EX-6 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

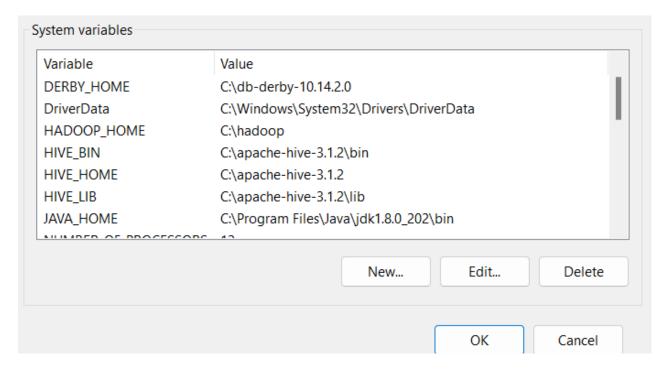
Install Apache Derby 10.14.2.0

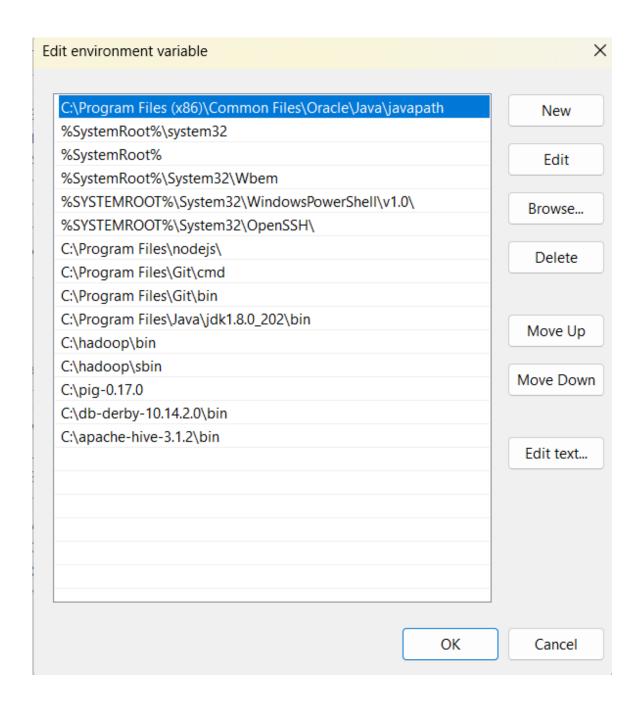
https://db.apache.org/derby/derby_downloads.html#For+Java+8+and+Higher

