Understanding The Concepts of Cloud Computing under Amazon Web Services Domain at Cognizant Digital Solutions

Project report submitted in partial fulfillment of the requirement for the Degree

of

Bachelor of Technology

In

Electronics and Communication Engineering

By

Stakshi Sharma (161084)

Under the supervision of

Mr. Karthick Selvam

Jaypee University of Information Technology, Waknaghat,

Solan-173234, Himachal Pradesh

May- 2020

Table of Contents

DECLARATION

This is to certify that the work titled "Understanding The Concepts of Cloud Computing Under Amazon Web Services Domain at Cognizant Digital Solutions", submitted by Stakshi Sharma under the pafrtial fulfillment for the award of Bachelor of Technology in Electronics and Communication of Jaypee University of Information Technology has been carried out under my supervision. This work has not been submitted partially or wholly to any other university or institute for the award of this or any other degree or diploma.

Signature of Supervisor

Mr. Karthick Selvam

Cognizant Technical Solutions

ACKNOWLEDGEMENT

List of Figures

ABSTRACT

There is no particulars capability quialification to learn Clouds Computing. The appllicants ought too ideally to be from an IT or PC related foundations with the goal that he/she have the general informations about PCs and programming. Information on distributeds computing essentials or any pertinent experience could be an exxtraordinary extraordinary to find a newly lines of works. Distributed computings contains a fews jobs. These could be identified with the executives, IT frameworks, end clients supported, appllication improvement, buisiness examination, systems, security and webbed advancement. Each job requires explicit aptitudes the know. Here's the gathering of the considerables number of aptitudes to know the required to investigate the universes of Clouds Computing. Cloud Computing this technology is wellcomed by many origanisation in India and they have even hiring people for the role of Cloud Support Engineer, Cloud computing Engineer, Cloud Solution Engineer, Cloud Administrator/ Architect, etc. But what I perisonally came to experience in this is that few origanisation doesn't allow feshers for Cloud related jobs they expect experienced candidates. And the reason for the same can be the teaching patterns, we are not taught what is required in the present new newly world and thus we are the reicruiters moves that are towards experience holder because they don't know what to want to spend much time and money in the training of candidates itself, where they can easily be getting their work done by experienced candidates. But, Everyone, was a fresher first, even the core industry now, where once just a startup. A fresher understands the pain of a fresher and understands it on how to do it. The world is changing and thus, there are few origanisation which are ready to test the feshers capability now-a-days so, don't lose hope. Learn the concepts with practical knowledge, have some programming as well as scripting knowledge (PowerShell, python) and have your basics clear. Do such courses from institutions which fulfils the industry demands (Certifications will be an add-on)? mCloud Computing is a fast growing technology and will hopefully bring opportunities for feshers.

CHAPTER 1

Introduction

The report is about the completion of the concepts such as MySQL, java etc. in order to understand the ultimate goal of learning cloud computing under the AWS domain. The first chapter is about the introduction of MySQL. It is basically defined which is used to communicates with the daatabases. Under this chapter I have covered all the important concepts that are required later in cloud computing. Like for instance, how to create a table, how to delete a table, how to insert values. The concept of primary key and the foreign key is also described in the below chapter. The concepts of joining multiple tables, what are the roles of different queries and aggregate functions such as count, sum etc. The important concepts of subqueries and normaliszation has also been covered. The chapter has also covered the the easy steps for a beginner to learn how to install the platform to practice the MySQL queries. In the next there is basic information about what is Hadoop and what is it history and why is it important in today's world. Apache Hadoop is the structure for disseminated capacity and figuring (mix of Processor and RAM) model to manage huge volume of information. Then after that there is HDFS which is the Hadoop distributed file system. Then in HDFS what is the architecture, what is it used for. As a result of which in the certain confinements that will be beniefit on every single record framework by methods for least amounting and most extreme volume it will support, least and greatest document size it will bolster, stage conditions will be there. The chapter also shows the concepts of big data, what is problem that is solved and what are the companies that are helping in managing such a huge amount of data.

Apache Hive is an information distribution centere programming ventture based on Apache Hadoop for giving information inquiry and exxamination. Hive gives a SQL-like interface to question information put always in different daatabases and record frameworks that coordinate with Hadoop. Then the last chapter is all about the cloud computing, what are the tools required and what is AWS and what are the different types of services that they provide to their users.

CHAPTER 2

Introduction to MySQL

SQL stands for Structured Query Language. It is a standard language which is widely used to communicate with the databases.

SQL is used to perform tasks on database. It allows to interact with the data.
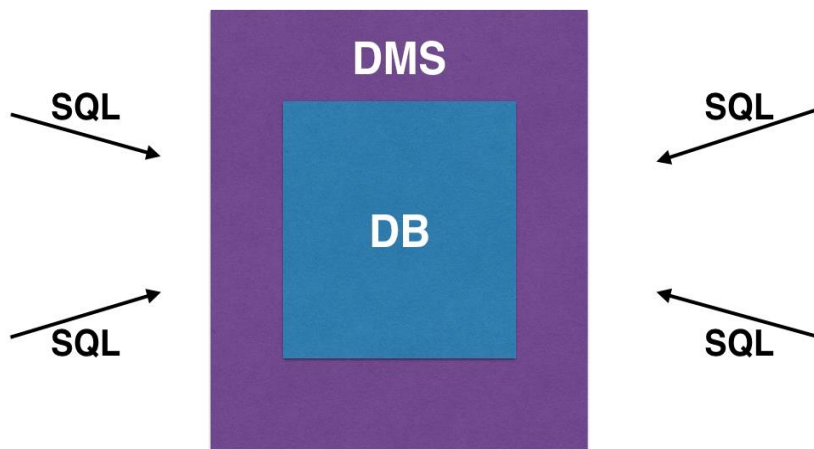
Figure 2.1: MySQL

2.1 What is MySQL

MySQL is a relational database management system. It provides a UI for uss to interacts with the database.

Figure 2.2: RDBMS

2.2 What is RDBMS

A relational database is a collection of data organized into tables. Tables contain columns of data categories and rows with particular instances of that data category. For example,

id	first_name	last_name	gender
1	Chris	Martin	M
2	Emma	Law	F
3	Mark	Watkins	M
4	Daniel	Williams	M
5	Sarah	Taylor	F

Figure:2.4

Tables in a relational database can be linked together. RDBMS is what we use to access and interact with the relational database.

id	first_name	last_name	gender
1	Chris	Martin	M
2	Emma	Law	F
3	Mark	Watkins	M
4	Daniel	Williams	M
5	Sarah	Taylor	F

customer_id	order_time
2	2017-01-01 08:05:16
12	2017-01-01 08:44:34
4	2017-01-01 09:20:02
9	2017-01-01 11:51:56
22	2017-01-01 13:07:10

Figure 2.5: Tables Example

2.3 Installation of MySQL

Step1:

The first step is to go to the website www.dev.sql.com and install MySQL for the required operating system.
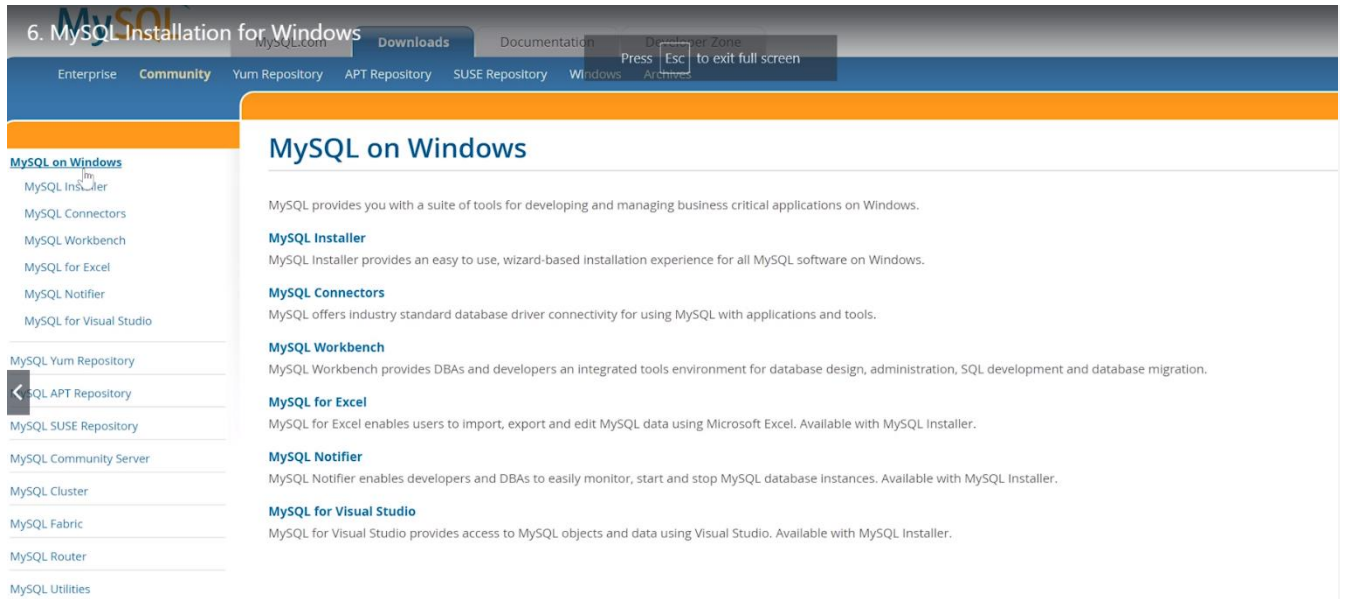
Figure 2.6: MySQL installation

Step2:

After installing the MySQL to your system, start setting up the workbench platform where you can write the queries.
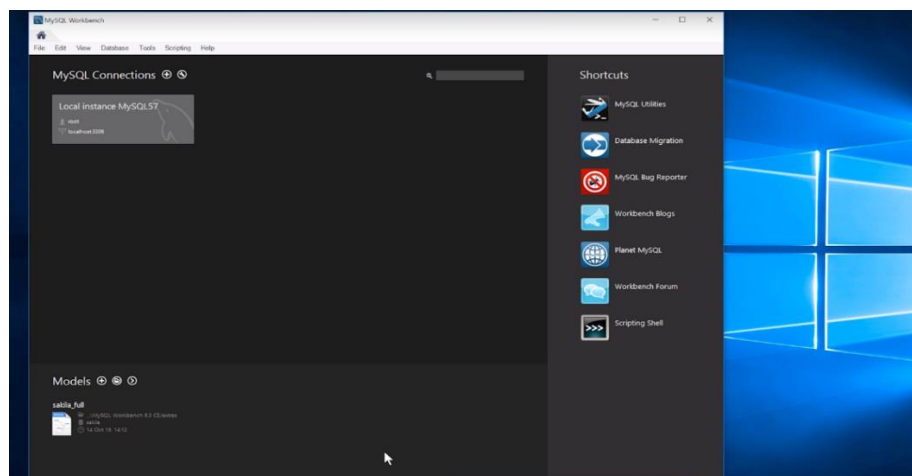
Figure 2.7: setting up the workbench

Step3:

When you open the workbench, a password prompt will pop up and you need to enter the root password that you created during the installation process.
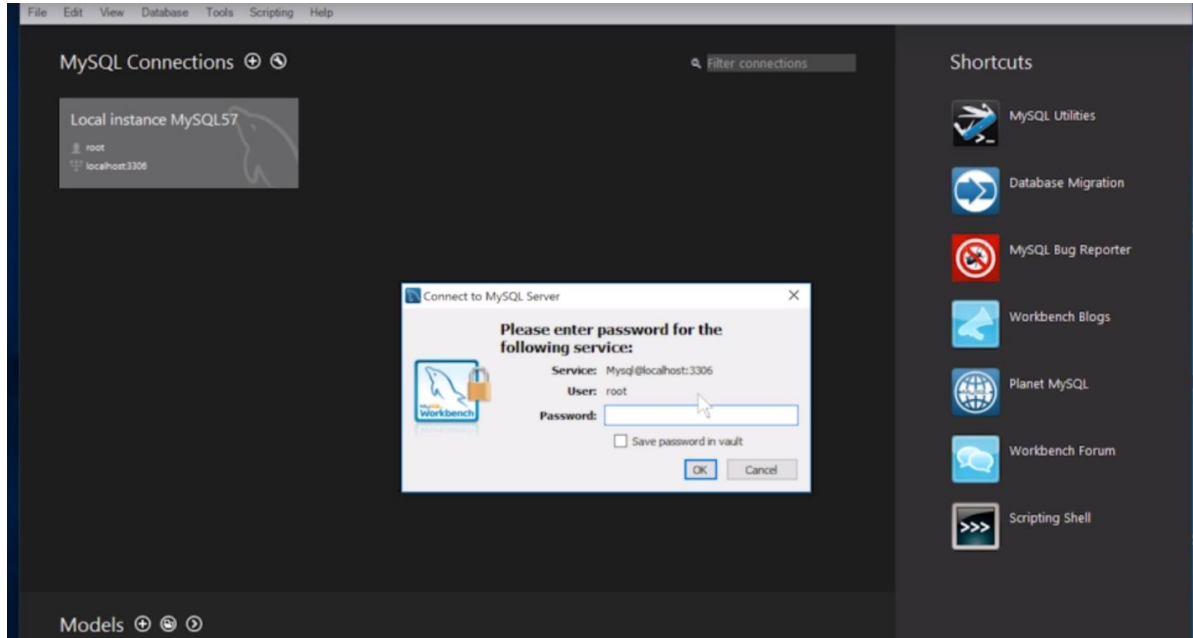
Figure 2.8: Entering the workbench root password

