ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

SCHOOL OF INFORMATION SCIENCE

DATA CENTER VIRTUALIZATION FRAMEWORK FOR AMHARA REGIONAL HEALTH BUREAU.

By

HAILEMARIAM ADMASU

JUNE, 2018
ADDIS ABABA, ETHIOPIA

ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

SCHOOL OF INFORMATION SCIENCE

DATA CENTER VIRTUALIZATION FRAMEWORK FOR AMHARA REGIONAL HEALTH BUREAU.

A Thesis Submitted to School of Graduate Studies of Addis Ababa University in Partial Fulfillment of the Requirements for the Degree of Master of Science in Information Science

By: Hailemariam Admasu

Advisor: Workshet Lamenew (PhD)

June, 2018

Addis Ababa, Ethiopia

ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE

SCHOOL OF INFORMATION SCIENCE

DATA CENTER VIRTUALIZATION FRAMEWORK FOR AMHARA REGIONAL HEALTH BUREAU.

By: Hailemariam Admasu

Name and signature of Members of the Examining Board

Workshet Lamenew (PhD)	_____	_____
Advisor	Signature	Date
_____	_____	_____
Examiner	Signature	Date
_____	_____	_____
Examiner	Signature	Date

Declaration

This thesis has not previously been accepted for any degree and is not being concurrently submitted in candidature for any degree in any university.

I declare that the thesis is a result of my own investigation, except where otherwise stated. I have undertaken the study independently with the guidance and support of my research advisor. Other sources are acknowledged by citations giving explicit references. A list of references is appended.

Signature: _____

Hailemariam Admasu

This thesis has been submitted for examination with my approval as university advisor.

Advisor's Signature: _____

Workshet Lamenew (PhD)

Acknowledgement

First and foremost, I am thankful to the Almighty GOD, without his blessing I wouldn't have been writing even this acknowledgement.

I would like to give my deepest gratitude to my advisor Dr. Workshet Lamenew for his patience, enthusiasm and constructive comment throughout my thesis writing. I could have not imagined this work without his immense contribution. Thank you Doctor.

Besides my advisor I would also like to thank all school of information science, AAU staffs and classmates for the support you provide in one or another way.

I cannot pass the contribution of Amhara Regional Health Bureau staffs for their cooperation during my work. Special Thanks should go to Mr. Getachew Abe who helps me during data collection and experimentation.

I am highly grateful to my family members for their moral support and inspiration to finish my study.

Last but not the least; I am thankful to Almaz, Haimanot, Friends @Jemmo, Hawassa School of informatics staffs and others who directly or indirectly contributes for this success.

Dedication

This thesis is dedicated to my family.

And above all, to the Almighty GOD.

Abstract

This research endeavor is about data center virtualization which is a critical issue in recent periods. It is critical because virtualization an enabler to other technologies including cloud computing. The ultimate goal of a data center is handling the organization data appropriately. The research area is Amhara regional health bureau data center.

The regional health bureau runs physical data center in which a one application for one server architecture is exercised which have numerous barriers. Certain researches have been conducted on virtualization especially client virtualization, application virtualization and network virtualization. But, little attention was given for server virtualization.

The general objective of the study is developing a data center virtualization framework for the regional health bureau.

Design science research methodology was used which is intended to design the proposed framework. Problem identification and motivation, objective of the solution, design and development, demonstration, evaluation and finally communications of the design science research process steps are used. The general approach of the research is a design science approach in which Amhara regional health bureau was selected for a case and data was collected from information technology staffs of the regional health bureau using purposive sampling. Semi-structured interview, observation and document analysis were the data collection instruments used to gather appropriate data. After the qualitative data was gathered, it was finally analyzed using thematic coding analysis technique in which the result of interview was grouped in to main categories. From the analysis result it was clear that the physical data center is 1) not cost effective, 2) not scalable, 3) difficult to manage, 4) difficult to recover, 5) suffers for security issues etc.

Experiment was used to check the resource utilization of both the physical machines and virtual machines and compared. By using both the experiment and the analysis result, virtualization framework was developed.

In this study expert validation methods through focus group discussion were used to evaluate the developed framework. Hence the research finding is believed to be valid which assures the usability and applicability of the output of the study.

Keywords: data center, virtualization, design science.

Table of contents

Contents

List of Tables

List of Figures

List of Acronyms

ICT	Information communication Technology
IT	Information Technology
LAN	Local Area Network
NDC	National Data Center
VMware	Company Providing Virtualization Software
VMs	Virtual Machine
UPS	Uninterruptible Power Supply
ISP	Internet Service Provider
ROI	Return On investment
DSRM	Design Science Research Methodology
DS	Design Science
ARHB	Amhara Regional Health Bureau
HVAC	Heating, Ventilating and Air-conditioning systems
CRAC	Computer Room Air Conditioning
VM-CP	Virtual machine Control Program
MVS	Multiple Virtual Storage
LPAR	Logical Partitioning
IBM	International Business Machine
JES	Job Entry Subsystem
PDU	Power Distribution Unit
VPN	Virtual Private Network
CIA	Confidentiality Integrity Availability

CHAPTER ONE

INTRODUCTION

1.1 Background of the study

In today's rapidly changing and competent world, data is the key to survival of both public and private organizations. The success and failure of an organization directly or indirectly depend on the availability and usage of data. So the desire to success and the competition among different organizations and also the fierce in security aspect forces organizations to build their own data centers.

Data center is the pillar for different services delivered through internet and intranet like web hosting, ecommerce, networking, software as a service, platform as a service, and grid/cloud computing. Among some of those generic platforms Microsoft azure platform, Google App engine, amazon's elastic cloud platform and sun's grid engines are the common ones and virtualization is the base for providing those services and increasingly used in data centers to provide better service utilization and flexible resource allocation even it makes data management challenging (Kant, 2009; Lowe et al., 2016).

Data center is not a choice rather an integral part and secret of success for modern organizations. It can be considered as an area that holds, a means of hosting critical data, applications, and servers, as well as contains basic assets of customer information, intellectual property, and other business critical data.

Data centers require high availability to run the business. But not all data centers meet the underlined criteria specified by ANSI/TIA-042 standards. According to this standard data centers can be classified into four tiers/ratings (Hossain, 2011).

Rated-1/Tier-1, basic site infrastructure with a down time of 28.8 hours a year, 99.671% available, data center which has single path for power and cooling distribution, no redundant component. Amhara regional health bureau is included in this category.

Rated -2/ Tier -2:- redundant capacity component site infrastructure with downtime of 22 hours per year, 99.741% available, a data center which has a single path for power and cooling distribution including redundant component.

Rated-3:- concurrently maintainable infrastructure with a downtime of 1.6 hours a year, 99.982% available a data center which has multiple paths for power and cooling distribution path with only one active.

Rated -4 fault tolerant site infrastructures with down time 0.8 hours a year 99.995% available, a data center which has multiple path for power and cooling distribution path with all are active.

A virtualized data center on the other hand is a data center where some or all of the infrastructures like server, network, application and desktop are virtualized (Hossain, 2011). Virtualization specifically server virtualization is a technology that helps us to install different Operating Systems on the same hardware. To do this a hypervisor is needed which is a thin software layer that intercepts operating system calls to the hardware. Virtualization hides the physical characteristics of computing resources from their users, their applications or end users. This includes making a single physical resource (such as a server, an operating system, an application or a storage device) appear to function as multiple virtual resources.

Currently data centers are being virtualized in different parts of the world. Amhara regional health bureau is one of a government organization which has a physical data center for accomplishing its mission under Ministry of Health supervision.

The regional health bureau data center has network active devices including two Aruba wireless LAN controllers, six active servers, one Cisco 6508 core switch and 9 access switches. It has also uninterrupted power supply (UPS), floor stand air conditioning system and racks for server placement as critical data center components.

HRIS (human resource information system), HMIS (health management information system) and smart card systems are among the applications used in the data center. Human Resource Information System/HRIS is a set of inter-related components working together to collect, process and store information to support Human Resource/HR decision making, coordination and control in an organization. Health Management Information System (HMIS) is the routine

collection, aggregation, analysis, presentation and utilization of health and health related data for evidence based decisions for health workers, managers, policy makers and others.

List of functional server in the data center

No	**Servers**	**Specifications**
1	Proxy server HP Proliant DL385 G6 Windows server 2008 R2 SP1	RAM 8GB, Processor 1 CPU 2.4 GHz Six-Core AMD Opteron™ Processor 2431, HDD (storage capacity) 600GB.
2	Windows Domain controller HP ProLiant DL385 G7 Windows server 2008 R2 SP1	RAM 32GB, Processor 2 CPU AMD Opteron (tm) Processor 6128, HDD (storage capacity) 300GB.
3	Anti-virus HP ProLiant DL360 G6 Windows server 2008 R2 SP1	RAM 4GB, Processor 1 CPU 2.0 Ghz Intel Xeon E5504 @2.00Ghz, 6128, HDD (storage capacity) 500GB.
4	Application server HP ProLiant DL360 G6 Windows server 2008 R2 SP1 (Running EHMIS)	RAM 8GB, Processor 2 CPU AMD Opteron (tm) Processor 6128, HDD (storage capacity) 300GB.
5	Application server HP ProLiant DL385 G7 Windows server 2008 R2 SP1 (Running HRIS)	RAM 8GB, Processor 1 CPU 2.0 Ghz Intel Xeon E5504 @2.00Ghz, 6128, HDD (storage capacity) 500GB.
6	Application server HP ProLiant DL360 G6 Windows server 2008 R2 SP1 (Running EMR)	RAM 8GB, Processor 1 CPU 2.0 Ghz Intel Xeon E5504 @2.00Ghz, 6128, HDD (storage capacity) 500GB.

Table1.1: List of physical servers in the health bureau

1.2 Statement of the problem

The organizations data exponentially increases from day to day and demands highly comparable infrastructures (including servers, networks, cooling equipment's and others). Even though those infrastructures are expensive, the regional health bureau forced to fulfill all requirements since the bureau is highly dependent on it.

Due to enhancement of technology, infrastructures are developed with high processing power but a small part of it about 10 to 15 percent is actually in use. This resulted in less utilization of the available resources even the space requirement and power usage continues. But by implementing virtualization almost 75 to 80 percent of the resources should be utilized which makes a better sense in many directions.

Hossain (2011) emphasized that by reducing 50 servers in a certain data center over 3 years we can lessen 1,790,982 Lbs (895tones) CO2, 37 Lbs methane, 25 Lbs nitrous oxide, 7143 Lbs sulfur dioxide, 2239 Lbs nitrogen oxide that is equivalent to 177 passenger not driven for one year. So, it is risky for environment sustainability and leads to global warming. It is also difficult for developing countries like Ethiopia to provide energy to these ever-growing data centers.

Certain researches have been conducted on virtualization especially client virtualization, application virtualization and network virtualization. But, as literature shows little attention was given for server virtualization. Although virtualization has its own overhead, it should not be a matter of choice these days because of its giant importance. Even the conducted researches have not adopted successfully in our country. The nature and practices of one organization might not fit with other organizations. The following non-local researches have been mentioned.

Calzolari (2006) has done his research on "High availability using Virtualization". It is on 3RC, which is the name of the project, acronym for 3 Re Cycle found in Geneva. The main objective of his work is achieving redundancy using virtualization technology. The methodology he followed is not clear and recommends additional work to improve the system performance.
Nyamweya (2013) develops a framework on virtualization in selected government ministries of Kenya. Exploratory method was used in which the Ministries of Energy and Finance, and also the Directorate of e-Government were used as cases. While developing the integrated

framework, implementation model the virtualization method is not mentioned and the framework is not self-descriptive.

I.T. Alliance (2010) has established framework for server virtualization and transformation projects. Here virtualization provides improved cost of ownership and can deliver cost savings through a reduction in servers and resources, such as space, power and cooling. But the final developed framework is poorly defined and doesn't show the required components and also looks a serious of steps than a framework.

Currently, there are limited researches on virtualization conducted in Ethiopia. Getnet (2017) conducted a research that has been focused on developing framework for banks specifically for Wogagen bank but his focus area is client or desktop virtualization which has completely different intention than server virtualization. Desktop virtualization is a technology in which the user's desktop is stored on a remote server allowing the user to access his desktop from any device and location using his/her username and password for the purpose of minimizing data theft. But server virtualization is partitioning of a physical server into a number of small, virtual servers with the help of virtualization software.

More over the earlier developed virtualization frameworks has limitations hence, did not bring successful changes. Because of this fact, local researches should be done on virtualization.

This research therefore, is geared to address the current problem of non-optimum usage of IT resources by developing an appropriate virtualization framework because there is no universal and best framework for virtualization.

1.3 Objective of the study

1.3.1 General Objective

The general objective of the study is to develop a data center virtualization framework for Amhara Regional Health Bureau.

1.3.2 Specific Objectives

The specific objectives devised to accomplish the general objective include:

- To explore the current traditional or physical data center in the bureau.
- To review literatures related with the general objective of the study.
- To identify components of the virtualized data center framework.
- To evaluate the effectiveness of the virtualized framework.
- To draw conclusion and forward directions for future the work.

1.4 Significance of the Study

This research is believed to produce a result that allows the Health Bureau and also other similar organizations to implement a data center virtualization appropriately to effectively solve the existed problems in the current data center and making scalable, reliable and efficient data center.

The result of this thesis will contribute to the ongoing researches in this domain area.

Since virtualization is a step for cloud computing it helps organization to transform for this giant technology.

1.5 Scope of the Study

- Because of Time constraint the thesis mainly focuses on developing a virtualization framework for Amhara Regional health bureau.
- It particularly focuses on hardware assisted server virtualization. So, some results of the thesis may not be applicable for other types of virtualization types like application, desktop/client and network virtualization. Since the regional health bureau is on an infant stage and on the way of expanding, the scope of the research was limited to server virtualization.
- The study does not consider the implementation of virtualization rather exploring the current data center and developing a virtualized framework.

1.6 Structure of the Thesis

This thesis report is organized in to six main chapters.

Chapter 1: is introduction which explains about data center, virtualization, Statement of the problem, objectives, significance and scope of the study.

Chapter 2: covers literature from different sources that support the work of the researcher. It covers main topics like data center, data center evolution, virtualization, types of virtualization, methods of virtualization, and the advantages of virtualization from different perspectives. **Chapter 3:** is generally about methodology like research design, sampling technique, data collection instruments, procedures etc. The researcher started by describing design science research methodology as the appropriate methodology for creation of artifacts.

Design science research methodology had six process models: problem identification and motivation, define objectives for a solution, design and development, demonstration, evaluation and lastly communication.

Chapter 4: is results and discussion: which defines the results after data is collected from the respected bodies and verified. It is also about experimentation which shows the importance of virtualized data center over physical data centers by using open source software.

Chapter 5: This chapter is about the designed artifact or framework which it's various components along with the evaluation mechanisms.

Chapter 6: is the last chapter which is the section of concluding the researchers result and giving direction for future work.

CHAPTER TWO

LITERATURE REVIEW

2.1 Overview

In this chapter different related literatures from different parts of the globe were reviewed. The researcher strongly finds related literatures that help to understand the theoretical background of data center virtualization using different search parameters. Google, Google scholar and IEEE were used to explore literatures.

2.2 Green Information Technology

Information and communication technology directly or indirectly enters into businesses and social practices and transformed it to e-businesses and e-economy so that it has the potential to create sustainable environments. Information communication technology products on the other hand contribute the lion share for global warming and ecological deteriorations. In recent years different business organizations witnessed to manufacture and use environment friendly products and technologies to dramatically decrease those hazards and achieve a sustainable business environment. Environmental impact of Information Technology (IT) under the banner of "Green IT" has started and being discussed by academia, media and government to encounter global warming and other environment offensive activities, but the consequences even growing (Uddin et al., 2012).

As explained by Murugesan (2008), information technology contributes to environmental problems even if peoples do not realize it. Green IT is a hot issue in today's world and should continue for the next years because IT affects the environment in several ways starting from computer production, throughout its use to final disposition. Every personal computer (PC) in use generates a ton of carbon dioxide every year. Even now a day's electrical consumption of servers, computers, monitors, data communication equipment's, and cooling systems in a data center significantly increasing. So greening the environment is not a matter of choice rather it should be individual's responsibility. Green IT refers to environmentally sound IT. It's the study

and practice of designing, manufacturing, using, and disposing of computers, servers, and associated subsystems such as monitors, printers, storage devices, networking and communications systems efficiently and effectively with minimal or no impact on the environment. It spans various focus areas including design for environmental sustainability, energy efficient computing, power management, data center design (layout and location), server virtualization, environment related risk mitigation and the like (Murugesan, 2008). From the above excellent explanation virtualization is one part of green IT.

In order to reduce these environmental problems the IT sectors, including other computer users should green their IT systems and the way they use the systems. All the peoples in the world are legally and ethically obliged to offer proper attention on the IT products, applications, services and practices.

Greening IT has the following major components.

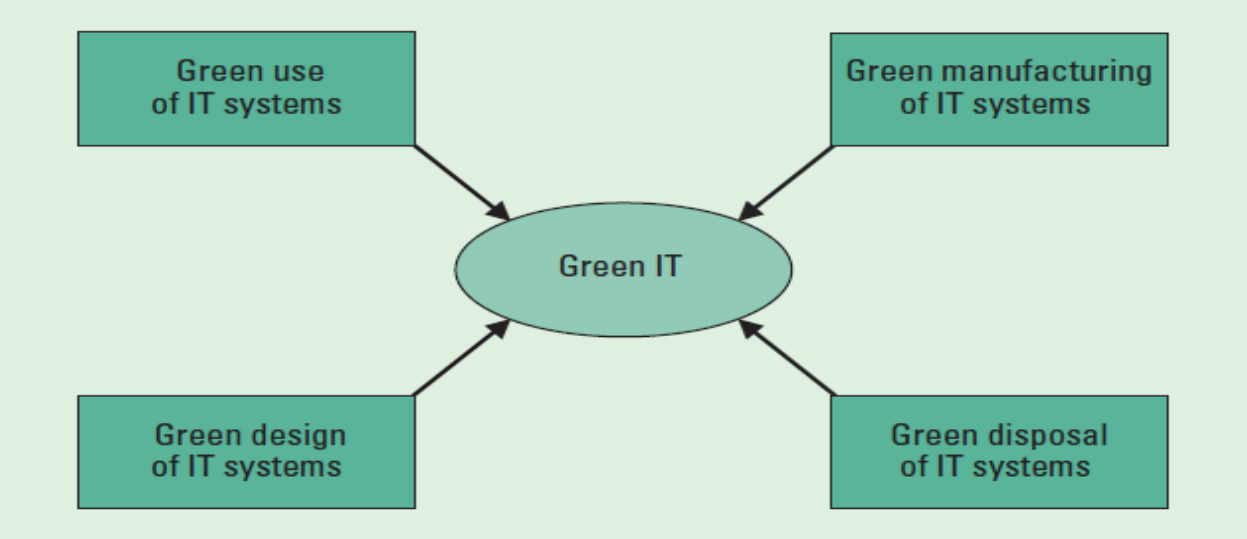


Figure 2.1: Holistic approach for green IT (Source: Murugesan, 2008).

The following figure also depicts the major reasons and benefits of using green IT.

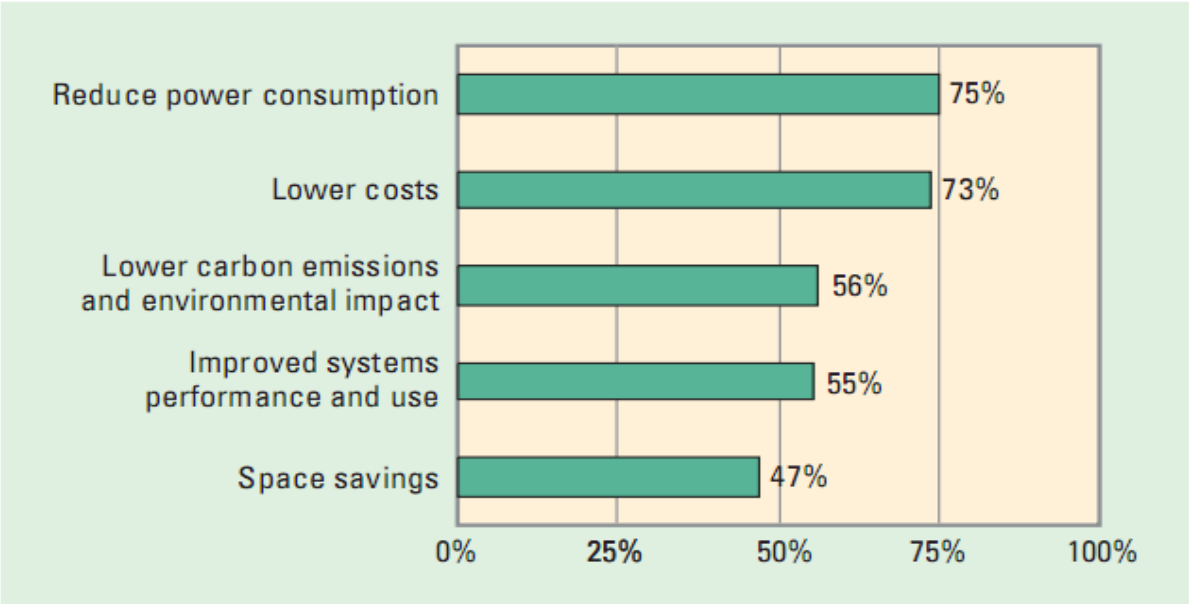


Figure 2.2: Reasons and benefits for using green IT practices (Source: Murugesan, 2008).

