

**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:**

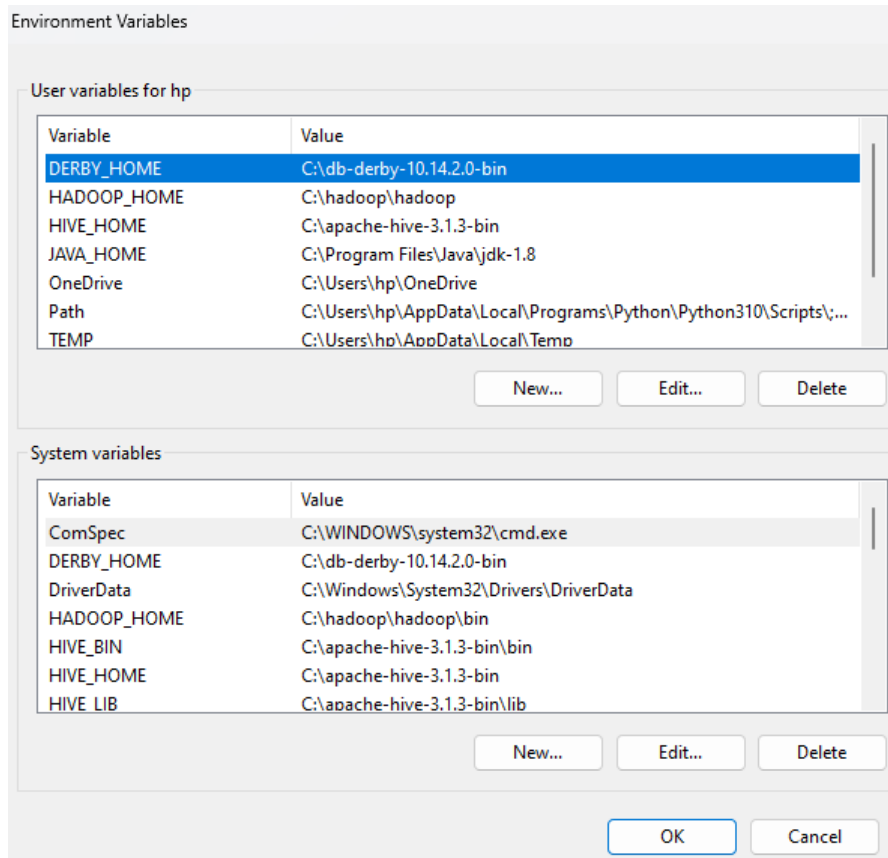
Hive Download and installation:

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

2. 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 -> (Inside Path)



C:\db-derby-10.14.2.0-bin\bin

C:\apache-hive-3.1.3-bin\bin

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=wDsoB>

5. Start the hadoop services.

#### start-all.cmd

```
Microsoft Windows [Version 10.0.22631.4037]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\System32>java -version
java version "1.8.0_421"
Java(TM) SE Runtime Environment (build 1.8.0_421-b09)
Java HotSpot(TM) 64-Bit Server VM (build 25.421-b09, mixed mode)

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
15904 DataNode
3296 Jps
4976 NodeManager
8532 ResourceManager
15560 NameNode
12540 NetworkServerControl

C:\Windows\System32>hive --service schematool -dbType derby -initSchema
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/C:/hadoop/hadoop/share/hadoop/common/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/C:/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]
2024-09-17 14:06:03,908 INFO conf.HiveConf: Found configuration file file:/C:/apache-hive-3.1.3-bin/conf/hive-site.xml
2024-09-17 14:06:04,720 INFO tools.HiveSchemaHelper: Metastore connection URL: jdbc:derby://localhost:1527/metastore_db;create=true
Metastore connection URL: jdbc:derby://localhost:1527/metastore_db;create=true
2024-09-17 14:06:04,721 INFO tools.HiveSchemaHelper: Metastore Connection Driver : org.apache.derby.jdbc.ClientDriver
Metastore Connection Driver : org.apache.derby.jdbc.ClientDriver
2024-09-17 14:06:04,722 INFO 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
```

6. Start the derby host in 0.0.0.0

#### StartNetworkServer -h 0.0.0.0

```
C:\Windows\System32>StartNetworkServer -h 0.0.0.0
Tue Sep 17 14:04:03 IST 2024 : Security manager installed using the Basic server security policy.
Tue Sep 17 14:04:03 IST 2024 : Apache Derby Network Server - 10.14.2.0 - (1828579) started and ready to accept connections on port 1527
```

```
Initialization script completed
schemaTool completed
```

```
C:\Windows\System32>
```

