

I am using cygwin to run all types of commands (shell and cmd)

HDFS file system running :

```
rajasekhharv@MyPC-RAJ ~/hadoop/hive
$ ../sbin/start-dfs.cmd

rajasekhharv@MyPC-RAJ ~/hadoop/hive
```

With above command it launch all name nodes and datanodes:

Name node:

```
Apache Hadoop Distribution - hadoop namenode
20/03/19 23:13:31 INFO hdfs.StateChange: DIR* completeFile: /hive-warehouse/servay_analysis_female_results/.hive-staging_hive_2020-03-19_23-13-25_124_12324353654651169
7-1/-ext-10000/_temporary/0/_temporary/attempt_20200319231326_0023_m_000011_1661/part-00011-b44d83b6-c337-4b19-ab72-beb0c9835729-c000 is closed by DFSClient_NONMAPREDU
E_1302770915_1
20/03/19 23:13:31 INFO hdfs.StateChange: DIR* completeFile: /hive-warehouse/servay_analysis_female_results/.hive-staging_hive_2020-03-19_23-13-25_124_12324353654651169
7-1/-ext-10000/_temporary/0/_temporary/attempt_20200319231326_0023_m_000012_1662/part-00012-b44d83b6-c337-4b19-ab72-beb0c9835729-c000 is closed by DFSClient_NONMAPREDU
E_1302770915_1
20/03/19 23:13:31 INFO hdfs.StateChange: DIR* completeFile: /hive-warehouse/servay_analysis_female_results/.hive-staging_hive_2020-03-19_23-13-25_124_12324353654651169
7-1/-ext-10000/_temporary/0/_temporary/attempt_20200319231326_0023_m_000010_1660/part-00010-b44d83b6-c337-4b19-ab72-beb0c9835729-c000 is closed by DFSClient_NONMAPREDU
E_1302770915_1
20/03/19 23:13:31 INFO hdfs.StateChange: DIR* completeFile: /hive-warehouse/servay_analysis_female_results/.hive-staging_hive_2020-03-19_23-13-25_124_12324353654651169
7-1/-ext-10000/_temporary/0/_temporary/attempt_20200319231326_0023_m_000013_1663/part-00013-b44d83b6-c337-4b19-ab72-beb0c9835729-c000 is closed by DFSClient_NONMAPREDU
E_1302770915_1
20/03/19 23:13:32 INFO hdfs.StateChange: BLOCK* allocate blk_1073741905_1081, replicas=127.0.0.1:50010 for /hive-warehouse/servay_analysis_female_results/.hive-staging
hive_2020-03-19_23-13-25_124_1232435365465116957-1/-ext-10000/_temporary/0/_temporary/attempt_20200319231326_0023_m_000014_1664/part-00014-b44d83b6-c337-4b19-ab72-beb0
c9835729-c000
20/03/19 23:13:32 INFO hdfs.StateChange: DIR* completeFile: /hive-warehouse/servay_analysis_female_results/.hive-staging_hive_2020-03-19_23-13-25_124_12324353654651169
7-1/-ext-10000/_temporary/0/_temporary/attempt_20200319231326_0023_m_000014_1664/part-00014-b44d83b6-c337-4b19-ab72-beb0c9835729-c000 is closed by DFSClient_NONMAPREDU
E_1302770915_1
20/03/19 23:13:33 INFO hdfs.StateChange: DIR* completeFile: /hive-warehouse/servay_analysis_female_results/.hive-staging_hive_2020-03-19_23-13-25_124_12324353654651169
7-1/-ext-10000/_SUCCESS is closed by DFSClient_NONMAPREDUCE_1302770915_1
20/03/20 05:16:09 WARN blockmanagement.BlockManager$HeartbeatManager: Skipping next heartbeat scan due to excessive pause
20/03/20 05:16:12 WARN util.JvmPauseMonitor: Detected pause in JVM or host machine (eg GC): pause of approximately 21456711ms
No GCs detected
20/03/20 05:16:15 INFO BlockStateChange: BLOCK* processReport 0xf93fc942d2acba1: from storage DS-775046d3-7343-4221-af21-01c3e0c13121 node DatanodeRegistration(127.0.
1:150010, datanodeuid=5e556c89-64b9-4e84-abed-ad64a3d26f9c, infoPort=50075, infoSecurePort=0, ipcPort=50020, storageInfo=lv=-57;cid=CID-fcb9e4c3-8aa4-46a7-8015-37beb
1a209nsid=1219704d43cc-1584629367534), blocks: 62, hasStaleStorage: false, processing time: 1485 msec, invalidatedBlocks: 0
20/03/20 05:16:15 INFO namenode.FSNamesystem: FSNamesystem write lock held for 1485 ms via
java.lang.Thread.getStackTrace(Thread.java:1556)
org.apache.hadoop.util.StringUtils.getStackTrace(StringUtils.java:1032)
org.apache.hadoop.hdfs.server.namenode.FSNamesystemLock.writeUnlock(FSNamesystemLock.java:233)
org.apache.hadoop.hdfs.server.namenode.FSNamesystemLock.writeUnlock(FSNamesystemLock.java:197)
org.apache.hadoop.hdfs.server.namenode.FSNamesystem.writeUnlock(FSNamesystem.java:1534)
org.apache.hadoop.hdfs.server.blockmanagement.BlockManager$BlockReportProcessingThread.processQueue(BlockManager.java:4001)
org.apache.hadoop.hdfs.server.blockmanagement.BlockManager$BlockReportProcessingThread.run(BlockManager.java:3973)
Number of suppressed write-lock reports: 0
Longest write-lock held interval: 1485
```