Step4:

Now you can create your first database inside the workbench.
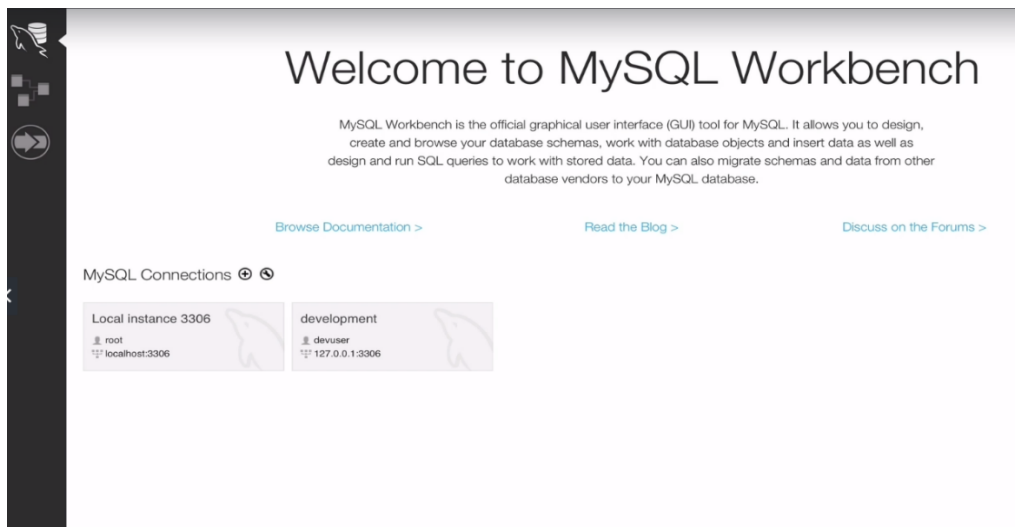
Figure 2.9: Opening screen of workbench

2.4 Data Definition Language

Data definition language is basically consisting of the sql commands which can be use to define the schema of the database. The commands including in this are: CREATE, ALTER, DROP, TRUNCATE.

2.4.1 Data Types

1) Numeric Data type:

- INT: Whole numbers
- FLOAT(M,D): Decimal numbers(approx.)
- DECIMAL(M,D): Decimal numbers(precise)

2) Non-Numeric Data type:

- CHAR(N): Fixed length character
- VARCHAR(N): Varying length character
- ENUM('M','F'): Value from a defined list
- BOOLEAN: True or False values

3) Date and Time type:

- DATE: Date(yyyy-mm-dd)
- DATETIME: Date and time (yyyy-mm-dd hh-mm-ss)
- TIME: Time(hh-mm-ss)
- YEAR: Year(yyyy)

2.4.2 Primary and Foreign Keys

Primary Key: It is a column or the sets of columns which very uniquely recognize a record within the given table. Following are the conditions that need to be satisfied for a column to be a primary key:

- It cannot be null
- It should be unique
- A table should have one primary key

Foreign Key: It is used to link two tables together. A foreign key is a column whose values match the values of another tables' primary keys column.

Figure 2.10: Create table command

2.4.3 Modifying Tables: Adding and Removing Tables

To add a new column in the existing created table the syntax is: "ADD COLUMN <COL NAME> <DATATYPE>".

To remove the column from a table the syntax is: "DROP COLUMN <COL NAME>".

And in order to delete all the data from inside the table truncate command is used.

Figure 2.11: Adding col to the table
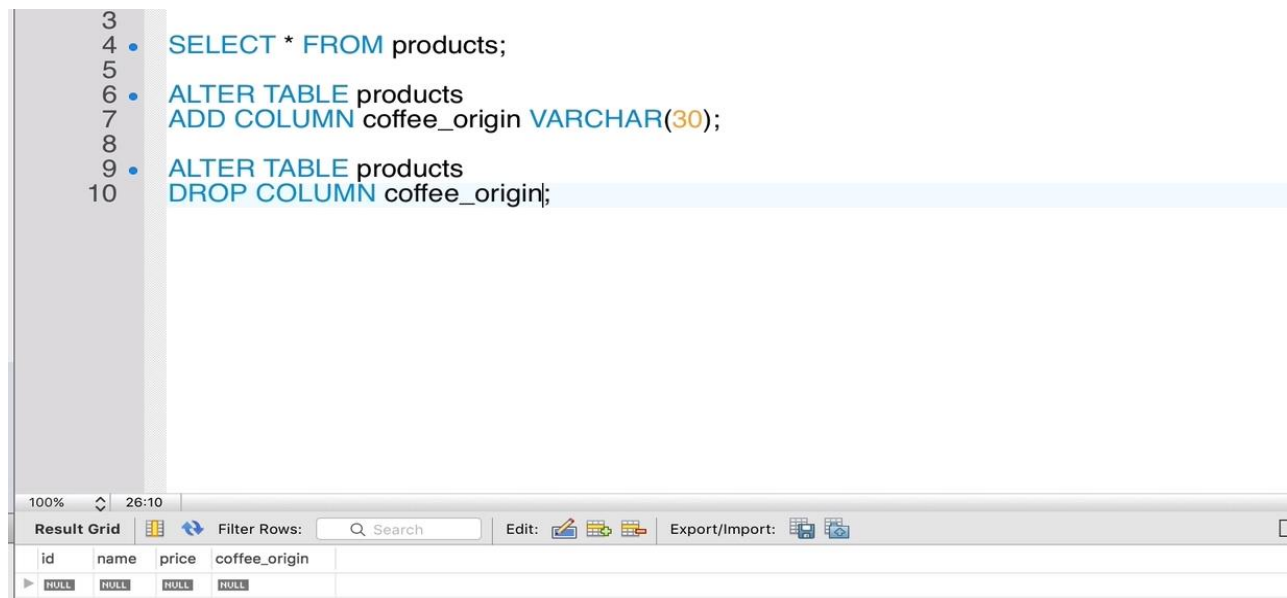
Figure 2.12: Removing column from table

2.4.4 Adding and Removing Primary key and Foreign key

To add the primary key, the syntax is: "ADD PRIMARY KEY(COL_NAME);".

And to drop the primary key from a given column the syntax is: "DROP PRIMARY KEY;".

In the same way to add the foreign key the syntax is:

And in order to remove the foreign key from the given table the syntax is:
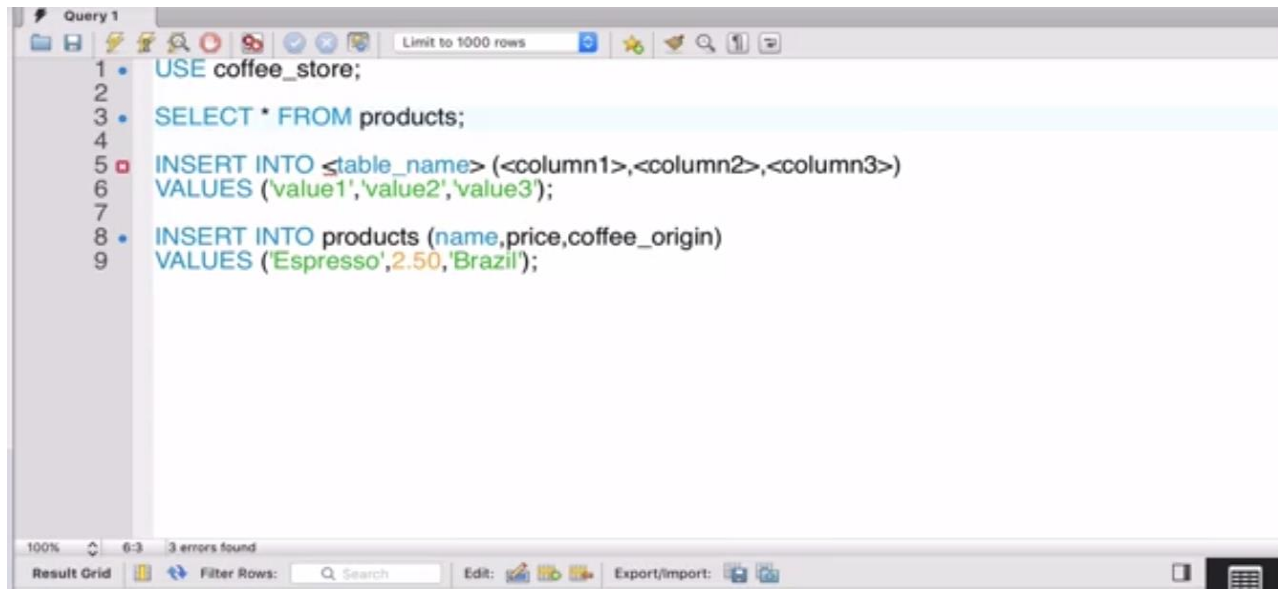
Figure 2.13: Adding and removing primary key

Figure 2.14: Adding and removing foreign key

2.5 Data Manipulation Language

The data manipulation language is used to access, modify or to retrieve the given amount of data from the database. It is basically used to populate and manipulate the data. The commands used under DML are: INSERT, SELECT, UPDATE, DELETE.
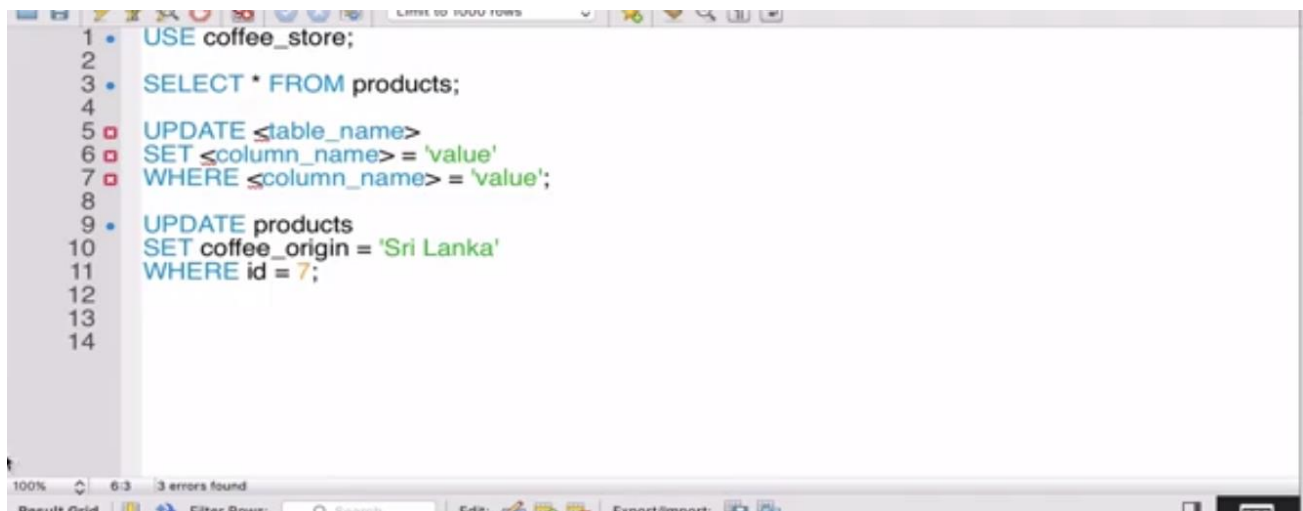
Figure 2.15: Inserting values into tables

Figure 2.16: Updating values into tables

Figure 2.17: Deleting from the table

2.6 About Joins

Joins allow to retrieve data from multiple tables in a single select statement. To join two tables there needs to be a related column between them. There are many different kinds of joins such as inner join, left join, right join.

