**Assignment 6**

**Big Data Analysis Lab**

Garvin Chanderia

MITU20BTCS0100

2203132

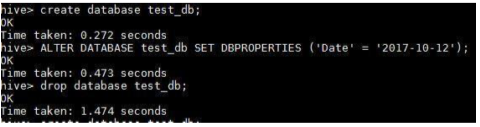
LYISA 1

**Problem Statement:**

Use Hive to create, alter, and drop databases, tables, views, functions, and indexes and create a web application using Hive and Hadoop to demonstrate any real-time application.

**Theory:**

Hadoop Hive is a database framework on the top of Hadoop distributed file systems (HDFS) developed by Facebook to analyse structured data. It supports almost all commands that regular databases support. Hadoop hive create, drop, alter, use database commands are database DDL commands. This article explains these commands with examples.



Hive contains a default database named default.

* Apache Hive is a data warehouse and an ETL tool which provides an SQL-like interface between the user and the Hadoop distributed file system (HDFS) which integrates Hadoop.
* It is built on top of Hadoop. It is a software project that provides data query and analysis.
* It facilitates reading, writing and handling wide datasets that stored in distributed storage and queried by Structure Query Language (SQL) syntax. It is not built for Online Transactional Processing (OLTP) workloads.
* It is frequently used for data warehousing tasks like data encapsulation, Ad hoc Queries, and analysis of huge datasets. It is designed to enhance scalability, extensibility, performance, fault-tolerance and loose-coupling with its input formats.

Features of Hive

These are the following features of Hive:

* + Hive is fast and scalable.
  + It provides SQL-like queries (i.e., HQL) that are implicitly transformed to MapReduce or Spark jobs.
  + It is capable of analyzing large datasets stored in HDFS.
  + It allows different storage types such as plain text, RCFile, and HBase.
  + It uses indexing to accelerate queries.
  + It can operate on compressed data stored in the Hadoop ecosystem.
  + It supports user-defined functions (UDFs) where user can provide its functionality.

## Hadoop Hive SHOW DATABASES commands

This command displays all the databases available in Hive. Below is the example of using show database command:

hive> show databases; OK

default

test\_db

## Hadoop Hive Create Database Command

Hadoop Hive create database is a statement used to create a databases. You can compare Hive database as a namespace in HBase. It is a collection of various tables.

The syntax for Hive create database statement is as follows:

CREATE (DATABASE|SCHEMA) [IF NOT EXISTS] database\_name

[COMMENT 'comment'] [LOCATION 'hdfs\_path']

[WITH DBPROPERTIES (property\_name=property\_value, ...)];

The uses of SCHEMA and DATABASE are interchangeable in Hive – they mean the same thing

Below is the example of create database in Hive:

hive> create database test\_db; OK

Time taken: 0.272 seconds

## Hadoop Hive USE Database Command

USE command sets the current database for all subsequent HiveQL query statements.

Below is the example of USE database command in Hive:

hive> use test\_db; OK

Time taken: 0.225 seconds

hive> use default; OK

Time taken: 0.222 seconds

To revert to the default database, use the keyword name “default” instead of a database name.

## Hadoop Hive Alter Database Command

Hadoop Hive alter database is a statement used to change the properties of a databases in Hive. You can add or remove the database comments, properties etc by using alter database statements.

The syntax for Hive alter database statement is as follows:

ALTER (DATABASE|SCHEMA) database\_name SET DBPROPERTIES

(property\_name=property\_value, ...);

ALTER (DATABASE|SCHEMA) database\_name SET OWNER [USER|ROLE] user\_or\_role;

ALTER (DATABASE|SCHEMA) database\_name SET LOCATION hdfs\_path; Below is the example of alter database in Hive:

hive> ALTER DATABASE test\_db SET DBPROPERTIES ('Date' = '2017-10-12'); OK

Time taken: 0.473 seconds

## Hive Drop Database Command

Hadoop Hive drop database is a statement used to drop the databases in Hive. The syntax for Hive drop database statement is as follows:

