COMP-SCI 5540: PRINCIPLES OF BIG DATA MANAGEMENT

SPRING 2016

Phase 2

Team Members:

Santosh Kumar

Sree Rama Raju Pericharla

Shanmukha Reddy Aalla

Source code github link:

https://github.com/sreeram66/PB.git

Principles of Big Data Management

INTRODUCTION:

We had collected the tweets using python and then analyzed it. The collected tweets are on the 2016 cricket world T20 and by analyzing the tweets using the hashtag “wt20”. Using the tweets tweeted represented important pictorial presentation of tweets on that period of time.

Queries used to analyze the collected tweets are shown below:

Query1: No of users based on the starting letter of their names

```
val query = sqlContext.sql("select count(user.name) as AtoB from querytable1 where user.name like 'A%' OR user.name like 'B%' OR user.name like 'C%' OR user.name like 'D%' OR user.name like 'E%'");
  val query = sqlContext.sql("select count(user.name) as FtoJ from querytable1 where user.name like 'F%' OR user.name like 'G%' OR user.name like 'H%' OR user.name like 'I%' OR user.name like 'J%'");
  val query = sqlContext.sql("select count(user.name) as KtoO from querytable1 where user.name like 'K%' OR user.name like 'L%' OR user.name like 'M%' OR user.name like 'N%' OR user.name like 'O%'");
  val query = sqlContext.sql("select count(user.name) as PtoT from querytable1 where user.name like 'P%' OR user.name like 'Q%' OR user.name like 'R%' OR user.name like 'S%' OR user.name like 'T%'");
  val query = sqlContext.sql("select count(user.name) as UtoZ from querytable1 where user.name like 'U%' OR user.name like 'V%' OR user.name like 'W%' OR user.name like 'X%' OR user.name like 'Y%' OR user.name like 'Z%'");
```

Query2: Average followers count of users based on language in descending order

```
val query = sqlContext.sql("select lang,avg(user.followers_count) as followers from querytable1 group by lang order by followers desc");
```

Query3: Name, followers count, favourites count for top 3 popular users

```
val query = sqlContext.sql("select user.screen_name, user.followers_count,user.favourites_count,user.friends_count from querytable1 order by user.followers_count desc limit 3");
```

Query4: No of users having statuses count more than 10000 but still are not verified and no of users with statuses less than 10000 but still verified

```
val query = sqlContext.sql("select count(user.name) as morecount from querytable1 where user.statuses_count>10000 and user.verified=false");
val query = sqlContext.sql("select count(user.name) as lesscount from querytable1 where user.statuses_count<10000 and user.verified=true");
```

Query5: No of users who tweeted about top 3 players from India.

```
val query = sqlContext.sql("select count(user.name) as kohli from querytable1 where text like '%kohli%' ");
val query = sqlContext.sql("select count(user.name) as dhoni from querytable1 where text like '%dhoni%' ");
val query = sqlContext.sql("select count(user.name) as ashwin from querytable1 where text like '%ashwin%' ");
```

Query6: No of users created from the beginning of twitters launching till now categorized per 3 years

```
val query = sqlContext.sql("select count(user.created_at) as one from querytable1 where user.created_at like '%2006%' or user.created_at like '%2007%' or user.created_at like '%2008%'");
  val query = sqlContext.sql("select count(user.created_at) as two from querytable1 where user.created_at like '%2009%' or user.created_at like '%2010%' or user.created_at like '%2011%'");
  val query = sqlContext.sql("select count(user.created_at) as three from querytable1 where user.created_at like '%2012%' or user.created_at like '%2013%' or user.created_at like '%2014%'");
  val query = sqlContext.sql("select count(user.created_at) as four from querytable1 where user.created_at like '%2015%' or user.created_at like '%2016%'");
```

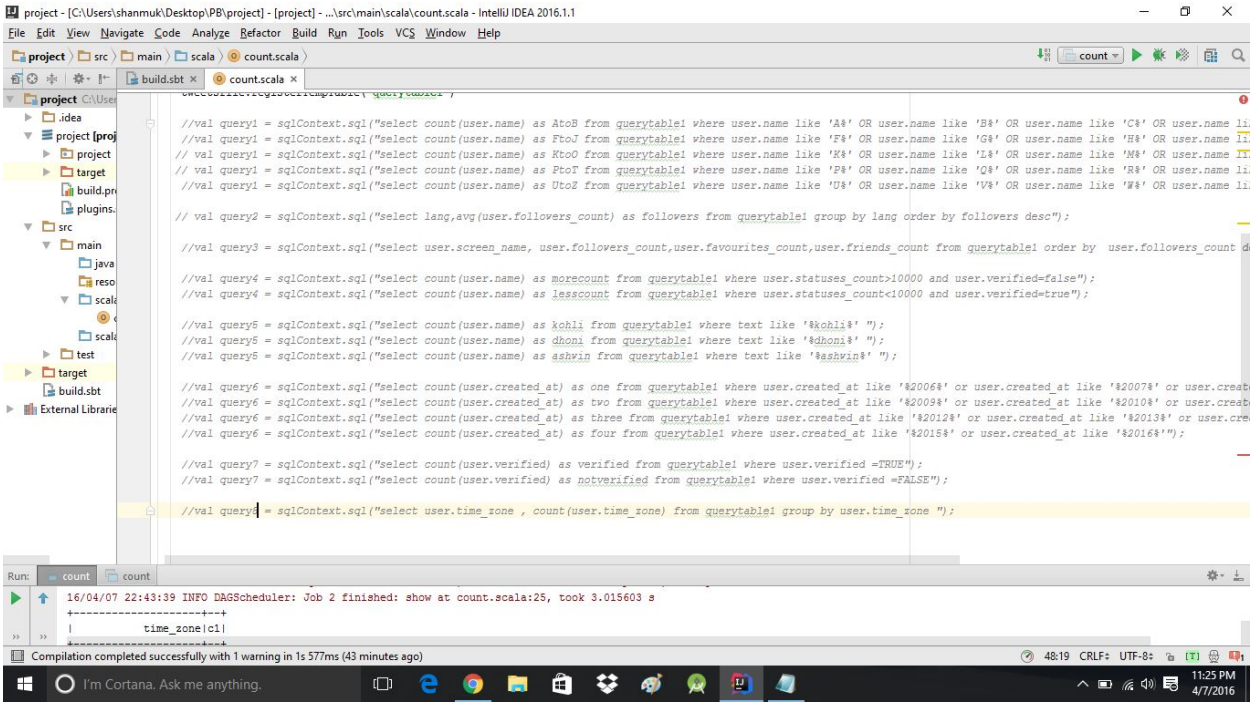
Query7: Total no of users who are verified vs non-verified

```
val query = sqlContext.sql("select count(user.verified) as verified from querytable1 where user.verified =TRUE");
  val query = sqlContext.sql("select count(user.verified) as notverified from querytable1 where user.verified =FALSE");
```

Query8: Highest no of tweets coming from different time zones of the world

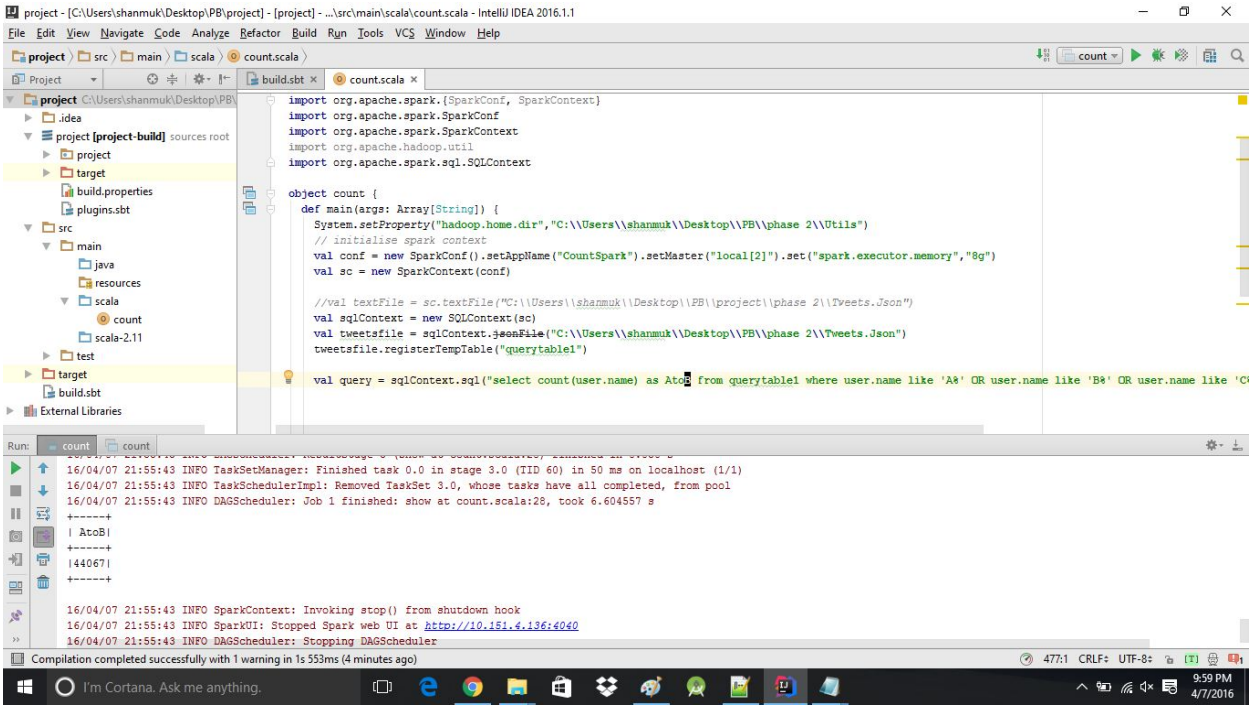
```
  val query = sqlContext.sql("select user.time_zone , count(user.time_zone) from querytable1 group by user.time_zone ");
```

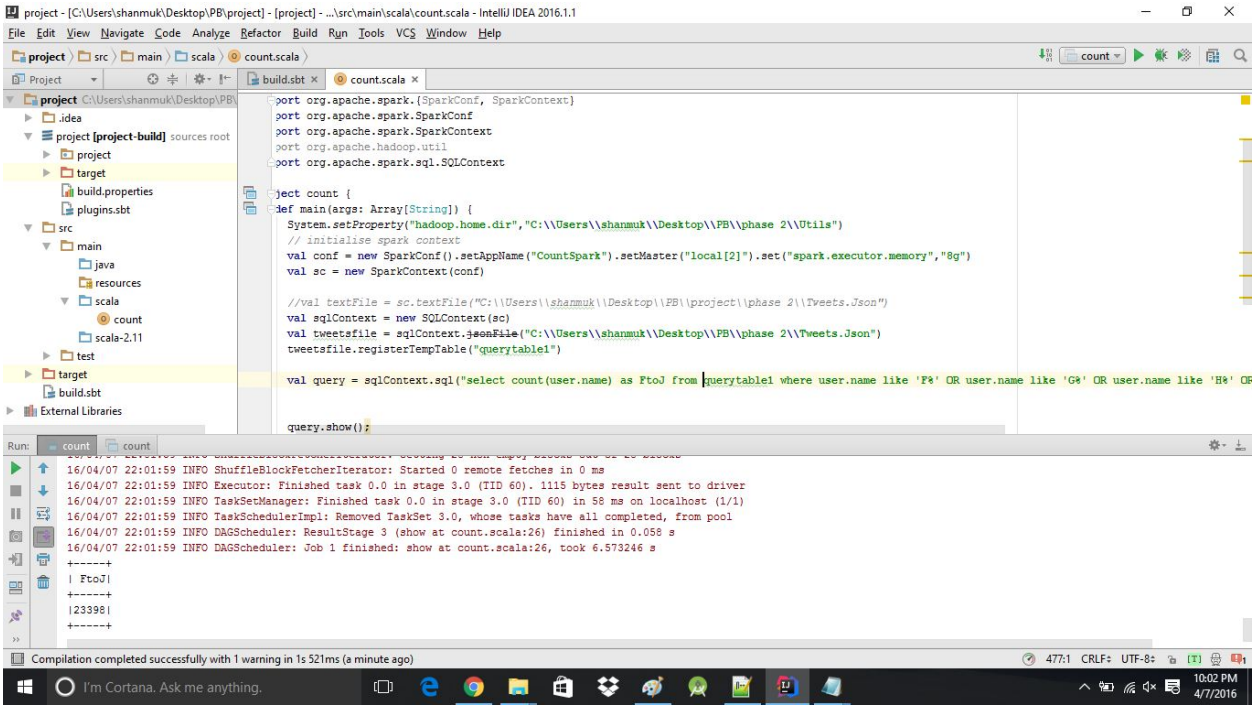
Screenshot of all the queries:

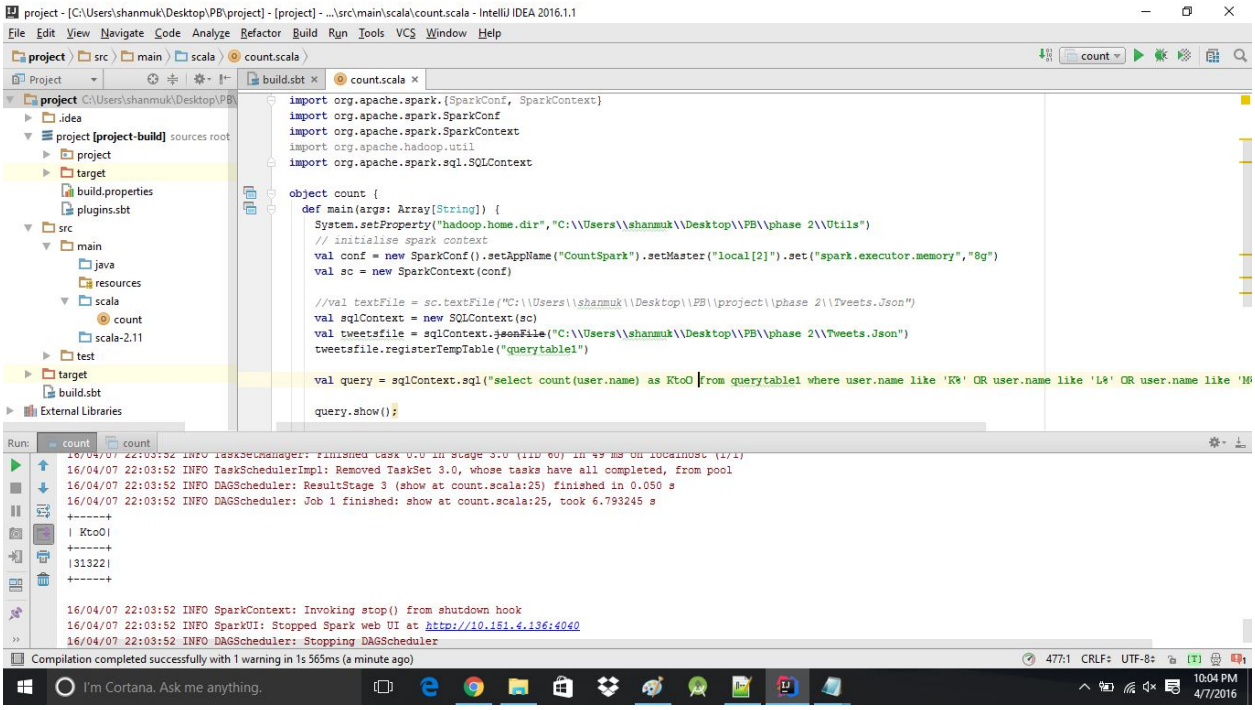


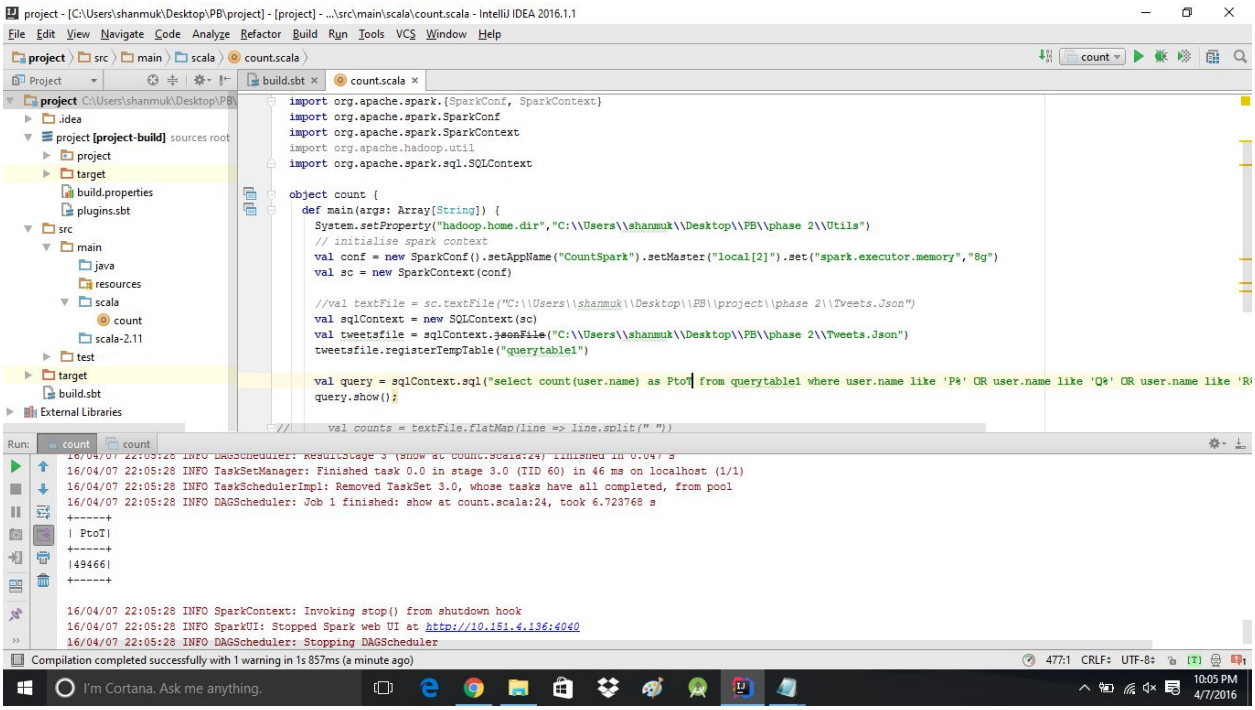
Output Screenshots:

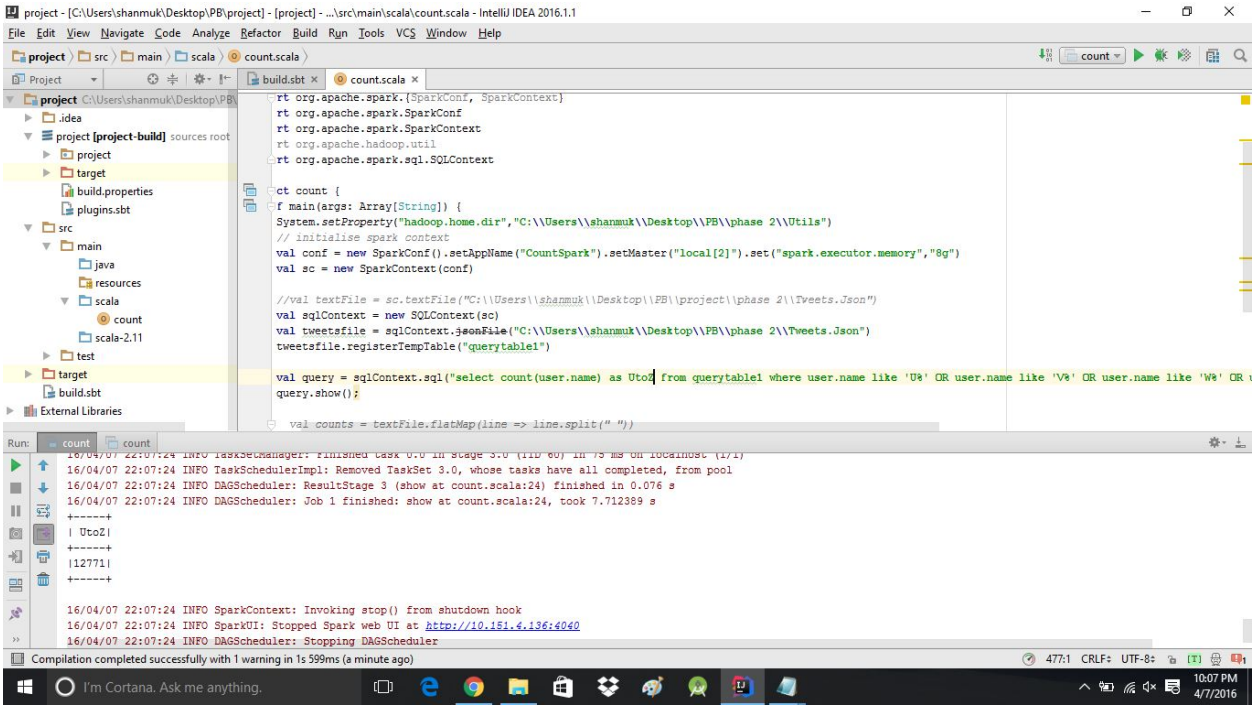
Query1:



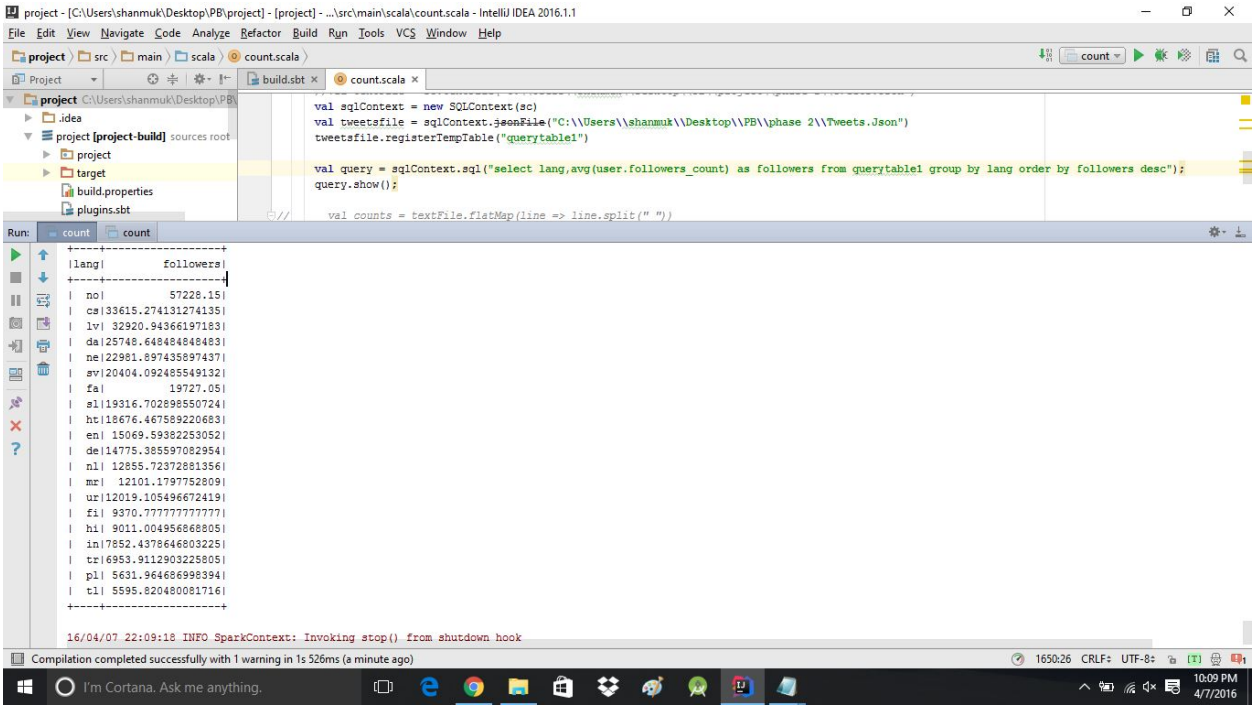

project - [C:\Users\shanmuk\Desktop\PB\project] - [project] - ...\src\main\scala\count.scala - IntelliJ IDEA 2016.1.1
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
project src main scala count.scala
count
Project
build.sbt
count.scala
project C:\Users\shanmuk\Desktop\PB\
.idea
project [project-build] sources root
project
target
build.properties
plugins.sbt
src
main
java
resources
scala
count
scala-2.11
test
target
build.sbt
External Libraries
port org.apache.spark.{SparkConf, SparkContext}
port org.apache.spark.SparkConf
port org.apache.spark.SparkContext
port org.apache.hadoop.util
port org.apache.spark.sql.SQLContext
ject count {
def main(args: Array[String]) {
System.setProperty("hadoop.home.dir","C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Utils")
// initialise spark context
val conf = new SparkConf().setAppName("CountSpark").setMaster("local[2]").set("spark.executor.memory","8g")
val sc = new SparkContext(conf)
//val textFile = sc.textFile("C:\\Users\\shanmuk\\Desktop\\PB\\project\\phase 2\\Tweets.Json")
val sqlContext = new SQLContext(sc)
val tweetsfile = sqlContext.jsonFile("C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Tweets.Json")
tweetsfile.registerTempTable("querytable1")
val query = sqlContext.sql("select count(user.name) as FtoJ from querytable1 where user.name like 'F%' OR user.name like 'G%' OR user.name like 'H%' OR
query.show();
Run: count count
16/04/07 22:01:59 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
16/04/07 22:01:59 INFO Executor: Finished task 0.0 in stage 3.0 (TID 60). 1115 bytes result sent to driver
16/04/07 22:01:59 INFO TaskSetManager: Finished task 0.0 in stage 3.0 (TID 60) in 58 ms on localhost (1/1)
16/04/07 22:01:59 INFO TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool
16/04/07 22:01:59 INFO DAGScheduler: ResultStage 3 (show at count.scala:26) finished in 0.058 s
16/04/07 22:01:59 INFO DAGScheduler: Job 1 finished: show at count.scala:26, took 6.573246 s
+-----+
| FtoJ|
+-----+
|23398|
+-----+
Compilation completed successfully with 1 warning in 1s 521ms (a minute ago)
477:1 CRLF: UTF-8:
I'm Cortana. Ask me anything.
10:02 PM
4/7/2016


project - [C:\Users\shanmuk\Desktop\PB\project] - [project] - ...\src\main\scala\count.scala - IntelliJ IDEA 2016.1.1
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
project src main scala count.scala
count
Project
build.sbt
count.scala
project C:\Users\shanmuk\Desktop\PB\
.idea
project [project-build] sources root
project
target
build.properties
plugins.sbt
src
main
java
resources
scala
count
scala-2.11
test
target
build.sbt
External Libraries
import org.apache.spark.{SparkConf, SparkContext}
import org.apache.spark.SparkConf
import org.apache.spark.SparkContext
import org.apache.hadoop.util
import org.apache.spark.sql.SQLContext
object count {
def main(args: Array[String]) {
System.setProperty("hadoop.home.dir","C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Utils")
// initialise spark context
val conf = new SparkConf().setAppName("CountSpark").setMaster("local[2]").set("spark.executor.memory","8g")
val sc = new SparkContext(conf)
//val textFile = sc.textFile("C:\\Users\\shanmuk\\Desktop\\PB\\project\\phase 2\\Tweets.Json")
val sqlContext = new SQLContext(sc)
val tweetsfile = sqlContext.jsonFile("C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Tweets.Json")
tweetsfile.registerTempTable("querytable1")
val query = sqlContext.sql("select count(user.name) as KtoO from querytable1 where user.name like 'K%' OR user.name like 'L%' OR user.name like 'M
query.show();
Run: count count
16/04/07 22:03:52 INFO TaskSetManager: Finished task 0.0 in stage 3.0 (TID 60) in 49 ms on localhost (1/1)
16/04/07 22:03:52 INFO TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool
16/04/07 22:03:52 INFO DAGScheduler: ResultStage 3 (show at count.scala:25) finished in 0.050 s
16/04/07 22:03:52 INFO DAGScheduler: Job 1 finished: show at count.scala:25, took 6.793245 s
+-----+
| KtoO|
+-----+
|31322|
+-----+
16/04/07 22:03:52 INFO SparkContext: Invoking stop() from shutdown hook
16/04/07 22:03:52 INFO SparkUI: Stopped Spark web UI at http://10.151.4.136:4040
16/04/07 22:03:52 INFO DAGScheduler: Stopping DAGScheduler
Compilation completed successfully with 1 warning in 1s 565ms (a minute ago)
477:1 CRLF: UTF-8:
I'm Cortana. Ask me anything.
10:04 PM
4/7/2016

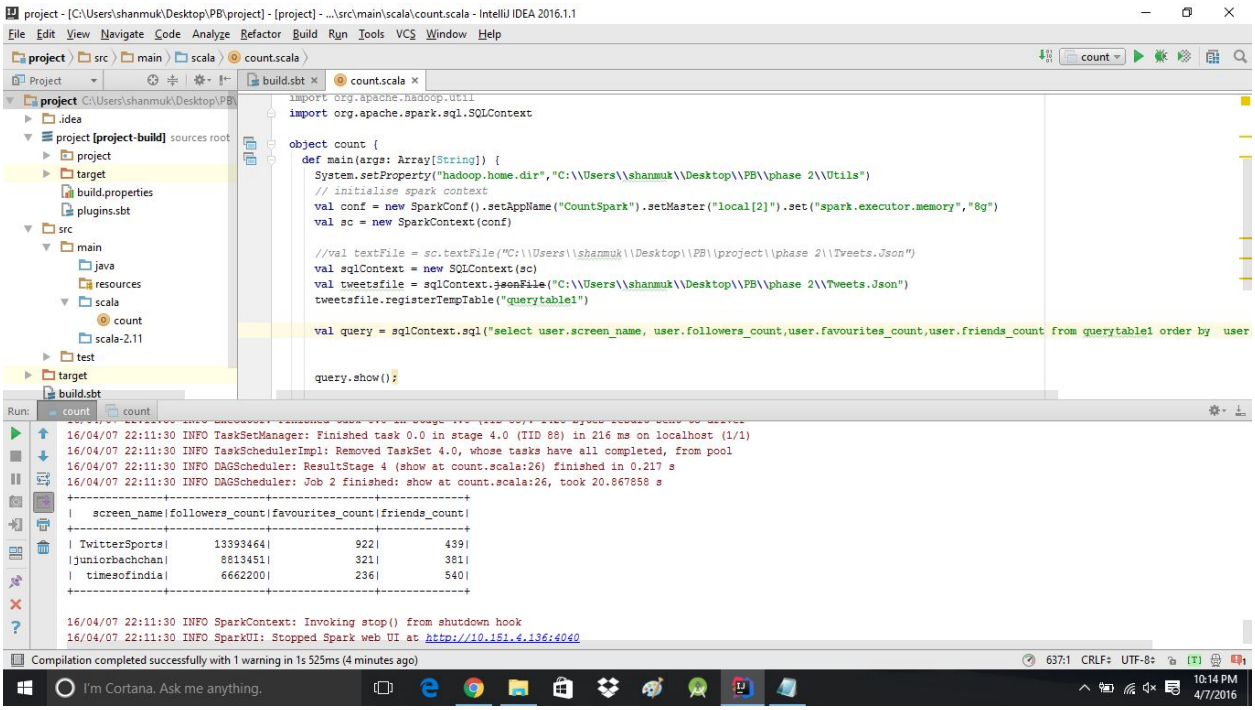

project - [C:\Users\shanmuk\Desktop\PB\project] - [project] - ...\src\main\scala\count.scala - IntelliJ IDEA 2016.1.1
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
project src main scala count.scala
count
Project
build.sbt
count.scala
project C:\Users\shanmuk\Desktop\PB\
.idea
project [project-build] sources root
project
target
build.properties
plugins.sbt
src
main
java
resources
scala
count
scala-2.11
test
target
build.sbt
External Libraries
import org.apache.spark.{SparkConf, SparkContext}
import org.apache.spark.SparkConf
import org.apache.spark.SparkContext
import org.apache.hadoop.util
import org.apache.spark.sql.SQLContext
object count {
def main(args: Array[String]) {
System.setProperty("hadoop.home.dir","C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Utils")
// initialise spark context
val conf = new SparkConf().setAppName("CountSpark").setMaster("local[2]").set("spark.executor.memory","8g")
val sc = new SparkContext(conf)
//val textFile = sc.textFile("C:\\Users\\shanmuk\\Desktop\\PB\\project\\phase 2\\Tweets.Json")
val sqlContext = new SQLContext(sc)
val tweetsfile = sqlContext.jsonFile("C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Tweets.Json")
tweetsfile.registerTempTable("querytable1")
val query = sqlContext.sql("select count(user.name) as PtoT from querytable1 where user.name like 'P%' OR user.name like 'Q%' OR user.name like 'R
query.show();
// val counts = textFile.flatMap(line => line.split(" "))
Run: count count
16/04/07 22:05:28 INFO TaskSetManager: Finished task 0.0 in stage 3.0 (TID 60) in 46 ms on localhost (1/1)
16/04/07 22:05:28 INFO TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool
16/04/07 22:05:28 INFO DAGScheduler: Job 1 finished: show at count.scala:24, took 6.723768 s
+-----+
| PtoT|
+-----+
|49466|
+-----+
16/04/07 22:05:28 INFO SparkContext: Invoking stop() from shutdown hook
16/04/07 22:05:28 INFO SparkUI: Stopped Spark web UI at http://10.151.4.136:4040
16/04/07 22:05:28 INFO DAGScheduler: Stopping DAGScheduler
Compilation completed successfully with 1 warning in 1s 857ms (a minute ago)
477:1 CRLF: UTF-8:
I'm Cortana. Ask me anything.
10:05 PM
4/7/2016


