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

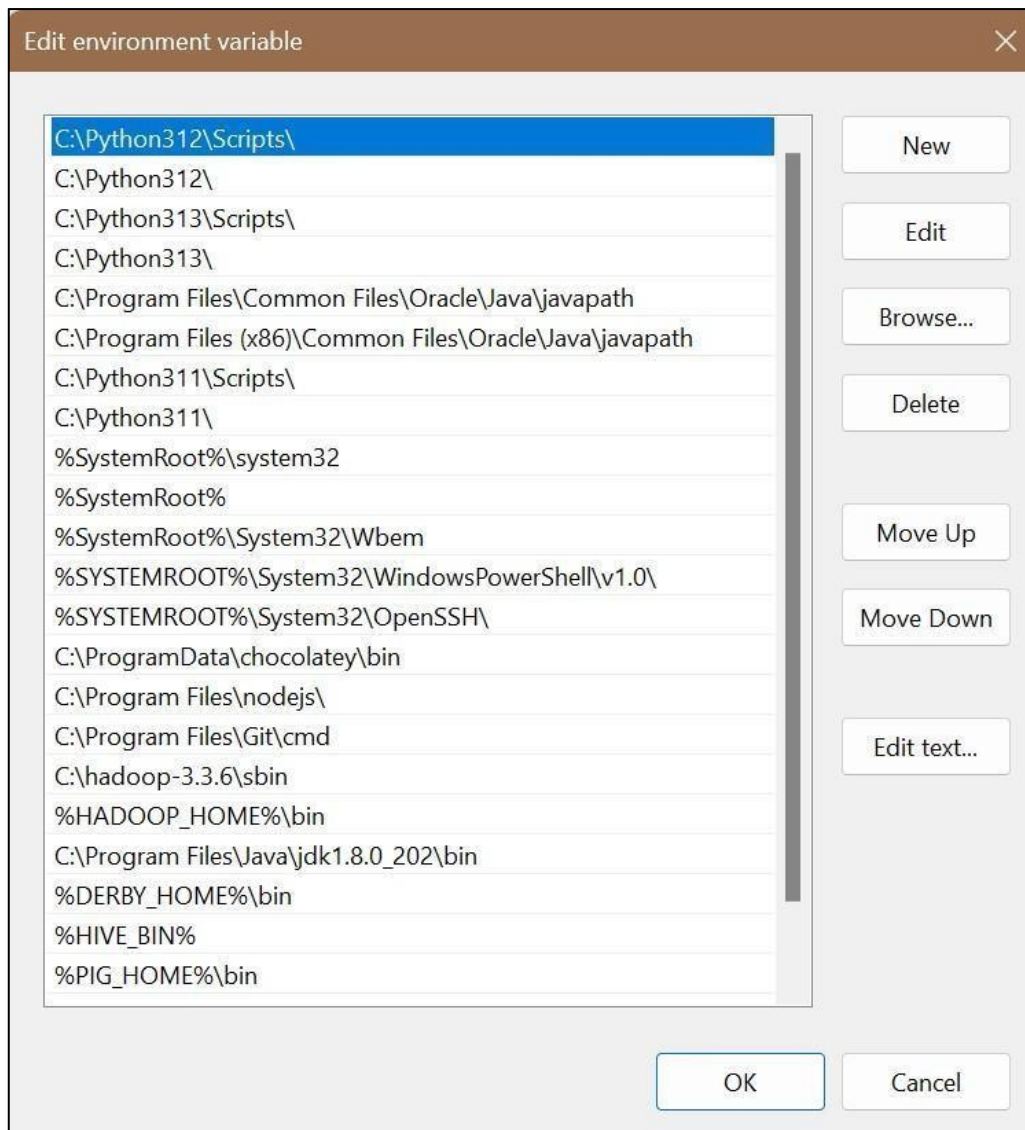   https://db.apache.org/derby/derby_downloads.html#For+Java+8+and+Higher

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-derby10.14.2.0\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 DATALOCAL 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:

```
C:\>hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/C:/hadoop-3.3.6/share/hadoop/common/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/C:/apache-hive-3.1.3-bin/lib/log4j-slf4j-impl-2.17.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Reload4jLoggerFactory]
2024-08-26 20:02:32,552 INFO conf.HiveConf: Found configuration file file:/C:/apache-hive-3.1.3-bin/conf/hive-site.xml
Hive Session ID = 6cbb1c57-f5d7-4a6e-9925-c62b4c956239
2024-08-26 20:02:33,378 INFO SessionState: Hive Session ID = 6cbb1c57-f5d7-4a6e-9925-c62b4c956239

Logging initialized using configuration in jar:file:/C:/apache-hive-3.1.3-bin/lib/hive-common-3.1.3.jar!/hive-log4j2.properties Async: true
2024-08-26 20:02:33,423 INFO SessionState:
Logging initialized using configuration in jar:file:/C:/apache-hive-3.1.3-bin/lib/hive-common-3.1.3.jar!/hive-log4j2.properties Async: true
2024-08-26 20:02:34,069 INFO session.SessionState: Created HDFS directory: /tmp/hive/mercy/6cbb1c57-f5d7-4a6e-9925-c62b4c956239
2024-08-26 20:02:34,074 INFO session.SessionState: Created local directory: C:/Users/mercy/AppData/Local/Temp/mercy/6cbb1c57-f5d7-4a6e-9925-c62b4c956239
2024-08-26 20:02:34,078 INFO session.SessionState: Created HDFS directory: /tmp/hive/mercy/6cbb1c57-f5d7-4a6e-9925-c62b4c956239/_tmp_space.db
