4. Creating Hive tables and queries in Apache Hue

Apache Hue

Apache Hue (Hadoop User Experience) is a web-based interface that makes it easier to work with Apache Hadoop. It provides a graphical user interface (GUI) for interacting with various components of the Hadoop ecosystem, such as Hive, Pig, MapReduce, HBase, and others. Hue simplifies the process of working with Hadoop technologies, making them more accessible to users who may not be comfortable with command-line interfaces.

Key features of Apache Hue include

1. Hive Query Editor: A user-friendly interface for executing Hive queries on Hadoop.
2. Pig Editor: An editor for creating and running Pig scripts.
3. File Browser: Provides a file system browser for HDFS (Hadoop Distributed File System).

Apache Hive

Apache Hive is a data warehouse infrastructure and query language that facilitates the analysis and querying of large datasets stored in Hadoop's distributed file system (HDFS) or other compatible storage systems. It is part of the Apache Hadoop project and provides a high-level, SQL-like interface to interact with big data.

Here are some key aspects of Apache Hive:

1. **SQL-Like Query Language**: Hive Query Language (HQL) is similar to SQL, making it familiar to users who are already accustomed to relational databases. Users can write queries to retrieve, analyze, and process data stored in Hadoop.
2. **Schema-on-Read**: Unlike traditional relational databases that use a schema-on-write approach, Hive follows a schema-on-read model. This means that data can be ingested into Hive without a predefined schema, and the schema is applied at the time of querying.
3. **Integration with Hadoop Ecosystem**: Hive seamlessly integrates with other components of the Hadoop ecosystem, such as HDFS, MapReduce, Apache HBase, and Apache Spark. This allows users to analyze and process data using various tools within the Hadoop ecosystem.

Task 1: Create a database and table in Hive using Hue browser

```
CREATE DATABASE hiveDB;
CREATE TABLE demo
(
Id INT,
Name STRING,
Age INT,
Salary INT
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ‘,’;
```

Task 2: Load the data from the sample.txt into the hive table

```
LOAD DATA INPATH ‘/user/cloudera/sample.txt’
INTO TABLE demo;
```

Task 3: Pull the records where age is more than 20 and salary is less than 40000. (Fill the data accordingly)