DROP (DATABASE|SCHEMA) [IF EXISTS] database\_name [RESTRICT|CASCADE];

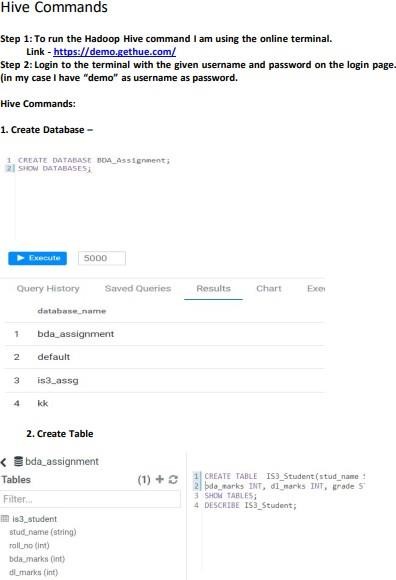
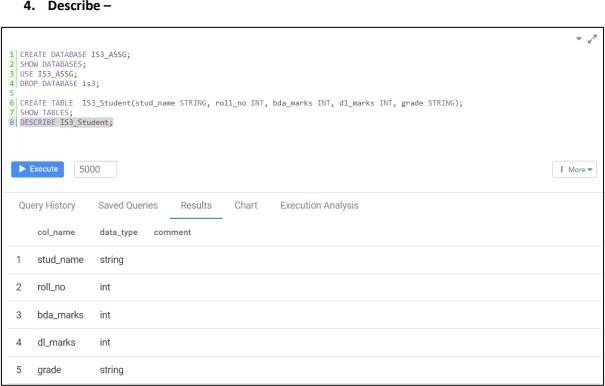
*The default behaviour is RESTRICT, where DROP DATABASE will fail if the database is not empty. To drop the tables in the database as well, use DROP DATABASE … with CASCADE option.*

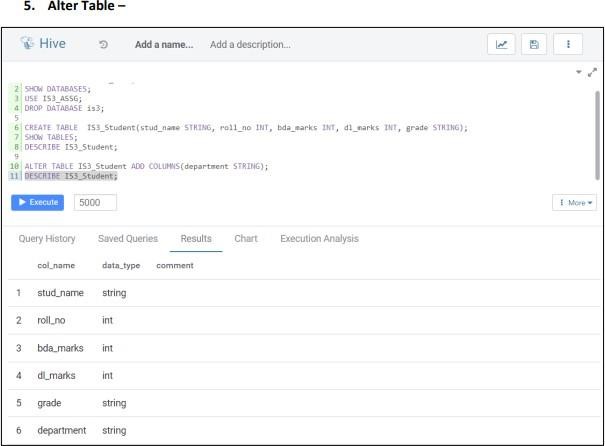
Below is the example of drop database in Hive:

