**Ex no: 5**

**Create tables in Hive and write queries to access the data in the table**  **Aim:**

To create tables in Hive and write queries to access the data in the table.

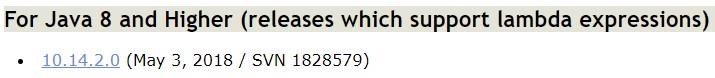
**Procedure:**

**Hive Download and installation:**

1. Hive Installation setup:

- Download and install Apache Derby version 10.14.2.0:

[https://db.apache.org/derby/derby\_downloads.html#For+Java+8+and+Higher](https://db.apache.org/derby/derby_downloads.html" \l "For+Java+8+and+Higher)



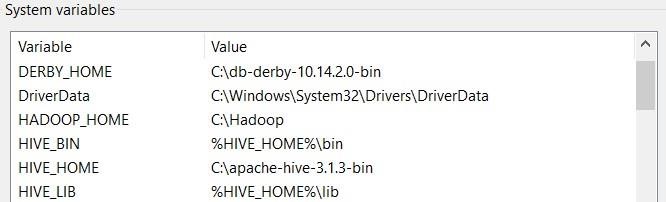
-Download and install Apache Hive version 3.1.3:

<https://downloads.apache.org/hive/hive-3.1.3/>



1. Add environment variables:

Environment variables > System variables > Add the below paths





-



>



(



Inside Path



)



1. Copy Derby libraries:

Go to the Derby libraries directory (db-derby-10.14.2.0\lib) and copy all \*.jar files. Then, paste them within the Hive libraries directory.

1. Configuring hive-site.xml and Hive’s Bin folder:

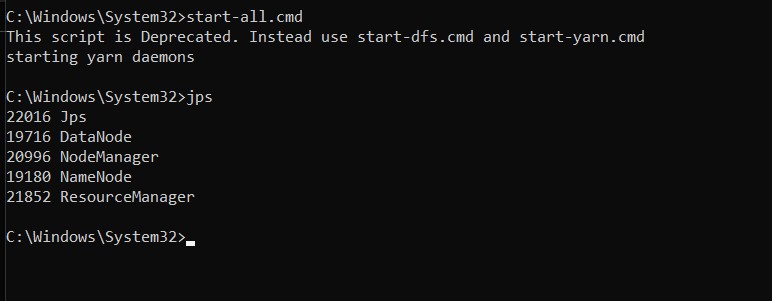
Refer following link to download the file. Also download the guava file. Put hive-site.xml file to hive’s conf location and replace hive’s current guava file with this one in lib location. Also download the bin folder from link and replace the existing hive’s bin folder.

https://1drv.ms/f/s!ArSg3Xpur4Grmw0SDqW0g44T7HYU?e=wDsoBn

1. Starting Hadoop Services

Open PowerShell as administrator and go to Hadoop sbin directory and start hadoop services using the following commands:

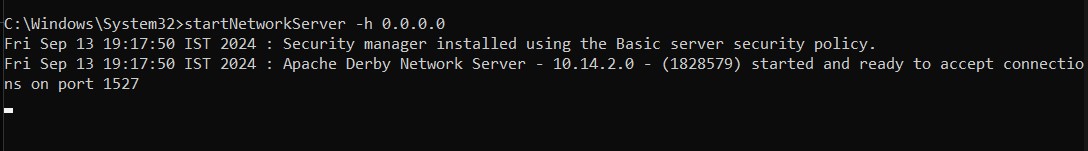
# Start-all.cmd



6. Derby Network Server:

Open another PowerShell window and run the following command to open Derby:

# StartNetworkServer -h 0.0.0.0

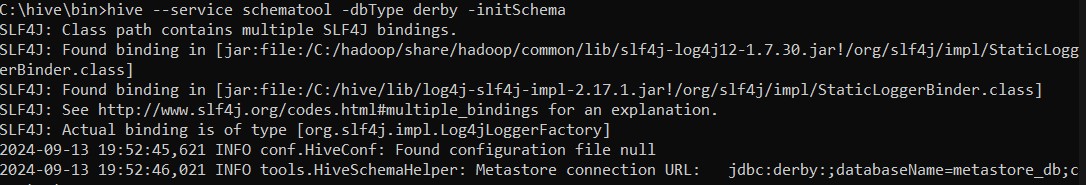


Go to first PowerShell window and check whether NetworkServerControl is running.



