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

HDFS file system running :

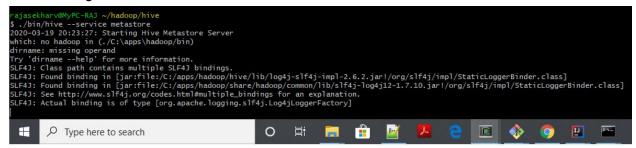With above command it launch all name nodes and datanodes:
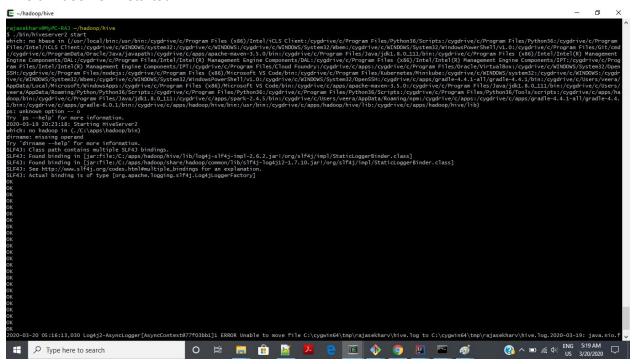Name node:

Data node:

Hive setup:

Initialize schema by selecting db using schema tool.


```
rajasekharv@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:


```
rajasekharv@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:

Started mysql server:

```
rajasekharv@MyPC-RAJ /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 resulsts:

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 .

Spark query results from logs:

Query : "select 'ALL' AS category,questionid,yearstart AS year,avg(data_value) AS avarage from servay2020 group by questionid,yearstart order by questionid,yearstart"

```
+--------+----------+----+------------------+
|category|questionid|year|           avarage|
+--------+----------+----+------------------+
|     ALL|      Q040|2008| 16.37172413793103|
|     ALL|      Q040|2010|16.231841652323595|
|     ALL|      Q040|2012|15.599483648881238|
|     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|
+--------+----------+----+------------------+
```

Query : "select 'FEMALE' AS category,questionid,yearstart AS year,avg(data_value) AS avarage from servay2020 where gender ='Female' group by questionid,yearstart order by questionid,yearstart"

20/03/21 07:17:06 INFO DAGScheduler: Job 9 finished: show at BatchDataAnalyser.java:89, took 1.252425 s

```
+--------+----------+----+------------------+
|category|questionid|year|           avarage|
+--------+----------+----+------------------+
|  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| 13.91666666666667|
|  FEMALE|      Q041|2010|14.161111111111111|
|  FEMALE|      Q041|2012|13.594444444444445|
|  FEMALE|      Q041|2014|13.198148148148144|
```

```
|  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|
+--------+----------+----+------------------+
```

① Not secure | mypc-raj:4040/SQL/execution/?id=3

Details for Query 3

Submitted Time: 2020/03/21 07:27:02
Duration: 3 s
Succeeded Jobs: 3

```
Scan hive default.servay_data_dump

number of output rows: 73,440
```

```
WholeStageCodegen
8.9 s (272 ms, 1.3 s, 1.5 s)
```

```
HashAggregate

aggregate time total (min, med, max):
8.6 s (271 ms, 1.1 s, 1.4 s)
number of output rows: 8
```

```
Exchange

data size total (min, med, max):
120.0 B (15.0 B, 15.0 B, 15.0 B)
```

```
WholeStageCodegen
62 ms (62 ms, 62 ms, 62 ms)
```

```
HashAggregate

aggregate time total (min, med, max):
61 ms (61 ms, 61 ms, 61 ms)
number of output rows: 1
```

Waiting for mypc-...

Type here to search

ENG 7:30 AM
IN 3/21/2020

---

① Not secure | mypc-raj:4040/SQL/execution/?id=3

```
   +- Project [cast(count(1)#321L as string) AS count(1)#324]
      +- Aggregate [count(1) AS count(1)#321L]
         +- SubqueryAlias `default`.`servay_data_dump`
            +- HiveTableRelation `default`.`servay_data_dump`, org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe, [yearstart#290, yearend#291, locationabbr#292, locationdesc#293, datasource#294,
class#295, topic#296, question#297, data_value_unit#298, data_value_type#299, data_value#300, data_value_alt#301, data_value_footnote_symbol#302, data_value_footnote#303, low_confidence_limit#304,
high_confidence_limit#305, sample_size#306, total#307, age(months)#308, gender#309, race/ethnicity#310, geolocation#311, classid#312, topicid#313, ... 7 more fields]


== Analyzed Logical Plan ==
count(1): string
GlobalLimit 21
+- LocalLimit 21
   +- Project [cast(count(1)#321L as string) AS count(1)#324]
      +- Aggregate [count(1) AS count(1)#321L]
         +- SubqueryAlias `default`.`servay_data_dump`
            +- HiveTableRelation `default`.`servay_data_dump`, org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe, [yearstart#290, yearend#291, locationabbr#292, locationdesc#293, datasource#294,
class#295, topic#296, question#297, data_value_unit#298, data_value_type#299, data_value#300, data_value_alt#301, data_value_footnote_symbol#302, data_value_footnote#303, low_confidence_limit#304,
high_confidence_limit#305, sample_size#306, total#307, age(months)#308, gender#309, race/ethnicity#310, geolocation#311, classid#312, topicid#313, ... 7 more fields]


