Experiment No.5
Create HIVE Database and Descriptive analytics-basic statistics.
Date of Performance: 14/08/2023
Date of Submission: 21/08/2023

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:

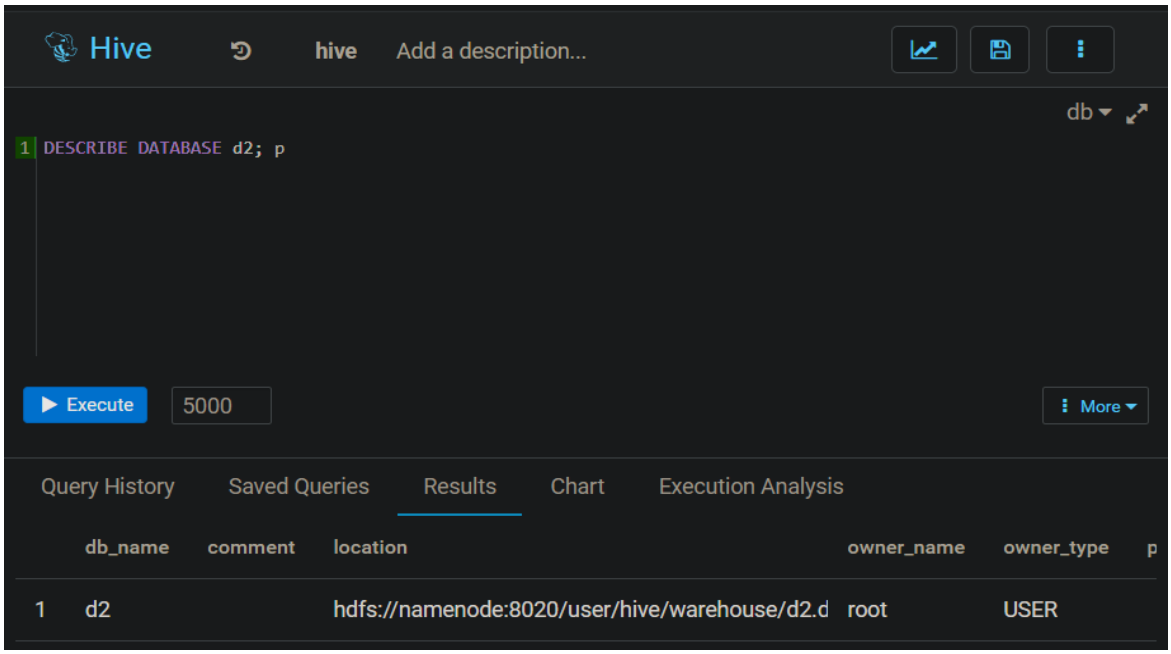
The JDBC program to create a database is given below.