2024-08-26 20:02:34,088 INFO conf.HiveConf: Using the default value passed in for log id: 6cbb1c57-f5d7-4a6e-9925-c62b4c956239
2024-08-26 20:02:34,088 INFO session.SessionState: Updating thread name to 6cbb1c57-f5d7-4a6e-9925-c62b4c956239 main
2024-08-26 20:02:34,780 INFO metastore.HiveMetaStore: 0: Opening raw store with implementation class:org.apache.hadoop.hive.metastore.ObjectStore
2024-08-26 20:02:34,806 WARN metastore.ObjectStore: datanucleus.autoStartMechanismMode is set to unsupported value null . Setting it to value: ignored
2024-08-26 20:02:34,810 INFO metastore.ObjectStore: ObjectStore, initialize called
2024-08-26 20:02:34,812 INFO conf.MetastoreConf: Found configuration file file:/C:/apache-hive-3.1.3-bin/conf/hive-site.xml
2024-08-26 20:02:34,812 INFO conf.MetastoreConf: Unable to find config file hivemetastore-site.xml
2024-08-26 20:02:34,813 INFO conf.MetastoreConf: Found configuration file null
2024-08-26 20:02:34,813 INFO conf.MetastoreConf: Unable to find config file metastore-site.xml
2024-08-26 20:02:34,814 INFO conf.MetastoreConf: Found configuration file null
2024-08-26 20:02:34,942 INFO DataNucleus.Persistence: Property datanucleus.cache.level2 unknown - will be ignored
2024-08-26 20:02:35,116 INFO hikari.HikariDataSource: HikariPool-1 - Starting...
2024-08-26 20:02:35,614 INFO pool.PoolBase: HikariPool-1 - Driver does not support get/set network timeout for connections. (Feature not implemented: getNetworkTimeout.)
2024-08-26 20:02:35,713 INFO hikari.HikariDataSource: HikariPool-1 - Start completed.
2024-08-26 20:02:35,767 INFO hikari.HikariDataSource: HikariPool-2 - Starting...