== Optimized Logical Plan ==
GlobalLimit 21
+- LocalLimit 21
   +- Aggregate [cast(count(1) as string) AS count(1)#324]
      +- Project
         +- HiveTableRelation `default`.`servay_data_dump`, org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe, [yearstart#290, yearend#291, locationabbr#292, locationdesc#293, datasource#294,
class#295, topic#296, question#297, data_value_unit#298, data_value_type#299, data_value#300, data_value_alt#301, data_value_footnote_symbol#302, data_value_footnote#303, low_confidence_limit#304,
high_confidence_limit#305, sample_size#306, total#307, age(months)#308, gender#309, race/ethnicity#310, geolocation#311, classid#312, topicid#313, ... 7 more fields]


== Physical Plan ==
CollectLimit 21
+- *(2) HashAggregate(keys=[], functions=[count(1)], output=[count(1)#324])
   +- Exchange SinglePartition
      +- *(1) HashAggregate(keys=[], functions=[partial_count(1)], output=[count#327L])
         +- Scan hive default.servay_data_dump HiveTableRelation `default`.`servay_data_dump`, org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe, [yearstart#290, yearend#291, locationabbr#292,
locationdesc#293, datasource#294, class#295, topic#296, question#297, data_value_unit#298, data_value_type#299, data_value#300, data_value_alt#301, data_value_footnote_symbol#302,
data_value_footnote#303, low_confidence_limit#304, high_confidence_limit#305, sample_size#306, total#307, age(months)#308, gender#309, race/ethnicity#310, geolocation#311, classid#312,
```
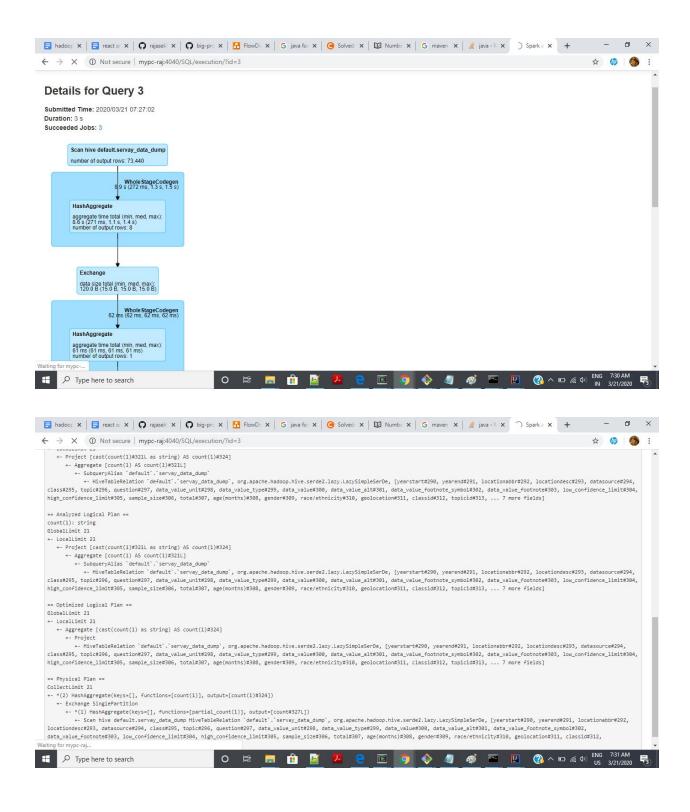
Waiting for mypc-raj...

Type here to search

ENG 7:31 AM
US 3/21/2020

User: rajasekharv
Total Uptime: 11 min
Scheduling Mode: FIFO
Active Jobs: 1
Completed Jobs: 6

▶ Event Timeline

▼ **Active Jobs (1)**

Job Id ▼	Description		Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
6	saveAsTable at BatchDataAnalyser.java:90 saveAsTable at BatchDataAnalyser.java:90	(kill)	2020/03/21 07:35:18	2 s	0/3	82/217 (4 running)

▼ **Completed Jobs (6)**

Job Id ▼	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
5	saveAsTable at BatchDataAnalyser.java:90 saveAsTable at BatchDataAnalyser.java:90	2020/03/21 07:35:16	2 s	2/2	202/202
4	show at BatchDataAnalyser.java:89 show at BatchDataAnalyser.java:89	2020/03/21 07:28:54	5 s	2/2	202/202
3	show at BatchDataAnalyser.java:78 show at BatchDataAnalyser.java:78	2020/03/21 07:27:03	3 s	2/2	9/9
2	saveAsTable at BatchDataAnalyser.java:74 saveAsTable at BatchDataAnalyser.java:74	2020/03/21 07:26:22	3 s	1/1	1/1
1	csv at BatchDataAnalyser.java:66 csv at BatchDataAnalyser.java:66	2020/03/21 07:25:46	1.0 s	1/1	1/1
0	csv at BatchDataAnalyser.java:66 csv at BatchDataAnalyser.java:66	2020/03/21 07:25:43	2 s	1/1	1/1

Spark 2.4.5   Jobs   Stages   Storage   Environment   Executors   SQL   **Spark app** application UI

Spark Jobs (?)

User: rajasekharv
Total Uptime: 12 min
Scheduling Mode: FIFO
Completed Jobs: 7

▶ Event Timeline

▼ **Completed Jobs (7)**

Job Id ▼	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
