

**EX.NO.5**

**Roll No: 210701173**

**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:**

1. Install Apache Derby using the following link.  
[https://db.apache.org/derby/derby\\_downloads.html#For+Java+8+and+Higher](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  
C:\db-derby-10.14.2.0-bin\bin  
C:\apache-hive-3.1.3-bin\bin
- 3.1 Copy Derby libraries and configure hive-site.xml and Hive's Bin Folder  
<https://1drv.ms/f/s!ArSg3Xpur4Grmw0SDqW0g44T7HYU?e=wDsoBn>
4. Start Hadoop using start-all.cmd
5. Derby Network Server: Run the following command in separate window to open Derby. **StartNetworkServer -h 0.0.0.0**

```
C:\>StartNetworkServer -h 0.0.0.0
Sun Sep 08 23:50:24 IST 2024 : Security manager installed using the Basic server security policy.
Sun Sep 08 23:50:25 IST 2024 : Apache Derby Network Server - 10.14.2.0 - (1828579) started and ready to accept connections on port 1527
```

```
C:\apache-hive-3.1.3-bin\bin>jps
16768 NameNode
22384 NodeManager
17540 DataNode
22212 NetworkServerControl
22404 ResourceManager
9564 Jps
```

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

## hive --service schematool -dbType derby -initSchema

```
C:\apache-hive-3.1.3-bin\bin>hive --service schematool -dbType derby -initSchema
2024-09-09 00:02:24,976 INFO conf.HiveConf: Found configuration file file:/C:/apache-hive-3.1.3-bin/conf/hive-site.xml
2024-09-09 00:02:25,669 INFO tools.HiveSchemaHelper: Metastore connection URL: jdbc:derby:metastore_db;create=true
Metastore connection URL: jdbc:derby:metastore_db;create=true
2024-09-09 00:02:25,670 INFO tools.HiveSchemaHelper: Metastore Connection Driver : org.apache.derby.jdbc.EmbeddedDriver
Metastore Connection Driver : org.apache.derby.jdbc.EmbeddedDriver
2024-09-09 00:02:25,670 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
```

```
Initialization script completed
schemaTool completed
```

### 7. Open Hive shell by typing: **hive**

```
C:\apache-hive-3.1.3-bin\bin>hive
2024-09-09 00:02:39,749 INFO conf.HiveConf: Found configuration file file:/C:/apache-hive-3.1.3-bin/conf/hive-site.xml
Unable to get Charset 'cp65001' for property 'sun.stderr.encoding', using default UTF-8 and continuing.
Hive Session ID = 872f7da8-be15-4062-b911-b351258b1657
2024-09-09 00:02:42,265 INFO SessionState: Hive Session ID = 872f7da8-be15-4062-b911-b351258b1657
```