```

```
hive> SHOW DATABASES;
2024-08-26 20:02:50,018 INFO conf.HiveConf: Using the default value passed in for log id: 6cbb1c57-f5d7-4a6e-9925-c62b4c956239
2024-08-26 20:02:50,096 INFO ql.Driver: Compiling command(queryId=mercy_20240826200250_6f3736af-a1b0-4590-9b97-93f24e1e4032): SHOW DATABASES
2024-08-26 20:02:50,653 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-08-26 20:02:50,683 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2024-08-26 20:02:50,740 INFO ql.Driver: Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:database_name, type:string, comment:from deserializer)], properties:null)
2024-08-26 20:02:50,829 INFO exec.ListSinkOperator: Initializing operator LIST_SINK[0]
2024-08-26 20:02:50,840 INFO ql.Driver: Completed compiling command(queryId=mercy_20240826200250_6f3736af-a1b0-4590-9b97-93f24e1e4032); Time taken: 0.77 seconds
2024-08-26 20:02:50,841 INFO reexec.ReExecDriver: Execution #1 of query
2024-08-26 20:02:50,846 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-08-26 20:02:50,846 INFO ql.Driver: Executing command(queryId=mercy_20240826200250_6f3736af-a1b0-4590-9b97-93f24e1e4032): SHOW DATABASES
2024-08-26 20:02:50,861 INFO ql.Driver: Starting task [Stage-0:DDL] in serial mode
2024-08-26 20:02:50,863 INFO metastore.HiveMetaStore: 0: get_databases: @hive#
2024-08-26 20:02:50,864 INFO HiveMetaStore.audit: ugi=mercy      ip=unknown-ip-addr      cmd=get_databases: @hive#

2024-08-26 20:02:50,870 INFO exec.DDLTask: results : 2
2024-08-26 20:02:50,907 INFO ql.Driver: Completed executing command(queryId=mercy_20240826200250_6f3736af-a1b0-4590-9b97-93f24e1e4032); Time taken: 0.061 seconds
OK
2024-08-26 20:02:50,909 INFO ql.Driver: OK
2024-08-26 20:02:50,910 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-08-26 20:02:50,926 INFO Configuration.deprecation: mapred.input.dir is deprecated. Instead, use mapreduce.input.fileinputformat.inputdir
2024-08-26 20:02:50,961 INFO mapred.FileInputFormat: Total input files to process : 1
2024-08-26 20:02:50,992 INFO exec.ListSinkOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_LIST_SINK_0:2,
default
sample
Time taken: 0.842 seconds, Fetched: 2 row(s)
2024-08-26 20:02:51,000 INFO CliDriver: Time taken: 0.842 seconds, Fetched: 2 row(s)
2024-08-26 20:02:51,001 INFO conf.HiveConf: Using the default value passed in for log id: 6cbb1c57-f5d7-4a6e-9925-c62b4c956239
2024-08-26 20:02:51,005 INFO session.SessionState: Resetting thread name to  main
```

```
hive> USE sample
    > ;
2024-08-26 20:05:59,795 INFO conf.HiveConf: Using the default value passed in for log id: 6cbb1c57-f5d7-4a6e-9925-c62b4c956239
2024-08-26 20:05:59,795 INFO session.SessionState: Updating thread name to 6cbb1c57-f5d7-4a6e-9925-c62b4c956239 main
2024-08-26 20:05:59,799 INFO ql.Driver: Compiling command(queryId=mercy_20240826200559_3b4f6151-de3b-4a30-8d29-0fe40576237e): USE sample

2024-08-26 20:05:59,818 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-08-26 20:05:59,819 INFO metastore.HiveMetaStore: 0: get_database: @hive#sample
2024-08-26 20:05:59,821 INFO HiveMetaStore.audit: ugi=mercy      ip=unknown-ip-addr      cmd=get_database: @hive#sample