2.6.1 Inner Join

Inner join will retrieve data only when there is matching values in both the tables.
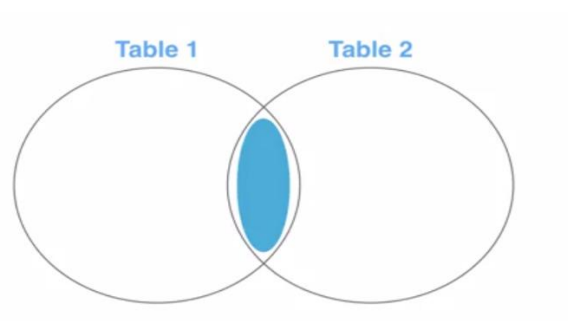
Figure 2.18: Inner joins

```
 1 •   USE coffee_store;
 2
 3
 4 •   SELECT * FROM products;
 5 •   SELECT * FROM orders;
 6
 7 •   SELECT products.name, orders.order_time FROM orders
 8     INNER JOIN products ON orders.product_id = products.id;
 9
10 •   SELECT p.name, p.price, o.order_time FROM orders o
11     JOIN products p ON o.product_id = p.id
12 ▫   ORDER BY ;
13
14
15
16
17
18
```

Figure 2.19: Example query of inner join

2.6.2 Left Join

Left join will get this info. Or get this amount of all the data from the left table and the matching rows from the right side of the table.
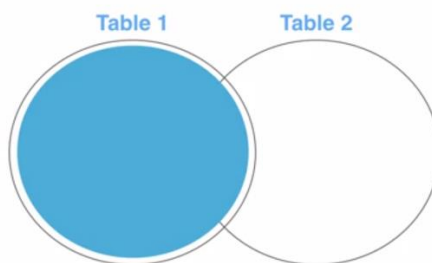
Figure 2.20: Left join example

```
 1 •  USE coffee_store;
 2
 3   ⊟/*
 4    | UPDATE orders
 5    | SET customer_id = 1
 6    | WHERE id = 1;
 7    └*/
 8
 9 •  SELECT * FROM customers;
10 •  SELECT * FROM orders;
11
12 •  UPDATE orders
13    SET customer_id = NULL
14    WHERE id = 1;
15
16 •  SELECT o.id, c.phone_number, c.last_name, o.order_time FROM orders o
17    LEFT JOIN customers c ON o.customer_id = c.id
18    ORDER BY o.order_time
19    LIMIT 10;
20
```

Figure 2.21: Example query of left join

2.6.3 Right Join

The right join will recover the information from the correct table and coordinating columns of the left table.
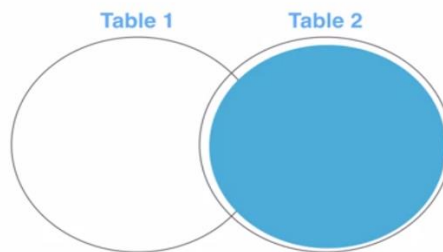
Figure 2.22: Right join

```
 1 •  USE coffee_store;
 2
 3  ⊟/*
 4  │ UPDATE orders
 5  │ SET customer_id = 1
 6  │ WHERE id = 1;
 7  └ */
 8
 9 •  SELECT * FROM customers;
10 •  SELECT * FROM orders;
11
12 •  UPDATE orders
13     SET customer_id = NULL
14     WHERE id = 1;
15
16 •  SELECT o.id, c.phone_number, c.last_name, o.order_time FROM customers c
17     RIGHT JOIN orders o ON c.id = o.customer_id
18     ORDER BY o.order_time
19     LIMIT 10;
20
```

Figure 2.23: Example query of right join

2.7 Subqueries

Subqueries are the queries that are nested within other queries. The subqueries can be used in a SELECT, INSERT, UPDATE, or DELETE query. The nested query could be in the WHERE clause or in the FROM. There are two types of subqueries namely, non-correlated subquery and correlated subquery.

2.7.1 Non-Correlated Subquery

In this type of subquery, the inner query can run independently of the outer query. Inner queries run first and produces a result set and which is then used by the outer query.


```
SELECT id, start_time FROM screenings
WHERE film_id IN
    (SELECT id FROM films
    WHERE length_min > 120)
;
```

Figure 2.24: Non-correlated query

2.7.2 Correlated Subquery

In this type of subquery, the inner query runs for every row in the outer query. Here the inner query cannot run independently of the outer query.

```
SELECT SCREENING_ID, CUSTOMER_ID,
(SELECT COUNT(SEAT_ID)
FROM RESERVED_SEAT WHERE BOOKING_ID = B.ID)
FROM BOOKINGS B;
```

Figure 2.25: Correlated subquery

2.8 Normalization

Normalization is the process of efficiently organizing the data into the database. It is basically done in order to remove the redundant data and to only store the related data in a table. The benefits of normalization are:

- Reduce the storage space
- Reduce insert, update, and deletion anomalies
- Improve query performance

Following are the levels of normalization:

- 1st normal form (1NF)
- 2nd normal form (2NF)
- 3rd normal form (3NF)
- Boyce and Codd normal form (BCNF)

2.8.1 First Normal Form (1NF)

The rules in the first normal form are that a table is in the 1NF if:

- There are no repeated rows of data
- The columns only contain a single value
- The table has a primary key

2.8.2 Second Normal Form (2NF)

The rules in the second normal form are that a table is in the 2NF if:

- They conform to 1NF
- Every section that is definitely not an essential key of the table is reliant in general of the essential key

2.8.3 Third Normal Form (3NF)

The rules in the third normal form are that a table is in the 3NF if:

- They conform to 2NF
- Every section that isn't the essential key is just reliant overall of the essential key

2.9 Relationships

The tables in the MySQL are related through primary key and foreign key. There are three types of relationships namely:

- One to One relationship
- One to Many relationships
- Many to Many relationships

2.9.1 One to One Relationship

In this type of relationship, a key to one table appears no more than once as the key in the another table or vice-versa.

2.9.2 One to Many Relationship

In this type of relationship, key of one table can be in multiple rows of a foreign key column of another table.

2.9.3 Many to Many Relationship

This is the most complicated type of relationship where the two tables can have many instances of each other.

2.9.4 Constraints

Constraints are something that are used to limit something or to put some condition on a column or the value of the table.

Some the constraints used in MySQL are:

- NOT NULL: A column cannot contain any null values
- UNIQUE: A column cannot contain any duplicate values
- PRIMARY KEY: A column that uniquely identifies each row of the data
- CHECK: It controls the values that can be inserted into a column
- DEFAULT: If no values are inserted into a column, you can set a default value

2.10 Aggregate Functions

Aggregate functions are those functions that calculate on data within a column and returns one result row. They group the results by one or more columns. They can use a HAVING clause in a similar way to filter the results sets. For example, count, sum, min, max, average.
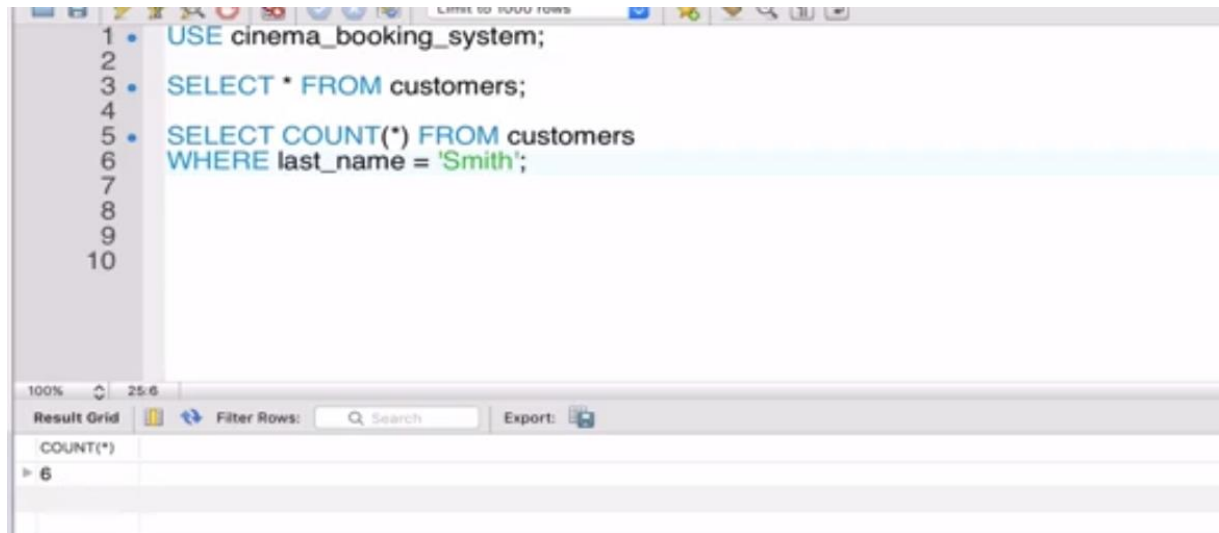
Figure 2.26: Count function example
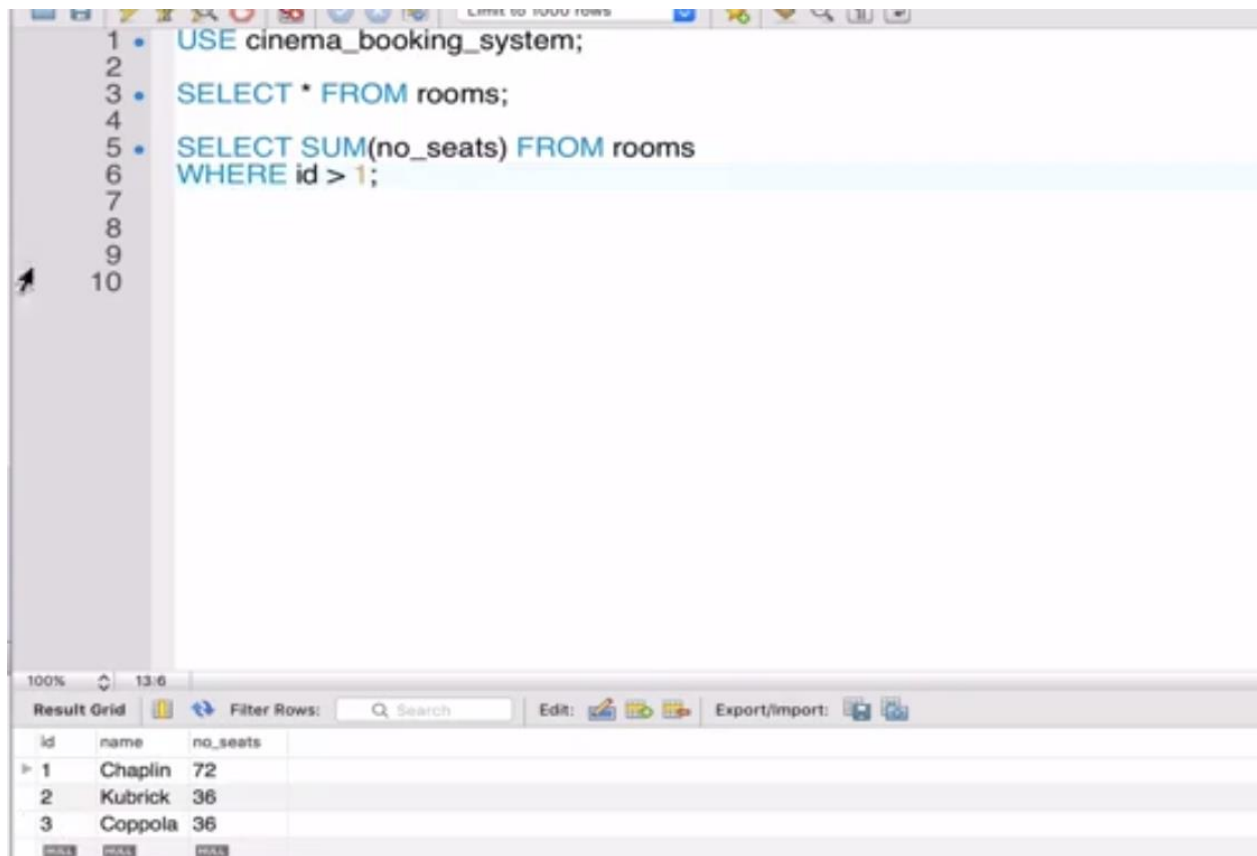
Figure 2.27: Sum function example

CHAPTER 3

Introduction to Hadoop

3.1 History of Hadoop

- Oct 2003: Google File system paper published

- Dec 2004: Jeffrey Dean & Sanjay Ghemawat from Google published MapReduce paper called "MapReduce: Simplified Data Processing on Large Clusters"

- Jan 2006: Above MapReduce Paper inspired Doug cutting, a yahoo employee then to develop an open source implementation of MapReduce framework

- Jan 2006: Hadoop subproject created as extension of Apache Nutch project, created by Doug Cutting.

