Optimization Of Multimedia Big Data Over Cloud Architecture

Submitted By

Shaktirajsinh Jadeja

13MCEN26

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2015

Optimization Of Multimedia Big Data Over Cloud Architecture

Major Project

Submitted in partial fulfillment of the requirements

for the degree of

Master of Technology in Computer Science and Engineering(Networking Technologies)

Submitted By

Shaktirajsinh I. Jadeja

(13MCEN26)

Guided By

Prof. Priyanka Sharma

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2015

Certificate

This is to certify that the major project entitled **"Optimization Of Multimedia Big Data Over Cloud Architecture"** submitted by **Shaktirajsinh Jadeja (Roll No: 13MCEN26)**, towards the partial fulfillment of the requirements for the award of degree of Master of Technology in Computer Science and Engineering of Institue of Technology, of Nirma University, Ahmedabad, is the record of work carried out by him under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this project, to the best of my knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Prof. Priyanka Sharma
Guide & Professor,
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Prof. Gaurang Raval
Associate Professor,
Coordinator M.Tech - NT
Institute of Technology,
Nirma University, Ahmedabad

Dr. Sanjay Garg
Professor and Head,
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr K Kotecha
Director,
Institute of Technology,
Nirma University, Ahmedabad

Statement of Originality

I, **Shaktirajsinh Jadeja**, Roll. No. **13MCEN26**, give undertaking that the Major Project entitled "**Optimization Of Multimedia Big Data Over Cloud Architecture**" submitted by me, towards the partial fulfillment of the requirements for the degree of Master of Technology in **Computer Science & Engineering** of Institute of Technology, Nirma University, Ahmedabad, contains no material that has been awarded for any degree or diploma in any university or school in any territory to the best of my knowledge. It is the original work carried out by me and I give assurance that no attempt of plagiarism has been made. It contains no material that is previously published or written, except where reference has been made. I understand that in the event of any similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

Signature of Student

Date:

Place:

Endorsed by

Prof. Priyanka Sharma

(Signature of Guide)

Acknowledgements

It gives me immense pleasure in expressing thanks and profound gratitude to **Prof. Priyanka Sharma**, Associate Professor, Computer Science Department, Institute of Technology, Nirma University, Ahmedabad for his valuable guidance and continual encouragement throughout this work. The appreciation and continual support she has imparted has been a great motivation to me in reaching a higher goal. Her guidance has triggered and nourished my intellectual maturity that I will benefit from, for a long time to come.

It gives me an immense pleasure to thank **Dr. Sanjay Garg**, Hon'ble Head of Computer Science and Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his kind support and providing basic infrastructure and healthy research environment.

A special thank you is expressed wholeheartedly to **Dr K Kotecha**, Hon'ble Director, Institute of Technology, Nirma University, Ahmedabad for the unmentionable motivation he has extended throughout course of this work.

I would also thank the Institution, all faculty members of Computer Engineering Department, Nirma University, Ahmedabad for their special attention and suggestions towards the project work.

- Shaktirajsinh Jadeja

13MCEN26

Abstract

Multimedia servers are quite difficult to handle when the frequency of data exchange is too large. In todays scenario multimedia industry is growing rapidly. Multimedia servers have lot of other application with multimedia broadcasting like handling of such big video files , retrival of files when query fired by user , securing the communication etc. One of the major application of multimedia server is to broadcast 3-D videos over network.

Holography is a emerging technology in the field of image projection and it is one of the main source for multimedia server as raw data. 3D movies are kind of holographic projection over a screen. Holography is a technique of projecting 3D images over air or on any screen creating a virtual reality environment of presence of that object.Usually 3D videos are quite large in data size then regular images. Similarly 3D motion pictures comprise of huge dataset behind it to plot the color pixel. To handle such large dataset is tedious task. Various frameworks are available to handle such large datasets. In a server system various platforms are available to handle and use such datasets quite quicker.

Cloud architecture provides flexible solution for implementing such systems and to minimize the work by implementing it by parallel manner over different computer systems. It also provides the flexibility of remote access. Instances can be created automatically and assigned to computing nodes. Horton framework does the job of making parallel processes.Horton also support hadoop to handle large data.Thus the proposed work is intended to provide solution for huge data over cloud.

Opensource Freeware Opennebula is proposed to work with software packages like FFmpeg using programming with Java and shell scripting. Combination of such architecture to optimize the working of multimedia servers by scheduling video retrivals over cloud. Improving extraction based on network parameters will boost the performance of multimedia server.

Abbreviations

3D	3 Dimensions.
ACL	Access Control List
BW	Bandwidth
DRM	Digital Rights Managment
DRS	Distributed Resource Schedular
GB	Gigabyte
GUI	Graphical User Managment
HDS	Http Dynamic Streaming
HLS	Http Live Streaming
HUD	Heads-Up Display
mp4	Motion Picture version 4
MPEG	Moving Picter Experts Group
NIC	Network Interface Card
OS	Operating System
QoS	Quality Of Service
RAM	Random Access Memory
RTP	Real Time Protocol
RTMP	Real Time Messaging Protocol
RTMFP	Real Time Media Flow Protocol
SQL	Structured Query Language

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Multimedia data refers to all data files starting from simple text to images , audio , video etc. Handling of such large files is a cumbersome task when the frequency of storing and retrieving of those data is high . Thus a proper scheduling algorithm is required to handle such data and its transmission.

1.1 Scheduling multimedia data

Scheduling of any task refers to setting the procedure of the task functions to execute in a particular sequence or pattern. Scheduling needs to be dynamic so that sequence of task can be changed in run time. Multi-media servers has functionality of broadcasting videos which are currently recording. For example live cricket match being broadcast on television or live video chat being broadcast for a conference.

1.1.1 Importance of scheduling for multi-media applications

Abundance of multi-media data is available over net to search , browse , access or store that data. Conventional techniques can handle the retrieving and broadcasting but due to advancements in recording technologies size of multimedia files is increasing day by day. Thus it become necessity to schedule the data processing with related task.

1.1.2 Challenges in today's scenario for multi-media data broadcasting

Technology evolves everyday , so do we. Thus media-servers also need to modify and upgrade themselves time-to-time. Various challenges faced for multi-media servers as

follows :

- **Data Availability** - for a server it may be possible that number of client request for same video over network thus it becomes difficult to handle the requests when video is already being used by other users at same time.
- **File format and size** - file format may differe from *.mp4 to *.avi in video *.wav to *.mp3 , playing those all files from single server requires all codecs to be installed and running over that server at same time.
- **Storage and Retrieval** - when such huge system is implemented a proper method is required to store and retrieve multi-media data from that server whenever that user requests.
- **Network configurations** - when its about multi-media it becomes default requirement to have large bandwidth as number of user are generally more at any time , requesting for data. Thus if network is not properly configured the system may hang and eventually get into failure.

1.2 Multi-media as Big-Data

Big data refers to large files containing text , image , video or raw information. It depends on two factors as size of data file and other is frequency of storing and retrieving those files. big data technologies works with lots of applications involving multi-media data. Broadcasting servers store and retrieve such video files in real time.

One of the case is of 3D videos files , which are quite large in size compared to regular video files even though there time-duration and codec is same. The difference is in GOP (group of pictures) in 3D files and 2D files and there frames.

Every video file is made by a codec which is short of compressor/decompressor. Codes are oprating system helpers to make or play or edit those video files. Without proper codecs no video file can be played or modified. Codecs are often misunderstood as containers. Codecs like H.264 are used in actually encoding and decoding the multi-media files while containers like *.MOV and *.AVI are just the way packing those multi-media data files.

Thus taking into consideration all difficulties a new algorithm for scheduling the multimedia data is proposed by taking into consideration network as well as framing parameters.

1.3 Objective Of Study

Objective of study behind this research is to understand working of different hardware with certain softwawres and clubbing there advantages to reform a whole new system for human computer interaction and making more simpler and easier for humans to interact with computer.

Market is already full of softwares available for different purposes, proper combination and usage will lead to fullfillments of all our requirments.

1.4 Thesis Organization

- Chapter 2 contains literature survey of various Holography technologies and their working.
- Chapter 3 Current trends and Challenges in current video transmission servers
- Chapter 4 contains basic architecture of cloud and its advantages
- Chapter 5 contains working of FFmpeg and modifications required for working it with 3D videos
- Chapter 6 contains Hadoop setup with Hive
- Chapter 7 contains proposed architecture with optimized algorithm
- Chapter 8 contains results of frame retrival with proposed architecture and algorithm
- Chapter 9 contains conclusion

Chapter 2

Literature Survey

2.1 Basics of Holography

Holography is science of lights , where projection of lights is made in such a way that it forms a real looking objects over screen or air.

Holography can also be described as the product of a light source scattered off objects, to be recorded and later reconstructed when the original light field is no longer present, due to the absence of the original objects.

