



Vidyavardhini's College of Engineering & Technology

Department of Computer Engineering

Experiment No. 8
Create HIVE Database and Descriptive analytics-basic statistics
Date of Performance: 25/09/2023
Date of Submission: 11/10/2023



Vidyavardhini's College of Engineering & Technology

Department of Computer Engineering

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;
```



Vidyavardhini's College of Engineering & Technology

Department of Computer Engineering

```
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();

    }

}
```



Vidyavardhini's College of Engineering & Technology

Department of Computer Engineering

Output:

Database userdb created successfully.

```
hive> SHOW DATABASES;
2023-10-02 16:14:49,020 INFO conf.HiveConf: Using the default value passed in for log id: 70073e24-e640-406e-9376-6316074738d3
2023-10-02 16:14:49,021 INFO session.SessionState: Updating thread name to 70073e24-e640-406e-9376-6316074738d3 main
2023-10-02 16:14:49,027 INFO ql.Driver: Compiling command(queryId=samar_20231002161449_940862b8-0e90-4d75-83ac-751114dcfe11): SHOW
DATABASES
2023-10-02 16:14:49,043 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2023-10-02 16:14:49,046 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2023-10-02 16:14:49,046 INFO ql.Driver: Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:database_name, type:string, c
omment:from deserializer)], properties:null)
2023-10-02 16:14:49,048 INFO exec.ListSinkOperator: Initializing operator LIST_SINK[0]
2023-10-02 16:14:49,049 INFO ql.Driver: Completed compiling command(queryId=samar_20231002161449_940862b8-0e90-4d75-83ac-751114dcfe11); Time taken: 0.023 seconds
2023-10-02 16:14:49,050 INFO reexec.ReExecDriver: Execution #1 of query
2023-10-02 16:14:49,050 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2023-10-02 16:14:49,051 INFO ql.Driver: Executing command(queryId=samar_20231002161449_940862b8-0e90-4d75-83ac-751114dcfe11): SHOW
DATABASES
2023-10-02 16:14:49,052 INFO ql.Driver: Starting task [Stage-0:DDL] in serial mode
2023-10-02 16:14:49,054 INFO metastore.HiveMetaStore: 0: get_databases: @hive#
2023-10-02 16:14:49,054 INFO HiveMetaStore.audit: ugi=samar ip=unknown-ip-addr cmd=get_databases: @hive#
2023-10-02 16:14:49,065 INFO exec.DDLTask: results : 2
2023-10-02 16:14:49,069 INFO ql.Driver: Completed executing command(queryId=samar_20231002161449_940862b8-0e90-4d75-83ac-751114dcfe11); Time taken: 0.018 seconds
OK
2023-10-02 16:14:49,070 INFO ql.Driver: OK
2023-10-02 16:14:49,074 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2023-10-02 16:14:49,079 INFO mapred.FileInputFormat: Total input files to process : 1
2023-10-02 16:14:49,083 INFO exec.ListSinkOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_LIST_SINK_0:2,
default
userdb
Time taken: 0.048 seconds, Fetched: 2 row(s)
2023-10-02 16:14:49,092 INFO CliDriver: Time taken: 0.048 seconds, Fetched: 2 row(s)
2023-10-02 16:14:49,093 INFO conf.HiveConf: Using the default value passed in for log id: 70073e24-e640-406e-9376-6316074738d3
2023-10-02 16:14:49,093 INFO session.SessionState: Resetting thread name to main
hive>
```

Conclusion:

In order to prepare structured data for analysis, this project was created to demonstrate how to establish a Hive database. Hive is a powerful tool for managing tabular data and running queries on it that resemble SQL queries. In addition to showing how to build a database in Hive, the experiment also emphasized the "IF NOT EXISTS" clause, which is optional and can be used to prevent the creation of duplicate databases. This is an experiment that will be useful to all Hive users since it lays the foundation for organized data management and storage, which will help with subsequent tasks including descriptive analytics and data processing within the Hadoop environment.