2.3 Data Center

A data center as a physical environment facility intended for housing computer systems and associated components. It comprises of components like servers, storage devices, chillers, switches, uninterruptible power supply (UPS), computer room air conditioners, direct expansion (DX) units, pumps, cooling tower, generators, power distribution units (PDUs), fire suppression, physical security devices and lightings which accounts approximately 2% of total global energy (Balodis & Opmane, n.d.; Koomey, 2007). Anixter (2012) also defines data center as a "nerve center" for both new economy and a special environment.

A data center has a serious of history and includes standards from tier one with small uptime to tier four which have ultimate uptimes (Balodis & Opmane, n.d.).

2.4 Data Center Evolution

According to Lowe et al. (2016), the need for consolidation in the data center didn't just occur overnight but it takes a long period of time and passes through different stages. It is also difficult to determine the exact date in history when data center was existed. The evolution started from the production of the first computer and encompasses different consolidation mechanisms. Consolidation is the process of bringing together separate and individual parts in to a unified

whole that has a great advantage for managing and minimization of complexities. Consolidation allows data center infrastructure employees to manage their data center more effectively and more consistently, thereby reducing the cost of managing the data center and reducing your total cost of ownership (TCO).

To better understand how we got to a point where consolidation has become necessary, it's worth taking a look at the evolution of today's computing environment. The data center has on the tremendous evolution for more than four decades. The emergence of different types of computers like mainframe computers, minicomputers and distributed computing plays a significant role in the history of data center evolution.

2.4.1 Mainframe Computers

Mainframes computer were the first computer to gain wide acceptance in commercial areas. The major vendors were known as IBM and the BUNCH (Burroughs, Univac, NCR, Control Data, and Honeywell). These players were controlled the commercial market for almost all U.S. companies for many years even if IBM is the only vendor survived in recent years.

The Strength of Mainframes

The strengths of mainframes make them valuable components to nearly every large-scale data center. These strengths include (Lowe et al., 2016).

- **Power**

Mainframes were the most powerful computers available, and each new generation got efficient and faster. Over the past several years the power and performance of distributed computing systems have improved dramatically and mainframes still play an important role in some data centers.

- **High utilization rates.**

Due to high expense for purchasing and also for building a data centers for it, users tend to use every bit of available computing power reaching peak utilization rate of 90 percent.

- **Running multiple applications through workload management.**

Since large amount of budget is required for purchasing mainframes many companies indebted to run multiple applications on a single machine.

2.4.2 The Introduction of Minicomputers

1970s and 1980s was the year for the entrance of attractive and portable mini computers to the entire market. Minis are much smaller and less expensive compared with mainframes and designed as scientific and engineering computers so that mostly used to run business applications.

The major players in this market were DEC, HP, Data General, and Prime. Initially, companies developed applications on minis because it gave them more freedom than they had in the mainframe environment (Lowe et al., 2016).

2.4.3 The Rise of Distributed Computing

Next to mini computers, the world revolves around a distributed system. As early users of UNIX systems moved out of undergraduate and postgraduate labs and into the corporate world, they wanted to take the computing freedom of their labs into the commercial world, and as they did, the commercial environment that they moved into evolved into today's distributed computing environment.

The availability of all the major Operating Systems on small and low servers is one of the characteristics of distributed computing environment. This characteristics permits various corporate groups (departments, work groups, etc.) to purchase servers outside the control of the traditional, centralized IT environment. As they shared applications with others in their departments, their workstations became servers that served many people.

Although distributed system provided great freedom of computing, it was also a major origin of the complication that has led to today's major trend towards consolidation.

Generally after a dramatically growing of computing power and development of different sophisticated software's, IT was able to provide simple to manage and more cost effective end

user computing resources by placing dumb devices as the endpoint and keeping the computing power in the data center (Lowe et al., 2016).

2.5 Characteristics of Basic Data Center Facilities

2.5.1 Electricity Supply System

Both IT equipment and other supporting facilities which consume the largest part of electricity are the two most critical aspects in the data center (Balodis & Opmane; Newcombe, n.d.). Even some data center facilities have power densities exceeding 100 times more than the power consumption in a typical office use. For higher power density facilities, electricity costs are a dominant operating expense and account for over 10% of the total cost of data center's ownership (TCO).

The different equipment's in the data center like cooling equipment, access control systems, backup power systems consisting of one or more uninterruptible power supplies (UPS), and/or diesel generators also consume a significant amount of electric power. Single point of failures arise if no redundancy is installed for most equipment's so, to prevent this situations all elements of the electrical systems, including the backup systems should be duplicated and computing facilities are connected to both power feeds as N+1 redundancy arrangement (Balodis & Opmane, n.d.; Anixter, 2012).

In addition to IT equipment, the other energy consuming equipment mainly consists of cooling systems, power delivery and other facility's infrastructure, like lighting. Because of that often IT equipment's consume only 33% of the power in the data center (Balodis & Opmane, n.d.). The most popular power components include generators, uninterruptible power supply systems (UPS), grounding, and power and environmental control.

Where does the energy go?

As explained by Rasmussen (2006), approximately half or less energy is consumed by the IT loads and half is going for other critical infrastructure including cooling equipment, power equipments, and lightening. This result is more closer with results of Balodis experiment explained above.

Now a day's electrical power usage become the total cost ownership (TCO) for most of the data center. It will also continue unless different data center design technologies used. Even though electric power consumption (KW) and data center efficiencies are used interchangeably as a metrics, researchers like Rasmussen (2006), recommends to use the former since it is un ambiguous. According to him saying two different devices with 50% and 80% efficiences is not clear to combine them in to a single number. Reductions of power consumptions of IT systems can be through various approches like operational actions and planning actions. Operational actions is held by retiring systems, using energy efficient equipments or by migrating to more energy efficient platforms. The other approach planning actions is reducing electrical power consumptions through virtualization and standardization.

Operational: Retiring IT systems

In many data centers there might be application servers but with no users. So, it is wise to take an inventory for a retirment plan and even we can make offline and powered down with out physically retire it.

Operational: Operating existing systems in an efficient manner

Many years ago servers did not have power management features meaning the power consumption of all devices are constant ragardless of their computational load. But, most of today's server has a power management features disabled by default. So, it is a responsibility of a system administrator to on this feature shaping the power consumption of the device according to computational loads. Even if power management features do not reduce the power requirement it significantly reduce the total electric usage.

Operational: Migration to energy efficient computing platforms

This is simply a method of migrating applications from high electric power consumer platform to low electric power consumer platform like using a single processor dual core server to replace two or more old servers

Planning: virtualization

Virtualization significantly decreases the power requirements of a device as well as a data center in general. Virtualization dramatically decreases by reducing the number of physically installedservers. Thesis concept is widely discussed in the next section.

Planning: Standardization

Even though virtualization is not used, standardization of energy efficient servers is an appropriate approach. As different study discloses, blade servers are the most electrically efficient form of server. But performance and power efficiencies may not much with the organization preferences meaning it is difficult to predict in advance the performance requirement for a server-based application, so users often specify the highest available performance at a substantial power consumption penalty.

When servers are virtualized, the strategy of using the highest performance server is generally the best approach to minimize overall power consumption. However, when servers are deployed application-by-application it can make sense to match the server performance to the application requirements to save energy.

For users that standardize on a blade server system and deploy servers per application, the option exists to standardize on two blades, a high performance / high power blade and a lower performance / lower power blade. A logical strategy is to deploy applications on the lower performance blade by default and only move to a higher performance blade if the need is demonstrated. This is facilitated by the ease of provisioning of blade servers.

Energy Efficient System Design

As Rasmussen (2006) described, people thought that electrical power consumption is regoresly decresed only by reducing the power consumption of individual devices. That is not always true and system design also has a great effect on the power consumption of data centers. Two data centers with the same IT equipment may have different electrical bills. So this shows that system design is even more important than the selection of power and cooling elements.

2.5.2 Low-Voltage Cable Routing

The availability of different equipment's in the data center including switches (access or core), ports and routers does not mean it is functional without communication among them. Communication is very mandatory and it is based on networking running internet protocols and special protocols for computing equipment interconnection. These ports, routers and switches transport traffic between computing equipment in the data center and the outside world. Redundancy of internet connection in the data centers is through using multiple internet service providers (Telecoms) but unfortunately does not exist in our country (Rasmussen, 2006; Anixter, 2012).

Data cabling is usually routed through overhead cable trays in modern data centers but, some are still recommending under raised floor cabling for higher security advantages and to consider additional cooling systems above the racks.

Structural cabling elements include service provider's cables, backbone cables (to data center and in between floors, horizontal cables (within floors) and zone distribution which is in a data center).

2.5.3 Raised Floors

The floor arrangement of the data center contributes more in different aspects like heat elimination. Mostly data centers have raised flooring made up of 60 cm removable square tiles but the trend is towards 80-100 centimeters void to provide better and uniform air distribution within it. These void spaces also used as part of air conditioning and provide plenty of spaces for cabling. Raised floors have a greater advantage than overhead cabling because it provides better cooling and also provides sufficient flexibility in location of cooling equipment's (Rasmussen, 2006).

Overhead cable trays are less expensive than raised floor and also provide more flexibility for supporting racks of varies heights. Simplicity in adding and removing racks in accordance with the requirement is also an added advantage for overhead cable trays arrangement.

2.5.4 Environmental Controls

Environment contributes a lot for better functionality and/or mal functionality of different equipment's in the data center. So, standard cooling system balances the temperature and humidity within the data centers. Different data center equipment's generate heats while in work and needs for sudden removal of it. Unless the generated heat is not removed the temperature is raised resulting in malfunctioning of different equipment's. The air conditioning systems balances the actual temperature with the manufacturer temperature/humidity range of servers and other facilities in the data centers (Rasmussen, 2006; Anixter, 2012).

Trends

Modern data centers try to use economizer cooling, where they use outside air to keep the data center cool. Many data centers now cool all of the servers using outside air. The industry is exploring progressive cooling solutions because the current generation, discussed earlier, has increased computing requirements. (Chillers, for instance, are estimated to consume 33% of a facility's total power in different layouts).

To optimize the cooling in the data center a good first step is an in-depth analysis of the current environment to gain a holistic understanding of the data center's environment, increased awareness of the critical risk factors, benchmark of performance metrics, and generate a punch list of opportunities for cooling improvement (Rasmussen, 2006).

2.5.5 Fire Protection

Fire is the most devastating hassle in both traditional and modern data centers. Different fire protection systems can be used including active and passive design elements as well as implementation of fire protections programs.

Smoke detectors are the most important system installed to provide early warning of a developing fire by detecting particles generated by smoldering components prior to the development of flame. This excellent alert allows investigation, interruption of power, and manual fire suppression using hand held fire extinguishers before the fire grows to a large size. A

fire sprinkler system is often provided to control a full scale fire if it develops. Fire sprinklers require 46 cm of clearance (free of cable trays, etc.) below the sprinklers.

Clean agent fire suppression gaseous systems are sometimes installed to suppress a fire earlier than the fire sprinkler system. Those fire protection elements can be active or passive. Active fire protection systems includes tasks like closing doors but, Passive fire protection elements include the installation of fire walls around the data center so a fire can be restricted to a portion of the facility for a limited time (Rasmussen, 2006; Anixter, 2012).

2.5.6 Security

Security is a critical aspect especially in this technological era. It is the life blood of information and should be treated as a customer data in a bank vault. Although security has different forms, physical security which is restricting physical access to sites should be the first activity. Physical access to the sites is usually limited for authorized personnel's only (Rasmussen, 2006).

Video camera surveillance (CCTV), card access control and permanent security guards have been always present if the data center is large or contains sensitive information on any of the systems within. The use of finger print recognition mantraps is starting to be commonplace (Anixter, 2012; Rasmussen, 2006).

Network security elements like firewalls, VPN gateways, intrusion detection systems, authentication and authorization are also deployed for securing the data center from different internal and external threats.

Lowe et al. (2016) stressed that, protecting the companies' data is the most challenging aspect in the data center. Delivering some data for different user at any time may be important but security becomes in danger. For this conflict the CIA Triad is mandatory.

Figure 2.3: The CIA Traid (Source: Lowe et al., 2016).

Confidentiality is sometimes called privacy and implemented by authorization, authentication and accounting (AAA). Authentication ensures who is going to accessing the data but authorization refers to what data can they access. Accounting on the other hand assures logs about what data is accessed, by whom and when.

Integrity: refers to the protection of data modification by unauthorized user through AAA and data encryption as data moves from place to place. High consideration should be given for integrity recently than ever before.

Availability: availability means the company data should be available to end users and application whenever and wherever needed.

2.5.7 Data Center Spaces

Data center can have different rooms. so it may have room for telecommunication equipment's, room for operation center, entrance facilities, rooms outside computer rooms for mechanical and electrical spaces for power generators, UPS, cooling refrigerants, structured cabling and communication entry cabling rooms may exist.

Even if rooms are different in accordance with the type of data centers the most important spaces in data centers include: computer room, computer room TR, operations center, entrance facility (Rasmussen, 2006).

2.6 International Standard for Data Centers

According to ADC Telecommunications (2006), standards are the basic instrument for the proper establishment, and effectively advancement of data centers. Few decades before cabling standards were the cornerstone for the proper design, installation and performance of networks. Telecommunication industry association (TIA-568) for commercial building telecommunications including the design, installation and performance requirements for telecommunication cabling systems were the first TIA standards.

Historically network administrators have been faced with different challenges in choosing technologies and also implementing it opening the gap for security and reliability issues. This is mainly due to lack of properly established standards.

TIA-942 was established in April 2005 by adding different features that specifically address data center infrastructure. This is also the first standard which specifically addresses data center infrastructure which includes the following:

I. Site Space and Layout

TIA-942 stated that proper space allocation for data centers starts with guaranteeing appropriate space reallocations for the future growth. It also recommends the designers to balance the initial costs of deployment with the space requirement for the reallocation of servers when the data center grows. The data center should have plenty of white spaces/ empty spaces for accommodation of cabinets and racks. Even the space around the data center should be appropriately examined to guarantee future expansions. Generally TIA-942 includes functional areas like main distribution areas, one or more horizontal distribution areas, equipment distribution areas, zone distribution area, back bone and horizontal cabling, and one or more entrance rooms.

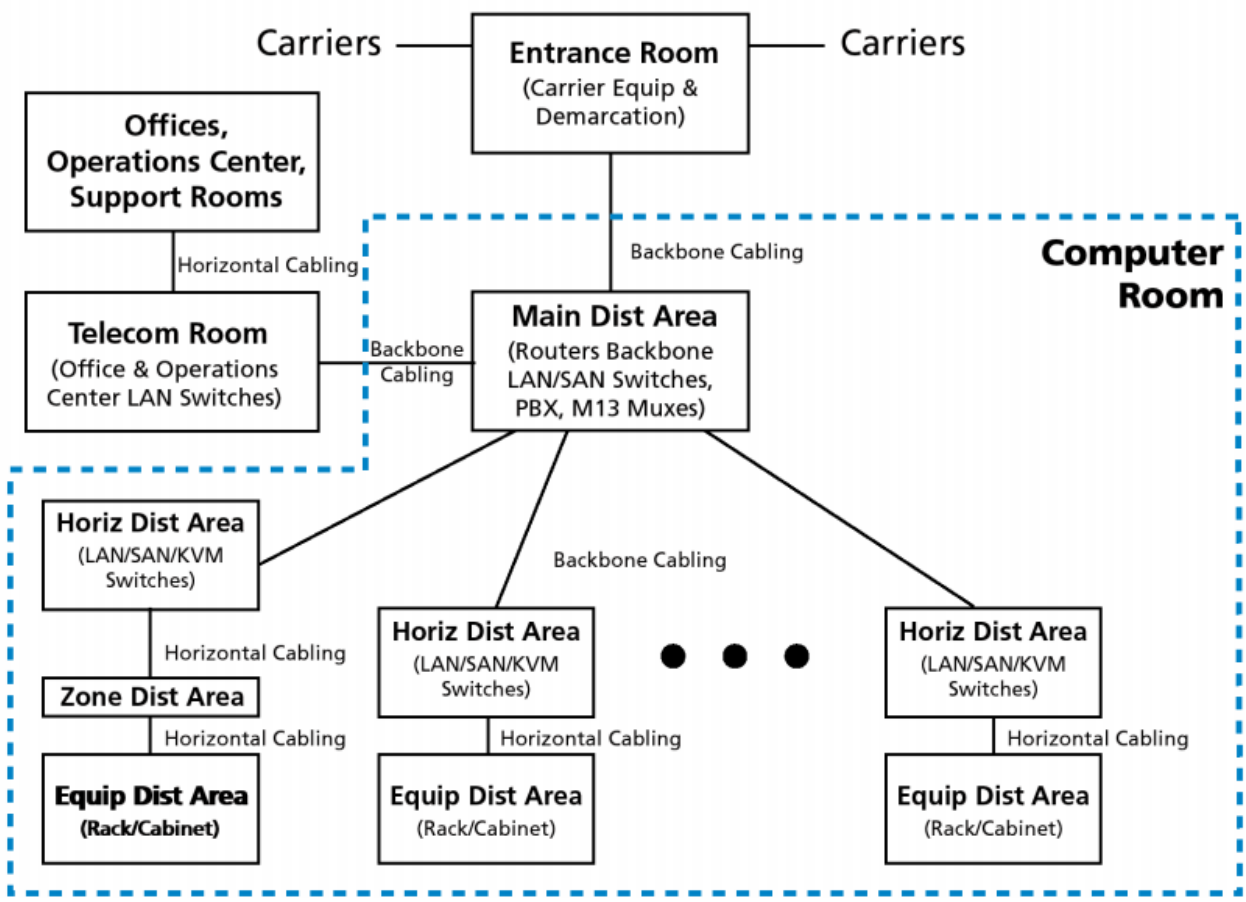


Figure 2.4: TIA-942 compliant Data Center showing key functional areas (Source: ADC Telecommunications, 2006)

II. Cabling Infrastructure

TIA-942 specifies a generic cabling system for data centers and also provides specifications including standard single mode fiber, 62.5 and 50um multimode fiber, 75-ohm coaxial cable, 4 pair UTP and ScTP cabling and laser-optimized 50um multimode fiber(recommended)

Laser-optimized 50um multimode fiber is recommended for backbone cabling because it is capable of supporting higher network speeds over long distances. UTP 6 is also recommended for horizontal cabling in many cases. Anixter (2012) stressed that appropriate cabling media should be selected not only for current need but also future requirements. Hence, standard cable limitations and performance have to be considered to prevent from additional cost and service disruption.

III. Tiered Reliability

Tiers are based on the information from uptime institute for improving the design and management of data centers. Tiers descriptions also include information like raised floor heights, watts per square fit and points of failure. Datacenter standards from tier 1 to 4 are used as a methodology of defining uptime of a data center (Hossain, 2011; ADC Telecommunications, 2006). These standards are useful for measuring data center performance, investment and return on investment.

Tier level one

This is the first tier which is single non-redundant distribution path serving the IT equipment. It is also a non-redundant capacity component with basic site infrastructure guaranteeing 99.671% availability.

Tier level two

It fulfills all tier 1 requirements with redundant site infrastructure capacity components guaranteeing 99.741% availability.

Tier level three

Tier three fulfills all Tier 1 and Tier 2 requirements with multiple independent distribution paths serving the IT equipment. All IT equipment must be dual-powered and fully compatible with the topology of a site's architecture. This is concurrently maintainable site infrastructure guaranteeing 99.982% availability

Tier level four

This tier is the last tier according to the uptime institute and fulfills all Tier 1, Tier 2 and Tier 3 requirements. It is a fault-tolerant site infrastructure with electrical power storage and distribution facilities guaranteeing 99.995% availability. All cooling equipment is independently dual-powered, including chillers and heating, ventilating and air-conditioning (HVAC) systems

IV. Environmental Considerations

Several environmental considerations are included in the TIA-942 standards. Humidity levels, operating temperatures, fire suppression; architectural electrical and mechanical system specifications are some of the components. This section focuses on the formation of proper environment for the data center especially by removing heat generated within it.

The standard provides specifications on encouraging proper air circulation and removal of the generated heat.

2.7 Data centers Design Models

As concisely explained by Hossain (2011), data center design model is a prerequisite for having a best model and selecting the best topology. Data center design models can be Multi-Tier model, application design models and web service models.

Multi-Tier Models

Most of the multi-tier models are dominated by HTTP applications in a tier approach indicating that most web based applications are mostly Multi-tiered. These approach consists of application servers, web servers and database servers by running a software on the same machine as inter process communication (IPC) or by a network.

Basic layered Design

The following diagram shows the most widely used network design in a data center showing the placement of different networking devices. There is also collapsed core architecture in which the core and aggregation layers are merged in to one. As the diagram shows the design includes the core, aggregation and access switches. In the aggregation layer the core switches are redundant which means one core switch take over the task if the other down (Hossain, 2011). For better understand each layer see figure 2.5.

