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/apachehive-3.1.3-bin/lib/https://ldrv.ms/f/s!ArSg3Xpur4Grmw0SDqW0g44T7HYU?e=wDsoB

n

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:

my 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;
```

```
New SCEATE DATABASE Flamacials;

2024-08-30723:55:44,068 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c29b-4d8f-b0b1-63904207bbe2 main org. apache.hadoop.hive.ql.session.SessionState - Updating thread name to 1c22e2df-c29b-4d8f-b0b1-63904207bbe2 main org. apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c29b-4d8f-b0b1-63904207bbe2 main org. apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c29b-4d8f-b0b1-63904207bbe2 main org. apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22e2df-c29b-4d8f-b0b1-63904207bbe2 main org. apache.hadoop.hive.gl.session.SessionState - Resetting thread name to main hive show databases;
10024-09-08 139:2814,778 INFO (1c22e2df-c29b-4d8f-b0b1-63904207bbe2 main) org. apache.hadoop.hive.gl.session.SessionState - Resetting thread name to main hive show databases;
10024-09-08 139:2814,778 INFO (1c22e2df-c29b-4d8f-b0b1-63904207bbe2 main) org. apache.hadoop.hive.gl.session.SessionState - Resetting thread name to main hive show databases;
10024-09-08 139:2814,778 INFO (1c22e2df-c29b-4d8f-b0b1-63904207bbe2 main) org. apache.hadoop.hive.gl.session.SessionState - Resetting thread name to main hive show databases;
10024-09-08 139:2815,06 INFO ql.Driver: Completed compiling command(queryId-user_2024098139218_5908b98c-9d8d-41cc-9ae0-4059062930a); show databases
10024-09-08 139:2815,06 INFO ql.Driver: Completed compiling command(queryId-user_20240908139218_5908b98c-9d8d-41cc-9ae0-4059062930a); show databases
10024-09-08 139:2816,00 INFO ql.Driver: Starting task [Stage-01001] in serial mode
10024-09-08 139:2816,00 INFO ql.Driver: Starting task [Stage-01001] in serial mode
10024-09-08 139:2816,00 INFO ql.Driver: Starting task [Stage-01001] in serial mode
10024-09-08 139:2816,00 INFO ql.Driver: Completed executing command(queryId-user_20240908139218_5908098c-9d8d-41cc-9ae0-4059062930a); Time taken: 0.1 seconds
10024-09-08 139:2816
```

8. Create a Table and Insert Data

Use the following to create a simple table in Hive:

```
USE financials;
```

```
hive> use financials;

2024-08-30723:36:00,002 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22c2df-c29b-4d8f-b0b3-63904207bbc2 2024-08-30723:36:00,002 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 1c22c2df-c29b-4d8f-b0b3-63904207bbc2 main log.

2024-08-30723:36:00,917 INFO [1c22c2df-c29b-4d8f-b0b3-63904207bbc2 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22c2df-c29b-4d8f-b0b3-63904207bbc2 [2024-08-30723:36:00,918 INFO [1c22c2df-c29b-4d8f-b0b3-63904207bbc2 main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to main

CREATE TABLE students table (id INT, name STRING);

2024-08-30723:36:32,350 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22c2df-c29b-4d8f-b0b3-63904207bbc2 2024-08-30723:35:32,351 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 1c22c2df-c29b-4d8f-b0b3-63904207bbc2 2024-08-30723:36:32,351 INFO [main] org.apache.hadoop.hive.ql.sessionState - Updating thread name to 1c22c2df-c29b-4d8f-b0b3-63904207bbc2 and OK IME taken: 0.078 seconds 2024-08-30723:36:32,439 INFO [1c22c2df-c29b-4d8f-b0b3-63904207bbc2 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 1c22c2df-c29b-4d8f-b0b3-63904207bbc2 2024-08-30723:36:32,449 INFO [1c22c2df-c29b-4d8f-b0b3-63904207bbc2 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');
2024-08-30723:7:30,130 INTO [asin] org.apache.hadoop.hive.comf-invectorf-lugicate default value passed in for log id: 1222016-CDb-4006-18003-03002070bc2
2024-08-30723:7:30,130 INTO [asin] org.apache.hadoop.hive.comf-invectorf-lugicate associated asso
```

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:

```
ativewit@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 db-derby-10.14.2.0-bin pig
Cellar derby-log sbin
Frameworks et share
Nomebrew 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
nativewitcNativewits-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
      ativewit@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 \sim/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 .
     ativewit@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 .
      ativewit@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
     ativewit@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
     ativewit@Nativewits-MacRook-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
NET.: Class path contains multiple SIFA; bindings.