project - [C:\Users\shanmuk\Desktop\PB\project] - [project] - ...\src\main\scala\count.scala - IntelliJ IDEA 2016.1.1
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
project src main scala count.scala
count
Project
build.sbt
count.scala
project C:\Users\shanmuk\Desktop\PB\
.idea
project [project-build] sources root
project
target
build.properties
plugins.sbt
src
main
java
resources
scala
count
scala-2.11
test
target
build.sbt
External Libraries
rt org.apache.spark.{SparkConf, SparkContext}
rt org.apache.spark.SparkConf
rt org.apache.spark.SparkContext
rt org.apache.hadoop.util
rt org.apache.spark.sql.SQLContext
ct count {
f main(args: Array[String]) {
System.setProperty("hadoop.home.dir","C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Utils")
// initialise spark context
val conf = new SparkConf().setAppName("CountSpark").setMaster("local[2]").set("spark.executor.memory","8g")
val sc = new SparkContext(conf)
//val textFile = sc.textFile("C:\\Users\\shanmuk\\Desktop\\PB\\project\\phase 2\\Tweets.Json")
val sqlContext = new SQLContext(sc)
val tweetsfile = sqlContext.jsonFile("C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Tweets.Json")
tweetsfile.registerTempTable("querytable1")
val query = sqlContext.sql("select count(user.name) as UtoZ from querytable1 where user.name like 'U%' OR user.name like 'V%' OR user.name like 'W%' OR
query.show();
val counts = textFile.flatMap(line => line.split(" "))
Run: count count
16/04/07 22:07:24 INFO TaskSetManager: Finished task 0.0 in stage 3.0 (TID 60) in 75 ms on localhost (1/1)
16/04/07 22:07:24 INFO TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool
16/04/07 22:07:24 INFO DAGScheduler: ResultStage 3 (show at count.scala:24) finished in 0.076 s
16/04/07 22:07:24 INFO DAGScheduler: Job 1 finished: show at count.scala:24, took 7.712389 s
+-----+
| UtoZ|
+-----+
|12771|
+-----+
16/04/07 22:07:24 INFO SparkContext: Invoking stop() from shutdown hook
16/04/07 22:07:24 INFO SparkUI: Stopped Spark web UI at http://10.151.4.136:4040
16/04/07 22:07:24 INFO DAGScheduler: Stopping DAGScheduler
Compilation completed successfully with 1 warning in 1s 599ms (a minute ago)
477:1 CRLF: UTF-8:
I'm Cortana. Ask me anything.
10:07 PM
4/7/2016

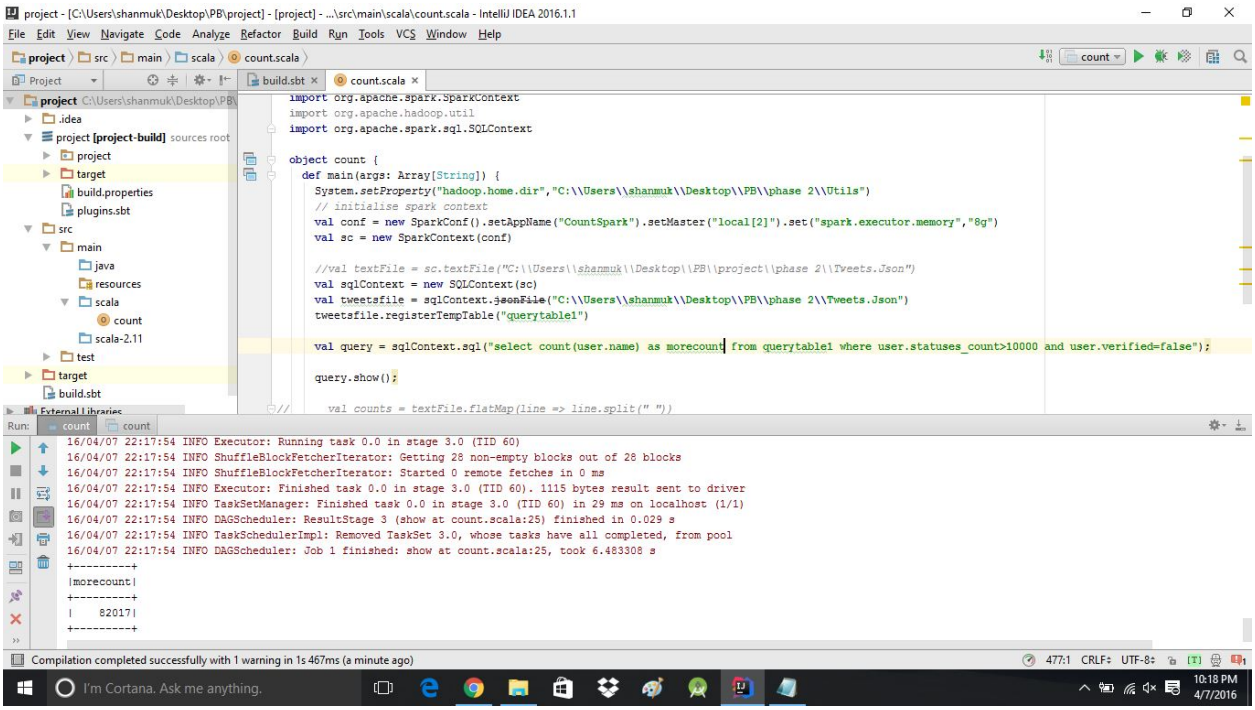
Query 2:

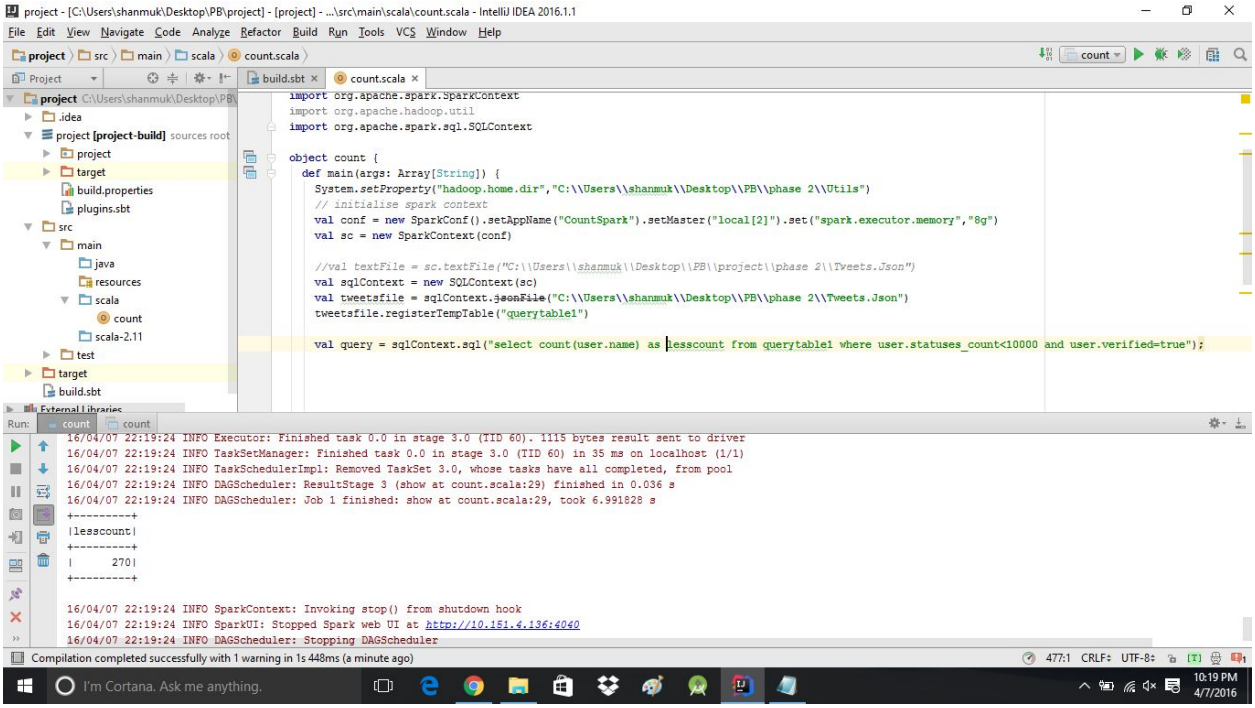

project - [C:\Users\shanmuk\Desktop\PB\project] - [project] - ...\src\main\scala\count.scala - IntelliJ IDEA 2016.1.1
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
project src main scala count.scala
count
Project
build.sbt
count.scala
project C:\Users\shanmuk\Desktop\PB\
.idea
project [project-build] sources root
project
target
build.properties
plugins.sbt
val sqlContext = new SQLContext(sc)
val tweetsfile = sqlContext.jsonFile("C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Tweets.Json")
tweetsfile.registerTempTable("querytable1")
val query = sqlContext.sql("select lang,avg(user.followers_count) as followers from querytable1 group by lang order by followers desc");
query.show();
// val counts = textFile.flatMap(line => line.split(" "))
Run: count count
+----+------------------+
|lang| followers|
+----+------------------+
no	57228.15
cs	33615.274131274135
lv	32920.94366197183
da	25748.648484848483
ne	22981.897435897437
sv	20404.092485549132
fa	19727.05
sl	19316.702898550724
ht	18676.467589220683
en	15069.59382253052
de	14775.385597082954
nl	12855.72372881356
mr	12101.1797752809
ur	12019.105496672419
fi	9370.777777777777
hi	9011.004956868805
in	7852.437864683225
tr	6953.9112903225805
pl	5631.964686998394
tl	5595.820480081716
+----+------------------+
16/04/07 22:09:18 INFO SparkContext: Invoking stop() from shutdown hook
Compilation completed successfully with 1 warning in 1s 526ms (a minute ago)
1650:26 CRLF: UTF-8:
I'm Cortana. Ask me anything.
10:09 PM
4/7/2016

Query3 :

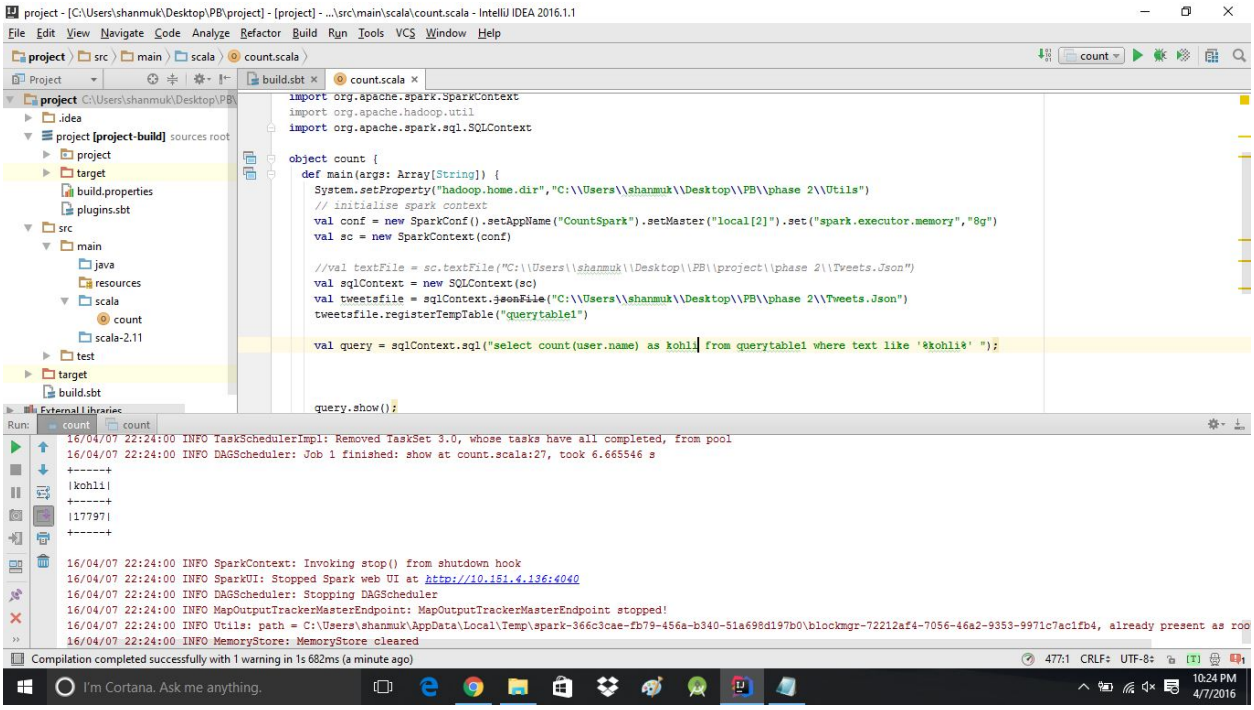

project - [C:\Users\shanmuk\Desktop\PB\project] - [project] - ...\src\main\scala\count.scala - IntelliJ IDEA 2016.1.1
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
project src main scala count.scala
count
Project
build.sbt
count.scala
project C:\Users\shanmuk\Desktop\PB\
.idea
project [project-build] sources root
project
target
build.properties
plugins.sbt
src
main
java
resources
scala
count
scala-2.11
test
target
build.sbt
import org.apache.hadoop.util
import org.apache.spark.sql.SQLContext
object count {
def main(args: Array[String]) {
System.setProperty("hadoop.home.dir","C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Utils")
// initialise spark context
val conf = new SparkConf().setAppName("CountSpark").setMaster("local[2]").set("spark.executor.memory","8g")
val sc = new SparkContext(conf)
//val textFile = sc.textFile("C:\\Users\\shanmuk\\Desktop\\PB\\project\\phase 2\\Tweets.Json")
val sqlContext = new SQLContext(sc)
val tweetsfile = sqlContext.jsonFile("C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Tweets.Json")
tweetsfile.registerTempTable("querytable1")
val query = sqlContext.sql("select user.screen_name, user.followers_count,user.favourites_count,user.friends_count from querytable1 order by user
query.show();
Run: count count
16/04/07 22:11:30 INFO TaskSetManager: Finished task 0.0 in stage 4.0 (TID 88) in 216 ms on localhost (1/1)
16/04/07 22:11:30 INFO TaskSchedulerImpl: Removed TaskSet 4.0, whose tasks have all completed, from pool
16/04/07 22:11:30 INFO DAGScheduler: ResultStage 4 (show at count.scala:26) finished in 0.217 s
16/04/07 22:11:30 INFO DAGScheduler: Job 2 finished: show at count.scala:26, took 20.867858 s
+--------------+---------------+----------------+-------------+
| screen_name|followers_count|favourites_count|friends_count|
+--------------+---------------+----------------+-------------+
| TwitterSports| 13393464| 922| 439|
|juniorbachchan| 8813451| 321| 381|
| timesofindia| 6662200| 236| 540|
+--------------+---------------+----------------+-------------+
16/04/07 22:11:30 INFO SparkContext: Invoking stop() from shutdown hook
16/04/07 22:11:30 INFO SparkUI: Stopped Spark web UI at http://10.151.4.136:4040
Compilation completed successfully with 1 warning in 1s 525ms (4 minutes ago)
637:1 CRLF: UTF-8:
I'm Cortana. Ask me anything.
10:14 PM
4/7/2016

