EX.NO.5

210701307 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:

Installation of Hive:

1. Installing Apache Derby

Install Apache Derby 10.14.2.0

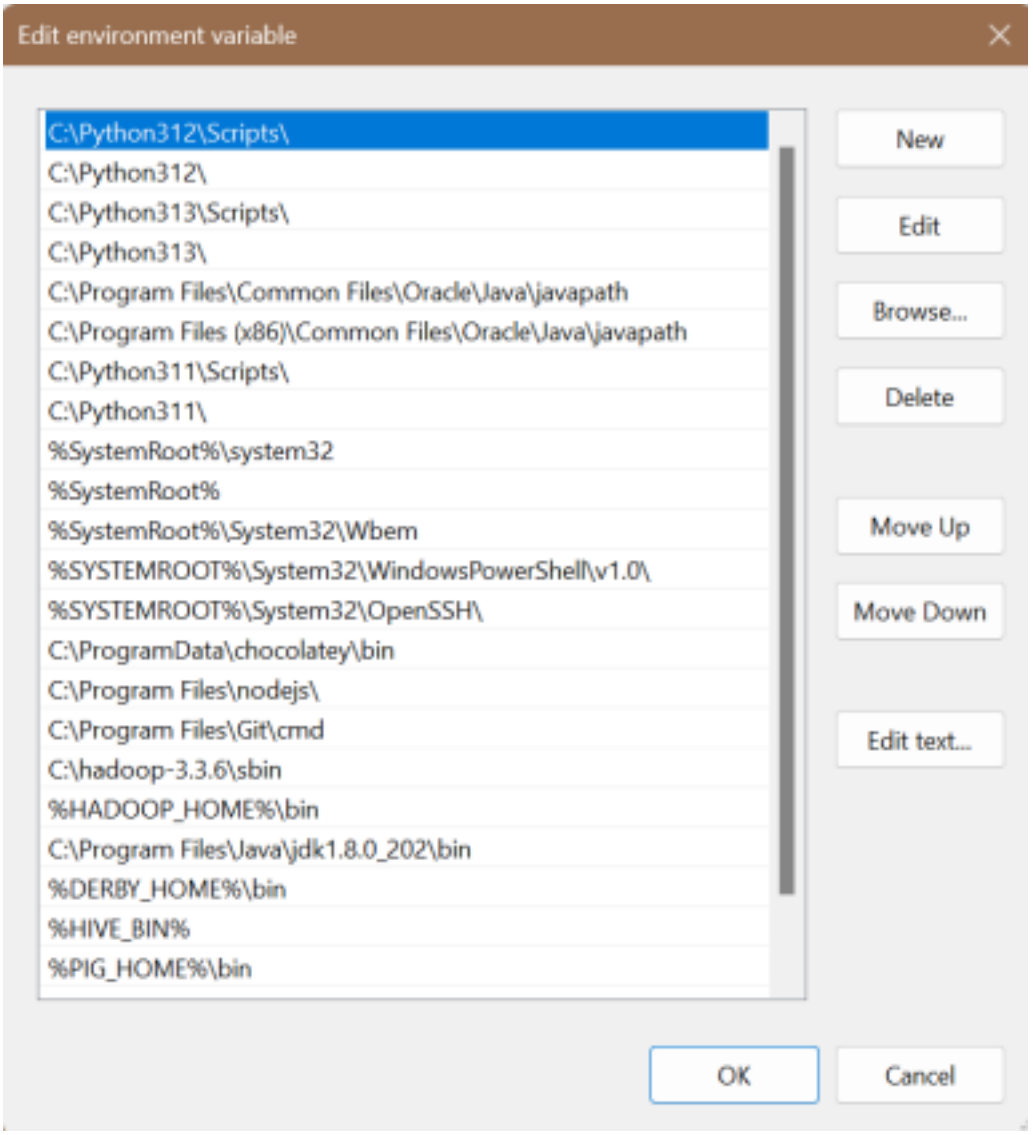
https://db.apache.org/derby/derby_downloads.html#For+Java+8+and+Higher

2. Downloading Apache Hive binaries

https://downloads.apache.org/hive/hive-3.1.2/

3. Setting environment variables

Variable	Value
HADOOP_HOME	C:\hadoop-3.3.6
HIVE_BIN	%HIVE_HOME%\bin
HIVE_HOME	C:\apache-hive-3.1.3-bin
HIVE_LIB	%HIVE_HOME%\lib
DERBY_HOME	C:\db-derby-10.14.2.0-bin