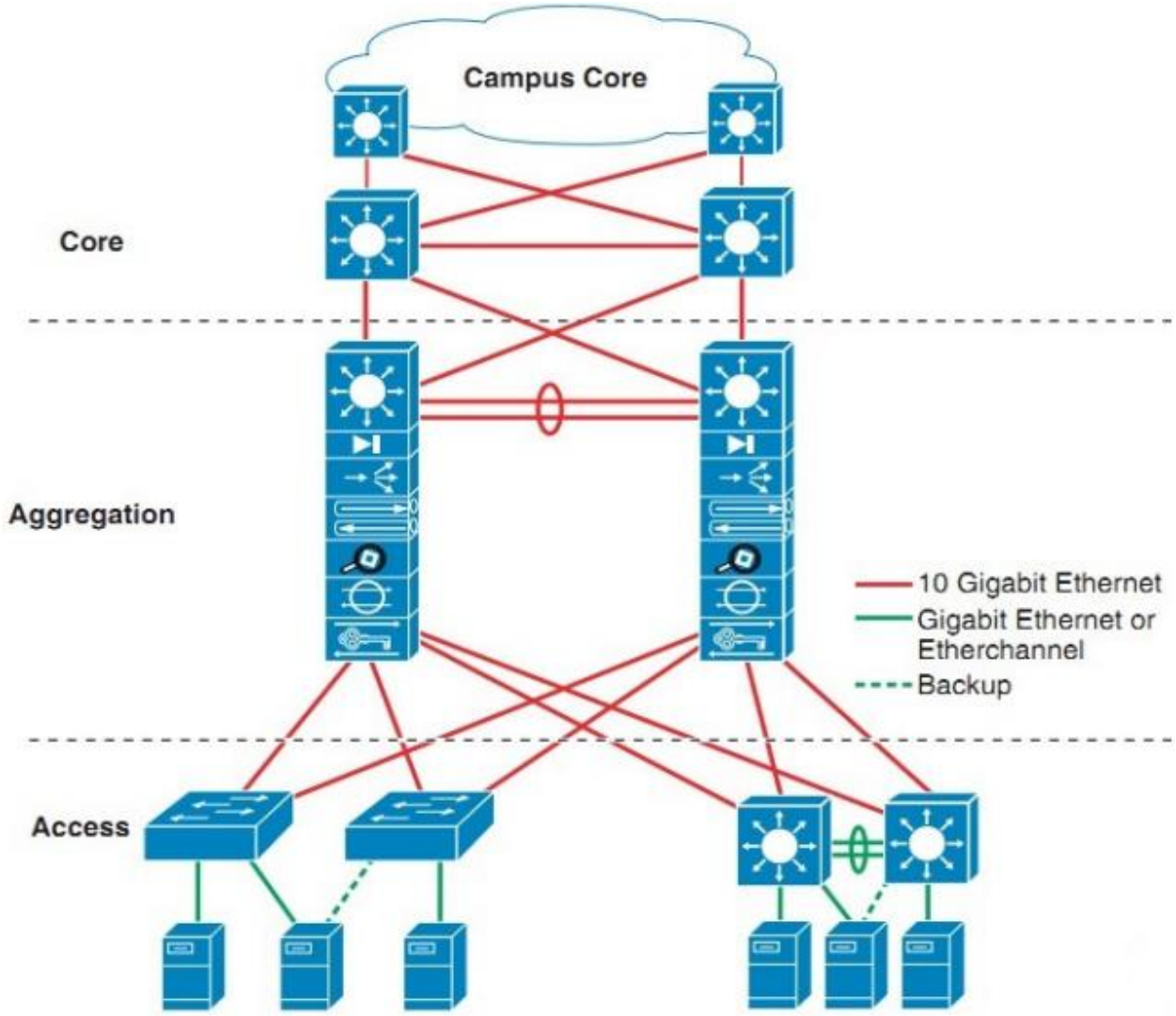


Figure 2.5: Basic Layered design (Source: Hossain, 2011)

Application Design Models

Application design models can be 1 tier, 2 tier, 3 tier and N-tier models.

One Tier Model

One-tier architecture: is used to describe systems in which all of the processing is done on a single host.

- Users can access such systems (aka mainframes) through display terminals (aka dumb terminals) but what is displayed and how it appears is controlled by the mainframe. The terminals have no storage and local processing database systems.

Figure 2.5a one tier Architecture

Two Tier Model

Two-tier architecture: is used to describe client/server systems, where clients request resources and servers respond directly to these requests, using their own resources

Or

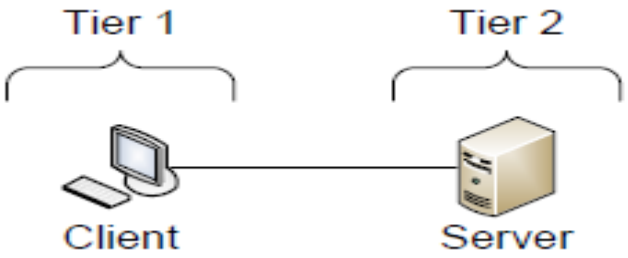


Figure 2.5b Two tier Model

Three-tier Model

It is used to describe client/server systems consisting of:

- Clients which request services
- Application servers whose task is to provide the requested resources, but by calling on database servers.
- Database servers which provide the application servers with the data they require.

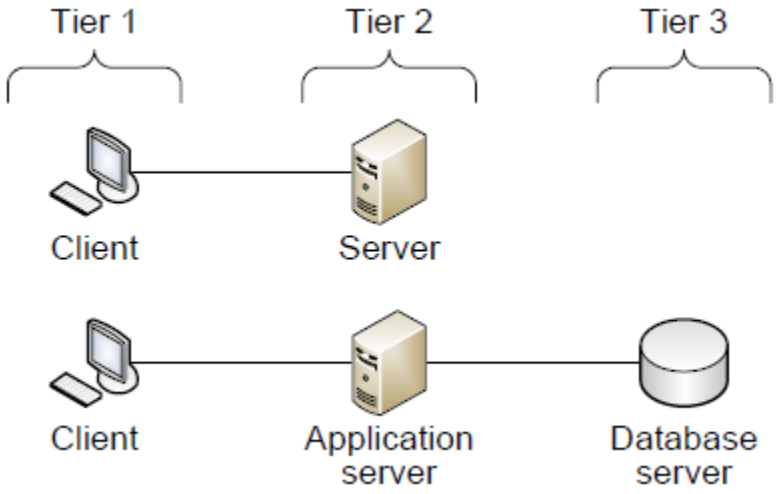


Figure 2.5c: Three Tier Model

N-tier Model

N-tier architecture: is used to describe client/server systems consisting of more than 3 tiers.

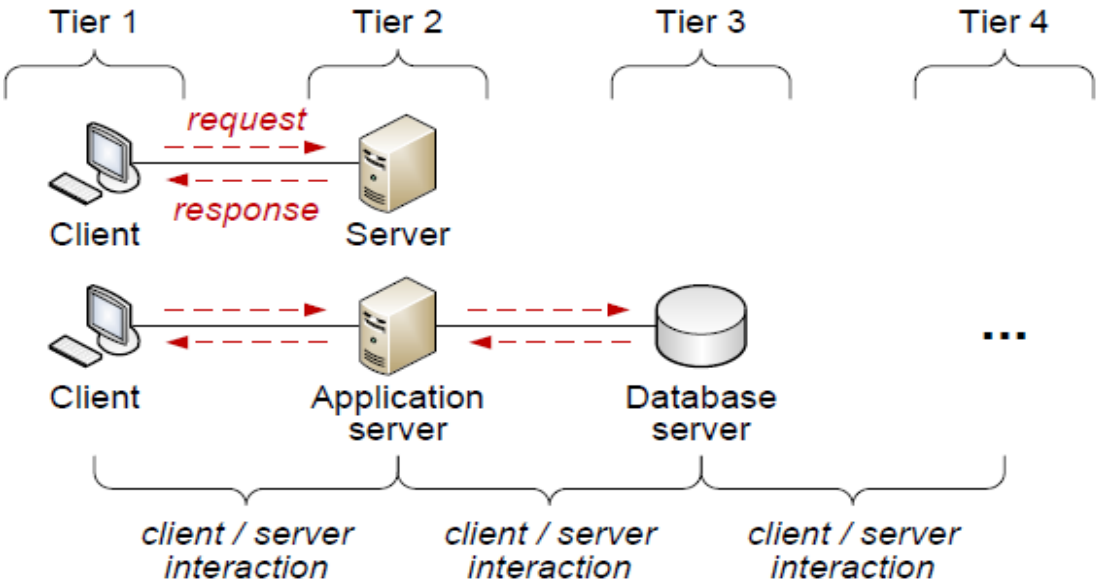


Figure 2.5d: N-tier

Generally it is better to look the summarized characteristics of each tier with the following table.

Tiers	**Characteristics**
Tier One	✓ limited scalability and lacks flexibility
Tier Two	✓ Limited scalability and also limited for some applications
Tier Three	✓ Scalability increases due to network insulation ✓ Lighter traffic to and from clients ✓ Heavier traffic to and from Servers
Tier N	✓ More scalability ✓ Robust N= (2+ (# of application servers))

Table 2.1: The key characteristics of Tiers.

Web service Model

In web service model different tasks were performed by multiple hosts which have specific roles. Moltchanov (2013) explains web service model as follows:

Presentation logic:

- Is responsible for displaying the information and interacting with the user (i.e., user interface).
- It must make the information available in a suitable form to the user and must respond appropriately to input from the user.
- Static versus dynamic content that service a web pages and can be Web 2.0 and SaaS

Application logic:

- Processes commands, makes logical decisions, performs calculations, and coordinates the application.
- It also moves and processes data between the presentation logic and database logic tiers. Implement business Logic, manipulate data and data mining

Database logic:

- Refers to the management of underlying databases. It is responsible for storing and retrieving the data according to the requirements of the application logic tier.
- Access data store, high transaction rate, high bandwidth and low latency.

2.8 Virtualization

Many researchers define virtualization in many different ways although they are rotating on the same spheres. The following are the most common definitions:

A framework or methodology of dividing the resources of a computer hardware into multiple execution environments, by applying one or more concepts or technologies such as hardware and software partitioning, time-sharing, partial or complete machine simulation, emulation, quality of service and many others (Padhy et al., 2011).

According to Hossain (2011), virtualization is the abstraction of system resources to allow multiple operating systems to run on one system at the same time. This can be accomplished by inserting either type one hypervisor or type two hypervisor.

As explained by IBM Global Education (2007), virtualization is a technique for hiding the physical characteristics of computing resources from the way in which other systems, applications or end users interact with those resources.

The process of decoupling the hardware from the operating system on a physical machine (Campbell, & Jeronimo, 2006).

2.9 Generation of Virtualization

Virtualization did not occur overnight rather it is the result of long period of time and progression. Padhy et al. (2011) briefly explain the generation in to three categories as follows.

From 1950 to 1970

It dates to the introduction of automatic page replacement in the atlas system. Page replacement is the method of storing and transmitting data up and down the memory hierarchy and the atlas system is the first to automate this system. The term virtual machine is first used in 1967 when IBM introduced the system/360 model 67 machine with virtual memory. This model uses the very old operating system by the name control program (cp-67) which intern changed to virtual machines operating systems. These virtual machines allowed many users to run multiple operating systems on a single processor which revealed that virtualization is not a very recent technology. Time sharing was also considered a new innovation at that time. Hardware virtualization technology that allows the virtual machine monitor to run virtual machines in an isolated protected environment was also a phenomenon in the year 1960s. Since the virtual machine monitor (VMM) is transparent to the software running in the virtual machine, the software thinks that it has exclusive control of the hardware. After some years in 1970 virtualization becomes a common technology and supported by many operating systems.

Virtualization was used for solving different problems at this time. For example the availability of virtual storage helps programs to get more storage than the machines already had. Although

virtualization was common in various areas it is radically disappeared in 1980s and 1990s. This is due to the fact that low cost mini computers and personal computers are largely come to the market.

From 1970 to 2000

Many technological changes were taken in the year 1970s and 1980s even the concept of hypervisor (formerly VM-CP) is widely used. Virtual machine monitor was used as type one hypervisor as they are installed on a bare metal for creating multiple instances. The roots of virtualization on the mainframe begin in MVS (Multiple Virtual Storage) was the most commonly used operating system on System/370 and System/390 machines. A component of MVS called JES3 allowed this sharing and separation to be managed from a single console. While MVS is no longer supported by IBM, the roots of virtualization can be traced to its support of these two key capabilities. In 1990s the concept of logical partitioning (LPAR) from IBM, java virtual machines (JVM) from sun micro system, and common language runtime are widely used. The beauty of those technologies is its portability in that it runs in java compiled programs regardless of the type of machines' platform. Zogaj (2012) on his thesis explained that even though virtualization was used as part of the IT landscape for many decades it was in 1998 that the benefit of virtualization becomes an industry standard in X86 platforms.

From 2000 to 2011

In this period virtualization became very common technology and changes everything. Here it becomes possible to run multiple applications in single virtual machines. Finally, in 2000, OS/390 turned into z/OS, which is the current production IBM mainframe operating system. It became highly expanded in unimaginable speed. At this period virtualization was used from small to complex data centers.

The most general and precise evolution of virtualization is also presented by Strassmann (2008). The first generation was intended to reduce cost through virtualization but the second and third generation is beyond that. The following figure shows the three generations of virtualization in detail.

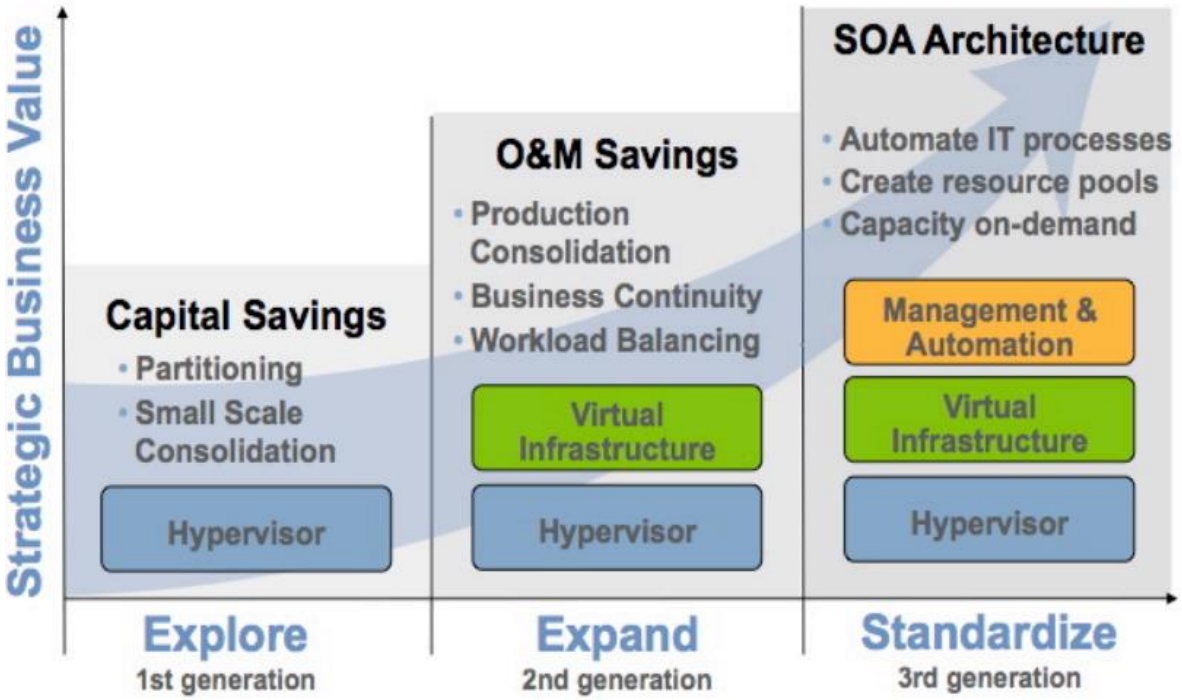


Figure 2.6: Evolution of virtualization (Source: Strassmann, 2009)

2.10 Issues to consider in virtualization

As stated by Nyamweya *(2013), b*efore implementing virtualization, critical planning is required to ensure that no huge problems are created from virtualization.

The first thing is, it is necessary to confirm compatibility issues of software and system with virtualization needs before continuing for large scale deployment. Some virtualization products claim to support unmodified applications. Cost of incompatibility might be more than benefits from virtualization so that proper attention should be given.

The second thing is every virtualization has its own performance overhead. This overhead should be measured and quantified during the evaluation stage to reach on decision (Nyamweya, 2013).

2.11 Components of Virtualization

Before discussing the different types of virtualization, it is very compulsory to identify the different key components of virtualization. As explained by Padhy et al. (2011), the following are the major components for virtualization to be accomplished.

Guest Operating System

The guest operating system is an operating system that runs in a virtual environment independently of dedicated hardware resources on a hypervisor.

Hypervisor

A hypervisor aka Virtual machine monitor (VMM) is software that allows different operating systems to run on a single physical machine. It can be either type 1(native /bare metal) or type 2 (hosted) hypervisor. Type 1 hypervisor as the name bare metal indicates runs directly on the surface of hardware and includes hyper-v from Microsoft, VMware ESX/ESXI, and Citrix from xen, oracle VM server and amazon EC2. Hosted hypervisor on the other hand runs on the host operating systems and includes Microsoft Virtual PC, VMWare workstation and Citrix Xen client. This research uses Microsoft Hyper-v for demonstration since it is open source easy and compatible with Microsoft products. For more clarification it can be shown as a table.

Hypervisor Type	**Key Characteristics**
Type 1 **(Bare Metal)**	▪ Runs directly on physical hardware's ▪ Does not have OS running below it ▪ Fully responsible for scheduling and allocating of the systems resources between virtual machines Example: VMware ESX (Enterprise), Xen, Hyper-V. ✓ More secure than Type II
Type 2 **(Hosted)**	▪ VMM runs as an application in a normal OS. ▪ The OS controls the real hardware resources called as Host OS. ▪ Host OS has no knowledge of the type II VMM, which is treated like any other process in the system. ▪ The OS run inside of the Type II VMM is referred to as the

	Guest OS. Example : VMware GSX (workstation), UML (User- Mode Linux) ✓ Less secure then Type 1 because any security Vulnerabilities that lead to the compromise of the host ✓ OS will also give full control of the guest OS. ✓ Host OS are heavyweight than Type II

Table 2.2: Types of hypervisor

Host machine

The host machine is a physical machine intended to be virtualized and consists of the physical memory, hard disk, CPU etc.

Virtual Machines

Virtual machines sometimes called virtual instances is virtualization representatives runs on the hypervisor and implemented as a single file or folders on the host machine.

2.12 Types of virtualization

There are five major types of virtualization each having its unique characteristics and significance. According to Padhy et al. (2011), virtualizations can be server, desktop/client, network virtualization, application virtualization and storage virtualizations.

Server virtualization is the process of dividing a single server in to multiple virtual machines using VMM. After virtualization each virtual servers have its own memory, CPU, hard disk and other peripheral interfaces. Each servers can act as a function like mail server, internet server, file server, enterprise resource planning server and the like that removes a one application per one server architecture barrier. This consolidation of multiple virtual servers as a single server helps for 70 to 80 % sometimes more utilization of the available hardware's. In other words it is the partitioning of a physical server into multiple smaller, virtual servers. Amazon Elastic Cloud 2 is the best example.

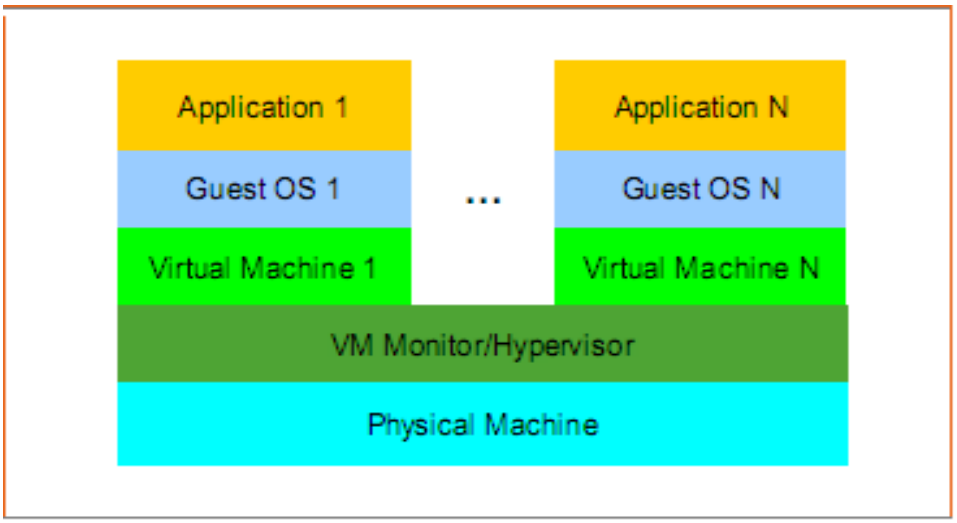


Figure 2.7: Basic architecture of server virtualization (Source: Metzler, 2011)

According to thesis conducted by Ronkainen (2003), server virtualization is the way of creating multiple virtual machines or virtual environments using physical hardware and virtualization software's. The primary focus of this virtualization software's is sharing the physical resources like hard disk, CPU, memory and others to the virtual instances or machines. This sharing of resources is done by creating a specialized virtualization layers which converts a physical resources in to virtual machines. The possibilities of running multiple operating systems concurrently on the same machine strengthen the above fact. The following figure clearly revealed the difference between physical servers and virtualized machines.