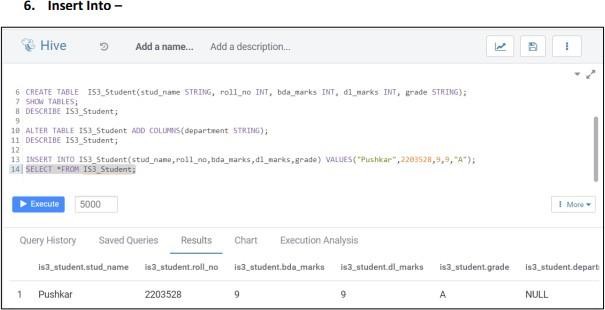
hive> drop database test\_db; OK

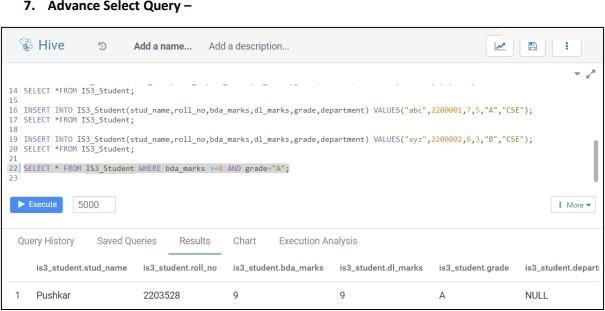
Time taken: 1.474 seconds

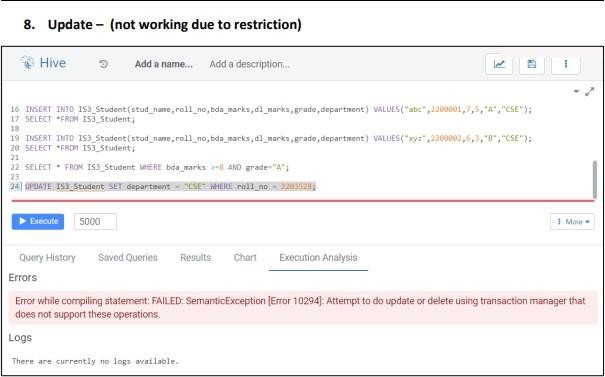
