

EXP: 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.

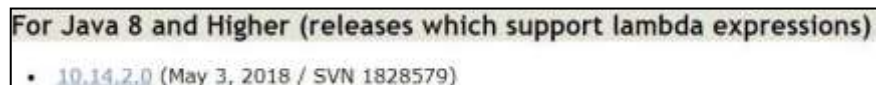
PROCEDURE:

Hive Download and installation:

1. Hive Installation setup:

- Download and install Apache Derby version 10.14.2.0:

https://db.apache.org/derby/derby_downloads.html#For+Java+8+and+Higher



-Download and install Apache Hive version 3.1.3:

<https://downloads.apache.org/hive/hive-3.1.3/>



2. Add environment variables:

Environment variables > System variables > Add the below paths

System variables	
Variable	Value
DERBY_HOME	C:\db-derby-10.14.2.0-bin
DriverData	C:\Windows\System32\Drivers\DriverData
HADOOP_HOME	C:\Hadoop
HIVE_BIN	%HIVE_HOME%\bin
HIVE_HOME	C:\apache-hive-3.1.3-bin
HIVE_LIB	%HIVE_HOME%\lib

C:\Java\jdk-1.8\bin
C:\Hadoop\bin
C:\Hadoop\sbin
C:\Python39\
%PIG_HOME%\bin
%DERBY_HOME%\bin
%HIVE_BIN%

-> (Inside Path)

3. Copy Derby libraries

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

4. 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://1drv.ms/f/s!ArSg3Xpur4Grmw0SDqW0g44T7HYU?e=wDsoBn>

5. Starting Hadoop Services

Open PowerShell as administrator and go to Hadoop sbin directory and start hadoop services using the following commands:

```
start-dfs.cmd
```

```
start-yarn.cmd
```

6. Derby Network Server:

Open another PowerShell window and run the following command to open Derby:

```
StartNetworkServer -h 0.0.0.0
```

Go to first PowerShell window and check whether NetworkServerControl is running.

7. Starting Apache Hive:

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

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

8. Open Hive shell by typing:

```
hive
```

Create a Database:

Start by creating a database. Open the Hive CLI and follow the steps below:

1. Use the **CREATE DATABASE** statement to create a new database:

```
CREATE DATABASE financials;
```

2. Verify the database is present:

```
SHOW DATABASES;
```

3. Switch to the new database:

Switch to the newly created database:

```
USE financials;
```

Create a Table in Hive:

Create a simple table in your database:

```
CREATE TABLE students_table (id INT, name STRING);
```

Insert values into the table:

You can insert sample data into the table:

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

Query your data:

Use SQL-like queries to retrieve data from your table:

```
CREATE VIEW myview AS SELECT name, id FROM finance_table;
```

List Hive Tables and Data:

To show all tables in a selected database, use the following statement:

```
SHOW TABLES;
```

To show table column names and data types, run:

```
DESCRIBE finance_table;
```

To display table data, use a **SELECT** statement. For example, to select everything in a table, run:

```
SELECT * FROM myview;
```

Alter a Table:

You can alter the table structure by adding a new column:

```
ALTER TABLE finance_table ADD COLUMNS (age INT);
```

To exit the Hive CLI, simply type:

quit;

OUTPUT SCREENSHOTS:

[illegible]


```
Administrator: Windows PowerShell
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\WINDOWS\system32> Start-NetWorkerServer -v 8.0.0.0
Wed Sep 11 17:46:51 IST 2024 : Security manager installed using the Basic server security policy.
Wed Sep 11 17:47:01 IST 2024 : Apache Derby Network Server - 10.10.3.0 - (1328370) started and ready to accept connections on port 1527

Administrator: Windows PowerShell
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\WINDOWS\system32> ipconfig
15576 DataNode
15948 HiveNode
21488 NetworkServerControl
28258 Zps
28888 ResourceManager
24188 HdfsManager
PS C:\WINDOWS\system32>
```

```

C:\> Select Administrator Command Prompt
C:\>
Time taken: 180.34 seconds
hive> (CREATE VIEW myview AS SELECT name, id FROM finance_table;
OK
Time taken: 0.104 seconds
hive> show tables;
OK
finance_table
myview
Time taken: 0.025 seconds, Fetched: 1 row(s)
hive> show databases;
OK
default
financials
Time taken: 0.017 seconds, Fetched: 1 row(s)
hive> SELECT*FROM myview;
OK
Alice 1
Bob 2
Charlie 3
Time taken: 0.000 seconds, Fetched: 1 row(s)
hive> DESCRIBE finance_table;
OK
id          int
name        string
Time taken: 0.027 seconds, Fetched: 1 row(s)
hive> ALTER TABLE finance_table ADD COLUMN (age INT);
OK
Time taken: 0.073 seconds
hive> DESCRIBE finance_table;
OK
id          int
name        string
age         int
Time taken: 0.026 seconds, Fetched: 1 row(s)
hive> quit;
C:\>

C:\> Select Administrator Command Prompt
hive> INSERT INTO finance_table VALUES (1, 'Alice'), (1, 'Bob'), (1, 'Charlie');
2024-09-11 17:53:02,154 WARN gl.Driver: Hive-mr-job is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X release.
Query ID = hqecrowd851b_20240911175300_jobRes2f-b0b-44a6-b4d0-4f152600ee24
Total jobs = 1
Launching job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.mapred.reduce.tasks.bytes.per.reducer.counters
In order to limit the maximum number of reducers:
  set hive.mapred.reduce.tasks.maxcounters
In order to set a constant number of reducers:
  set mapreduce.job.reducecounters
2024-09-11 17:53:03,164 WARN mapreduce.JobDriver: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
Starting Job = job_1726056038621_0001, Tracking URL = http://localhost:8088/proxy/application_1726056038621_0001/
Kill Command = D:\hadoop\bin\hadoop job -kill job_1726056038621_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2024-09-11 17:53:03,292 WARN mapreduce.Counters: Group org.apache.hadoop.mapred.TaskCounter is deprecated. Use org.apache.hadoop.mapreduce.TaskCounter instead
2024-09-11 17:53:03,290 Stage-1 map = 0%, reduce = 0%
2024-09-11 17:54:04,707 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.478 sec
2024-09-11 17:54:25,324 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 1.984 sec
MapReduce Total cumulative CPU time: 1 seconds 984 msec
Ended Job = job_1726056038621_0001
Stage-4 is selected 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_2024-09-11_17-53-00_832_471788647301748129-1/-ext-10000
Loading data to table financials.finance_table:
2024-09-11 17:54:00,641 WARN metastore.ObjectStore: datacatalog.autoStartMechanismMode is set to unsupported value null. Setting it to value: ignored
2024-09-11 17:54:00,737 WARN metastore.ObjectStore: datacatalog.autoStartMechanismMode is set to unsupported value null. Setting it to value: ignored
MapReduce jobs launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU 1.984 sec HDFS Read: 15722 HDFS Write: 293 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 984 msec
OK
Time taken: 180.34 seconds
hive> CREATE VIEW myview AS SELECT name, id FROM finance_table;
OK
Time taken: 0.104 seconds
hive> show tables;
OK
finance_table
myview
Time taken: 0.025 seconds, Fetched: 1 row(s)
hive> show databases;
OK
default
financials
Time taken: 0.017 seconds, Fetched: 1 row(s)

```

RESULT:

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