Data node:

```
Apache Hadoop Distribution - hadoop datanode
20/03/19 23:13:31 INFO datanode.DataNode: PacketResponder: BP-331090368-192.168.56.1-1584629367534:blk_1073741899_1075, type=LAST_IN_PIPELINE terminating
20/03/19 23:13:31 INFO datanode.ClientTrace: src: /127.0.0.1:50565, dest: /127.0.0.1:50010, bytes: 36, op: HDFS_WRITE, cliID: DFSClient_NONMAPREDUCE_1302770915_1, offse
t: 0, srvID: 5e556c89-64b9-4e84-abed-ad64a3d26f9c, blockid: BP-331090368-192.168.56.1-1584629367534:blk_1073741900_1076, duration: 2459701
20/03/19 23:13:31 INFO datanode.DataNode: PacketResponder: BP-331090368-192.168.56.1-1584629367534:blk_1073741900_1076, type=LAST_IN_PIPELINE terminating
20/03/19 23:13:31 INFO datanode.DataNode: Receiving BP-331090368-192.168.56.1-1584629367534:blk_1073741901_1077 src: /127.0.0.1:50566 dest: /127.0.0.1:50010
20/03/19 23:13:31 INFO datanode.ClientTrace: src: /127.0.0.1:50566, dest: /127.0.0.1:50010, bytes: 35, op: HDFS_WRITE, cliID: DFSClient_NONMAPREDUCE_1302770915_1, offse
t: 0, srvID: 5e556c89-64b9-4e84-abed-ad64a3d26f9c, blockid: BP-331090368-192.168.56.1-1584629367534:blk_1073741901_1077, duration: 2752900
20/03/19 23:13:31 INFO datanode.DataNode: PacketResponder: BP-331090368-192.168.56.1-1584629367534:blk_1073741901_1077, type=LAST_IN_PIPELINE terminating
20/03/19 23:13:31 INFO datanode.DataNode: Receiving BP-331090368-192.168.56.1-1584629367534:blk_1073741902_1078 src: /127.0.0.1:50567 dest: /127.0.0.1:50010
20/03/19 23:13:31 INFO datanode.ClientTrace: src: /127.0.0.1:50567, dest: /127.0.0.1:50010, bytes: 35, op: HDFS_WRITE, cliID: DFSClient_NONMAPREDUCE_1302770915_1, offse
t: 0, srvID: 5e556c89-64b9-4e84-abed-ad64a3d26f9c, blockid: BP-331090368-192.168.56.1-1584629367534:blk_1073741902_1078, duration: 3453999
20/03/19 23:13:31 INFO datanode.DataNode: PacketResponder: BP-331090368-192.168.56.1-1584629367534:blk_1073741902_1078, type=LAST_IN_PIPELINE terminating
20/03/19 23:13:31 INFO datanode.DataNode: Receiving BP-331090368-192.168.56.1-1584629367534:blk_1073741903_1079 src: /127.0.0.1:50568 dest: /127.0.0.1:50010
20/03/19 23:13:31 INFO datanode.ClientTrace: src: /127.0.0.1:50568, dest: /127.0.0.1:50010, bytes: 36, op: HDFS_WRITE, cliID: DFSClient_NONMAPREDUCE_1302770915_1, offse