- Apr 2006: Hadoop 0.1.0 released

- May 2006: Yahoo deploys 300 machine Hadoop cluster

- 2008: Cloudera, one of the major distributor of Hadoop founded

- Apr 2007: Yahoo runs 2 clusters of 1,000 machines

- Dec 2011: Apache Hadoop release 1.0.0 available

- 2011: Hortonworks, another major Hadoop distributor founded

- Dec 2015: Apache Hadoop release 2.6.3 available

3.2 What is Hadoop?

Processing of this Bulk volume of data which is available on earth with such amount of limited computing resource means how long it will take?
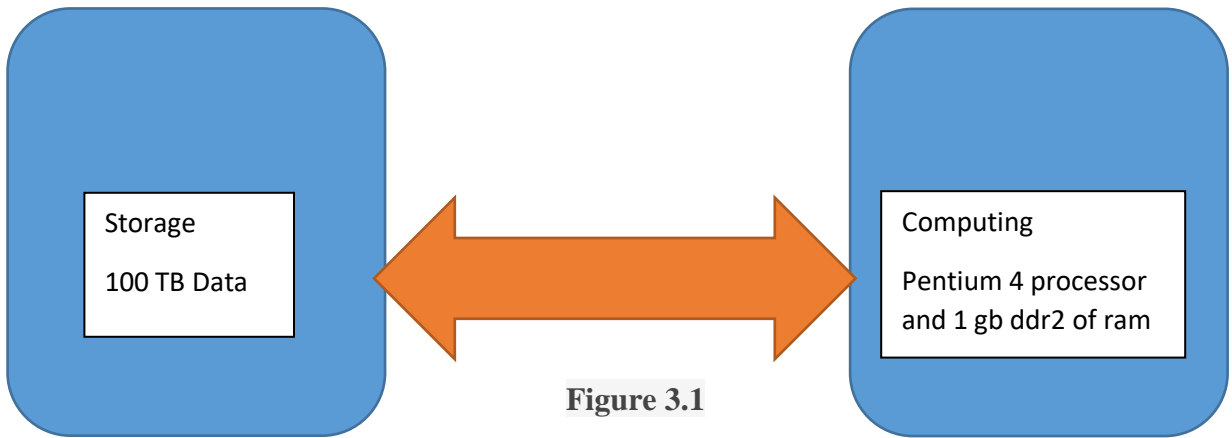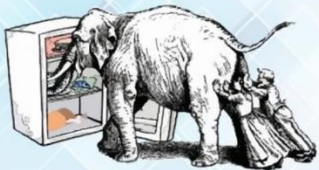
Figure 3.1

Apache Hadoop is the structure for disseminated capacity and figuring (mix of Processor and RAM) model to manage huge volume of information.

Apache Hadoop is an arrangement of opening of the source of the programming utilities that support using an arrangement of kind of the computers or lapitops to deal with issues including gigantic proportions of data and computation. It gives an item structure to scattered limit and getting ready of gigantic data using the MapReduce programming model.

3.3 Problem in storing Big amount of Data

Figure 3.2: Data storing problem

Huge data is a mixture of the composed and unstructured data assembled by affiliations that can be delved for information and used in AI adventures, insightful exhibiting and other advanced assessment applications. Structures that method and store enormous data have become a normal piece of data the officials' plans in affiliations. Colossal data is consistently depicted by the 3Vs: the enormous volume of data in various circumstances, the wide arrangement of data types set aside in gigantic data structures and the speed at which the data is created, assembled and arranged. These characteristics were first perceived by Doug Laney, Gartner further advanced them after it got it. Even more starting late, a couple of various Vs have been added to different portrayals of enormous data, including veracity, worth and change.

3.4 Examples of Big Data

Utilizing client information for instance, the various parts of examination that should be possible with the data found in sets of enormous information incorporate the accompanying:

Comparative analysis. This includes the examination of user behavior metrics and the observation of real-time customer engagement in order to compare one company's products, services and brand authority with those of its competition.

Social media listening. This is information about what people are saying on social media about a specific business or product that goes beyond what can be delivered in a poll or survey. This data can be used to help identify target audiences for marketing campaigns by observing the activity surrounding specific topics across various sources.

Marketing analysis. This includes information that can be used to make the promotion of new products, services and initiatives more informed and innovative.

Customer satisfaction and sentiment analysis. The entirety of the data accumulated can uncover how clients are feeling about an organization or brand, if any potential issues may emerge, how brand dedication may be safeguarded and how client assistance endeavors may be improved.

3.5 HDFS Introduction

HDFS stands for Hadoop Distributed File System. As a result of certain confinements that will be benefit on every single record framework by methods for least and most extreme volume it will support, least and greatest document size it will bolster, stage conditions will be there.

To defeat with every one of those challenges and with some additional highlights to deal with the enormous volume of information Hadoop concocted some filesystem called as Hadoop Distributed document framework that will just utilized in Hadoop structure. Hadoop has been developed by the base language is java. In Hadoop if u want to do any work means, give the input as java program code. It will revert u back with output also as java format.

3.5.1 MapReduce

MapReduce is the data processing model in Hadoop framework. It is the logical combination of CPU and RAM. It is used for cluster resource management and data processing.

If in the case of handle the Hadoop frame work data's this computational unit will act as processing model and it will complete the work as early as possible with the help of logical unit.

8 core processor

4 GB of RAM

▶ Logical unit of Separate computing and storage point

▶ 10 TB of Storage(HDFS)

▶ 800 Core Processor
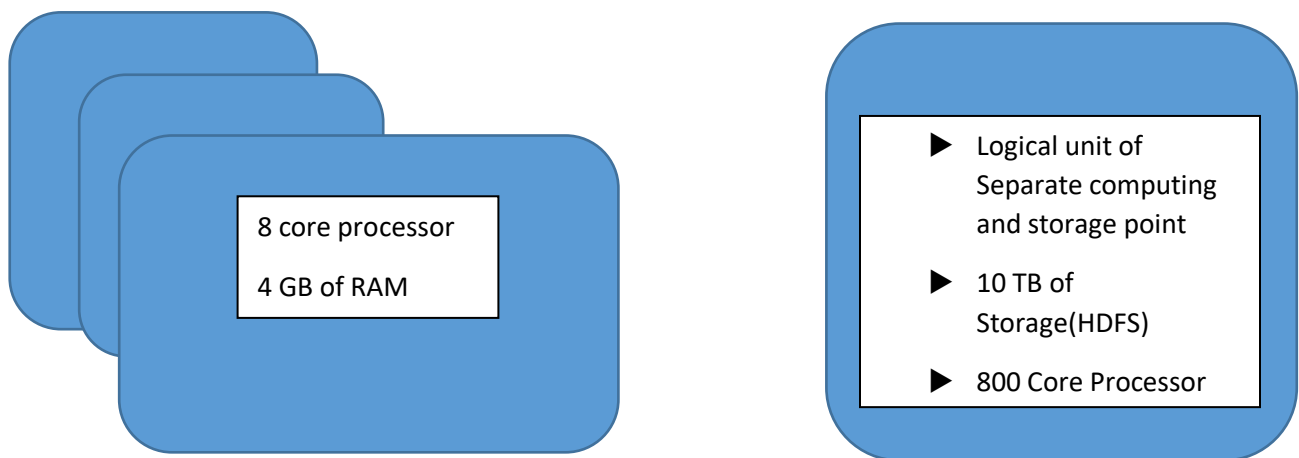
Figure 3.3: block diagram

Figure 3.4: Hadoop solves the big data

Hadoop come up with two components that are:

- Storage (HDFS)
- MapReduce

If in the case of Cluster Architecture, it will be in the structure master/slave architecture. One master/multiple slave architecture means if master node fails what will happen?

In Master slave architecture based software, especially in the case of single master architecture, they will write an algorithm to select the new master from one of the slave node in the cluster. Election process will be based on latency, performance (latency is nothing but time taken for one successful communication).

3.5.2 Hadoop 1.0

MapReduce (processing unit) for Cluster resource management and data processing:

- Master
- Slave

HDFS (storage unit) for redundant and reliable storage:

- Master
- Slave

3.5.3 Components of Hadoop 1.0

Some of the basic components of Hadoop 1.0 are as follows:

- HDFS(Storage)

- Name Node

- Data Node

- MapReduce(Processing)

- Job Tracker

- Task Tracker
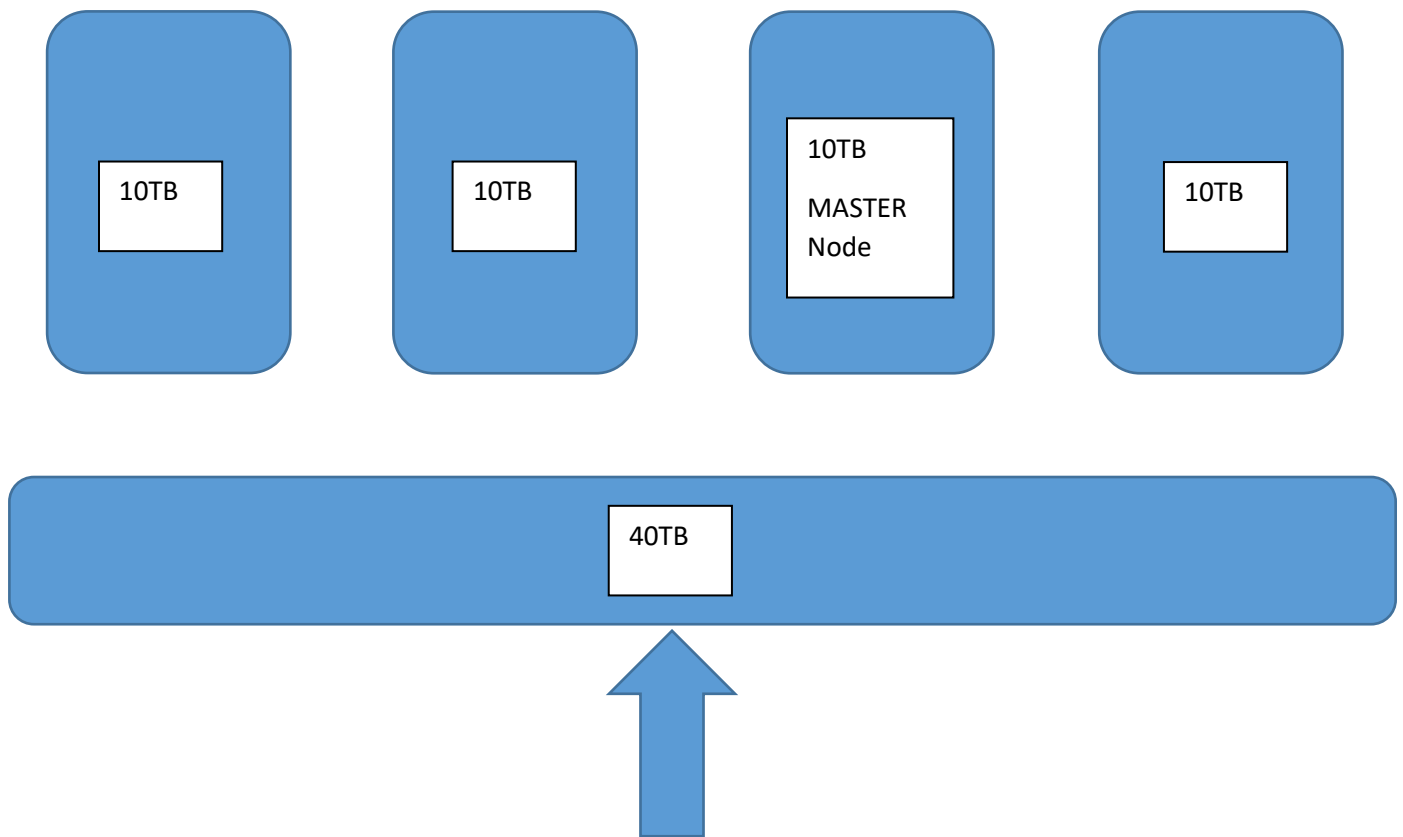
Figure 3.5: Master and slave nodes

Name node and data node are sub components of HDFS. Name node present only in Master and Data node present in all the Slave.

Name Node:

- Name Node just stores the metadata of HDFS – the index tree of all documents in the record framework, and tracks the records over the bunch.
- Name Node doesn't store the real information or the dataset.