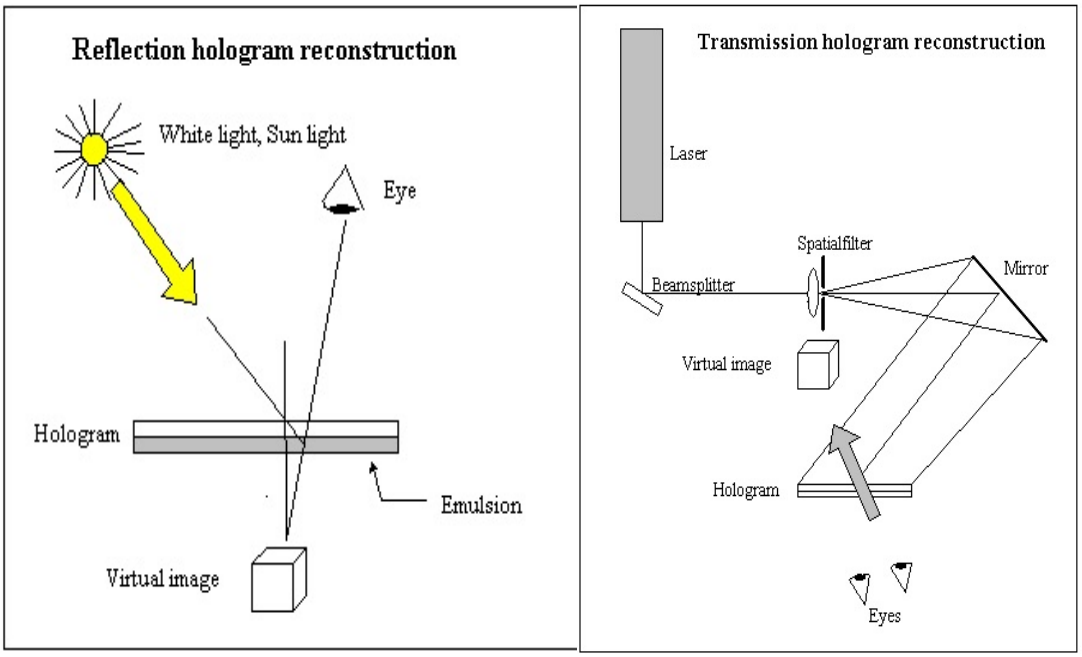
Depending upon how holographs are formed there are various types of holograms are categorized. Few of them are :

- Reflective Holograms
- Transmission Holograms
- Hybrid Holograms

Holograms are considered as 3D image of any object which forms the source image with true colors. Its like fake image formation of some real world entity.Holograms are formed by linear source of light like laser beams. Holograms are formed by BEAM splitter which divides BEAM into two identical beams. One directed to real object to be projected and other directs towards screen it has to be projected.

2.1.1 Reflective Holograms

Reflective holograms are formed by reflecting the object and reference beams are incident on the plate from opposite sides of the plate. The reconstructed object is then viewed



(a) Reflective Holograph [4] (b) Transmission Holograph [5]

Figure 2.1: Basic type of holographic projection

from the same side of the plate as that at which the re-constructing beam is incident.
For eg. forming of holograms by keeping a glass at 45 Deg angel to projecting screen like telephone and having a screen as background support after the glass or all 3D movies come into reflection holograms.

2.1.2 Transmission Holograms

Transmission Holograms are formed by keeping a laser after laser and then the viewer , in-short the object and reference beams are incident on the recording medium from the same side.
For eg. forming of scene of eagle on visa card or other debit and credit cards. This sticker is used for security purpose so that no one can forge the cards.

2.1.3 Hybrid Holograms

There are 5 types of hybrid holograms which comes somewhere between transmission and reflection holograms as they uses some common properties of them :

- Embossed Holograms : To mass produce cheap holograms for security application such as the eagle on VISA cards, a two-dimensional interference pattern is pressed

onto thin plastic foils. The original hologram is usually recorded on a photosensitive material called photo-resist. When developed, the hologram consists of grooves on the surface. A layer of nickel is deposited on this hologram and then peeled off, resulting in a metallic shim. More secondary shims can be produced from the first one. The shim is placed on a roller. Under high temperature and pressure, the shim presses (embosses) the hologram onto a roll of composite material similar to Mylar.

- Integral holograms: A transmission or reflection hologram can be made from a series of photographs (usually transparencies) of an objectwhich can be a live person, an outdoor scene, a computer graphic, or an X-ray picture. Usually, the object is scanned by a camera, thus recording many discrete views. Each view is shown on an LCD screen illuminated with laser light and is used as the object beam to record a hologram on a narrow vertical strip of holographic plate . The next view is similarly recorded on an adjacent strip, until all the views are recorded. When viewing the finished composite hologram, the left and right eyes see images from different narrow holograms; thus, a stereoscopic image is observed. Recently, video cameras have been used for the original recording, which allows images to be manipulated through the use of computer software.

- Holographic interferometry: Microscopic changes on an object can be quantitatively measured by making two exposures on a changing object. The two images interfere with each other and fringes can be seen on the object that reveal the vector displacement. In real-time holographic interferometry, the virtual image of the object is compared directly with the real object. Even invisible objects, such as heat or shock waves, can be rendered visible. There are countless engineering applications in this field of holometry.

- Multichannel holograms: With changes in the angle of the viewing light on the same hologram, completely different scenes can be observed. This concept has enormous

potential for massive computer memories.

- Computer-generated holograms: The mathematics of holography is now well understood. Essentially, there are three basic elements in holography: the light source, the hologram, and the image. If any two of the elements are predetermined, the third can be computed. For example, if we know that we have a parallel beam of light of certain wavelength and we have a double-slit system (a simple hologram), we can calculate the diffraction pattern. Also, knowing the diffraction pattern and the details of the double-slit system, we can calculate the wavelength of the light. Therefore, we can dream up any pattern we want to see. After we decide what wavelength we will use for observation, the hologram can be designed by a computer. This computer-generated holography (CGH) has become a sub-branch that is growing rapidly. For example, CGH is used to make holographic optical elements (HOE) for scanning, splitting, focusing, and, in general, controlling laser light in many optical devices such as a common CD player.

2.2 Applications Of Holography

2.2.1 Animation and Movies

Film industry is one of the most active area where new technology is been tried out over every movies to attract more and more viewers. Trends are set every year for adding more and more effects in a movie. Few of the famous movies and scene with holography are :

- **Star wars(series)** - movie series introduced in 1977 based on space and disputes among people living on other planets.
- **Total Recall** - movie of 1990 where actor uses wrist watch to project himself as 3D-hologram and use it as Decoy
- **Ocean's Twelve** - movie of 2004 where few thieves project a royal egg to depict the presence of EGG in museum while they steal the original egg

- **Iron man** - movie of 2008 where actor uses his computer system to project all screen as holograms over air and he interacts with his computer system named jarvis from those holograms
- **Avatar** - movie of 2009 where exclusive use of HUD is done to project the teritory of alien world and a big tree.
- **The Giver** - movie of 2014 where the chief talks with 3D-hologram video calling
- **The Hunger games - mocking jay part -1** - movie of 2014 where players practise there skills with holographic projection of opponents. [6]

2.2.2 Medical Science

Holography has a promising scope for medical science , where doctors can project live a beating heart for the medical students. It can help learn students faster about the organs of human body and thier working. As 3D image can plot the organ in air thus it can give clear perception to students about size , shape , colour etc. of the organ. Not onlu human organs but any animals projection can also be done. One of the leading scientist of this filed is Dr. Garner , as per his comments in newsroom website [7] holography is tommrow's future , data of different bio-organs can be taken from 3D-projection of organs and micro-processor chips can be programmed to take actions based on those data.

