Page No.
Date

Devanshu Surana
PC-23, Panel C
BDT 1, Batch 2
10322210755

BDT Lab Assignment 7

Problem statement:
Create Hive Database and 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 and SQL-like query lang tool that is built on top of the HDFS. It was developed to provide a high-level, user-friendly interface for querying and analyzing large datasets stored in Hadoop. Hive is part of the Apache Hadoop ecosystem. and enables users to write SQL-like queries known as HiveSQL 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 and 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 UPFs: Users can create custom functions for complex data transformation.

Hive Commands:

1) 'Create Database' : Creates a new db in Hive.
2) 'Use Database' : Sets the current working db.
3) 'Create Table' : Defines a new schema Hive for table.
4) 'Insert into Table' : Inserts data into Hive table.
5) 'Select' : Performs data retrieval and querying.
6) 'Alter Table' : Modifies structure of an existing table.
7) 'Drop Table' : Deletes a table and its data.
8) 'Describe Table' : Provides metadata about a table's structur

Platform : 64-bit Open Source Linux / Windows.

Conclusion : Hence, I learned to create Hive Database and performed data analytics on it.

FAQ's.

1. Who developed Hive?

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

2. 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;
```

3. State the difference between Hive and MySQL.

Ans.

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