Data Node:
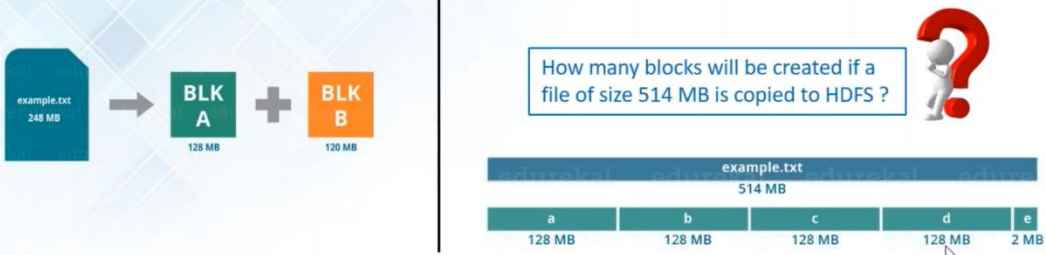
- The data itself is actually stored in the Data Nodes.

All the files that are stored in the format of block. Each and every blocks will be in the size of 64MB in Hadoop 1.0 and 128MB in Hadoop 2.0. For example, I got file size will in 500 Mb means: In 1.0

architecture first 7 blocks will 64mb last block will be in 52mb size and in 2.0 architecture first 3 will be 128MB last 116 MB.

Figure 3.6: HDFS blocks

3.5.4 Rack Awareness Algorithm

- Rack Awareness is an algorithm to main the effective replication policy over the Hadoop architecture
- in Hadoop each and every blocks have 1+2 replication policy by default
- Ex 2 tb that 2 tb has been saved as 6tb data (need of replication)
- Computing Resource is cheaper no Storage resource is cheaper?
- Storage is Cheaper
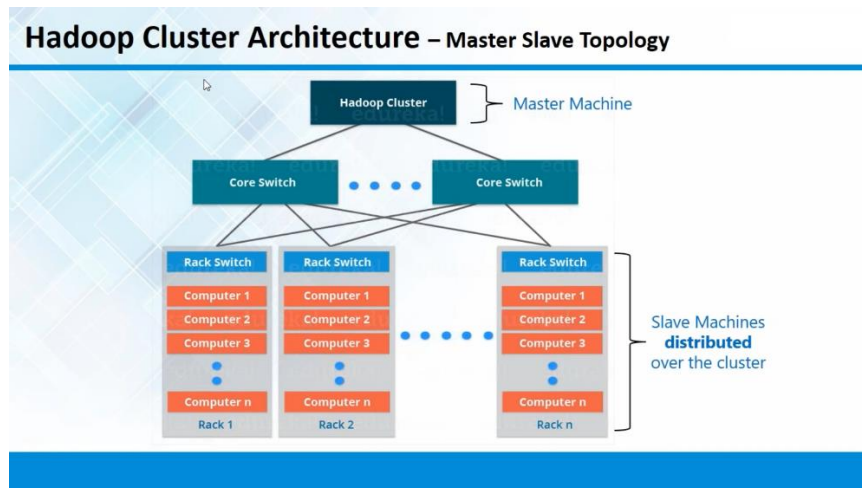
Figure 3.7

3.6 HDFS Architecture
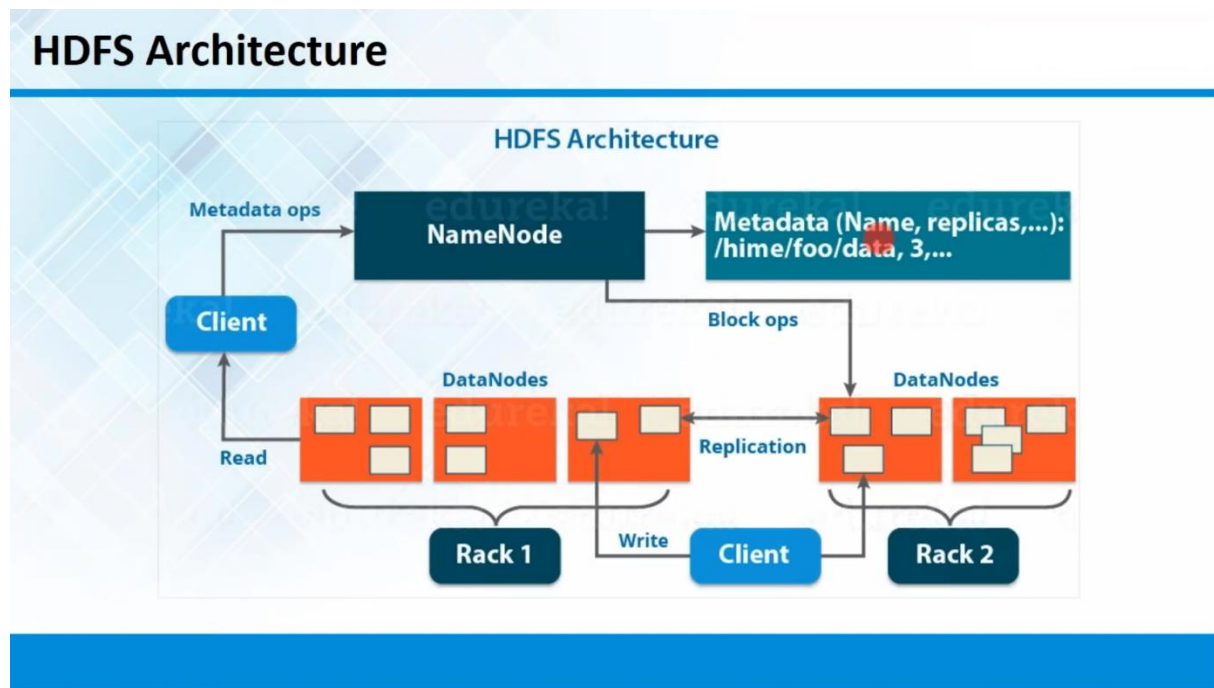
Figure 3.8: Block diagram of HDFS

3.6.1 Job Tracker and Task Tracker in MapReduce

- Job Tracker and Task Tracer are the components of MapReduce.
- Job Tracker always available on Master

- The JobTracker is the administration inside Hadoop that homesteads out MapReduce assignments to explicit hubs in the bunch, preferably the hubs that have the information..
- TaskTrackers is always available on all the slave nodes
- TaskTrackers is a node in the cluster that accepts tasks from JobTracker

3.6.2 Hadoop 1.0 vs 2.0
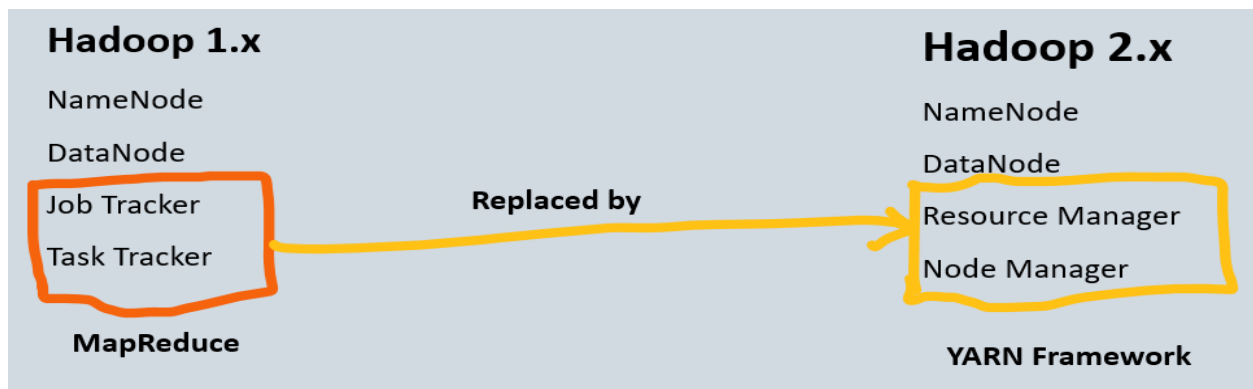
Figure 3.9

In 1.0 there are following components:

- Naame Nodes
- Daata Nodes
- Jobs Trackers
- Tasking Tracker

In 2.0 there are following components:

- Names Node
- Data Node
- Resources Manager compute
- Node Manager
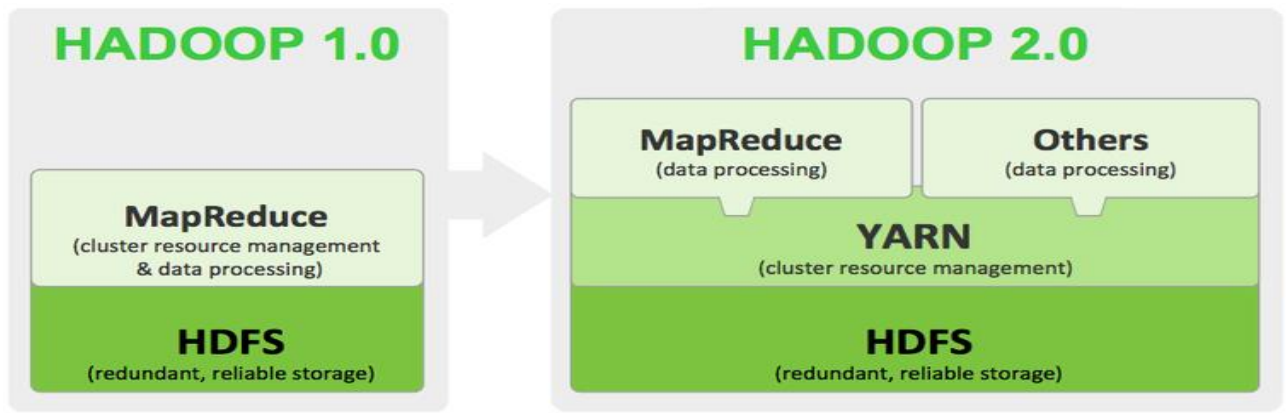
Figure 3.10: Block diagram of Hadoop 1.0 and 2.0

- Because of limitations that will be there in 1.0, it will accept java as input unit

- Job tracker will accept the input only as java language

- In Hadoop Java programming guys only able to work on this architecture

- In 2.0 YARN it will accept the input as multi programming language and software's

- Ex: Python/Hive

- Datacentre

- Datacenter are group of servers avail for storing and retrieving data in same location

- Datacentre Rack

- rack is the electronic framework of switches assembling different servers in datacenter

3.6.3 Installation of Hadoop

Two method of running Hadoop:

- Standalone

- Cluster

Types of Hardware's used for installation:

- Manual Installation

- Ready to run Environment in Public Clouds(EMR)

- Ready to run in Private Cloud/ Virtual Machine (Cloud ERA/Horton Works)

Need of JVM

Java Virtual Machine (**JVM**) is a medium between the operating system and the Java application (**Hadoop** in this case).

- System will deal all the data's as byte codes
- Chrome -> YouTube video
- Bytes by byte(Streaming)
- How fast it will stream your video quality will be good
- We are going to deal with huge amount of data, if I go by traditional processing it will consume more and more time even though very good processing unit will be there
- JVM- will deal the kernel data will be in the blocks format (it will be much faster than traditional approach)

Hadoop is only meant for Analytical purpose. All the application that will concentrate on read they won't concentrate on write. All the commands we execute in Hadoop architecture ad nothing but predefined scripts. If I get any customized requirement means what will I do?

Ex migrate MySQL database – need to write java script to import all the data/ specified data from the destination to hdfs file system vi jdbc connection. Apache Sqoop -> Apache Sqoop will be the tool that will sit on top of yarn, that is mainly used for import and export the database. Import -> database to hdfs. Export -> hdfs – database. Database – performance because of heavy load, application maximum capability it reached storage/computing.

Apache Sqoop is a tool designed for efficiently transferring bulk data between and non bulk informations.

3.7 Sqoop

Apache Sqoop(TM) is a device intended for proficiently moving mass information between Apache Hadoop and organized datastores, for example, social databases.

3.7.1 Instructions to Import Sqoop

- o sqoop imports
- o --connecting
- o //karthick1808.c5e69p24nuic.ap-south-1.rds.amazonaws.com/aman2910
- o --username karthick1808
- o --password 12345678
- o --table cus_tbl1
- o --incremental append
- o --check-column cus_id
- o --last-value 5
- o --target-dir /amansql3
- o -m 1

One of method to import structural data base to HDFS file system is Sqoop?

- o Importing as new/particular and auto increment is possible or not?
- o Can we import multi flavours of structural databases?
- o Ex MySQL, Oracle, Microsoftsql, Postgres, Mariadb along with drivers
- o Client is coming migrate all my data from oracle into MySQL?
- o Why customers will come for data migration?
- o Performance Issue
- o Application incapability (Max Volume, Cpu Core)
- o Xyz, you migrated all the data to hdfs file system, how will you run the query against the migrated data
- o Hive – min and max volume it will support – My cluster size volume is 115.9 GB

3.7.2 Hive

A hive is a data appropriation focus programming setting dependent on hadoop for giving information question and assessment and is basically used in the command line prompt. Hive gives sql like to address data hive is an information distribution center programming venture based on Apache Hadoop for giving information inquiry.

