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 <u>Hive Download</u>
- Apache Derby: Download from <u>Derby Download</u>

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 <code>Downloads</code> 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:

```
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;
2024-08-30723-35-44,067 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c29b-4d8f-b0b3-63904207bbe2
2024-08-30723-35-44,068 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main
OK
Time taken: 0.046 seconds
2024-08-30723-35-44,19 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
```

```
Niver show databases;
2024-0-80 19:28:14,775 INFO conf.HiveConf: Using the default value passed in for log id: 168ffab1-50a4-42fa-b931-df2184794e2e
2024-0-80 19:28:14,965 INFO ql.Driver: Computing command(queryId-user_20240908192814_5a0bb98c-9d8d-41ec-9ae0-4059b629930a); show databases
2024-0-9-08 19:28:15,605 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-0-9-08 19:28:15,605 INFO ql.Driver: Securing River schematic Analysis Completed (retrial = false)
2024-0-9-08 19:28:15,005 INFO ql.Driver: Returning River schematic Scheman(fieldScheman:[FieldScheman(name:database_name, type:string, comment:from deserializer)], properties:null)
2024-0-9-08 19:28:15,005 INFO ql.Driver: Returning River schematic Analysis (completed compiling command(queryId-user_20240908192814_5a0bb98c-9d8d-41ec-9ae0-4b59b629930a); Time taken: 1.137 seconds
2024-0-9-08 19:28:16,007 INFO ql.Driver: Completed compiling command(queryId-user_20240908192814_5a0bb98c-9d8d-41ec-9ae0-4b59b629930a); Fine taken: 1.137 seconds
2024-0-9-08 19:28:16,009 INFO ql.Driver: Executing command(queryId-user_20240908192814_5a0bb98c-9d8d-41ec-9ae0-4b59b629930a); show databases
2024-0-9-08 19:28:16,009 INFO ql.Driver: Executing command(queryId-user_20240908192814_5a0bb98c-9d8d-41ec-9ae0-4b59b629930a); show databases
2024-0-9-08 19:28:16,009 INFO ql.Driver: Executing command(queryId-user_20240908192814_5a0bb98c-9d8d-41ec-9ae0-4b59b629930a); Time taken: 0.1 seconds
2024-0-9-08 19:28:16,009 INFO ql.Driver: Completed executing command(queryId-user_20240908192814_5a0bb98c-9d8d-41ec-9ae0-4b59b629930a); Time taken: 0.1 seconds
2024-0-9-08 19:28:16,009 INFO ql.Driver: Completed executing command(queryId-user_20240908192814_5a0bb98c-9d8d-41ec-9ae0-4b59b629930a); Time taken: 0.1 seconds
2024-0-9-08 19:28:16,009 INFO ql.Driver: Completed executing command(queryId-user_20240908192814_5a0bb98c-9d8d-41ec-9ae0-4b59b629930a); Time taken: 0.1 seconds
2024-0-9-08 19:28:16,009 INFO ql.Driver: Completed executing command(queryId-user_20240908192814_5a0bb98
```

8. Create a Table and Insert Data

Use the following to create a simple table in Hive:

```
USE financials;
```

```
hive> use financials;
2024-08-30723:36:00,882 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c29b-4d8f-b0b3-63904207bbe2
2024-08-30723:36:00,862 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
2024-08-30723:36:00,917 INFO [1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c
2024-085-30723:36:00,918 INFO [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);
2024-08-30T23:36:32,350 IMFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c29b-4d8f-b0b3-63904207bbe2
2024-08-30T23:36:32,351 IMFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main
OK
Time taken: 0.078 seconds
2024-08-30T23:36:32,439 IMFO [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
2024-08-30T23:36:32,440 IMFO [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,'80b'), (3,'Charlie');

2024-08-30723:37:36,149 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c29b-4d8f-b0b3-63904207bbe2

2024-08-30723:37:36,149 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main

2024-08-30723:37:41,255 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/financials.db/finance_table/.hive-staging_hive_2024-08-30_23-37-36_169_2114510206597313358-1

2024-08-30723:741,461 INFO [1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localh

ost:9000/tmp/hive/user/ic22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localh

ost:9000/tmp/hive/user/ic22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localh

ost:9000/tmp/hive/user/ic22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localh

ost:9000/tmp/hive/user/ic22e2df-c29b-4d8f-b0b3-63904207bbe2 main]

org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localh

ost:9000/tmp/hive/user/ic22e2df-c29b-4d8f-b0b3-63904207bbe2 main]

org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localh

ost:9000/tmp/hive/user/ic22e2df-c29b-4d8f-b0b3-63904207bbe2 main]

org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localh

ost:9000/tmp/hive/user/ic22e2df-c29b-4d8f-b0b3-63904207bbe2 main]

org.apache.hadoop.micleUtils - Creating directory if it doesn't exist: hdfs://localh

ost:9000/tmp/hive/user/ic22e2df-c29b-4d8f-b0b3-63904207bbe2 m
```

```
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. reducers.
set hive exec. reducers. maker.
In order to limit the maximum number of reducers:
set hive exec. reducers. maker.
set hive exec. reducers. maker.

10 redre to set a constant number of reducers:
set mapreduce. job. reducers. rounders.

2024-08-30723:37:42,285 INFO [1c22e2df-c29b-4d8f-bbb3-63904207bb2 main] org.apache.hadoop.conf.Configuration.deprecation - mapred.submit.replication is deprecated. Instead, use mapreduce.client. usbmit.file.replication
2024-08-30723:37:40,235 INFO [1c22e2df-c29b-4d8f-bbb3-63904207bb2 main] org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use semine-risc-spublisher.enabled as deprecated. Instead, use yarn.system-metrics-publisher.enabled as deprecated. Instead, use yarn.system-metrics-publisher.enabled 2024-08-3072:37:40,791 INFO [1c22e2df-c29b-4d8f-bbb3-63904207bb2 main] org.apache.hadoop.conf.Configuration - resource-types.xml not found 5tarting Job - job.1725040228025.001 Instead. Job - Job. 1725040228025.001 Instead. Job - Job.1725040228025.001 Instead. Job - Job -
```