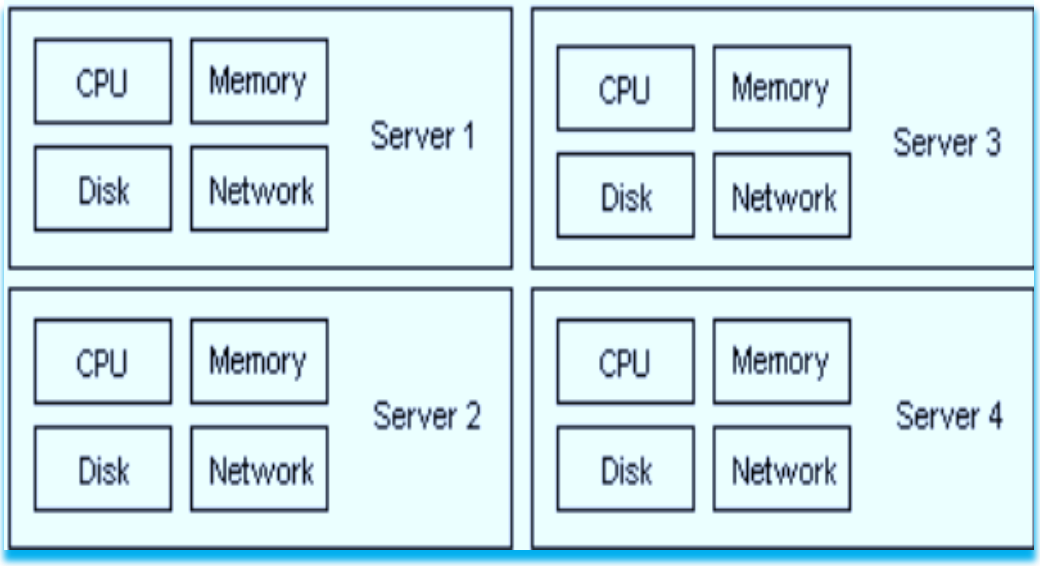


Figure 2.8: Physical machines (Source: Ronkainen, 2003).

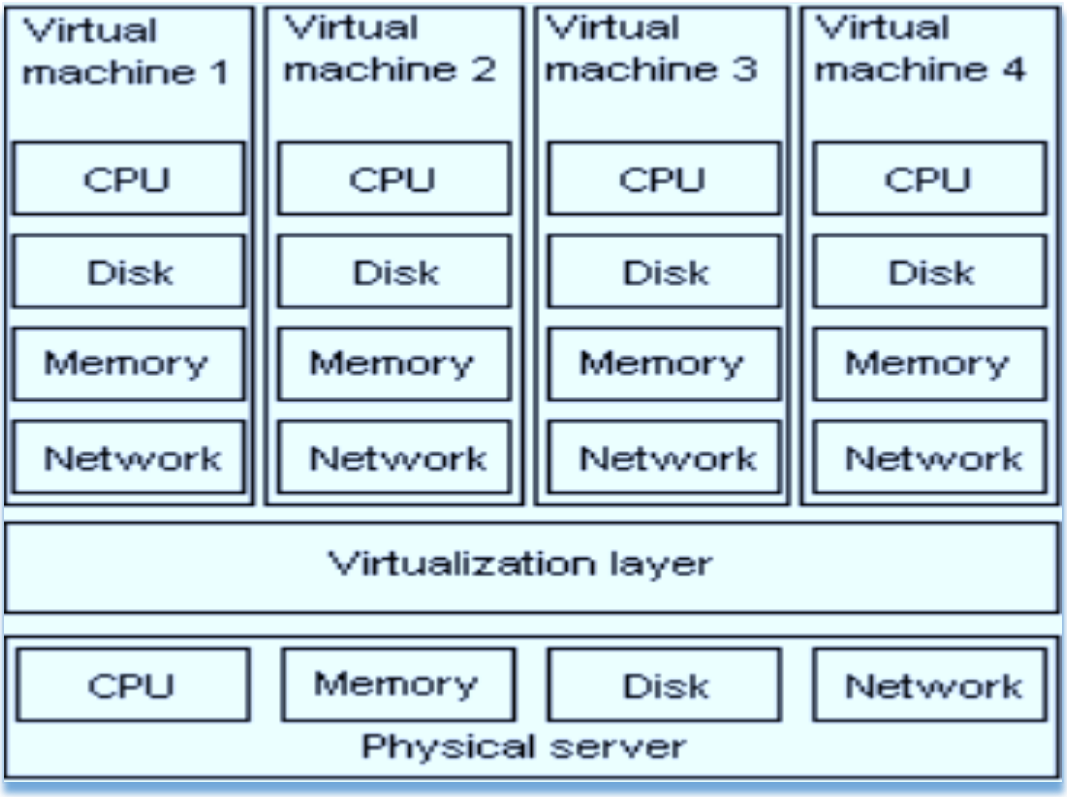


Figure 2.9: Virtual machines in a single virtual server (Source: Ronkainen, 2003).

In figure 2.8 above each of the four physical servers has its own processor, memory, hard disk and network but figure 2.9 shows four virtual machines running on a single physical machine with virtualization software's.

Advantages of Server Virtualization:

Server virtualization has advantages including Reduce operational costs (Hardware, Energy, and Space), improved uptime and availability, enables robust Disaster Recovery, reduced maintenance disruption, enables server consolidation, enables legacy software's on every machine, streamline resource provisioning and scale.

Disadvantages of Server Virtualization:

Different technology has its own drawbacks hence; server virtualization has the following disadvantages.

- Only a few processors that support virtualization can be used to virtualize servers.
- The resource allocation for each virtual system needs to be planned carefully. If very less resource is allocated, the application performance might be affected and if too much resource is allocated, it will result in under-utilization.

Storage virtualization

Storage virtualization is the pooling of multiple physical storage resources into what appears to be a single storage resource that is centrally managed. Storage virtualization is commonly used in file systems, storage area network (SANS), switches and virtual tape systems. The act of abstracting, hiding, or isolating the internal function of a storage system, subsystem or service from applications, compute servers or general network resources for the purpose of enabling application and network independent management of storage or data.

Storage Virtualization goes hand in hand with server virtualization, as when both are used together they provide tremendous flexibility. It makes managing storage from multiple sources to be managed and utilized as a single repository. This simplifies utilization of Storage Area Networks (SAN's). Because this storage isn't married to any particular server, many servers can access the data stored on the SAN. Because servers (VM's) are now in the form of a file, they

can be stored on the SAN. If a host server goes down, another host can access that VM from the SAN and turn it on. This improves the high availability of these VM's by facilitating mobility between which host servers it will run on.

Desktop/Client Virtualization

Creates a separate OS environment over and above the existing running OS on the desktop. Desktop virtualization is the creation of a virtual computer environment that is then delivered to a user in place of a physical computer. The virtual computer is stored on a remote server and delivered to the user's device. They operate it in approximately the same way as they would a physical machine. A single server can deliver multiple personalized virtual desktop images.

Vitalizing a desktop consists of removing the operating system from a traditional workstation and relocating a virtual copy of it on a host server. Users can then access the virtualized workstation, all of the programs, applications and data through a remote desktop client application from workstations, laptops, Smartphone's or thin client terminals.

Network Virtualization

Network virtualization proposes decoupling of functionalities in a networking environment by separating the role of the traditional Internet Service Providers (ISPs) into two: infrastructure providers (InPs), who manage the physical infrastructure, and service providers (SPs), who create virtual networks by aggregating resources from multiple infrastructure providers and offer end-to-end network services.

In a Network virtualization environment one or more virtual machines can access the local or external network using the physical network adaptor attached to physical machine. It can also be connected without physical network adaptor and it uses logical network adaptor.

Network virtualization was developed by using the same concepts of server virtualization. Software Defined Networking uses virtual switches, routers, firewalls and load balancers. This allows IT staff to provision networks without disruption to the physical network while running traffic over the physical network. This allows VM's to retain their security properties when moved from one host server to another that may be located on a different network. Server

managers have the ability to configure virtual switches, routers, firewalls, load balancers, etc. without having to bother the network administrator.

Application Virtualization

This is a method of providing a specific application to an end user that is virtualized from the desktop OS and which is not installed in a traditional manner. An application can be installed and/or executed locally within a container that controls how it interacts with other system and application components. An application can be isolated in its own virtualized "sandbox" to prevent interaction with other system and application components it can be streamed across a network. Applications can be delivered across the network to a web browser with most processing executed on a centralized web server. Application virtualization separates the application layer from the OS in a desktop environment which reduce application conflicts.

Application virtualization allows users to run applications from devices that don't possess the operating system the application requires. Another possible reason to use application virtualization is to run conflicting programs that can't coexist on the same device. An example of this would be a user who needs to run two different versions of the same software. Two of the main application virtualization technologies are hosted applications and packaged applications.

Hosted solutions use servers to host applications and allow users to connect to the server from their device. The user sends keystrokes and mouse clicks to the server and the server sends screen changes to the user, while the server is actually running the applications. Many users can run applications on the server simultaneously so these servers must have a lot of resources. The user's device doesn't require a lot of resources since it is not doing the work (one of the main benefits of application virtualization).

Packaged solutions package software within their own environment so they can run on devices with a different operating system than the application was built to run on like running an XP application on Windows 8. Packaged applications are often stored on network shares and then streamed to the user's device.

2.13 Methods of Virtualization

Two methods of virtualization are widely accepted. Calzorali (2006) stated these methods as full virtualization and Para virtualization.

Full Virtualization

This is first used in 1966 by IBM mainframes. These permit the virtual machines to have enough hardware to allow unmodified guest operating system to run. In full virtualization the guest operating system even not aware as it is virtualized so that it abstracts the guest operating system from the underlying hardware resources. The greatest advantage of full virtualization is better security, portability and migration of virtual machines. Full Virtualization is a virtualization in which the guest operating system is unaware that it is in a virtualized environment, and therefore hardware is virtualized by the host operating system so that the guest can issue commands to what it thinks is actual hardware, but really are just simulated hardware devices created by the host.

Para Virtualization

In these types of virtualization the role of providing resources to the virtual machines is held by the hypervisor. In Para virtualization the guest and host operating system should be modified. Portability and compatibility are critical because Para virtualization do not support unmodified operating system. Para-virtualization is a virtualization technique in which the guest operating system is aware that it is a guest and accordingly has drivers that, instead of issuing hardware commands, simply issues commands directly to the host operating system.

2.14 Virtualization for cloud computing

The idea of virtualization and cloud computing mostly creates confusion but there is a huge bridge among the two technologies. Cloud computing is the technology of delivering different resources as pay per use fashion but virtualization is an enabling technology. The US institutes of standard and technology (NIST) defines the different service delivery (deployment) models and service delivery layers as follows (Mell, 2011).

Cloud deployment Models

Service delivery or deployment models of cloud computing includes private, public, hybrid and community clouds.

Private cloud: infrastructure is provisioned for use from a single organization that comprises multiple tenants. It can be operated on or off premises and are behind the companies firewall.

Public: a cloud service provider offers services to multiple businesses, academic institutions, government agencies, and other organizations with access via the internet.

Hybrid: this combines the above two cloud delivery models that remains unique as entities but are bound together by technology that enables data and application portability. Cloud bursting is an example of one way enterprises use hybrid clouds to balance loads during peak demand periods.

Community: this is a cloud infrastructure provisioned for the exclusive use of a specific community of user organizations with shared computing requirements such as security, policy and compliance.

Cloud service delivery Layers

The following are the main and significant the service delivery layers for the above service delivery models.

Infrastructure as a service (IaaS): infrastructure as a service is a collection of different hardware and software available on the cloud which enables users to self-provision these resources in order to run platform and applications.

Platform as a service (PaaS):

Platform as a service enables users to adapt legacy applications to a cloud environment or develop cloud aware applications using programming languages, services, libraries and other developer tools.

Software as a service (SaaS):

It is one of cloud computing layers in which users can run applications via multiple devices on a cloud infrastructure.

Here the researcher intention is to clearly show the distinction between the two technologies. Virtualization abstracts compute resources as virtual machines with associated storage and network connectivity. On the other hand cloud determines how those virtualized resources are allocated, delivered and presented. Virtualization is not necessarily important for cloud computing environments but enables rapid scaling of resources in a way that non virtualized environments find hard to achieve. 1 &1, Azure, City cloud, Amazon, Century Link, Cloud Sigma, Cloud Watt, data centered, phoenix NAP, profit Bricks, vCloud Air and Verizon are among the cloud service providers located in different parts of the world.

2.15 Advantages of virtualization

Virtualization has giant benefits for the business and end user managements. As Malek (2013) briefly described in his thesis, the following are some but not all benefits of virtualizing ICT infrastructures as a workstation or data centers in general.

- **Consolidation of Workload to fewer machines**

Generally speaking having fewer server means less cost expenditure for maintenance, energy consumption, and service level agreement (LSA). Zainelabden, & IBRAHIM (2015) explains Service level agreement as the document contract between cloud customers and cloud services providers. Service level agreement is mostly used in cloud computing between the cloud provider and users. SLA is the part of a contract which defines exactly what services a service provider will provide and the required level or standard for those services. Consolidation is the process of merging different virtual machines on single physical machines. This helps for the appropriate use of the physical machines central processing unit, RAM, hard disk at recommended threshold and increases return on investment (ROI).

➢ Running Legacy Applications in a new hardware

Small to large organizations evolved for some years had different old applications that can be used before. Since, new hardware and operating systems were developed dynamically the legacy applications become difficult or totally impossible to run on the new machines/servers. But both the client and server virtualization permits to run those legacy applications on the incompatible hardware's.

➢ Isolated Operating Environment

Sometimes it might be important to run multiple operating systems on the same machines especially for laboratory or experiment systems. By using virtualization it is possible to run multiple operating system in which each operating systems has its own setting, code registry, and code libraries so one applications or data cannot corrupt others.

➢ Simplifying Backup and Recovery

Virtualization rationalizes and facilitates recovery processes in a simple way and effectively. Since recovery is faster and effective the organization data is recovered with less impact on the organizations business.

➢ Enable Storage Provisioning

Virtualization reduces cost through reducing underutilization of storage resources and also prevent running out of storage resources. This is because of the flexibility of virtualization to add or remove storage resources as needed.

➢ Breaking Hardware Dependency

The base of virtualization is laid on the creation of hypervisor which is used as an Abstraction layer between the operating system and the underlying hardware.

Because of this fact critical virtual machines can be easily shifted to other hardware's. Shifting of virtual machines can be either for emergency cases or for planned maintenance of production servers.

- **Quick Recovery of End user Devices**

It is factual that no data is stored locally while using virtual desktop infrastructure (VDI). These revealed that failed hardware can be easily swapped to new VDI client.

Tvaruzek (2009) also explained the key benefits of virtualization as better server utilization, compatibility, isolation, novel application delivery model and less energy consumption.

2.16 Disadvantage of virtualization

As indicated through life, different new technologies have its own drawbacks even if it is not significant when compared with its benefits. Accordingly, virtualization has the following drawbacks (Malek, 2013; Tvaruzek, 2009).

- **License Management**

Licensing in virtual environment is more complicated compared with a standard model. This is because running virtual machines need additional licenses and also software vendors have different licensing policy for virtualizing platforms or environments.

- **Support from Application Vendors**

Sometimes software products can be protected by individual hardware keys or even cannot be virtualized. This indicates that software should be supported by different vendors to be run in a virtualized environment.

2.17 Virtualization Software Vendors

There are numerous vendors of virtualization software in the market. But now a day's three vendors are the most widely known, prominent and competent in providing this software.

VMware is the leading technology and innovation leader focusing on virtualization from the beginning. It is still dominant because it has great installation in enterprise environment, broad product range and technology advantage. Even if it is dominant in the market it is on the way of highly weakness because Microsoft produced its own virtualization software known as hyper-V.

The second dominant virtualization software vendor is Microsoft which is widely focus in virtualization business after the introduction of Microsoft server 2008 and hyper-V. It becomes very competent with VMware. The big secret for VMware to be the leader in the market is because it was 7 years before Microsoft started virtualization business. **Changing vendor** is very **expensive** and also risky so that many organizations still use VMware. Microsoft on the other hand has high number of server deployments, broad range of products, known brand and financial strength but it is late start compared to VMware.

Citrix is the third virtualization software vendor in the market share and widely known by desktop visualization called Citrix xenDesktop. Due to rich product capabilities and the growing of VDI market it follows the two companies. Citrix was tightly cooperated with Microsoft since 1989 by providing virtualization and remote connection capabilities. The greatest weakness started after Microsoft developed virtualization software (Hyper -V) by its own.

2.18 The concept of framework and its representation

As concisely explained by Hevner (2004), the major output of an IS research is an artifact which is intended to address an organizational problem. These IT artifacts can broadly classified as constructs, methods, models and instantiations.

According to Hevner (2004), constructs represent vocabulary and symbols; models represent abstractions and representations; methods incorporates algorithms and practices; instantiations stands for implemented and prototype systems. Framework is a model artifact that provides a broad overview or skeleton of interlinked items which helps as a guide to achieve a specific objective.

2.19 Related work

Various works were reviewed from related literatures written by different scholars at different periods around the globe. Some works are presented as follows:

According to research conducted by Getnet (2017), framework is developed for the banking sector specifically for Wogagen bank using design science methodology. By reviewing different literatures the researcher design a framework which shows different components that fit

with the banking industry. The main intention of the researcher is resource sharing and isolation of data between users.

The main components of the designed artifact include:

User: is a customer who is served or gain services from the bank in various ways including VISA and Master Card users.

Channel: these are the mediums by which users get services from certain banks. This channel includes branch, ATM, POS and the like.

Web interface: it is an integrated interface to the core banking system to provide banking services using web address.

Virtual client: virtual client is a client with its own operating system which is managed locally through the existed domain controller to access the core banking activities.

What to virtualize/to whom /by what: these are the main variables used by the researcher to observe the continuous operation with virtualization.

Here, the scope of his research is client virtualization which is completely different with server virtualization.

There is also another investigation on green IT framework using virtualization done at Pakistan. They designed a framework as a serious of steps from developing plan for green data center to measuring the performance in terms of energy efficiency, resource utilization and CO2 emissions. The ultimate goal of the research is saving huge amount of energy, minimization of CO2 emissions and as a result reducing global warming and other consequences from it using virtualization technology. The researcher uses five steps for proposing the framework using virtualization (Uddin et al., 2012). The following steps also clearly show the main components included by the researcher:

i. Develop plan for green data center

Under this the researcher shows that green data centers is the result of different challenges from theoretical to realistic manners. In order to develop a plan for green data centers the net

advantage should be identified. The following is the advantage of IT framework using virtualization: lower server storage temperature, better system reliability, better uptime, lower total cost of ownership, maximizes hardware and software utilization, lower heating and cooling cost etc.

ii. Categorize data center in to measurable components

Data center is the name given for the collection of server, storage device, UPS, switch gear, computer room air conditioners, pumps, lightening, generator and many others. So measuring the performance of data center is the sum of measuring the real performance of individual components.

iii. Identify green metrics and set components

A good metric would be something that measures the efficiency, the sustainability and the cost of a green initiative. The biggest problem any metrics faces when applied for calculating energy efficiency is the lack of standardized system of categorizing different resources of data center.

iv. Identify and implement Virtualization type

The type of virtualization is as per the requirement of certain organizations. These can be desktop, server, network, storage, application and operating systems. The virtual machine monitor will be from many different vendors like Vmware, Cisco, Citrix, Hawllet Packard, IBM, Microsoft, NetApp, Novel, Oracle, Quest, Sun, Symantec and virtual Iron.

v. Measure the performance in terms of energy efficiency and CO2 emissions

This stage is used to measure the performance of the data center from time to time using the selected metrics. The phase has the following two sub steps:

a. Collect and categorize data (Facilities data, Infrastructure data, IT equipment data)
b. Analysis using energy and CO2 efficiency calculators.