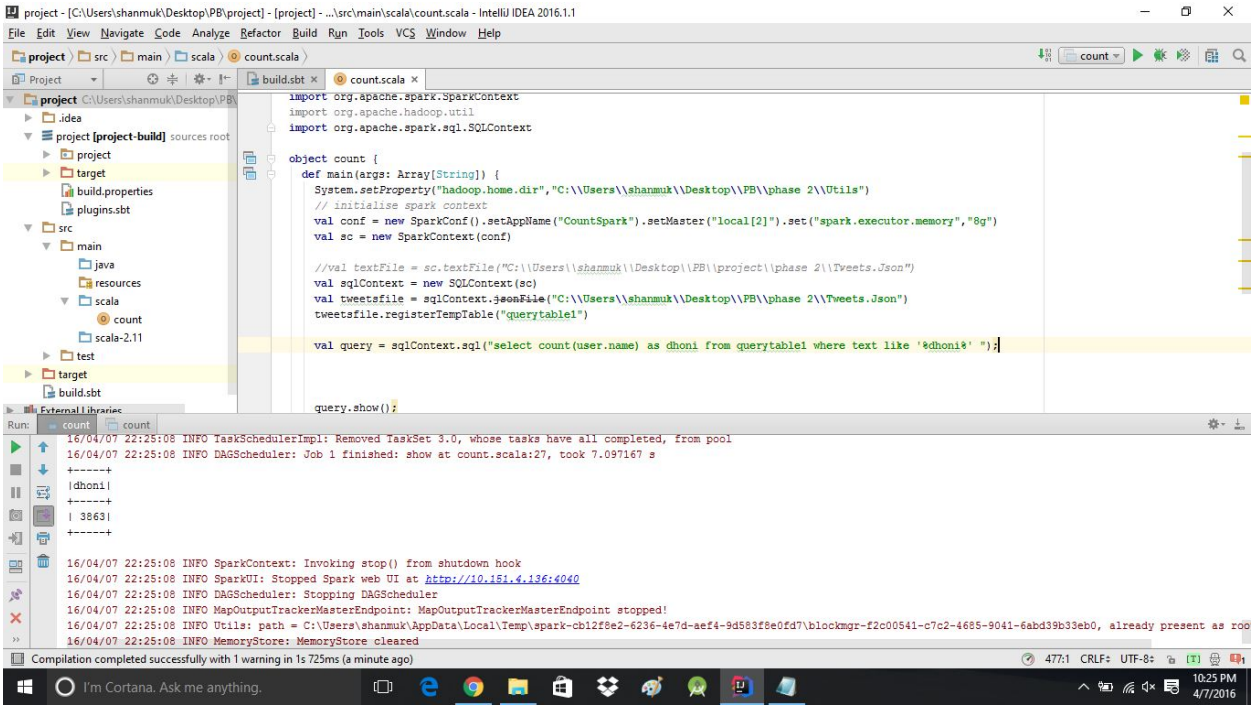
Query4:

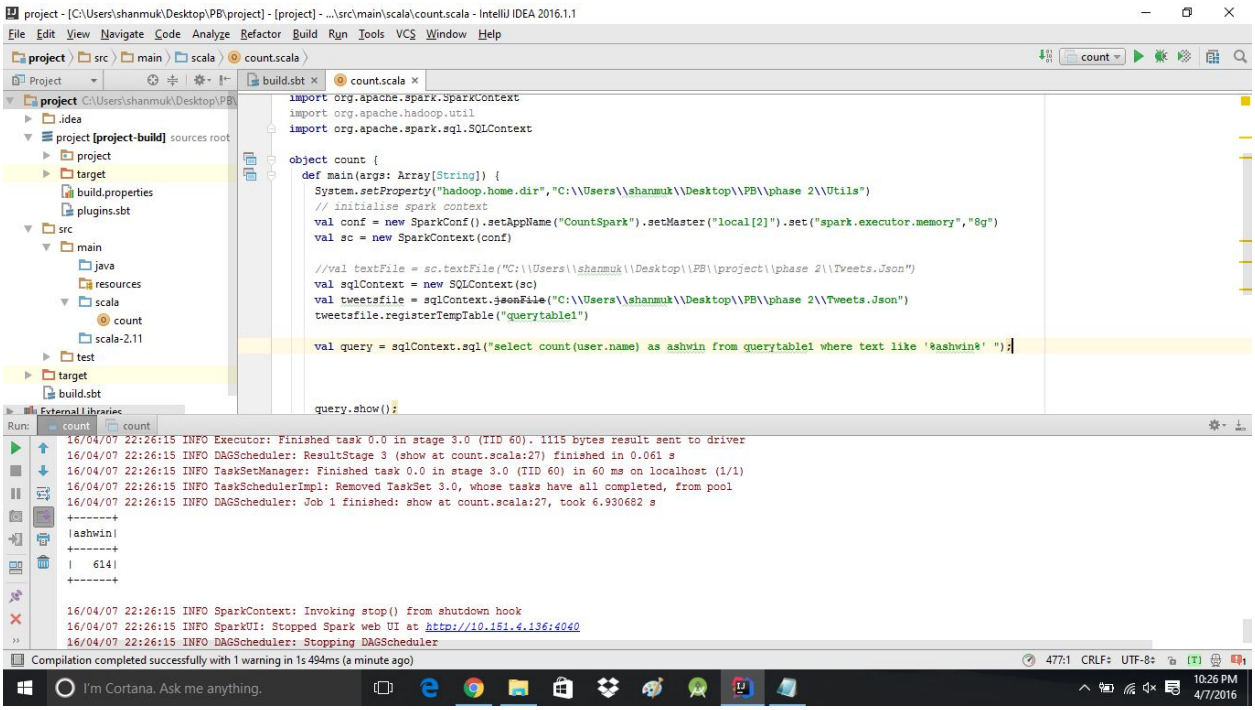

project - [C:\Users\shanmuk\Desktop\PB\project] - [project] - ...\src\main\scala\count.scala - IntelliJ IDEA 2016.1.1
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
project src main scala count.scala
count
Project
build.sbt
count.scala
import org.apache.spark.SparkContext
import org.apache.hadoop.util
import org.apache.spark.sql.SQLContext
object count {
def main(args: Array[String]) {
System.setProperty("hadoop.home.dir","C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Utils")
// initialise spark context
val conf = new SparkConf().setAppName("CountSpark").setMaster("local[2]").set("spark.executor.memory","8g")
val sc = new SparkContext(conf)
//val textFile = sc.textFile("C:\\Users\\shanmuk\\Desktop\\PB\\project\\phase 2\\Tweets.Json")
val sqlContext = new SQLContext(sc)
val tweetsfile = sqlContext.jsonFile("C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Tweets.Json")
tweetsfile.registerTempTable("querytable1")
val query = sqlContext.sql("select count(user.name) as morecount from querytable1 where user.statuses_count>10000 and user.verified=false");
query.show();
// val counts = textFile.flatMap(line => line.split(" "))
Run: count count
16/04/07 22:17:54 INFO Executor: Running task 0.0 in stage 3.0 (TID 60)
16/04/07 22:17:54 INFO ShuffleBlockFetcherIterator: Getting 28 non-empty blocks out of 28 blocks
16/04/07 22:17:54 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
16/04/07 22:17:54 INFO Executor: Finished task 0.0 in stage 3.0 (TID 60). 1115 bytes result sent to driver
16/04/07 22:17:54 INFO TaskSetManager: Finished task 0.0 in stage 3.0 (TID 60) in 29 ms on localhost (1/1)
16/04/07 22:17:54 INFO DAGScheduler: ResultStage 3 (show at count.scala:25) finished in 0.029 s
16/04/07 22:17:54 INFO TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool
16/04/07 22:17:54 INFO DAGScheduler: Job 1 finished: show at count.scala:25, took 6.483308 s
+---------+
|morecount|
+---------+
| 82017|
+---------+
Compilation completed successfully with 1 warning in 1s 467ms (a minute ago)
477:1 CRLF: UTF-8:
I'm Cortana. Ask me anything.
10:18 PM
4/7/2016


project - [C:\Users\shanmuk\Desktop\PB\project] - [project] - ...\src\main\scala\count.scala - IntelliJ IDEA 2016.1.1
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
project src main scala count.scala
count
Project
build.sbt
count.scala
import org.apache.spark.SparkContext
import org.apache.hadoop.util
import org.apache.spark.sql.SQLContext
object count {
def main(args: Array[String]) {
System.setProperty("hadoop.home.dir","C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Utils")
// initialise spark context
val conf = new SparkConf().setAppName("CountSpark").setMaster("local[2]").set("spark.executor.memory","8g")
val sc = new SparkContext(conf)
//val textFile = sc.textFile("C:\\Users\\shanmuk\\Desktop\\PB\\project\\phase 2\\Tweets.Json")
val sqlContext = new SQLContext(sc)
val tweetsfile = sqlContext.jsonFile("C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Tweets.Json")
tweetsfile.registerTempTable("querytable1")
val query = sqlContext.sql("select count(user.name) as lesscount from querytable1 where user.statuses_count<10000 and user.verified=true");
project C:\Users\shanmuk\Desktop\PB
.idea
project [project-build] sources root
project
target
build.properties
plugins.sbt
src
main
java
resources
scala
count
scala-2.11
test
target
build.sbt
External Libraries
Run: count count
16/04/07 22:19:24 INFO Executor: Finished task 0.0 in stage 3.0 (TID 60). 1115 bytes result sent to driver
16/04/07 22:19:24 INFO TaskSetManager: Finished task 0.0 in stage 3.0 (TID 60) in 35 ms on localhost (1/1)
16/04/07 22:19:24 INFO TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool
16/04/07 22:19:24 INFO DAGScheduler: ResultStage 3 (show at count.scala:29) finished in 0.036 s
16/04/07 22:19:24 INFO DAGScheduler: Job 1 finished: show at count.scala:29, took 6.991828 s
+---------+
|lesscount|
+---------+
| 270|
+---------+
16/04/07 22:19:24 INFO SparkContext: Invoking stop() from shutdown hook
16/04/07 22:19:24 INFO SparkUI: Stopped Spark web UI at http://10.151.4.136:4040
16/04/07 22:19:24 INFO DAGScheduler: Stopping DAGScheduler
Compilation completed successfully with 1 warning in 1s 448ms (a minute ago)
477:1 CRLF: UTF-8:
I'm Cortana. Ask me anything.
10:19 PM
4/7/2016

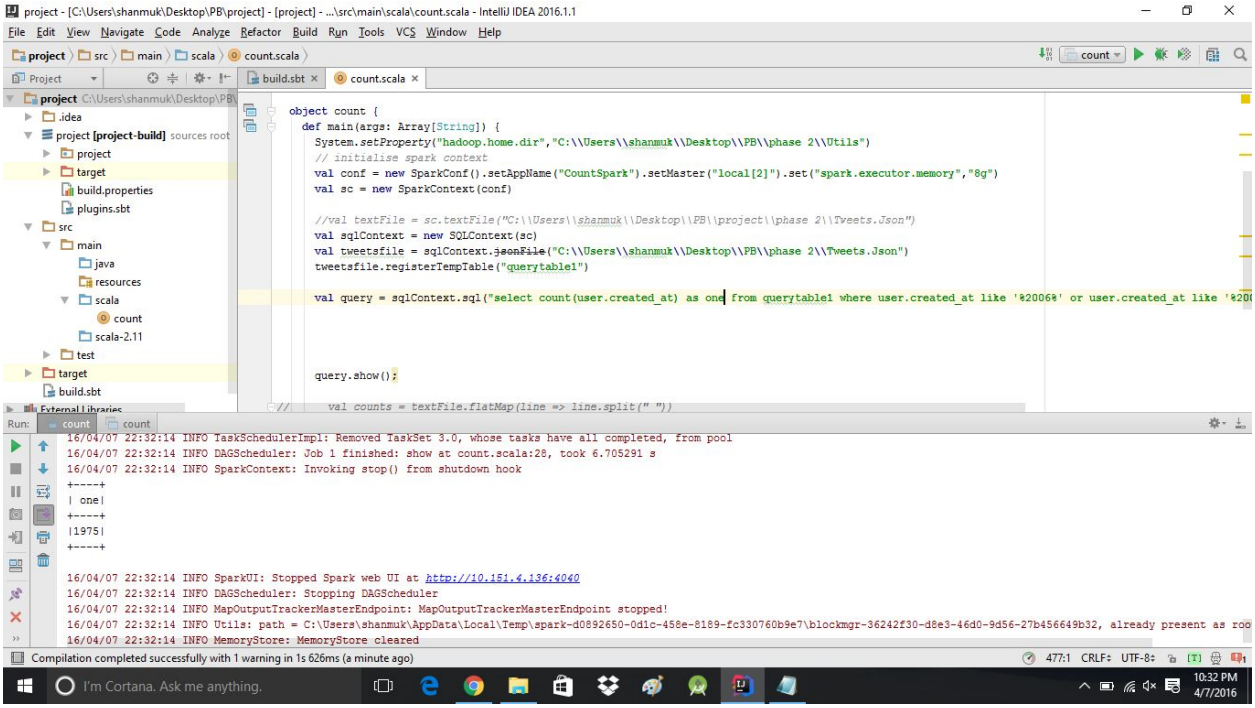
Query5:

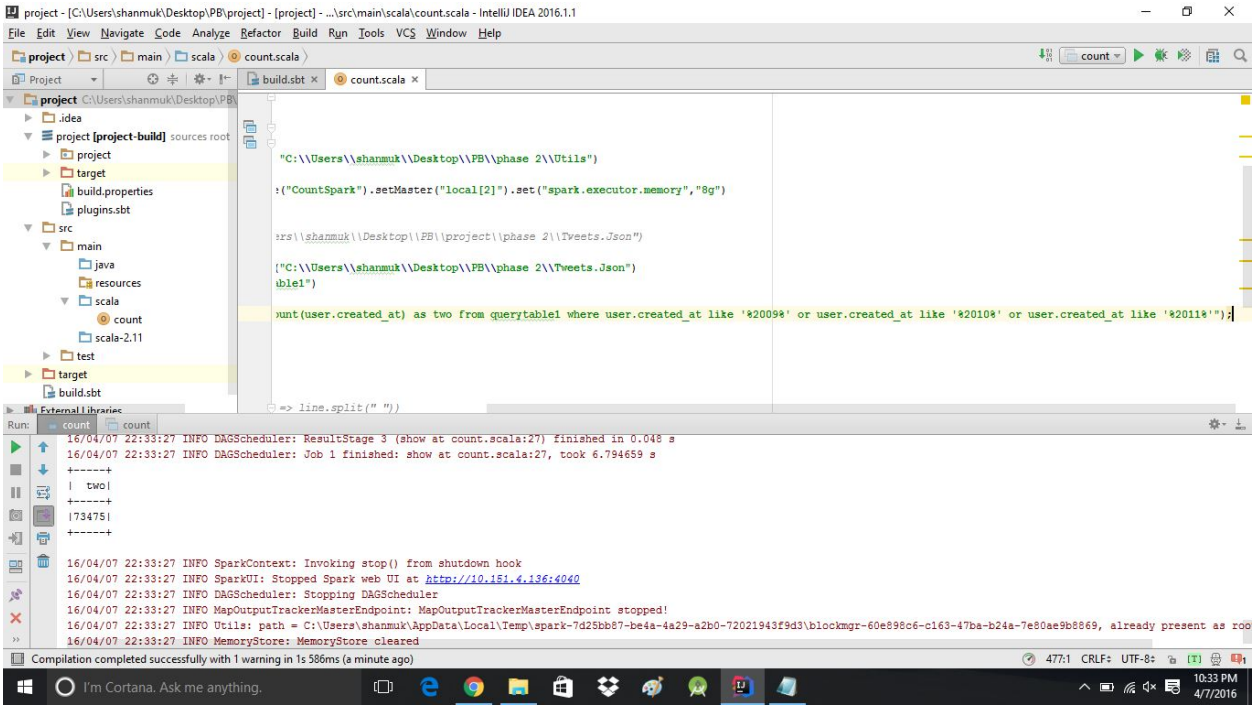

project - [C:\Users\shanmuk\Desktop\PB\project] - [project] - ...\src\main\scala\count.scala - IntelliJ IDEA 2016.1.1
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
build.sbt
count.scala
import org.apache.spark.SparkContext
import org.apache.hadoop.util
import org.apache.spark.sql.SQLContext
object count {
def main(args: Array[String]) {
System.setProperty("hadoop.home.dir","C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Utils")
// initialise spark context
val conf = new SparkConf().setAppName("CountSpark").setMaster("local[2]").set("spark.executor.memory","8g")
val sc = new SparkContext(conf)
//val textFile = sc.textFile("C:\\Users\\shanmuk\\Desktop\\PB\\project\\phase 2\\Tweets.Json")
val sqlContext = new SQLContext(sc)
val tweetsfile = sqlContext.jsonFile("C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Tweets.Json")
tweetsfile.registerTempTable("querytable1")
val query = sqlContext.sql("select count(user.name) as kohli from querytable1 where text like '%kohli%' ");
query.show();
16/04/07 22:24:00 INFO TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool
16/04/07 22:24:00 INFO DAGScheduler: Job 1 finished: show at count.scala:27, took 6.665546 s
+------+
|kohli|
+------+
|17797|
+------+
16/04/07 22:24:00 INFO SparkContext: Invoking stop() from shutdown hook
16/04/07 22:24:00 INFO SparkUI: Stopped Spark web UI at http://10.151.4.136:4040
16/04/07 22:24:00 INFO DAGScheduler: Stopping DAGScheduler
16/04/07 22:24:00 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
16/04/07 22:24:00 INFO MemoryStore: MemoryStore cleared
Compilation completed successfully with 1 warning in 1s 682ms (a minute ago)
I'm Cortana. Ask me anything.
10:24 PM
4/7/2016

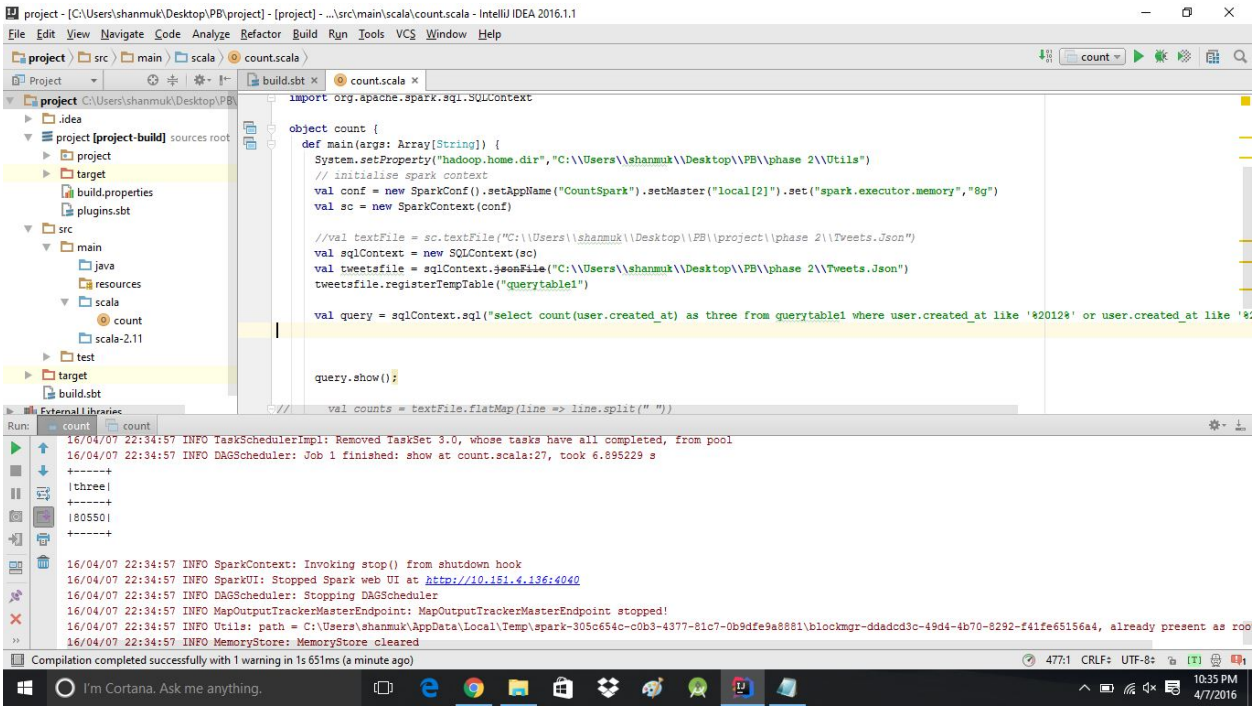

project - [C:\Users\shanmuk\Desktop\PB\project] - [project] - ...\src\main\scala\count.scala - IntelliJ IDEA 2016.1.1
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
build.sbt
count.scala
import org.apache.spark.SparkContext
import org.apache.hadoop.util
import org.apache.spark.sql.SQLContext
object count {
def main(args: Array[String]) {
System.setProperty("hadoop.home.dir","C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Utils")
// initialise spark context
val conf = new SparkConf().setAppName("CountSpark").setMaster("local[2]").set("spark.executor.memory","8g")
val sc = new SparkContext(conf)
//val textFile = sc.textFile("C:\\Users\\shanmuk\\Desktop\\PB\\project\\phase 2\\Tweets.Json")
val sqlContext = new SQLContext(sc)
val tweetsfile = sqlContext.jsonFile("C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Tweets.Json")
tweetsfile.registerTempTable("querytable1")
val query = sqlContext.sql("select count(user.name) as dhoni from querytable1 where text like '%dhoni%' ");
query.show();
16/04/07 22:25:08 INFO TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool
16/04/07 22:25:08 INFO DAGScheduler: Job 1 finished: show at count.scala:27, took 7.097167 s
+------+
|dhoni|
+------+
| 3863|
+------+
16/04/07 22:25:08 INFO SparkContext: Invoking stop() from shutdown hook
16/04/07 22:25:08 INFO SparkUI: Stopped Spark web UI at http://10.151.4.136:4040
16/04/07 22:25:08 INFO DAGScheduler: Stopping DAGScheduler
16/04/07 22:25:08 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
16/04/07 22:25:08 INFO MemoryStore: MemoryStore cleared
Compilation completed successfully with 1 warning in 1s 725ms (a minute ago)
I'm Cortana. Ask me anything.
10:25 PM
4/7/2016

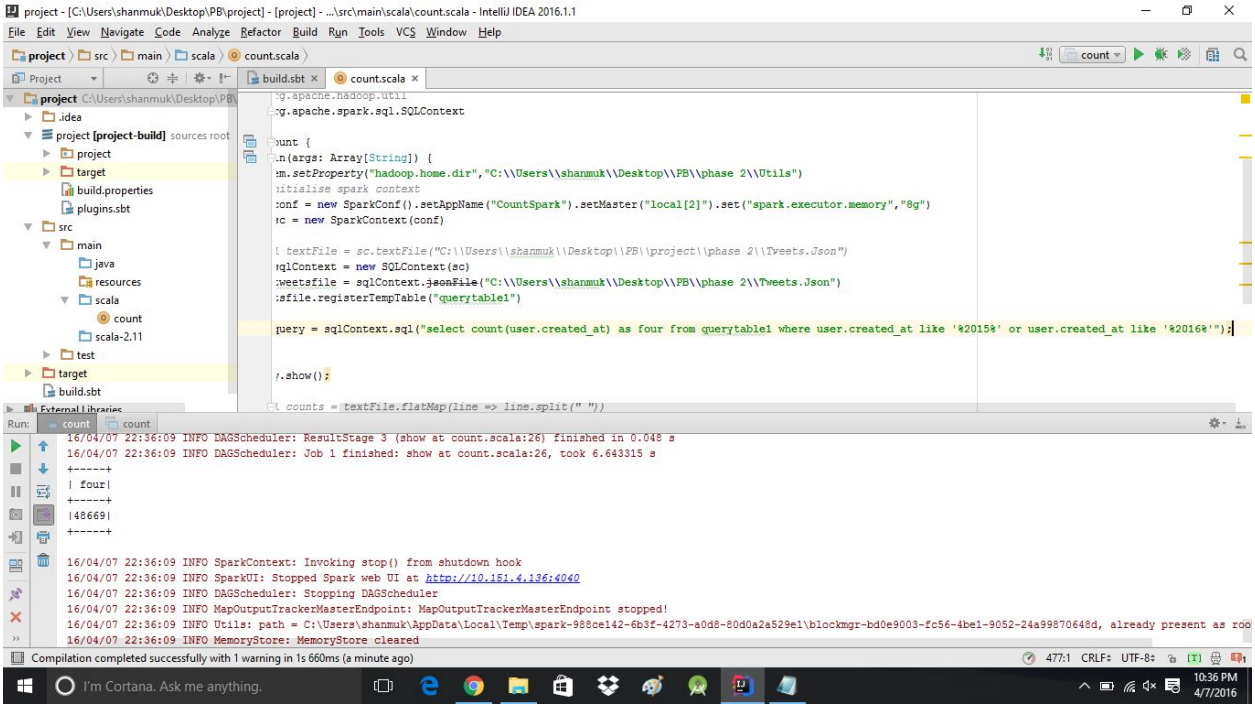

project - [C:\Users\shanmuk\Desktop\PB\project] - [project] - ...\src\main\scala\count.scala - IntelliJ IDEA 2016.1.1
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
project src main scala count.scala
count
Project
build.sbt
count.scala
project C:\Users\shanmuk\Desktop\PB\
.idea
project [project-build] sources root
project
target
build.properties
plugins.sbt
src
main
java
resources
scala
count
scala-2.11
test
target
build.sbt
External Libraries
import org.apache.spark.SparkContext
import org.apache.hadoop.util
import org.apache.spark.sql.SQLContext
object count {
def main(args: Array[String]) {
System.setProperty("hadoop.home.dir","C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Utils")
// initialise spark context
val conf = new SparkConf().setAppName("CountSpark").setMaster("local[2]").set("spark.executor.memory","8g")
val sc = new SparkContext(conf)
//val textFile = sc.textFile("C:\\Users\\shanmuk\\Desktop\\PB\\project\\phase 2\\Tweets.Json")
val sqlContext = new SQLContext(sc)
val tweetsfile = sqlContext.jsonFile("C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Tweets.Json")
tweetsfile.registerTempTable("querytable1")
val query = sqlContext.sql("select count(user.name) as ashwin from querytable1 where text like '%ashwin%' ");
query.show();
Run: count count
16/04/07 22:26:15 INFO Executor: Finished task 0.0 in stage 3.0 (TID 60). 1115 bytes result sent to driver
16/04/07 22:26:15 INFO DAGScheduler: ResultStage 3 (show at count.scala:27) finished in 0.061 s
16/04/07 22:26:15 INFO TaskSetManager: Finished task 0.0 in stage 3.0 (TID 60) in 60 ms on localhost (1/1)
16/04/07 22:26:15 INFO TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool
16/04/07 22:26:15 INFO DAGScheduler: Job 1 finished: show at count.scala:27, took 6.930682 s
+------+
|ashwin|
+------+
| 614|
+------+
16/04/07 22:26:15 INFO SparkContext: Invoking stop() from shutdown hook
16/04/07 22:26:15 INFO SparkUI: Stopped Spark web UI at http://10.151.4.136:4040
16/04/07 22:26:15 INFO DAGScheduler: Stopping DAGScheduler
Compilation completed successfully with 1 warning in 1s 494ms (a minute ago)
477:1 CRLF: UTF-8:
I'm Cortana. Ask me anything.
10:26 PM
4/7/2016

Query 6:

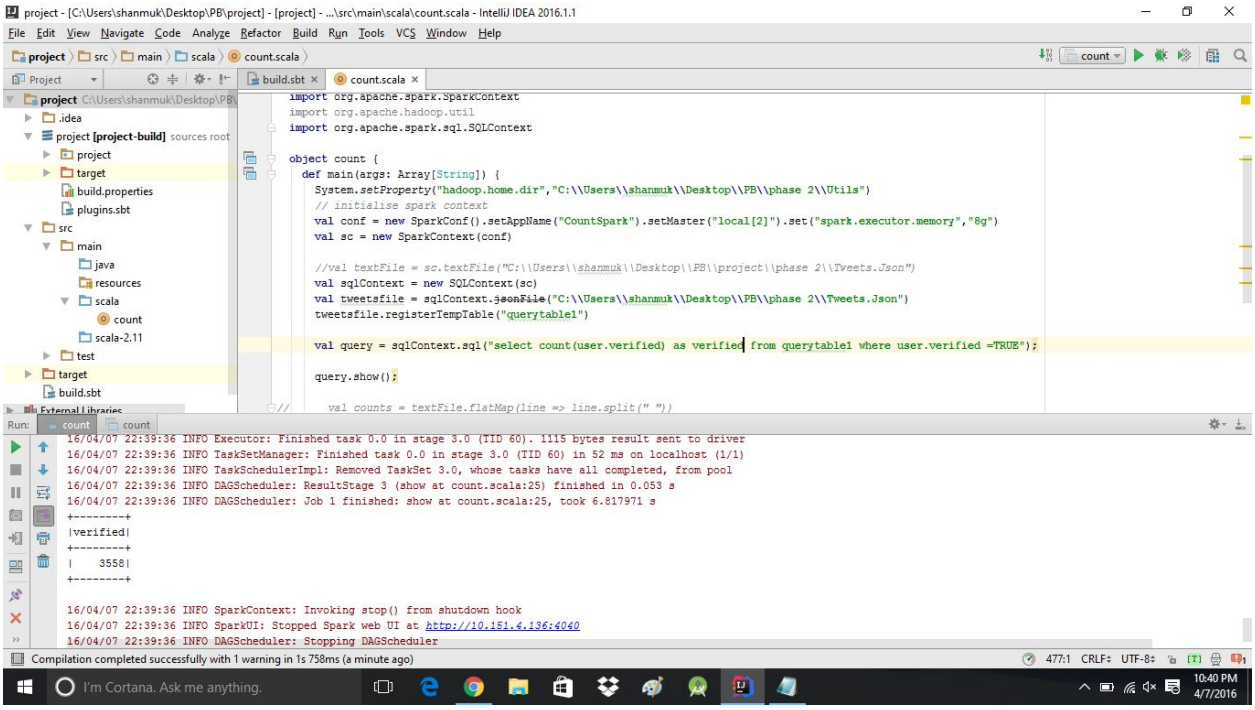

project - [C:\Users\shanmuk\Desktop\PB\project] - [project] - ...\src\main\scala\count.scala - IntelliJ IDEA 2016.1.1
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
build.sbt
count.scala
object count {
def main(args: Array[String]) {
System.setProperty("hadoop.home.dir","C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Utils")
// initialise spark context
val conf = new SparkConf().setAppName("CountSpark").setMaster("local[2]").set("spark.executor.memory","8g")
val sc = new SparkContext(conf)
//val textFile = sc.textFile("C:\\Users\\shanmuk\\Desktop\\PB\\project\\phase 2\\Tweets.Json")
val sqlContext = new SQLContext(sc)
val tweetsfile = sqlContext.jsonFile("C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Tweets.Json")
tweetsfile.registerTempTable("querytable1")
val query = sqlContext.sql("select count(user.created_at) as one from querytable1 where user.created_at like '%2006%' or user.created_at like '%20
query.show();
// val counts = textFile.flatMap(line => line.split(" "))
Run: count count
16/04/07 22:32:14 INFO TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool
16/04/07 22:32:14 INFO DAGScheduler: Job 1 finished: show at count.scala:28, took 6.705291 s
16/04/07 22:32:14 INFO SparkContext: Invoking stop() from shutdown hook
+----+
| one|
+----+
|1975|
+----+
16/04/07 22:32:14 INFO SparkUI: Stopped Spark web UI at http://10.151.4.136:4040
16/04/07 22:32:14 INFO DAGScheduler: Stopping DAGScheduler
16/04/07 22:32:14 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
16/04/07 22:32:14 INFO Utils: path = C:\Users\shanmuk\AppData\Local\Temp\spark-d0892650-0d1c-458e-8189-fc330760b9e7\blockmgr-36242f30-d8e3-46d0-9d56-27b456649b32, already present as roo
16/04/07 22:32:14 INFO MemoryStore: MemoryStore cleared
Compilation completed successfully with 1 warning in 1s 626ms (a minute ago)
I'm Cortana. Ask me anything.
10:32 PM
4/7/2016