2. Downloading Apache Hive binaries

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

3. Setting environment variables





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://ldrv.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:

```
Administrator: Command Prompt - hive
                                                                                                                                                                                                                         C:\Windows\System32>jps
16784 NodeManager
                                                                                                                                                                                                                                           П
 7904 Jps
11780 DataNode
20488 NameNode
 25144 NetworkServerControl
2648 ResourceManager
 C:\> hive --service schematool -dbType derby -initSchema
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/C:/hadoop/share/hadoop/common/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLog
  lF4J: Found binding in [jar:file:/C:/apache-hive-3.1.2/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBind
 strag. Found Blooming to perceive and the process of the process o
  letastore connection URL:
 Metastore connection URL: jdbc:derby:;databaseName=metastore_db;create=true
2024-09-01 15:02:19,982 INFO tools.HiveSchemaHelper: Metastore Connection Driver :
                                                                                                                                                                             org.apache.derby.jdbc.EmbeddedD
river
Metastore Connection Driver : org.apache.derby.jdbc.EmbeddedDriver
2024-09-01 15:02:19,982 INFO tools.HiveSchemaHelper: Metastore connection User: APP
Metastore connection User: APP
Starting metastore schema initialization to 3.1.0
Initialization script hive-schema-3.1.0.derby.sql
                                                                                                                                                                                                                           X
Administrator: Command Prompt - StartNetworkServer -h 0.0.0.0
Microsoft Windows [Version 10.0.22631.4037]
(c) Microsoft Corporation. All rights reserved.
 :\Windows\System32>StartNetworkServer -h 0.0.0.0
 un Sep 01 14:55:05 IST 2024 : Security manager installed using the Basic server security policy.
 un Sep 01 14:55:05 IST 2024 : Apache Derby Network Server - 10.14.2.0 - (1828579) started and ready to accept connectio
 s on port 1527
 Administrator: Command Prompt - hive
2024-09-01 15:03:13,034 INFO session.SessionState: Created HDFS directory: /tmp/hive/monik/3e59fbed-f58f-41ab-8862-11722
ba3445f/_tmp_space.db
2024-09-01 15:03:13,036 INFO metastore.HiveMetaStore: 1: get_databases: @hive#
2024-09-01 15:03:13,036 INFO HiveMetaStore.audit: ugi=monik
                                                                                                                          ip=unknown-ip-addr
                                                                                                                                                                             cmd=get databases: @hive#
2024-09-01 15:03:13,036 INFO metastore.HiveMetaStore: 1: Opening raw store with implementation class:org.apache.hadoop.h
ive.metastore.ObjectStore
2024-09-01 15:03:13,036 INFO metastore.ObjectStore: ObjectStore, initialize called
2024-09-01 15:03:13,050 INFO metastore.MetaStoreDirectSql: Using direct SQL, underlying DB is DERBY
2024-09-01 15:03:13,050 INFO metastore.ObjectStore: Initialized ObjectStore
2024-09-01 15:03:13,067 INFO metastore.HiveMetaStore: 1: get_tables_by_type: db=@hive#default pat=.*,type=MATERIALIZED_V
IEW
2024-09-01 15:03:13,067 INFO HiveMetaStore.audit: ugi=monik
                                                                                                                                                                             cmd=get_tables_by_type: db=@hive
                                                                                                                             ip=unknown-ip-addr
#default pat=.*,type=MATERIALIZED_VIEW
2024-09-01 15:03:13,099 INFO metastore.HiveMetaStore: 1: get_multi_table : db=default tbls=
2024-09-01 15:03:13,099 INFO HiveMetaStore.audit: ugi=monik
                                                                                                                             ip=unknown-ip-addr
                                                                                                                                                                             cmd=get_multi_table : db=default
 tbls=
2024-09-01 15:03:13,099 INFO metadata.HiveMaterializedViewsRegistry: Materialized views registry has been initialized
nive> _
```

```
Administrator: Command Prompt - hive
                                                                                                                [nitialization script completed
schemaTool completed