2.2.3 Case - studies

Person/Company	**Event**
Coca-cola	Sales conference presentation in Prague where senior directors were projected [8]
Bill gates	Made an virtual apperance at "World Congress on Information and Technology 2008" [8]
Narendra Modi	Total of 1500 locations where Narendra will project himself for election campaign [9]
One River Side Park	Setup for attracting clients to buy space at park by dislaying the area and lifestyle at the park [9]
Live Park 4d	Theme park at korea which projects hologram for visitors for entertainment [10]
Cisco	Telepresence conferences by integrating with musion technology [9]

Table 2.1: Case-studies of few famous events/places projecting Holograms

2.3 Scheduling algorithms for cloud

Cloud platforms provide inbuilt scheduling for the processes running on the system. Few of them are shown in Table 2.2

Cloud	Algorithm	Description	Advantage	Disadvantage
Open-nebula	Match-making Scheduler	Uses rank for getting free resources and expanding jobs based on available free VMs	manual ranking can be done	Master-node has to process rank everytime new job is added
Openstack	nova	Weighting mechanism applied so instances are weighted before allocating to hosts and filters are ran before weighting.	runtime calculation of weigths thus dynamic algorithms are used and optimization is achived	Lot of filters are used which gives extra load on runtime
Cloudstack	VMware DRS	Prioritazation and modules like DPM continuously monitor resources	Complete buisness package , no downtime	Not opensource
Eucalyptus	Round-robin and greedy	finding suitable nodes and allocating jobs	simple to code , easy to implement	lags user flexibility

Table 2.2: Comparision of Scheduling Algorithm

Various technologies have been introduced to improve the scheduling process of cloud systems. One of them is Blind Online Scheduling for Mobile cloud multimedia services , it serves current user by taking into account the last user request. They also introduce content recommendation system to achieve asymptotic optimality. Introduction to dynamic scheme for improving Quality of Service(QoS) and response time. They improved the response time and optimized the working of practical mobile cloud. [11]

Scheduling is also improved by introducing greedy algorithms over cloud and reduces response time and minimization of cost of resource utilization. In a paper by Xiaoming Nan[12] , they tried the simulations on Amazon's EC2 cloud structure with their optimization in Greedy algorithm for Joint workload scheduling and VM allocation. For practical aspect M/M/1 queuing model is selected with poission's ratio as incoming rate of

request. Clusters are set with corresponding weight according to there handling capacity and then jobs are divided accordingly.

2.4 Retriving and Handling data on hadoop

Hadoop is basically a distributed programming framework to execute programs in parallel. Minimum requirments of installing hadoop is Java 1.6 and higher with appropriate JDK. Two basic components of Hadoop are :

- **HDFS** - HDFS stands for Hadoop File System which basically keeps track of how files are broken in data sets and distributed over network
- **MapReduce** - it does the of job of mapping data chunks and then reducing it for processsing over cluster.

Many advance overlaying platforms are formed to work upon Hadoop for developer and user simplicity , one of them is Horton. Horton platform processes large interactive query over graphs which are now a days common for web-data like social-networking graphs [13]. In this paper they have worked upon data of codebook, which is application for maintaing information about software engineers, software components and their interactions in software project. Horton supports both directed as well as undirected graphs for query processing. Query language is processed in form of *regular language expression*.

Integrating cloud and hadoop is explained in paper by Anirban Mandal [14]. It presents working with the cloud-platforms working on Infrastructure-as-service (IAS). Open Resource Cloud Architecture(ORCA) is used to implement on-demand cluster sets for practical purpose. Eucalyptus cloud framework are used over two systems and it was connected with different inter and intranet connections. Three test cases were taken with varying network speed of 10 Mbps , 100 Mbps and 1000 Mbps and common request was sent to form clusters with 1 master node and 7 slave nodes each time. Experimentation showed multi-cloud environments can perform better to handle clusters.

2.5 Video transmission techniques

Video transmission over network is already a hot topic for researchers. Scenario where 3D videos are to be transmitted is even more challenging. A paper by Hong-Hsu Yen [15] shows how routes can be bifurcated to form scheme for seprating and then recombining the

packets over network. They proposed IV/IPower algorithm to improvise the transmission power to have largest ratio of the incremental received video at the receivers. They used shortest path algorithm with weight assigment with identification of minimum power routing. Numerial result shows a substantial growth of around 10

Chapter 3

Transmission Servers

3.1 Current trends and Challenges in current video transmission servers

Current video streaming servers work upon combinations of following components

- Distributed File Storage
- Edge servers
- DRM systems
- Client GUI

All the components have their own importance and required to provide seamless functionalities to user

3.1.1 Distributed File Storage

Videos are stored in specific structured manner to provide video delivery from storage whenever required within shortest time of request with proper retrieval technique.

The basic reason behind using such specific structure is providing fail-safe mechanism so that at time of system failure recovery should be easy and copies are saved at different regions so that transmission time can be reduced while broadcasting the videos.

3.1.2 EDGE servers

Edge server performs various functionalities depending upon the requirements configured by server Administrator.As origin servers are only used at time of failovers because they

are designed for providing content available 24*7 , but many times due to heavy traffic over network it is certain at some point the server will break and stop its services of video broadcasting. [16]

3.1.3 DRM systems

DRM stands for Digital Rights Management , it allows content providers to control the distribution and usage of content over Internet. It is the system which scrambles the video broadcasting signals and it may also encrypt the video and then broadcast it , according to system requirement. It supports all video delivery protocols from HLS , MPEGDASH and RTMP

3.1.4 Client GUI

It basically refers to the video players working over client machine for video decryption and running those video fragments sent by server. Almost all player supports DRM standards but there may be flows while working with older version and cross-platform players. In some cases it requires custom players like JWplayer, FLOWplayer etc. while going to specific propritory devices like AppleTV , android-box etc.

3.1.5 Present scenario

Presently two options are most widely used for broadcasting the videos over internet

- **RTMP/RTMFP** - which is propritory of Adobe Inc. It does not have any limitations over connected clients and can be used over both p2p networks as well as NAT traversals.
- **HTTP-based protocols** - it includes HLS/HDS to MPEG-DASH , easy scaling for CDNs and m3u8. Live and videoondemand supported. May produce delay from 2 sec or longer in real-time broadcasting.

3.2 Challenges in current scenario

- **Difference of formats** - on todays date this challenge is still growing as ever as for example BLU-Ray vs HDDVD both provide high quality videos to viewer. RTP/RTSP was introduced for sole purpose of broadcasting videos without any delay. Later with introduction of RTMP , standard of video foramating started changing due to use of propriotory protocols

- **Difference of terminal devices** - now-a-days video streaming capable devices come in vast variety of models starting from smartphones , laptops , tablets , smart TVs , embedded devices , desktops etc. All of this devices are manufactured with their own softwares with supoorting codecs and media formats. Unfortunately there is no uniformely supported videos solution

- **Low priority for 3D videos** - 3D videos are not retireved unless exclusively asked for. This challege is due to low availability of 3D- Data.

Chapter 4

Cloud Computing

4.1 Present scenario for cloud architectures

4.1.1 Microsoft Azure

Microsoft has developed cloud service for its users which provides functionalites of developing Web-apps using there platform on demand, for creating and deploying virtual machines with any windows or Linux OS , for deploying and managing SQL database services , platform for building mobile apps and many more.

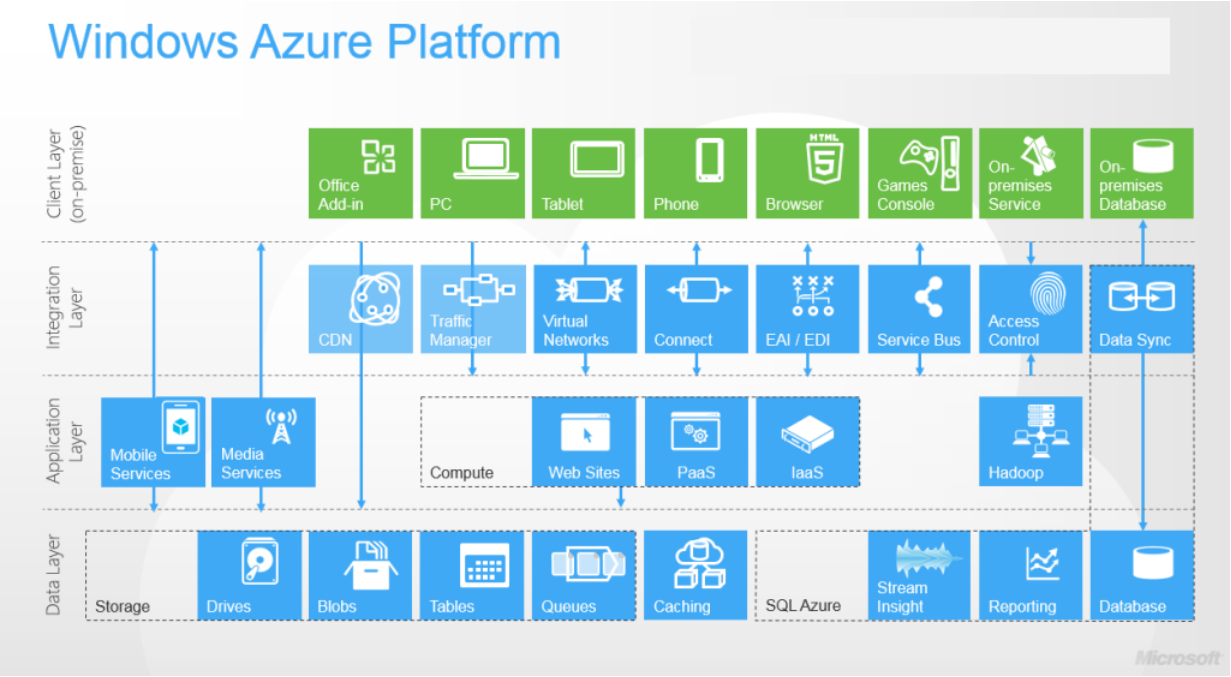


Figure 4.1: Azure Architecture [1]

Windows server also provide base for presentation servers of Citrix which helps clonning of servers and databases in rapid manner. [17] Microsoft SQL server provides base architecture for storing and retriving videos. It may prove beneficial to use microsoft resources but as we all know it requires initial investment or good financial support.