3.1. Copy Derby libraries

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

3.2. 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

4. Starting Hadoop Services

start-all.cmd

5. Derby Network Server: Run the following command in separate window to open

Derby StartNetworkServer -h 0.0.0.0

6. Starting Apache Hive

Go to Apache Hive's bin location with cd command and run the following command:

hive --service schematool -dbType derby -initSchema

7. 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 sample;

2. Verify the database is present:

SHOW DATABASES;

3. Switch to the new database:

USE sample;

Create a Table in Hive

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

Add Data

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

LOAD DATA INPATH '/C:/Users/mercy/OneDrive/Documents/DataAnalytics/Hive/students.csv' WRITE INTO TABLE students;

List Hive Tables and Data

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

SHOW TABLES;

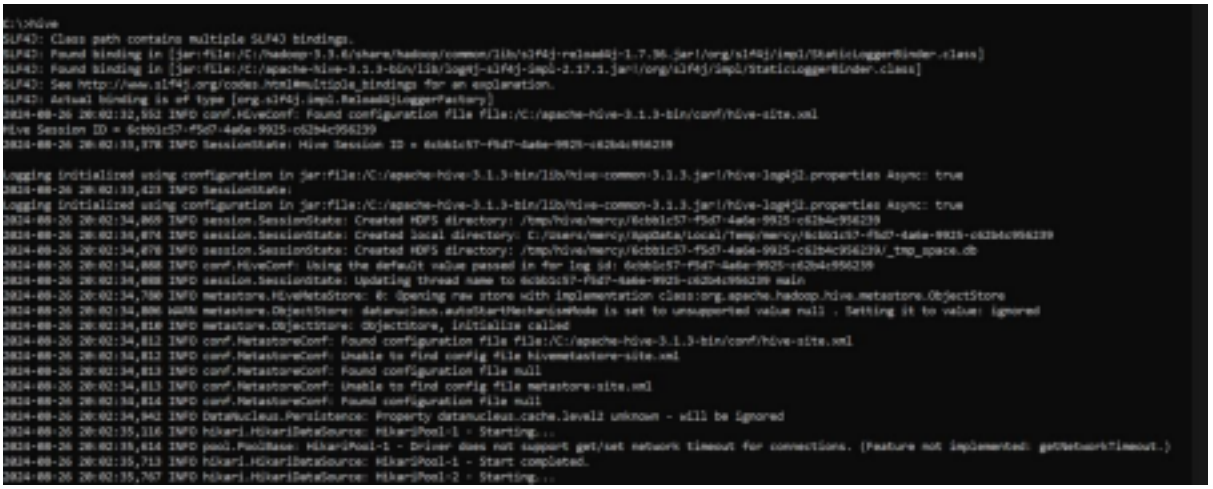
To show table column names and data types, run:

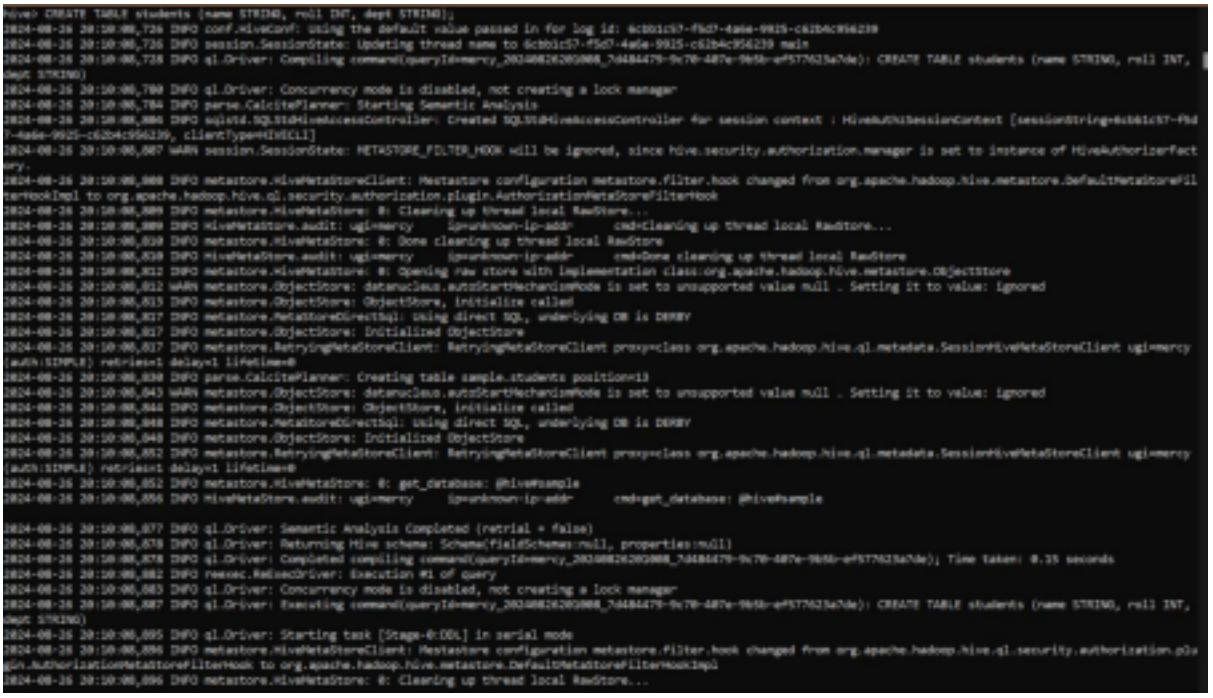
DESC students;

