

## Ex 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 using hive.

### Procedure:

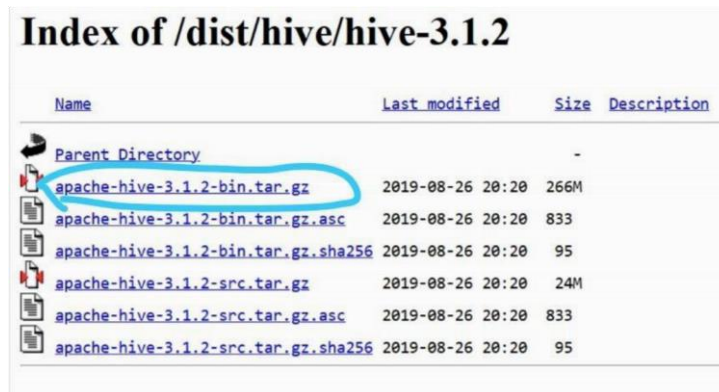
1. Check if Hadoop is installed.
2. Download and configure **Apache Derby**.
  - a. Download Derby 10.14.2.0 from [https://db.apache.org/derby/derby\\_downloads.html](https://db.apache.org/derby/derby_downloads.html)

**For Java 8 and Higher (releases which support lambda expressions)**

- [10.14.2.0](#) (May 3, 2018 / SVN 1828579)
- [10.13.1.1](#) (October 25, 2016 / SVN 1766613)

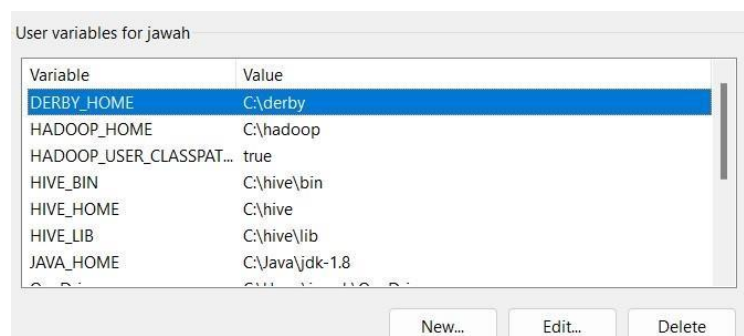
b. Extract the files and save to 'C://derby'

3. Download and configure **Apache Hive**
  - a. Download Hive from <https://archive.apache.org/dist/hive/hive-3.1.2/> and extract to C://hive



Name	Last modified	Size	Description
Parent Directory	-	-	-
<a href="#">apache-hive-3.1.2-bin.tar.gz</a>	2019-08-26 20:20	266M	
<a href="#">apache-hive-3.1.2-bin.tar.gz.asc</a>	2019-08-26 20:20	833	
<a href="#">apache-hive-3.1.2-bin.tar.gz.sha256</a>	2019-08-26 20:20	95	
<a href="#">apache-hive-3.1.2-src.tar.gz</a>	2019-08-26 20:20	24M	
<a href="#">apache-hive-3.1.2-src.tar.gz.asc</a>	2019-08-26 20:20	833	
<a href="#">apache-hive-3.1.2-src.tar.gz.sha256</a>	2019-08-26 20:20	95	

- b. Navigate to C://derby/lib and copy all the \*jar files and paste them in C://hive/lib
4. Check for proper version of guava file in Hadoop, if not make necessary changes. 5. Set up the Environment Variables



6. Download the windows version of hive bin folder and replace it.
7. Run the **Hadoop daemons** using the command **start-all.cmd**
8. Run **Derby Server** using the command

**StartNetworkServer -h 0.0.0.0**

```
C:\>StartNetworkServer -h 0.0.0.0
Sun Sep 08 20:04:25 IST 2024 : Security manager installed using the Basic server security policy.
Sun Sep 08 20:04:26 IST 2024 : Apache Derby Network Server - 10.14.2.0 - (1828579) started and ready to accept connections on port 1527
```

9. To **Initialize hive Schema**, Navigate to C://hive/bin and type the command **hive --service schematool -dbType derby -initSchema**

