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
```

Saveandexit(`Ctrl+O` tosave,then `Ctrl+X` toexit).

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> show databases;
2024-09-08 19:28:14,778 INFO conf.HiveConf: Using the default value passed in for log id: 168ffab1-50a4-42fa-b931-df2184794e2e
2024-09-08 19:28:14,966 INFO ql.Driver: Compiling command(queryId=user_20240908192814_5a0bb98c-9d8d-41ec-9ae0-4b59b629930a): show databases
2024-09-08 19:28:15,646 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-08 19:28:15,687 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2024-09-08 19:28:15,816 INFO ql.Driver: Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:database_name, type:string, comment:from deserializer)], properties:null)
2024-09-08 19:28:15,996 INFO exec.ListSinkOperator: Initializing operator LIST_SINK[0]
2024-09-08 19:28:16,017 INFO ql.Driver: Completed compiling command(queryId=user_20240908192814_5a0bb98c-9d8d-41ec-9ae0-4b59b629930a); Time taken: 1.137 seconds
2024-09-08 19:28:16,017 INFO reexec.ReExecDriver: Execution #1 of query
2024-09-08 19:28:16,020 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-08 19:28:16,020 INFO ql.Driver: Executing command(queryId=user_20240908192814_5a0bb98c-9d8d-41ec-9ae0-4b59b629930a): show databases
2024-09-08 19:28:16,047 INFO ql.Driver: Starting task [Stage-0:DDL] in serial mode
2024-09-08 19:28:16,050 INFO metastore.HiveMetaStore: 0: get_databases: @hive#
2024-09-08 19:28:16,050 INFO HiveMetaStore.audit: ugi=user      ip=unknown-ip-addr      cmd=get_databases: @hive#
2024-09-08 19:28:16,053 INFO exec.DDLTask: results : 3
2024-09-08 19:28:16,120 INFO ql.Driver: Completed executing command(queryId=user_20240908192814_5a0bb98c-9d8d-41ec-9ae0-4b59b629930a); Time taken: 0.1 seconds
OK
2024-09-08 19:28:16,122 INFO ql.Driver: OK
2024-09-08 19:28:16,125 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-08 19:28:16,146 INFO Configuration.deprecation: mapred.input.dir is deprecated. Instead, use mapreduce.input.fileinputformat.inputdir
2024-09-08 19:28:16,216 INFO mapred.FileInputFormat: Total input files to process : 1
2024-09-08 19:28:16,268 INFO exec.ListSinkOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_LIST_SINK_0:3,
default
financials
test
Time taken: 1.25 seconds, Fetched: 3 row(s)
2024-09-08 19:28:16,288 INFO CliDriver: Time taken: 1.25 seconds, Fetched: 3 row(s)
2024-09-08 19:28:16,289 INFO conf.HiveConf: Using the default value passed in for log id: 168ffab1-50a4-42fa-b931-df2184794e2e
2024-09-08 19:28:16,291 INFO session.SessionState: Resetting thread name to  main
```

