100.57**	100.5	100.6	100.6
8	**100.73**	100.4	100.7	101.1
9	**101.07**	100.8	101.3	101.1
10	**100.87**	100.5	101	101.1
11	**100.33**	100.6	100	100.4
12	**100.93**	100.8	100.9	101.1
13	**100.9**	100.8	101	100.9
14	**101.1**	100.8	101.3	101.2
15	**100.43**	100	100.7	100.6
16	**100.7**	100.6	100.6	100.9
17	**101.47**	101.7	101.6	101.1
18	**101.07**	100.6	100.4	102.2
19	**100.2**	101.3	100.3	99
20	**100.07**	99.6	99.6	100.7
21	**100.43**	101.4	100.1	99.8
22	**101.43**	101.8	102	100.5
23	**100.83**	100	101.5	101
24	**100.9**	101.2	100.7	100.8
25	**101.67**	101.5	102	101.5

Table B.7: OpenSSL-Amazon EC2
Values in RSA computations/second

Test No	Average	**Result 1**	**Result 2**	**Result 3**
1	**96.73**	90.09	92.11	108
2	**95.09**	90.09	96.09	99.1
3	**96.56**	96.09	100.09	93.5
4	**93.51**	93.4	91.06	96.08
5	**95.09**	90.06	92.5	102.7
6	**94.8**	96.05	95.45	92.9
7	**99.9**	101.1	95.6	102.99
8	**96.9**	102.82	95.19	92.7
9	**95.34**	93.01	101.02	92
10	**94.19**	92.1	92.45	98.01
11	**94.77**	91.36	99.54	93.4
12	**93.5**	97.12	90.29	93.1
13	**97.93**	95.6	101.97	96.21
14	**93.9**	93.92	92.79	94.99
15	**97.87**	102.49	95.12	95.99
16	**91.21**	91.69	91.39	90.54
17	**99.97**	102.16	99.2	98.54
18	**96.84**	94.54	93.89	102.1
19	**96.13**	96.7	97.5	94.19
20	**99.73**	100.1	101.99	97.1
21	**101.51**	100.1	101.49	102.95
22	**95.36**	92.9	101.5	91.69
23	**100.63**	100.69	99.09	102.1
24	**98.01**	101.03	96.9	96.1
25	**96.35**	94.1	92.15	102.79

Table B.8: OpenSSL-VMware
Values in RSA Computations/second

Test No	Average	Result 1	Result 2	Result 3	Result 4
1	**11.56**	11.229	11.827	11.285	11.877
2	**11.4**	11.129	12.11	11.235	11.136
3	**11.83**	12.127	11.695	11.652	11.865
4	**11.46**	11.267	11.643	11.784	11.15
5	**11.79**	11.994	11.94	11.274	11.96
6	**11.76**	11.476	12.118	12.098	11.36
7	**11.71**	11.73	11.33	11.768	12.007
8	**11.64**	11.226	11.71	11.51	12.1
9	**11.81**	12.075	11.942	11.269	11.96
10	**11.46**	11.113	11.501	11.362	11.881
11	**11.65**	12.119	11.793	11.527	11.174
12	**11.8**	11.815	12.036	11.305	12.051
13	**11.36**	11.381	11.134	11.438	11.477
14	**11.35**	11.178	11.634	11.141	11.45
15	**11.56**	11.678	11.168	11.797	11.61
16	**11.32**	11.337	11.156	11.346	11.457
17	**11.67**	12.106	11.436	11.454	11.665
18	**11.72**	12.001	11.322	12.075	11.475
19	**11.53**	11.947	11.415	11.28	11.477
20	**11.98**	12.069	11.636	12.119	12.094
21	**11.22**	11.404	11.104	11.144	11.227
22	**11.58**	11.403	11.796	11.097	12.041
23	**11.47**	11.575	11.962	11.182	11.156
24	**11.98**	11.575	12.109	12.112	12.11
25	**12.09**	12.125	12.118	12.004	12.098

Table B.9: Unpack-Amazon EC2
Values in seconds

Test No	Average	Result 1	Result 2	Result 3	Result 4
1	**12.41**	12.315	12.14	13.025	12.177
2	**12.41**	12.086	13.15	12.267	12.129
3	**12.62**	12.858	12.155	12.593	12.875
4	**12.45**	12.115	12.593	12.857	12.239
5	**13**	12.35	13.48	13.169	13.008
6	**12.3**	12.146	12.561	12.272	12.226
7	**12.93**	11.913	13.41	13.35	13.04
8	**12.47**	13.16	12.226	12.359	12.153
9	**13**	12.278	12.942	13.85	12.951
10	**13.14**	13.292	12.878	13.045	13.378
11	**12.85**	13.197	13.069	12.522	12.622
12	**12.85**	12.69	13.355	12.521	12.85
13	**12.67**	12.686	12.567	12.536	12.879
14	**12.77**	12.683	12.677	12.67	13.069
15	**12.6**	12.522	12.622	12.69	12.567
16	**12.94**	12.536	12.879	13.683	12.677
17	**13.03**	12.67	12.684	12.981	13.771
18	**12.85**	12.748	12.793	13.335	12.521
19	**12.99**	12.85	12.686	12.843	13.576
20	**12.9**	12.707	13.619	12.609	12.674
21	**12.73**	12.71	12.846	12.827	12.541
22	**12.89**	12.861	13.418	12.501	12.771
23	**13.01**	12.697	12.687	12.778	13.875
24	**13.06**	13.583	13.379	12.596	12.677
25	**12.96**	12.616	12.814	12.802	13.615

Table B.10: Unpack-VMware
Values in seconds