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 the Apache Derby .

2. Download Apache Hive binaries and set the environment variables.

2a. Copy Derby libraries

we should go to the Derby libraries directory (db-derby-10.14.2.0\lib) and copy all *.jar files. Then, we should paste them within the Hive libraries directory.

2b. Configuring hive-site.xml and Hive's Bin folder

Refer following link to download the file. Also download the guava file. Put hive-site.xml file to hive's conf location and replace hive's current guava file with this one in lib location. Also download the bin folder from link and replace the existing hive's bin folder.

https://drive.google.com/drive/folders/1r1OTDNcw8aF74V21k8GlD2IUGMJ_v11o?usp=drive_link

3. Starting Hadoop Services.

```
C:\Windows\System32>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons

C:\Windows\System32>jps
10068 NodeManager
30616 Jps
21308 DataNode
5612 ResourceManager
5836 NameNode
```

4. Start Derby Network Server: Run the following command in separate window to open Derby

```
C:\Windows\System32>StartNetworkServer -h 0.0.0.0
Mon Sep 09 21:06:55 IST 2024 : Security manager installed using the Basic server security policy.
Mon Sep 09 21:06:55 IST 2024 : Apache Derby Network Server - 10.12.1.1 - (1704137) started and ready to accept connections on port 1527
```

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

hive --service schematool -dbType derby -initSchema

6. Open Hive shell by typing:

hive

Creating tables and accessing its data:

1. Create database name joshin and use that database to create table. Create a table named students with columns for student name, department, and age.

CREATE DATABASE kanna;

USE kanna;

CREATE TABLE students (student_name STRING,department STRING,age INT);

2. Insert data into the students table.

INSERT INTO TABLE students VALUES ('Kanna', 'Computer Science', 150);

3. View the structure of the students table using the DESCRIBE command:

DESCRIBE students;

4. List all tables in the current database:

SHOW TABLES;

5. View the contents of the students table using a SELECT query:

SELECT * FROM students;

OUTPUT:

1. Creating Database:

```
hive> CREATE DATABASE kanna;
2024-09-17T23:08:31,442 INFO [38dd860c-35f8-4303-b55e-aba0121c8d1e main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 38dd860c-35f8-
4303-b55e-aba0121c8d1e
2024-09-17T23:08:31,812 WARN [38dd860c-35f8-4303-b55e-aba0121c8d1e main] org.apache.hadoop.hive.ql.session.SessionState - METASTORE_FILTER_HOOK will be ignored, since hive.
security.authorization.manager is set to instance of HiveAuthorizerFactory.
OK
Time taken: 0.378 seconds
2024-09-17T23:08:31,859 INFO [38dd860c-35f8-4303-b55e-aba0121c8d1e main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 38dd860c-35f8-
4303-b55e-aba0121c8d1e
2024-09-17T23:08:31,859 INFO [38dd860c-35f8-4303-b55e-aba0121c8d1e main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to  main
```

```
hive> SHOW TABLES;
2024-09-17T23:21:31,504 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 38dd860c-35f8-4303-b55e-aba0121c8d1e
2024-09-17T23:21:31,505 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 38dd860c-35f8-4303-b55e-aba0121c8d1e main
OK
students
Time taken: 0.023 seconds, Fetched: 1 row(s)
2024-09-17T23:21:31,544 INFO [38dd860c-35f8-4303-b55e-aba0121c8d1e main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 38dd860c-35f8-
4303-b55e-aba0121c8d1e
2024-09-17T23:21:31,544 INFO [38dd860c-35f8-4303-b55e-aba0121c8d1e main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to  main
hive>
```

2. Use created Database to create table:

```
hive> USE kanna;
2024-09-17T23:08:43,682 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 38dd860c-35f8-4303-b55e-aba0121c8d1e
2024-09-17T23:08:43,682 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 38dd860c-35f8-4303-b55e-aba0121c8d1e main
OK
Time taken: 0.025 seconds
2024-09-17T23:08:43,709 INFO [38dd860c-35f8-4303-b55e-aba0121c8d1e main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 38dd860c-35f8-
4303-b55e-aba0121c8d1e
2024-09-17T23:08:43,709 INFO [38dd860c-35f8-4303-b55e-aba0121c8d1e main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to   main
```

3. Creating table to store data:

```
hive> CREATE TABLE students (student_name STRING,department STRING,age INT);
2024-09-17T23:09:30,257 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 38dd860c-35f8-4303-b55e-aba0121c8d1e
2024-09-17T23:09:30,257 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 38dd860c-35f8-4303-b55e-aba0121c8d1e main
OK
Time taken: 0.414 seconds
2024-09-17T23:09:30,673 INFO [38dd860c-35f8-4303-b55e-aba0121c8d1e main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 38dd860c-35f8-
4303-b55e-aba0121c8d1e
2024-09-17T23:09:30,674 INFO [38dd860c-35f8-4303-b55e-aba0121c8d1e main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to   main
```

