



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.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.");

        Console();

    }

}
```



}

## Output:

Database userdb created successfully.

The image shows the Hive web interface. The query entered is `DESCRIBE DATABASE d2; p`. The results tab is active, displaying a table with columns: `db_name`, `comment`, `location`, `owner_name`, `owner_type`, and `p`. The results show one entry for database `d2`.

	db_name	comment	location	owner_name	owner_type	p
1	d2		hdfs://namenode:8020/user/hive/warehouse/d2.d	root	USER	



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



The screenshot shows the Hive web interface. At the top, there's a header with the Hive logo, a 'hive' label, and an 'Add a description...' button. Below the header, a SQL query is entered in a text area:

```
1 CREATE TABLE IF NOT EXISTS student(Student_Name STRING,Student_Rollno INT,Student_Marks FLOAT)
2
3 INSERT INTO TABLE student VALUES ('Prathmesh Malvi',1,'95'),('Rutuja Malvi',2,'96');
4
5 SELECT * FROM student;
6
```

Below the query, there's an 'Execute' button and a '5000' character limit. To the right, there's a 'More' button. Below the query area, there's a tabbed interface with 'Query History', 'Saved Queries', 'Results', 'Chart', and 'Execution Analysis'. The 'Results' tab is selected, showing a table with the following data:

	student.student_name	student.student_rollno	student.student_marks
1	Prathmesh Malvi	1	95
2	Rutuja Malvi	2	96

## 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.