6	saveAsTable at BatchDataAnalyser.java:90 saveAsTable at BatchDataAnalyser.java:90	2020/03/21 07:35:18	12 s	2/2 (1 skipped)	215/215 (2 skipped)
5	saveAsTable at BatchDataAnalyser.java:90 saveAsTable at BatchDataAnalyser.java:90	2020/03/21 07:35:16	2 s	2/2	202/202
4	show at BatchDataAnalyser.java:89 show at BatchDataAnalyser.java:89	2020/03/21 07:28:54	5 s	2/2	202/202
3	show at BatchDataAnalyser.java:78 show at BatchDataAnalyser.java:78	2020/03/21 07:27:03	3 s	2/2	9/9
2	saveAsTable at BatchDataAnalyser.java:74 saveAsTable at BatchDataAnalyser.java:74	2020/03/21 07:26:22	3 s	1/1	1/1
1	csv at BatchDataAnalyser.java:66 csv at BatchDataAnalyser.java:66	2020/03/21 07:25:46	1.0 s	1/1	1/1
0	csv at BatchDataAnalyser.java:66	2020/03/21 07:25:43	2 s	1/1	1/1

Spark 2.4.5   Jobs   Stages   Storage   Environment   Executors   SQL

Spark app application UI

Spark Jobs [?]

User: rajasekharv
Total Uptime: 13 min
Scheduling Mode: FIFO
Completed Jobs: 12

▸ Event Timeline

▾ **Completed Jobs (12)**

Job Id ▾	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
11	saveAsTable at BatchDataAnalyser.java:90 saveAsTable at BatchDataAnalyser.java:90	2020/03/21 07:36:59	4 s	2/2 (1 skipped)	215/215 (2 skipped)
10	saveAsTable at BatchDataAnalyser.java:90 saveAsTable at BatchDataAnalyser.java:90	2020/03/21 07:36:57	2 s	2/2	202/202
9	show at BatchDataAnalyser.java:89 show at BatchDataAnalyser.java:89	2020/03/21 07:36:44	2 s	2/2	202/202
8	save at BatchDataAnalyser.java:95 save at BatchDataAnalyser.java:95	2020/03/21 07:36:07	3 s	2/2 (1 skipped)	215/215 (2 skipped)
7	save at BatchDataAnalyser.java:95 save at BatchDataAnalyser.java:95	2020/03/21 07:36:05	2 s	2/2	202/202
6	saveAsTable at BatchDataAnalyser.java:90 saveAsTable at BatchDataAnalyser.java:90	2020/03/21 07:35:18	12 s	2/2 (1 skipped)	215/215 (2 skipped)
5	saveAsTable at BatchDataAnalyser.java:90	2020/03/21 07:35:16	2 s	2/2	202/202

Spark 2.4.5   Jobs   Stages   Storage   Environment   Executors   SQL

Spark app application UI

SQL

Running Queries: 1
Completed Queries: 9

Running Queries (1)

ID	Description		Submitted	Duration	Running Job IDs	Succeeded Job IDs	Failed Job IDs
9	save at BatchDataAnalyser.java:95	+details	2020/03/21 07:37:48	0.2 s			

Completed Queries (9)

ID	Description		Submitted	Duration	Job IDs
8	saveAsTable at BatchDataAnalyser.java:90	+details	2020/03/21 07:36:56	8 s	[10][11]
7	show at BatchDataAnalyser.java:89	+details	2020/03/21 07:36:44	2 s	[9]
6	save at BatchDataAnalyser.java:95	+details	2020/03/21 07:36:03	7 s	[7][8]
5	saveAsTable at BatchDataAnalyser.java:90	+details	2020/03/21 07:35:15	18 s	[5][6]
4	show at BatchDataAnalyser.java:89	+details	2020/03/21 07:28:46	13 s	[4]
3	show at BatchDataAnalyser.java:78	+details	2020/03/21 07:27:02	3 s	[3]
2	saveAsTable at BatchDataAnalyser.java:74	+details	2020/03/21 07:26:20	6 s	[2]
1	createOrReplaceTempView at BatchDataAnalyser.java:71	+details	2020/03/21 07:25:47	5 ms	
0	csv at BatchDataAnalyser.java:66	+details	2020/03/21 07:25:40	6 s	[0]

Tales 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 | 13.91666666666667  |
| 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)
```