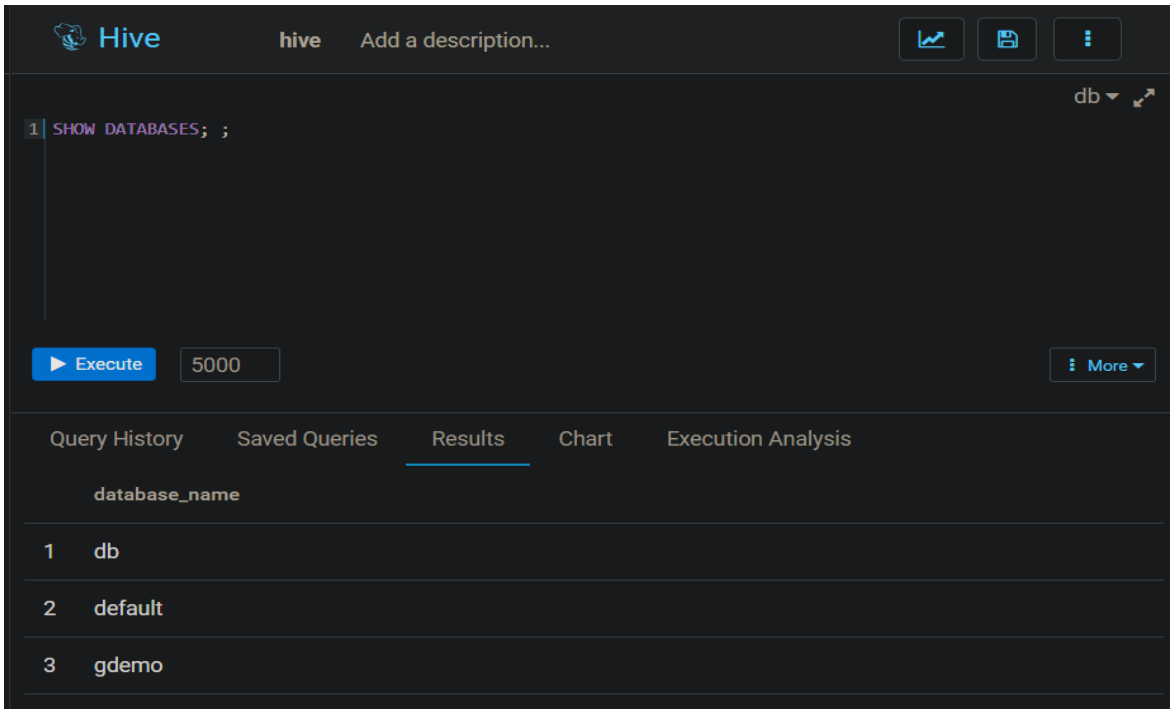
```
Import java.sql.SQLException;

import java.sql.Connection;

import java.sql.ResultSet;

import java.sql.Statemnt;

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

 Connectioncon=
DriverManager.getConnection("jdbc:hive://localhost:10000/default","", "");

Statement stmt = con.createStatement();

stmt.executeQuery("CREATE DATABASE userdb");

System.out.println("Database userdb created successfully.");

 Console();

}
```