**Code & Output:**

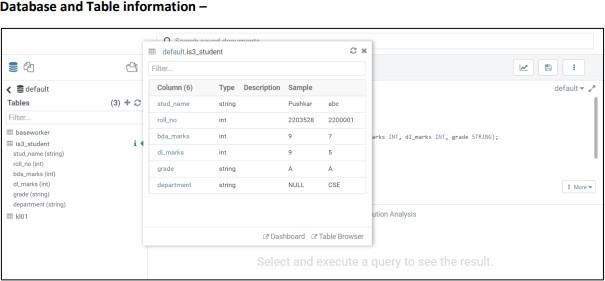






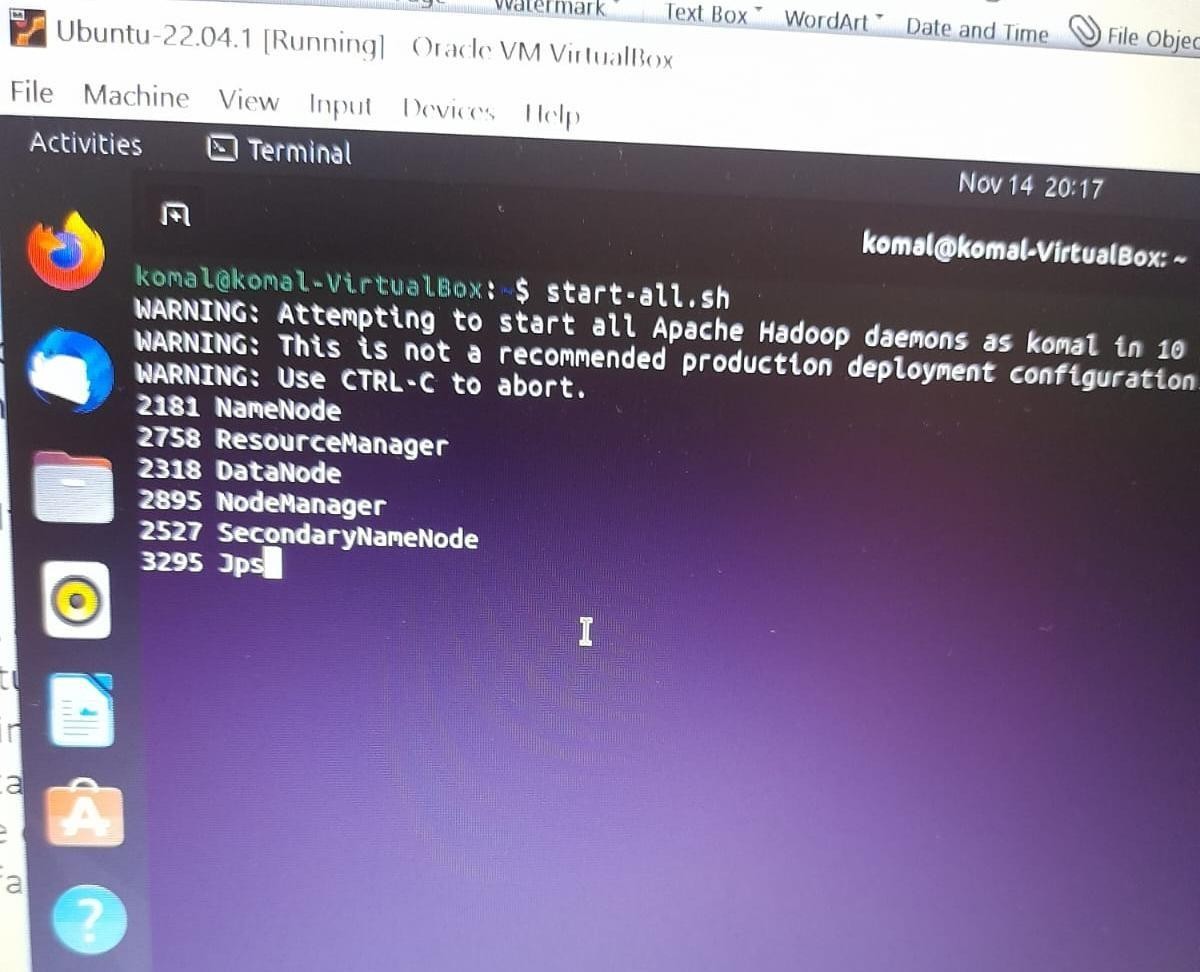
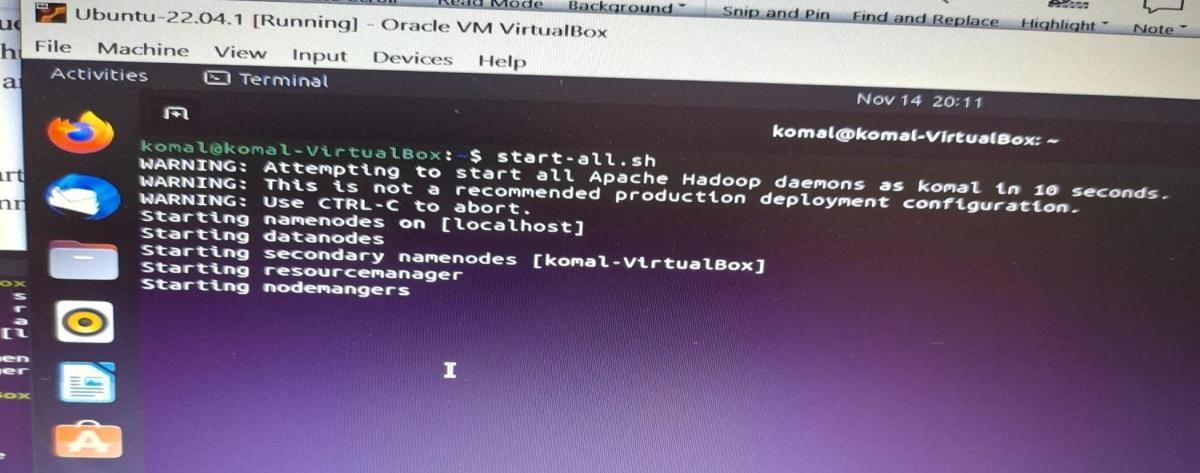