::\>nive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/C:/hadoop/share/hadoop/common/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLog
gerBinder.class]
stf4]: Found binding in [jar:file:/C:/apache-hive-3.1.2/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBind
r.class]
SILF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SILF4J: Actual binding is of type [org.slf4j.impl.Reload4jloggerFactory]
2024-09-01 15:02:56,242 INFO conf.HiveConf: Found configuration file null
2024-09-01 15:02:59,189 WARN common.LogUtils: hive-site.xml not found on CLASSPATH
dive Session ID = 612a00af-e4af-4260-a4dd-72304ffd9426
2024-09-01 15:02:59,271 INFO SessionState: Hive Session ID = 612a00af-e4af-4260-a4dd-72304ffd9426
 ogging initialized using configuration in jar:file:/C:/apache-hive-3.1.2/lib/hive-common-3.1.2.jar!/hive-log4j2.propert
ies Async: true
2024-09-01 15:02:59,436 INFO SessionState:
ogging initialized using configuration in jar:file:/C:/apache-hive-3.1.2/lib/hive-common-3.1.2.jar!/hive-log4j2.propert
es Async: true
2024-09-01 15:03:01,426 INFO session.SessionState: Created HDFS directory: /tmp/hive/monik
2024-09-01 15:03:01,440 INFO session.SessionState: Created HDFS directory: /tmp/hive/monik/612a00af-e4af-4260-a4dd-72304
2024-09-01 15:03:01,446 INFO session.SessionState: Created local directory: C:/Users/monik/AppData/Local/Temp/monik/612a
00af-e4af-4260-a4dd-72304ffd9426
2024-09-01 15:03:01,453 INFO session.SessionState: Created HDFS directory: /tmp/hive/monik/612a00af-e4af-4260-a4dd-72304
Select Administrator: Command Prompt - hive
hive> CREATE DATABASE financials;
2024-09-01 15:13:01,654 INFO conf.HiveConf: Using the default value passed in for log id: 612a00af-e4af
4260-a4dd-72304ffd9426
2024-09-01 15:13:01,942 INFO ql.Driver: Compiling command(queryId=monik_20240901151301_3f644507-c532-48d
7-a39e-c69e22c5aaf8): CREATE DATABASE financials
2024-09-01 15:13:02,564 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-01 15:13:02,596 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2024-09-01 15:13:02,603 INFO ql.Driver: Returning Hive schema: Schema(fieldSchemas:null, properties:null
2024-09-01 15:13:02,623 INFO ql.Driver: Completed compiling command(queryId=monik_20240901151301_3f64450
7-c532-48d7-a39e-c69e22c5aaf8); Time taken: 0.795 seconds
2024-09-01 15:13:02,625 INFO reexec.ReExecDriver: Execution #1 of query
2024-09-01 15:13:02,627 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-01 15:13:02,630 INFO ql.Driver: Executing command(queryId=monik_20240901151301_3f644507-c532-48d
Select Administrator: Command Prompt - hive
2024-09-01 15:13:02,844 INFO session.SessionState: Resetting thread name to main
hive> use financials;
2024-09-01 15:13:33,260 INFO conf.HiveConf: Using the default value passed in for log id: 612a00af-e4af
4260-a4dd-72304ffd9426
2024-09-01 15:13:33,261 INFO session.SessionState: Updating thread name to 612a00af-e4af-4260-a4dd-72304
ffd9426 main
2024-09-01 15:13:33,263 INFO ql.Driver: Compiling command(queryId=monik_20240901151333_27a2f212-1e1f-47e
b-a9eb-4770da973259): use financials
2024-09-01 15:13:33,288 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-01 15:13:33,288 INFO metastore.HiveMetaStore: 0: get_database: @hive#financials
2024-09-01 15:13:33,289 INFO HiveMetaStore.audit: ugi=monik
                                                                        ip=unknown-ip-addr
                                                                                                    cmd=get database
: @hive#financials
2024-09-01 15:13:33,324 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2024-09-01 15:13:33,325 INFO ql.Driver: Returning Hive schema: Schema(fieldSchemas:null, properties:null
2024-09-01 15:13:33,325 INFO ql.Driver: Completed compiling command(queryId=monik_20240901151333_27a2f21
2-1e1f-47eb-a9eb-4770da973259); Time taken: 0.062 seconds
 Select Administrator: Command Prompt - hive
hive> CREATE TABLE finance_table(id INT, name STRING);
2024-09-01 15:14:52,808 INFO conf.HiveConf: Using the default value passed in for log id: 612a00af-e4af-
4260-a4dd-72304ffd9426
2024-09-01 15:14:52,808 INFO session.SessionState: Updating thread name to 612a00af-e4af-4260-a4dd-72304
ffd9426 main
2024-09-01 15:14:52,812 INFO ql.Driver: Compiling command(queryId=monik_20240901151452_6694b761-fe8b-4e2
4-9a0b-bf91f9bdedc4): CREATE TABLE finance_table(id INT, name STRING)
2024-09-01 15:14:52,912 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-01 15:14:52,919 INFO parse.CalcitePlanner: Starting Semantic Analysis
2024-09-01 15:14:52,938 INFO parse.CalcitePlanner: Creating table financials.finance_table position=13
2024-09-01 15:14:52,971 INFO metastore.HiveMetaStore: 0: get_database: @hive#financials
2024-09-01 15:14:52,971 INFO HiveMetaStore.audit: ugi=monik
                                                                        ip=unknown-ip-addr
                                                                                                    cmd=get database
@hive#financials
2024-09-01 15:14:52,995 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2024-09-01 15:14:53,006 INFO ql.Driver: Returning Hive schema: Schema(fieldSchemas:null, properties:null
```