7. Start the hive services:

#### hive --service schematool -dbType derby -initSchema

```
C:\Windows\System32>hive --service schematool -dbType derby -initSchema
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/C:/hadoop/hadoop/share/hadoop/common/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/C:/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]
2024-09-17 14:06:03,908 INFO conf.HiveConf: Found configuration file file:/C:/apache-hive-3.1.3-bin/conf/hive-site.xml
2024-09-17 14:06:04,720 INFO tools.HiveSchemaHelper: Metastore connection URL: jdbc:derby://localhost:1527/metastore_db;create=true
Metastore connection URL: jdbc:derby://localhost:1527/metastore_db;create=true
2024-09-17 14:06:04,721 INFO tools.HiveSchemaHelper: Metastore Connection Driver : org.apache.derby.jdbc.ClientDriver
Metastore Connection Driver : org.apache.derby.jdbc.ClientDriver
2024-09-17 14:06:04,722 INFO 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
```

8. To execute the SQL query open hive shell

hive

```
C:\Windows\System32>cd C:\apache-hive-3.1.3-bin\bin
C:\apache-hive-3.1.3-bin\bin>hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/C:/hadoop/hadoop/share/hadoop/common/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/C:/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]
2024-09-17 14:08:05,973 INFO conf.HiveConf: Found configuration file file:/C:/apache-hive-3.1.3-bin/conf/hive-site.xml
Hive Session ID = 3c26b629-b286-465a-a0fc-03f0264456e5
2024-09-17 14:08:09,202 INFO SessionState: Hive Session ID = 3c26b629-b286-465a-a0fc-03f0264456e5
```

Create a database:

**CREATE DATABASE mydb;**

```
hive> CREATE database mydb;
2024-09-17 14:08:41,903 INFO conf.HiveConf: Using the default value passed in for log id: 3c26b629-b286-465a-a0fc-03f0264456e5
2024-09-17 14:08:42,159 INFO ql.Driver: Compiling command(queryId=hp_20240917140841_2082b18e-f98a-404f-823c-3df184184c3d): CREATE database mydb
2024-09-17 14:08:43,034 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-17 14:08:43,067 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2024-09-17 14:08:43,075 INFO ql.Driver: Returning Hive schema: Schema(fieldSchemas:null, properties:null)
2024-09-17 14:08:43,091 INFO ql.Driver: Completed compiling command(queryId=hp_20240917140841_2082b18e-f98a-404f-823c-3df184184c3d); Time taken: 1.018 seconds
2024-09-17 14:08:43,092 INFO rexec.ReExecDriver: Execution #1 of query
```

