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:

1. Download Hive and Derby

- Apache Hive: Download from Hive Download
- Apache Derby: Download from Derby Download

Extract both .tar.gz files after downloading.

2. Move Hive and Derby to /usr/local/

Open Terminal.

Navigate to the folder where the downloaded files are located. For example, if they are in your Downloads folder:

```
cd ~/Downloads
```

Move the extracted Hive and Derby folders to /usr/local/.

For Hive:

```
sudo mv apache-hive-3.1.3-bin /usr/local/
For Derby:
```

```
sudo mv db-derby-10.14.2.0-bin /usr/local/
```

Verify that the folders were successfully moved:

```
cd /usr/local/
ls
```

You should see apache-hive-3.1.3-bin and db-derby-10.14.2.0-bin listed.

3. Set Environment Variables

Open your ._profile or .zshrc (depending on your shell) in a text editor. Most likely, you are using Zsh on newer macOS versions, so edit .zshrc:

```
nano ~/.zshrc
```

Add the following lines to set the environment variables for Hive and Derby:

```
# Set Hive environment variables
export HIVE_HOME=/usr/local/apache-hive-3.1.3-bin
export PATH=$PATH:$HIVE_HOME/bin

# Set Derby environment variables
export DERBY_HOME=/usr/local/db-derby-10.14.2.0-bin
export PATH=$PATH:$DERBY_HOME/bin
```

Save and exit (Ctrl + 0 to save, then Ctrl + X to exit).

Apply the changes by running the following command:

```
source ~/.zshrc
cd /usr/local/db-derby-10.14.2.0-bin/lib
cp *.jar /usr/local/apache-hive-3.1.3-bin/lib/
```

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

Download all

Move and Replace hive-site.xml

Navigate to Hive Configuration Directory:

```
cd /usr/local/apache-hive-3.1.3-bin/conf
```

Backup Existing hive-site.xml (if it exists):

```
mv hive-site.xml hive-site.xml.bak
```

Copy the Downloaded hive-site.xml from Downloads:

```
cp ~/Downloads/hive-site.xml .
```

Replace the Guava Library

Navigate to Hive's Libraries Directory:

```
cd /usr/local/apache-hive-3.1.3-bin/lib
```

Backup Existing Guava Library (if it exists):

```
mv guava-*.jar guava-*.jar.bak
```

Copy the Downloaded Guava Library from Downloads:

```
cp ~/Downloads/guava-*.jar .
```

Ensure you replace guava-*.jar with the actual filename of the Guava library you downloaded.

Replace the Bin Folder

Navigate to Hive Installation Directory:

```
cd /usr/local/apache-hive-3.1.3-bin
```

Backup Existing bin Directory:

mv bin bin.bak

Copy the Downloaded bin Directory from Downloads:

```
cp -r ~/Downloads/bin .
```

Verify Installation

Navigate to Hive's Bin Directory:

```
cd /usr/local/apache-hive-3.1.3-bin/bin
```

Check Hive Version to Verify Installation:

```
cd /usr/local/apache-hive-3.1.3-bin/bin
chmod +x hive
```

4. Test Hive and Derby Installation

To test if Hive is working, run:

```
hive --version
```

You should see the version of Hive that you installed.

To test if Derby is working, run:

```
sudo java -jar $DERBY_HOME/lib/derbyrun.jar server start
```

You should see a message saying the Derby server started successfully.

5. Start Hadoop Services (if Hadoop is Installed)

If Hadoop is set up on your system, you'll need to start the Hadoop services for Hive to run:

Open a new terminal and go to Hadoop's sbin directory:

```
cd /path/to/hadoop/sbin
Start the Hadoop services:
```

```
start-dfs.sh
start-yarn.sh
```

6. Configure Hive Schema with Derby

Now, initialize the Hive schema with Derby as the metastore:

Open a new terminal and run the following command to initialize the schema:

```
hive --service schematool -dbType derby -initSchema
```

7. Open the Hive Shell

Now, you can open the Hive shell and start working with Hive:

Open the Hive shell by typing:

hive

Verify that Hive is running correctly by creating a database and listing it:

```
CREATE DATABASE financials;
SHOW DATABASES;
```

```
hive> CREATE DATABASE financials;
2024-08-30123:35:44,065 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c29b-4d8f-b0b3-63904207bbe2
2024-08-30123:35:44,066 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main
DK
Time taken: 0.046 seconds
2024-08-30123:35:44,119 INFO [1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c
20b-4d8f-b0b3-63904207bbe2
2024-08-30123:35:44,119 INFO [1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.gl.session.SessionState - Resetting thread name to main
```

```
Alton Show databases;
2824-89-80 19:28:14,966 INFO Q1.Driver: Compiling command(queryId-user_2024098192814_5a0bb98c-9d8d-41ec-9ae-4b59b619930a): show databases
2824-89-80 19:28:14,966 INFO Q1.Driver: Compiling command(queryId-user_2024098192814_5a0bb98c-9d8d-41ec-9ae-4b59b619930a): show databases
2824-89-80 19:28:15,666 INFO Q1.Driver: Compiling command(queryId-user_2024098192814_5a0bb98c-9d8d-41ec-9ae-4b59b619930a): show databases
2824-89-80 19:28:15,958 INFO Q1.Driver: Returning Hive schema: Schema(fieldSchemas): [FieldSchema(name:database_name, type:string, comment:from deserializer)], properties:null)
2824-89-80 19:28:15,959 INFO Q1.Driver: Compiling command(queryId-user_20240986192814_5a0bb98c-9d8d-41ec-9ae0-4b59b629930a); Time taken: 1.137 seconds
2824-89-80 19:28:16,020 INFO Q1.Driver: Compiled compiling command(queryId-user_20240986192814_5a0bb98c-9d8d-41ec-9ae0-4b59b629930a); Time taken: 1.137 seconds
2824-89-80 19:28:16,020 INFO Q1.Driver: Executing command(queryId-user_20240986192814_5a0bb98c-9d8d-41ec-9ae0-4b59b629930a); show databases
2824-89-80 19:28:16,020 INFO Q1.Driver: Executing command(queryId-user_20240986192814_5a0bb98c-9d8d-41ec-9ae0-4b59b629930a); show databases
2824-89-80 19:28:16,692 INFO Q1.Driver: Executing command(queryId-user_20240986192814_5a0bb98c-9d8d-41ec-9ae0-4b59b629930a); show databases
2824-89-80 19:28:16,695 INFO HiveNetSatore.audit: ugi-user ip-unknown-ip-addr cmd=get_databases: (Mive#
2824-89-80 19:28:16,695 INFO HiveNetSatore.audit: ugi-user ip-unknown-ip-addr cmd=get_databases: (Mive#
2824-89-80 19:28:16,695 INFO Q1.Driver: Completed executing command(queryId-user_20240908192814_5a0bb98c-9d8d-41ec-9ae0-4b59b629930a); Time taken: 0.1 seconds
0K
2824-89-80 19:28:16,595 INFO Q1.Driver: Completed executing command(queryId-user_20240908192814_5a0bb98c-9d8d-41ec-9ae0-4b59b629930a); Time taken: 0.1 seconds
0K
2824-89-80 19:28:16,595 INFO Q1.Driver: Completed executing command(queryId-user_20240908192814_5a0bb98c-9d8d-41ec-9ae0-4b59b629930a); Time taken: 0.1 seconds
0K
2824-89-
```

8. Create a Table and Insert Data

Use the following to create a simple table in Hive:

USE financials;

```
hive> use financials;
2024-08-30123:36:00,802 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c29b-4d8f-b0b3-63904207bbe2
2024-08-30123:36:00,802 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main

OK
Time taken: 0.048 seconds
2024-08-30123:36:00,917 INFO [1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to main
```

CREATE TABLE students_table (id INT, name STRING);

```
hive> CREATE TABLE finance_table(id INT, name STRING);
2024-08-30123:36:32,350 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c29b-4d8f-b0b3-63904207bbe2
2024-08-30123:36:32,351 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main
OK
Time taken: 0.678 seconds
2024-08-30123:36:32,439 INFO [1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c29b-4d8f-b0b3-63904207bbe2
2024-08-30123:36:32,439 INFO [1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to main
```

Insert some sample data:

```
INSERT INTO students_table VALUES (1, 'Alice'), (2, 'Bob'), (3,
'Charlie');
```

```
hive> INSERT INTO finance_table VALUES (1,'Alice'), (2,'Bob'), (3,'Charlie');
2024-08-30T23:37:36,349 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c29b-4d8f-b0b3-63904207bbe2
2024-08-30T23:37:36,149 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main
2024-08-30T23:37:41,265 INFO [1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localh
ost:9000/user/hive/warehouse/finance_table_hive-staging_nive_2024-08-30_23-37-6_109_211810260597313358-1
2024-08-30T23:37:41,461 INFO [1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localh
ost:9000/Imp/hive/user/1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localh
ost:9000/Imp/hive/user/1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localh
ost:9000/Imp/hive/user/1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localh
ost:9000/Imp/hive/user/1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localh
ost:9000/Imp/hive/user/1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localh
ost:9000/Imp/hive/user/1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localh
ost:9000/Imp/hive/user/1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localh
ost:9000/Imp/hive/user/1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoo
```

```
Launching Job 1 out of 3
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
set hive.exec.reducers.bytes.per.reducer-cnumber>
In order to change the average load for a reducer (in bytes):
set hive.exec.reducers.max-cnumber of reducers:
set hive.exec.reducers.max-cnumber>
In order to set a constant number of reducers:
set mipreduce.job.reduces-cnumber>
2024-08-3012;37:42,285 INFO [1622e2df-c20b-d408f-b0b3-63904207bbe2 main] org.apache.hadoop.conf.Configuration.deprecation - mapred.submit.replication is deprecated. Instead, use mappeduce.job.Info [162e2df-c20b-d408f-b0b3-63904207bbe2 main] org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher consider is deprecated. Instead, use yarn.system-metrics-publisher enabled is deprecated. Instead use yarn.system-metrics-publisher enabled is deprecated. Instead use yarn.system-metrics-publisher enabled is deprecated. Instead use yarn.system-metrics-publisher enabled 2024-08-30 / job 1725e002728625 0001. Tracking URL = http://Doss/0756570-PURSS(07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.8008/07.
```