### 8. Use the CREATE DATABASE statement to create a new database: **CREATE DATABASE financials;**

```
hive> CREATE DATABASE financials;
2024-09-09 00:10:07,732 INFO conf.HiveConf: Using the default value passed in for log id: 872f7da8-be15-4062-b911-b351258b1657
2024-09-09 00:10:07,732 INFO session.SessionState: Updating thread name to 872f7da8-be15-4062-b911-b351258b1657 main
2024-09-09 00:10:07,735 INFO ql.Driver: Compiling command(queryId=sajjad_20240909001007_3b770d5f-369a-4096-a70f-e348df9ca71d): CREATE DATABASE financials
2024-09-09 00:10:07,767 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-09 00:10:07,767 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2024-09-09 00:10:07,769 INFO ql.Driver: Returning Hive schema: Schema(fieldsSchemas:null, properties:null)
2024-09-09 00:10:07,770 INFO ql.Driver: Completed compiling command(queryId=sajjad_20240909001007_3b770d5f-369a-4096-a70f-e348df9ca71d); Time taken: 0.034 seconds
2024-09-09 00:10:07,770 INFO rexec.ReExecDriver: Execution #1 of query
2024-09-09 00:10:07,771 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-09 00:10:07,771 INFO ql.Driver: Executing command(queryId=sajjad_20240909001007_3b770d5f-369a-4096-a70f-e348df9ca71d): CREATE DATABASE financials
2024-09-09 00:10:07,772 INFO ql.Driver: Starting task [Stage:0:001] in serial mode
2024-09-09 00:10:07,773 INFO metastore.HiveMetaStore: 0: create_database: Database(name=financials, description:null, locationUri:hdfs://localhost:9000/user/hive/warehouse/financials.db, parameters:null, ownerName:sajjad, ownerType:USER, catalogName=hive)
2024-09-09 00:10:07,773 INFO HiveMetaStore.audit: ugi=sajjad ip=unknown-ip-addr cmd=create_database: Database(name=financials, description:null, locationUri:hdfs://localhost:9000/user/hive/warehouse/financials.db, parameters:null, ownerName:sajjad, ownerType:USER, catalogName=hive)
2024-09-09 00:10:07,776 WARN metastore.ObjectStore: Failed to get database hive.financials, returning NoSuchObjectException
2024-09-09 00:10:07,788 INFO util.FileUtils: Creating directory if it doesn't exist: hdfs://localhost:9000/user/hive/warehouse/financials.db
2024-09-09 00:10:07,795 INFO ql.Driver: Completed executing command(queryId=sajjad_20240909001007_3b770d5f-369a-4096-a70f-e348df9ca71d); Time taken: 0.024 seconds
OK
2024-09-09 00:10:07,796 INFO ql.Driver: OK
2024-09-09 00:10:07,803 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
Time taken: 0.069 seconds
2024-09-09 00:10:07,805 INFO CliDriver: Time taken: 0.069 seconds
2024-09-09 00:10:07,806 INFO conf.HiveConf: Using the default value passed in for log id: 872f7da8-be15-4062-b911-b351258b1657
2024-09-09 00:10:07,807 INFO session.SessionState: Resetting thread name to main
```

### 9. Switch to the new database: **USE financials;**

```
hive> use financials;
2024-09-09 00:10:16,109 INFO conf.HiveConf: Using the default value passed in for log id: 872f7da8-be15-4062-b911-b351258b1657
2024-09-09 00:10:16,110 INFO session.SessionState: Updating thread name to 872f7da8-be15-4062-b911-b351258b1657 main
2024-09-09 00:10:16,114 INFO ql.Driver: Compiling command(queryId=sajjad_20240909001016_d1375dc9-8a10-4386-8e84-45110ac09cf3): use financials
2024-09-09 00:10:16,135 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-09 00:10:16,137 INFO metastore.HiveMetaStore: 0: get_database: @hive#financials
2024-09-09 00:10:16,138 INFO HiveMetaStore.audit: ugi=sajjad ip=unknown-ip-addr cmd=get_database: @hive#financials
2024-09-09 00:10:16,140 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2024-09-09 00:10:16,140 INFO ql.Driver: Returning Hive schema: Schema(fieldsSchemas:null, properties:null)
2024-09-09 00:10:16,141 INFO ql.Driver: Completed compiling command(queryId=sajjad_20240909001016_d1375dc9-8a10-4386-8e84-45110ac09cf3); Time taken: 0.027 seconds
2024-09-09 00:10:16,141 INFO rexec.ReExecDriver: Execution #1 of query
2024-09-09 00:10:16,142 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-09 00:10:16,142 INFO ql.Driver: Executing command(queryId=sajjad_20240909001016_d1375dc9-8a10-4386-8e84-45110ac09cf3): use financials
2024-09-09 00:10:16,147 INFO ql.Driver: Starting task [Stage:0:001] in serial mode
2024-09-09 00:10:16,147 INFO metastore.HiveMetaStore: 0: get_database: @hive#financials
2024-09-09 00:10:16,148 INFO HiveMetaStore.audit: ugi=sajjad ip=unknown-ip-addr cmd=get_database: @hive#financials
2024-09-09 00:10:16,150 INFO metastore.HiveMetaStore: 0: get_database: @hive#financials
2024-09-09 00:10:16,151 INFO HiveMetaStore.audit: ugi=sajjad ip=unknown-ip-addr cmd=get_database: @hive#financials
2024-09-09 00:10:16,160 INFO ql.Driver: Completed executing command(queryId=sajjad_20240909001016_d1375dc9-8a10-4386-8e84-45110ac09cf3); Time taken: 0.018 seconds
OK
2024-09-09 00:10:16,161 INFO ql.Driver: OK
2024-09-09 00:10:16,164 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
Time taken: 0.051 seconds
2024-09-09 00:10:16,165 INFO CliDriver: Time taken: 0.051 seconds
2024-09-09 00:10:16,166 INFO conf.HiveConf: Using the default value passed in for log id: 872f7da8-be15-4062-b911-b351258b1657
2024-09-09 00:10:16,166 INFO session.SessionState: Resetting thread name to main
```