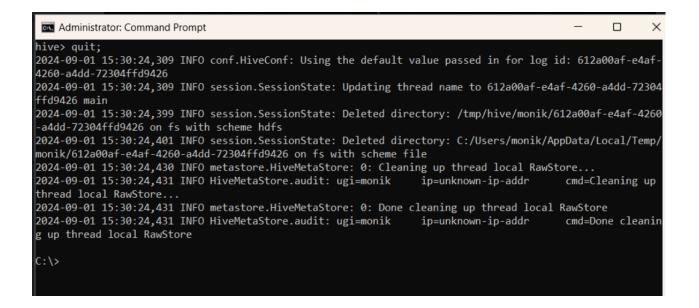
```
Select Administrator: Command Prompt - hive
hive> INSERT INTO finance_table VALUES (1,'MONIKA'), (2,'UMA');
2024-09-01 15:17:33,183 INFO conf.HiveConf: Using the default value passed in for log id: 612a00af-e4af
4260-a4dd-72304ffd9426
2024-09-01 15:17:33,184 INFO session.SessionState: Updating thread name to 612a00af-e4af-4260-a4dd-72304
ffd9426 main
2024-09-01 15:17:33,189 INFO ql.Driver: Compiling command(queryId=monik_20240901151733_e08d9e40-e7bb-471
6-86dd-0c93d5fb0265): INSERT INTO finance_table VALUES (1,'MONIKA'), (2,'UMA')
2024-09-01 15:17:33,217 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-01 15:17:33,218 INFO parse.CalcitePlanner: Starting Semantic Analysis
2024-09-01 15:17:33,262 INFO metastore.HiveMetaStore: 0: get_table : tbl=hive.financials.finance_table
2024-09-01 15:17:33,262 INFO HiveMetaStore.audit: ugi=monik
                                                               ip=unknown-ip-addr
                                                                                       cmd=get_table :
tbl=hive.financials.finance_table
2024-09-01 15:17:33,459 INFO parse.CalcitePlanner: Completed phase 1 of Semantic Analysis
2024-09-01 15:17:33,459 INFO parse.CalcitePlanner: Get metadata for source tables
2024-09-01 15:17:33,459 INFO parse.CalcitePlanner: Get metadata for subqueries
 Select Administrator: Command Prompt - hive
nive> CREATE VIEW myview AS SELECT name, id FROM finance_table;
2024-09-01 15:19:10,400 INFO conf.HiveConf: Using the default value passed in for log id: 612a00af-e4af
4260-a4dd-72304ffd9426
2024-09-01 15:19:10,401 INFO session.SessionState: Updating thread name to 612a00af-e4af-4260-a4dd-7230
ffd9426 main
2024-09-01 15:19:10,404 INFO ql.Driver: Compiling command(queryId=monik_20240901151910_b547082e-4de0-45
f-9a8d-48b66bc6edd4): CREATE VIEW myview AS SELECT name, id FROM finance_table
2024-09-01 15:19:10,431 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-01 15:19:10,431 INFO parse.CalcitePlanner: Starting Semantic Analysis
2024-09-01 15:19:10,432 INFO parse.CalcitePlanner: Creating view financials.myview position=12
2024-09-01 15:19:10,437 INFO metastore.HiveMetaStore: 0: get_database: @hive#financials
2024-09-01 15:19:10,437 INFO HiveMetaStore.audit: ugi=monik
                                                                 ip=unknown-ip-addr
                                                                                         cmd=get databas
: @hive#financials