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

SELECT * FROM students;

OUTPUT:

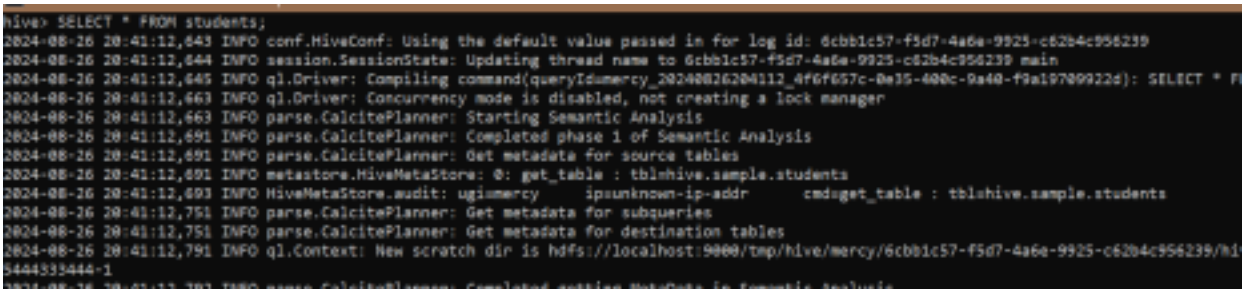


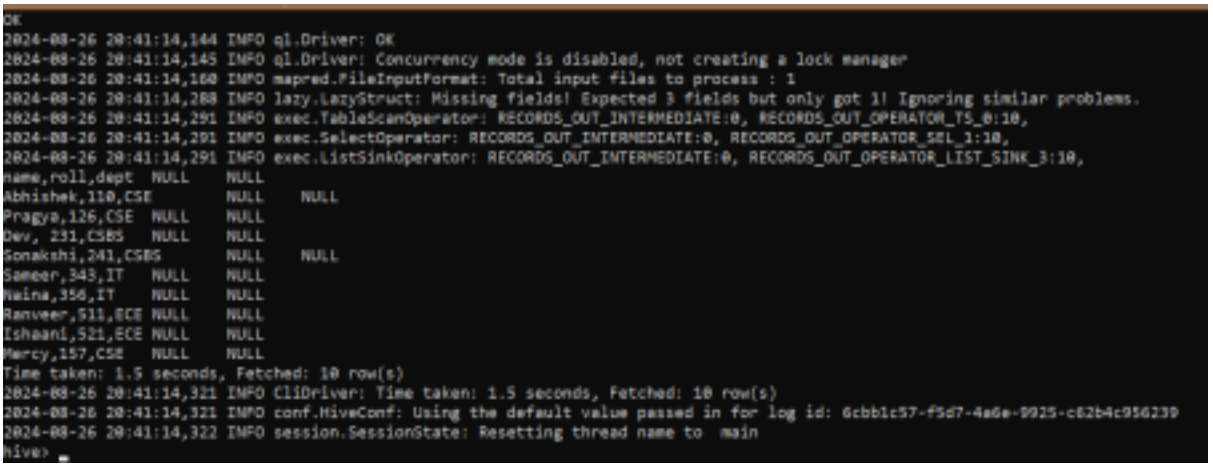


```
hive> LOAD DATA LOCAL INPATH '/C:/Users/mercy/OneDrive/Documents/DataAnalytics/Hive/students.csv' OVERWRITE INTO TABLE students;
2024-08-26 20:39:08,846 INFO conf.HiveConf: Using the default value passed in for log id: 6cbb1c57-f5d7-4a6e-9925-c62b4c956239
2024-08-26 20:39:08,846 INFO session.SessionState: Updating thread name to 6cbb1c57-f5d7-4a6e-9925-c62b4c956239 main
2024-08-26 20:39:08,849 INFO ql.Driver: Compiling command(queryId=mercy_20240826203908_2518a7ae-a17b-44a3-a9dd-7ec2ce964eba): LOAD DATA LOCAL INPATH '/C:/Users/mercy/OneDri
ve/Documents/DataAnalytics/Hive/students.csv' OVERWRITE INTO TABLE students
2024-08-26 20:39:08,867 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-08-26 20:39:08,867 INFO metastore.HiveMetaStore: 0: get_table : tbl=hive.sample.students
2024-08-26 20:39:08,868 INFO HiveMetaStore.audit: ugi=mercy     ip=unknown-ip-addr      cmd=get_table : tbl=hive.sample.students
2024-08-26 20:39:08,975 INFO io.HiveFileFormatUtils: Failed to check mode for C:\Users\mercy\OneDrive\Documents\DataAnalytics\Hive\students.csv: View 'unix' not available (
class java.lang.UnsupportedOperationException)
2024-08-26 20:39:09,038 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2024-08-26 20:39:09,039 INFO ql.Driver: Returning Hive schema: Schema(fieldSchemas:null, properties:null)
2024-08-26 20:39:09,039 INFO ql.Driver: Completed compiling command(queryId=mercy_20240826203908_2518a7ae-a17b-44a3-a9dd-7ec2ce964eba); Time taken: 0.191 seconds
2024-08-26 20:39:09,040 INFO reexec.ReExecDriver: Execution #1 of query
2024-08-26 20:39:09,041 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-08-26 20:39:09,041 INFO ql.Driver: Executing command(queryId=mercy_20240826203908_2518a7ae-a17b-44a3-a9dd-7ec2ce964eba): LOAD DATA LOCAL INPATH '/C:/Users/mercy/OneDri
ve/Documents/DataAnalytics/Hive/students.csv' OVERWRITE INTO TABLE students
2024-08-26 20:39:09,042 INFO ql.Driver: Starting task [Stage-0:MOVE] in serial mode
2024-08-26 20:39:09,042 INFO metastore.HiveMetaStore: 0: Cleaning up thread local RawStore...
2024-08-26 20:39:09,042 INFO HiveMetaStore.audit: ugi=mercy     ip=unknown-ip-addr      cmd=Cleaning up thread local RawStore...
2024-08-26 20:39:09,043 INFO metastore.HiveMetaStore: 0: Done cleaning up thread local RawStore
2024-08-26 20:39:09,045 INFO HiveMetaStore.audit: ugi=mercy     ip=unknown-ip-addr      cmd=Done cleaning up thread local RawStore
Loading data to table sample.students
2024-08-26 20:39:09,046 INFO exec.Task: Loading data to table sample.students from file:/C:/Users/mercy/OneDrive/Documents/DataAnalytics/Hive/students.csv
