

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:**Installation of Hive:****1. Installing Apache Derby**

Install Apache Derby 10.14.2.0

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

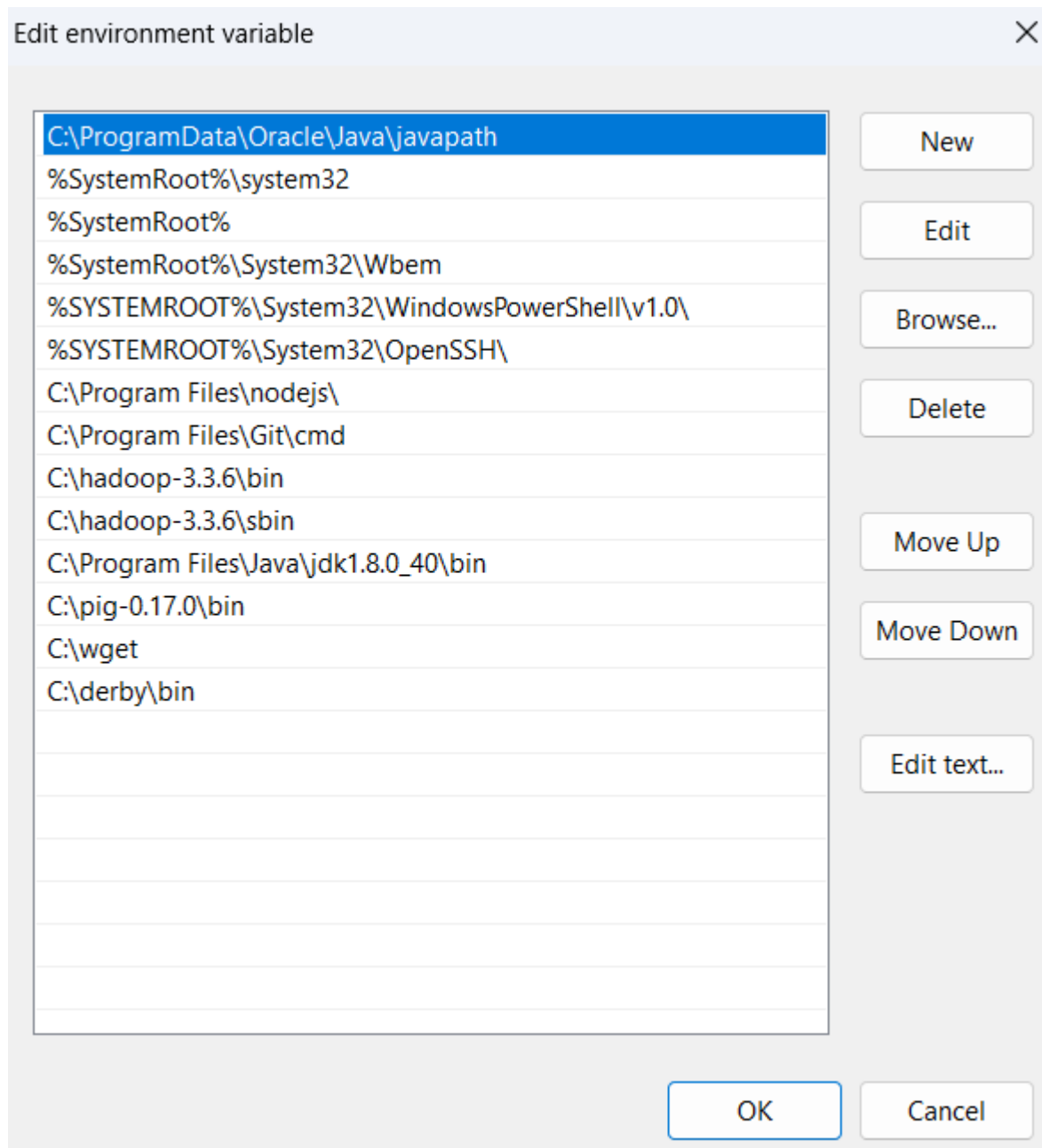
2. Downloading Apache Hive binaries

<https://downloads.apache.org/hive/hive-3.1.2/>

3. Setting environment variables

Variable	Value
HADOOP_HOME	C:\hadoop-3.3.6
HIVE_BIN	%HIVE_HOME%\bin
HIVE_HOME	C:\apache-hive-3.1.3-bin
HIVE_LIB	%HIVE_HOME%\lib

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



3.1. 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.