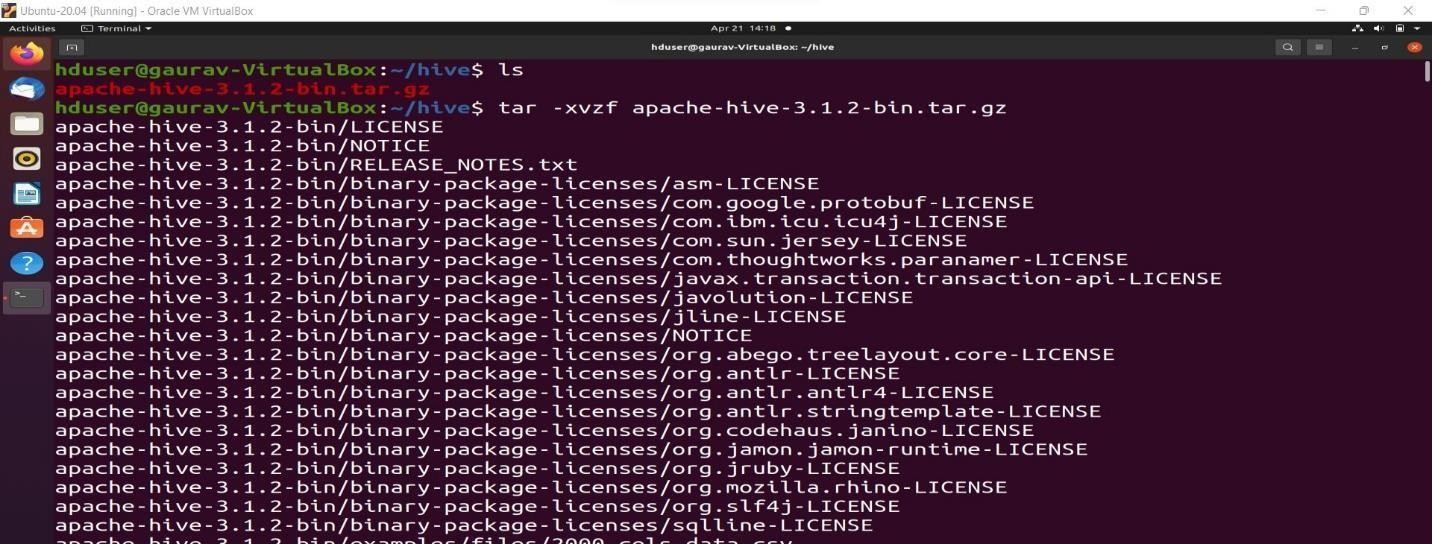


Implementation:

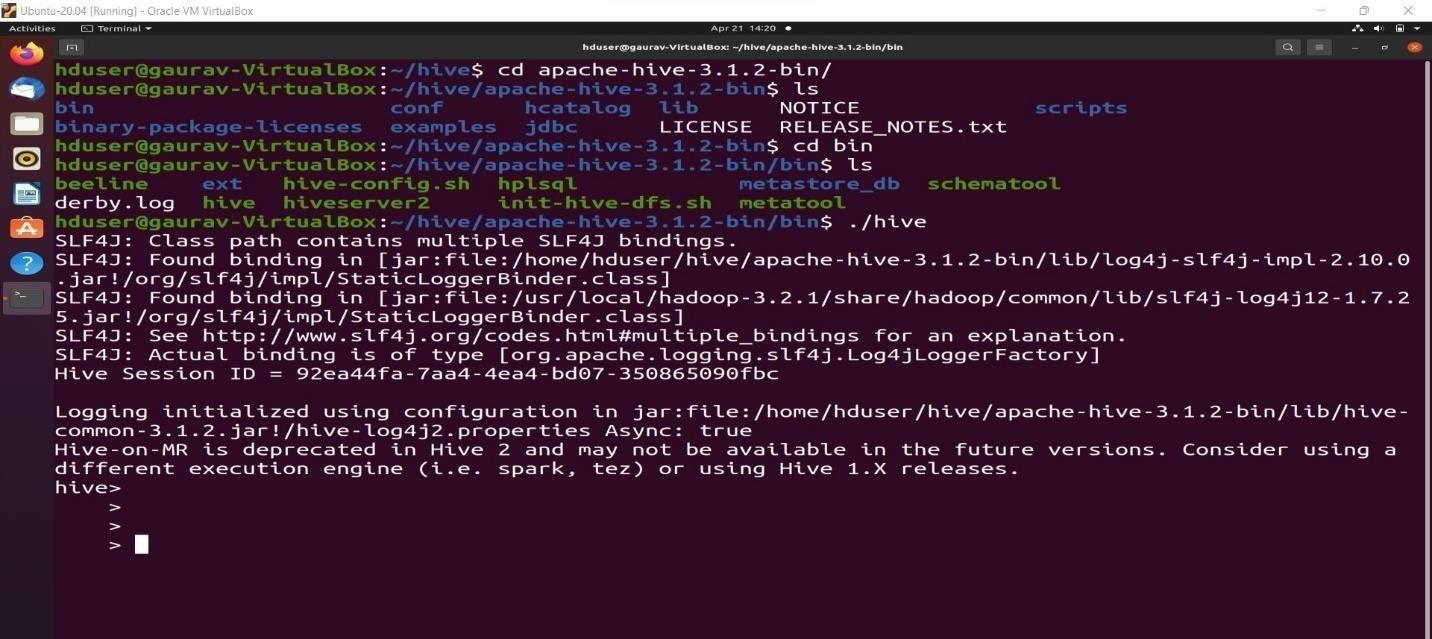
1. At the most we need to start the cluster using the start-all.sh script further check all daemons are running using the jsp command.