8. Create a Table and Insert Data

Use the following to create a simple table in Hive:

```
USE financials;
```

```
hive> use financials;
2024-08-30T23:36:00,862 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c29b-4d8f-b0b3-63904207bbe2
2024-08-30T23:36:00,864 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
2024-08-30T23:36:00,917 INFO [1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c
29b-4d8f-b0b3-63904207bbe2
2024-08-30T23:36:00,918 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 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c29b-4d8f-b0b3-63904207bbe2
2024-08-30T23:36:32,351 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main
OK
Time taken: 0.078 seconds
2024-08-30T23:36:32,439 INFO [1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c
29b-4d8f-b0b3-63904207bbe2
2024-08-30T23:36:32,440 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');
2024-08-30T23:37:36,149 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c29b-4d8f-b0b3-63904207bbe2
2024-08-30T23:37:36,149 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/financials.db/finance_table/.hive-staging_hive_2024-08-30_23-37-36_169_2114510206597313358-1
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/tmp/hive/user/1c22e2df-c29b-4d8f-b0b3-63904207bbe2/hive_2024-08-30_23-37-41_328_123745459230859352 8-1/-mr-10000/.hive-staging_hive_2024-08-30_23-37-41_328_1237
45459230859352 8-1
Query ID = user_20240830233736_01a5737e-9654-4013-bcdd-9071c1f42090
Total jobs = 3
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=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
```

```
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=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
2024-08-30T23:37:42,285 INFO [1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.conf.Configuration.deprecation - mapred.submit.replication is deprecated. Ins
tead, use mapreduce.client.submit.file.replication
2024-08-30T23:37:49,235 INFO [1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publishe
r.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
2024-08-30T23:37:49,791 INFO [1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.conf.Configuration - resource-types.xml not found
Starting Job = job_1725040228625_0001, Tracking URL = http://DESKTOP-DKU83QM:8088/proxy/application_1725040228625_0001/
Kill Command = C:\hadoop\bin\mapred job  -kill job_1725040228625_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2024-08-30 23:38:30,410 Stage-1 map = 0%,  reduce = 0%
2024-08-30 23:38:51,438 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.092 sec
2024-08-30 23:39:16,522 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 11.215 sec
MapReduce Total cumulative CPU time: 11 seconds 215 msec
Ended Job = job_1725040228625_0001
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://localhost:9000/user/hive/warehouse/financials.db/finance_table/.hive-staging_hive_2024-08-30_23-37-36_169_2114510206597313358-1/-ext-100
00
Loading data to table financials.finance_table
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 11.215 sec   HDFS Read: 15666 HDFS Write: 291 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 215 msec
OK
Time taken: 107.363 seconds
2024-08-30T23:39:23,519 INFO [1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c
29b-4d8f-b0b3-63904207bbe2
2024-08-30T23:39:23,519 INFO [1c22e2df-c29b-4d8f-b0b3-63904207bbe2 main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to  main
```

9. Query the Data

Query the data to verify everything is working:

```
SELECT * FROM students_table;
```

10. Shut Down Hive and Derby

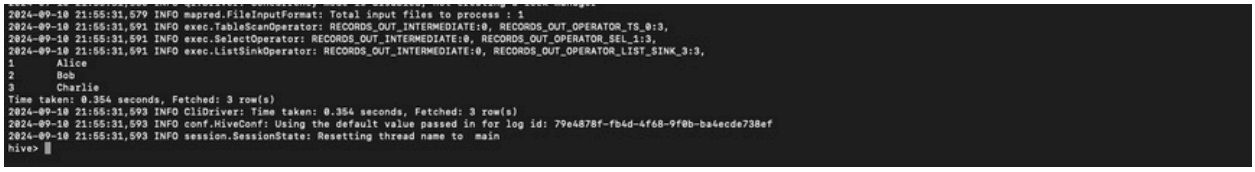
Exit the Hive shell by typing `quit;`.

Stop the Derby server by running:

```
sudojava -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
nativewit@Nativewits-MacBook-Air ~ % cd /usr/local/
nativewit@Nativewits-MacBook-Air local % ls
Caskroom db-derby-10.14.2.0-bin pig
Cellar derby.log sbin
Frameworks etc share
Homebrew include var
apache-hive-3.1.3-bin lib
bin opt
nativewit@Nativewits-MacBook-Air local % cd /usr/local/db-derby-10.14.2.0-bin/lib
nativewit@Nativewits-MacBook-Air lib % cp *.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 ~/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
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 % mv bin bin.bak
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 % chmod +x hive
nativewit@Nativewits-MacBook-Air bin % ./hive --version
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/Cellar/hadoop/3.4.0/libexec/share/hadoop/common/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/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]
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/Cellar/hadoop/3.4.0/libexec/share/hadoop/common/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/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]
Hive 3.1.3
Git git://MacBook-Pro.fios-router.home/Users/ngangam/commit/hive -r 4df4d75bf1e16fe0af75aad0b4179c34c07fc975
Compiled by ngangam on Sun Apr 3 16:58:16 EDT 2022
From source with checksum 5da234766db5dfbe3e92926c9bbab2af
nativewit@Nativewits-MacBook-Air bin % sudo java -jar $DERBY_HOME/lib/derbyrun.jar server start



RESULT:

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