SIFA; Class path contains multiple SIFA; bindings.

SIFA; Class path contains multiple SIFA; bindings.

SIFA; See http://www.sif4;.org/codes.httple.bindings.

SIFA; See http://www.sif4;.org/codes.httple.bindings.

SIFA; See http://www.sif4;.org/codes.httple.bindings for an explanation.

SIFA; See http://www.sif4;.org/codes.httple.bindings for an explanation.

SIFA; See http://www.sif4;.org/codes.httple.bindings.

SIFA; Actual binding is of type (org.sif4;.impl.Reload6/loggerFactory)

SIFA; Class path contains multiple SIFA; bindings.

SIFA; See http://www.sif4;.org/codes.httple.bindings.

SIFA; Found binding in [jsrifile:/usr/local/opapath-hive-3.1.3-bin/lib/logi-sif4;-impl-2.17.1.jarl/org/sif4j/impl/StaticloggerBinder.class]

SIFA; Found binding in [jsrifile:/usr/local/opapath-hive-3.1.3-bin/lib/logi-sif4;-impl-2.17.1.jarl/org/sif4j/impl/StaticloggerBinder.class]

SIFA; See http://www.sif4j.org/codes.httplmultiple_bindings for an explanation.

SIFA; See http://www.sif4j.org/codes.htmlemultiple_bindings for an explanation.

SIFA; See http://www.sif4j.org/codes.htmlemultiple_bindings.

S
 Last login: Tue Sep 10 21:23:19 on ttys002 nativewit@Nativewits-MacBook-Air ~ %
     d /usr/local/Cellar/hadoop/3.4.0/libexec/sbin
    nativewit@Nativewits-MacBook-Air sbin % ./start-dfs.sh
   Starting namenodes on [localhost]
localhost: namenode is running as process 71574. Stop it first and ensure /tmp/hadoop-nativewit-namenode.pid file is empty before retry.
Starting datanodes is running as process 71574. Stop it first and ensure /tmp/hadoop-nativewit-datanode.pid file is empty before retry.
Starting secondary namenodes [Nativewits-MacBook-Air.local]
Nativewits-MacBook-Air.local: secondarynamenode is running as process 71812. Stop it first and ensure /tmp/hadoop-nativewit-secondarynamenode.pid file is empty before retry.
2024-09-10 2176-1080-886 MAMFW util.NativeCodeloader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
nativewiteNativewits-MacBook-Air sbin % ./start-yarn.sh
  Starting resourcemanager
resourcemanager is running as process 72011. Stop it first and ensure /tmp/badoop-nativewit-resourcemanager.pid file is empty before retry.
Starting nodemanagers
localhost: nodemanager is running as process 72114. Stop it first and ensure /tmp/hadoop-nativewit-nodemanager.pid file is empty before retry.

Rativesti@Nativewit=MacBook-Air sbin % hadoop fs =hkdir /input_dir

Rativesti@Nativewit=MacBook-Air sbin % hadoop fs =hkdir /input_dir

Rativesti@Nativewit=MacBook-Air sbin % hadoop fs =but ~/input_file.txt /input_dir