2024-08-26 20:39:09,045 INFO metastore.HiveMetaStore: 0: Opening raw store with implementation class:org.apache.hadoop.hive.metastore.ObjectStore
2024-08-26 20:39:09,046 WARN metastore.ObjectStore: datanucleus.autoStartMechanismMode is set to unsupported value null . Setting it to value: ignored
2024-08-26 20:39:09,046 INFO metastore.ObjectStore: ObjectStore, initialize called
2024-08-26 20:39:09,050 INFO metastore.MetaStoreDirectSql: Using direct SQL, underlying DB is DERBY
2024-08-26 20:39:09,050 INFO metastore.ObjectStore: Initialized ObjectStore
2024-08-26 20:39:09,051 INFO metastore.RetryingHMSHandler: RetryingHMSHandler proxy=class org.apache.hadoop.hive.ql.metadata.SessionHiveMetaStoreClient ugi=mercy
(auth:SIMPLE) retries=1 delay=1 lifetime=0
2024-08-26 20:39:09,051 INFO metastore.HiveMetaStore: 0: get_table : tbl=hive.sample.students
2024-08-26 20:39:09,052 INFO HiveMetaStore.audit: ugi=mercy     ip=unknown-ip-addr      cmd=get_table : tbl=hive.sample.students
2024-08-26 20:39:09,086 INFO io.HiveFileFormatUtils: Failed to check mode for C:\Users\mercy\OneDrive\Documents\DataAnalytics\Hive\students.csv: View 'unix' not available (
class java.lang.UnsupportedOperationException)
2024-08-26 20:39:09,089 INFO metastore.HiveMetaStore: 0: get_table : tbl=hive.sample.students
2024-08-26 20:39:09,090 INFO HiveMetaStore.audit: ugi=mercy     ip=unknown-ip-addr      cmd=get_table : tbl=hive.sample.students
2024-08-26 20:39:09,123 INFO metastore.HiveMetaStore: 0: get_database: @hive#sample
2024-08-26 20:39:09,124 INFO HiveMetaStore.audit: ugi=mercy     ip=unknown-ip-addr      cmd=get_database: @hive#sample
2024-08-26 20:39:09,193 INFO common.FileUtils: Creating directory if it doesn't exist: hdfs://localhost:9000/user/hive/warehouse/sample.db/students
2024-08-26 20:39:09,966 INFO metastore.HiveMetaStore: 0: alter_table: hive.sample.students newtbl=students
2024-08-26 20:39:09,967 INFO HiveMetaStore.audit: ugi=mercy     ip=unknown-ip-addr      cmd=alter_table: hive.sample.students newtbl=students
2024-08-26 20:39:10,085 INFO ql.Driver: Starting task [Stage-1:STATS] in serial mode
2024-08-26 20:39:10,086 INFO metastore.HiveMetaStore: 0: Cleaning up thread local RawStore...
2024-08-26 20:39:10,086 INFO HiveMetaStore.audit: ugi=mercy     ip=unknown-ip-addr      cmd=Cleaning up thread local RawStore...
```

```
hive> SHOW TABLES;
2024-08-26 20:40:00,827 INFO conf.HiveConf: Using the default value passed in for log id: 6cbb1c57-f5d7-4a6e-9925-c62b4c956239
2024-08-26 20:40:00,828 INFO session.SessionState: Updating thread name to 6cbb1c57-f5d7-4a6e-9925-c62b4c956239 main