2024-08-26 20:05:59,842 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2024-08-26 20:05:59,842 INFO ql.Driver: Returning Hive schema: Schema(fieldSchemas:null, properties:null)
2024-08-26 20:05:59,845 INFO ql.Driver: Completed compiling command(queryId=mercy_20240826200559_3b4f6151-de3b-4a30-8d29-0fe40576237e); Time taken: 0.045 seconds
2024-08-26 20:05:59,846 INFO reexec.ReExecDriver: Execution #1 of query
2024-08-26 20:05:59,850 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-08-26 20:05:59,851 INFO ql.Driver: Executing command(queryId=mercy_20240826200559_3b4f6151-de3b-4a30-8d29-0fe40576237e): USE sample

2024-08-26 20:05:59,852 INFO ql.Driver: Starting task [Stage-0:DDL] in serial mode
2024-08-26 20:05:59,853 INFO metastore.HiveMetaStore: 0: get_database: @hive#sample
2024-08-26 20:05:59,854 INFO HiveMetaStore.audit: ugi=mercy      ip=unknown-ip-addr      cmd=get_database: @hive#sample

2024-08-26 20:05:59,858 INFO metastore.HiveMetaStore: 0: get_database: @hive#sample
2024-08-26 20:05:59,858 INFO HiveMetaStore.audit: ugi=mercy      ip=unknown-ip-addr      cmd=get_database: @hive#sample

2024-08-26 20:05:59,863 INFO ql.Driver: Completed executing command(queryId=mercy_20240826200559_3b4f6151-de3b-4a30-8d29-0fe40576237e); Time taken: 0.013 seconds
OK
2024-08-26 20:05:59,864 INFO ql.Driver: OK
2024-08-26 20:05:59,864 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
Time taken: 0.066 seconds
2024-08-26 20:05:59,865 INFO CliDriver: Time taken: 0.066 seconds
2024-08-26 20:05:59,865 INFO conf.HiveConf: Using the default value passed in for log id: 6cbb1c57-f5d7-4a6e-9925-c62b4c956239
2024-08-26 20:05:59,866 INFO session.SessionState: Resetting thread name to  main
```

```
hive> CREATE TABLE students (name STRING, roll INT, dept STRING);
2024-08-26 20:10:08,726 INFO conf.HiveConf: Using the default value passed in for log id: 6cbb1c57-f5d7-4a6e-9925-c62b4c956239
2024-08-26 20:10:08,726 INFO session.SessionState: Updating thread name to 6cbb1c57-f5d7-4a6e-9925-c62b4c956239 main
2024-08-26 20:10:08,728 INFO ql.Driver: Compiling command(queryId=mercy_20240826201008_7d484479-9c70-407e-9b5b-ef577623a7de): CREATE TABLE students (name STRING, roll INT,
dept STRING)
2024-08-26 20:10:08,780 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-08-26 20:10:08,784 INFO parse.CalcitePlanner: Starting Semantic Analysis
2024-08-26 20:10:08,804 INFO sqlstd.SQLStdHiveAccessController: Created SQLStdHiveAccessController for session context : HiveAuthzSessionContext [sessionString=6cbb1c57-f5d
7-4a6e-9925-c62b4c956239, clientType=HIVECLI]
2024-08-26 20:10:08,807 WARN session.SessionState: METASTORE_FILTER_HOOK will be ignored, since hive.security.authorization.manager is set to instance of HiveAuthorizerFact
ory.
2024-08-26 20:10:08,808 INFO metastore.HiveMetaStoreClient: Mestastore configuration metastore.filter.hook changed from org.apache.hadoop.hive.metastore.DefaultMetaStoreFil
terHookImpl to org.apache.hadoop.hive.ql.security.authorization.plugin.AuthorizationMetaStoreFilterHook
2024-08-26 20:10:08,809 INFO metastore.HiveMetaStore: 0: Cleaning up thread local RawStore...
2024-08-26 20:10:08,809 INFO HiveMetaStore.audit: ugi=mercy      ip=unknown-ip-addr      cmd=Cleaning up thread local RawStore...