4.1.2 Openstack

Openstack is cloud operating system that controls large pool of computer systems, datacenters , network resources etc. through a dashboard accessed through a web-interface. Openstack also provides functionalities of creting clusters over network. Openstack can be installed on any opensource server OS.

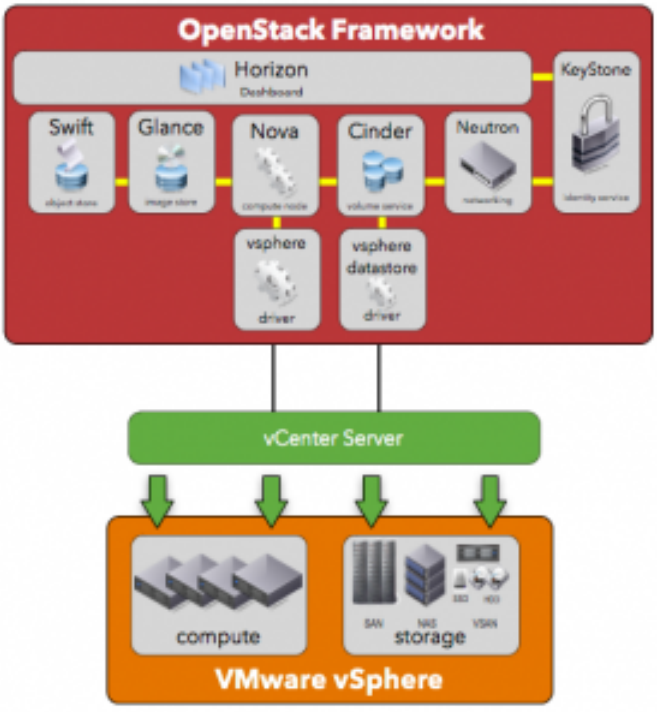


Figure 4.2: OpenStack Architecture [2]

Openstack also provides API's for controlling database, network resources , Image services etc. The problem faced for using Openstack is that it requires minimal of 8 GB RAM with 1 GB of NIC card for basic setup.

4.1.3 OpenNebula

Opennebula is also opensource cloud platform for deploying virtual machines.It is too flexible , simple and robust. The project architecture consist of opennebula architecture. It provides virtual datacenter management to integrate with exsisting data-centers.

Opennebula supports both KVM and XEN hypervisors according to requirments. It can be deployed upon Centos as base OS or with Ubuntu. Thus it was compatible for research purpose from all prespective.

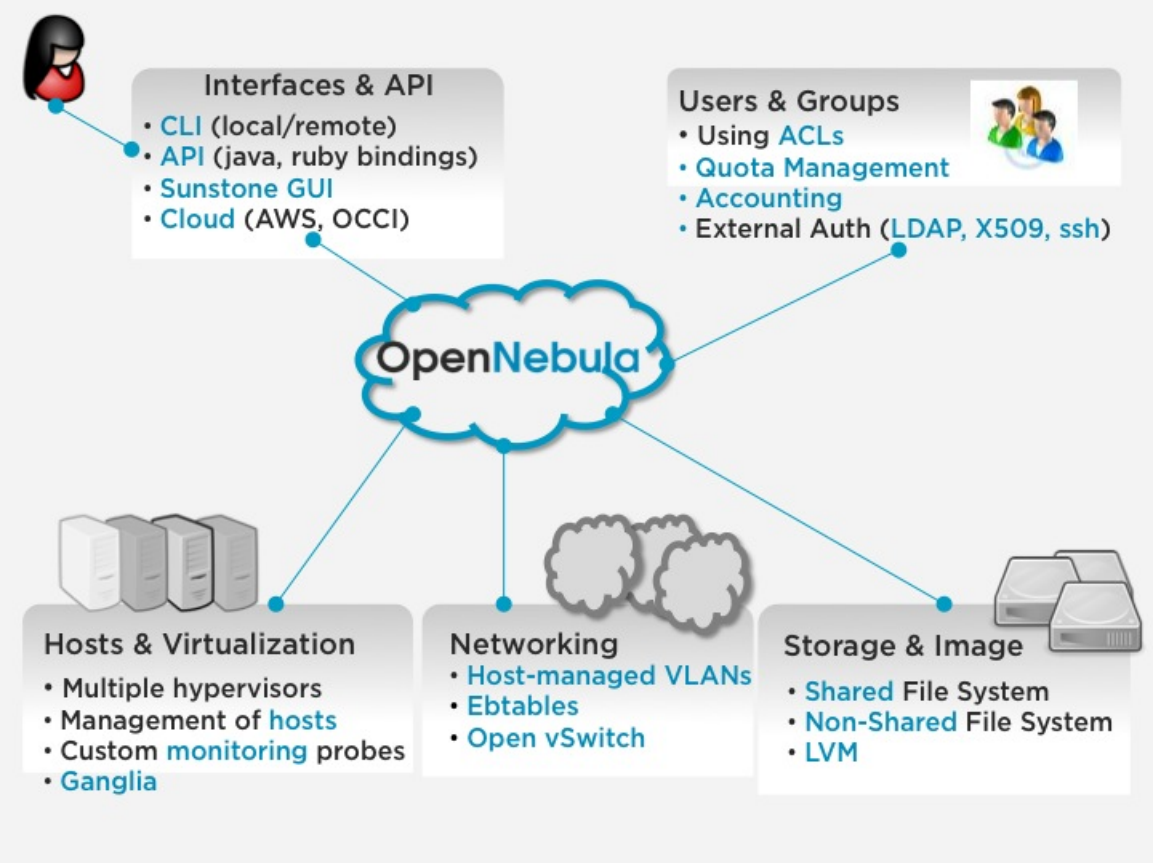


Figure 4.3: Opennebula Architecture [3]

Following are the main components of opennebula cloud with their working

- **Networking** - it provides customization for administrator to integrate it with other specific networks like datacenters , VLANs and internet.
- **Interfaces and API's** - Two basic interfaces are command line interface and Sunstone GUI over web-browsers. Most of the administrators work with sunstone GUI as it very difficult to rememeber all the commands for command line interface. Both GUI and command line can harness the control of opennebula cloud easily.
- **Hosts** - Xen , KVM and VMWare hypervisor are supported for creating hosts over opennebula , which provides great flexibility for virtualization manager for monitoring.

- **Clusters** - opennebula provides unique feature by supporting the deployment of clusters over pool of hosts
- **Users and Groups** - admin can create users and groups with proper rights. Access is controled by Access Control List(ACL) for all users and groups to finely control the access of services of the cloud components.

Chapter 5

Working of FFmpeg and 3D videos

5.1 Introduction to FFMpeg

FFmpeg project tries to provide the technically possible solution for developers of video related applications and end users alike. To achieve this they use open-source libraries and code through them to take advantage of already built fuctions.

5.1.1 Functionalities of FFmpeg

Following are the functionalities of FFMpeg

- **Convert** - it provides functionalities of converting both audios and videos from one format to another for eg. from mp4 to avi.
- **Streaming** - it provides streaming services for both audio and video over network. It can stream live audio and video also option of broadcasting recorded file is also available. FFserver accepts POST request from ffmpeg and acquires stream for publishing it over RTP/RTSP.
- **Player** - it plays videos and auidos of all codecs. It can also be used to filter colors , adjust color level , etc. It can also dump the reading into a specified file as output.
- **Probe** - ffprobe gets the information from audio video files and presents the information into machine-human readable format. It can show packets , frames , streams , programs , chapters etc. from a file.

It contains libavcodec, libavutil, libavformat, libavfilter, libavdevice, libswscale and libswresample which can be used by applications. As well as ffmpeg, ffserver, ffplay and

ffprobe which can be used by end users for transcoding, streaming and playing.

5.2 FFmpeg Libraries

- libavutil - It is a library containing functions for simplifying programming, including random number generators, data structures, mathematics routines, core multimedia utilities, and much more.
- libavcodec - It is a library containing decoders and encoders for audio/video codecs.
- libavformat - It is a library containing demuxers and muxers for multimedia container formats.
- libavdevice - It is a library containing input and output devices for grabbing from and rendering to many common multimedia input/output software frameworks, including Video4Linux, Video4Linux2, VfW, and ALSA.
- libavfilter - It is a library containing media filters.
- libswscale - It is a library performing highly optimized image scaling and color space/pixel format conversion operations.
- libswresample - It is a library performing highly optimized audio resampling, rematrixing and sample format conversion operations.

