LAB - 9

Name	Keval D Gandevia
Roll Number	CE046
ID	19CEUEG017
Subject	Big Data Analytics

Aim: CRUD operations and more using Hive query language.

❖ **Checking the java version:**

```
C:\Users\HP>java --version
java 11.0.8 2020-07-14 LTS
Java(TM) SE Runtime Environment 18.9 (build 11.0.8+10-LTS)
Java HotSpot(TM) 64-Bit Server VM 18.9 (build 11.0.8+10-LTS, mixed mode)
```

❖ **Checking the Hadoop version:**

```
C:\Users\HP>hadoop version
Hadoop 3.3.3
Source code repository https://github.com/apache/hadoop.git -r d37586cbda38c338d9fe481addda5a05fb516f71
Compiled by stevel on 2022-05-09T16:36Z
Compiled with protoc 3.7.1
From source with checksum eb96dd4a797b6989ae0cdb9db6efc6
This command was run using /F:/hadoop-3.3.3/hadoop-3.3.3/share/hadoop/common/hadoop-common-3.3.3.jar
```

❖ Starting the Hadoop server:

```
F:\hadoop-3.3.3\hadoop-3.3.3\sbin>start-dfs.sh

F:\hadoop-3.3.3\hadoop-3.3.3\sbin>
[main 2022-10-05T07:33:28.501Z] update#setState idle
[main 2022-10-05T07:33:30.399Z] Starting extension host with pid 15584 (fork() took 21 ms).
[main 2022-10-05T07:33:34.717Z] Waiting for extension host with pid 15584 to exit.
[main 2022-10-05T07:33:34.935Z] Extension host with pid 15584 exited with code: 0, signal: null.


F:\hadoop-3.3.3\hadoop-3.3.3\sbin>start-dfs.cmd

F:\hadoop-3.3.3\hadoop-3.3.3\sbin>jps
8688 NameNode
17068 DataNode
4908 Jps

F:\hadoop-3.3.3\hadoop-3.3.3\sbin>start-yarn.cmd
starting yarn daemons

F:\hadoop-3.3.3\hadoop-3.3.3\sbin>jps
8688 NameNode
20820 Jps
14488 NodeManager
17068 DataNode
9564 ResourceManager
```

❖ Starting the network server for hive:

```
C:\WINDOWS\system32>F:

F:\>cd db-derby-10.15.2.0-bin

F:\db-derby-10.15.2.0-bin>cd bin

F:\db-derby-10.15.2.0-bin\bin>startNetworkServer -h 0.0.0.0
Wed Oct 05 13:08:35 IST 2022 : Security manager installed using the Basic server security policy.
Wed Oct 05 13:08:35 IST 2022 : Apache Derby Network Server - 10.15.2.0 - (1873585) started and ready to accept connectio
ns on port 1527
```

❖ Ensuring that hive is properly installed or not:

```
E:\hadoop-env\apache-hive-3.1.2\bin>hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/E:/hadoop-env/apache-hive-3.1.2/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/E:/hadoop-env/hadoop-3.2.1/share/hadoop/common/lib/slf4j-log4j12-1.7.25.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]
2020-05-04T01:32:53,067 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Found configuration file file:/E:/hadoop-env/apache-hive-3.1.2/conf/hive-site.xml
2020-05-04T01:32:53,881 WARN [main] org.apache.hadoop.hive.conf.HiveConf - HiveConf of name hive.server2.enable.impersonation does not exist
2020-05-04T01:32:56,344 WARN [main] org.apache.hadoop.hive.conf.HiveConf - HiveConf of name hive.server2.enable.impersonation does not exist
Hive Session ID = 868ef6ea-bf7e-464b-969b-f75e1f453587

Logging initialized using configuration in jar:file:/E:/hadoop-env/apache-hive-3.1.2/lib/hive-common-3.1.2.jar!/hive-log4j2.properties Async: true
2020-05-04T01:32:59,638 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Created HDFS directory: /tmp/hive/HFadl
2020-05-04T01:32:59,649 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Created HDFS directory: /tmp/hive/HFadl/868ef6ea-bf7e-464b-969b-f75e1f453587
2020-05-04T01:32:59,664 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Created local directory: C:/Users/HFadl/AppData/Local/Temp/HFadl/868ef6ea-bf7e-464b
-969b-f75e1f453587
2020-05-04T01:32:59,673 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Created HDFS directory: /tmp/hive/HFadl/868ef6ea-bf7e-464b-969b-f75e1f453587/_tmp_s
pace.db
2020-05-04T01:32:59,701 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Using the default value passed in for log id: 868ef6ea-bf7e-464b-969b-f75e1f453587
2020-05-04T01:32:59,702 INFO [main] org.apache.hadoop.hive.ql.session.SessionState - Updating thread name to 868ef6ea-bf7e-464b-969b-f75e1f453587 main
2020-05-04T01:32:59,799 WARN [868ef6ea-bf7e-464b-969b-f75e1f453587 main] org.apache.hadoop.hive.conf.HiveConf - HiveConf of name hive.server2.enable.impersonation does
not exist
Hive-on-MR 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 rele
ases.
2020-05-04T01:36:36,797 INFO [868ef6ea-bf7e-464b-969b-f75e1f453587 main] CliDriver - Hive-on-MR 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 releases.
hive>
```

CRUD operations in Hive.

❖ Creating table:

```
> CREATE TABLE STUDENT
> (
> STD_ID INT,
> STD_NAME STRING,
> AGE INT,
> ADDRESS STRING
> )
> CLUSTERED BY (ADDRESS) into 3 buckets
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ','
> STORED as orc tblproperties('transactional'='true');
OK
Time taken: 1.458 seconds
hive> INSERT INTO TABLE STUDENT VALUES (101,'JAVACHAIN',30,'PAUL REVERE RD'),
>     (102,'ANTO',18,'29 NATHAN HALE'),
>     (103,'PRABU',23,'34 henry road'),
>     (104,'KUMAR',24,'gandhi road'),
>     (105,'jack',35,'Modi street');
Query ID = root_20151028113939_11fbd791-e036-471a-8a56-f1ad3c5571aa
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 3
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.reduces=<number>
Job running in-process (local Hadoop)
Hadoop job information for Stage-1: number of mappers: 0; number of reducers: 0
2015-10-28 11:40:30,405 Stage-1 map = 0%,   reduce = 0%
2015-10-28 11:40:32,198 Stage-1 map = 100%,  reduce = 0%
2015-10-28 11:40:38,046 Stage-1 map = 100%,  reduce = 33%
2015-10-28 11:40:41,471 Stage-1 map = 100%,  reduce = 67%
2015-10-28 11:40:44,843 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local816469325_0005
Loading data to table default.student
Table default.student stats: [numFiles=3, numRows=5, totalSize=1523, rawDataSize
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 0 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
Time taken: 56.464 seconds
```

❖ Update operation:

```
hive> select * from STUDENT;
OK
104     KUMAR    24       gandhi road
103     PRABU    23       34 henry road
102     ANTO     18       29 NATHAN HALE
101     JAVACHAIN        30        PAUL REVERE RD
Time taken: 2.237 seconds, Fetched: 4 row(s)
hive>
    >   update STUDENT SET std_id = 110 WHERE std_id = 101;
```

❖ Verifying the entries after update operation:

```
hive> select * from STUDENT;
OK
104     KUMAR    24       gandhi road
103     PRABU    23       34 henry road
102     ANTO     18       29 NATHAN HALE
110     JAVACHAIN        30        PAUL REVERE RD        updated
Time taken: 0.139 seconds, Fetched: 4 row(s)
```