BIG DATA PROJECT 2

CIIC 8995/5995 5016 – Big Data Analytics Spring 2017

Project 2: Streaming for Twitter Analysis

By

Jessica Cotrina Revilla

University of Puerto Rico

Mayaguez Campus
2017

Consultas Hive

4.

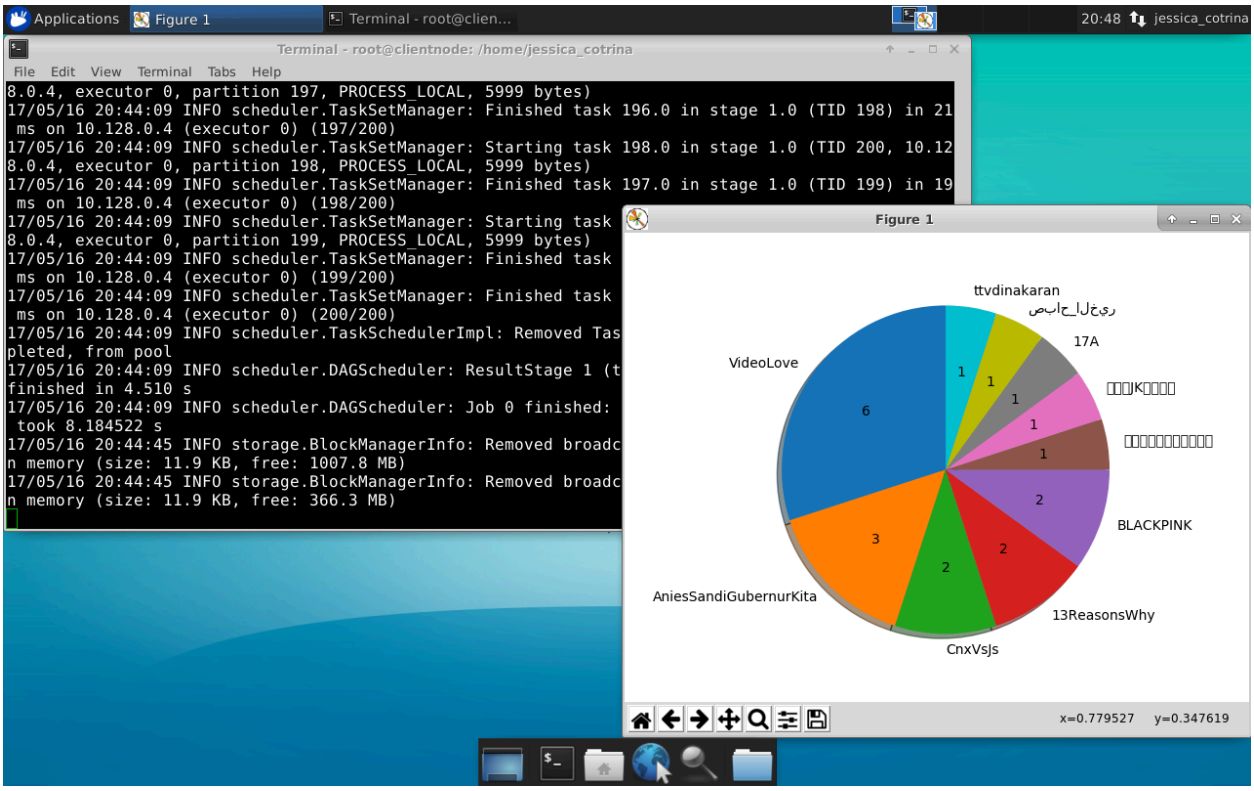
a) **Capture the top 10 trending hashtags**