Using the above five stages the researcher develop the following framework:

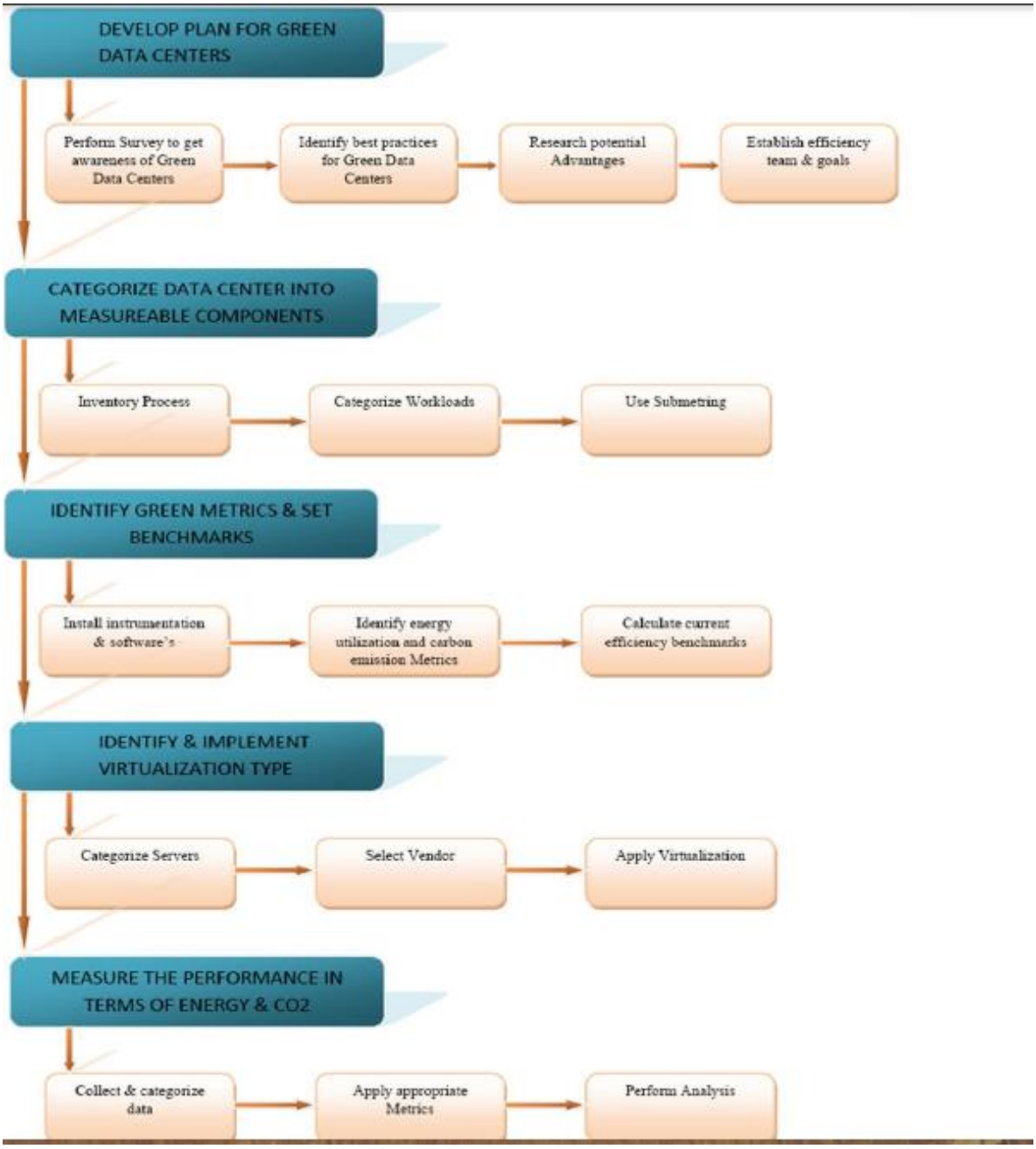


Figure 2.10: Green IT framework using virtualization (Uddin et al., 2012)

But, the developed framework lacks showing the specific components of a virtualized framework.

I.T. Alliance (2010), established a framework for server virtualization and transformation projects. Accordingly server transformation through virtualization provides improved total cost of ownership, cost saving through reduction server and resources. The I.T. alliance framework of virtualization helps organizations to structure their virtualization program in accordance with best industry practices. Each server virtualization projects has three phases including assessment, planning and execution each of them has its own sub activities. Finally the following I.T. alliance framework for server virtualization is illustrated. This study shows the specific stages followed while virtualizing a server but does not purely develop a server virtualization framework.

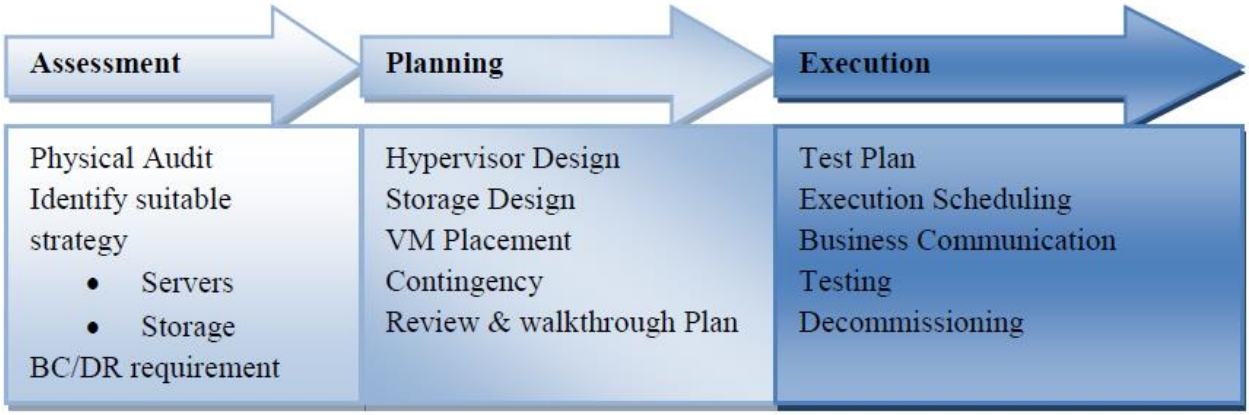


Figure 2.11: I.T. Alliance frameworks for server virtualization (I.T. Alliance, 2010)

Calzolari (2006) has been done his research on "High availability using Virtualization". The main objective of his work is achieving redundancy using virtualization technology. The methodology he followed is not clear and recommends additional work to improve the system performance.

Lastly, Nyamweya (2013) develops a framework on virtualization in selected government ministries of Kenya. Exploratory method was used in which the Ministries of Energy and Finance, and also the Directorate of e-Government were used as cases. While developing the integrated framework, implementation model is not mentioned and the framework is not self-descriptive.

The researcher summarized the related works that was discussed above using the following table for better understanding.

Author, Title & Year	Objective/purpose	Approaches/Methodologies	Key Findings	Recommendations And future work	Remark
Getnet Worku Data center virtualization framework in banking sectors. (2017)	To propose a virtualized data center framework,	Design science research methodology with qualitative data analysis	Develop a framework for the banking sectors	Further work is recommended to be done in order to recognize the proposed framework	The research is focused on desktop/client virtualization
Nyamweya, B Server virtualization framework: case of selected government ministries, Kenya (2013)	This research aimed at developing a framework to be used for implementation of server virtualization.	exploratory and applied a case study method where the Ministries of Energy and Finance, also the Directorate of e-Government were used as cases	She develops a framework as a series of steps from planning to support	Further research recommended on impact of the Integrated server virtualization framework and assessment of customer satisfaction	The developed integrated framework implementation model is not mentioned and the framework is not self-descriptive.
Calzolari, F. High availability using Virtualization	Achieving a redundancy system for all services running on the data center using	The methodology is not that much clear. The general research approach is case study in which 3RC	High availability service 3RC is developed which requires less human effort.	Recommends additional work to improve the system performance and add	The research is focused on high availability of virtual servers and on disaster recovery but the

(2006)	virtualization	project is selected as a case.		extra features. He also recommends other virtual environment solutions like Xen…	methodology is somehow un understandable and it should be better to be clarified with model or framework.
Uddin, et al. Green IT framework using virtualization (2012)	Greening IT using virtualization. It focuses on decreasing CO2 emissions and other related environmental hazards using virtualizations.	To achieve some of the objectives highlighted, Interviews were conducted from top managers including directors from IT data centers operations, environmental initiatives and enterprise infrastructures. The survey results from the interviews are then normalized and it seemed surprising that performance and cost were the most important attributes.	He designed a framework The ultimate goal of the research is saving huge amount of energy, minimization of CO2 emissions and as a result reducing global warming and other consequences from it using virtualization technology.	The research did not identify any future work	He developed framework as a serious of steps from developing plan for green data center to measuring the performance in terms of energy efficiency, resource utilization and CO2 emissions but doesn't show components required.

Table 2.3 Summary of related work

2.20 Chapter summary

In this chapter, various literatures related with the study are reviewed. This includes topics on green data center, evolution of data center, virtualization types, advantages and disadvantages, implementation models and security issues. Above all, various works related with the objective of the study were reviewed and summarized.

And in the next chapter methodology will follow.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Overview

In this chapter issues related to the research approach, research design, research method, samples of the study with the selection mechanisms, instruments and procedures of data collection etc. are anticipated.

3.2 Design Science Research Methodology

The objective of this study is exploring the traditional or physical data center infrastructure and developing a virtualized data center framework for Amhara Regional Health Bureau. In order to achieve the specific objectives design science research methodology (DSRM) was used which is a system of principles, practices and procedures required to carry out a study. Information system can draw advantage from DSRM by often using theories from diverse disciplines, such as social sciences, engineering, computer science, economics and philosophy to address problems at the intersection of IT and organizations (Peffers et al., 2007). Design science approach was used because; the researcher is intended to develop a framework that guided the regional health bureau to implement server virtualization.

Design science research is highly used in disciplines that are concerned with creation of artifacts. Many Information system researchers pioneered design science but little Design science researches were conducted in more than a decades. Because of this fact many Information System researches are mainly found on the engineering journals than IS journals. Lack of a methodology to serve as a commonly accepted framework for DS research and of template for its presentation may have contributed to its slow adoption. The design science research methodology (DSRM) incorporates principles, practices, and procedures required to carry out such research and should meet three objectives: it is consistent with prior literature, it provides a

nominal process model for doing DS research, and it provides a mental model for presenting and evaluating DS research in Information Systems. The more widely accepted and also used design science process includes six steps: problem identification and motivation, definition of the objectives for a solution, design and development, demonstration, evaluation, and communication (Peffers et al., 2007).

Qualitative research methods was used which is an empirical research method which is widely used in social, behavioral, organizational and evaluative research. Most data in qualitative research are not normally in the form of numbers and to be collected from various sources, which involves many techniques such as data description, decoding, translation, etc. to understand their meaning in a natural setting (Zhang & Fever, 2013).

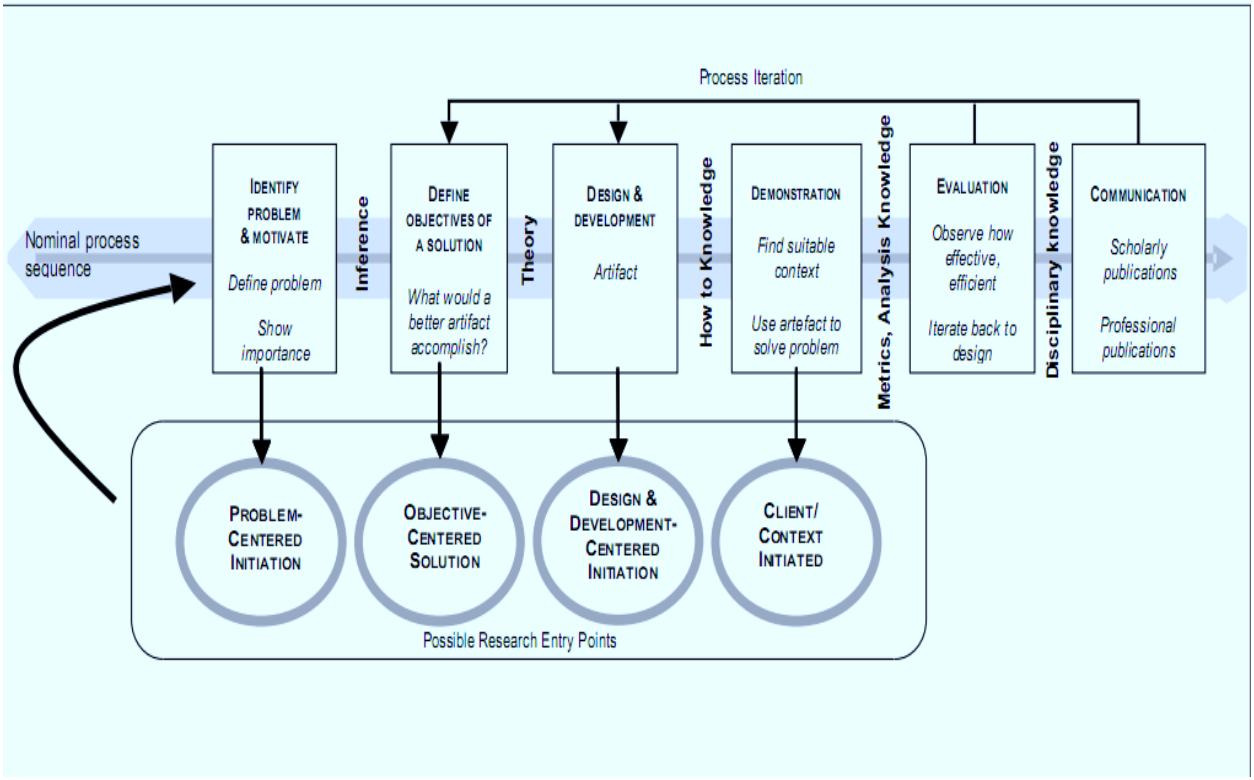


Figure 3.1: Design science research methodology process models (Peffers, et al., 2007)

3.2.1 Problem Identification and Motivation

3.2.1.1 Identify Problems

This is the first step in the process model of design science research methodology. In this stage the specific problems in the traditional or physical data center is critically identified and the researcher strongly work to solve those problems. Both government and non-governmental organizations fund for the health bureau. Nevertheless, the regional health bureau does not fulfill the major requirements. The existed data center in the regional health bureau is still more or less traditional. A one server for one application architecture is exercised in the regional health bureau which is not cost effective, difficult for migration, difficult for backup and recovery, take more space, and also difficult for management. After the problems are clearly identified the researcher showed the specific importance of his/her work to solve the existed problems. The researcher's objective is developing an artifact (in this case framework) that can effectively and efficiently provide a solution for the current problem in the data center. In order to come up with the solution, the researcher followed qualitative data collection instruments. Generally one server for one application in the regional health bureau data center is a great barrier in various situations.

3.2.1.2 Expert Interview

In order to develop the study, the researcher started with literature review to gather relevant requirements and aspects of existing data center. The researcher used both primary and secondary data collection techniques to gather appropriate data from the respected bodies.

Observation and interview (preferably semi structured interview) are among the primary techniques but document analysis is the secondary data collection techniques which is used to gather appropriate data. The following description shows how to use those instruments in this thesis.

Primary Data Collection Techniques/Instruments

Primary data collection techniques provide the researcher with appropriate first-hand information that contributes more for solving the existed problems.

Observation

Observation is a technique that involves systematically selecting, watching, and recording behavior and characteristics of phenomena at the specific research area (Marczyk et al., 2005). It is one data collection instrument by looking the actual work conducted in the regional health bureau data center. The researcher carefully watches actual health bureau data center to gather information on how to use the different information technology infrastructures.

Interview

According to Marczyk et al. (2005), interview is a form of self-report which is relatively simple method but producing wealth information. The different types of interview, includes structured interview, unstructured interview and semi structured interview. A researcher selected semi - structured interview because it provides information which assures the comparability of data. It also requires a few interviewing skills than does unstructured interview. Concentrated interview with information technology managers, network Administrators, database administrators, system administrators, and ICT officers was made for better understanding of the current physical data center infrastructure architecture and trends using well prepared interview guide which is adopted from literatures and modified accordingly.

Secondary Data Collection Techniques/Instruments

Secondary data source is also used by reviewing different documents relevant to the study, including literatures on related works and other documents in the bureau.

So, document analysis, semi-structured interview and observation are used to gather rich and in-depth understanding of practices in the health bureau data centers. Qualitative data collection techniques are particularly effective at gaining insight into the processes and events that led up to the observed variation (Castellan, 2010).

But, Gathering data from the total population is time taking and expensive. So, sampling method is preferable. Sampling techniques provide a range of methods that enable to restrict the amount of data needed to collect by considering only data from a subgroup rather than all the population and finally generalizing the whole population.

Types of sampling

As explained by Acharya et al. (2013), sampling can be broadly classified as probabilistic and non-probabilistic. Accordingly, in probability sampling each individual in the population has equal chance of being selected in the study. But, in non -probabilistic sampling method the probability that a subject is selected is unknown.

Alvi (2016) also defines probability sampling which is based on the concept of random selection, and non-probability sampling which is non-random sampling. In probability sampling design, every item of the population has an equal chance of inclusion in the sample. But, in non-probability sampling things are not left to the chance.

Among the non-probability sampling techniques purposive sampling was selected for this study. This is because purposive sampling is appropriate to gather data from the experts who are familiar with the topics in detail or it is where the researcher selects information rich cases for study in depth. Purposive is also better because it is a method where by a researcher selects sample based on experience or knowledge of the group to be sampled and also based on the convenience of the investigator. For conducting purposive sampling, a researcher has something in mind and participants that suit the purpose of the study are included. Purposive or judgmental sampling enables to use judgment to select case that best enable to answer questions and to meet the objectives. Hence, purposive sampling was selected to gather appropriate data from 9 Information technology staffs of the regional health bureau. These include information technology managers, network Administrators, system administrators, database administrators, and ICT officers in the specified health bureau. The following table shows the information technology staffs involved for gathering data.

Respondents	Total population	Target population	Total selected
ICT Manager	**1**	**1**	**100%**
Network Administrator and System Administrator	**3**	**3**	**100%**
ICT Officers	**3**	**3**	**100%**
Database Administrator	**2**	**2**	**100%**

Table 3.1 Sampling

Interview questions are prepared by adapting others work from various literatures literatures and modifying for the specified case.

3.2.2 Define the Objective of the Solution

This is the second step in design science research process models. In this stage the study aims to identify the requirements for developing the proposed data center virtualization framework from knowledge of the state of problems. It desires to provide an answer for the questions "what a better artifact accomplishes?" After identifying a problem and pre-evaluating its relevance, a solution has to be developed in the form of a framework. Qualitative objectives are used, which described how a proposed framework is expected to support solutions to problems not previously addressed. After data was collected from respected bodies, it should be narrated and summarized. It involves data preparation, analysis, and finally data interpretation. Qualitative data analysis method is used to analyze the data collected from interview, observation and document analysis.

The analysis technique used in this study is thematic coding analysis technique in which the result of interview was grouped in to main categories. Finally appropriate generalization is made and presented accordingly for the qualitative data by way of narrating and interpreting the situations.

3.2.3 Design and Development

Design science research should produce a feasible artifact in the form of models, methods, constructs and instantiations (Hevner, 2004). A framework is a model artifact that provides a broad overview or skeleton of interlinked items which helps as a guide to achieve the specific objective. After all the above stages are completed the appropriate and suitable framework was developed using software tools Edraw-Max and Visio. The framework is developed based on best experiences from various literatures and possible modifications. Edraw Max and visio were selected because both software's have rich features and easy to understand. The developed framework shows each component with its functionalities to solve the existed problems in the health regional health bureau. Users, web server, virtual servers, selected hypervisors, security aspects and storage servers are the main components which are discussed in detail on chapter five. The developed server virtualization framework is unique to ARHB but can be used for other organizations by applying minor modifications.

3.2.4 Demonstration

In the demonstration section the researcher used the developed artifact to solve the existed problems in a better manner through a prototyping tool including experimentation and/ or simulation. As stated above in the literature document, virtualization software's or hypervisors can convert a single computer into many virtual instances or machines. Hypervisors are two types: - Native or Bare Metal Hypervisor and Hosted Hypervisor. Native or Bare Metal hypervisor are software systems that run directly on the host's hardware to control the hardware and to monitor the Guest Operating Systems. The guest operating system runs on a separate level above the hypervisor. Examples of this virtual machine architecture are Oracle VM, Microsoft Hyper-V, VMware ESX and Xen. The second type is Hosted hypervisors which is designed to run within a traditional operating system. In other words, a hosted hypervisor adds a distinct software layer on top of the host operating system. Because of this fact, the guest operating system just becomes a third software level above the hardware. A well-known example of a hosted hypervisor is Oracle VM Virtual Box. Others include VMware Server and Workstation, Microsoft Virtual PC, KVM, QEMU and Parallels. Among the Native hypervisor Microsoft

Hyper-V from Microsoft is selected for this study demonstration. Hyper –V is selected because according to Microsoft (2013), it has greater scalability and performance, secure, flexible infrastructure, and also high availability and resiliency.

Here in a demonstration part of this thesis the utilization of central processing unit (CPU), random access memory (RAM) and other scenarios of both the physical servers and the virtualized machines (instances) were measured. After demonstration was completed the measuring values of the two approaches (physical & virtual) machines are compared. For the purpose of taking the results, both graphical user interfaces (GUI) and command line interface tools can be used. Non-GUI testing tools can be windows power shell, windows management instrumentation etc. On the contrary server manager, task manager, performance monitor and resource monitors are among the GUI tools to take and compare the utilization of physical and virtual machines on the required components explained above. The researcher selected the graphical user interface tools called resource monitor to capture the utilization of physical machines and the virtual instances. Experiment was done to check the resource utilization of the virtual instances and determine the number of instances to be deployed on a single physical server. Demonstration and analysis result was presented in a separate chapter for its simplicity.

3.2.5 Evaluation

Here the study aims to observe and measure how well the framework supports a solution to the problem. Evaluation proves how the proposed framework is effective and efficient to solve the existed problem in the physical data center. This activity can be done by comparing the objectives of a solution to actual observed results from use of the framework. To Develop and propose the virtualization framework, an open source Edraw-Max version 8.4 and Microsoft Visio was used and the developed artifact is evaluated by the respected professionals. Artifact can be evaluated using its goal, environment, structure, activity and evolution which are briefly described in the following figure (Prat et al., n.d).