5.3 FFmpeg Image Extraction

FFmpeg takes Group Of Pictures(GOP) of Video Clip as input and will give output as image file as frame extracted from 3D video clip.

Chapter 6

Hadoop with Hive

6.1 Hadoop introduction

Hadoop system is originated from a paper published by google employees on how they are managing there huge data. Presently Hadoop is owned by Apache software foundations. Hadoop runs on any server system which requires handling of huge data over a network. The main two components of hadoop are HDFS (Hadoop distributed file system) and Map Reduce

6.1.1 Hadoop Distributed File System (HDFS)

File system deployed over a cluster of computers are nodes running in parallel for a single task. A job file is submitted by client to master node by client . Data is processed parallel over cluster of computers in a reliable , fault-tolerant manner.

HDFS can be deployed on low cost hard-wares with high throughput access to application data. HDFS consist of client-server model as Master-slave. Master node regulates the file system namespace and access to files by clients. Data is given in the form of file by user , this data is been split and stored upon datanodes.

Datanodes send heartbeat signal to Namenode to notify that they are alive and connected to network. It prevents failure of processing while distributing the workload.HDFS can be accessed through many different ways like API provided by HDFS in Java , wrapper in C language is also available and most common is web-browsers using WebDAV protocol.

6.1.2 MapReduce

MapReduce is key component for splitting work into parallel parts. It splits the input data into independent chunks . Output of Map is used as input for Reduce and task is split upon datanodes for processing. Components of MapReduce are JobTracker and TaskTracker.

JobTracker manages and tracks the submission of jobs over nodes. Functions performed by JobTracker are

- Checking the input and output specifications of the job.
- Computing the Input , split values for the job.
- Setting up the requisite accounting information for the Distributed Cache of the job, if necessary.
- Copying the job's jar and configuration to the MapReduce system directory on the File System.
- Submitting the job to the JobTracker and optionally monitoring it's status.

TaskTracker creates a local library which can be used as cache for data. TaskTracker also logs the stdout which is standard output and errors into logs. TaskTracker also provides credentials for job to nodes so that proper data fetching can be performed.

6.2 Hortonworks Framework

Hortonworks provide a framework including hadoop with HIVE and PIG as additional packages. Hortonworks have developed Hortonworks Data Platform(HDP) which provides singular platform for implementation of hadoop , Hive and Pig.

Hortonworks is basically a organization which provides solution for building modern database systems with reduced costs. It also supports various data analysis apps for data analyst.

Hive explores the tables of big-data through hadoop. Data units are stored in taxonomic way from larger to granular units. Data is retrieved through simple query language which is similar to SQL. It is developed such way that database managers working with

SQL can easily shift to HDL i.e. Hive query Language. It provides GUI over web browser so that anyone can remotely access the data with proper credential.

Pig was designed to perform below three types of job :

1. Extract , Transform and Load - data pipeline
2. Research on raw data
3. Iterative Data processing

Pig Latin can run scripts which are also compatible with other scripting language like java,ruby,python etc.

Chapter 7

Proposed Architecture and scheduling algorithm

7.1 System Architecture

A simple i3 processor of 2.30 GHz clock rate is used with 4 GB DDR3 ram of 660 MHz. System also contains a GPU of NVIDIA GeForce 410M of size 512 MB. This was the hardware configuration and for software , CentOs 6.6 is used as base OS. For framework setup OpenNebula v4.10 cloud is configured with sunstone as frontend. It will be working on port 9869 by default. FFMpeg is comipled and installed on base os so that its functionality can be used by other resources.

7.2 Proposed Scheduling Algorithm

In proposed algorithm various parameters like BW which is to signify Bandwidth for connected client to server. Bandwidth varies by different factors like connection type , user device , user location, natural noise etc. BW_0 indicates the initial bandwidth of the connection.

SVC (scalable video coding) is used which has basic two layers as BL - base layer and EL - enhancement layer. Now in most case there will only one base layer and there may be multiple enhancement layer according to requirement of system or architecture. Here BL_i indicates SVC segment of BL.

Certain parameters obtained from user as P which is packet loss rate , RTT as round-trip-time of network , SINR as signal to noise interference ratio. This information is

Algorithm 1 : *multimedia scheduling*$(R_{bl}, BL_0, BW_0^{practical}, T_{win})$

Initialize ***Parameters[]***
i=0;
$BW_0 = R_{BL}$
Transmit BL_0
Monitor $BW_0^{practical}$
repeat
Sleep for T_{win}
$Obtain\, P_i, RTT_i, SINR_i from client$
Predict $BW_{i+1}^{estimate}$(or $BW_{i+1}^{estimate} = BW_i^{practical}$)
k=0
$BW_{el} = 0$
repeat
extract frame (I-frames)
k++
$if k \geq j \quad break$
$BW_{EL} = BW_{EL} + R_{EL}$
until $BW_{EL} \geq BW_{i+1}^{estimate} - R_{BL}$
Transmit BL_{i+1} and $EL_{i+1}^1, EL_{i+1}^2, EL_{i+1}^3, ..., EL_{i+1}^k - 1$
Monitor $BW_{i+1}^{practical}$
i++
Until *all video segments are transmitted*

obtained by server when a client connects to it. This are network parameters which vary according to atmosphere , distance , noise and based on other factors.

From this algorithm a change in bandwidth of user according to his connection type and change in frame structure to provide a seem less quality video to user. Frames will be extracted as per device type and connection speed. Prediction on upcoming bandwidth of system from history that is being going on with current communication which practically being going on by equation as below

$$BW_{i+1}^{estimate} = BW_i^{practical}.[\alpha.f(p_i, p_{i-1}) + \beta.g(RTT_i, RTT_{i-1}) + \gamma.h(SINR_i, SINR)] \tag{7.1}$$

Here pre-assumption is taken of $\alpha + \beta + \gamma = 1$. As they are practical constants which varies into there limits making there sum as 1. The functions of f() , g() , h() are reflecting the value change of each factor compared with that of the last time window.