### 10. Create a Table in Hive using **CREATE TABLE students (name STRING, roll INT, dept STRING);** and insert values into table.

```
hive> CREATE TABLE finance_table( id INT, name STRING );
2024-09-09 00:10:25,738 INFO conf.HiveConf: Using the default value passed in for log id: 872f7da8-be15-4062-b911-b351258b1657
2024-09-09 00:10:25,739 INFO session.SessionState: Updating thread name to 872f7da8-be15-4062-b911-b351258b1657 main
2024-09-09 00:10:25,742 INFO ql.Driver: Compiling command(queryId=sajjad_20240909001025_c3a1aff4-09e5-4081-9069-131426915f75): CREATE TABLE finance_table( id INT, name STRI
NG )
2024-09-09 00:10:25,742 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2024-09-09 00:10:25,742 INFO ql.Driver: Returning Hive schema: Schema(fieldsSchemas:null, properties:null)
2024-09-09 00:10:25,742 INFO ql.Driver: Completed compiling command(queryId=sajjad_20240909001025_c3a1aff4-09e5-4081-9069-131426915f75); Time taken: 0.024 seconds
2024-09-09 00:10:25,742 INFO rexec.ReExecDriver: Execution #1 of query
2024-09-09 00:10:25,742 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-09 00:10:25,742 INFO ql.Driver: Executing command(queryId=sajjad_20240909001025_c3a1aff4-09e5-4081-9069-131426915f75): CREATE TABLE finance_table( id INT, name STRI
NG )
2024-09-09 00:10:25,742 INFO ql.Driver: Starting task [Stage:0:001] in serial mode
2024-09-09 00:10:25,742 INFO metastore.HiveMetaStore: 0: create_table: Table(name=finance_table, description:null, locationUri:hdfs://localhost:9000/user/hive/warehouse/finance_table.tbl, parameters:null, ownerName:sajjad, ownerType:USER, catalogName=hive)
2024-09-09 00:10:25,742 INFO HiveMetaStore.audit: ugi=sajjad ip=unknown-ip-addr cmd=create_table: Table(name=finance_table, description:null, locationUri:hdfs://localhost:9000/user/hive/warehouse/finance_table.tbl, parameters:null, ownerName:sajjad, ownerType:USER, catalogName=hive)
2024-09-09 00:10:25,742 INFO ql.Driver: Completed executing command(queryId=sajjad_20240909001025_c3a1aff4-09e5-4081-9069-131426915f75); Time taken: 0.024 seconds
OK
2024-09-09 00:10:25,742 INFO ql.Driver: OK
2024-09-09 00:10:25,742 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
Time taken: 0.051 seconds
2024-09-09 00:10:25,742 INFO CliDriver: Time taken: 0.051 seconds
2024-09-09 00:10:25,742 INFO conf.HiveConf: Using the default value passed in for log id: 872f7da8-be15-4062-b911-b351258b1657
2024-09-09 00:10:25,742 INFO session.SessionState: Resetting thread name to main
hive> INSERT INTO finance_table VALUES (1, 'Alice'), (2, 'Bob'), (3, 'Charlie');
2024-09-09 00:11:09,038 INFO conf.HiveConf: Using the default value passed in for log id: 872f7da8-be15-4062-b911-b351258b1657
2024-09-09 00:11:09,038 INFO session.SessionState: Updating thread name to 872f7da8-be15-4062-b911-b351258b1657 main
2024-09-09 00:11:09,040 INFO ql.Driver: Compiling command(queryId=sajjad_20240909001109_bec55097-1459-41bd-8a7a-40a76790b1a7): INSERT INTO finance_table VALUES (1, 'Alice')
, (2, 'Bob'), (3, 'Charlie')
2024-09-09 00:11:09,040 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2024-09-09 00:11:09,040 INFO ql.Driver: Returning Hive schema: Schema(fieldsSchemas:null, properties:null)
2024-09-09 00:11:09,040 INFO ql.Driver: Completed compiling command(queryId=sajjad_20240909001109_bec55097-1459-41bd-8a7a-40a76790b1a7); Time taken: 0.024 seconds
2024-09-09 00:11:09,040 INFO rexec.ReExecDriver: Execution #1 of query
2024-09-09 00:11:09,040 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-09 00:11:09,040 INFO ql.Driver: Executing command(queryId=sajjad_20240909001109_bec55097-1459-41bd-8a7a-40a76790b1a7): INSERT INTO finance_table VALUES (1, 'Alice')
, (2, 'Bob'), (3, 'Charlie')
2024-09-09 00:11:09,040 INFO ql.Driver: Starting task [Stage:0:001] in serial mode
2024-09-09 00:11:09,040 INFO metastore.HiveMetaStore: 0: insert_into: Table(name=finance_table, description:null, locationUri:hdfs://localhost:9000/user/hive/warehouse/finance_table.tbl, parameters:null, ownerName:sajjad, ownerType:USER, catalogName=hive)
2024-09-09 00:11:09,040 INFO HiveMetaStore.audit: ugi=sajjad ip=unknown-ip-addr cmd=insert_into: Table(name=finance_table, description:null, locationUri:hdfs://localhost:9000/user/hive/warehouse/finance_table.tbl, parameters:null, ownerName:sajjad, ownerType:USER, catalogName=hive)
2024-09-09 00:11:09,040 INFO ql.Driver: Completed executing command(queryId=sajjad_20240909001109_bec55097-1459-41bd-8a7a-40a76790b1a7); Time taken: 0.024 seconds
OK
2024-09-09 00:11:09,040 INFO ql.Driver: OK
2024-09-09 00:11:09,040 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
Time taken: 0.051 seconds
2024-09-09 00:11:09,040 INFO CliDriver: Time taken: 0.051 seconds
2024-09-09 00:11:09,040 INFO conf.HiveConf: Using the default value passed in for log id: 872f7da8-be15-4062-b911-b351258b1657
2024-09-09 00:11:09,040 INFO session.SessionState: Resetting thread name to main
```