4. Inserting data into the table:

```
hive> INSERT INTO TABLE students VALUES('Kanna','Computer Science',150);
2024-09-17T23:12:06,104 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 38dd860c-35f8-4303-b55e-aba0121c8d1e
2024-09-17T23:12:06,104 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 38dd860c-35f8-4303-b55e-aba0121c8d1e main
2024-09-17T23:12:07,990 INFO [38dd860c-35f8-4303-b55e-aba0121c8d1e main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localhost:
9000/user/hive/warehouse/kanna.db/students/.hive-staging_hive_2024-09-17_23-12-06_117_4189796483909259758-1
2024-09-17T23:12:08,075 INFO [38dd860c-35f8-4303-b55e-aba0121c8d1e main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localhost:
9000/tmp/hive/Manoj/38dd860c-35f8-4303-b55e-aba0121c8d1e/hive_2024-09-17_23-12-08_018_3071303827969161845-1/-mr-10000/.hive-staging_hive_2024-09-17_23-12-08_018_30713038279
69161845-1
Query ID = Manoj_20240917231206_7f273aa8-b25e-45b9-a6e3-8509d042ccaa
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>
In order to set a constant number of reducers:
```
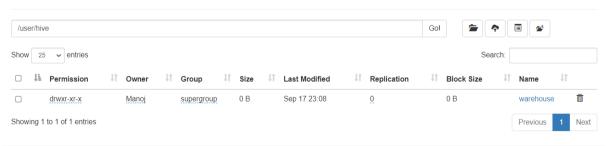
5. Viewing the created table:

```
2024-09-17T23:12:12,02? INFO [38dd860c-35f8-4303-b55e-aba0121c8d1e main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to   main
hive> DESCRIBE students;
2024-09-17T23:14:20,873 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 38dd860c-35f8-4303-b55e-aba0121c8d1e
2024-09-17T23:14:20,873 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 38dd860c-35f8-4303-b55e-aba0121c8d1e main
OK
student_name            string
department              string
age                     int
Time taken: 0.043 seconds, Fetched: 3 row(s)
2024-09-17T23:14:20,949 INFO [38dd860c-35f8-4303-b55e-aba0121c8d1e main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 38dd860c-35f8-
4303-b55e-aba0121c8d1e
2024-09-17T23:14:20,949 INFO [38dd860c-35f8-4303-b55e-aba0121c8d1e main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to   main
hive>
```
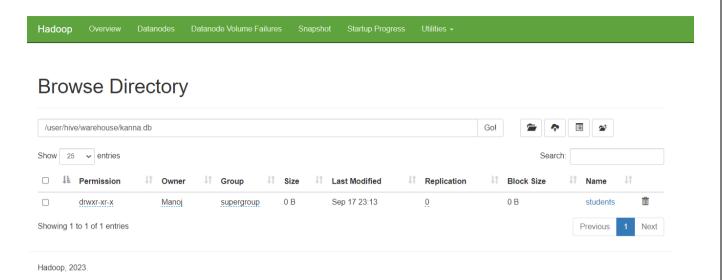
```
hive> SELECT * FROM students;
2024-09-17T23:23:06,984 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 38dd860c-35f8-4303-b55e-aba0121c8d1e
2024-09-17T23:23:06,984 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 38dd860c-35f8-4303-b55e-aba0121c8d1e main
2024-09-17T23:23:07,070 INFO [38dd860c-35f8-4303-b55e-aba0121c8d1e main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localhost:
9000/tmp/hive/Manoj/38dd860c-35f8-4303-b55e-aba0121c8d1e/hive_2024-09-17_23-23-06_998_475503907806911757-1/-mr-10001/.hive-staging_hive_2024-09-17_23-23-06_998_475503907806
911757-1
OK
Kanna    Computer Science        150
Time taken: 0.103 seconds, Fetched: 1 row(s)
2024-09-17T23:23:07,109 INFO [38dd860c-35f8-4303-b55e-aba0121c8d1e main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 38dd860c-35f8-
4303-b55e-aba0121c8d1e
2024-09-17T23:23:07,110 INFO [38dd860c-35f8-4303-b55e-aba0121c8d1e main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to   main
hive>
```

Hadoop	Overview	Datanodes	Datanode Volume Failures	Snapshot	Startup Progress	Utilities ▾

Browse Directory

/user/hive   Go!

Show 25 entries   Search:

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
☐	drwxr-xr-x	Manoj	supergroup	0 B	Sep 17 23:08	0	0 B	warehouse	🗑

Showing 1 to 1 of 1 entries   Previous | 1 | Next

Hadoop, 2023.

Hadoop | Overview | Datanodes | Datanode Volume Failures | Snapshot | Startup Progress | Utilities ▾

Browse Directory

/user/hive/warehouse/kanna.db — Go!

Show 25 entries — Search:

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
☐	drwxr-xr-x	Manoj	supergroup	0 B	Sep 17 23:13	0	0 B	students	🗑

Showing 1 to 1 of 1 entries — Previous 1 Next

Hadoop, 2023.

File information - 000000_0 ×

Download | Head the file (first 32K) | Tail the file (last 32K)

Block information — Block 0 ▾

Block ID: 1073741919
Block Pool ID: BP-373262882-192.168.56.1-1724824016975
Generation Stamp: 1095
Size: 27
Availability:
- LEGION

File contents

Kanna Computer Science 150

Close
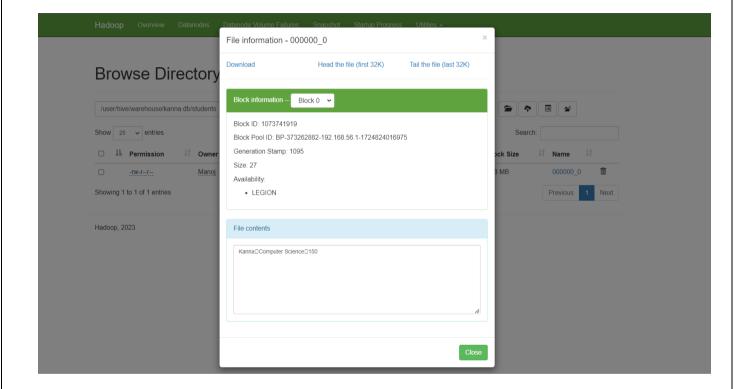

RESULT:

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