- Sql like tool (Command line tool) that will run on top of hdfs file system.
- We can write the java program for each and every searching what we are making so far for the replace of programmatic interface they came up with API based tool called hive.
- Hive is the distributed tool that will run on Hadoop cluster that will support for both Mapreduce and hdfs.
- If in the case of searching large data sets also it will acquire the resource from Mapreduce.
- If in the case of large data sets also because of Hadoop supports elasticity we can do anything possible here.
- By default, whatever the table we create in hive default location it will save in hdfs file system is /user/hive/warehouse
- Import and export all the data are in the form *. Sql
- Comma separated
- Import that csv to Hadoop master machine depends on the data's going to create tables in hive and trying to import that csv file to hive

CHAPTER 4
AWS Overview

Distributed computing has two implications. The most well-known alludes to running outstanding tasks at hand remotely over the web in a business supplier's server farm, otherwise called the "open cloud" model. Well known open cloud contributions, for example, Amazon Web Services (AWS), Salesforce's CRM framework, and Microsoft Azure—all epitomize this recognizable idea of distributed computing. Today, most organizations adopt a multicloud strategy, which basically implies they utilize more than one open cloud administration.

The second significance of distributed computing portrays how it functions: a virtualized pool of assets, from crude register capacity to application usefulness, accessible on request. At the point when clients obtain cloud benefits, the supplier satisfies those solicitations utilizing propelled mechanization as opposed. Companies offering hypervisors (Type 1/ Type 2) as an service are meant to be Cloud Providing Companies

- By means of 3 types of methods they will offer you the service
- Public – Server that used as hypervisor that has been shared to group of companies nor group from different sectors are defined as public cloud
- Private – Server that has been used for same set of people/Company
- Hybrid – Same community people sharing the same server

4.1 Public Clouds

A public cloud is a cloud service offered to multiple customers by a cloud provider. The term "public cloud" is used to differentiate between the original cloud model of services accessed over the Internet and the private cloud model. Public clouds include SaaS, PaaS, and IaaS services.

- AMAZON WEB SERVICES(AWS)
- GOOGLE
- AZURE(MICROSOFT)
- ORACLE
- IBM

- DELL Emc
- DIGITAL OCEAN

4.2 Amazon Web Services

i.   AWS Will OFFER you the services in multiple Region as well as in Multiple Technologies

ii. How Region that AWS Will offer you the Service?

iii. Total of 24 Regions it will support and in that 2 will be used only for govt use and remaining 22 is available for general normal users.

iv. In which case we will give preference to regions

v.   Ex: I am running an company from India but all my clients belong to Australia locations where will I deploy my server?

vi. Performance -> Latency, Cost -> Australia cost of AWS is lesser than US COST

vii. Rates will vary for each and every services in AWS as well as it will vary for each and every regions

4.2.1 Services Offered by AWS

i.   EC2 - COMPUTE
ii.   EMR – Analytics (HADOOP, SPARK/SCALA)
iii. Kinesis - Analytics
iv.   S3 - STORAGE
v.   Dynamo DB - DATABASE
vi.   LAMDA - COMPUTE
vii. REDSHIFT - DATABASE
viii. GLUE - Analytics
ix.   VPC - Networking & Content Delivery
x.   IAM - Security, Identity, & Compliance
xi.   ATHENNA - Analytics

4.2.2 IAM (Identity and Access Management)

AWS Identity and Access Management (IAM) empowers you to oversee access to AWS administrations and assets safely. Utilizing IAM, you can make and oversee AWS clients and gatherings, and use authorizations to permit and deny their entrance to AWS assets. IAM is a component of your AWS account offered at no extra charge. You will be charged distinctly for utilization of different AWS benefits by your clients.

Active Directory which helps your employees sign in and access resources in secure manner. It manages permission and privilege control. Create own network under that put all the users on same network for securely accessing the resources under the same network by the users belonging to the same. Root account I am going to create sub accounts.

What will be login url? Created some user with password. We gave some read-only privilege for IAM access.

```
ROOT account in AWS

500 Services

Full privilege for all the 500 services
for all the users
```
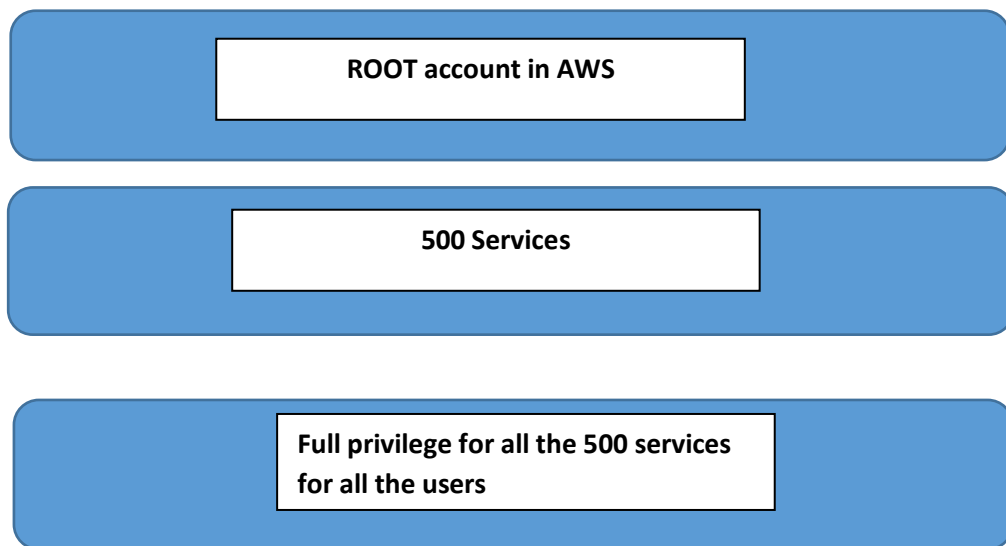
Figure 4.1: Block diagram for Root account

4.2.3 EC2 (Elastic Compute Cloud)

Versatile Compute Cloud – Elastic (can broaden and contract relies on it needs up until now) they are giving the flexibility highlight to figuring Resources, for example, CPU, RAM, Storage and Networking ETC. in the cloud.

It as an IAAS sort of administration. Framework as Service (We are taking RAM, CPU, Networking and Storage as rental and introducing OS on that base layer and going to utilize it). Foundation as a Service (IaaS) is a distributed computing administration where endeavors lease or rent servers for figure and capacity in the cloud. Clients can run any working framework or applications on the leased servers without the upkeep and working expenses of those servers.

- Selecting your OS Flavour - **Microsoft Windows Server 2019 Base**

- Selecting your OS RAM, CPU, Type of Hard drive and Type of Network

- Configure Instance Details – Networking Details

- Configure your Storage Size C Drive Space – 30 Gb

- Add Tags

- Select an **existing** security group – Default Security Group

4.2.4 AMI (Amazon Machine Image)

- Amazon Machine Image – VMWARE Machine Template File
- Amazon Market Place – Search for image and select your image
- Security Group – EC2 Firewall – External Firewall means – IAAS – wile launching the machine itself you need to take care of the responsibility of firewall if something goes wrong we won't take in charge because of it its IAAS

4.3 Storage

Like Electrical signals and all we transporting from one end to another we are using wires. They tried to transmit the computer data from one end to another end through wires. Electrical signals are easy to carry because of those are analogy signals. Computer data are digital medium; Lan Cable 4 pairs of twisted copper wires will be there to transit the data in a fast manner. Ethernet Adaptor - > converting the digital to analogue and analogue to digital.

- LAN – within 100 to 200 meters without issue they shared the data in efficient manner and less cost
- MAN – 50 – 100 kms range it will transmit the data
- WAN – Country to Country Transmission along internet, intranet

Various storage devices used are:

- Flash Cards
- Magnetic Tape
- Hard Drive (Magnetic Drive)
- Floppy Drive
- Cd Drive
- Pen Drive/ Cards
- DVD Drive
- Hard Drive(SSD)

4.3.1 Principles of Filesystem

1) File Based Sharing (Bit Based Storage)

2) Block Based Sharing (ISCSI)

3) Object Based Storage

4.3.2 Object Based Storage

It's a mechanism to store and retrieve the data, it will consider all the files as objects. For each and every object they will maintain separate privilege control mechanism. Place where all the objects will save they will call it as Buckets.
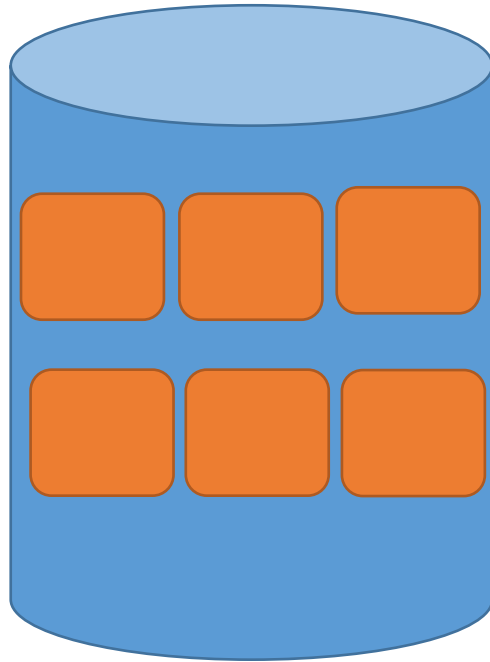
Figure 4.2: Bucket containing objects

4.3.3 S3 (Simple Storage Service)

- S3 is kind of PAAS (we are not going to install any OS here, NO Application Installation Procedure also there) that is available on AWS.

- S3 is a storage mechanism that will work on the basis of Object Oriented Storage

- With the help of S3, we will provide fine granular permission, Encryption, Security and features (log, Versioning)

- S3 is a Global service that is available on AWS

- If u want to start with s3, we need to create our own storage location called bucket.

Some of the S3 bucket norms are:

- Bucket names must be between 3 and 63 characters long.
- Bucket names can consist only of lowercase letters, numbers, dots (.), and hyphens (-).
- Bucket names must begin and end with a letter or number.

- Bucket names must not be formatted as an IP address (for example, 192.168.5.4).
- Bucket names must be unique across global
- Buckets name should not be in the format any company names

4.3.4 S3 Standard Storage (default)

Key Features:

- Low latency and high throughput performance
- Designed for durability of 99.999999999% of objects across multiple Availability Zones
- Resilient against events that impact an entire Availability Zone
- Designed for 99.99% availability over a given year
- Backed with the Amazon S3 Service Level Agreement for availability
- Supports SSL for data in transit and encryption of data at rest
- S3 Lifecycle management for automatic migration of objects to other S3 Storage Classes

4.3.5 AWS VPC

Amazon Virtual Private Cloud (Amazon VPC) lets you provision a logically isolated section of the AWS Cloud where you can launch AWS resources in a virtual network that you define. You have complete control over your virtual networking environment, including selection of your own IP address range, creation of subnets, and configuration of route tables and network gateways. You can use both IPv4 and IPv6 in your VPC for secure and easy access to resources and applications.

