Experiment No.8
Create HIVE Database and Descriptive analytics-basic statistics.
Date of Performance:
Date of Submission.

Aim: Create HIVE Database and Descriptive analytics-basic statistics.

Theory:

Hive is a database technology that can define databases and tables to analyze structured data. The theme for structured data analysis is to store the data in a tabular manner, and pass queries to analyze it. This chapter explains how to create Hive database. Hive contains a default database named default.

Create Database Statement

Create Database is a statement used to create a database in Hive. A database in Hive is a namespace or a collection of tables. The syntax for this statement is as follows:

CREATE DATABASE|SCHEMA [IF NOT EXISTS] <database name>

Here, IF NOT EXISTS is an optional clause, which notifies the user that a database with the same name already exists. We can use SCHEMA in place of DATABASE in this command. The following query is executed to create a database named userdb:

hive> CREATE DATABASE [IF NOT EXISTS] userdb;

hive> CREATE SCHEMA userdb;

The following query is used to verify a databases list:

hive> SHOW DATABASES;

default userdb

Program:

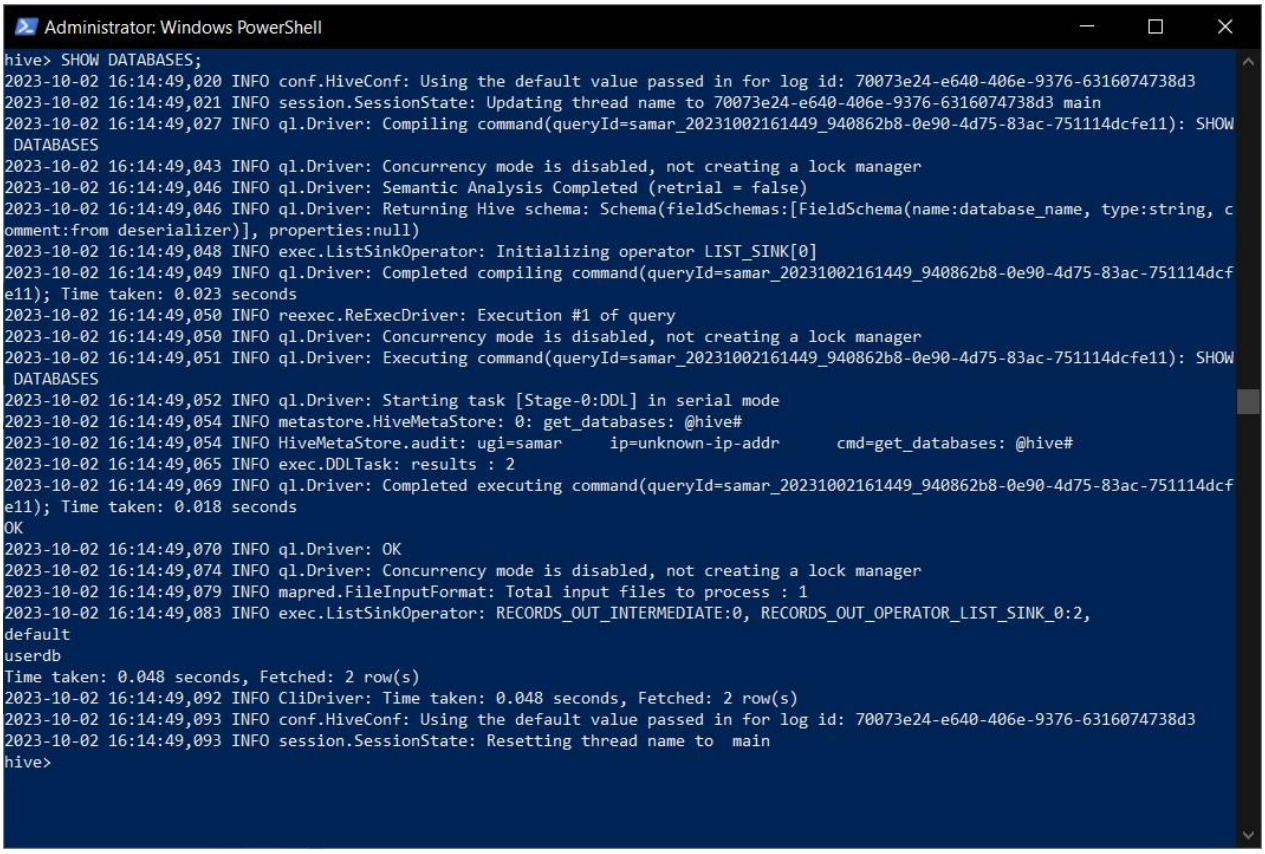
The JDBC program to create a database is given below.

```
import  java.sql.SQLException;
import      java.sql.Connection;
import         java.sql.ResultSet;
import        java.sql.Statement;
```

```
import
java.sql.DriverManager;


public class HiveCreateDb {
   private static String driverName = "org.apache.hadoop.hive.jdbc.HiveDriver";


   public static void main(String[] args) throws SQLException {
      // Register driver and create driver instance


      Class.forName(driverName);
      // get connection


      Connection con = DriverManager.getConnection("jdbc:hive://localhost:10000/default",
"", "");
      Statement stmt = con.createStatement();


      stmt.executeQuery("CREATE DATABASE userdb");
      System.out.println("Database userdb created successfully.");


      con.close();
   }
}
```

Output:

Database userdb created successfully.



CONCLUSION:

The experiment revolved around the establishment of a HIVE database and the application of fundamental statistics for descriptive analytics. We methodically arranged and carefully prepped the data, effectively handling missing values and outliers. Basic statistical indicators, including mean, median, standard deviation, along with data visualization techniques, played a pivotal role in summarizing data trends and rendering them visually. These efforts yielded valuable insights that could inform decision-making and serve as a foundation for further analysis. While basic statistics provide initial insights, more advanced analytical methods may become necessary. Consistent data quality monitoring remains essential. This experiment underscores the pivotal role of proper data management and analysis in facilitating informed decision-making, applicable across a wide range of industries.