2024-08-26 20:10:08,810 INFO metastore.HiveMetaStore: 0: Done cleaning up thread local RawStore
2024-08-26 20:10:08,810 INFO HiveMetaStore.audit: ugi=mercy      ip=unknown-ip-addr      cmd=Done cleaning up thread local RawStore
2024-08-26 20:10:08,812 INFO metastore.HiveMetaStore: 0: Opening raw store with implementation class:org.apache.hadoop.hive.metastore.ObjectStore
2024-08-26 20:10:08,812 WARN metastore.ObjectStore: datanucleus.autoStartMechanismMode is set to unsupported value null . Setting it to value: ignored
2024-08-26 20:10:08,813 INFO metastore.ObjectStore: ObjectStore, initialize called
2024-08-26 20:10:08,817 INFO metastore.MetaStoreDirectSql: Using direct SQL, underlying DB is DERBY
2024-08-26 20:10:08,817 INFO metastore.ObjectStore: Initialized ObjectStore
2024-08-26 20:10:08,817 INFO metastore.RetryingMetaStoreClient: RetryingMetaStoreClient proxy=class org.apache.hadoop.hive.ql.metadata.SessionHiveMetaStoreClient ugi=mercy
(auth:SIMPLE) retries=1 delay=1 lifetime=0
2024-08-26 20:10:08,830 INFO parse.CalcitePlanner: Creating table sample.students position=13
2024-08-26 20:10:08,843 WARN metastore.ObjectStore: datanucleus.autoStartMechanismMode is set to unsupported value null . Setting it to value: ignored
2024-08-26 20:10:08,844 INFO metastore.ObjectStore: ObjectStore, initialize called
2024-08-26 20:10:08,848 INFO metastore.MetaStoreDirectSql: Using direct SQL, underlying DB is DERBY
2024-08-26 20:10:08,848 INFO metastore.ObjectStore: Initialized ObjectStore
2024-08-26 20:10:08,852 INFO metastore.RetryingMetaStoreClient: RetryingMetaStoreClient proxy=class org.apache.hadoop.hive.ql.metadata.SessionHiveMetaStoreClient ugi=mercy
(auth:SIMPLE) retries=1 delay=1 lifetime=0
2024-08-26 20:10:08,852 INFO metastore.HiveMetaStore: 0: get_database: @hive#sample
2024-08-26 20:10:08,856 INFO HiveMetaStore.audit: ugi=mercy      ip=unknown-ip-addr      cmd=get_database: @hive#sample

2024-08-26 20:10:08,877 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2024-08-26 20:10:08,878 INFO ql.Driver: Returning Hive schema: Schema(fieldSchemas:null, properties:null)
2024-08-26 20:10:08,878 INFO ql.Driver: Completed compiling command(queryId=mercy_20240826201008_7d484479-9c70-407e-9b5b-ef577623a7de); Time taken: 0.15 seconds
2024-08-26 20:10:08,882 INFO reexec.ReExecDriver: Execution #1 of query
2024-08-26 20:10:08,883 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-08-26 20:10:08,887 INFO ql.Driver: Executing command(queryId=mercy_20240826201008_7d484479-9c70-407e-9b5b-ef577623a7de): CREATE TABLE students (name STRING, roll INT,
dept STRING)
2024-08-26 20:10:08,895 INFO ql.Driver: Starting task [Stage-0:DDL] in serial mode
2024-08-26 20:10:08,896 INFO metastore.HiveMetaStoreClient: Mestastore configuration metastore.filter.hook changed from org.apache.hadoop.hive.ql.security.authorization.plu
gin.AuthorizationMetaStoreFilterHook to org.apache.hadoop.hive.metastore.DefaultMetaStoreFilterHookImpl
2024-08-26 20:10:08,896 INFO metastore.HiveMetaStore: 0: Cleaning up thread local RawStore...