1. Once we have the cluster running we also need to download the tar.gz file which we will be using to install the hive in our system. To install the hive we will be just extracting the data in the file tar -xzfs <tar file name>



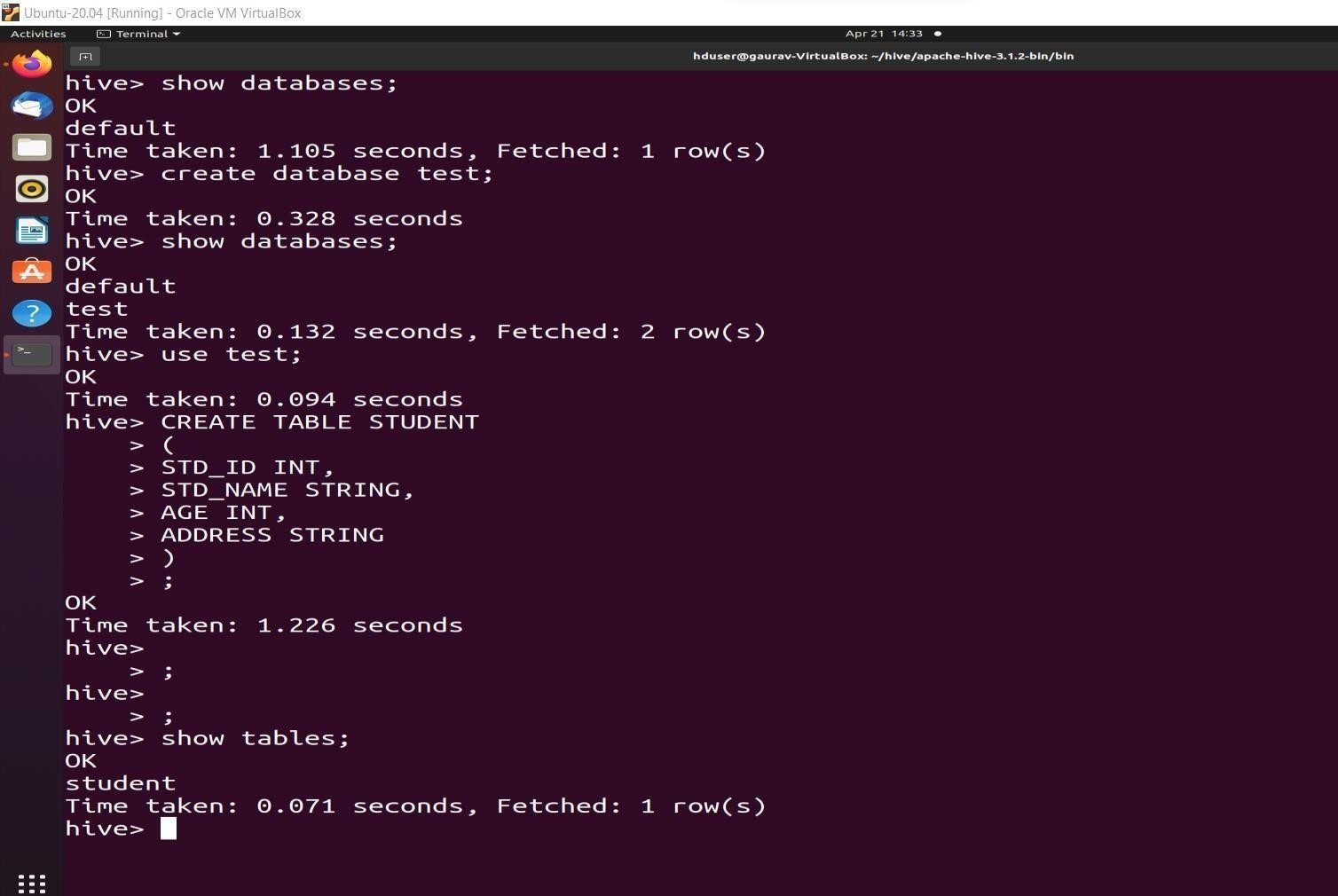
1. Now folder will be visible for Apache-hive in which we have the bin where the executable for the hive is present which will be using to get the hive shell