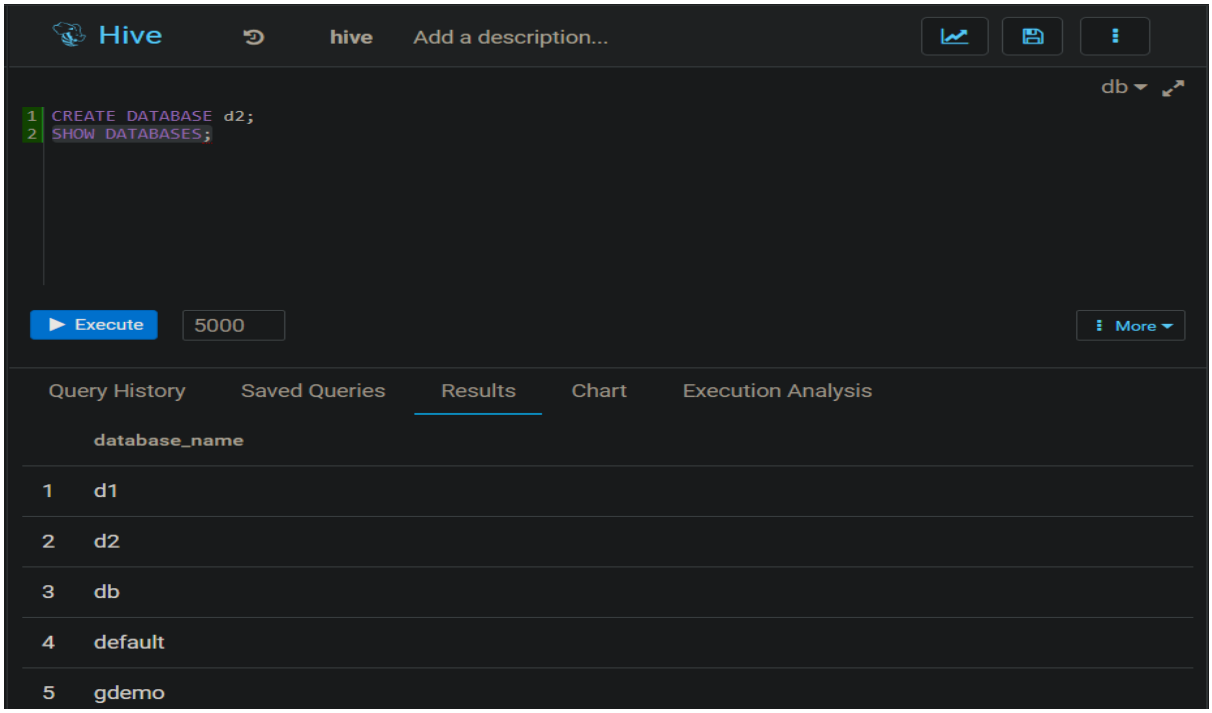
}

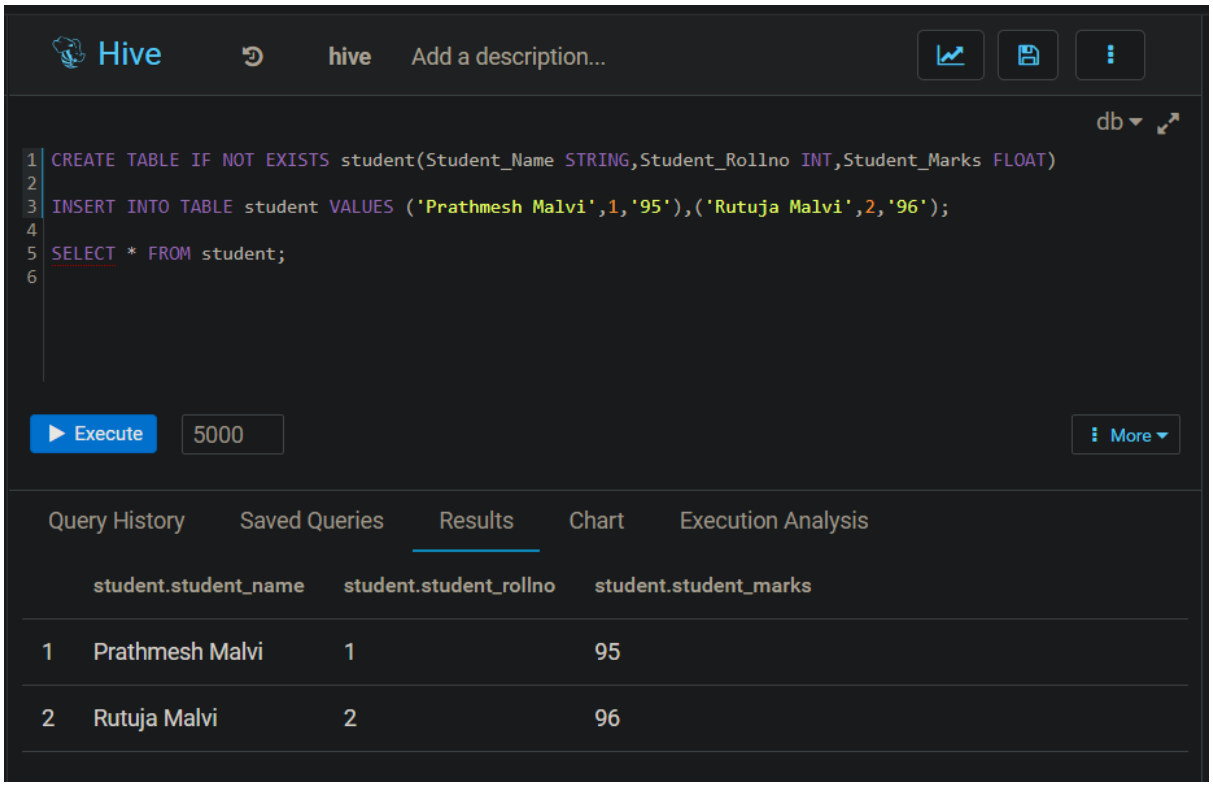
Output:

Database userdb created successfully.




Hive
hive
Add a description...
db
1 SHOW DATABASES; ;
Execute
5000
More
Query History
Saved Queries
Results
Chart
Execution Analysis
database_name
1 db
2 default
3 gdemo


Hive
hive
Add a description...
db
1 CREATE DATABASE d2;
2 SHOW DATABASES;
Execute
5000
More
Query History
Saved Queries
Results
Chart
Execution Analysis
database_name
1 d1
2 d2
3 db
4 default
5 gdemo



Conclusion:

Hive is an open-source data warehousing and SQL-like query language system designed for processing and analyzing large datasets in a distributed computing environment. Hive is commonly used for data warehousing, ad-hoc querying, log analysis, and large-scale data processing in applications like business intelligence, reporting, and data exploration. It abstracts the complexities of distributed data processing and allows users to work with big data using familiar SQL-like syntax.

In this experiment, we successfully created hive database and executed basic hive queries.