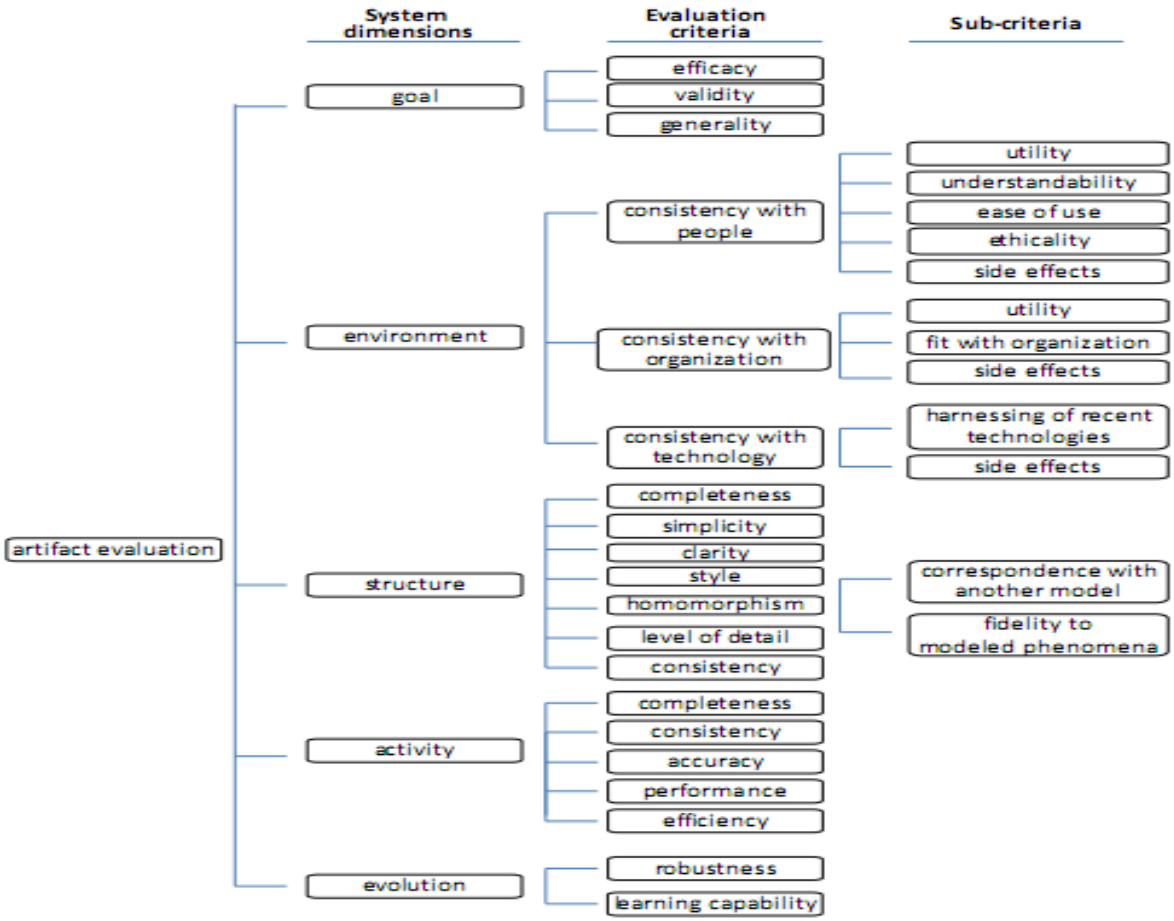


Figure 3.2: Artifact evaluation criteria (Prat et al.).

Hence, the developed framework was evaluated comparing with some of this artifact evaluation criteria. Evaluation can include evaluating its structure, environment etc. and also evaluation it through its soundness to the solution.

3.3 Chapter Summary

This chapter ties to cover the following main points.

Design science research methodology was selected which is a methodology intended to create a framework. Judgmental/ purposive sampling was used to collect data from the ICT staffs through semi- structured interview.

The next chapter will cover analysis of the collected data and experiments on utilization of resources in the virtual and physical machines.

CHAPTER FOUR

INTERVIEW DISCUSSION AND INTERPRETATION

4.1 Overview

The main purpose of this chapter was to present the data gained from semi structured interview, observation and document analysis. For this, data is properly collected from Information Technology and other respective staffs and finally verified. The data collection process was for solving the different problems existed while using physical data centers and solving through virtualized data center. Hence, this chapter contributes a lot for the design and development phase of design science research methodology.

4.2 Demographic Data

Demographic variables are crucial for collecting individual respondent's profile. For this research, work experience and education status are considered.

Educational Level and Work Experience	Frequency	Percent
Education Level		
BSC	8	89%
MSC	1	11%
Total	9	100%
Work Experience		
<5 Years	5	56%
>=5 Years	4	44%
Total	9	100%

Table 4.1: Respondents Educational Level and Work Experience.

4.3 Interview Interpretation

Interview questions was prepared by adopting from literatures and modified according to the researcher's case. The result of the interview helps for identifying each and every problem in the regional health bureau. It is based on the interview results, observation and literatures that the framework was developed.

After the interview was conducted with the information technology staffs the next step is preprocessing in an appropriate manner and finally analysis is performed. An interview finding is broadly categorized in to 5 sections. This includes: current physical data center, virtualization plan, pros of virtualization, challenges and securities of physical data center and virtualization, data center infrastructure management (DCIM).

I. Current Physical Data Center

The data center has many components like servers, switches, cooling components, powering equipment's, lightening and many more. These components are highly integrated for the functioning of the data center and the health bureau in general. The regional health bureau data center is working as a one server for one application architecture which is not recommended. Three of the respondents replied that "the current data center is basic site infrastructure which does not have a disaster recovery data center". Generally almost all respondents agree that the current data center is traditional and has various problems. Physical data center is: 1) not cost effective, 2) not scalable, 3) difficult to manage, 4) difficult to recover, 5) suffers for security issues etc. Hence virtualization should be implemented in the regional health bureau data center.

II. Virtualization Plan

Planning is the preliminary task for undertaking different activities. Need assessment is also mandatory for organization in order to go with the specified plan. The regional health bureau has a future plan for implementing a virtual data center. Virtualization is now one component of the current technology cloud computing which makes things easier even if security is the barrier. All the respondents stressed the advantage of virtualization from various perspectives. The health bureau should implement virtualization after evaluating the developed framework.

III. Pros of virtualization

All the respondents indicated that the health bureau will benefit from server virtualization. All of the respondents had of the opinion that server virtualization would enable the health bureau to lower total cost of operation. Eight of the respondents were of the opinion that server virtualization would enable achievement of increased reliability and availability by providing automatic failover. Consolidating server hardware via virtualization offers ways to increase utilization of existing hardware resource from 5-15% up to 80%. Virtualization reduces hardware requirements by a 10-to-1 ratio or better; reducing space requirements; reduce human power for managing; easily disaster recovery; reduce energy cost by 80%; easily migrating machines; power down servers without affecting applications or users; greening the data center through decreasing costs and improving service levels. Almost all the respondents agree on the importance of virtualization over physical data centers. Scalability, flexibility and redundancy of server are an added advantages provided through virtualization.

IV. Challenges and Security of Virtualization

As a technology, virtualization may have its own drawbacks or challenges due to different occasions. As pointed by Randell (n.d.), configuration risk is the biggest risk to virtual infrastructure. Following best practices is an advisable mechanism to avoid the above mentioned challenges according to the researcher. One of the interviewee replied “the network topology and design is not clear so that it is difficult for management”. Two of the interviewees also stressed that lack of dedicated specialized experts is among the challenges. Lack of various security equipment’s might be also the trepidation. Hardware or software firewalls, virtual private network (VPN), intrusion prevention and intrusion detection systems, authorization and authentication and logs should be considered. Video camera surveillance and permanent guards in addition to the above mechanisms should be available. Giving awareness for the privileged user on how to use the system is also another part of securing the data center.

V. Data Center Infrastructure Management (DCIM)

A data center infrastructure management system is a system which collects and manages information about a data center asset, resource use and operational status. DCIM is proactively identified issues before escalating in to problems, modeling the data center configuration and status, finally plan for and accommodate future IT needs (Vertiv, 2017). Infrastructure management also includes managing IT devices, cooling system, power system, data center room and distributed infrastructures. Among the respondents almost all, highly believes data center infrastructure management requires an experienced expert.

Observation is also one data collection instrument by actually looking what is happened in the regional health bureau data center. The data center is traditional and did not have any standard design. The data center did not have a free space for future expansions and still uses a normal key which is not recommended in developed countries.

4.4 Performance Analysis Experiment

4.4.1 Configuration of virtual machines

Confirming or verifying a theory with an experiment is a good supporter for the study conducted. Related literatures shown that virtual machines (VM) are created using abstracting software called hypervisor. Hyper-v from Microsoft was selected to the experiment for the purpose of creating different virtual machines or servers on one physical server. Microsoft hyper-v is installed directly on the physical machine as a role of Microsoft windows server. Proper Hard disk space and memory size should be allocated for the virtual machines. Microsoft Hyper-v can be run like active directory, DHCP, and DNS as a role on windows server 2008 or as a standalone application. Hyper-V has not been part of the Microsoft windows roles prior to windows server 2008. In this research the researcher configure hyper-v as one role of the Microsoft server 2008 shown in the following figure.

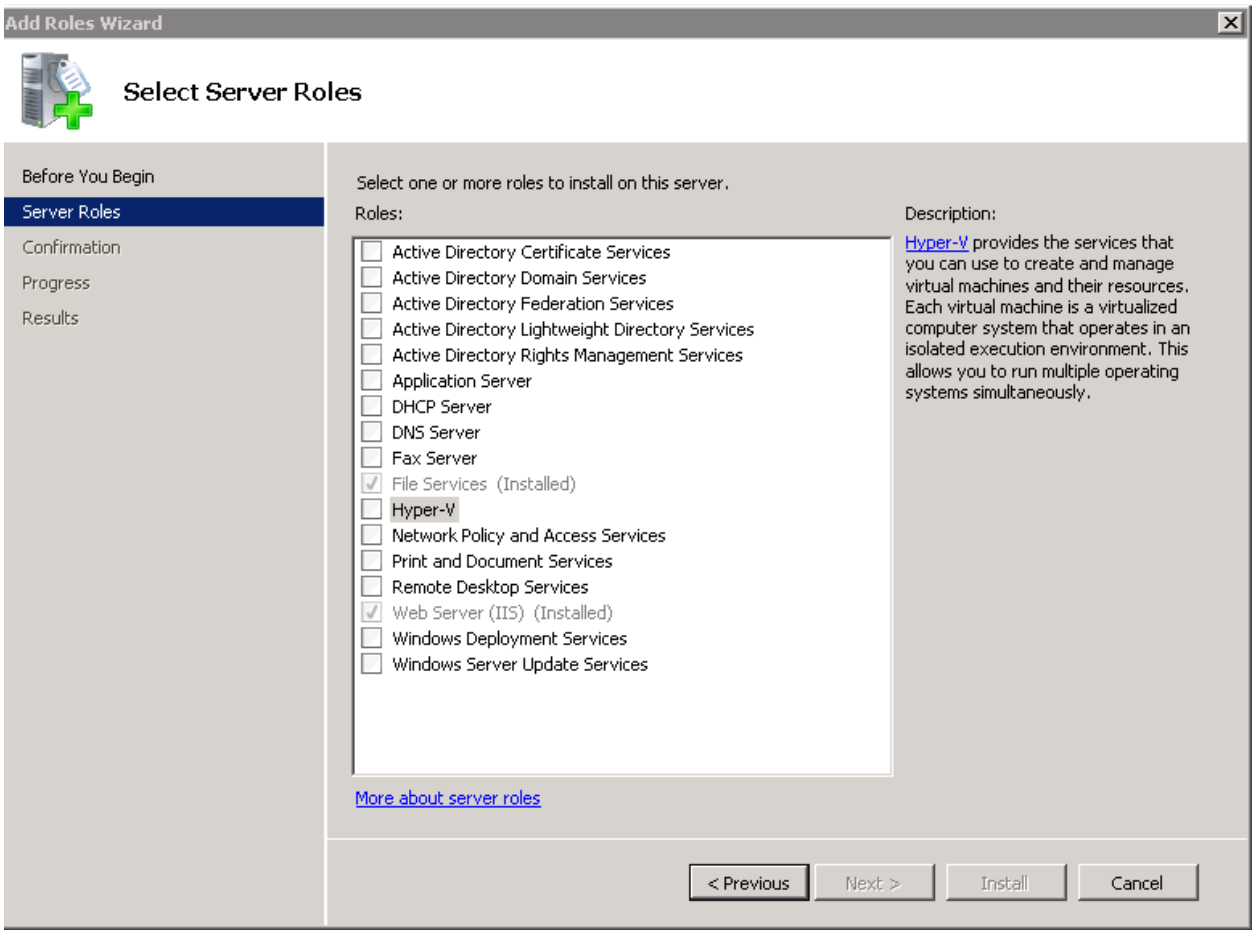


Figure 4.1: Hyper-V installation

As literature shows, proper attention should be given while configuration of hyper-V. Hence the researcher configures hyper-v properly. Next to that the application used in the regional health bureau for different purposes like electronic health management information system (eHMIS), human resource information system (HRIS) and EMR are created as servers. As indicated on the disadvantage of virtualization above the resource allocation for each virtual system needs to be planned carefully. If very less resource is allocated, the application performance might be affected and if too much resource is allocated, it will result in under-utilization. The allocation of the resources depends on the workload that the specific server performs. The user process communicates with the server which intern communicates with the actual database to implement what is required. The following figure shows the three virtual servers created on a single physical machine.

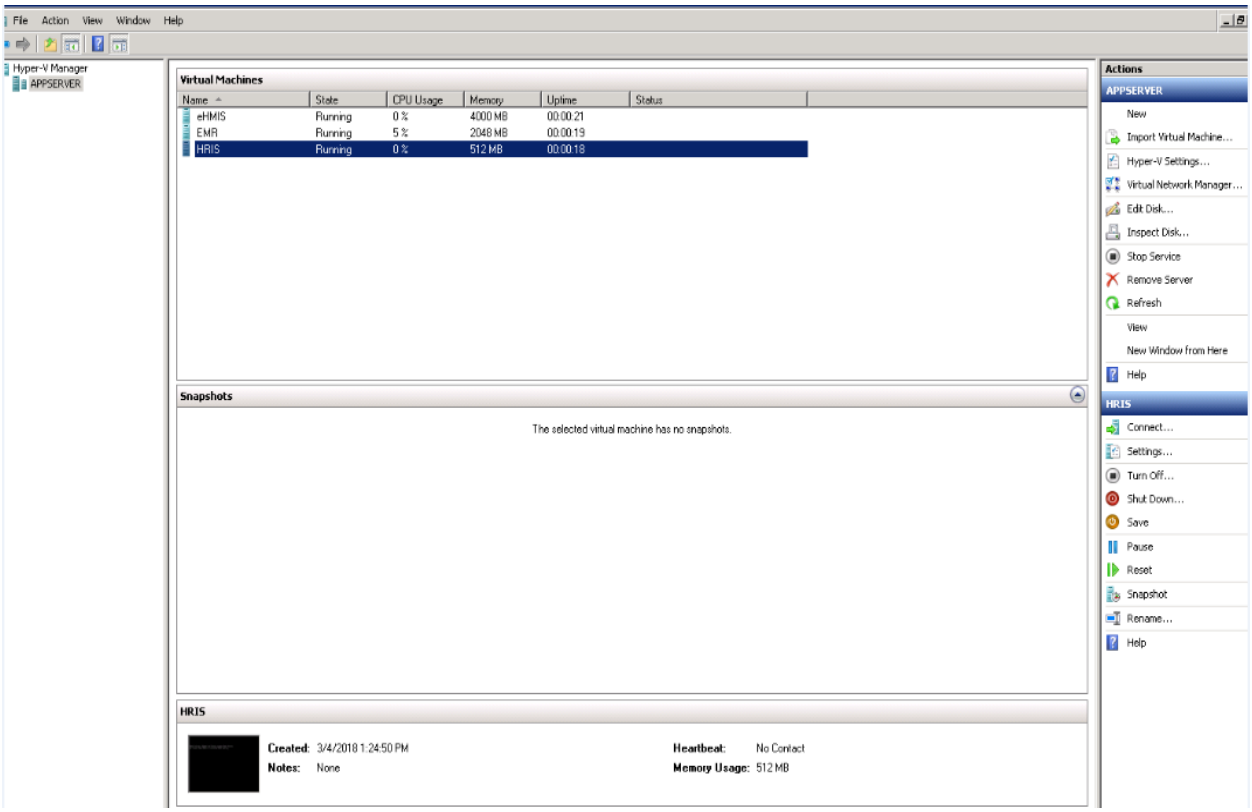


Figure 4.2: Created Virtual Machines

4.4.2 Performance Comparison Test

The intention of this section is to compare the resource utilization of physical servers with the virtual machines. For comparing the two different machines resource utilization, the researcher can use both graphical user interfaces (GUI) and command line interface tools. Non-GUI testing tools can be windows power shell, windows management instrumentation etc. On the other hand, server manager, task manager, performance monitor and resource monitors are among the GUI tools to test and compare the utilization of physical and virtual machines using different components explained above. The researcher selected one of the graphical user interface tools called resource monitor. The physical servers has RAM 8GB, Processor 2 CPU AMD Opteron (tm) Processor 6128, HDD (storage capacity) 300GB which is running one of the application in the health bureau eHMIS. While the single application is running in the physical server it shows the following usage.

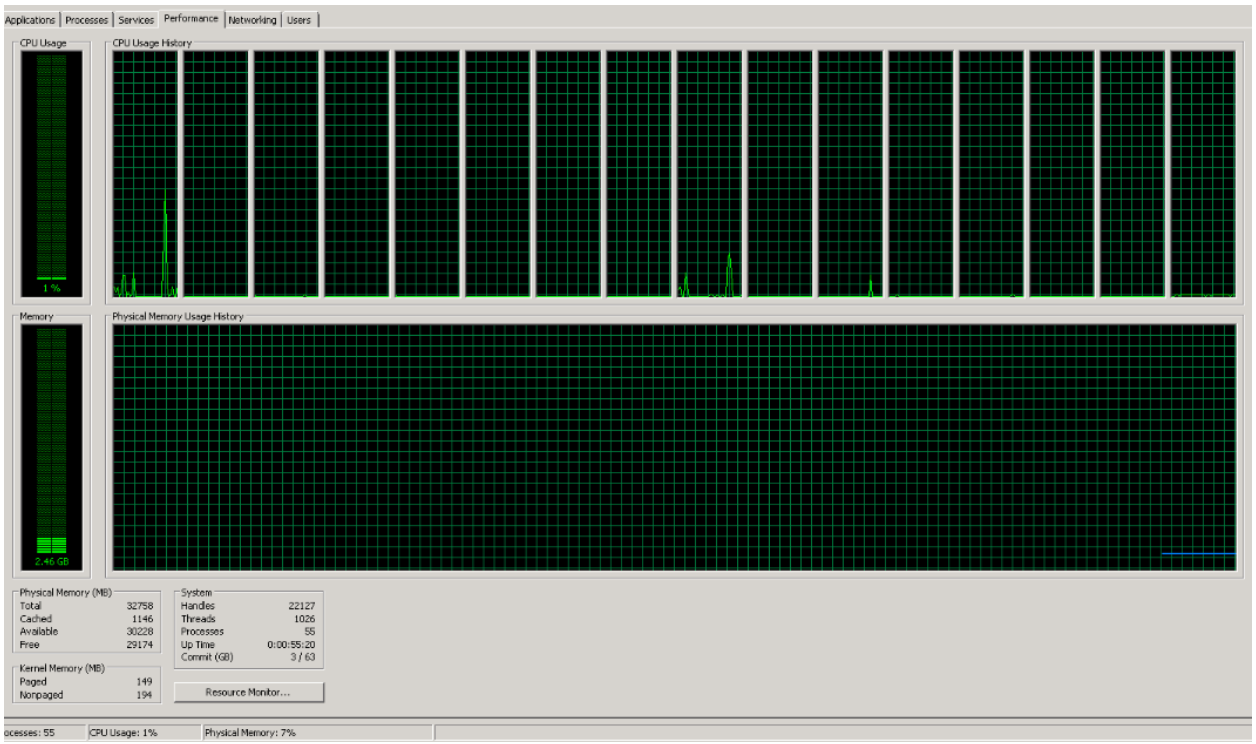


Figure 4.3: Performance of physical servers before virtualization

From the above figure the application eHMIS uses 1% of CPU and 7% of memory which is very low. As indicated in the literature review part, the utilization of physical resources of most physical servers is averagely 5-15%. But it is recommended to use up to 80% of its resources (CPU, physical memory etc). Here the experiment reveals that a one server to one application architecture is not business oriented. Hence, it is advisable to deploy many servers/ virtual machines on a single physical server. The researcher deployed three virtual instances on the same machine running three distinct applications eHMIS, HRIS and EMR. In this experiment the physical memory, CPU and hard disk is given for the individual virtual machines but network is left to be shared. In the following figure the performance of a server after virtualization is shown.