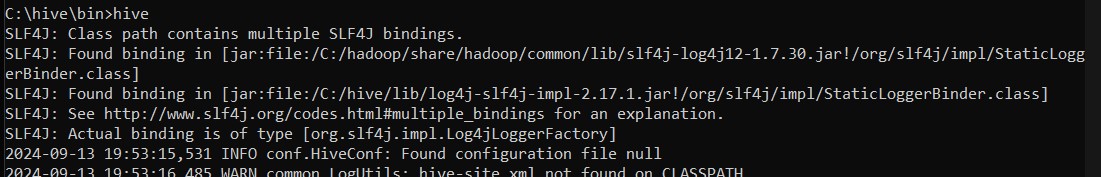
1. Starting Apache Hive:

Go to Apache Hive’s bin location with cd command and run the following command: hive --service schematool -dbType derby –initSchema



1. Open Hive shell by typing:

# hive



**Create a Database:**

Start by creating a database. Open the Hive CLI and follow the steps below:

1. Use the **CREATE DATABASE** statement to create a new database:

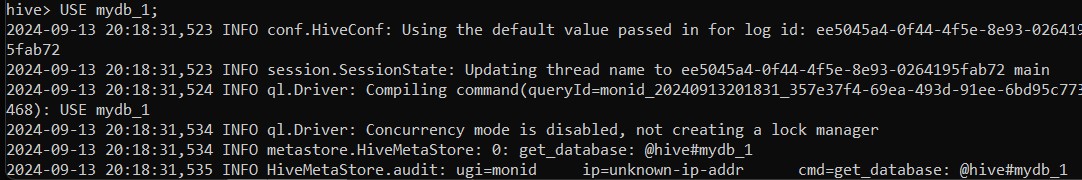
CREATE DATABASE mydb\_1;

1. Verify the database is present:

SHOW DATABASES;

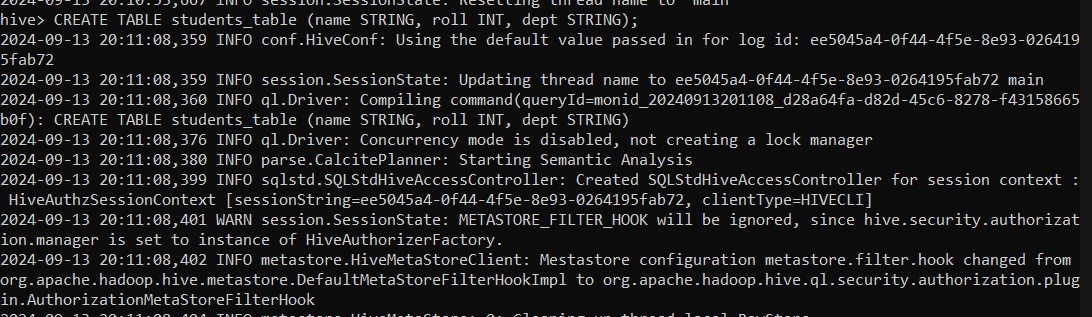
1. Switch to the new database:

# USE mydata;

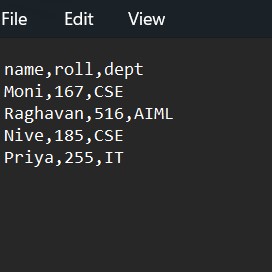


**Create a Table in Hive:**

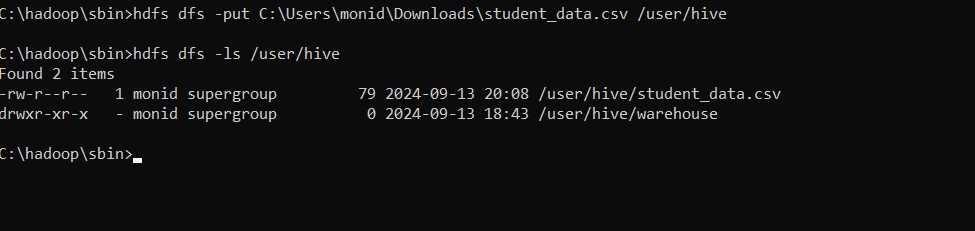
# CREATE TABLE students\_table (name STRING, roll INT, dept STRING);



**Create a .csv file:**



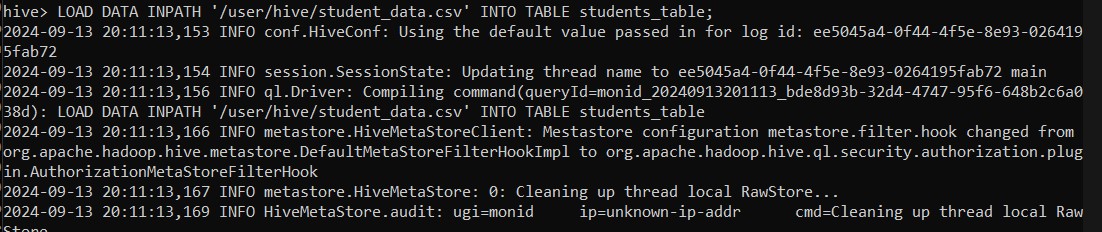
**Add it to hadoop using –put command:**



**Add Data to the TABLE:**

Run the **LOAD DATA LOCAL INPATH** command:

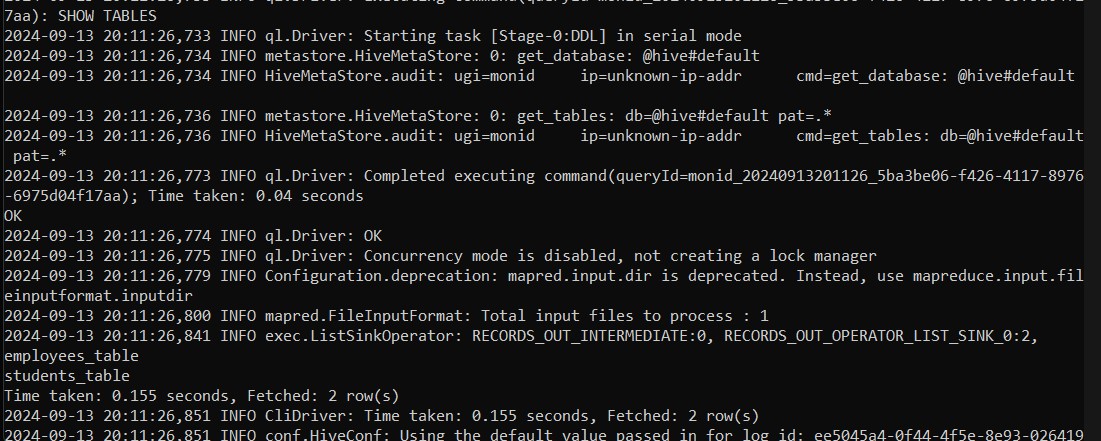
# LOAD DATA INPATH '/user/hive/student\_data.csv' INTO TABLE students\_table;



**List Hive Tables and Data:**

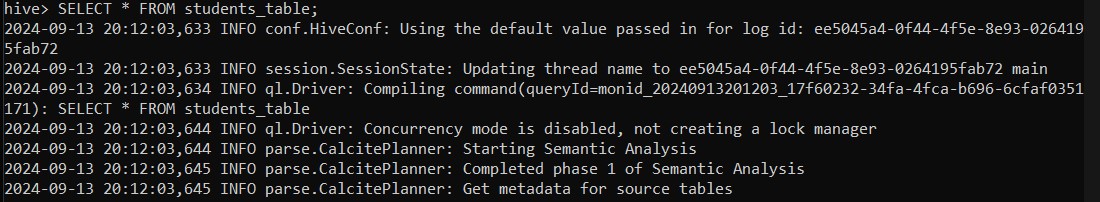
To show all tables in a selected database, use the following statement:

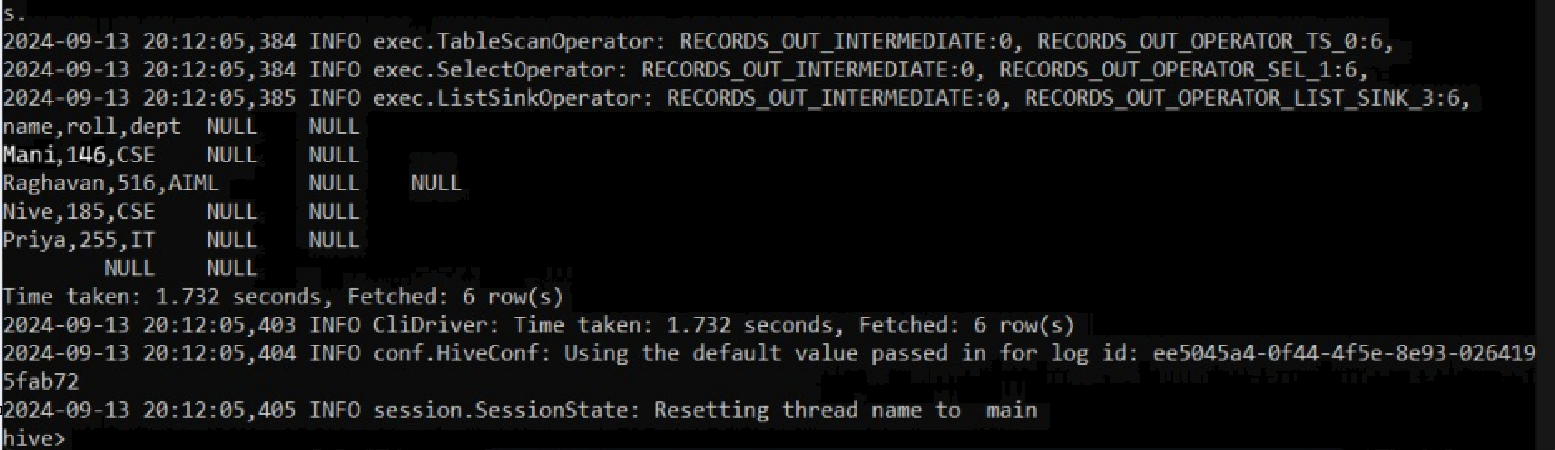
# SHOW TABLES;



To display table data, use a **SELECT** statement. For example, to select everything in a table, run:

# SELECT \* FROM students;





**Result:**

Thus, to create tables in Hive and write queries to access the data in the table was completed successfully.