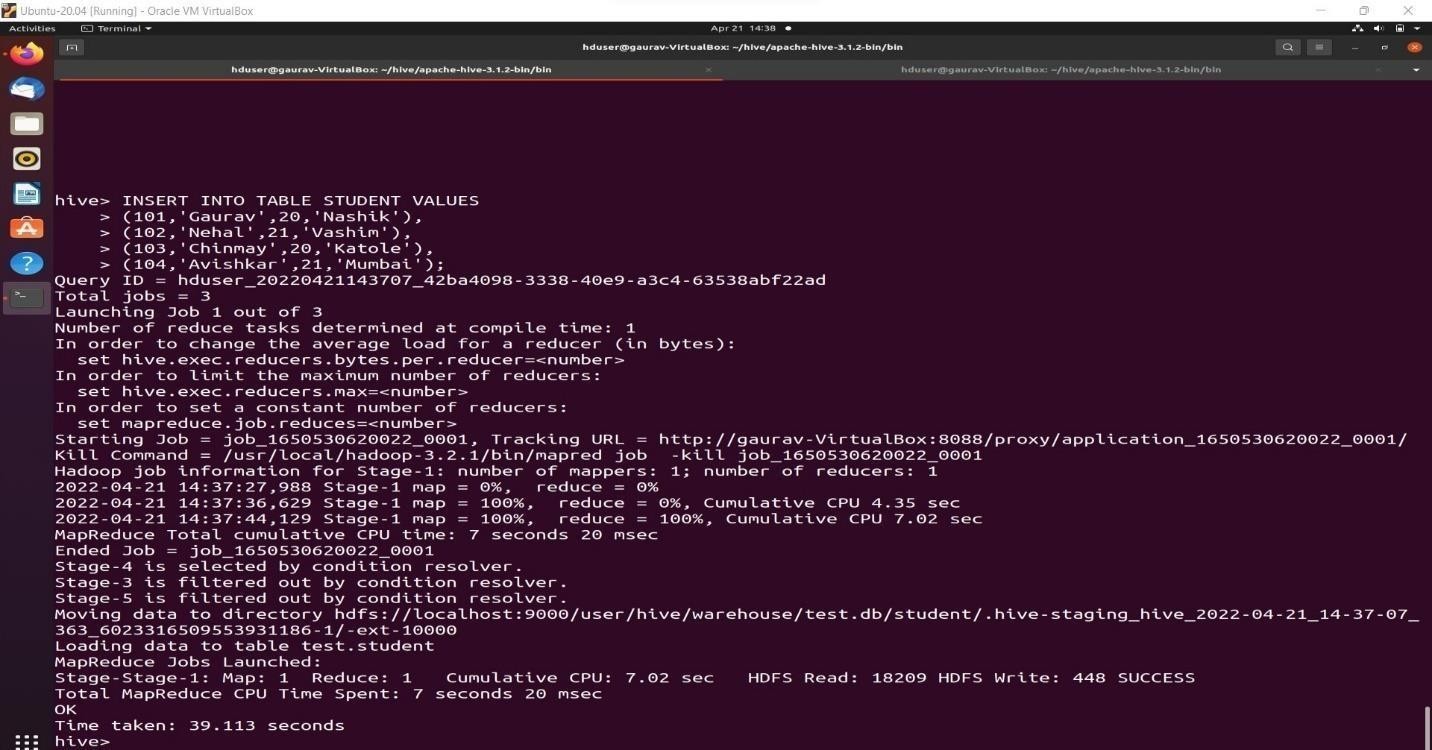
1. Once we get the hive shell ready we will be using the HQL command first we will be creating our database using the use command which will switch us to the new database and if not existed it will create and then switch.