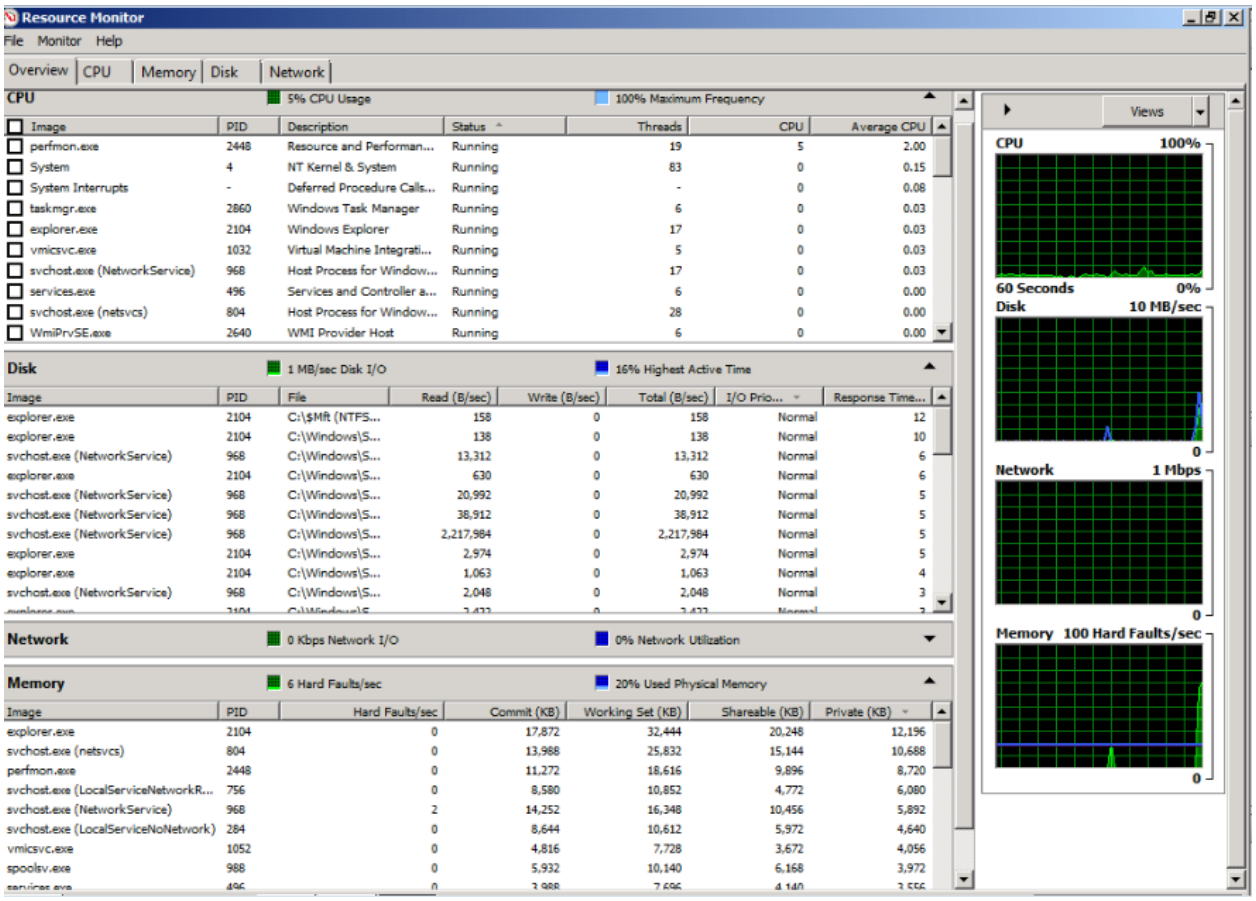


Figure 4.4: Performance of eHMIS virtual server

As the above figure 4.4 discloses, virtual machine running eHMIS uses 5% of CPU and memory. But the CPU usage and memory of any system or server is dependent on the state. While the server is running and many users are interacted with the system the usage of resources are increasing. Servers with different workloads and computations will have different resource usages without ignoring other facts like the resources the original servers have and also the version of servers.

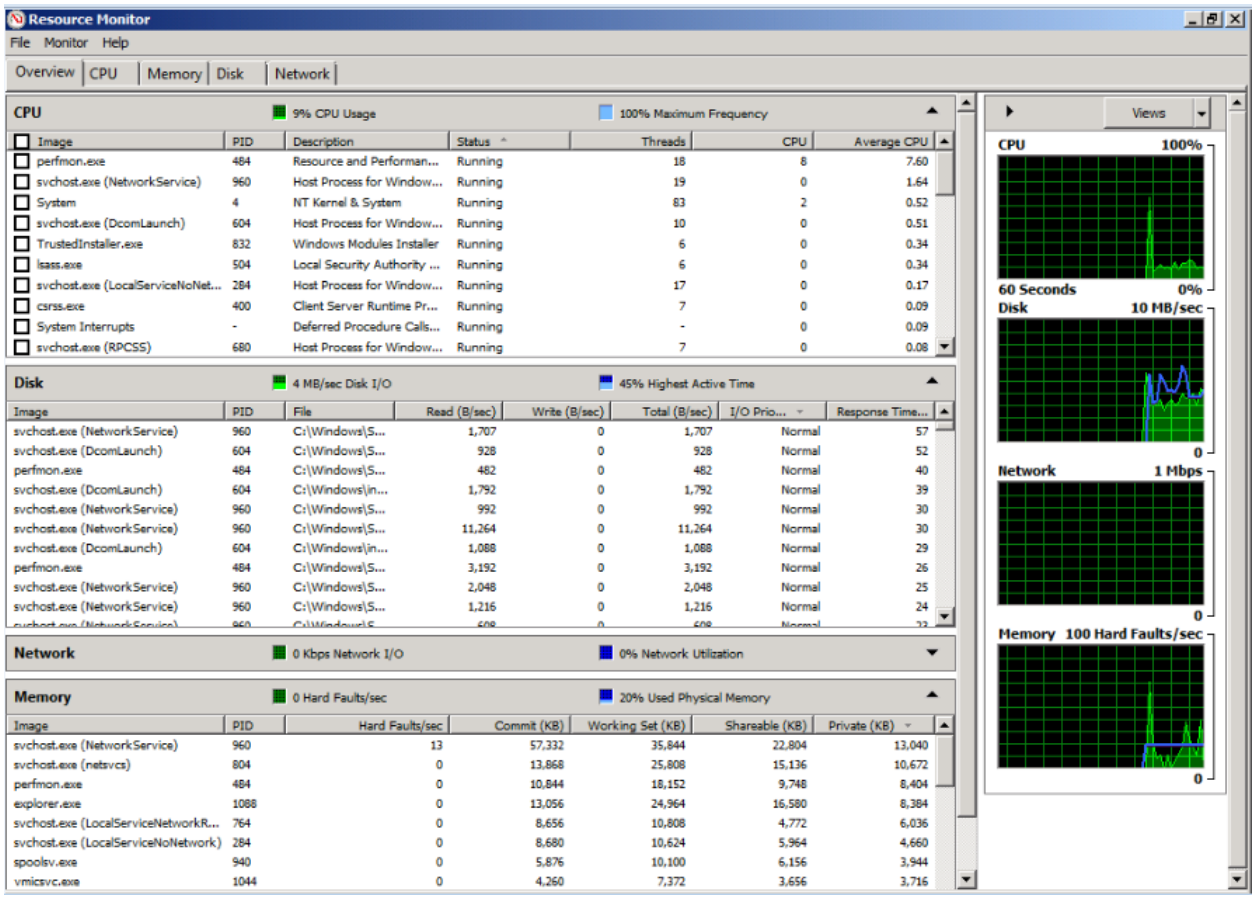


Figure 4.5: Performance of EMR virtual server

As indicated in the figure 4.5 above, by deploying EMR it consumes 9% of central processing unit and 20% physical memory which is less significant when compared with maximum utilization supported.

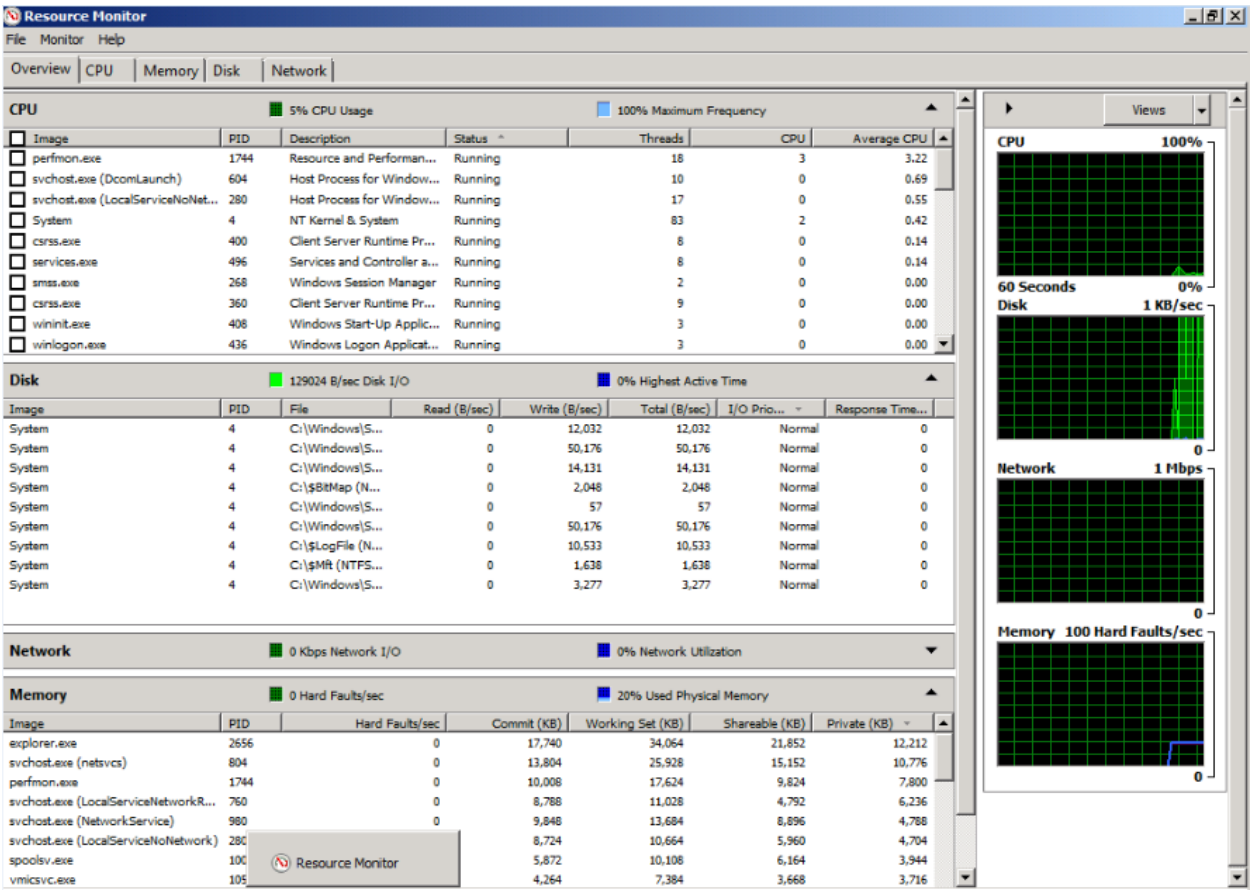


Figure 4.6: Performance of HRIS virtual server

The other virtual server that the researcher deployed is Human Resource information system which is used by the health bureau to collect process and store information to support human resource/HR decision making, coordination and control in an organization. It also consumes 5% of the shared central processing unit, and 20% of the actual memory.

From the above experiment it is business oriented and recommended to deploy more than one virtual server within a single physical machine to utilize the server appropriately. The number of virtual machines deployed on a single server depends on the amount of workload it had and the number of users interacted with the server. Therefore by using this experiment it is possible implement virtualization in various arenas.

4.5 Chapter summary

In this chapter the data from interview, observation and document review is properly analyzed and verified. The experiment using virtual machine monitor (VMM) reveals that not all resources can effectively use their resources. It also discloses the possibility to run multiple virtual machines on the same physical machines which have multiple advantages

The next chapter shows the developed data center virtualization framework and its evaluation results.

CHAPTER FIVE

THE PROPOSED VIRTUALIZATION FRAMEWORK

After the collected data from interview and observation is analyzed proper data center virtualization framework should be designed for the regional health bureau.

5.1 Type of virtualization selected for the health center

Virtualization technology is the one and the best step for cloud computing. There are different types of virtualization which is selected according to the need of an organization.

Server virtualization: is a virtualization technique involving the partitioning of a physical server into a number of small, virtual servers with the help of virtualization software. Server virtualization is selected for the regional health bureau because its data center is at infancy stage and needs more servers if deployed traditionally. By implementing server virtualization the health bureau can get scalability, availability, better uptime, easy migration, sustainable backup and recovery and many more.

5.2 Datacenter Virtualization Framework Implementation Model

Implementing virtualization for the health bureau is for the aim of providing less spending, easier backup and disaster recovery, better business continuity and more efficient IT operations. Virtualization divides the computing resources of a server based environment to provide different operating environments using different methodologies and techniques like hardware and software partitioning or aggregation, partial or complete machine simulation, emulation and time sharing (Uddin, & Rahman, 2011). Data center can be implemented either in bare or hosted model. Native model is used to implement a framework for the regional health bureau to achieve its vision.

5.3 Data center virtualization method selected for the regional health bureau

There are two widely known and accepted server virtualization methods. These are full virtualization and Para –virtualization (Calzolari, 2006; Lingeswaran, 2017).

When the virtual machine (created servers) simulates enough hardware to allow an unmodified guest operating system to be run in isolation from the other operating system it is full virtualization. Full virtualization fully abstracts the guest operating environment from the physical hardware. Even the guest operating system did not aware that it is run as a virtual environment and did not require modification. But, in Para-virtualization method the guest operating system should be modified which creates portability and compatibility issues. Even non modified operating system (like windows) cannot be supported by Para virtualization.

Full virtualization is deployed in the regional health bureau data center due to its giant advantage over Para virtualization. Providing best isolation and security for virtual machines and allowing procedures for migration and portability as the same guest operating system can run in the virtualized environment.

Virtual box, parallels, virtual pc, virtual server, Microsoft hyper-V (used in this thesis), QEMU and VMware are among the recent products supporting full virtualization. Full virtualization method was proposed in the regional health bureau data center for creating virtual servers. Modifying the guest operating in the case of Para virtualization creates several complexities so that it is not selected for the paper.

5.4 Proposed Server Virtualization Framework for Amhara Regional Health bureau

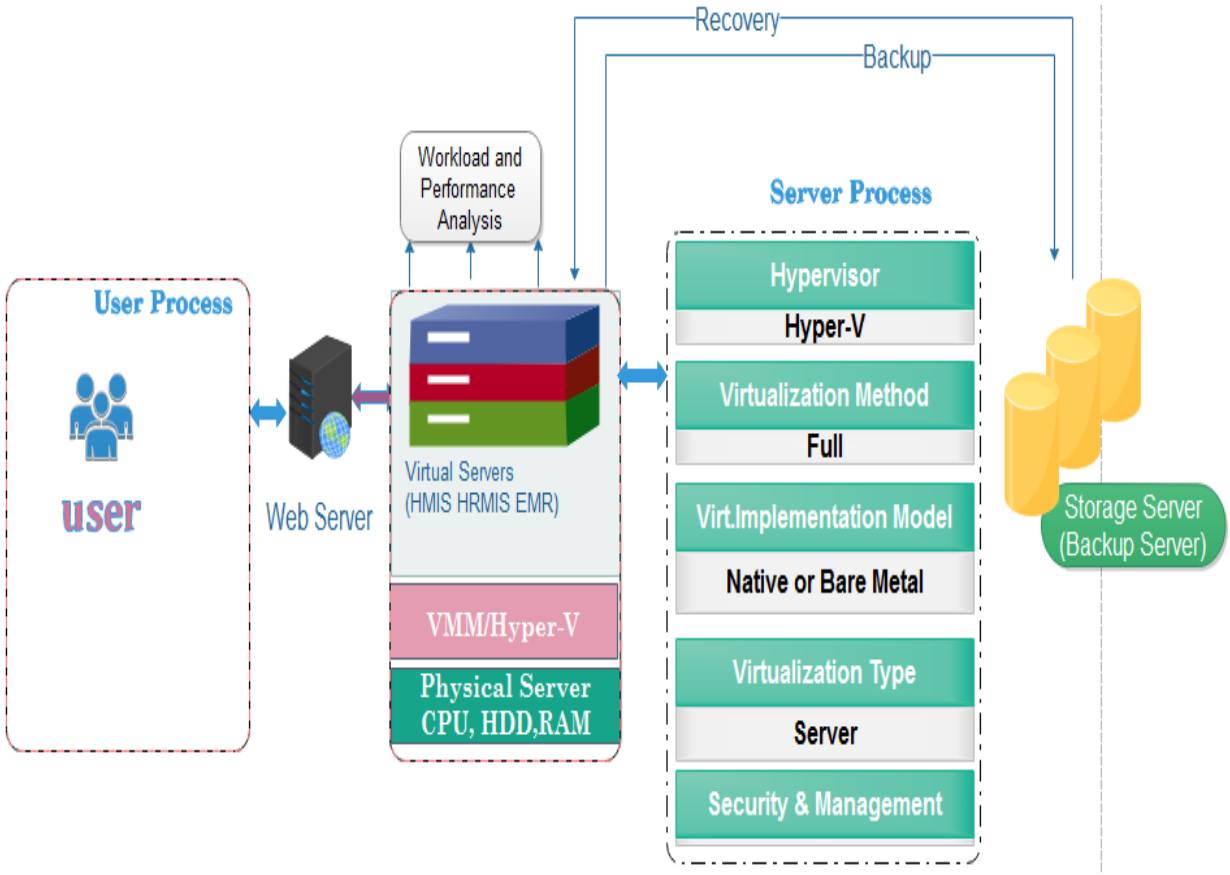


Figure 5.1: The proposed framework

The above figure clearly shows the developed framework for the health bureau. The developed framework is not generic but, can be implemented in similar organizations by modifying some of its features. It comprises of many different components required in the data center. The main components include:

Users: they are authorized peoples who are responsible for entering transactional data or who wants the data for research purpose. These include the ICT manager, network and system

administrators, ICT officers, database administrators and others who takes a mandate by the regional health bureau.

Application: Accessing any of the virtualized machines (servers) directly from the user's computer is through the application. The applications in the regional health bureau include human resource management information system (HRMIS) which is a set of inter-related components working together to collect, process and store information to support Human Resource decision making, coordination and control in an organization. Health management information system (HMIS)which helps to provide accurate, timely and complete data to support decision making at each level of the health system.

Web Server: in a client server architecture a client computer requests a service and a server responds to the request. A web server is a system that delivers content or services to the end user over the internet. Application server (HP ProLiant DL385 G7Windows server 2008 R2 SP1) with specification: RAM 8GB, Processor 1 CPU 2.0 Ghz Intel Xeon E5504 @3.00Ghz, 6128, HDD (storage capacity) 750GB and above are required for better functionality.

Virtual Servers: virtual machines (instances) are servers that are created on the physical server having its own hard disk and memory. As indicated in the chapter three above virtual servers are created on the same physical host. The first server running HMIS, the second runs HRMIS and the third EMR. Other required virtual instances can be deployed as the same fashion with the other virtual machines.

Hypervisor: Hyper –V is selected because according to Microsoft (2013), it has greater scalability, performance, secure & flexible infrastructure, and also has high availability and resiliency.

Management and Security

Management and security is the pillar for any data center without which a data center virtualization becomes meaningless. These two major components are really required for both physical and virtual data centers. Hardware and software firewalls, virtual private network, intrusion detection and intrusion prevention systems, video camera surveillance, permanent guards are among the mechanisms to secure certain data center. Permanent security guards, video surveillance camera and firewalls were recommended for the regional health due to its

status. ARHB also needs permanent information communication Technology experts for managing and controlling the data center.

Storage Server (Backup)

It is wise to take a backup at any time. Enabling this feature on the virtual machines saves an organization from data loos during system crash. Generally backup servers enables backup of data, application and database both in a specialized in-house and a remote server. Backup can be full or incremental (How to choose the correct backup type, 2008). Full back up means a strategy of copying all the data to an appropriate media like tape, disk or DVD but, incremental back up is copying data that is only changed since the last backup operation. Incremental backup was recommended for the regional health bureau since it saves resources of the backup server. To store backup at a remote server proxy server might be required. The process of putting copy of a data, application or database on another place is called back upping but restoring from the backup server is called recovering.

5.4.1 Guidelines of full server virtualization implementation

For implementing full virtualization, let's start from a physical machine running server 2008. Now follow the following to install Microsoft hyper-v as a window role.

1) Click Start, and then click Server Manager.

2) In the Roles Summary area of the Server Manager main window, click Add Roles.

3) On the Select Server Roles page, click Hyper-V.

4) On the Create Virtual Networks page, click one or more network adapters if you want to make their network connection available to virtual machines.

5) On the Confirm Installation Selections page, click Install.

6) The computer must be restarted to complete the installation. Click Close to finish the wizard, and then click yes to restart the computer.

7) After you restart the computer, log on with the same account you used to install the role. After the Resume Configuration Wizard completes the installation, click Close to finish the wizard.

After you have installed Hyper-V, you can create a virtual machine and set up an operating system on the virtual machine.

Before you create the virtual machine, you may find it helpful to consider the following questions. You can provide answers to the questions when you use the New Virtual Machine Wizard to create the virtual machine.

- Is the installation media available for the operating system you want to install on the virtual machine? You can use physical media, a remote image server, or an .ISO file. The method you want to use determines how you should configure the virtual machine
- How much memory will you allocate to the virtual machine?
- Where do you want to store the virtual machine and what do you want to name it?

Now, it is time for creating virtual instances using the following steps.

1) Open Hyper-V Manager. Click Start, point to Administrative Tools, and then click Hyper-V Manager.

2) From the Action pane, click New, and then click Virtual Machine.

3) From the New Virtual Machine Wizard, click Next.

4) On the Specify Name and Location page, specify what you want to name the virtual machine and where you want to store it.

5) On the Memory page, specify enough memory to run the guest operating system you want to use on the virtual machine.

6) On the Networking page, connect the network adapter to an existing virtual network if you want to establish network connectivity at this point.

Note

If you want to use a remote image server to install an operating system on your test virtual machine, select the external network.

7) On the Connect Virtual Hard Disk page, specify a name, location, and size to create a virtual hard disk so you can install an operating system on it.