**11.** To display table data, use a **SELECT** statement. For example, to select everything in a table, run: **SELECT \* FROM myview;**

```
hive> SELECT*FROM myview;
2024-09-09 00:12:43,848 INFO conf.HiveConf: Using the default value passed in for log id: 872f7da8-be15-4062-b911-b351258b1657
2024-09-09 00:12:43,849 INFO session.SessionState: Updating thread name to 872f7da8-be15-4062-b911-b351258b1657 main
2024-09-09 00:12:43,851 INFO ql.Driver: Compiling command(queryId=sajjad_20240909001243_ae937c50-a94a-41a4-8a44-05cfa1547dd6): SELECT*FROM myview
2024-09-09 00:12:43,872 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-09 00:12:44,248 INFO ql.Driver: OK
2024-09-09 00:12:44,250 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-09 00:12:44,262 INFO mapred.FileInputFormat: Total input files to process : 1
2024-09-09 00:12:44,288 INFO exec.TableScanOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_TS_0:3,
2024-09-09 00:12:44,288 INFO exec.SelectOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_SEL_1:3,
2024-09-09 00:12:44,294 INFO exec.ListSinkOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_LIST_SINK_3:3,
Alice 1
Bob 2
Charlie 3
Time taken: 0.4 seconds, Fetched: 3 row(s)
2024-09-09 00:12:44,309 INFO CliDriver: Time taken: 0.4 seconds, Fetched: 3 row(s)
2024-09-09 00:12:44,311 INFO conf.HiveConf: Using the default value passed in for log id: 872f7da8-be15-4062-b911-b351258b1657
2024-09-09 00:12:44,312 INFO session.SessionState: Resetting thread name to main
```

## RESULT:

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