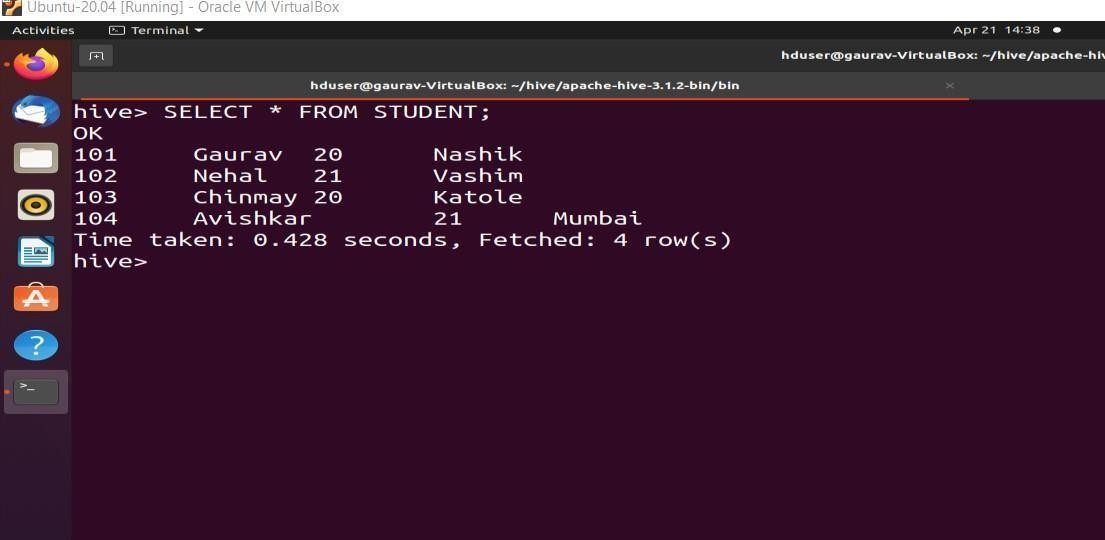
$create database <database\_name>

**$use <database\_name>**



1. We can ask to show the tables using the $show tables command further we can also create the table using the $create table <table\_name> (<column\_name> <type of value>, …) query.
2. We can now insert necessary values in the table using the insert values command.





**Conclusion:**

We have successfully implemented the installation and actions on Hive database.