Mostrando un ejemplo de una consulta de hive; de algunos datos registrados en la tabla.

```
+---------------------------+-----------------+---------------------
----+--+
|      hashtag1.hashtag      | hashtag1.total  |    hashtag1.timestamp
|
+---------------------------+-----------------+---------------------
----+--+
| 17A                        | 1               | 2017-04-18
03:19:30.531  |
| ذوق_لشخص_كلمه_قل           | 1               | 2017-04-18
03:19:30.531  |
| mermaid                    | 1               | 2017-04-18
03:19:30.531  |
| 66حياتك                    | 1               | 2017-04-18
03:18:11.074  |
| ColombiaNeeds방탄소년단         | 1               | 2017-04-18
03:18:11.074  |
| سيرته_نشر                  | 1               | 2017-04-18
03:18:11.074  |
| FelizCumpleGastonSardelli  | 2               | 2017-04-18
03:23:20.597  |
| للدعم_عشيرته_قروب          | 1               | 2017-04-18 03:23:20.597
|
| DemiLovato                 | 1               | 2017-04-18
03:23:20.597  |
| giveaway                   | 2               | 2017-04-18
03:22:10.635  |
| SF9                        | 2               | 2017-04-18
03:22:10.635  |
| CandyCelebs                | 2               | 2017-04-18
03:22:10.635  |
| MerchInBio                 | 1               | 2017-04-18
03:18:30.519  |
| FPJAPHatol                 | 2               | 2017-04-18
03:22:20.623  |
| Economía                   | 2               | 2017-04-18
03:22:20.623  |
| زوجيه_نصيحه_عطنا           | 2               | 2017-04-18 03:22:20.623
|
| トレクル                       | 2               | 2017-04-18
```

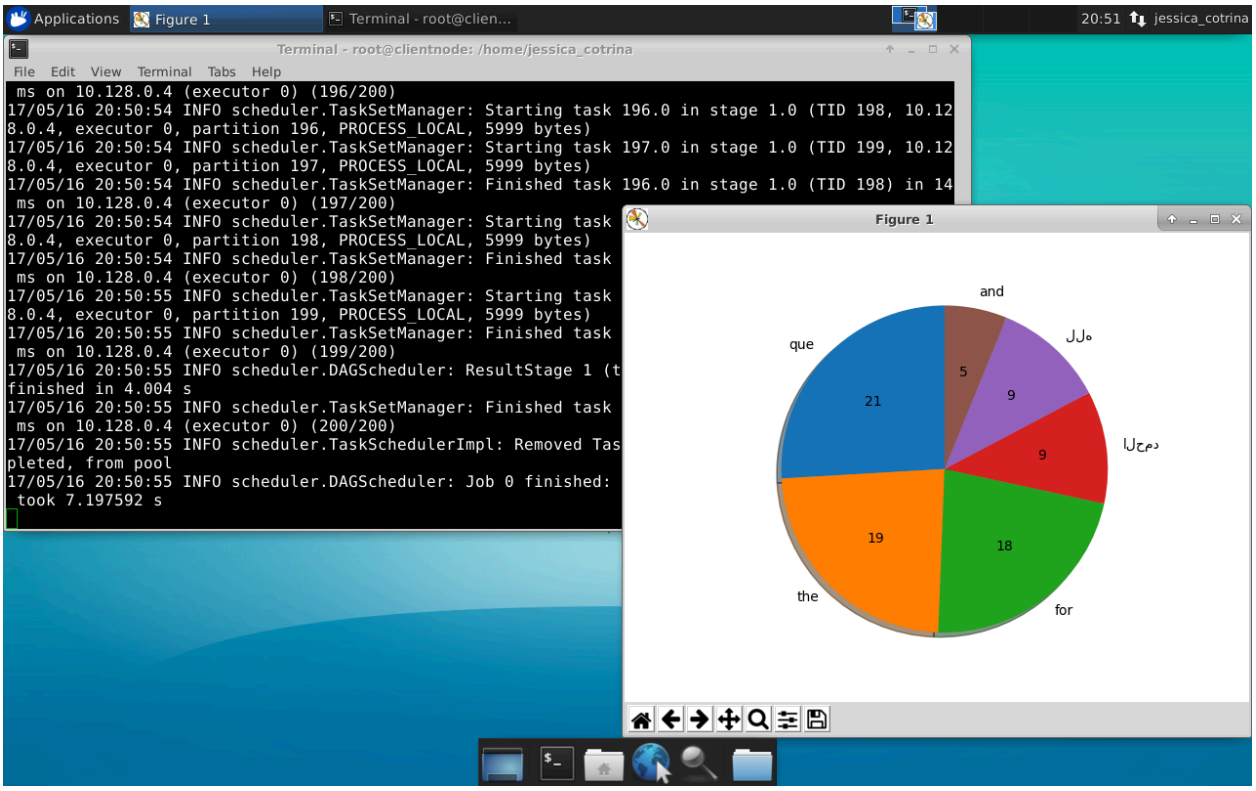
```
03:20:20.728  |
| حياتك66                    | 1               | 2017-04-18
03:20:20.728  |
| MARQUEE                    | 1               | 2017-04-18
03:20:20.728  |
| FPJAPHatol                 | 3               | 2017-04-18
03:21:10.671  |
| MayWard                    | 2               | 2017-04-18
03:21:10.671  |
| TrumpRussia                | 2               | 2017-04-18
03:21:10.671  |
| ニコ生JKミスコン                  | 1               | 2017-04-18
03:17:39.495  |
| 여성최음제구입방법                  | 1               | 2017-04-18
03:17:39.495  |
| FHOOTERS                   | 1               | 2017-04-18
03:17:39.495  |
| AniesSandiGubernurKita     | 3               | 2017-04-18
03:17:20.478  |
| BLACKPINK                  | 2               | 2017-04-18
03:17:20.478  |
| CnxVsJs                    | 2               | 2017-04-18
03:17:20.478  |
| زوجيه_نصيحه_عطنا           | 1               | 2017-04-18 03:23:30.347
|
| VideoLove                  | 1               | 2017-04-18
03:23:30.347  |
| ChopOn                     | 1               | 2017-04-18
03:23:30.347  |
| VideoLove                  | 2               | 2017-04-18
03:19:20.722  |
```

