Name : Tejas V. Redkar
PRN: 1032210937
RollNo: BDT-2 (22)
Panel-C, Batch-1

BDT Lab Assignment -7

* Problem Statement

Create HIVE Database & perform data analytics on it.

* Objectives:

1) To learn Hive concept
2) To perform data analytics on it.

* Theory:

Explain -
Introduction to HIVE:

HIVE is a data warehousing & SQL-like query language tool that is built on top of the HDFS. It was developed to provide a high-level, user-friendly interface for querying & analyzing large datasets stored in Hadoop. Hive is part of the Apache Hadoop ecosystem & enables users to write SQL-like queries known as HiveQL on top

of Hadoop data.

- Key features of Hive:

1) SQL-like Query Language: Hive provides HiveQL, making it easy for SQL users to work with big data.
2) Scalability - It can handle massive datasets & scales horizontally.
3) Schema on Read: Hive interprets schema when Data is read, allowing for flexibility with data structure.
4) Integration with Hadoop: Hive seamlessly integrates with Hadoop components.
5) Custom UDFs: Users can create custom functions for complex data transformations.

- HIVE commands

- 'CREATE DATABASE': Creates a new db in Hive
- 'USE DATA BASE': sets the current working db.
- 'CREATE TABLE': Defines a new scheme Hive for table.
- 'INSERT INTO TABLE': Inserts data into Hive table
- 'SELECT': Performs data retrival & querying
- 'ALTER TABLE': Modifies the structure of an existing tab.
- ~~'SELECT': Performs data retrieval & querying~~
- 'DROP TABLE': Deletes a table & its data

'DESCRIBE TABLE' : Provides metadata about a table's structure.

* Platform : 64-bit open source Linux/ Windows

* Conclusion : Hence, I learned to create HIVE Database & performed data analytics on it.

* FAQ's

Q1) Who developed HIVE?

Ans HIVE was initially developed by Facebook & later contributed to the Apache software foundation. It is now maintained as an open-source project by the Apache Hive community.

Q2) How to load data in a Hive table?

Ans Hive provides the functionality to load pre-created data table entities either from our local file system or from HDFS. The LOAD DATA statement is used to load data into the hive table

Syntax:

LOAD DATA INPATH '< The table data location'>
INTO TABLE table_name;

Q3) State the difference between HIVE & MySQL

	Hive	MySQL
1)	Hive stores data in HDFS	MySQL stores data in traditional relational databases
2)	Hive uses HiveQL for querying	MySQL uses standard SQL
3)	Hive follows schema-on-read	MySQL follows schema-on-write
4)	Hive is for big data analytics	MySQL for traditional databases.
5)	Hive is highly scalable	MySQL has limitations
6)	Hive is designed for OLAP on large datasets, often slower for real-time transactions	MySQL is optimized for OLAP

21/11/23