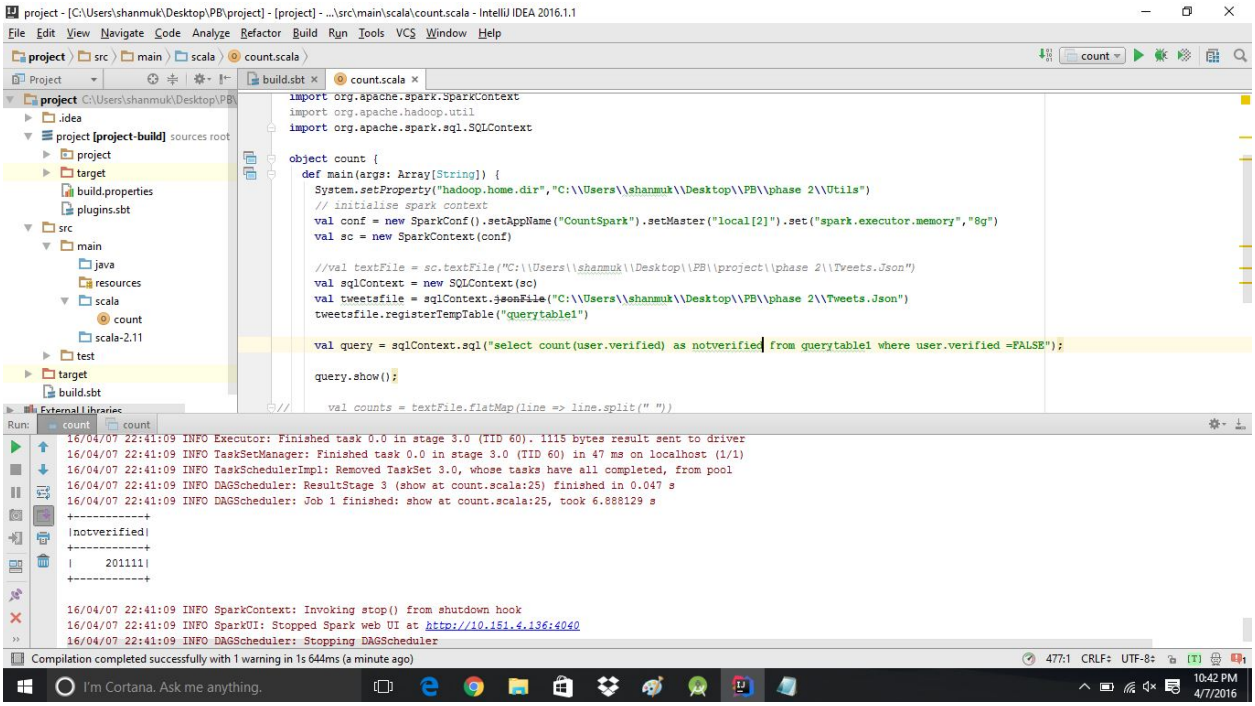

project - [C:\Users\shanmuk\Desktop\PB\project] - [project] - ...\src\main\scala\count.scala - IntelliJ IDEA 2016.1.1
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
project src main scala count.scala
count
Project
build.sbt
count.scala
project C:\Users\shanmuk\Desktop\PB\
.idea
project [project-build] sources root
project
target
build.properties
plugins.sbt
src
main
java
resources
scala
count
scala-2.11
test
target
build.sbt
External Libraries
"C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Utils")
("CountSpark").setMaster("local[2]").set("spark.executor.memory","8g")
rs\\shanmuk\\Desktop\\PB\\project\\phase 2\\Tweets.Json")
("C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Tweets.Json")
able1")
ount(user.created_at) as two from querytable1 where user.created_at like '%2009%' or user.created_at like '%2010%' or user.created_at like '%2011%'");
=> line.split(" "))
Run: count count
16/04/07 22:33:27 INFO DAGScheduler: ResultStage 3 (show at count.scala:27) finished in 0.048 s
16/04/07 22:33:27 INFO DAGScheduler: Job 1 finished: show at count.scala:27, took 6.794659 s
+-----+
| two|
+-----+
|73475|
+-----+
16/04/07 22:33:27 INFO SparkContext: Invoking stop() from shutdown hook
16/04/07 22:33:27 INFO SparkUI: Stopped Spark web UI at http://10.151.4.136:4040
16/04/07 22:33:27 INFO DAGScheduler: Stopping DAGScheduler
16/04/07 22:33:27 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
16/04/07 22:33:27 INFO Utils: path = C:\Users\shanmuk\AppData\Local\Temp\spark-7d25bb87-be4a-4a29-a2b0-72021943f9d3\blockmgr-60e898c6-c163-47ba-b24a-7e80ae9b8869, already present as roo
16/04/07 22:33:27 INFO MemoryStore: MemoryStore cleared
Compilation completed successfully with 1 warning in 1s 586ms (a minute ago)
477:1 CRLF: UTF-8:
I'm Cortana. Ask me anything.
10:33 PM
4/7/2016


project - [C:\Users\shanmuk\Desktop\PB\project] - [project] - ...\src\main\scala\count.scala - IntelliJ IDEA 2016.1.1
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
project src main scala count.scala
count
Project
build.sbt
count.scala
import org.apache.spark.sql.SQLContext
object count {
def main(args: Array[String]) {
System.setProperty("hadoop.home.dir","C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Utils")
// initialise spark context
val conf = new SparkConf().setAppName("CountSpark").setMaster("local[2]").set("spark.executor.memory","8g")
val sc = new SparkContext(conf)
//val textFile = sc.textFile("C:\\Users\\shanmuk\\Desktop\\PB\\project\\phase 2\\Tweets.Json")
val sqlContext = new SQLContext(sc)
val tweetsfile = sqlContext.jsonFile("C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Tweets.Json")
tweetsfile.registerTempTable("querytable1")
val query = sqlContext.sql("select count(user.created_at) as three from querytable1 where user.created_at like '%2012%' or user.created_at like '%
query.show();
// val counts = textFile.flatMap(line => line.split(" "))
project C:\Users\shanmuk\Desktop\PB\
.idea
project [project-build] sources root
project
target
build.properties
plugins.sbt
src
main
java
resources
scala
count
scala-2.11
test
target
build.sbt
External Libraries
Run: count count
16/04/07 22:34:57 INFO TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool
16/04/07 22:34:57 INFO DAGScheduler: Job 1 finished: show at count.scala:27, took 6.895229 s
+-----+
|three|
+-----+
|80550|
+-----+
16/04/07 22:34:57 INFO SparkContext: Invoking stop() from shutdown hook
16/04/07 22:34:57 INFO SparkUI: Stopped Spark web UI at http://10.151.4.136:4040
16/04/07 22:34:57 INFO DAGScheduler: Stopping DAGScheduler
16/04/07 22:34:57 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
16/04/07 22:34:57 INFO Utils: path = C:\Users\shanmuk\AppData\Local\Temp\spark-305c654c-c0b3-4377-81c7-0b9dfe9a8881\blockmgr-ddadcd3c-49d4-4b70-8292-f41fe65156a4, already present as roo
16/04/07 22:34:57 INFO MemoryStore: MemoryStore cleared
Compilation completed successfully with 1 warning in 1s 651ms (a minute ago)
477:1 CRLF: UTF-8:
I'm Cortana. Ask me anything.
10:35 PM
4/7/2016

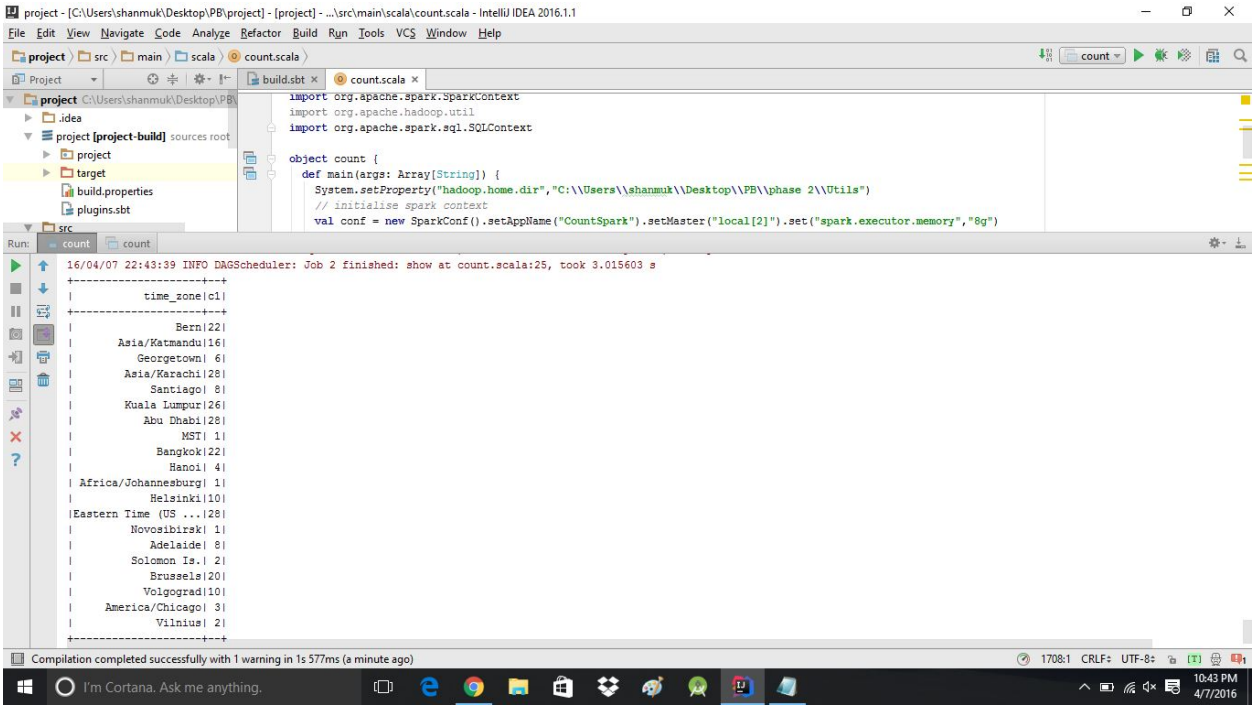

project - [C:\Users\shanmuk\Desktop\PB\project] - [project] - ...\src\main\scala\count.scala - IntelliJ IDEA 2016.1.1
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
project src main scala count.scala
count
Project
build.sbt
count.scala
project C:\Users\shanmuk\Desktop\PB
.idea
project [project-build] sources root
project
target
build.properties
plugins.sbt
src
main
java
resources
scala
count
scala-2.11
test
target
build.sbt
External Libraries
.g.apache.hadoop.util
.g.apache.spark.sql.SQLContext
unt {
.n(args: Array[String]) {
:m.setProperty("hadoop.home.dir","C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Utils")
itialise spark context
:onf = new SparkConf().setAppName("CountSpark").setMaster("local[2]").set("spark.executor.memory","8g")
:c = new SparkContext(conf)
textFile = sc.textFile("C:\\Users\\shanmuk\\Desktop\\PB\\project\\phase 2\\Tweets.Json")
:qlContext = new SQLContext(sc)
:weetsfile = sqlContext.jsonFile("C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Tweets.Json")
:sfile.registerTempTable("querytable1")
query = sqlContext.sql("select count(user.created_at) as four from querytable1 where user.created_at like '%2015%' or user.created_at like '%2016%'");
.show();
counts = textFile.flatMap(line => line.split(" "))
Run: count count
16/04/07 22:36:09 INFO DAGScheduler: ResultStage 3 (show at count.scala:26) finished in 0.048 s
16/04/07 22:36:09 INFO DAGScheduler: Job 1 finished: show at count.scala:26, took 6.643315 s
+-----+
| four|
+-----+
|48669|
+-----+
16/04/07 22:36:09 INFO SparkContext: Invoking stop() from shutdown hook
16/04/07 22:36:09 INFO SparkUI: Stopped Spark web UI at http://10.151.4.136:4040
16/04/07 22:36:09 INFO DAGScheduler: Stopping DAGScheduler
16/04/07 22:36:09 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
16/04/07 22:36:09 INFO Utils: path = C:\Users\shanmuk\AppData\Local\Temp\spark-988ce142-6b3f-4273-a0d8-80d0a2a529e1\blockmgr-bd0e9003-fc56-4be1-9052-24a99870648d, already present as roo
16/04/07 22:36:09 INFO MemoryStore: MemoryStore cleared
Compilation completed successfully with 1 warning in 1s 660ms (a minute ago)
477:1 CRLF: UTF-8:
I'm Cortana. Ask me anything.
10:36 PM
4/7/2016

Query 7:


project - [C:\Users\shanmuk\Desktop\PB\project] - [project] - ...\src\main\scala\count.scala - IntelliJ IDEA 2016.1.1
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
project > src > main > scala > count.scala
count
Project
build.sbt
count.scala
project C:\Users\shanmuk\Desktop\PB\
.idea
project [project-build] sources root
project
target
build.properties
plugins.sbt
src
main
java
resources
scala
count
scala-2.11
test
target
build.sbt
External Libraries
import org.apache.spark.SparkContext
import org.apache.hadoop.util
import org.apache.spark.sql.SQLContext
object count {
def main(args: Array[String]) {
System.setProperty("hadoop.home.dir","C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Utils")
// initialise spark context
val conf = new SparkConf().setAppName("CountSpark").setMaster("local[2]").set("spark.executor.memory","8g")
val sc = new SparkContext(conf)
//val textFile = sc.textFile("C:\\Users\\shanmuk\\Desktop\\PB\\project\\phase 2\\Tweets.Json")
val sqlContext = new SQLContext(sc)
val tweetsfile = sqlContext.jsonFile("C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Tweets.Json")
tweetsfile.registerTempTable("querytable1")
val query = sqlContext.sql("select count(user.verified) as verified from querytable1 where user.verified =TRUE");
query.show();
// val counts = textFile.flatMap(line => line.split(" "))
Run: count count
16/04/07 22:39:36 INFO Executor: Finished task 0.0 in stage 3.0 (TID 60). 1115 bytes result sent to driver
16/04/07 22:39:36 INFO TaskSetManager: Finished task 0.0 in stage 3.0 (TID 60) in 52 ms on localhost (1/1)
16/04/07 22:39:36 INFO TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool
16/04/07 22:39:36 INFO DAGScheduler: ResultStage 3 (show at count.scala:25) finished in 0.053 s
16/04/07 22:39:36 INFO DAGScheduler: Job 1 finished: show at count.scala:25, took 6.817971 s
+--------+
|verified|
+--------+
| 3558|
+--------+
16/04/07 22:39:36 INFO SparkContext: Invoking stop() from shutdown hook
16/04/07 22:39:36 INFO SparkUI: Stopped Spark web UI at http://10.151.4.136:4040
16/04/07 22:39:36 INFO DAGScheduler: Stopping DAGScheduler
Compilation completed successfully with 1 warning in 1s 758ms (a minute ago)
477:1 CRLF: UTF-8:
I'm Cortana. Ask me anything.
10:40 PM
4/7/2016