Rativexti@Nativewit=MacBook-Air sbin % hadoop fs =put ~/input_file.txt /input_dir
    2024-09-10 21:41:14,209 WARN util.NativeCodeloader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
ut: '/input_dir/input_file.tut': file exists
ativewit(Pativewit=MeaCod-Air sbin X hive--service schematool -dbType derby -initSchema
  instivewite Mactivewite MacBook-Air sbin % hive —mervice schematool —dbType derby -initSchema
SiF42: Class path contains multiple SiF43 bindings.
SiF43: Class path contains multiple SiF43 bindings.
SiF44: Found binding in [ser:file:/usr/local/dellar/hadoop/3.4.0/libexec/share/hadoop/common/lib/siF4j-reload4j-1.7.36.jar/org/siF4j/impl/StaticLoggerBinder.class]
SiF43: Found binding in [ser:file:/usr/local/dellar/hadoop/3.4.0/libexec/share/hadoop/common/lib/siF4j-reload4j-1.7.36.jar/org/siF4j/impl/StaticLoggerBinder.class]
SiF43: Found binding in [ser:file:/usr/local/dellar/hadoop/siF4j-reload4j-1.7.1.jar/org/siF4j/impl/StaticLoggerBinder.class]
SiF43: Service schematool not found
Macsilable Services: bealine cli help hiveburninclient hiveserver hiveserver2 hei jar lineage metastore metastool corfiledump refilecat schemaTool version
nativewiteNativewits-MacBook-Air sbin % hive
SiF43: Class sath contains multiple SiF44b bindings.
SiF43: Found binding in [ser:file:/usr/local/dellar/hadoop/3.4.0/libexec/share/hadoop/common/lib/siF4j-reload4j-1.7.36.jar/org/siF4j/impl/StaticLoggerBinder.class]
SiF43: Found binding in [ser:file:/usr/local/pache-hive-3.1.3-bin/lib/log4j-siF4j-impl-2.17.1.jar/org/siF4j/impl/StaticLoggerBinder.class]
SiF443: See http://www.siF4j.org/codes.html@multipl_bindings for an explanation.
SiF43: Found binding in [ser:file:/usr/local/dellar/hadoop/3.4.0/libexec/share/hadoop/common/lib/siF4j-reload4j-1.7.36.jar/org/siF4j/impl/StaticLoggerBinder.class]
SiF43: Found binding in [ser:file:/usr/local/dellar/hadoop/3.4.0/libexec/share/hadoop/common/lib/siF4j-reload4j-1.7.36.jar/org/siF4j/impl/StaticLoggerBinder.class]
SiF43: Found binding in [ser:file:/usr/local/spache-hive-3.1.3-bin/lib/log4j-siF4j-impl-2.17.1.jar/org/sif4j/impl/StaticLoggerBinder.class]
SiF43: Found binding in [ser:file:/usr/local/spache-hive-3.1.3-bin/lib/log4j-sif4j-impl-2.17.1.jar/org/sif4j/impl/StaticLoggerBinder.class]
SiF43: Found binding in [ser:file:/usr/local/spache-hive-3.1.3-bin/lib/log4j-sif4j-impl-2.17.1.jar/org/sif4j/impl/
```

```
2024-09-20 21:05:03,597 INFO maprod-File/InduFormat: Tetal imput files to process 1 2
2024-09-20 21:05:03,591 INFO exec.TableScanDeprator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_TS_0:3,
2024-09-20 21:05:03,591 INFO exec.ListSinKOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_EL_1:3,
2024-09-20 21:05:03,591 INFO exec.ListSinKOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_EL_1:3,
2024-09-20 21:05:03,591 INFO exec.ListSinKOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_LIST_SINK_3:3,
2024-09-20 21:05:03,591 INFO exec.ListSinKOperator: RECORDS_OUT_OPERATOR_LIST_SINK_3:3,
2024-09-20 21:05:03,593 INFO exec.ListSinKOperator: RECORDS_OUT_OPERATOR_LIST_SINK_3:3,
2024-09-20 21:05:03,593 INFO exec.ListSinKOperator: Time taken: 0.354 seconds, Fetched: 3 row(s)
2024-09-20 21:05:03,593 INFO exec.ListSinKOperator: Time taken: 0.354 seconds, Fetched: 3 row(s)
2024-09-20 21:05:03,593 INFO exec.ListSinKOperator: Time taken: 0.354 seconds, Fetched: 3 row(s)
2024-09-20 21:05:03,593 INFO exec.ListSinKOperator: Time taken: 0.354 seconds, Fetched: 3 row(s)
2024-09-20 21:05:03,593 INFO exec.ListSinKOperator: Time taken: 0.354 seconds, Fetched: 3 row(s)
2024-09-20 21:05:03,593 INFO exec.ListSinKOperator: Time taken: 0.354 seconds, Fetched: 3 row(s)
2024-09-20 21:05:03,593 INFO exec.ListSinKOperator: Time taken: 0.354 seconds, Fetched: 3 row(s)
2024-09-20 21:05:03,593 INFO exec.ListSinKOperator: Time taken: 0.354 seconds, Fetched: 3 row(s)
2024-09-20 21:05:03,593 INFO exec.ListSinKOperator: Time taken: 0.354 seconds, Fetched: 3 row(s)
2024-09-20 21:05:03,593 INFO exec.ListSinKOperator: Time taken: 0.354 seconds, Fetched: 3 row(s)
2024-09-20 21:05:03,593 INFO exec.ListSinKOperator: Time taken: 0.354 seconds, Fetched: 3 row(s)
2024-09-20 21:05:03,593 INFO exec.ListSinKOperator: Time taken: 0.354 seconds, Fetched: 3 row(s)
2024-09-20 21:05:03,593 INFO exec.ListSinKOperator: Time taken: 0.354 seconds, Fetched: 2 row(s)
2024-09-20 21:05:03,593 INFO exec.ListSinKOperator: Time taken: 0.354 seconds,
```

RESULT:

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