3.2. 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>

4. Starting Hadoop Services

`start-all.cmd`

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

```
StartNetworkServer -h 0.0.0.0
```

6. Starting Apache Hive

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

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

7. 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 sample;
```

2. Verify the database is present:

```
SHOW DATABASES;
```

3. Switch to the new database:

```
USE sample;
```

Create a Table in Hive

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

Add Data

```
INSERT INTO employee VALUES(1,"john");
```

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:

```
DESC employee;
```

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

```
SELECT * FROM employee;
```

OUTPUT:

```
Administrator: Command Prompt - StartNetworkServer -h 0.0.0.0
Microsoft Windows [Version 10.0.22631.4037]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\System32>StartNetworkServer -h 0.0.0.0
Sun Sep 08 12:50:16 IST 2024 : Security manager installed using the Basic server security policy.
Sun Sep 08 12:50:16 IST 2024 : Apache Derby Network Server - 10.14.2.0 - (1828579) started and ready to accept connections on port 1527
```

```
Administrator: Command Prompt - hive
Microsoft Windows [Version 10.0.22631.4037]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\System32>cd C:\hive\bin
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-3.3.6/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]
2024-09-08T12:54:07,769 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Found configuration file file:/C:/hive/conf/hive-site.xml
2024-09-08T12:54:07,950 WARN [main] org.apache.hadoop.hive.conf.HiveConf - HiveConf of name hive.server2.enable.doAS does not exist
2024-09-08T12:54:08,001 INFO [main] org.apache.hadoop.hive.conf.HiveConf - HiveConf of name hive.server2.enable.doAS does not exist
Metastore connection URL: jdbc:derby://localhost:1527/metastore_db2;create=true
2024-09-08T12:54:08,001 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
2024-09-08T12:54:08,001 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
```

```
Initialization script completed
schemaTool completed

C:\hive\bin>hive
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-3.3.6/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]
2024-09-08T12:54:54,137 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Found configuration file file:/C:/hive/conf/hive-site.xml
2024-09-08T12:54:54,315 WARN [main] org.apache.hadoop.hive.conf.HiveConf - HiveConf of name hive.server2.enable.doAS does not exist
2024-09-08T12:54:55,100 WARN [main] org.apache.hadoop.hive.conf.HiveConf - HiveConf of name hive.server2.enable.doAS does not exist
Hive Session ID = aa56317c-a2fe-4357-bd38-918e789e107b
Logging initialized using configuration in jar:file:/C:/hive/lib/hive-common-3.1.2.jar!/hive-log4j2.properties Async: true
2024-09-08T12:54:55,919 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Created HDFS directory: /tmp/hive/MONASHREE/aa56317c-a2fe-4357-bd38-918e789e107b
2024-09-08T12:54:55,924 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Created local directory: C:/hive/MONASHREE/aa56317c-a2fe-4357-bd38-918e789e107b
```

```
hive> CREATE TABLE Employees (EmployeeID INT, EMPLOYEENAME STRING);
2024-09-08T13:01:26,019 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: aa56317c-a2fe-4357-bd38-918e789e107b
2024-09-08T13:01:26,019 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to aa56317c-a2fe-4357-bd38-918e789e107b main
OK
Time taken: 0.493 seconds
2024-09-08T13:01:26,514 INFO [aa56317c-a2fe-4357-bd38-918e789e107b main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: aa56317c-a2fe-4357-bd38-918e789e107b
```

```
hive> INSERT INTO Employees (EmployeeID, EMPLOYEENAME) VALUES (1, 'John Doe');
2024-09-08T13:03:13,576 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: aa56317c-a2fe-4357-bd38-918e789e107b
2024-09-08T13:03:13,576 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to aa56317c-a2fe-4357-bd38-918e789e107b main
2024-09-08T13:03:15,513 INFO [aa56317c-a2fe-4357-bd38-918e789e107b main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localhost:9000/user/hive/warehouse/employees/.hive-staging_hive_2024-09-08_13-03-15_513_5508999880735781994-1
2024-09-08T13:03:15,665 INFO [aa56317c-a2fe-4357-bd38-918e789e107b main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localhost:9000/tmp/hive/MONASHREE/aa56317c-a2fe-4357-bd38-918e789e107b/hive_2024-09-08_13-03-15_574_738169738181158168-1/-mr-10000/.hive-staging_hive_2024-09-08_13-03-15_574_738169738181158168-1
Query ID = MONASHREE_20240908130313_2498a453-88c9-48ea-8f04-e3b28e9f1608
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):
```