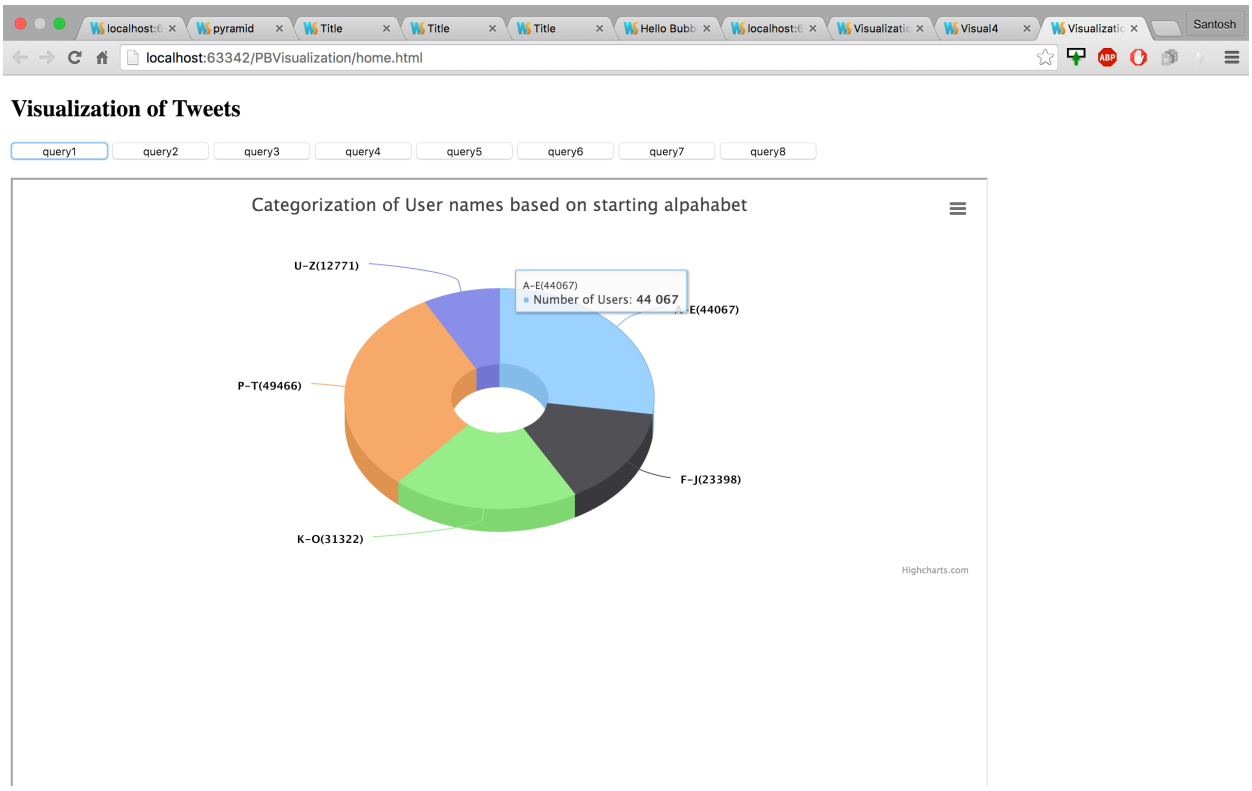

project - [C:\Users\shanmuk\Desktop\PB\project] - [project] - ...\src\main\scala\count.scala - IntelliJ IDEA 2016.1.1
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
project src main scala count.scala
count
Project
build.sbt
count.scala
project C:\Users\shanmuk\Desktop\PB\
.idea
project [project-build] sources root
project
target
build.properties
plugins.sbt
src
main
java
resources
scala
count
scala-2.11
test
target
build.sbt
External Libraries
import org.apache.spark.SparkContext
import org.apache.hadoop.util
import org.apache.spark.sql.SQLContext
object count {
def main(args: Array[String]) {
System.setProperty("hadoop.home.dir","C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Utils")
// initialise spark context
val conf = new SparkConf().setAppName("CountSpark").setMaster("local[2]").set("spark.executor.memory","8g")
val sc = new SparkContext(conf)
//val textFile = sc.textFile("C:\\Users\\shanmuk\\Desktop\\PB\\project\\phase 2\\Tweets.Json")
val sqlContext = new SQLContext(sc)
val tweetsfile = sqlContext.jsonFile("C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Tweets.Json")
tweetsfile.registerTempTable("querytable1")
val query = sqlContext.sql("select count(user.verified) as notverified from querytable1 where user.verified =FALSE");
query.show();
// val counts = textFile.flatMap(line => line.split(" "))
Run: count count
16/04/07 22:41:09 INFO Executor: Finished task 0.0 in stage 3.0 (TID 60). 1115 bytes result sent to driver
16/04/07 22:41:09 INFO TaskSetManager: Finished task 0.0 in stage 3.0 (TID 60) in 47 ms on localhost (1/1)
16/04/07 22:41:09 INFO TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool
16/04/07 22:41:09 INFO DAGScheduler: ResultStage 3 (show at count.scala:25) finished in 0.047 s
16/04/07 22:41:09 INFO DAGScheduler: Job 1 finished: show at count.scala:25, took 6.888129 s
+----------+
|notverified|
+----------+
| 201111|
+----------+
16/04/07 22:41:09 INFO SparkContext: Invoking stop() from shutdown hook
16/04/07 22:41:09 INFO SparkUI: Stopped Spark web UI at http://10.151.4.136:4040
16/04/07 22:41:09 INFO DAGScheduler: Stopping DAGScheduler
Compilation completed successfully with 1 warning in 1s 644ms (a minute ago)
477:1 CRLF: UTF-8:
I'm Cortana. Ask me anything.
10:42 PM
4/7/2016

Query 8:

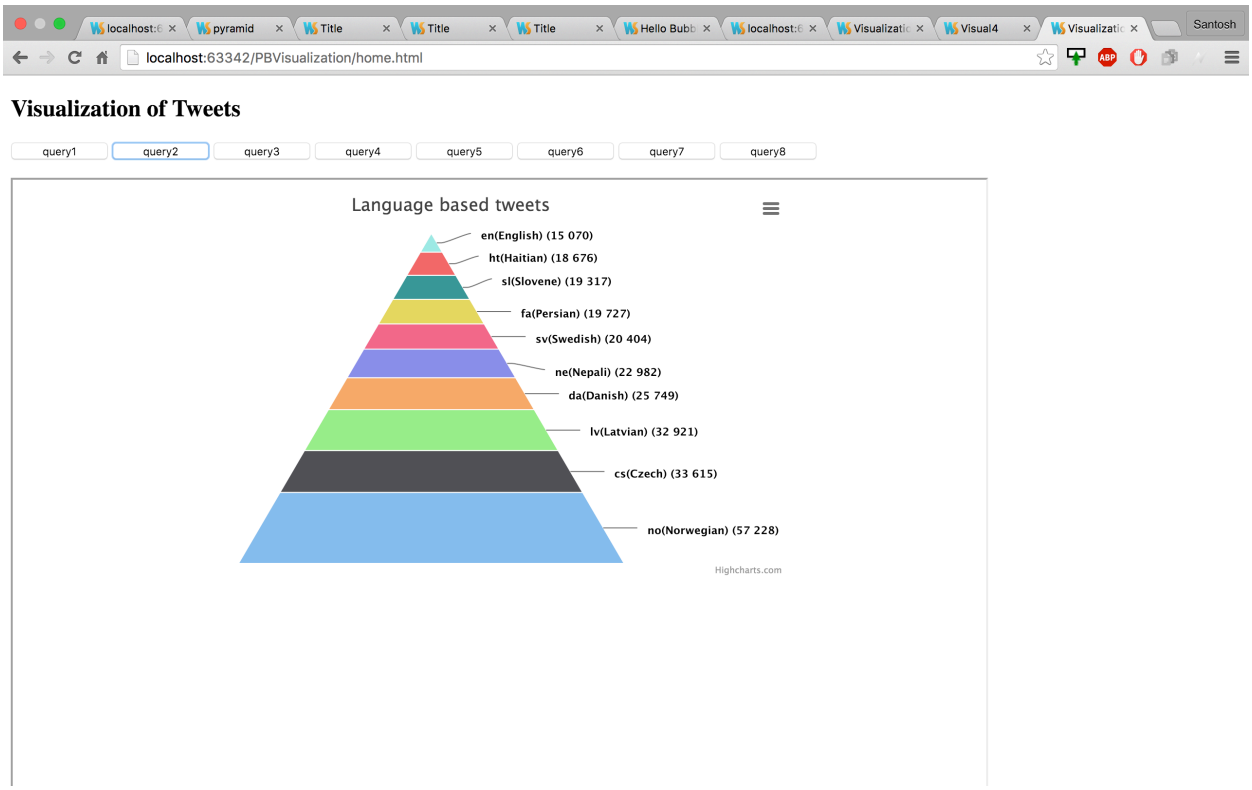

project - [C:\Users\shanmuk\Desktop\PB\project] - [project] - ...\src\main\scala\count.scala - IntelliJ IDEA 2016.1.1
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
project src main scala count.scala
count
Project
build.sbt
count.scala
project C:\Users\shanmuk\Desktop\PB\
.idea
project [project-build] sources root
project
target
build.properties
plugins.sbt
src
import org.apache.spark.SparkContext
import org.apache.hadoop.util
import org.apache.spark.sql.SQLContext
object count {
def main(args: Array[String]) {
System.setProperty("hadoop.home.dir","C:\\Users\\shanmuk\\Desktop\\PB\\phase 2\\Utils")
// initialise spark context
val conf = new SparkConf().setAppName("CountSpark").setMaster("local[2]").set("spark.executor.memory","8g")
Run:
count
count
16/04/07 22:43:39 INFO DAGScheduler: Job 2 finished: show at count.scala:25, took 3.015603 s
+--------------------+--+
| time_zone|c1|
+--------------------+--+
Bern	22
Asia/Katmandu	16
Georgetown	6
Asia/Karachi	28
Santiago	8
Kuala Lumpur	26
Abu Dhabi	28
MST	1
Bangkok	22
Hanoi	4
Africa/Johannesburg	1
Helsinki	10
Eastern Time (US ...	28
Novosibirsk	1
Adelaide	8
Solomon Is.	2
Brussels	20
Volgograd	10
America/Chicago	3
Vilnius	2
+--------------------+--+
Compilation completed successfully with 1 warning in 1s 577ms (a minute ago)
1708:1 CRLF: UTF-8:
I'm Cortana. Ask me anything.
10:43 PM
4/7/2016

Visualization:

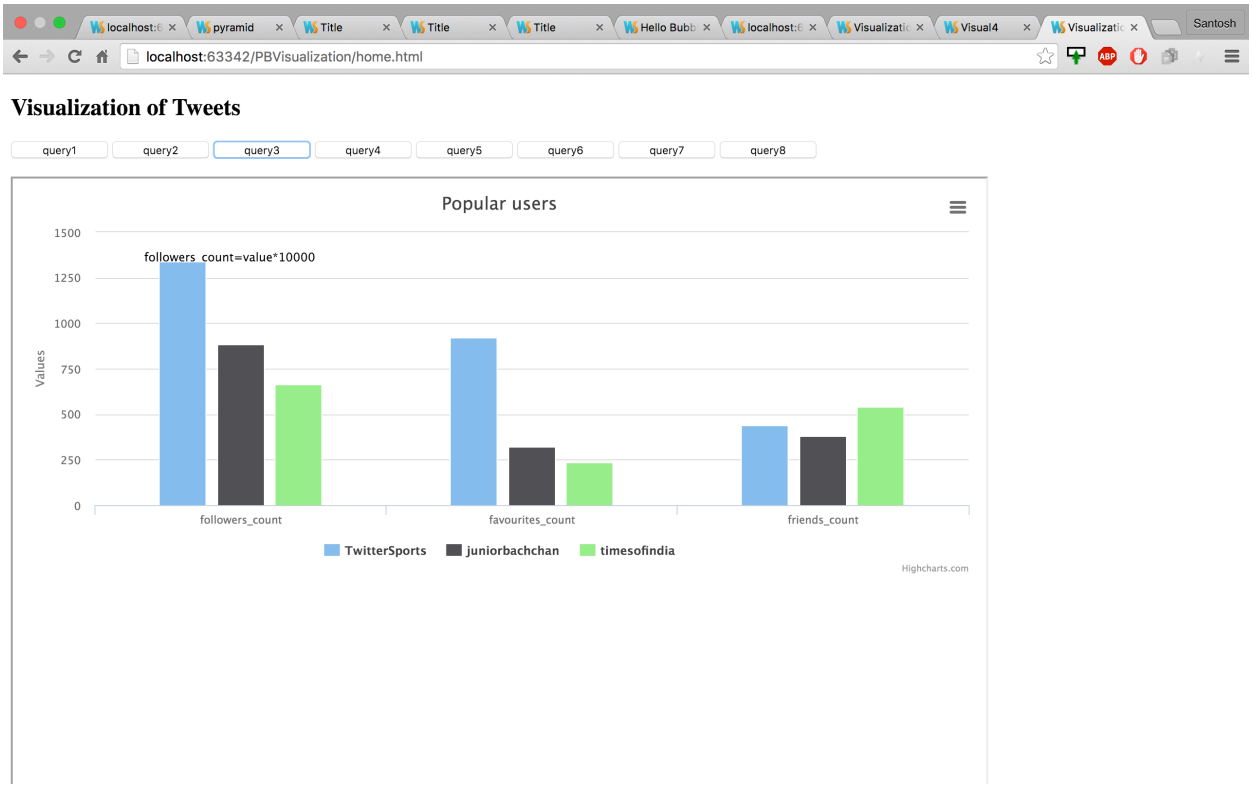
Query1



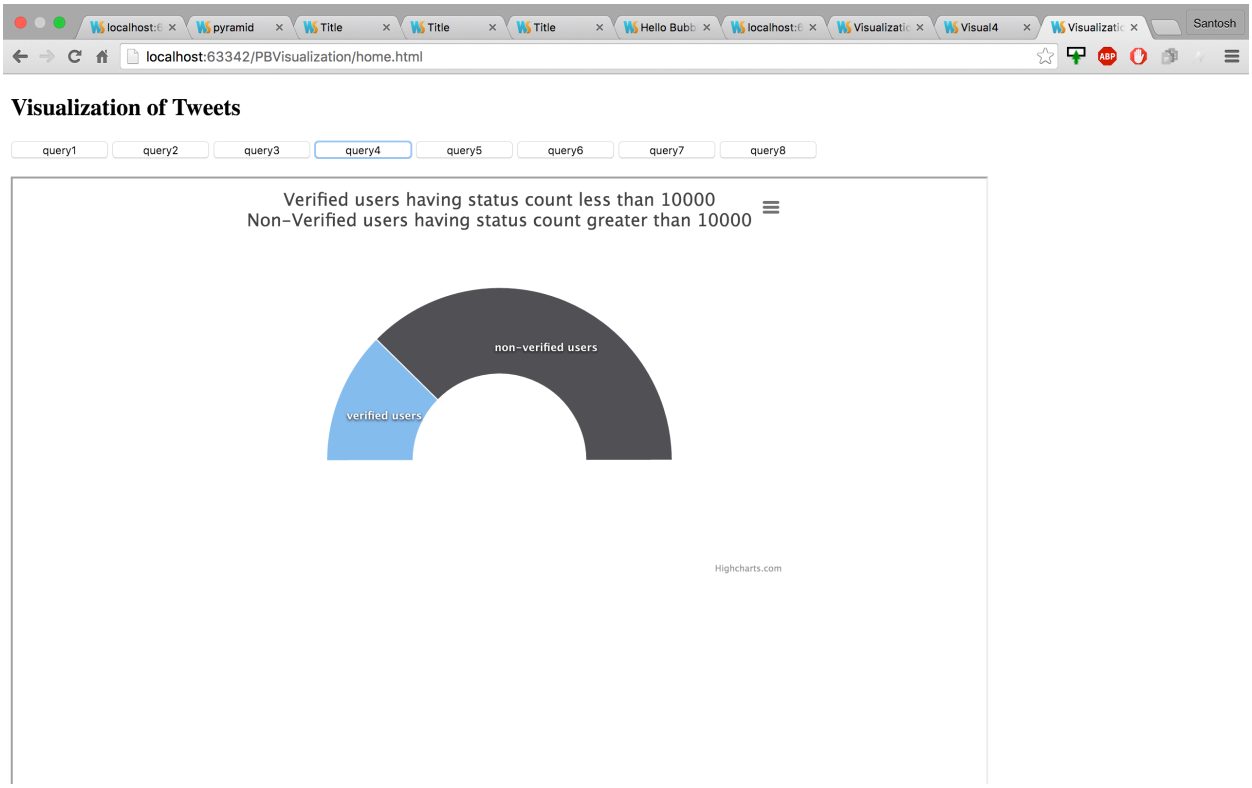
Query2:


localhost:6
pyramid
Title
Title
Title
Hello Bubb
localhost:6
Visualizatic
Visual4
Visualizatic
Santosh
localhost:63342/PBVisualization/home.html
Visualization of Tweets
query1
query2
query3
query4
query5
query6
query7
query8
Language based tweets
en(English) (15 070)
ht(Haitian) (18 676)
sl(Slovene) (19 317)
fa(Persian) (19 727)
sv(Swedish) (20 404)
ne(Nepali) (22 982)
da(Danish) (25 749)
lv(Latvian) (32 921)
cs(Czech) (33 615)
no(Norwegian) (57 228)
Highcharts.com