8) On the Installation Options page, choose the method you want to use to install the operating system:

Install an operating system from a boot CD/DVD-ROM. You can use either physical media or an image file (.iso file).

Install an operating system from a network-based installation server. To use this option, you must configure the virtual machine with a network adapter connected to the same network as the image server.

After you create the virtual machine, you can start the virtual machine and install the operating system.

In the final step of this process, you connect to the virtual machine to set up the operating system. As part of the setup, you install a software package that improves integration between the virtualization server and the virtual machine.

The instructions in this step assume that you specified the location of the installation media when you created the virtual machine. The instructions also assume that you are installing an operating system for which integration services are available.

1. From the Virtual Machines section of the results pane, right-click the name of the virtual machine you created in step 2 and click Connect. The Virtual Machine Connection tool will open.

2. From the Action menu in the Virtual Machine Connection window, click Start.

3. Proceed through the installation.

Tips:

When you are at the point where you need to provide input to complete the process, move the mouse cursor over the image of the setup window. After the mouse pointer changes to a small dot, click anywhere in the virtual machine window. This action "captures" the mouse so that keyboard and mouse input is sent to the virtual machine. To return the input to the physical computer, press Ctrl-Alt-Left arrow and then move the mouse pointer outside of the virtual machine window.

After the operating system is set up, you are ready to install the integration services. From the Action menu of Virtual Machine Connection, click Insert Integration Services Setup Disk. If Autorun does not start the installation automatically, you can start it manually. From a command prompt, type:

%windir%\support\amd64\setup.exe.

After you have completed the setup and integration services are installed, you can proceed to test the virtual machine by customizing it to suit your testing goals. For example, you can view or modify the virtual hardware that is configured for the virtual machine. From the Virtual Machines pane, right-click the name of the virtual machine you created in step 3 and click Settings. From the Settings window, click the name of the hardware to view or change it.

5.5 Evaluation of the proposed Framework

After the required framework is developed it is mandatory to be evaluated by different ICT experts in the regional health bureau. The developed framework has evaluated in two phases. In the first phase it was presentedd for information communication technology staffs of the regional health bureau for further and in-depth evaluation. After the necessary commens and feedbacks have been revised and corrected the above framework was developed. The second phase of evaluation is held by including one from finance and all the information communication Technology staffs having experiences in different areas. The reason why the researcher and the ICT staffs wants a person from the finance to be engaged in the presentation, is because he will help to facilitate the implementation process.

The group considers the framework from different directions and evaluates using business continuty metrics and used the artifact evaluation criteria above.

Business continuity and Importance

As evaluators approved that the developed framework is understandable, simple and easy to use. This features is consistent with the future expansion of the data center. The data center even does not have a website indicating that it requires to support running of various applications for the future inorder to meet the bureau's requirement. They suggest that the developed framework is consistent with the organization and helps to achieve the mission. The technology (Virtualization) definetely helps to reduce the infrastructure requirements by consolidating different machine to be run on a single physical host so that dramatically decrease cost. Apart from cost reduction virtualization can:

- ✓ Reduce lectric power consumptionresulting in less bill payment
- ✓ Reduce the number of man power for managing the data center
- ✓ Green the environment and contributes for creating suitable world
- ✓ Reduce the space required in the data center
- ✓ Facilitate delivery of quality of service
- ✓ Enhance migration and recovery of virtual mschines
- ✓ Increased flexibility so that additional clients can be added on demand by creating session.

More over, the evaluators stressed the value of virtualization on harnessing the environment and contributes for creating suitable world.

5.6 Chapter Summary

In this chapter the virtualization framework is developed for the regional health bureau which believes to solve the existing problems. After the framework was developed it was evaluated by the regional health bureau Information Technology Staffs using different Metrics.

The next chapter will conclude the researchers work and sets directions for further study.

CHAPTER SIX

CONCLUSION AND RECOMMENDATIONS

6.1 Overview

The previous chapter was focused on analyzing data gathered from different data collection instruments and finally an appropriate framework was developed. The ultimate purpose of this chapter is to conclude and summarize the study and indicating future direction for fellow researchers.

6.2 Conclusion

The research assessed the physical data center in the regional health bureau. The current data center is a one server-one application architecture which is not currently advisable. The result from interview and observation highly confirmed the above problems in the data center which is an obstacle for the mission of the bureau.

This study is conducted to achieve the general objective which is developing a virtual data center framework for Amhara regional health bureau. To achieve this objective specific objectives are used. Reviewing different literature related with the study from different sources like articles, books and the like are among the sources.

The literature review briefly describes various topics such as what a data center is, major components of data center, from where the technology comes, virtualization definition, evolution, types of virtualization, method of virtualization, implementation model and other related topics.

The current data center is properly observed and interview with the network and system administrators, ICT manager, and ICT officers discloses the different problems in the data center. Data center should require a special house having special technology which protects the organizations data from any disaster but the health bureau data center is located in one of the room which is not proper for it. There is no advanced fire protection equipment installed in the

data center meaning it is less resistance to fire if it happens. Since the data center does not have automatic generators power interruption and shutting of servers is common which is not recommended. The data center does not have extra white spaces which is difficult to accommodate other infrastructure for the future. Generally the data center does not follow modern international standards.

Based on observation, literatures and interview results a server virtualization framework was developed for the regional health bureau. The framework was developed for ARHB but, can be easily modified for other organizations that are in need of server virtualization. The developed framework constitutes all features but, depends on the statues of the organization.

In this technological era servers, networks, applications, operating systems and other infrastructures can be virtualized. Server virtualization is the main focus of this study which is a technology of dividing a single physical machine to run multiple virtual machines. Server virtualization has various advantages including lower infrastructure cost, easy management, faster and easy disaster recovery, reduce human power, better security and greening environment.

Two major virtualization methods are concisely discussed in the literature review part above. These methods are full virtualization and Para - virtualization according to some scenarios. Hardware assisted full virtualization method which is used in this study does not allow modification of either the host or guest operating system. The virtual machine even does not aware as it runs in a virtual environment. This could be achieved using hypervisors like Microsoft hyper-V which is an integral part of windows server 2008. Full virtualization can avoid compatibility and portability issues raised from operating system modification. Para – virtualization on the other hand does not support unmodified guest operating system in which compatibility and portability becomes a big issue.

The server in a data center utilizes its resources including RAM, CPU and other resources below average. As literature reveals only 5-10% of physical resources are utilized by the system so that more than 90% are lost which is not suggested. This study also confirmed that it is possible to install multiple virtual machines on a single physical machines resulted in getting reduced cost

and other advantages. In fact the usage of resources of the machine highly depends on the number of user using the particular system and the type of workload existed on it.

One major problem in virtualization is assigning resources to virtual machines appropriately. Randell (n.d.) also confirmed that configuration risk is the biggest risk to virtual infrastructure. Planned amounts of hardware resources should be assigned for each virtual machine. Allotting extremely low amount of resources during configuration may result in low utilization of the server.

Security and management practices are not properly implemented in the current data center. The bureau focuses on expanding infrastructures and adapting technologies but much attention is not given for security issues. Since data center has critical and sensitive information, security practice and management is indisputable. Even there is a barrier in the physical security tasks. Entry controls to limit physical access of sensitive resources like servers using video camera should be implemented. As Susanto et al. (2011) describes, physical and environmental security are used to protect systems, buildings, and related supporting infrastructure against threats associated with their physical environment, buildings and rooms to protect the environment and to avoid damage or unauthorized access to information and systems.

Implementing physical security alone does not make the data center secure hence, integration of logical and physical security mechanisms are important. Access control, firewall, Intrusion detection system (IDS), intrusion prevention system (IPS), wireless scanning, virtual private network (VPN), password complexity, Virtual Local area network (VLAN), changing manufacturers default passwords, encryption, log file analysis and using licensed antivirus in combination with physical security should be implemented in the data center.

6.3 Recommendations

Based on the study findings and the above conclusions the following recommendations are forwarded.

a) The researcher recommended the regional health bureau to implement server virtualization for their data center to be benefited from the above mentioned advantages.
b) The health bureau should prepare appropriate training program for the respected staffs.
c) Other organizations should implement data center virtualization to increase level of utilization of servers and decrease server hardware purchase expenditure.

6.4 Limitation

- The accessibility of recent literatures published on server virtualization ultimately affects the researchers work.
- Since there were no free servers in the regional health bureau, it was difficult to do the experiment more than once.

6.5 Suggestion for Future Research

There are still several issues regarding virtualization that warrant further research. Hence, the researcher recommends study on the following areas:

- Studying other virtualization types like storage, network and application virtualizations.
- Implement full server virtualization using different Algorithms to minimize performance overhead.
- Assessment of user's satisfaction on the virtualized environments.
- Exploring incompatibility between virtualization software and the underlining Hardware's.
- Exploring security aspects of server virtualization by developing a model.

REFERENCES

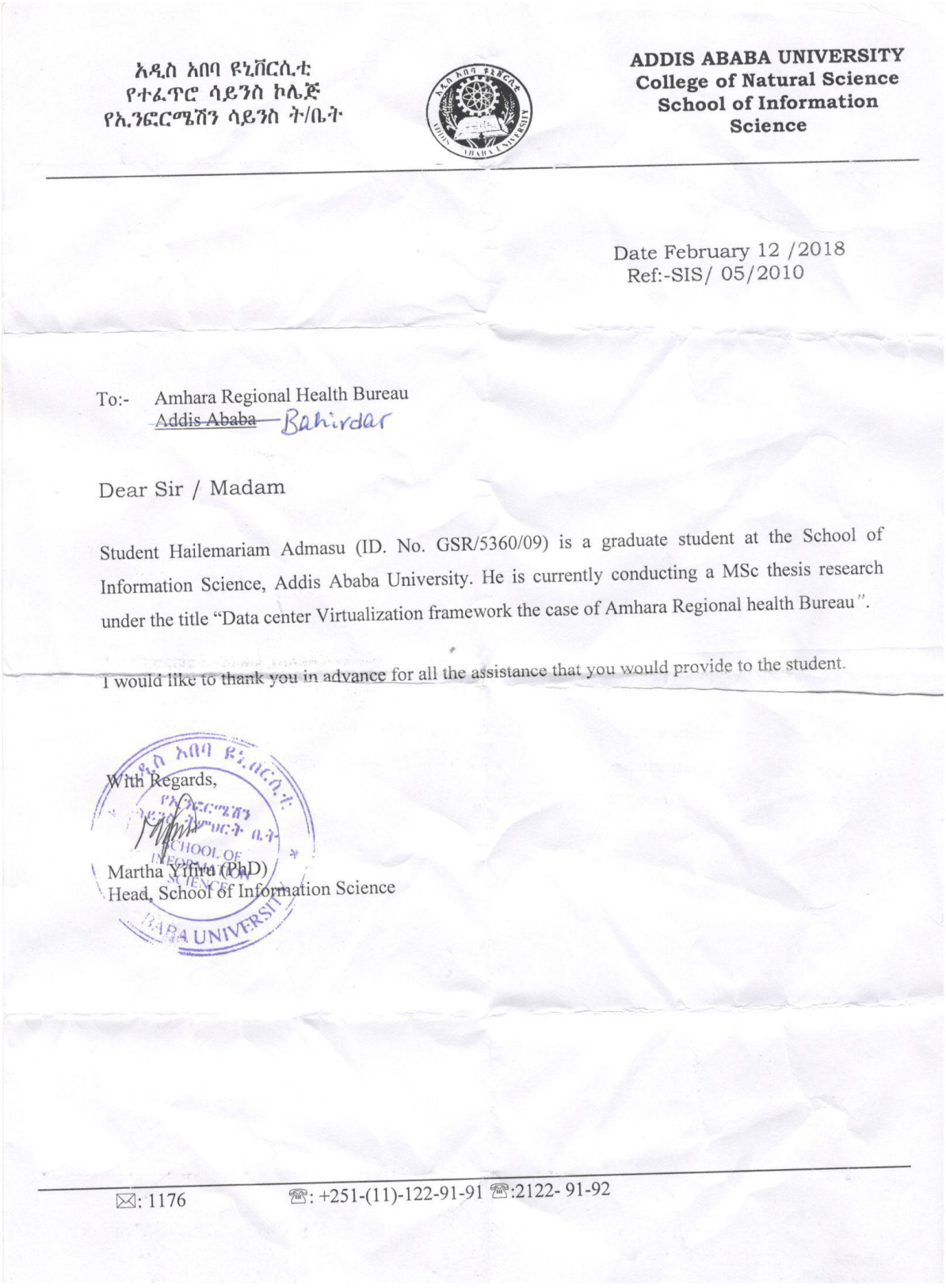
1. Acharya, A. S., Prakash, A., Saxena, P., & Nigam, A. (2013). Sampling: Why and How of it? Indian journal of medical specialties, 4(2), 330-333.
2. ADC Telecommunications (2006). TIA-942 Data Center Standards Overview. USA.
3. Alvi, M. (2016). A Manual for Selecting Sampling Techniques in Research.
4. Anixter (2012). Data Center Infrastructure Resource Guide.
5. Balodis R, & Opmane. I. (n.d.). History of Data Center Development. SANOG: Open Communication Limited.
6. Calzolari, F. (2006). High Availability Using Virtualization. Anno.
7. Campbell, S., & Jeronimo, M. (2006). An Introduction to Virtualization. Hillsboro: Intel Corporation.
8. Castellan, C. M. (2010). Quantitative and Qualitative Research: A View for Clarity. International Journal of Education, 2(2).
9. Dolores, M. and Tongco, C. (2007). Purposive Sampling as a Tool for Informant Selection Research Methods. Journal of plants people and applied research, ethnobatany.148-158.
10. Getnet, W. (2017). Data Center Virtualization Framework in banking sector. Addis Ababa University: Addis Ababa.
11. Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science In Information Systems Research. Journal of Information System Quarterly, 28(1).
12. Hossain, J. (2011). Data Center Design and Virtualization. Sanog.
13. How to choose the correct backup type. (2008). Retrieved may 10, 2018, from https://searchdatabackup.techtarget.com/feature/Full-incremental-or-differential-How-to-choose-the-correct-backup-type.
14. I.T. Alliance (2010). Alliance framework for for server virtualization.
15. IBM Global Education (2007). Virtualization in Education. United States of America: IBM Corporation.
16. Kant, K. (2009). Data center evolution. Computer Networks. Retrieved from www.elsevier.com/locate/comnet.

17. Koomey, J. G. (2011). Growth in data center electricity from 2005 to 2010. Analytics Press.
18. Lacey, A. and Luff, D. (2007) Qualitative Research Analysis. The NIHR RDS for the EastMidlands/Yorkshire & the Humber, 2007.
19. Lingeswaran, R. (2017). Para virtualization vs Full virtualization vs. Hardware assisted Virtualization. Retrieved April 11, 2018, from https://www.unixarena.com/2017/12/para-virtualization-full-virtualization-hardware-assisted-virtualization.html.
20. Lowe, S. D., Green, J., & Davis, D. (2016). Building a Modern Data Center: Principles and Strategies of Design. Atlantis Computing.
21. Malek, J. (2013). Centralized deployment of end user stations in an enterprise environment. Brno.
22. Marczyk, G., DeMatteo, D., & Festinger, D. (2005). Essentials of Research Design &Methodology. John Wiley & Sons, Inc.
23. Mell, P.,& Grance, T. (2011). The NIST Definition of Cloud Computing:Recommendations of the National Institute of Standards and Technology .
24. Metzler, J. (2011). Virtualization: Benefits, Challenges, and Solutions. Riverbed Technology.
25. Microsoft (2013). Why Hyper -V? Competitive Advantages of Microsoft Hyper -V Server 2012. Microsoft Corporation.
26. Moltchanov, D. (2013). Client/server and peer-to-peer models: basic concepts.
27. Murugesan, S. (2008). Harnessing Green IT: Principles and practices. I E E E Comp u t e r S o c i e t y.
28. Newcombe, L. (n.d.). Data centre energy efficiency metrics: Existing and proposed metrics to provide effective understanding and reporting of data center energy. Data Centre Specialist Group.
29. Padhy, R. P., Patra, M. R., & Satapathy, S. C. (2011). Virtualization techniques & technologies: state-of-the-art. Journal of Global Research in Computer Science, 2(12).
30. Peffers, K., Tuunanen, T., Gengler, C.E., Rossi, m., Hui, W., Virtanen, V., & Bragge, J. (2006). The Design Science Research Process:A Model For Producing And Presenting Information Systems Research . Claremont.

31. Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems, 24(3), pp. 45-78.
32. Prat, N., Comyn-Wattiau, I., & Akoka, J. (n.d.). Artifact evaluation in information systems design-science research – a holistic view. France.
33. Randell, R. (nd) Virtualization Security and Best Practices. VmWare.
34. Rasmussen, N. (2006). Implementing energy efficient data centers. American Power Conversion.
35. Ronkainen, N. (2003). Server Virtualization: master‟s thesis
36. Strassmann, P. A. (2008). Introduction to Virtualization. George Mason University.
37. Susanto, H., Almunawar, M. N., & Tuan, Y. C. (n.d.). Information Security Management System Standards: A Comparative Study of the Big Five. International Journal of Electrical & Computer Sciences, 11(05).
38. Tvaruzek, S. (2009). The Shift Towards Cloud Computing. brno.
39. Uddin, M., & Rahman, A. A. (2011). Virtualization Implementation Model for Cost Effective & Efficient Data Centers. International Journal of Advanced Computer Science and Applications, 2(1).
40. Uddin, M., Talha, M., Rahman, A. A., Shah, A., Khader, J. A., & Memon, J. (2012). Green Information Technology (IT) framework for energy efficient data centers using virtualization. International Journal of Physical Sciences, 7(13).
41. Vertiv (2017). Data Center Infrastructure management,Managing the Physical Infrastructure. Vertiv Headquarters, 1050 Dearborn Drive, Columbus, OH, 43085, USA: Vertiv.
42. Zainelabden, A. A., & IBRAHIM, A. (2015). Service Level Agreement Assurance in Cloud Computing Data Centers. Luxembourg.
43. Zhang, S., & Fever, H. L. (2013). An Examination of the Practicability of COBIT Framework and the Proposal of a COBIT-BSC Model. Journal Economics, Business and management, 1(4).
44. Znyamweya, B. Z. (2013). Server virtualization framework: case of selected government ministries, Kenya.

45. Zhang, S., & Fever, H. L. (2013). An Examination of the Practicability of COBIT Framework and the Proposal of a COBIT-BSC Model. Journal Economics, Business and management, 1(4).
46. Zogaj, G. (2012). Virtualization and shared Infrastructure data storage for IT in Kosovo institutions. Thesis. Rochester Institute of Technology.

Appendix A: Permission Letter



አዲስ አበባ ዩኒቨርሲቲ
የተፈጥሮ ሳይንስ ኮሌጅ
የኢንፎርሜሽን ሳይንስ ት/ቤት

ADDIS ABABA UNIVERSITY
College of Natural Science
School of Information Science

Date February 12 /2018
Ref:-SIS/ 05/2010

To:- Amhara Regional Health Bureau
~~Addis Ababa~~ Bahirdar

Dear Sir / Madam

Student Hailemariam Admasu (ID. No. GSR/5360/09) is a graduate student at the School of Information Science, Addis Ababa University. He is currently conducting a MSc thesis research under the title "Data center Virtualization framework the case of Amhara Regional health Bureau".

I would like to thank you in advance for all the assistance that you would provide to the student.

With Regards,

Martha Yifiru (PhD)
Head, School of Information Science

✉: 1176 ☎: +251-(11)-122-91-91 ☎:2122- 91-92

ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

SCHOOL OF INFORMATION SCIENCE

Appendix B: Interview Guide

Interview Questions

1. Does the idea of virtualization started before in your organizations? If not what would be the reason? Which type? Server, Application, network, storage.
2. Would you please mention which type of Data center Architecture the health bureau uses? Traditional (physical), full Virtualized or Hyper converged.
3. Which type of data center standard the health bureau follows? Tier 1, Tier 2, Tier 3, or Tier 4?

   Tier 1 basic site infrastructure

   Tier 2 redundant capacity component site infrastructure

   Tier 3 concurrently maintainable site infrastructure

   Tier 4 Fault Tolerant Site Infrastructure

4. What are the challenges in your data center infrastructure architecture?
5. What is annual budget for information technology department related to your data center? And do you have standard and legal framework or policy guideline for new technological infrastructure implementation? To what extent its functional level?
6. What kinds of benefits could you get from virtualized infrastructure?
7. Do you believe most of your data center infrastructure are outdated and has limited capabilities for health bureau services? Yes/ No and Why?
8. Do you believe health institutions should follow the traditional data center infrastructure or should run with virtualized technology? Yes/ No Why?

9. Can you confidently suggest your organization can use virtualized data center? Yes/ No Why?
10. Do you believe most top management members and IT infrastructure decision makers are usually change resistant? Yes/ No Why? If so what would be the reason?
11. Do you believe virtualized data center solution can cause unnecessary complication for the health bureau and data center administrators? Yes/ No Why?
12. How do you rate the existing traditional (physical) data center infrastructure related to security and management issues?
13. Do you believe the current infrastructure is fully capable of holding health bureau service which satisfies the mission and vision of the health bureau? Yes/ No Why?
14. How the health bureau identifies its requirements prior to select best practices in data center? Does it use a framework?
15. Do you believe the Infrastructures in the data center capable of supporting virtualization technology?
16. Is that easy to have one application per a server if the services in the health bureau expand?
17. Do you have a specialized data center expert who is responsible for managing the different activities? If not who is the responsible body?