```
hive> INSERT INTO Employees (EmployeeID, EMPLOYEENAME) VALUES (2, 'mary');
2024-09-08T13:04:53,387 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: aa56317c-a2fe-4357-bd38-918e789e107b
2024-09-08T13:04:53,388 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to aa56317c-a2fe-4357-bd38-918e789e107b main
2024-09-08T13:04:53,538 INFO [aa56317c-a2fe-4357-bd38-918e789e107b main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localhost:9000/user/hive/warehouse/employees/.hive-staging_hive_2024-09-08_13-04-53_400_5580999880735781994-1
2024-09-08T13:04:53,586 INFO [aa56317c-a2fe-4357-bd38-918e789e107b main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localhost:9000/tmp/hive/MONASHREE/aa56317c-a2fe-4357-bd38-918e789e107b/hive_2024-09-08_13-04-53_555_461141872240483956-1/-mr-10000/.hive-staging_hive_2024-09-08_13-04-53_555_461141872240483956-1
Query ID = MONASHREE_20240908130453_b7258b38-9c2a-460f-ae06-b8ab9fe18d16
```

```
hive>
> SELECT * FROM Employees;
2024-09-08T13:05:51,368 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: aa56317c-a2fe-4357-bd38-918e789e107b
2024-09-08T13:05:51,368 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to aa56317c-a2fe-4357-bd38-918e789e107b main
2024-09-08T13:05:51,474 INFO [aa56317c-a2fe-4357-bd38-918e789e107b main] org.apache.hadoop.hive.common.FileUtils - Creating directory if it doesn't exist: hdfs://localhost:9000/tmp/hive/MONASHREE/aa56317c-a2fe-4357-bd38-918e789e107b/hive_2024-09-08_13-05-51_388_4171213250699526575-1/-mr-10001/hive-staging_hive_2024-09-08_13-05-51_388_4171213250699526575-1
OK
1      John Doe
2      Mary
Time taken: 0.122 seconds, Fetched: 2 row(s)
2024-09-08T13:05:51,525 INFO [aa56317c-a2fe-4357-bd38-918e789e107b main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: aa56317c-a2fe-4357-bd38-918e789e107b
2024-09-08T13:05:51,526 INFO [aa56317c-a2fe-4357-bd38-918e789e107b main] org.apache.hadoop.hive.ql.session.SessionState - Resetting thread name to main
hive>
```

Browse Directory

Show 25 entries

Search:

<input type="checkbox"/>		Permission		Owner		Group		Size		Last Modified		Replication		Block Size		Name	
<input type="checkbox"/>		drwxr-xr-x		MONASHREE		supergroup		0 B		Sep 08 13:01		0		0 B		hive	
<input type="checkbox"/>		-rw-r--r--		MONASHREE		supergroup		54 B		Aug 19 19:09		1		128 MB		input.txt	

Showing 1 to 2 of 2 entries

Browse Directory

Show 25 entries

Search:

<input type="checkbox"/>		Permission		Owner		Group		Size		Last Modified		Replication		Block Size		Name	
<input type="checkbox"/>		drwxr-xr-x		MONASHREE		supergroup		0 B		Sep 08 13:01		0		0 B		warehouse	

Showing 1 to 1 of 1 entries

Browse Directory

Show 25 entries

Search:

<input type="checkbox"/>		Permission		Owner		Group		Size		Last Modified		Replication		Block Size		Name	
<input type="checkbox"/>		drwxr-xr-x		MONASHREE		supergroup		0 B		Sep 08 13:05		0		0 B		employees	

Showing 1 to 1 of 1 entries

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Utilities

Browse Directory

/user/hive/warehouse/employees

Go!

Show

25

entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	MONASHREE	supergroup	11 B	Sep 08 13:03	1	128 MB	000000_0	
<input type="checkbox"/>	-rw-r--r--	MONASHREE	supergroup	7 B	Sep 08 13:05	1	128 MB	000000_0_copy_1	

Showing 1 to 2 of 2 entries

Previous

1

Next

Block information --

Block 0

Block ID: 1073741966

Block Pool ID: BP-468074218-172.16.11.91-1722913037250

Generation Stamp: 1142

Size: 11

Availability:

LAPTOP-VG848917

File contents

1

John Doe

Block information --

Block 0 ▾

Block ID: 1073741981

Block Pool ID: BP-468074218-172.16.11.91-1722913037250

Generation Stamp: 1157

Size: 7

Availability:

- LAPTOP-VG848917

File contents

2□mary

RESULT:

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