t: 0, srvID: 5e556c89-64b9-4e84-abed-ad64a3d26f9c, blockid: BP-331090368-192.168.56.1-1584629367534:blk_1073741903_1079, duration: 2044600
20/03/19 23:13:31 INFO datanode.DataNode: PacketResponder: BP-331090368-192.168.56.1-1584629367534:blk_1073741903_1079, type=LAST_IN_PIPELINE terminating
20/03/19 23:13:31 INFO datanode.DataNode: Receiving BP-331090368-192.168.56.1-1584629367534:blk_1073741904_1080 src: /127.0.0.1:50569 dest: /127.0.0.1:50010
20/03/19 23:13:31 INFO datanode.ClientTrace: src: /127.0.0.1:50569, dest: /127.0.0.1:50010, bytes: 36, op: HDFS_WRITE, cliID: DFSClient_NONMAPREDUCE_1302770915_1, offse
t: 0, srvID: 5e556c89-64b9-4e84-abed-ad64a3d26f9c, blockid: BP-331090368-192.168.56.1-1584629367534:blk_1073741904_1080, duration: 2600600
20/03/19 23:13:31 INFO datanode.DataNode: PacketResponder: BP-331090368-192.168.56.1-1584629367534:blk_1073741904_1080, type=LAST_IN_PIPELINE terminating
20/03/19 23:13:32 INFO datanode.DataNode: Receiving BP-331090368-192.168.56.1-1584629367534:blk_1073741905_1081 src: /127.0.0.1:50570 dest: /127.0.0.1:50010
20/03/19 23:13:32 INFO datanode.ClientTrace: src: /127.0.0.1:50570, dest: /127.0.0.1:50010, bytes: 36, op: HDFS_WRITE, cliID: DFSClient_NONMAPREDUCE_1302770915_1, offse
t: 0, srvID: 5e556c89-64b9-4e84-abed-ad64a3d26f9c, blockid: BP-331090368-192.168.56.1-1584629367534:blk_1073741905_1081, duration: 4064499
20/03/19 23:13:32 INFO datanode.DataNode: PacketResponder: BP-331090368-192.168.56.1-1584629367534:blk_1073741905_1081, type=LAST_IN_PIPELINE terminating
20/03/20 05:16:12 WARN util.JvmPauseMonitor: Detected pause in JVM or host machine (eg GC): pause of approximately 21456711ms
No GCs detected
20/03/20 05:16:15 INFO datanode.DataNode: Successfully sent block report 0xf93fc942d2acba1, containing 1 storage report(s), of which we sent 1. The reports had 62 tot
al blocks and used 1 RPC(s). This took 4840 msec to generate and 3555 msec for RPC and NN processing. Got back one command: FinalizeCommand/5.
20/03/20 05:16:15 INFO datanode.DataNode: Got finalize command for block pool BP-331090368-192.168.56.1-1584629367534
```

