PRACTICAL-8

AIM:

To Install and Run Hive. Use Hive to create, alter, and drop databases, tables, views, functions, and indexes. To create HDFS tables and load them in Hive and implement joining of tables in Hive

THEORY:

Hive:

• Hive is a data warehouse infrastructure tool to process structured data in Hadoop. It resides on top of Hadoop to summarize Big Data, and makes querying and analyzing easy.

• This is a brief tutorial that provides an introduction on how to use Apache Hive HiveQL with Hadoop Distributed File System. This tutorial can be your first step towards becoming a successful Hadoop Developer with Hive.

• Apache Hive is a data warehouse software project built on top of Apache Hadoop for providing data query and analysis.

• Hive gives an SQL-like interface to query data stored in various databases and file systems that integrate with Hadoop. Traditional SQL queries must be implemented in the MapReduce Java API to execute SQL applications and queries over distributed data. Hive provides the necessary SQL abstraction to integrate SQL-like queries (HiveQL) into the underlying Java without the need to implement queries in the low-level Java API. Since most data warehousing applications work with SQL-based querying languages, Hive aids portability of SQL-based applications to Hadoop.

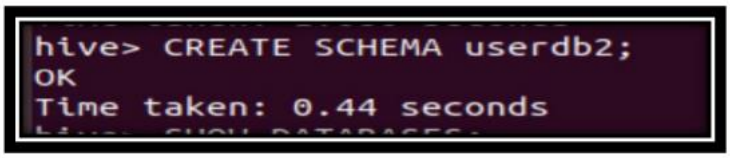
• While initially developed by Facebook, Apache Hive is used and developed by other companies such as Netflix and the Financial Industry Regulatory Authority (FINRA).

• Amazon maintains a software fork of Apache Hive included in Amazon Elastic MapReduce on Amazon Web Services.

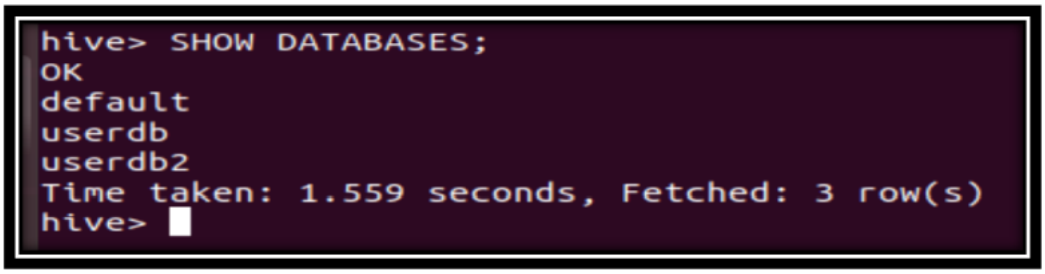
CODE:

CREATE DATABASE [IF NOT EXISTS] userdb;

CREATE SCHEMA userdb;

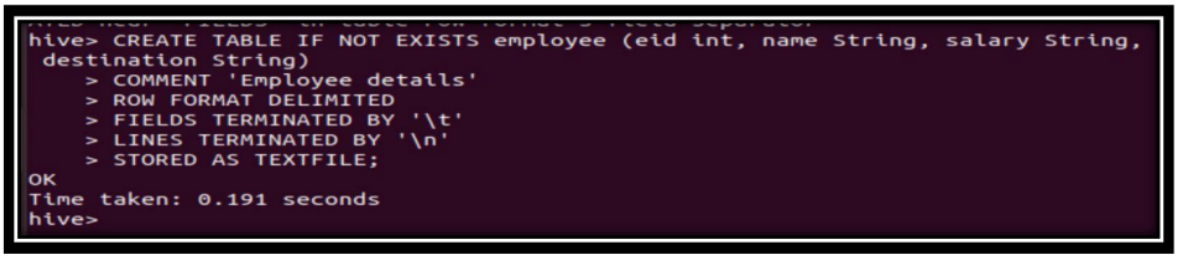


SHOW DATABASES;



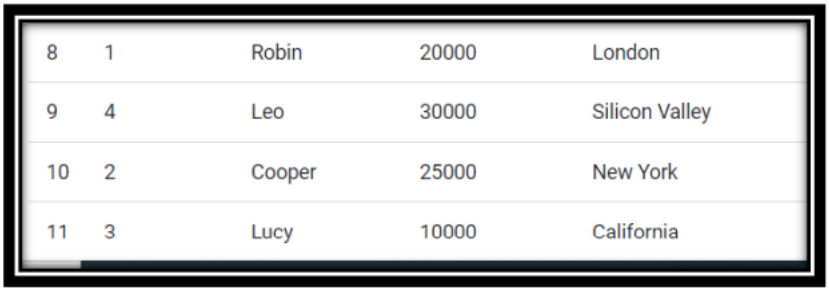
CREATE TABLE IF NOT EXISTS employee (eid int, name String, salary String, destination String) COMMENT ‘Employee details’ ROW FORMAT DELIMITED FIELDS TERMINATED BY ‘\t’ LINES TERMINATED BY ‘\n’

STORED AS TEXTFILE;



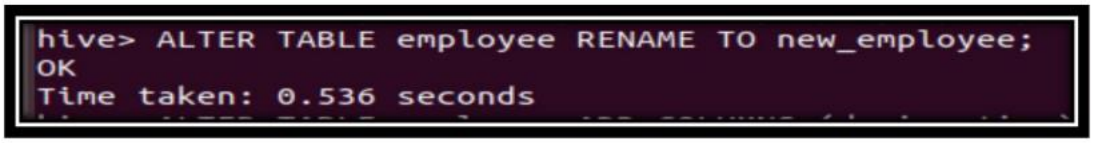
INSERT INTO employee VALUES (1, ‘Robin’, ‘20000’, ‘London’);

SELECT * FROM employee;



8	1	Robin	20000	London
9	4	Leo	30000	Silicon Valley
10	2	Cooper	25000	New York
11	3	Lucy	10000	California

ALTER TABLE name RENAME TO new_name;



ALTER TABLE name ADD COLUMNS (col_spec[, col_spec ...]);

ALTER TABLE name CHANGE column_name new_name new_type;

CONCLUSION:

In this practical, we learnt about hive and performed different operations using it.