Chapter 8

Experiments and Results

8.1 Experimental Setup

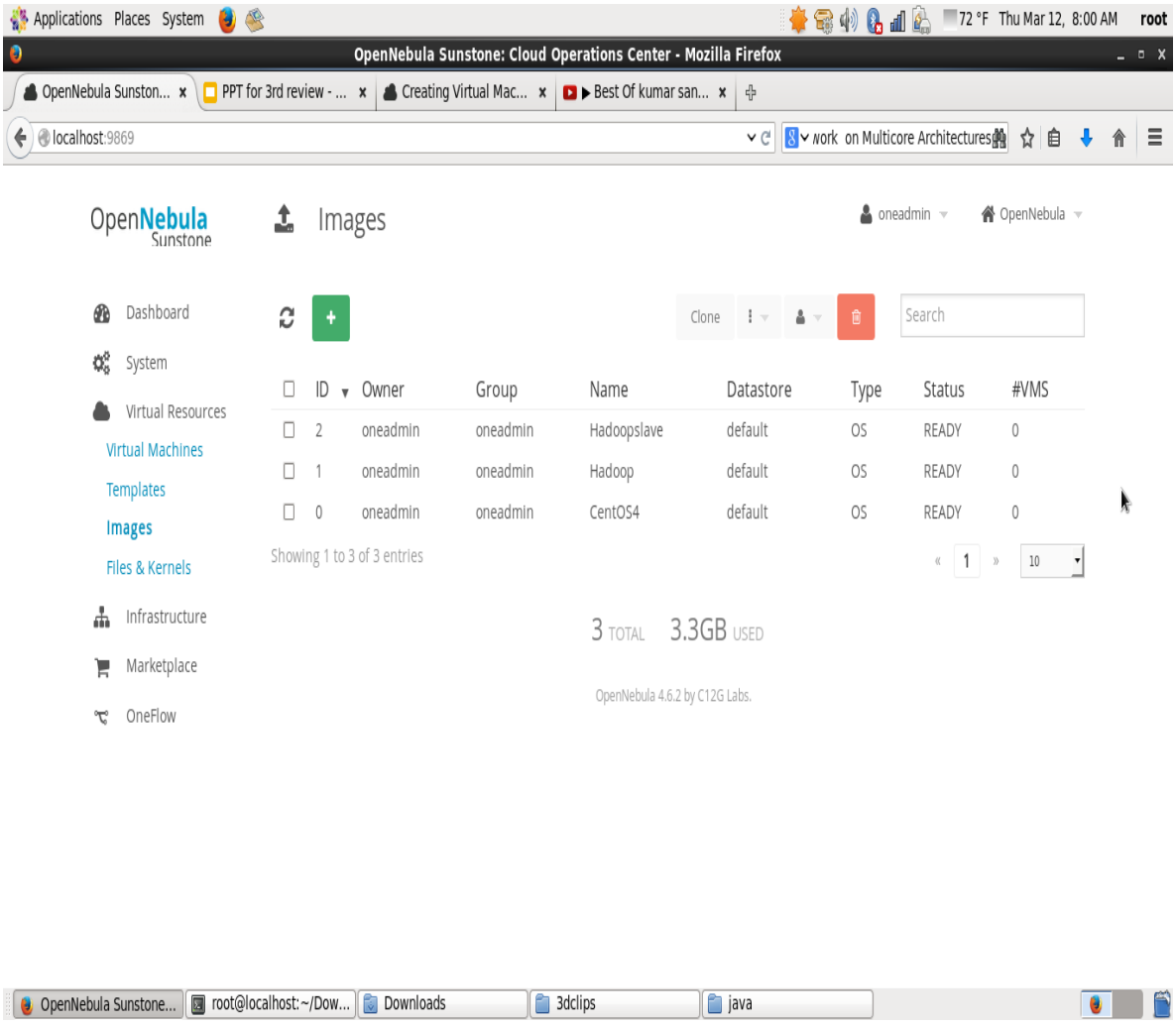


Figure 8.1: OS images Stored over OpenNebula

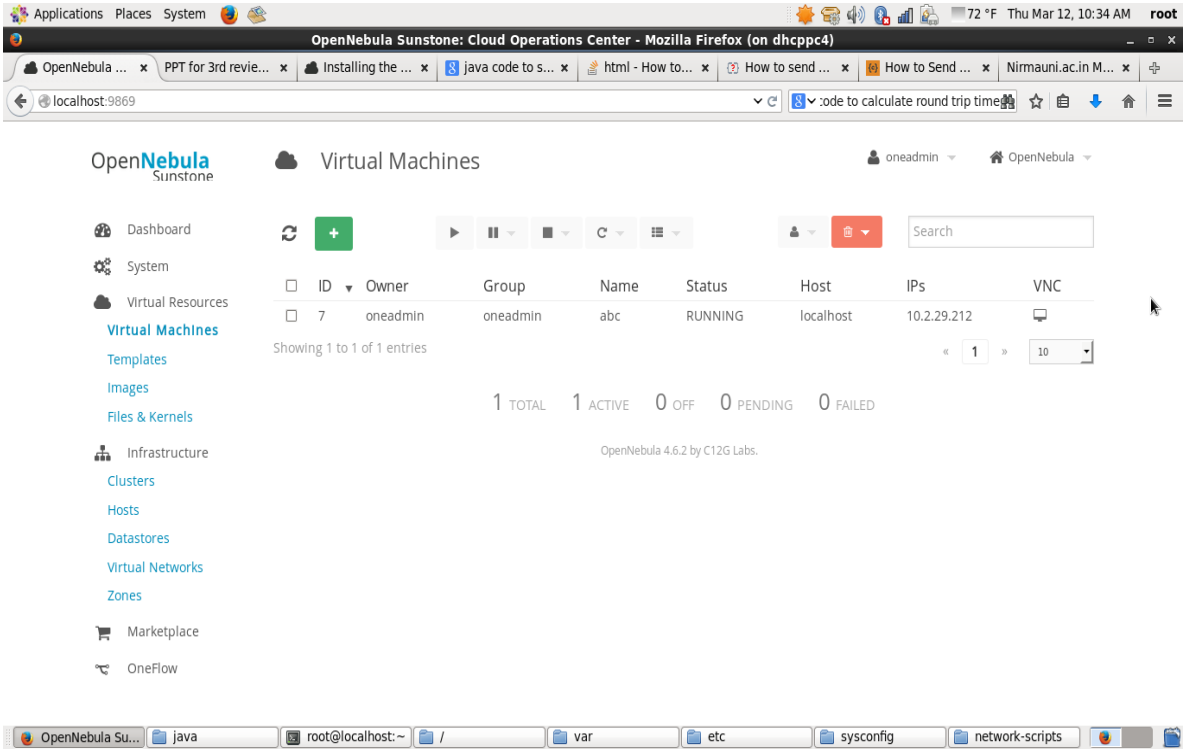


Figure 8.2: Virtual Host running over Cloud

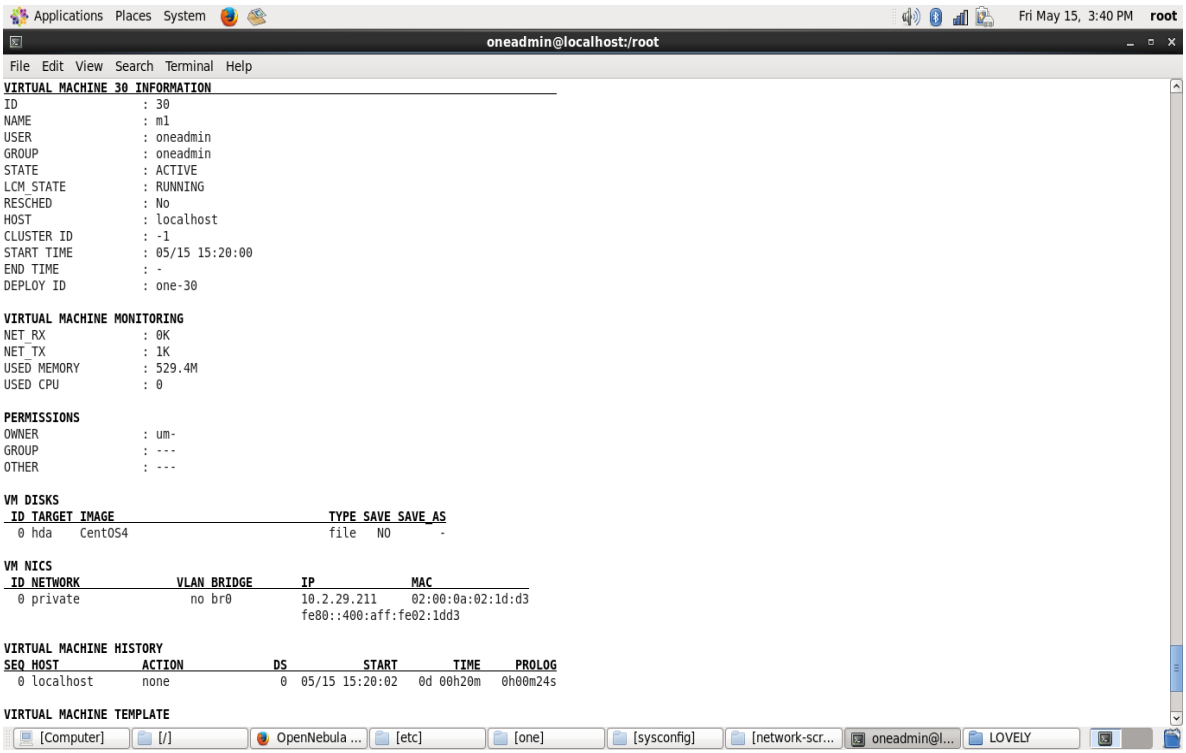


Figure 8.3: Host Information Through Command Prompt

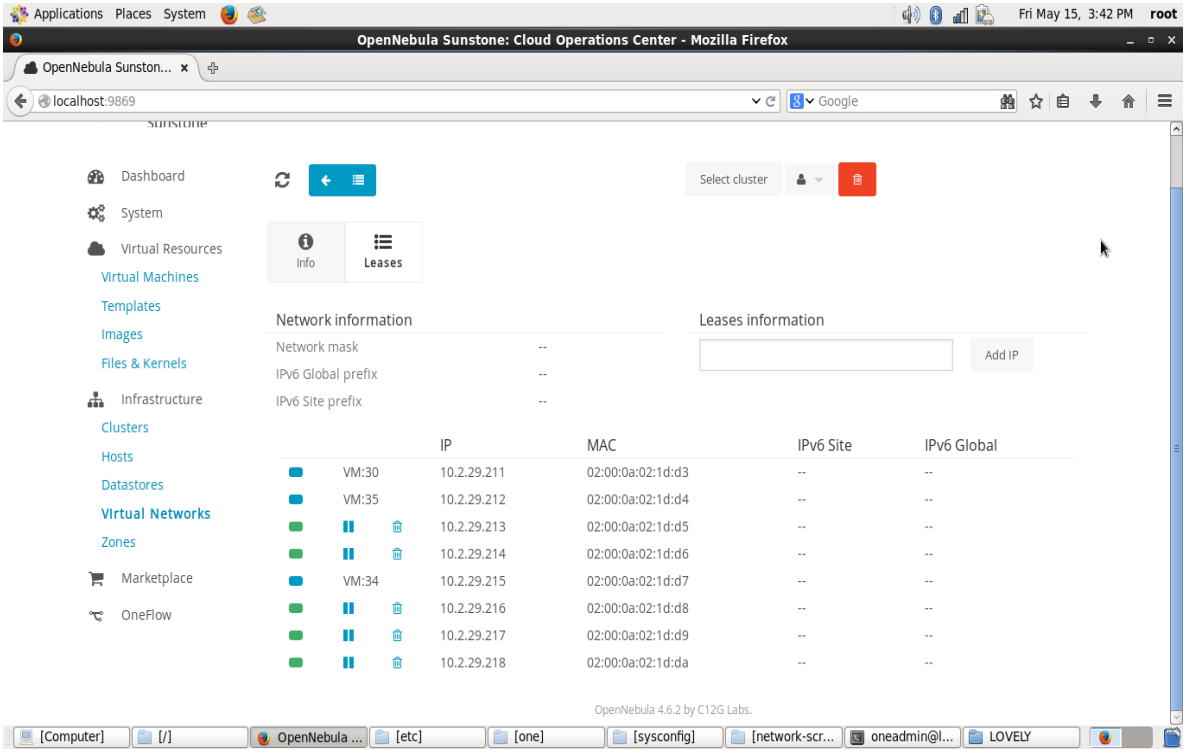


Figure 8.4: Network Pool of IPs

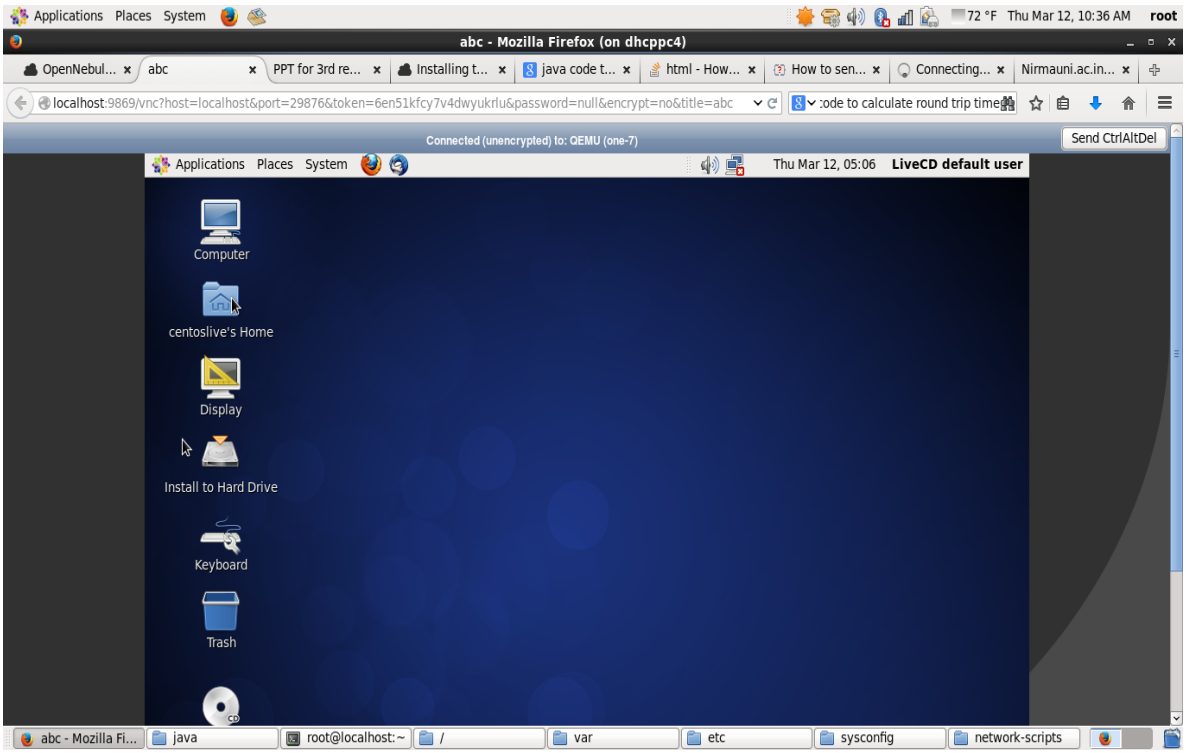


Figure 8.5: Image of Cent Running on Virtual Machine

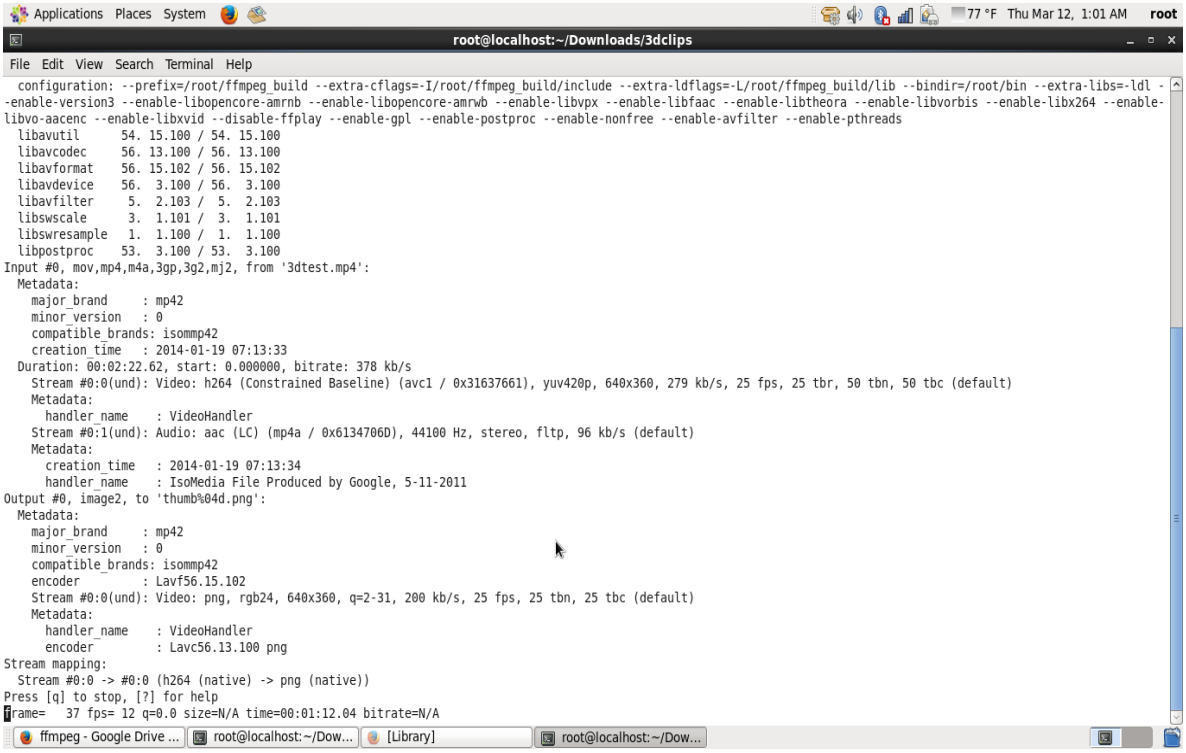


Figure 8.6: FFMpeg Extracting I-Frames from 3D-clips

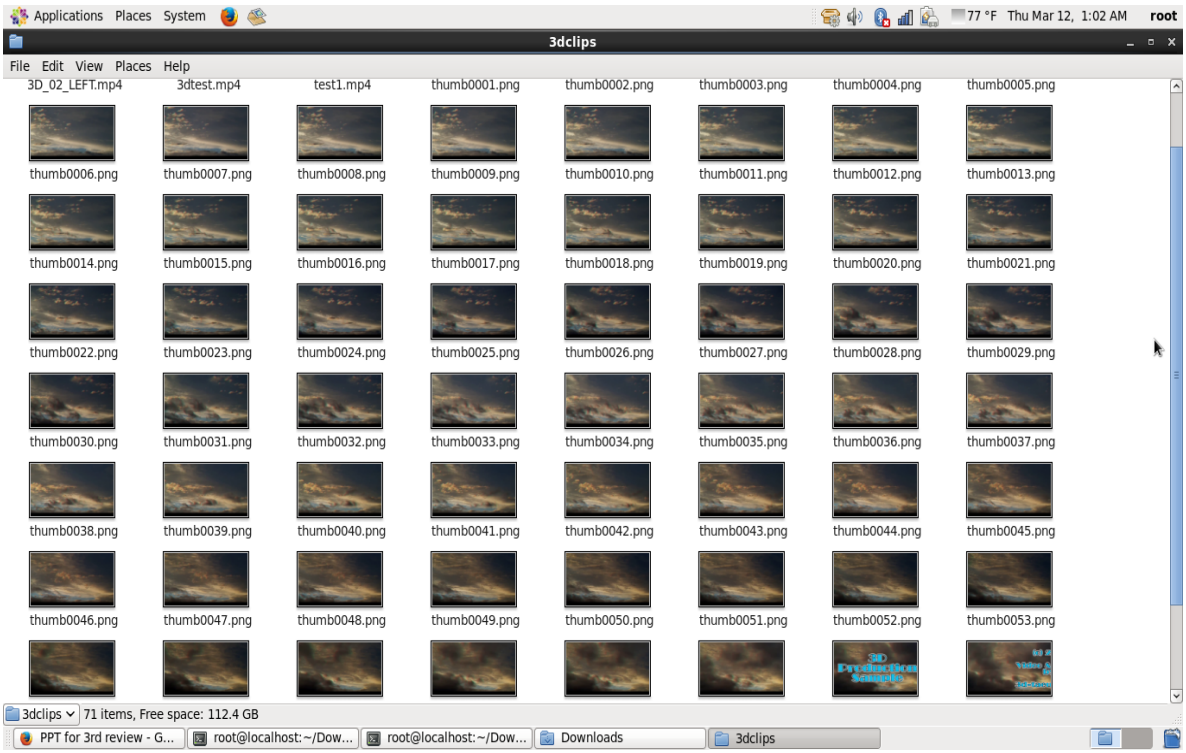


Figure 8.7: FFMpeg Extracted Images

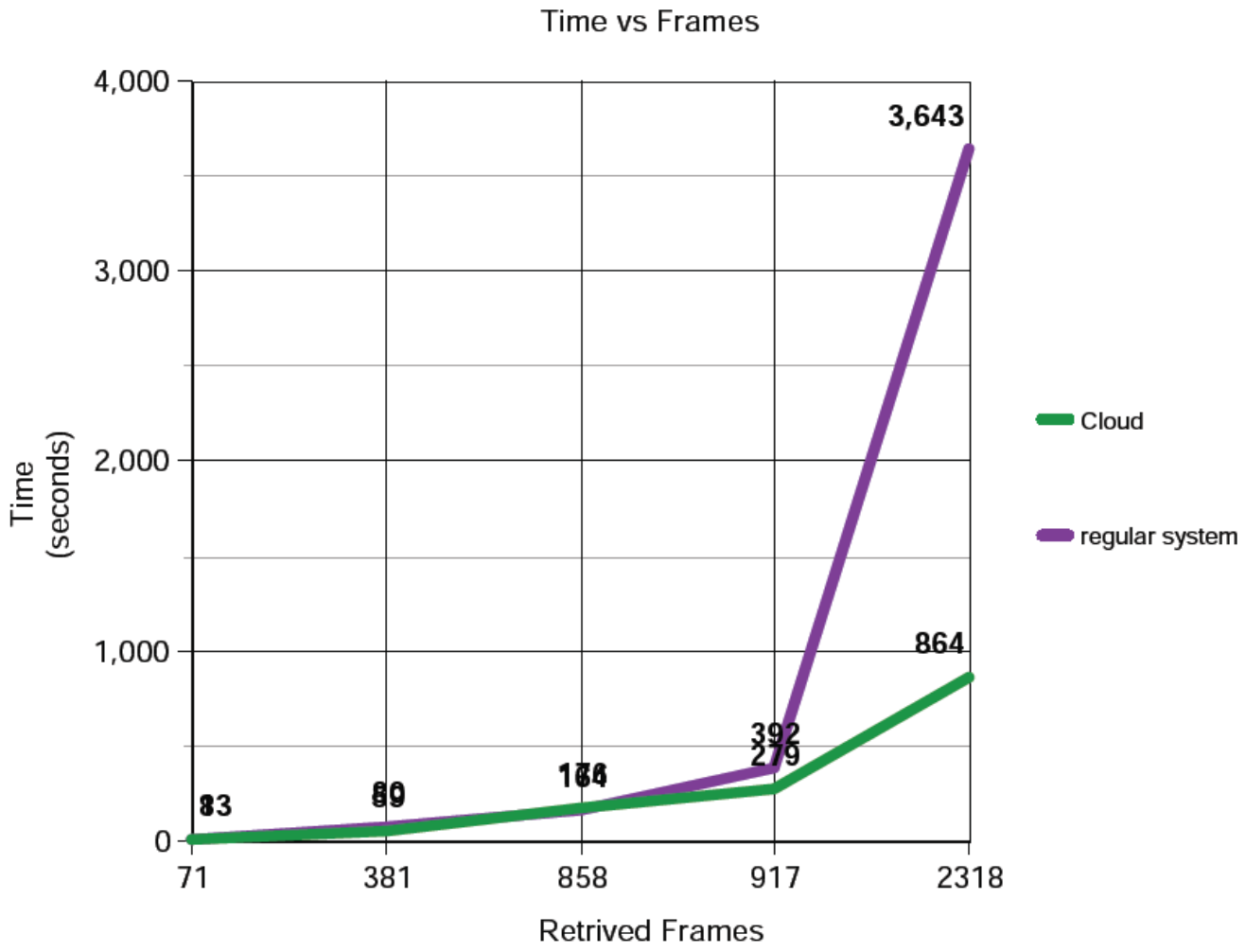


Figure 8.8: Output Graph with comparision to regular Systems

The Graph 8.8 shows difference in retrival times of cloud system and regular machine. To optimize more of results more number of systems can be included in cloud system and then retrival can be done. In this practical cloud and database was implemented on same system for simplicity but in future seprate database server can be configured to improve the results in more better way.

Chapter 9

Summary and Conclusion

9.1 Summary

In filed of video transmission we have tried to improve broadcast services by improving retrival of videos and making a dynamic system for broadcast based on network parameters and frame extraction of particular video. Cloud computing is used hadoop to improvise the performance. FFmpeg is used for retrieval of frames. And transmission is done based on number of I-frames extracted and accordigly setting the bandwidth of transmission.