2024-08-26 20:40:00,830 INFO ql.Driver: Compiling command(queryId=mercy_20240826204000_a3d158d3-c6f2-47b4-8af0-7301c6420533): SHOW TABLES
2024-08-26 20:40:00,865 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-08-26 20:40:00,868 INFO metastore.HiveMetaStore: 0: get_database: @hive#sample
2024-08-26 20:40:00,868 INFO HiveMetaStore.audit: ugi=mercy     ip=unknown-ip-addr      cmd=get_database: @hive#sample
2024-08-26 20:40:00,873 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2024-08-26 20:40:00,874 INFO ql.Driver: Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:tab_name, type:string, comment:from deserializer)], properties:null)
2024-08-26 20:40:00,876 INFO exec.ListSinkOperator: Initializing operator LIST_SINK[0]
2024-08-26 20:40:00,876 INFO ql.Driver: Completed compiling command(queryId=mercy_20240826204000_a3d158d3-c6f2-47b4-8af0-7301c6420533); Time taken: 0.046 seconds
2024-08-26 20:40:00,876 INFO reexec.ReExecDriver: Execution #1 of query
2024-08-26 20:40:00,877 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-08-26 20:40:00,877 INFO ql.Driver: Executing command(queryId=mercy_20240826204000_a3d158d3-c6f2-47b4-8af0-7301c6420533): SHOW TABLES
2024-08-26 20:40:00,878 INFO ql.Driver: Starting task [Stage-0:DDL] in serial mode
2024-08-26 20:40:00,880 INFO metastore.HiveMetaStore: 0: get_database: @hive#sample
2024-08-26 20:40:00,881 INFO HiveMetaStore.audit: ugi=mercy     ip=unknown-ip-addr      cmd=get_database: @hive#sample
2024-08-26 20:40:00,885 INFO metastore.HiveMetaStore: 0: get_tables: db=@hive#sample pat=.*
2024-08-26 20:40:00,886 INFO HiveMetaStore.audit: ugi=mercy     ip=unknown-ip-addr      cmd=get_tables: db=@hive#sample pat=.*
2024-08-26 20:40:00,918 INFO ql.Driver: Completed executing command(queryId=mercy_20240826204000_a3d158d3-c6f2-47b4-8af0-7301c6420533); Time taken: 0.041 seconds
OK
2024-08-26 20:40:00,919 INFO ql.Driver: OK
2024-08-26 20:40:00,920 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-08-26 20:40:00,924 INFO mapred.FileInputFormat: Total input files to process : 1
2024-08-26 20:40:00,954 INFO exec.ListSinkOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_LIST_SINK_0:1,
students
Time taken: 0.091 seconds, Fetched: 1 row(s)
2024-08-26 20:40:00,961 INFO CliDriver: Time taken: 0.091 seconds, Fetched: 1 row(s)
2024-08-26 20:40:00,963 INFO conf.HiveConf: Using the default value passed in for log id: 6cbb1c57-f5d7-4a6e-9925-c62b4c956239
2024-08-26 20:40:00,964 INFO session.SessionState: Resetting thread name to  main
hive> 
```

```
hive> DESC students;
2024-08-26 20:40:30,768 INFO conf.HiveConf: Using the default value passed in for log id: 6cbb1c57-f5d7-4a6e-9925-c62b4c956239
2024-08-26 20:40:30,769 INFO session.SessionState: Updating thread name to 6cbb1c57-f5d7-4a6e-9925-c62b4c956239 main