You can without much of a stretch tweak the system design of your Amazon VPC. For instance, you can make an open confronting subnet for your web servers that approach the web. You can likewise put your backend frameworks, for example, databases or application servers, in a private-confronting subnet with no web get to. You can utilize numerous layers of security, including security gatherings and system get to control records, to assist control with getting to Amazon EC2 cases in each subnet.
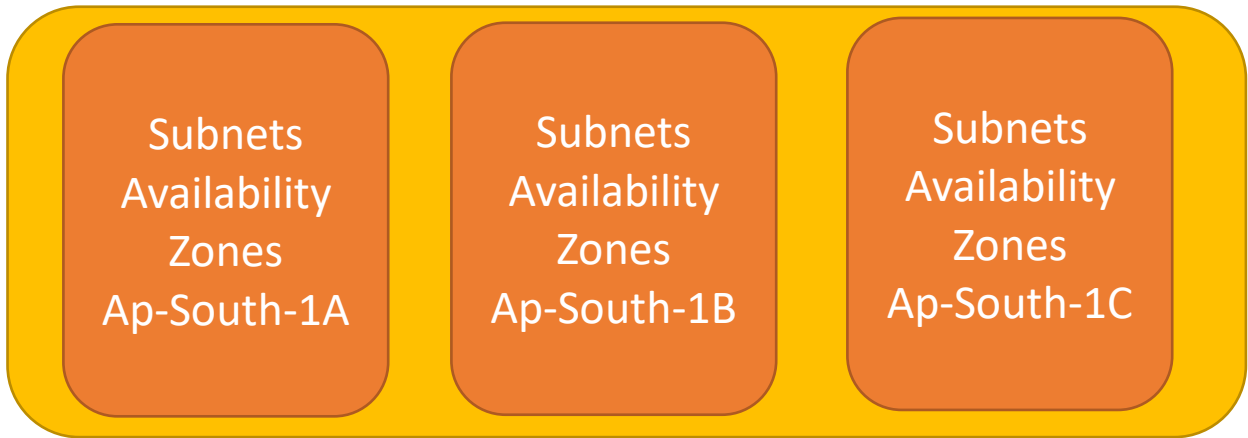
Figure 4.3: Block diagram of the availability regions

4.4 Amazon Redshift

Amazon Redshift is an information distribution center item which structures some portion of the bigger distributed computing stage Amazon Web Services. The name intends to move away from Oracle, red being a reference to Oracle, whose corporate shading is red and is casually alluded to as "Large Red." It is based on innovation from the huge equal preparing (MPP) information distribution center organization ParAccel (later obtained by Actian), to deal with huge scope informational indexes and database relocations. Redshift varies from Amazon's other facilitated database offering, Amazon RDS, in its capacity to deal with logical remaining tasks at hand on huge information, informational indexes put away by a section situated DBMS rule.

Amazon Web Services (AWS). It is a productive answer for gather and store every one of your information and empowers you to examine it utilizing different business knowledge devices to get new experiences for your business and clients. Columnar upheld database.

CHAPTER 5

<u>JAVA</u>

Java is an extensively valuable programming language that is class-based, object-arranged, and planned to have as very few execution conditions as could sensibly be normal. It is relied upon to let application engineers create once, run wherever (WORA), suggesting that arranged Java code can run on all phases that help Java without the necessity for recompilation. Java applications are regularly accumulated to bytecode that can run on any Java virtual machine (JVM) paying little notice to the fundamental PC building. The syntax of Java resembles C and C++, anyway it has less low-level workplaces than them two. Beginning at 2019, Java was one of the most notable programming vernaculars being utilized by GitHub, particularly for client server web applications.
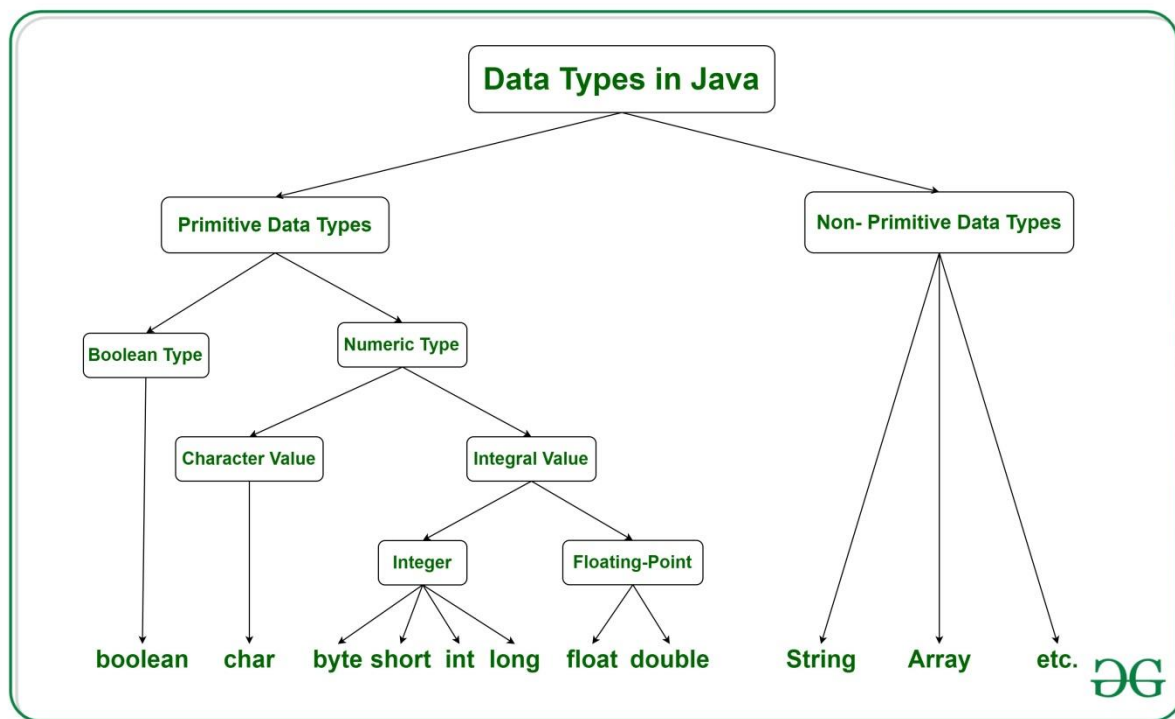
5.1 Data Types in Java

Figure 5.1: Data types

5.2 Some Basic Concepts of OOPS

- **Inheritance**: Classes can share, acquire or "acquire" properties and techniques that have a place with existing classes. This lets you reuse existing code and lessens the time you spend coding. A class that acquires from another is known as an inferred class or a youngster class. A class that shares its properties and strategies is known as a base class or parent class.

- **Polymorphism** : alludes to the capacity of a strategy to be utilized in various manners, that is, it can take various structures at various occasions (poly + morphos). There are two sorts of polymorphism: compile time polymorphism and run time polymorphism.

- **Abstraction** : Information deliberation alludes to the procedure of just showing applicable properties and techniques to deal with an item, while concealing the rest. Information deliberation lets you lessen the multifaceted nature (clearly) of a program and is an enormous bit of leeway offered by classes in OOP dialects.

- **Encapsulation**: Embodiment alludes to keeping objects with their strategies in a single spot. It likewise ensures the trustworthiness of the information – keeps it from being unnecessarily modified by confining access to the information, ideally by concealing it from outside components. Epitome is regularly mistaken for information reflection, however they are various ideas totally. Information stowing away, or information reflection, has more to do with get to specifiers. A software engineer should initially epitomize the information and afterward he can find a way to conceal it.

5.3 Strings

String is a game plan of characters. In java, objects of String are perpetual which suggests a consistent and can't be changed once made. Making a String. There are two distinct approaches to make string in Java:

String literal
String s = "STAKSHI '';

Using *new* keyword
String s = new String ("STAKSHI");

5.4 Arrays in Java

A cluster is a holder object that holds a fixed number of estimations of a single kind. The length of a group is set up when the display is made. After creation, its length is fixed.

• In Java all displays are effectively dispensed.

• Since displays are dissents in Java, we can find their length using part length. This is extraordinary according to C/C++ where we find length using sizeof.

• A Java bunch variable can in like manner be articulated like various elements with [] after the data type.

• The elements in the bunch are mentioned and each have a record beginning from 0.

• Java bunch can be in like manner be used as a static field, a close by factor or a methodology parameter.

• The size of a display must be dictated by an int regard and not long or short.

• The immediate superclass of a display type is Object.

• Every show type completes the interfaces Cloneable and java.io.Serializable.

Show can contains locals data types similarly as objects of a class dependent upon the significance of group. On the off chance that there ought to emerge an event of locals data types, the genuine characteristics are taken care of in flanking memory zones. In the event that there ought to be an event of objects of a class, the veritable things are taken care of in load segment.

Creating, Initializing, and Accessing an Array

```
// create an array of integers
aaaa = new int[33];

//Initialization
aaaa[0] = 100;/
aaaa[1] = 200;/
aaaa[2] = 300;/, etc

//Accesing an array
System.out.println("Element 1 at file 0: " + aaaa[0]);
```

5.5 JDBC Connectivity

To interface Java application with the MySQL database, we need to follow 5 after advances. In this model we are using MySql as the database. So we need to know following informations for the mysql database:

1.Driver class: The driver class for the mysql database is com.mysql.jdbc.Driver.

2.Connection URL: The affiliation URL for the mysql database is

where jdbc is the API, mysql is the database, localhost is the server name on which mysql is running, we may likewise utilize IP address, 3306 is the port number and sonoo is the database name. We may utilize any database, in such case, we have to supplant the sonoo with our database name.

5.6 Exception Handling

The Exiception Handling in Java is one of the unbelievable instrument to manages the runtime bumbles so average movement of the appllication can be kept up. JAVA special case Hierarchy. The java.lang. Throw able class is the root class of Java Exception request which is then gainned by two subclasses: Exception and Error. A request for Java Exception classes are given underneath:

Kinds of Exceptions :

- Built-in Exceptions
- User Defined Exceptions

Some case of inherent exemptions are as following :

1.ArithmeticException : It is hurled when a wonderful condition has occurred in a number shuffling movement.

2.ArrayIndexOutOfBoundsException : It is hurled to show that a display has been gotten to with an illegal record. The rundown is either negative or more noticeable than or proportional to the size of the bunch.

3.ClassNotFoundException : This Exception is raised when we endeavor to get to a class whose definition isn't found .

4.FileNotFoundException: This Exception is raised when a record isn't accessible or doesn't open.

5.IOException: It is hurled when a data yield action failed or ruined.

5.7 Java Exception Keywords

There are 5 watchwords which are used in managing extraordinary cases in Java:

1. The "attempt" watchword is used to decide a square where we should put exclusion code. The endeavor square ought to be trailed by either get or finally. In other words, we can't use attempt square alone.

2. Catch: The "get" square is used to manage the exclusion. It must be gone before by endeavor square which infers we can't use get square alone. It will in general be trailed by finally square later.

3. Finally: The "at last" square is used to execute the noteworthy code of the program. It is executed whether a unique case is managed or not.

4. Throw: The "toss" catchphrase is used to hurl an exceptional case.

5. Throws: The "tosses" catchphrase is used to declare exclusions. It doesn't hurl an extraordinary case. It discovers that there may happen an extraordinary case in the procedure. It is continually used with methodology signature.
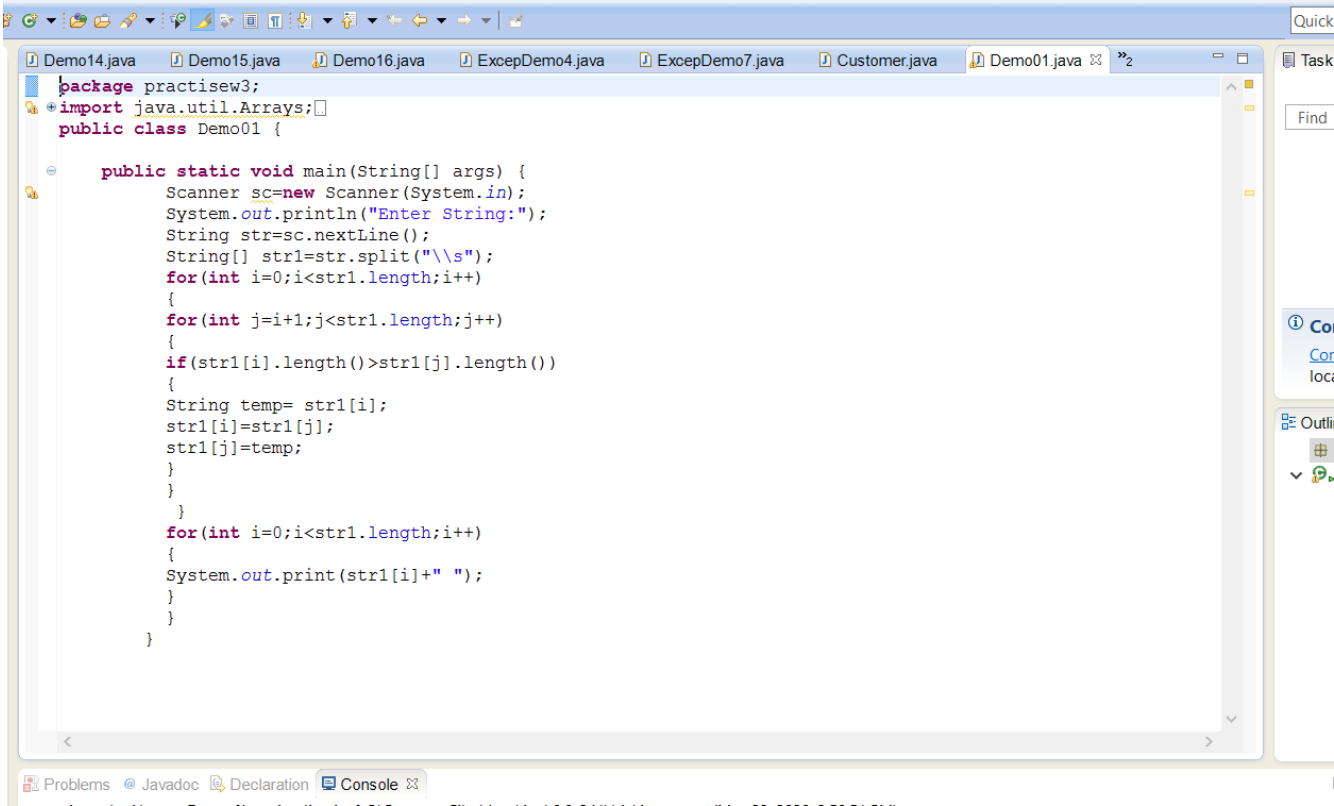
5.8 Some example codes of Java