Hive setup:

Initialize schema by selecting db using schema tool.

```

Fajasekharv@MyPC-RAJ ~/hadoop/hive
$ ./bin/schematool -dbType mysql -initSchema
which: no hadoop in (.:/C:/apps/hadoop/bin)
dirname: missing operand
Try 'dirname --help' for more information.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:C:/apps/hadoop/hive/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:C:/apps/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.10.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]
Metastore connection URL:      jdbc:mysql://localhost:3306/hivedb?createDatabaseIfNotExist=true
Metastore Connection Driver :  com.mysql.cj.jdbc.Driver
Metastore connection User:    root
Starting metastore schema initialization to 2.3.0
Initialization script hive-schema-2.3.0.mysql.sql
Initialization script completed
schemaTool completed

```

Hive- running meta store service:

```
rajasekhara@MyPC-RAJ ~/hadoop/hive
$ ./bin/hive --service metastore
2020-03-19 20:23:27: Starting Hive Metastore Server
which: no hadoop in (./C:\apps\hadoop\bin)
dirname: missing operand
Try 'dirname --help' for more information.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/C:/apps/hadoop/hive/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/C:/apps/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.10.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]
```

Hive thrift server 2 started:

[illegible]

Started mysql server:


```

rajasekharv@MYPC-KAJ /cygdrive/c/Program Files/mysql/bin
$ ./mysqld.exe --console
mysqld: Could not create or access the registry key needed for the MySQL application
to log to the Windows EventLog. Run the application with sufficient
privileges once to create the key, add the key manually, or turn off
logging for that application.
2020-03-18T17:50:28.554394Z 0 [Warning] TIMESTAMP with implicit DEFAULT value is deprecated. Please use --explicit_defaults_for_timestamp se
2020-03-18T17:50:28.554488Z 0 [Note] --secure-file-priv is set to NULL. Operations related to importing and exporting data are disabled
2020-03-18T17:50:28.555837Z 0 [ERROR] Cannot open Windows EventLog; check privileges, or start server with --log_syslog=0
2020-03-18T17:50:28.555884Z 0 [Note] C:\Program Files\mysql\bin\mysqld.exe (mysqld 5.7.29) starting as process 16820 ...
2020-03-18T17:50:28.564694Z 0 [Note] InnoDB: Mutexes and rw_locks use Windows interlocked functions
2020-03-18T17:50:28.564999Z 0 [Note] InnoDB: Uses event mutexes
2020-03-18T17:50:28.565368Z 0 [Note] InnoDB: Memory barrier is not used
2020-03-18T17:50:28.565746Z 0 [Note] InnoDB: Compressed tables use zlib 1.2.11
2020-03-18T17:50:28.566406Z 0 [Note] InnoDB: Number of pools: 1
2020-03-18T17:50:28.567007Z 0 [Note] InnoDB: Not using CPU crc32 instructions
2020-03-18T17:50:28.571559Z 0 [Note] InnoDB: Initializing buffer pool, total size = 128M, instances = 1, chunk size = 128M
2020-03-18T17:50:28.580582Z 0 [Note] InnoDB: Completed initialization of buffer pool
2020-03-18T17:50:28.765232Z 0 [Note] InnoDB: Highest supported file format is Barracuda.
2020-03-18T17:50:28.780642Z 0 [Note] InnoDB: Log scan progressed past the checkpoint lsn 2630804
2020-03-18T17:50:28.780783Z 0 [Note] InnoDB: Doing recovery: scanned up to log sequence number 2630823
2020-03-18T17:50:28.780928Z 0 [Note] InnoDB: Database was not shutdown normally!
2020-03-18T17:50:28.781037Z 0 [Note] InnoDB: Starting crash recovery.
2020-03-18T17:50:29.303721Z 0 [Note] InnoDB: Removed temporary tablespace data file: "ibtmp1"
2020-03-18T17:50:29.304089Z 0 [Note] InnoDB: Creating shared tablespace for temporary tables
2020-03-18T17:50:29.305191Z 0 [Note] InnoDB: Setting file 'ibtmp1' size to 12 MB. Physically writing the file full; Please wait ...
2020-03-18T17:50:29.440623Z 0 [Note] InnoDB: File 'ibtmp1' size is now 12 MB.
2020-03-18T17:50:29.488897Z 0 [Note] InnoDB: 96 redo rollback segment(s) found. 96 redo rollback segment(s) are active.
2020-03-18T17:50:29.489182Z 0 [Note] InnoDB: 32 non-redo rollback segment(s) are active.
2020-03-18T17:50:29.490096Z 0 [Note] InnoDB: Waiting for purge to start
2020-03-18T17:50:29.561689Z 0 [Note] InnoDB: 5.7.29 started; log sequence number 2630823
2020-03-18T17:50:29.562423Z 0 [Note] InnoDB: Loading buffer pool(s) from C:\Program Files\mysql\data\ib_buffer_pool
2020-03-18T17:50:29.562455Z 0 [Note] Plugin 'FEDERATED' is disabled.
2020-03-18T17:50:29.939909Z 0 [Note] Found ca.pem, server-cert.pem and server-key.pem in data directory. Trying to enable SSL support using
2020-03-18T17:50:29.940117Z 0 [Note] Skipping generation of SSL certificates as certificate files are present in data directory.
2020-03-18T17:50:29.941689Z 0 [Warning] CA certificate ca.pem is self signed.
2020-03-18T17:50:29.942056Z 0 [Note] Skipping generation of RSA key pair as key files are present in data directory.
2020-03-18T17:50:30.094620Z 0 [Note] Server hostname (bind-address): '*'; port: 3306
2020-03-18T17:50:30.095490Z 0 [Note] IPv6 is available.