Gráfica usando Pandas



b) Capture the top 10 trending keywords

```
0: jdbc:hive2://localhost:10000> select * from keywordt1;
+--------------------+-------------------+--------------------------+--
+
| keywordt1.keyword  | keywordt1.total   |   keywordt1.timestamp    |
+--------------------+-------------------+--------------------------+--
+
| que                | 11                | 2017-05-03 21:26:12.317  |
| the                | 10                | 2017-05-03 21:26:12.317  |
| and                | 5                 | 2017-05-03 21:26:12.317  |
| for                | 13                | 2017-05-03 21:25:13.037  |
| لله                | 9                 | 2017-05-03 21:25:13.037  |
| الحمد              | 9                 | 2017-05-03 21:25:13.037  |
| que                | 10                | 2017-05-03 21:24:22.151  |
| the                | 9                 | 2017-05-03 21:24:22.151  |
| for                | 5                 | 2017-05-03 21:24:22.151  |
```

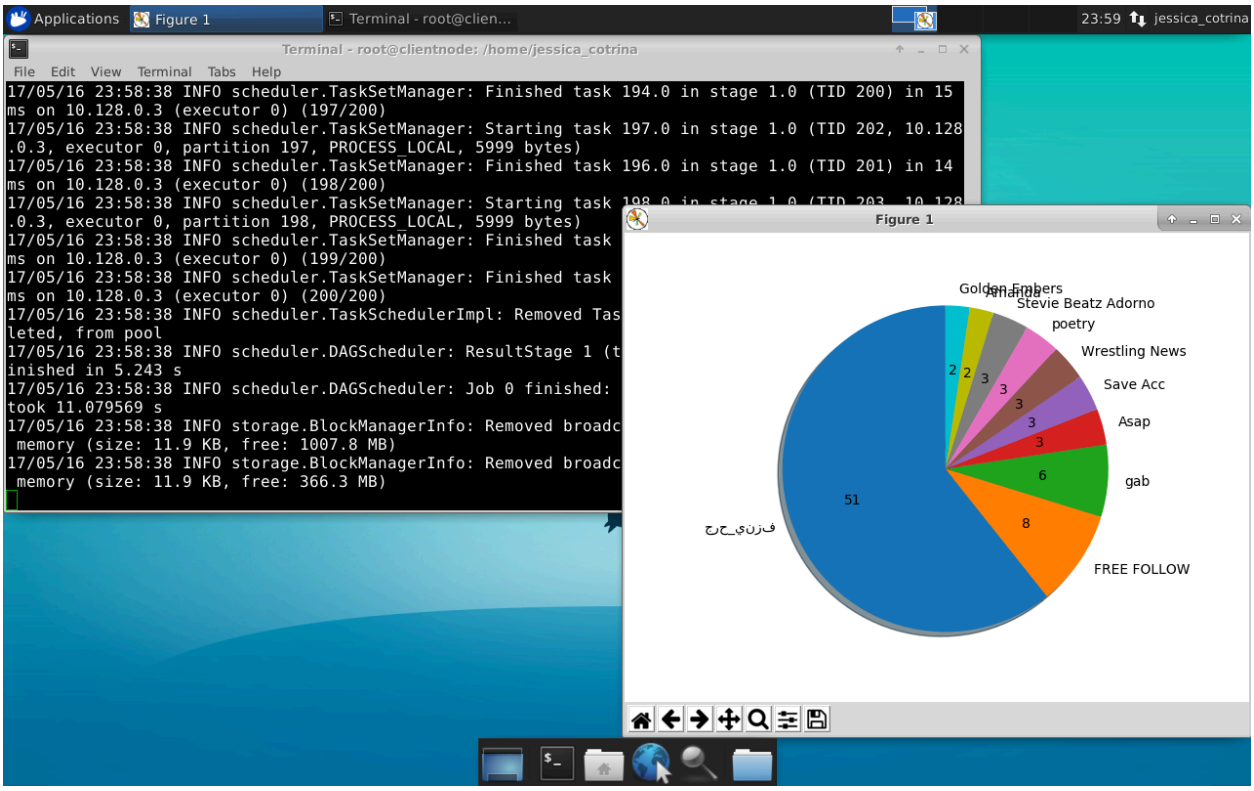


c) Capture the top 10 participants

id1.idparticipant	id1.total	id1.timestamp
Mari	2	2017-05-16 23:39:40.788
👽spooky jim	1	2017-05-16 23:39:40.788
Playmaker Hell	1	2017-05-16 23:39:40.788
Laurinha	2	2017-05-16 23:32:30.952
Rach	1	2017-05-16 23:32:30.952
Zakiyyah	1	2017-05-16 23:32:30.952
JOHN	2	2017-05-16 23:46:50.821
ふわ ⿻	2	2017-05-16 23:46:50.821