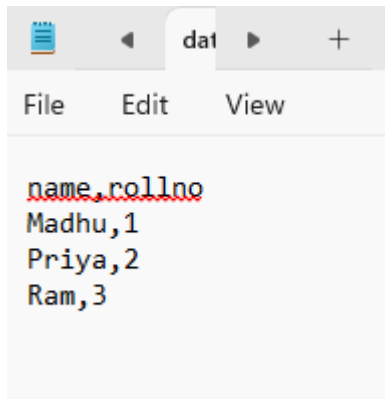
**SHOW DATABASES;**

```
hive> SHOW DATABASES;
2024-09-17 14:09:10,482 INFO conf.HiveConf: Using the default value passed in for log id: 3c26b629-b286-465a-a0fc-03f0264456e5
2024-09-17 14:09:10,483 INFO session.SessionState: Updating thread name to 3c26b629-b286-465a-a0fc-03f0264456e5 main
2024-09-17 14:09:10,486 INFO ql.Driver: Compiling command(queryId=hp_20240917140910_841e2de5-3150-4c88-9c59-25c77214142a): SHOW DATABASES
2024-09-17 14:09:10,510 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-17 14:09:10,528 INFO ql.Driver: Semantic Analysis Completed (retrial = false)
2024-09-17 14:09:10,664 INFO ql.Driver: Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name=database_name, type:string, comment:from deserializer)], properties:null)
2024-09-17 14:09:10,866 INFO exec.ListSinkOperator: Initializing operator LIST_SINK[0]
2024-09-17 14:09:10,882 INFO ql.Driver: Completed compiling command(queryId=hp_20240917140910_841e2de5-3150-4c88-9c59-25c77214142a); Time taken: 0.396 seconds
2024-09-17 14:09:10,882 INFO rexec.ReExecDriver: Execution #1 of query
2024-09-17 14:09:10,882 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-17 14:09:10,887 INFO ql.Driver: Executing command(queryId=hp_20240917140910_841e2de5-3150-4c88-9c59-25c77214142a): SHOW DATABASES
2024-09-17 14:09:10,888 INFO ql.Driver: Starting task [Stage-0:DDL] in serial mode
2024-09-17 14:09:10,889 INFO metastore.HiveMetaStore: @: get_databases: @hive#
2024-09-17 14:09:10,890 INFO HiveMetaStore.audit: ugi=hp ip=unknown-ip-addr cmd=get_databases: @hive#
2024-09-17 14:09:10,907 INFO exec.DDLTask: results : 2
2024-09-17 14:09:11,003 INFO ql.Driver: Completed executing command(queryId=hp_20240917140910_841e2de5-3150-4c88-9c59-25c77214142a); Time taken: 0.116 seconds
OK
2024-09-17 14:09:11,004 INFO ql.Driver: OK
2024-09-17 14:09:11,005 INFO ql.Driver: Concurrency mode is disabled, not creating a lock manager
2024-09-17 14:09:11,024 INFO Configuration.deprecation: mapred.input.dir is deprecated. Instead, use mapreduce.input.fileinputformat.inputdir
2024-09-17 14:09:11,111 INFO mapred.FileInputFormat: Total input files to process : 1
2024-09-17 14:09:11,175 INFO exec.ListSinkOperator: RECORDS_OUT_INTERMEDIATE:0, RECORDS_OUT_OPERATOR_LIST_SINK_0:2,
default
mydb
Time taken: 0.52 seconds, Fetched: 2 row(s)
2024-09-17 14:09:11,195 INFO CliDriver: Time taken: 0.52 seconds, Fetched: 2 row(s)
2024-09-17 14:09:11,196 INFO conf.HiveConf: Using the default value passed in for log id: 3c26b629-b286-465a-a0fc-03f0264456e5
2024-09-17 14:09:11,197 INFO session.SessionState: Resetting thread name to main
```

**CREATE TABLE student(name STRING,rollno INT);**

```
CREATE TABLE student(name STRING,rollno INT);
2024-09-17 14:37:35,448 INFO conf.HiveConf: Using the
2024-09-17 14:37:35,459 INFO session.SessionState: Upd
2024-09-17 14:37:35,559 INFO ql.Driver: Compiling com
inWritableFormatException(2027051)
```

Create a csv file to load the data in table.



```
name,rollno
Madhu,1
Priya,2
Ram,3
```

Load the data :

```
hive> LOAD DATA INPATH '/hivee/dataa.csv' INTO TABLE student;
2024-09-17 14:43:36,223 INFO conf.HiveConf: Using the default value p
2024-09-17 14:43:36,223 INFO session.SessionState: Updating thread na
2024-09-17 14:43:36,232 INFO ql.Driver: Compiling command(queryId=hp
E student
```

Put the csv file inside the hadoop directory:

```
C:\Windows\System32>hdfs dfs -mkdir /user/hive
mkdir: `/user/hive': File exists

C:\Windows\System32>hdfs dfs -mkdir /hivee

C:\Windows\System32>hdfs dfs -put C:\text\dataa.csv /hivee

C:\Windows\System32>hdfs dfs -cat /hivee/dataa.csv
name,rollno
Madhu,1
Priya,2
Ram,3
C:\Windows\System32>
```

**SHOW TABLES;** -Display the table in mydb.

```
hive> SHOW TABLES;
2024-09-17 14:43:52,449 IN
2024-09-17 14:43:52,577 INFO mapred.FileInputFormat
2024-09-17 14:43:52,634 INFO exec.ListSinkOperator
student
Time taken: 0.115 seconds, Fetched: 1 row(s)
```

**SELECT \* FROM student ;** -Display all rows of the table student.

```
): SELECT * FROM student
2024-09-17 14:44:40,380 INFO ql.Driver: Completed ex
2e98ba99fd); Time taken: 0.002 seconds
```

```
name,rollno      NULL
Madhu,1 NULL
Priya,2 NULL
Ram,3  NULL
Time taken: 4.905 seconds, Fetched: 4 row(s)
```

**DESC student;** -Display the structure of the table.

```
hive> DESC student;
2024-09-17 14:44:18,398 INFO conf.HiveConf:
```

```
2024-09-17 14:44:18,709 INFO exec.ListSinkOperato
name                string
rollno              int
Time taken: 0.245 seconds, Fetched: 2 row(s)
```

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