2024-08-26 20:40:30,770 INFO ql.Driver: Compiling command(queryId=mercy_20240826204030_742ecece-f18c-4465-82ba-0c86bbbaeae2): DESC students
2024-08-26 20:40:30,781 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-08-26 20:40:30,784 INFO metastore.HiveMetaStore: 0: get_table : tbl=hive.sample.students
2024-08-26 20:40:30,786 INFO HiveMetaStore.audit: ugi=mercy     ip=unknown-ip-addr      cmd=get_table : tbl=hive.sample.students
2024-08-26 20:40:30,831 INFO parse.BDLSemanticAnalyzer: analyzeDescribeTable done
2024-08-26 20:40:30,831 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2024-08-26 20:40:30,832 INFO ql.Driver: Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:col_name, type:string, comment:from deserializer), FieldSchema(
name:data_type, type:string, comment:from deserializer), FieldSchema(name:comment, type:string, comment:from deserializer)], properties:null)
2024-08-26 20:40:30,834 INFO exec.ListSinkOperator: Initializing operator LIST_SINK[0]
2024-08-26 20:40:30,834 INFO ql.Driver: Completed compiling command(queryId=mercy_20240826204030_742ecece-f18c-4465-82ba-0c86bbbaeae2); Time taken: 0.064 seconds
2024-08-26 20:40:30,834 INFO reexec.ReExecDriver: Execution #1 of query
2024-08-26 20:40:30,834 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-08-26 20:40:30,834 INFO ql.Driver: Executing command(queryId=mercy_20240826204030_742ecece-f18c-4465-82ba-0c86bbbaeae2): DESC students
2024-08-26 20:40:30,834 INFO ql.Driver: Starting task [Stage-0:DDL] in serial mode
2024-08-26 20:40:30,838 INFO metastore.HiveMetaStore: 0: get_table : tbl=hive.sample.students
2024-08-26 20:40:30,838 INFO HiveMetaStore.audit: ugi=mercy     ip=unknown-ip-addr      cmd=get_table : tbl=hive.sample.students
2024-08-26 20:40:30,871 INFO ql.Driver: Completed executing command(queryId=mercy_20240826204030_742ecece-f18c-4465-82ba-0c86bbbaeae2); Time taken: 0.037 seconds
OK
2024-08-26 20:40:30,872 INFO ql.Driver: OK
2024-08-26 20:40:30,872 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-08-26 20:40:30,875 INFO mapred.FileInputFormat: Total input files to process : 1
2024-08-26 20:40:30,890 INFO exec.ListSinkOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_LIST_SINK_0:3,
name                    string
roll                    int
dept                    string
Time taken: 0.103 seconds, Fetched: 3 row(s)
2024-08-26 20:40:30,915 INFO CliDriver: Time taken: 0.103 seconds, Fetched: 3 row(s)
2024-08-26 20:40:30,917 INFO conf.HiveConf: Using the default value passed in for log id: 6cbb1c57-f5d7-4a6e-9925-c62b4c956239
2024-08-26 20:40:30,923 INFO session.SessionState: Resetting thread name to  main
hive> 
```





RESULT:

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