**hive --service schematool -dbType derby --upgradeSchema**

```
C:\hive\bin>hive --service schematool -dbType derby --upgradeSchema
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/C:/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/C:/hadoop/share/hadoop/common/lib/slf4j-reload4j-1.7.36.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.apache.logging.slf4j.Log4jLoggerFactory]
2024-09-08T20:08:59,130 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Found configuration file file:/C:/hive/conf/hive-site.xml
2024-09-08T20:08:59,549 WARN [main] org.apache.hadoop.hive.conf.HiveConf - HiveConf of name hive.server2.enable.doAS does not exist
No schema upgrade required from version 3.1.0
schemaTool completed
```

10. To run the hive use command **hive**

```
hive> _
```

```
C:\hive\bin>hive --service schematool -dbType derby -initSchema
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/C:/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/C:/hadoop/share/hadoop/common/lib/slf4j-reload4j-1.7.36.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.apache.logging.slf4j.Log4jLoggerFactory]
2024-09-08T20:07:16,514 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Found configuration file file:/C:/hive/conf/hive-site.xml
2024-09-08T20:07:17,070 WARN [main] org.apache.hadoop.hive.conf.HiveConf - HiveConf of name hive.server2.enable.doAS does not exist
2024-09-08T20:07:17,215 INFO [main] org.apache.hadoop.hive.metastore.tools.HiveSchemaHelper - Metastore connection URL: jdbc:derby://localhost:1527/metastore_db;create=true
Metastore connection URL: jdbc:derby://localhost:1527/metastore_db;create=true
2024-09-08T20:07:17,217 INFO [main] org.apache.hadoop.hive.metastore.tools.HiveSchemaHelper - Metastore Connection Driver: org.apache.derby.jdbc.ClientDriver
Metastore Connection Driver: org.apache.derby.jdbc.ClientDriver
2024-09-08T20:07:17,219 INFO [main] org.apache.hadoop.hive.metastore.tools.HiveSchemaHelper - Metastore connection User: APP
Metastore connection User: APP
Starting metastore schema initialization to 3.1.0
Initialization script hive-schema-3.1.0.derby.sql
```

if Partially initialized use

**Hive is successfully installed and now we can run commands in hive**

## Executing commands

1. Creating database **create database company;**

```
hive> create database company;
2024-09-08T20:09:57,773 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: dcf4bc2d-647f-4cfc-8029-4f8d4367e434
2024-09-08T20:09:57,774 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main
2024-09-08T20:09:57,980 WARN [dcf4bc2d-647f-4cfc-8029-4f8d4367e434 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: 0.48 seconds
2024-09-08T20:09:58,259 INFO [dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: dcf4bc2d-647f-4cfc-8029-4f8d4367e434
2024-09-08T20:09:58,260 INFO [dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to main
```

## 2. Check if database is created Show databases;

```
hive> Show databases;
2024-09-08T20:12:03,100 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: dcf4bc2d-647f-4cfc-8029-4f8d4367e434
2024-09-08T20:12:03,100 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main
OK
2024-09-08T20:12:03,527 INFO [dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main] org.apache.hadoop.conf.Configuration.deprecation - mapred.input.dir is deprecated. Instead, use mapreduce.input.fileinputformat.inputdir
company
default
Time taken: 0.403 seconds, Fetched: 2 row(s)
2024-09-08T20:12:03,775 INFO [dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: dcf4bc2d-647f-4cfc-8029-4f8d4367e434
2024-09-08T20:12:03,776 INFO [dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to main
```

## 3. Create table employee create table employee(name varchar(10), age int, empId int, designation varchar(10));

```
hive> create table employee(name varchar(10), age int, empId int, designation varchar(20));
2024-09-08T20:58:29,228 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: dcf4bc2d-647f-4cfc-8029-4f8d4367e434
2024-09-08T20:58:29,229 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main
OK
Time taken: 1.751 seconds
2024-09-08T20:58:30,982 INFO [dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: dcf4bc2d-647f-4cfc-8029-4f8d4367e434
2024-09-08T20:58:30,983 INFO [dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to main
hive>
```