9. Query the Data

Query the data to verify everything is working:

```
SELECT * FROM students_table;
```

```
2024-09-10 21:55:31.579 INFO mapred_FileInputFormat: Total input files to process: 1
2024-09-10 21:55:31.591 INFO exec.TableScanGperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_TS_0:3,
2024-09-10 21:55:31.591 INFO exec.SlactOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_SEL_1:3,
2024-09-10 21:55:31.591 INFO exec.SlactOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_LIST_SINK_3:3,

Alice
1 Alice
2 Charlie
3 Charlie
1 Time taken: 0.356 seconds, Fetched: 3 row(s)
2024-09-10 21:55:31.593 INFO CliDriver: Time taken: 0.354 seconds, Fetched: 3 row(s)
2024-09-10 21:55:31.593 INFO Conf.HiveConf. Using the default value passed in for log id: 79e4878f-fb4d-4f68-9f0b-ba4ecde738ef
2024-09-10 21:55:31.593 INFO conf.HiveConf. Using the default value passed in for log id: 79e4878f-fb4d-4f68-9f0b-ba4ecde738ef
```

10. Shut Down Hive and Derby

Exit the Hive shell by typing quit;.

Stop the Derby server by running:

sudo java -jar \$DERBY_HOME/lib/derbyrun.jar server shutdown

OUTPUT:

```
nativewit@Nativewits-MacBook-Air ~ % cd /usr/local/db-derby-10.14.2.0/lib
  cd: no such file or directory: /usr/local/db-derby-10.14.2.0/lib
nativewitcNetivewits-MacBook-Air - % cd /usr/local/db-derby-10.14.2.0/lib
nativewitcNetivewits-MacBook-Air local % le
Caskroom
Gellar derby-10g sbin
Frameworks et share
Homebrew include var
spache-hive-3.1.3-bin lib
spache-hive-3.1.3-bin lib
nativewitcNetivewits-MacBook-Air local % cd /usr/local/spache-hive-3.1.3-bin/lib
nativewitcNetivewits-MacBook-Air lib % cp *.jar /usr/local/spache-hive-3.1.3-bin/lib
    nativewit@Nativewits-MacBook-Air lib % cd /usr/local/apache-hive-3.1.3-bin/conf
   nativewit@Nativewits-MacBook-Air conf % mv hive-site.xml hive-site.xml.bak
mv: rename hive-site.xml to hive-site.xml.bak: No such file or directory nativewit@Nativewits-MacBook-Air conf % cp ~/Downloads/hive-site.xml .
cp: /Users/nativewit/Downloads/hive-site.xml: No such file or directory
nativewit@Nativewits-MacBook-Air conf % cp ~/Downloads/hive-site.xml .
  nativewit@Nativewits-MacBook-Air conf % cd /usr/local/apache-hive-3.1.3-bin/lib
     ativewit@Nativewits-MacBook-Air lib % mv guava-*.jar guava-*.jar.bak
 zsh: no matches found: guava-*.jar.bak
|nativewit@Nativewits-MacBook-Air lib % cp ~/Downloads/guava-*.jar .
  nativewit@Nativewits-MacBook-Air lib % cd /usr/local/apache-hive-3.1.3-bin
   nativewit@Nativewits-MacBook-Air apache-hive-3.1.3-bin % cp -r ~/Downloads/bin .
     nativewit@Nativewits-MacBook-Air apache-hive-3.1.3-bin % cd /usr/local/apache-hive-3.1.3-bin/bin
   nativewit@Nativewits-MacBook-Air bin % ./hive --version
zsh: permission denied: ./hive
nativewit@Nativewits-MacBook-Air bin % cd /usr/local/apache-hive-3.1.3-bin/bin
    nativewit@Nativewits-MacBook-Air bin % ./hive --version
  NEAT: Class path contains multiple SLFAD bindings.

SLFAD: Isse path contains multiple SLFAD bindings.

SLFAD: A part of the substitute of
```

```
The control of the Set 38 212234 on type 18 control of the set 38 control of the set 38
```

RESULT:

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