2024-09-01 15:19:10,446 INFO parse.CalcitePlanner: Completed phase 1 of Semantic Analysis
2024-09-01 15:19:10,446 INFO parse.CalcitePlanner: Get metadata for source tables
2024-09-01 15:19:10,447 INFO metastore.HiveMetaStore: 0: get_table : tbl=hive.financials.finance_table
2024-09-01 15:19:10,451 INFO HiveMetaStore.audit: ugi=monik
                                                                 ip=unknown-ip-addr
                                                                                         cmd=get table
tbl=hive.financials.finance table
 Select Administrator: Command Prompt - hive
hive> SELECT * FROM myview;
2024-09-01 15:19:28,583 INFO conf.HiveConf: Using the default value passed in for log id: 612a00af-e4af
4260-a4dd-72304ffd9426
2024-09-01 15:19:28,585 INFO session.SessionState: Updating thread name to 612a00af-e4af-4260-a4dd-72304
ffd9426 main
2024-09-01 15:19:28,587 INFO ql.Driver: Compiling command(queryId=monik_20240901151928_e5b39971-edef-4a2
0-a37a-ab49a36a3d1b): SELECT * FROM myview
2024-09-01 15:19:28,610 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-01 15:19:28,611 INFO parse.CalcitePlanner: Starting Semantic Analysis
2024-09-01 15:19:28,612 INFO parse.CalcitePlanner: Completed phase 1 of Semantic Analysis
2024-09-01 15:19:28,615 INFO parse.CalcitePlanner: Get metadata for source tables
2024-09-01 15:19:28,616 INFO metastore.HiveMetaStore: 0: get_table : tbl=hive.financials.myview
                                                               ip=unknown-ip-addr
2024-09-01 15:19:28,617 INFO HiveMetaStore.audit: ugi=monik
                                                                                       cmd=get_table :
tbl=hive.financials.myview
2024-09-01 15:19:28,641 INFO parse.CalcitePlanner: Get metadata for subqueries
2024-09-01 15:19:28,641 INFO parse.CalcitePlanner: Get metadata for source tables
2024-09-01 15:19:28,642 INFO metastore.HiveMetaStore: 0: get_table : tbl=hive.financials.finance_table
2024-09-01 15:19:28,643 INFO HiveMetaStore.audit: ugi=monik
                                                               ip=unknown-ip-addr
                                                                                       cmd=get table :
     CUZ4-UJ-UI IJ.IJ.ZO, JUJ INIU EACC. JETECLUPEI ALU
    :2,
    2024-09-01 15:19:28,964 INFO exec.ListSinkOpera
      SINK 3:2,
    MONIKA
                  1
    Time taken: 0.342 seconds, Fetched: 2 row(s)
    2024-09-01 15:19:28,980 INFO CliDriver: Time ta
```