9.2 Conclusion

This report has showed the improved results of retrival timining and broadcast QoS can be improved. Use of cloud with parallel computing platform like hadoop can improvise the overall performance of streaming servers. In most cases till now research is done on either extractions policies or either improving network performances , here focus on both is done and scheduling is based on both parameters. The speedup achived is quite significant. Mergining various platforms is cumbersome but after proper setups working of structure is much more efficient then regular servers.

References

[1] "Microsoft Azure Architecture." http://www.notsotrivial.net/. [Online; accessed 2-may-2015].

[2] "Openstack cloud Architecture." http://cloudarchitectmusings.com/. [Online; accessed 2-may-2015].

[3] "Opennebula cloud Architecture." http://archives.opennebula.org/. [Online; accessed 2-may-2015].

[4] K. E. Olsen, "Reflective Holography." http://www.fou.uib.no/fd/1996/h/404001/kap02.htm. [Online; accessed 8-may-2015].

[5] K. E. Olsen, "Transmission Holography." http://www.fou.uib.no/fd/1996/h/404001/kap02.htm. [Online; accessed 8-may-2015].

[6] Wikipedia, "Holography in fiction — wikipedia, the free encyclopedia," 2015. [Online; accessed 14-May-2015].

[7] A. Siegfried, "Medical applications of holography," 2014. [Online; accessed 14-May-2015].

[8] A. Elmorshidy, "Holographic projection technology: The world is changing," *arXiv preprint arXiv:1006.0846*, 2010.

[9] B. Narayan, "Modis modus operandi in the 2014 elections," *Economic and Political Weekly*, vol. 49, no. 20, pp. 12–14, 2014.