Pachii💗👑	1	2017-05-16 23:46:50.821
パチメン@盲目の賢者様に仕え隊	1	2017-05-16 23:43:40.774
Trev	1	2017-05-16 23:43:40.774
rt pinned pls	1	2017-05-16 23:43:40.774
حنان	2	2017-05-16 23:43:30.811
Alex Perkins	1	2017-05-16 23:43:30.811
الوافي 🇰🇼	1	2017-05-16 23:43:30.811
□□□	2	2017-05-16 23:42:40.752
April Harper	1	2017-05-16 23:42:40.752
EriAnna Graham🌸	1	2017-05-16 23:42:40.752
Asap	3	2017-05-16 23:32:40.838
janainaluzia	2	2017-05-16 23:32:40.838

Alvaro	2	2017-05-16 23:32:40.838
ترفه احساس 200K	2	2017-05-16 23:40:20.728
민들레처럼	1	2017-05-16 23:40:20.728
Üzerli Emre	1	2017-05-16 23:40:20.728
□□□□□	2	2017-05-16 23:30:51.236
Alexie Slatta	2	2017-05-16 23:30:51.236
Chalong Bdeerath	1	2017-05-16 23:30:51.236
ينزف_جرح	2	2017-05-16 23:42:20.857
Regina	2	2017-05-16 23:42:20.857
InfoTrinos	1	2017-05-16 23:42:20.857
なおえもん=花田直樹@鬼本ちょこ神推し	1	2017-05-16 23:30:01.331
Muhamad Haziq	1	2017-05-16 23:30:01.331
selo 🤔	1	2017-05-16 23:30:01.331
柵（やな）	2	2017-05-16 23:44:00.851
Deplorable A.B.	1	2017-05-16 23:44:00.851
Zaineb	1	2017-05-16 23:44:00.851
ash	3	2017-05-16 23:45:50.775
caro WILL see harry!	2	2017-05-16 23:45:50.775
黒澤アメジスト	1	2017-05-16 23:45:50.775
Kit Rjd HB	2	2017-05-16 23:44:50.834
Uraz Ertepınar	1	2017-05-16 23:44:50.834
zu72O9xmag	1	2017-05-16 23:44:50.834
laura	2	2017-05-16 23:45:00.842
[H5] 오르페 라우레라	1	2017-05-16 23:45:00.842
Steven Stroh	1	2017-05-16 23:45:00.842