```

Results from local execution stored in hive and mysql :

Hive tale and results:

```

~/hadoop/hive
lservay2020
lservay_analysis_all_results
lservay_analysis_female_results
lservay_analysis_results
lservay_data_dump

```

5 rows selected (1.615 seconds)

```

0: jdbc:hive2://localhost:10000/default> select * from servay_analysis_all_results;

```

| servay_analysis_all_results.category | servay_analysis_all_results.questionid | servay_analysis_all_results.year | servay_analysis_all_results.average |
|--------------------------------------|--|----------------------------------|-------------------------------------|
| ALL | Q040 | 2008 | 16.37172413793103 |
| ALL | Q040 | 2010 | 16.23184185185185 |
| ALL | Q040 | 2012 | 15.59483648881238 |
| ALL | Q040 | 2014 | 15.436269430051802 |
| ALL | Q040 | 2016 | 15.148615916955029 |
| ALL | Q041 | 2008 | 14.983793103448296 |
| ALL | Q041 | 2010 | 15.064716006884682 |
| ALL | Q041 | 2012 | 14.276592082616187 |
| ALL | Q041 | 2014 | 13.943350604490504 |
| ALL | Q041 | 2016 | 13.53235294117647 |
| ALL | Q060 | 2008 | 13.218181818181813 |
| ALL | Q060 | 2010 | 13.046984126984121 |
| ALL | Q060 | 2012 | 11.963057324840753 |
| ALL | Q060 | 2014 | 11.450638977635784 |
| ALL | Q060 | 2016 | 11.577635782747599 |

15 rows selected (3.148 seconds)

```

0: jdbc:hive2://localhost:10000/default> select * from servay_analysis_female_results;

```

| servay_analysis_female_results.category | servay_analysis_female_results.questionid | servay_analysis_female_results.year | servay_analysis_female_results.average |
|---|---|-------------------------------------|--|
| FEMALE | Q040 | 2008 | 16.2962962962963 |
| FEMALE | Q040 | 2010 | 16.23518518518518 |
| FEMALE | Q040 | 2012 | 15.648148148148145 |
| FEMALE | Q040 | 2014 | 15.411111111111113 |
| FEMALE | Q040 | 2016 | 15.39259259259259 |
| FEMALE | Q041 | 2008 | 13.916666666666667 |
| FEMALE | Q041 | 2010 | 14.161111111111111 |
| FEMALE | Q041 | 2012 | 13.594444444444445 |
| FEMALE | Q041 | 2014 | 13.198148148148144 |
| FEMALE | Q041 | 2016 | 12.957407407407413 |
| FEMALE | Q060 | 2008 | 12.72862962962963 |
| FEMALE | Q060 | 2010 | 12.38148148148148 |
| FEMALE | Q060 | 2012 | 11.264814814814816 |
| FEMALE | Q060 | 2014 | 10.725925925925926 |
| FEMALE | Q060 | 2016 | 10.996296296296292 |