9. Query the Data

Query the data to verify everything is working:

```
SELECT * FROM students table;
```

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
   od: no such file or directory (sur/local/db-datby-0.14.2.0/lib
nativewitiNativewitt-MacBook-Air - % od /usr/local/db-datby-0.14.2.0/lib
nativewitiNativewitt-MacBook-Air local %
Caskroom db-datby-10.14.2.0-bin pig
Callar derby-10g sbin
Frameworks etc share
flowebrew include var
niclude var
nativewitiNativewitt-MacBook-Air local % od /usr/local/db-datby-10.14.2.0-bin/lib
nativewitiNativewitt-MacBook-Air local % od /usr/local/dpache-hive-3.1.3-bin/lib
nativewitiNativewitt-MacBook-Air local % x + jar /usr/local/apache-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 \mbox{\ensuremath{^{\prime\prime}}}\xspace cp \mbox{\ensuremath{^{\prime\prime}}}\xspace cp \mbox{\ensuremath{^{\prime\prime}}}\xspace conf % cp \mbox{\ensuremath{^{\prime\prime}}}\xspace con % conf % cp \mbox{\ensuremath{^{\prime\prime}}}\xspace conf % cp \mbox{\ensuremath{^{\prime\prime}}}\xspace con % cp \mbox{\ensuremath{^{\prime\prime}}}\xspace con % conf % conf % cp \mbox{\ensuremath{^
cp: /Users/nativewit/Downloads/hive-site.xml: No such file or directory nativewit@Nativewits-MacBook-Air conf % cp \sim/Downloads/hive-site.xml .
    nativewit@Nativewits-MacBook-Air conf % cd /usr/local/apache-hive-3.1.3-bin/lib
    nativewit@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 .
       ativewit@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
      ativewit@Nativewits-MacBook-Air bin % chmod +x hive
    nativewit@Nativewits-MacBook-Air bin % ./hive --version
 Institution of the Company of the Co
```

```
Inst legic. The fee 38 1131370 on trys00 minimized (Minimize Angloon, Ar. 8 )

of /mar/local/Callar/Angloo/1.4 4/lbsene/shi
mativest@Minimize Angloon, Ar. 8 )

string nameson of localboat!

Starting nameson of localboat!
```

```
2024—09-10 21:55:31,579 INFO mapred.FileInputFormat: Total input files to process: 1
2024—09-10 21:55:33,591 INFO exec.lableScanOparator: RECOMDS_OUT_INTERMEDIATE:0, RECOMDS_OUT_OPERATOR_TS_0:3,
2024—09-10 21:55:33,591 INFO exec.sleatOparator: RECOMDS_OUT_INTERMEDIATE:0, RECOMDS_OUT_OPERATOR_SEL_1:3,
2024—09-10 21:55:31,591 INFO exec.ListSinkOparator: RECOMDS_OUT_INTERMEDIATE:0, RECOMDS_OUT_OPERATOR_LIST_SINK_3:3,
1 Alice
2 Bob
3 Charlie
Time taken: 0.356 seconds, Fetched: 3 row(s)
2024—09-10 21:55:31,593 INFO Clibriver: Time taken: 0.356 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
hive> 1
```

RESULT:

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