```

```
2024-08-26 20:10:09,162 INFO utils.FileUtils: Creating directory if it doesn't exist: hdfs://localhost:9000/user/hive/warehouse/sample.db/students
2024-08-26 20:10:09,718 INFO ql.Driver: Completed executing command(queryId=mercy_20240826201008_7d484479-9c70-407e-9b5b-ef577623a7de); Time taken: 0.831 seconds
OK
2024-08-26 20:10:09,719 INFO ql.Driver: OK
2024-08-26 20:10:09,720 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
Time taken: 0.993 seconds
2024-08-26 20:10:09,722 INFO CliDriver: Time taken: 0.993 seconds
2024-08-26 20:10:09,723 INFO conf.HiveConf: Using the default value passed in for log id: 6cbb1c57-f5d7-4a6e-9925-c62b4c956239
2024-08-26 20:10:09,723 INFO session.SessionState: Resetting thread name to  main
hive>
```

```
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_251843ae-e17b-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_251843ae-e17b-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_251843ae-e17b-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 20:39:09,043 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 20:39:09,044 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.RetryingMetaStoreClient: RetryingMetaStoreClient 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 20:39:09,153 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-c6f2-47b4-8afd-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-c6f2-47b4-8afd-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-c6f2-47b4-8afd-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-c6f2-47b4-8afd-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.DDLSemanticAnalyzer: 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(
a_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>
```

```
hive> SELECT * FROM students;
2024-08-26 20:41:12,643 INFO conf.HiveConf: Using the default value passed in for log id: 6cbb1c57-f5d7-4a6e-9925-c62b4c956239
2024-08-26 20:41:12,644 INFO session.SessionState: Updating thread name to 6cbb1c57-f5d7-4a6e-9925-c62b4c956239 main
2024-08-26 20:41:12,645 INFO ql.Driver: Compiling command(queryId=mercy_20240826204112_4f6f657c-0e35-400c-9a40-f9a19709922d): SELECT * F
2024-08-26 20:41:12,663 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-08-26 20:41:12,663 INFO parse.CalcitePlanner: Starting Semantic Analysis
2024-08-26 20:41:12,691 INFO parse.CalcitePlanner: Completed phase 1 of Semantic Analysis
2024-08-26 20:41:12,691 INFO parse.CalcitePlanner: Get metadata for source tables
2024-08-26 20:41:12,691 INFO metastore.HiveMetaStore: 0: get_table : tbl=hive.sample.students
2024-08-26 20:41:12,693 INFO HiveMetaStore.audit: ugi=mercy      ip=unknown-ip-addr      cmd=get_table : tbl=hive.sample.students
2024-08-26 20:41:12,751 INFO parse.CalcitePlanner: Get metadata for subqueries
2024-08-26 20:41:12,751 INFO parse.CalcitePlanner: Get metadata for destination tables
2024-08-26 20:41:12,791 INFO ql.Context: New scratch dir is hdfs://localhost:9000/tmp/hive/mercy/6cbb1c57-f5d7-4a6e-9925-c62b4c956239/hi
5444333444-1
2024-08-26 20:41:12,792 INFO parse.CalcitePlanner: Completed getting MetaData in Semantic Analysis
```

```
OK
2024-08-26 20:41:14,144 INFO ql.Driver: OK
2024-08-26 20:41:14,145 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-08-26 20:41:14,160 INFO mapred.FileInputFormat: Total input files to process : 1
2024-08-26 20:41:14,288 INFO lazy.LazyStruct: Missing fields! Expected 3 fields but only got 1! Ignoring similar problems.
2024-08-26 20:41:14,291 INFO exec.TableScanOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_TS_0:10,
2024-08-26 20:41:14,291 INFO exec.SelectOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_SEL_1:10,
2024-08-26 20:41:14,291 INFO exec.ListSinkOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_LIST_SINK_3:10,
name,roll,dept  NULL        NULL
Abhishek,110,CSE        NULL    NULL
Pragya,126,CSE  NULL    NULL
Dev, 231,CSBS   NULL    NULL
Sonakshi,241,CSBS       NULL    NULL
Sameer,343,IT   NULL    NULL
Naina,356,IT    NULL    NULL
Ranveer,511,ECE NULL    NULL
Ishaani,521,ECE NULL    NULL
Mercy,157,CSE   NULL    NULL
Time taken: 1.5 seconds, Fetched: 10 row(s)
2024-08-26 20:41:14,321 INFO CliDriver: Time taken: 1.5 seconds, Fetched: 10 row(s)
2024-08-26 20:41:14,321 INFO conf.HiveConf: Using the default value passed in for log id: 6cbb1c57-f5d7-4a6e-9925-c62b4c956239
2024-08-26 20:41:14,322 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.