## 4. Insert values in the table employee insert into employee values("Amar",45,1,"DataAnalyst");

```
hive> insert into employee values("Amar",45,1,"DataAnalyst");
2024-09-08T21:01:26,553 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: dcf4bc2d-647f-4cfc-8029-4f8d4367e434
2024-09-08T21:01:26,554 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main
2024-09-08T21:01:31,743 INFO [dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localhost:9000/user/hive/warehouse/employee/hive-staging_hive_2024-09-08_21-01-26_580_3515185677712923735-1
2024-09-08T21:01:32,063 INFO [dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localhost:9000/tmp/hive/jawah/dcf4bc2d-647f-4cfc-8029-4f8d4367e434/hive_2024-09-08_21-01-31_850_604811567303269016-1/-mr-10000/hive-staging_hive_2024-09-08_21-01-31_850_604811567303269016-1
Query ID = jawah_20240908210126_1d8f3456-70c9-4975-9e75-f02675e6da0a
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:
  set mapreduce.job.reducers=<number>
2024-09-08T21:01:32,941 INFO [dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main] org.apache.hadoop.conf.Configuration.deprecation - mapred.submit.replication is deprecated. Instead, use mapreduce.client.submit.file.replication
2024-09-08T21:01:34,936 INFO [dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main] org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
2024-09-08T21:01:35,397 INFO [dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main] org.apache.hadoop.conf.Configuration - resource-types.xml not found
Starting Job = job_1725805694850_0001, Tracking URL = http://jawahar:8088/proxy/application_1725805694850_0001/
Kill Command = C:\hadoop\bin\mapred job -kill job_1725805694850_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2024-09-08 21:01:52,854 Stage-1 map = 0%, reduce = 0%
2024-09-08 21:02:02,768 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.499 sec
2024-09-08 21:02:12,283 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.718 sec
MapReduce Total cumulative CPU time: 4 seconds 718 msec
Ended Job = job_1725805694850_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/employee/hive-staging_hive_2024-09-08_21-01-26_580_3515185677712923735-1/-ext-10000
Loading data to table default.employee
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.718 sec HDFS Read: 18323 HDFS Write: 326 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 718 msec
OK
Time taken: 48.081 seconds
2024-09-08T21:02:14,644 INFO [dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: dcf4bc2d-647f-4cfc-8029-4f8d4367e434
2024-09-08T21:02:14,645 INFO [dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to main
```

**insert into employee values("Divya",25,3,"Marketing");**



5. Use the select command to access the data **select \* from employee;**

```
hive> select * from employee;
2024-09-08T21:11:39,480 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: dcf4bc2d-647f-4cfc-8029-4f8d4367e434
2024-09-08T21:11:39,482 INFO [main] org.apache.hadoop.hive.q1.session.SessionState - Updating thread name to dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main
2024-09-08T21:11:39,733 INFO [dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localhost:
9000/tmp/hive/jawah/dcf4bc2d-647f-4cfc-8029-4f8d4367e434/hive_2024-09-08_21-11-39_510_5468558375529413220-1/-mr-10001/.hive-staging_hive_2024-09-08_21-11-39_510_54685583755
29413220-1
OK
Amar 45 1 DataAnalyst
Divya 25 3 Marketing
John 30 5 QA
Karthik 27 7 Developer
Time taken: 0.29 seconds, Fetched: 4 row(s)
2024-09-08T21:11:39,924 INFO [dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: dcf4bc2d-647f-
4cfc-8029-4f8d4367e434
2024-09-08T21:11:39,926 INFO [dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main] org.apache.hadoop.hive.q1.session.SessionState - Resetting thread name to main
hive>
```

**select \* from employee where age<30;**

```
hive> select * from employee where age<30;
2024-09-08T21:13:10,257 INFO [main] org.apache.
2024-09-08T21:13:10,260 INFO [main] org.apache.
2024-09-08T21:13:10,533 INFO [dcf4bc2d-647f-4cfc-8029-4f8d4367e434 main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localhost:
9000/tmp/hive/jawah/dcf4bc2d-647f-4cfc-8029-4f8d4367e434/hive_2024-09-08_21-13-10_533_5468558375529413220-1/-mr-10001/.hive-staging_hive_2024-09-08_21-13-10_533_54685583755
692180-1
OK
Divya 25 3 Marketing
Karthik 27 7 Developer
Time taken: 0.326 seconds, Fetched: 2 row(s)
```

**Stop the Derby server:**

**java -jar %DERBY\_HOME%\lib\derbyrun.jar server shutdown**

**Result:**

Thus the creation of tables in Hive and writing queries to access the data in the table is executed successfully.