Figure 5.3: Sample code of java

Figure 5.4: Sample code of java

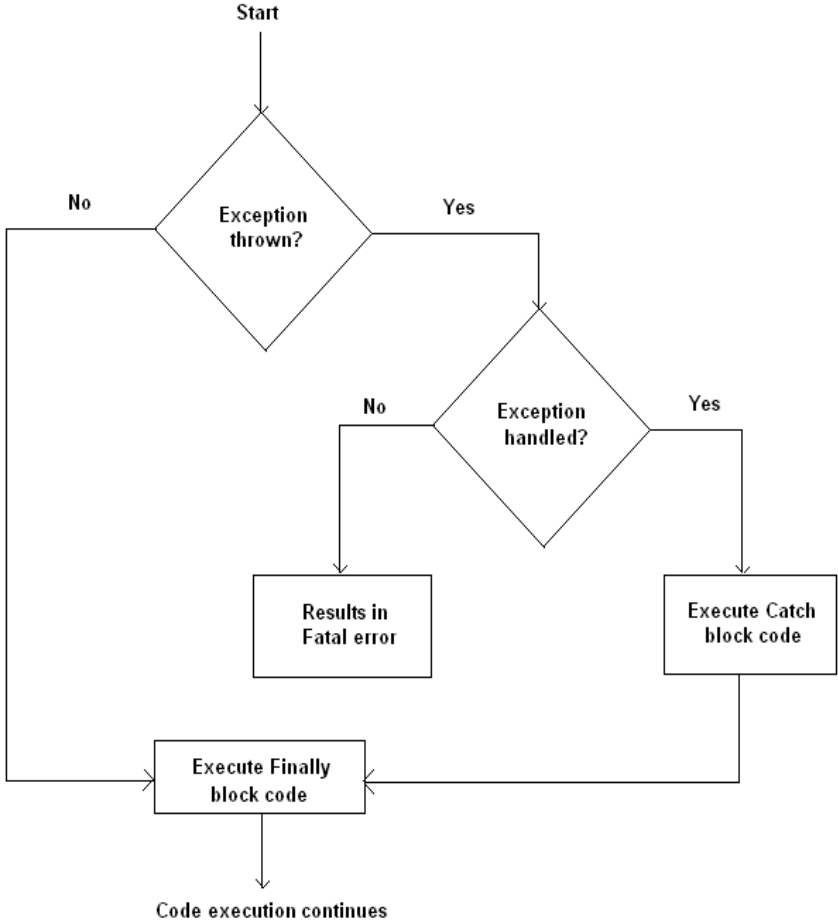## 5.9 Exception Handling explained through block diagram

Figure 5.2: Block diagram of exception Handling

CHAPTER 6

<u>CONCLUSION</u>

To make a report on AWS cloud computing I get know about the need of this technology in today's world. It is very important to manage such a huge amount of data that is prevailing around us and technologies such as AWS are the ones that are exceeding in doing so. They remove the problems of big data and also tells the efficient ways to store, compute and analyze your client's data. I also get to know about the various knew terms and get to know that one can have a great future and scope in this technology. There are a lot more players in the on-request advertise that numerous reports recognize These range from essential foundation contributions (IaaS), through stage support (PaaS) to full applications (SaaS). The drawn out expense of proprietorship may from the start not appear to include, however contemplate different factors, for example, decreased hazard and included worth and for some associations on-request benefits bode well.

Circulated figuring is evolving the way in which IT deparrtments get IT Associations have a scope of ways to the cloud, counting system, stages and applications that are available from cloud providers as on the web organizations. Various people may be dumbfounded by the scope of commitments and the wording utilized to depict them and will be dubious of the peril and points of interest. This display, formed for Computer Weekly, by Bob Tarzey of Quocirca implies to give further information.

CHAPTER 7

<u>FUTURE SCOPE</u>

There are a huge number of conceivable outcomes starting to shape as the fate of distributed computing starts to truly take off.

For example, merchants and specialist organizations can jump aboard to grow new and various methods of offering their merchandise and enterprises to the cloud clients through the cloud innovation. It opens up a totally different stage for architects and web engineers. Organizations and associations can sort out themselves and lead business substantially more moderate and expertly. Long range informal communication and staying in contact with companions gets significantly simpler also. The primary explanation that the eventual fate of distributed computing will be as incredible and sweeping as it predicts to be is that cloud innovation is very helpful. For a certain something, the outrageous readiness and availability of a cloud is far better than the utilization of current innovation. Regardless of where on the planet somebody happens to be, or what gadget they are utilizing, they can get to their cloud and keep on accomplishing their work or offer their data. That, yet cloud innovation is very financially savvy, and an organization could wind up sparing thousands by picking this alternative. For the unwavering quality a cloud offers, the security it gives, and the exhibition it brags of, the expense of a cloud makes it a fantastic choice for people and partnerships the same. The fate of distributed computing is splendid, and savvy individuals of any sort should start to jump aboard with believed distributed computing suppliers like Apprenda.

REFERENCES

1. Achieving the Cloud Computing Vision. (2010). Oracle White Paper in Enterprise Architecture.

2. Anthony T. Velte, T. J. (2010). Cloud Computing a Practical Approach. TATA McGRAW HILL Edition.

3. Antti, P. and Jukka, K. (2010). Energy Efficiency of Mobile Clients in Cloud Computing. HotCloud'10 Proceedings of the 2nd USENIX Conference on Hot Topics in Cloud Computing.

4. Architecture of the Kernel-based Virtual Machine (KVM). Retrieved from: http://www.linux-kongress.org/2010/slides/KVM-Architecture-LK2010.pdf and http://it20.info/2007/06/a-brief-architecture-overview-of-vmware-esx-xenand-ms-viridian/

5. Arora, P., Chaudhry, R. and Ahuja, S. (2012). Cloud Computing Security Issues in Infrastructure as a Service. International Journal of Advanced Research in Computer Science and Software Engineering. Volume 2, Issue 1, ISSN: 2277 128X

6. Badger, L., Grance, T., Robert, P. and Voas, J. Computer Security. DRAFT Cloud Computing Synopsis and Recommendations, NIST Special publication 800-146.

7. Barbara, J., Jo, A., Cynthia, S. and Adam, T. (2009). Collaboration Using Cloud Computing and Traditional Systems. Western Carolina University, Vol. 10, No. 2.

8. Ben, K. (2011). Revolution not Evolution How Cloud Computing Differs from Traditional IT And Why It Matters. Diversity Limited White Paper.

9. CISCO. CISCO Cloud Enablement Services for Adopting Clouds. Company Literature.

10. Cloud billing: The missing link for cloud providers. White Paper. Retrieved from: http://www.cgi.com/files/white-papers/cloud-billing-white-paper.pdf

11. CSA. (2011). Security Guidance for Critical Areas of Focus in Cloud Computing v3.0. 177.

12. Daryl, P. Gartner INC. The Business Landscape of Cloud Computing. Retrieved from http://www.ft.com/cms/5e231aca-a42b-11e1-a701- 00144feabdc0.pdf 256

13. Das, G. (2012). Silver Lining. Business Today. Retrieved from http://businesstoday.intoday.in/story/cloud-computing-create-two-millionnew-jobs-in-india/1/23502.html

14. Deborah, D. and Giacomo, T. (2010). Evaluating Cloud Computing: How It Differs To Traditional IT Outsourcing (Master Thesis in Information Technology and Management). Jonkoping International Business School, Jonkoping University.

15. Dykstra, J. and Alan, T. (2012). Acquiring Evidence form Infrastructure-as-aService Cloud Computing: Exploring The Evaluation Tools, Trust and Techniques. Digital Investigation S90-S98.

16. E. Toews, B. S. (2011). An internationally distributed cloud for science: the cloud-enabled space weather platform. 2nd International Workshop on Software Engineering for Cloud Computing, (pp. 1-7). New York, NY, USA.

17. Erdogmus, H. (2009). Cloud Computing: Does Nirvana Hide Behind the Nebula IEEE Software (26)2, pp. (4–6).

18. Ernst and Young. (2012). Ready for Takeoff-Preparing For Your Journey into the Cloud. Insights on Governance, Risk and Compliance.

19. Ernst and Young. Cloud Computing Issues and Impact. Retrieved from http://www.ey.com/Publication/vwLUAssets/Cloudcomputing_issues_and_impacts/$FILE/Cloud_computing_issues_and_impacts .pdf

20. Firat, A., Andre, C., Wiktoria, W., Martina, Z., Stefan, D. and Andreas, H. (2010). From Cloud Computing to Mobile Internet, From User Focus to Culture and Hedonism. The Crucible of Mobile Health Care and Wellness Applications, 978-1-4244-9142-1/10/$26.00©2010IEEE

21. Flavio, L. (2010). Transparent Security for Cloud. In SAC '10: Proceedings of the 2010 ACM Symposium on Applied Computing.

22. Gerald, J. and Goldberg, R. (1974). Formal Requirements for Virtualizable Third Generation Architectures. Communications of the ACM 17 (7): 412– 421. doi:10.1145/361011.361073

23. Gibson, J., Rondeau, R., Qing, T. (2012). Benefits and Challenges of Three Cloud Computing Service Models. Fourth International Conference on Computational Aspects of Social Networks (CASoN), pp. 198-205, 257

24. Gurudatt Kulkarni, M. J. (2013). Communication As Service Cloud . International Journal of Computer Networking, Wireless and Mobile Communications (IJCNWMC), Vol.3, Issue 1, 150-154.

25. History.com. (2010). The Invention of the Internet.

26. Hofer, C., Karagiannis, G. (2011). Cloud Computing Services: Taxonomy and Comparison. Received: 1 February 2011 / Accepted: 18 May 2011 / Published online: 19 June 2011 © The Author(s) 2011. Retrieved from http://link.springer.com/article/10.1007%2Fs13174-011-0027-x

27. Hong-Linh, T. and Schahram, D. (2010). Cloud Computing for Small Research Groups in Computational Science and Engineering: Current Status and Outlook. ©Springer –Verlag 2010; 25-09-2010.

28. http://www.cisco.com

29. http://apprenda.com/library/software-on-demand/saas-billing-pricing-models/

30. http://arstechnica.com/information-technology/2010/01/googles-mapreducepatent-what-does-it-mean-for-hadoop/

31. http://blog.softlayer.com/2013/virtual-magic-the-cloud/

32. http://developer.yahoo.com/hadoop/tutorial/module2.html#intro

33. http://en.wikipedia.org/wiki/Amazon.com

34. http://en.wikipedia.org/wiki/Computer_network_diagram

35. http://en.wikipedia.org/wiki/File:Hyperviseur.png

36. http://en.wikipedia.org/wiki/File:Point_cloud_torus.gif

37. http://en.wikipedia.org/wiki/File:Telephoneschematic.gif

38. http://en.wikipedia.org/wiki/File:Virtual_Private_Network_overview.svg

39. http://hadoop.apache.org/docs/r1.0.4/hdfs_design.html#Introduction

40. http://huawei.com/ecommunity/bbs/10167597.html

41. http://iaesjournal.com/online/index.php/IJ-CLOSER/article/view/1960

42. http://programmers.stackexchange.com/questions/124345/what-areframeworks-stacks-and-middleware

43. http://redmonk.com/sogrady/2012/05/18/paas-is-the-new-middleware/

44. http://searchcloudprovider.techtarget.com/definition/Eucalyptus

45. http://www. smartinsights.com

46. http://www.ciol.com/ciol/features/141226/six-technology-trends-2013- citrix/page/2

47. http://www.clearcube.com 258

48. http://www.cloudcredential.org/images/pdf_files/nist%20reference%20archite cture.pdf)

49. http://www.computerworld.com/s/article/9190878/Financial_services_firm_tu rning_to_a_private_cloud

50. http://www.ehow.com/how-does_4621707_device-driver-work.html

pres_satakshi