15 rows selected (1.758 seconds)

```

0: jdbc:hive2://localhost:10000/default>

```

My sql data bases and tables: (hived is created by schematool), metricsdb created by spark-App. Tables are from metrics db. Implementation stores each query results in specific table servay_analysis_all_results,servay_analysis_female_results .

```
mysql> show databases;
+-----+
| Database |
+-----+
| information_schema |
| hivedb |
| metricsdb |
| mysql |
| performance_schema |
| sys |
+-----+
6 rows in set (0.00 sec)

mysql> show tables;
+-----+
| Tables_in_metricsdb |
+-----+
| servay_analysis_all_results |
| servay_analysis_female_results |
| servay_analysis_results |
+-----+
3 rows in set (0.00 sec)

mysql> |
```

Hive db tables :

```
| part_col_privs |
| part_col_stats |
| part_privs |
| partition_events |
| partition_key_vals |
| partition_keys |
| partition_params |
| partitions |
| role_map |
| roles |
| sd_params |
| sds |
| sequence_table |
| serde_params |
| serdes |
| skewed_col_names |
| skewed_col_value_loc_map |
| skewed_string_list |
| skewed_string_list_values |
| skewed_values |
| sort_cols |
| tab_col_stats |
| table_params |
| tbl_col_privs |
| tbl_privs |
| tbls |
| txn_components |
| txns |
| type_fields |
| types |
| version |
| write_set |
+-----+
57 rows in set (0.00 sec)

mysql> show tables;
```

Mysql results tables:

```
mysql> select * from servay_analysis_all_results;
```

| category | questionid | year | avarage |
|----------|------------|------|--------------------|
| ALL | Q040 | 2008 | 16.37172413793103 |
| ALL | Q040 | 2014 | 15.436269430051802 |
| ALL | Q040 | 2012 | 15.599483648881238 |
| ALL | Q040 | 2010 | 16.231841652323595 |
| ALL | Q040 | 2016 | 15.148615916955029 |
| ALL | Q041 | 2008 | 14.983793103448296 |
| ALL | Q041 | 2014 | 13.943350604490504 |
| ALL | Q041 | 2010 | 15.064716006884682 |
| ALL | Q041 | 2012 | 14.276592082616187 |
| ALL | Q041 | 2016 | 13.53235294117647 |
| ALL | Q060 | 2008 | 13.218181818181813 |
| ALL | Q060 | 2010 | 13.046984126984121 |
| ALL | Q060 | 2012 | 11.963057324840753 |
| ALL | Q060 | 2014 | 11.450638977635784 |
| ALL | Q060 | 2016 | 11.577635782747599 |

```
15 rows in set (0.00 sec)
```

```
mysql> select * from servay_analysis_female_results;
```

| category | questionid | year | avarage |
|----------|------------|------|--------------------|
| FEMALE | Q040 | 2010 | 16.23518518518518 |
| FEMALE | Q040 | 2012 | 15.648148148148145 |
| FEMALE | Q040 | 2008 | 16.2962962962963 |
| FEMALE | Q040 | 2016 | 15.39259259259259 |
| FEMALE | Q040 | 2014 | 15.411111111111113 |
| FEMALE | Q041 | 2008 | 13.916666666666667 |
| FEMALE | Q041 | 2010 | 14.161111111111111 |
| FEMALE | Q041 | 2014 | 13.198148148148144 |
| FEMALE | Q041 | 2012 | 13.594444444444445 |
| FEMALE | Q041 | 2016 | 12.957407407407413 |
| FEMALE | Q060 | 2008 | 12.72962962962963 |
| FEMALE | Q060 | 2010 | 12.38148148148148 |
| FEMALE | Q060 | 2012 | 11.264814814814816 |
| FEMALE | Q060 | 2014 | 10.725925925925926 |
| FEMALE | Q060 | 2016 | 10.996296296296292 |

```
15 rows in set (0.00 sec)
```