2024-09-01 15:19:28,980 INFO conf.HiveConf: Usi

2021-09-01 15:19:28 981 TNFO session SessionSta

4260-a4dd-72304ffd9426

```
Select Administrator: Command Prompt - hive
hive> DESCRIBE finance_table;
2024-09-01 15:19:53,646 INFO conf.HiveConf: Using the default value passed in for log id: 612a00af-e4af
4260-a4dd-72304ffd9426
2024-09-01 15:19:53,647 INFO session.SessionState: Updating thread name to 612a00af-e4af-4260-a4dd-72304
ffd9426 main
2024-09-01 15:19:53,651 INFO ql.Driver: Compiling command(queryId=monik_20240901151953_98924e18-edbc-42c
f-b56c-f84ee90199ed): DESCRIBE finance table
2024-09-01 15:19:53,674 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-01 15:19:53,685 INFO metastore.HiveMetaStore: 0: get_table : tbl=hive.financials.finance_table
2024-09-01 15:19:53,686 INFO HiveMetaStore.audit: ugi=monik
                                                               ip=unknown-ip-addr
                                                                                       cmd=get_table :
tbl=hive.financials.finance_table
2024-09-01 15:19:53,709 INFO parse.DDLSemanticAnalyzer: analyzeDescribeTable done
2024-09-01 15:19:53,710 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2024-09-01 15:19:53,711 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 de_
serializer), FieldSchema(name:comment, type:string, comment:from deserializer)], properties:null)
2024-09-01 15:19:53,714 INFO exec.ListSinkOperator: Initializing operator LIST_SINK[0]
2024-09-01 15:19:53,716 INFO ql.Driver: Completed compiling command(queryId=monik_20240901151953_98924e1
8-edbc-42cf-b56c-f84ee90199ed); Time taken: 0.065 seconds
2024-09-01 15:19:53,720 INFO reexec.ReExecDriver: Execution #1 of query
2024-09-01 15:19:53,721 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-01 15:19:53,722 INFO ql.Driver: Executing command(queryId=monik_20240901151953_98924e18-edbc-42c
f-b56c-f84ee90199ed): DESCRIBE finance_table
2024-09-01 15:19:53,724 INFO ql.Driver: Starting task [Stage-0:DDL] in serial mode
2024-09-01 15:19:53,725 INFO metastore.HiveMetaStore: 0: get_table : tbl=hive.financials.finance_table
2024-09-01 15:19:53,726 INFO HiveMetaStore.audit: ugi=monik
                                                               ip=unknown-ip-addr
                                                                                       cmd=get_table :
tbl=hive.financials.finance table
2024-09-01 15:19:53,756 INFO ql.Driver: Completed executing command(queryId=monik_20240901151953_98924e1
8-edbc-42cf-b56c-f84ee90199ed); Time taken: 0.034 seconds
OK
2024-09-01 15:19:53,758 INFO ql.Driver: OK
2024-09-01 15:19:53,759 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-01 15:19:53,764 INFO mapred.FileInputFormat: Total input files to process : 1
2024-09-01 15:19:53,772 INFO exec.ListSinkOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_LIS
T_SINK_0:2,
id
                        int
                        string
name
Time taken: 0.109 seconds, Fetched: 2 row(s)
2024-09-01 15:19:53,778 INFO CliDriver: Time taken: 0.109 seconds, Fetched: 2 row(s)
2024-09-01 15:19:53,778 INFO conf.HiveConf: Using the default value passed in for log id: 612a00af-e4af
4260-a4dd-72304ffd9426
 Select Administrator: Command Prompt - hive
hive> ALTER TABLE finance_table ADD COLUMNS (age INT);
2024-09-01 15:21:08,854 INFO conf.HiveConf: Using the default value passed in for log id: 612a00af-e4af
4260-a4dd-72304ffd9426
2024-09-01 15:21:08,855 INFO session.SessionState: Updating thread name to 612a00af-e4af-4260-a4dd-72304
2024-09-01 15:21:08,858 INFO ql.Driver: Compiling command(queryId=monik_20240901152108_fe3fbea7-11fb-433
4-b0bd-ef4cc6746f05): ALTER TABLE finance_table ADD COLUMNS (age INT)
2024-09-01 15:21:08,884 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-01 15:21:08,889 INFO metastore.HiveMetaStore: 0: get_table : tbl=hive.financials.finance_table
2024-09-01 15:21:08,890 INFO HiveMetaStore.audit: ugi=monik
                                                                 ip=unknown-ip-addr
                                                                                           cmd=get_table :
tbl=hive.financials.finance table
2024-09-01 15:21:08,911 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2024-09-01 15:21:08,911 INFO ql.Driver: Returning Hive schema: Schema(fieldSchemas:null, properties:null
2024-09-01 15:21:08,913 INFO ql.Driver: Completed compiling command(queryId=monik_20240901152108_fe3fbea
7-11fb-4334-b0bd-ef4cc6746f05); Time taken: 0.054 seconds
2024-09-01 15:21:08,916 INFO reexec.ReExecDriver: Execution #1 of query
2024-09-01 15:21:08,919 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
```



RESULT:

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