[10] "Theme park with holography," 2015. [Online; accessed 14-May-2015].

[11] L. Zhou, Z. Yang, J. J. Rodrigues, and M. Guizani, "Exploring blind online scheduling for mobile cloud multimedia services," *Wireless Communications, IEEE*, vol. 20, no. 3, pp. 54–61, 2013.

[12] X. Nan, Y. He, and L. Guan, "Optimization of workload scheduling for multimedia cloud computing," in *Circuits and Systems (ISCAS), 2013 IEEE International Symposium on*, pp. 2872–2875, IEEE, 2013.

[13] M. Sarwat, S. Elnikety, Y. He, and G. Kliot, "Horton: Online query execution engine for large distributed graphs," in *Data Engineering (ICDE), 2012 IEEE 28th International Conference on*, pp. 1289–1292, IEEE, 2012.

[14] A. Mandal, Y. Xin, I. Baldine, P. Ruth, C. Heerman, J. Chase, V. Orlikowski, and A. Yumerefendi, "Provisioning and evaluating multi-domain networked clouds for hadoop-based applications," in *Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference on*, pp. 690–697, Nov 2011.

[15] H.-H. Yen and F. Yap, "Power efficient routing algorithm for 3d video transmission in bandwidth-limited and radio-limited wireless networks," in *Ubiquitous and Future Networks (ICUFN), 2012 Fourth International Conference on*, pp. 411–416, July 2012.

[16] C.-T. Yen, H.-T. Liaw, N.-W. Lo, T.-C. Liu, and J. Stu, "Transparent digital rights management system with superdistribution," in *Broadband, Wireless Computing, Communication and Applications (BWCCA), 2010 International Conference on*, pp. 435–440, Nov 2010.

[17] "Citrix presentation server." http://support.citrix.com/article/CTX107406. [Online; accessed 3-march-2015].