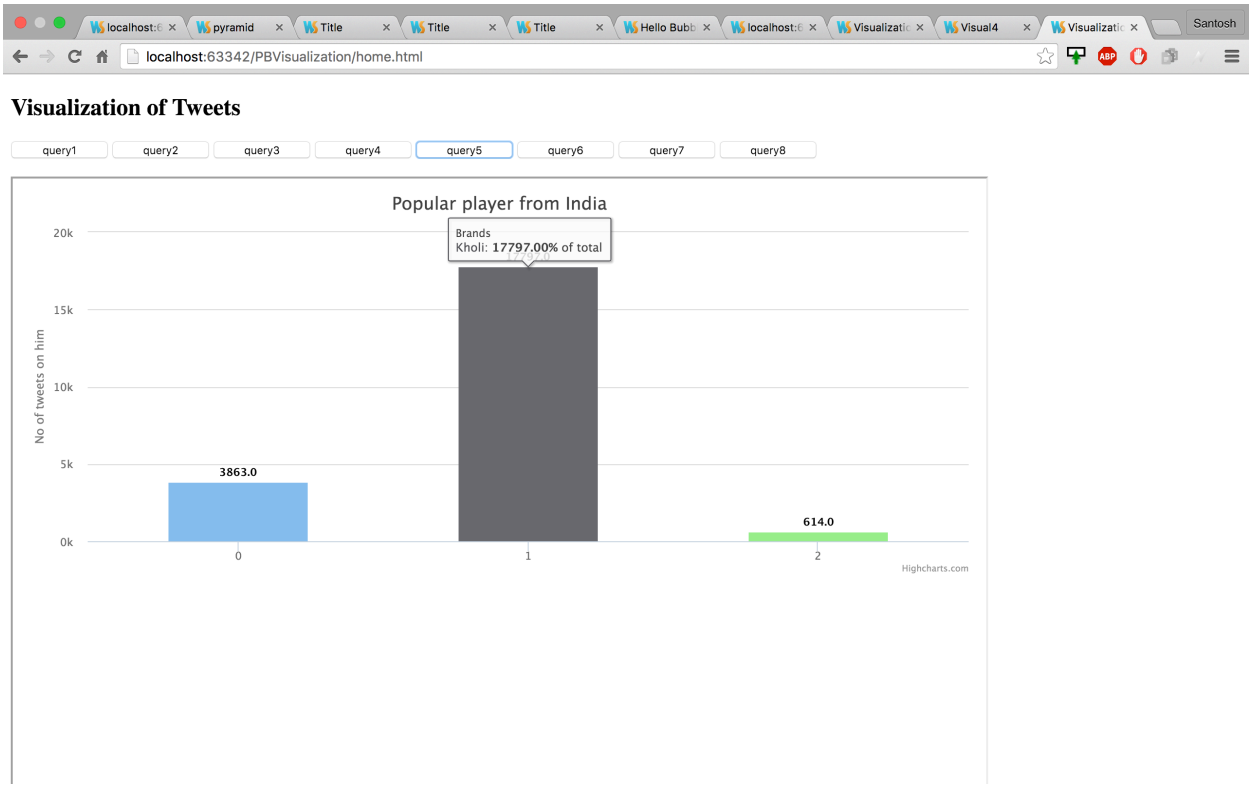
Query3


localhost:6
pyramid
Title
Title
Title
Hello Bubb
localhost:6
Visualizatic
Visual4
Visualizatic
Santosh
localhost:63342/PBVisualization/home.html
Visualization of Tweets
query1
query2
query3
query4
query5
query6
query7
query8
Popular users
followers_count=value*10000
1500
1250
1000
750
500
250
0
Values
followers_count
favourites_count
friends_count
TwitterSports
juniorbachchan
timesofindia
Highcharts.com

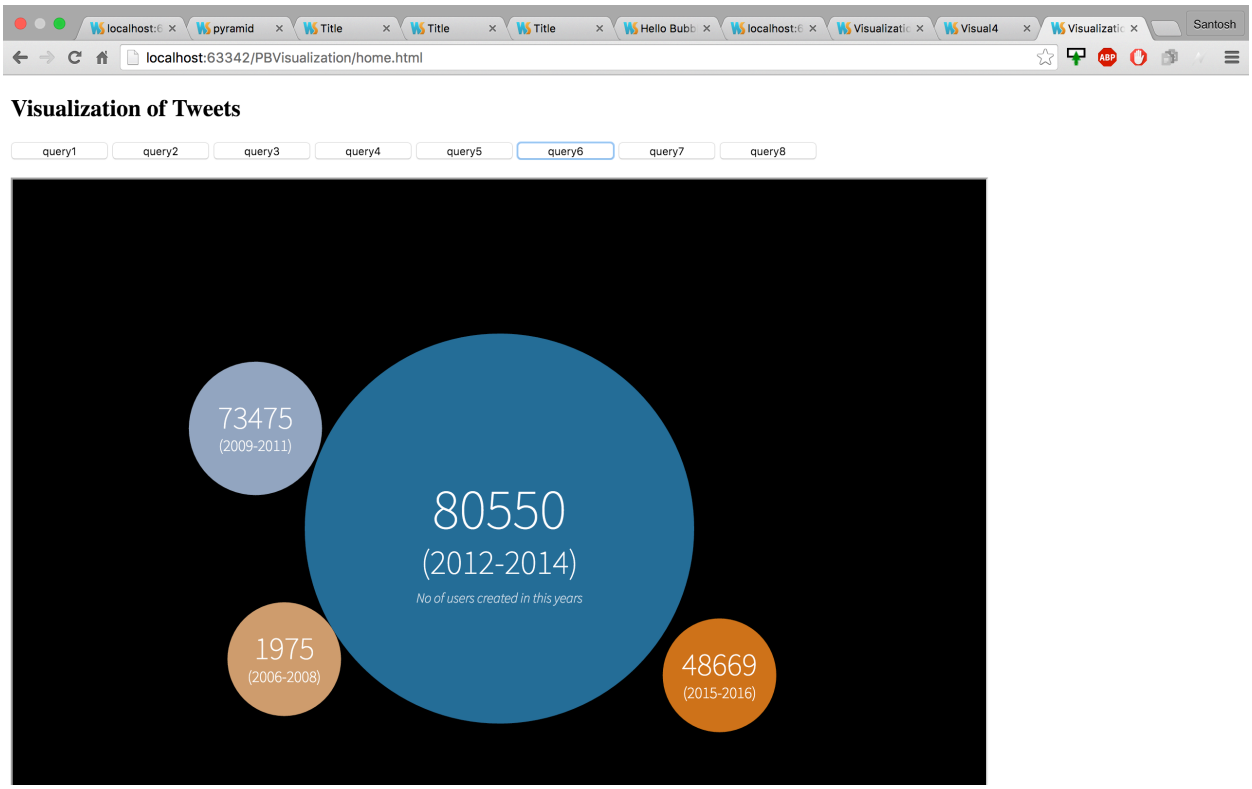
Query4


localhost:6
pyramid
Title
Title
Title
Hello Bubb
localhost:6
Visualizatic
Visual4
Visualizatic
Santosh
localhost:63342/PBVisualization/home.html
Visualization of Tweets
query1
query2
query3
query4
query5
query6
query7
query8
Verified users having status count less than 10000
Non-Verified users having status count greater than 10000
non-verified users
verified users
Highcharts.com

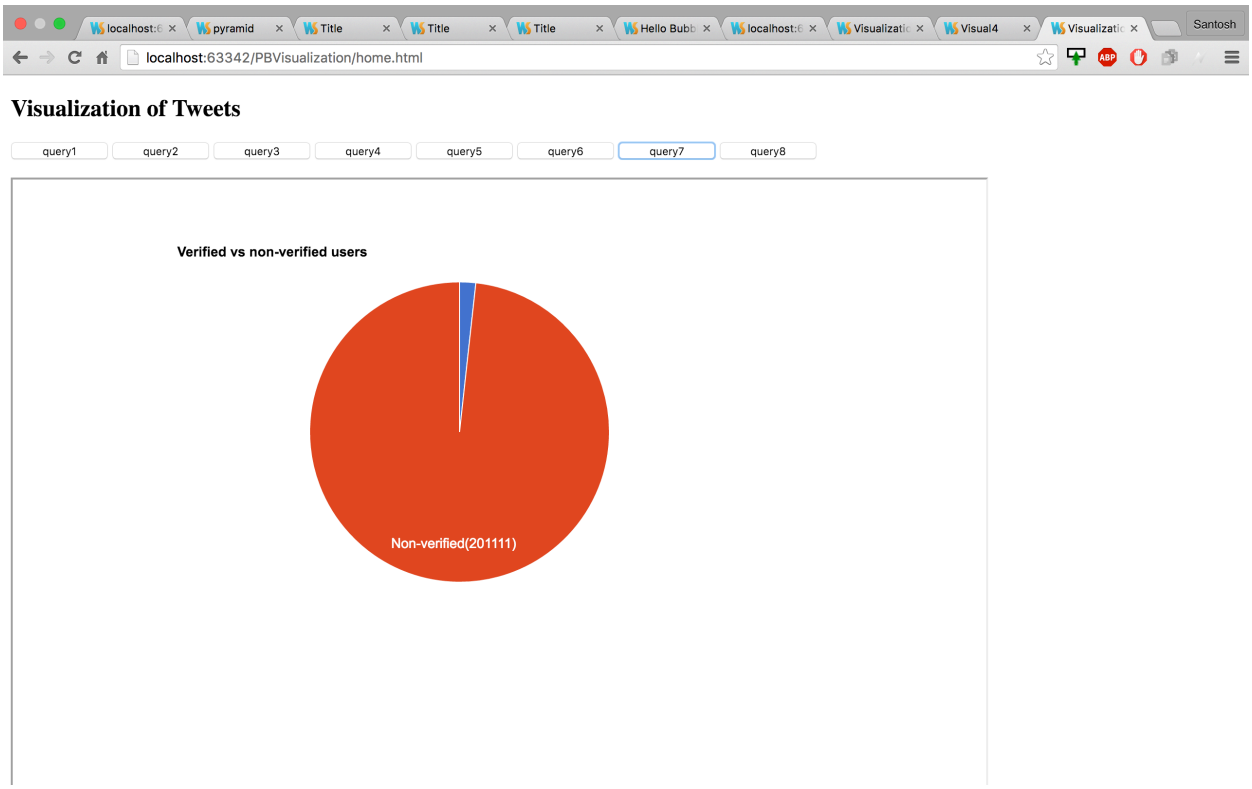
Query5


localhost:6
pyramid
Title
Title
Title
Hello Bubb
localhost:6
Visualizatic
Visual4
Visualizatic
Santosh
localhost:63342/PBVisualization/home.html
Visualization of Tweets
query1
query2
query3
query4
query5
query6
query7
query8
Popular player from India
Brands
Kholi: 17797.00% of total
No of tweets on him
20k
15k
10k
5k
0k
3863.0
614.0
0
1
2
Highcharts.com

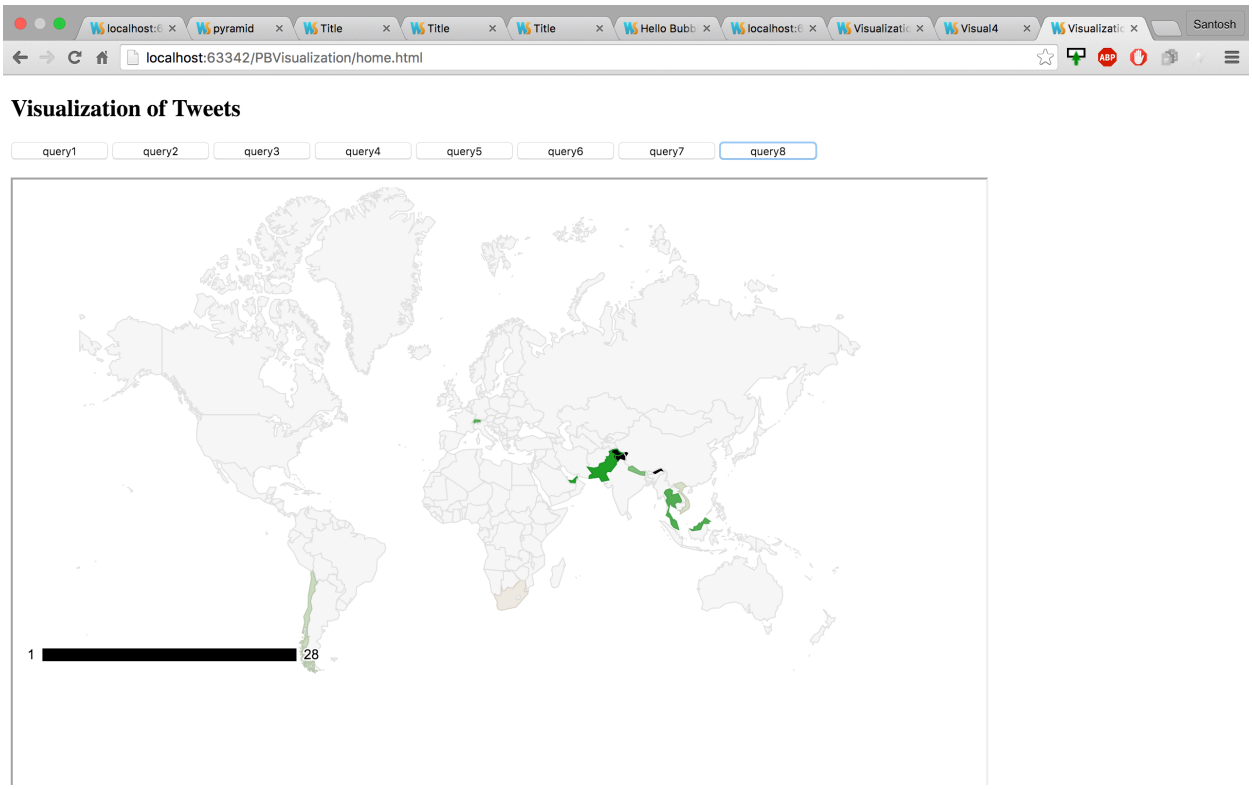
Query6


localhost:63342/PBVisualization/home.html
Visualization of Tweets
query1
query2
query3
query4
query5
query6
query7
query8
73475
(2009-2011)
80550
(2012-2014)
No of users created in this years
1975
(2006-2008)
48669
(2015-2016)

Query 7


localhost:63342/PBVisualization/home.html
Santosh
Visualization of Tweets
query1
query2
query3
query4
query5
query6
query7
query8
Verified vs non-verified users
Non-verified(201111)

Query 8



Source code github link:

https://github.com/sreeram66/PB.git