Gráfica



5. Count the number of occurrences for these keyword

a. Trump

b. MAGA

c. Dictator

d. Impeach

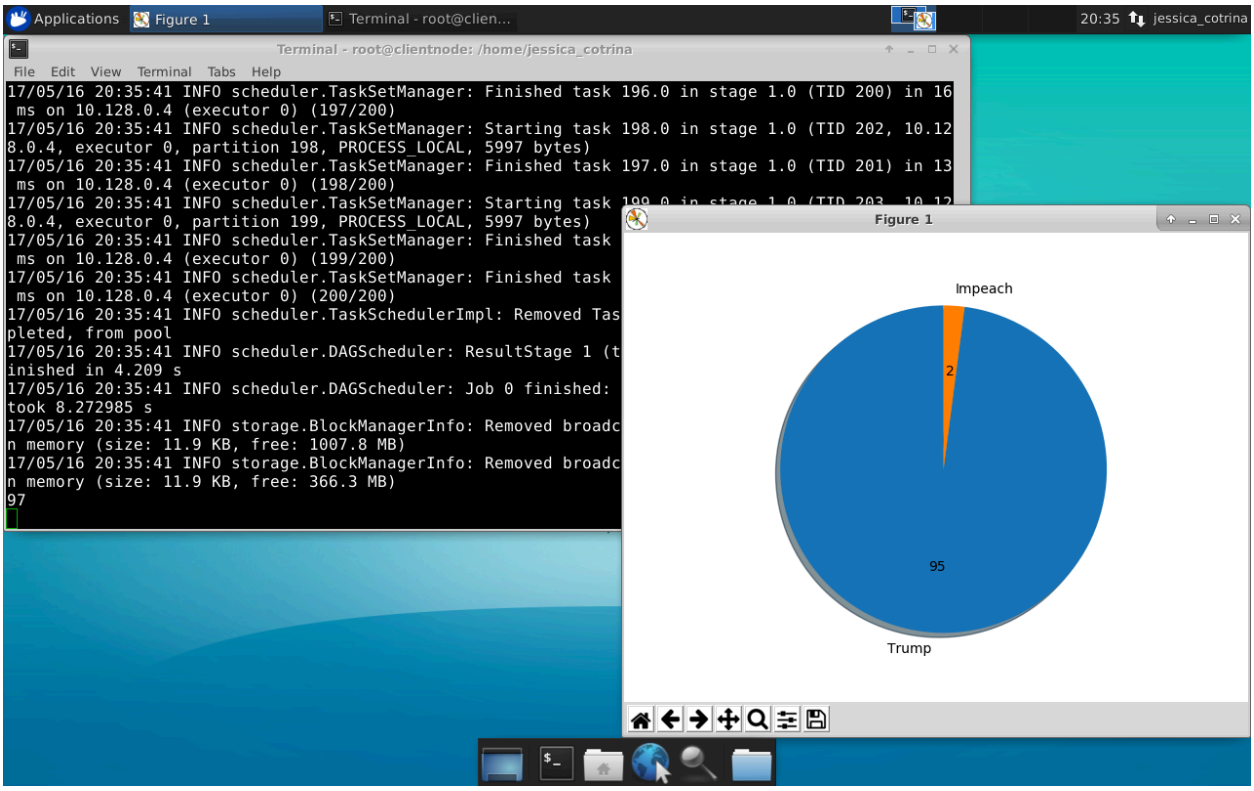
e. Drain

f. Swamp

```
0: jdbc:hive2://localhost:10000> select * from keywordt5;
+---------------------+-------------------+--------------------------+-
-+
| keywordt5.keywordt  | keywordt5.total   |   keywordt5.timestamp    |
+---------------------+-------------------+--------------------------+-
-+
| Trump               | 9                 | 2017-05-15 21:46:32.21   |
| Trump               | 4                 | 2017-05-15 21:56:01.852  |
| Trump               | 4                 | 2017-05-15 21:51:21.853  |
```

Trump	1	2017-05-16 15:37:03.362
Trump	1	2017-05-15 21:48:01.852
Trump	4	2017-05-15 21:50:01.792
Trump	6	2017-05-15 21:47:41.911
Trump	6	2017-05-15 21:55:41.77
Trump	5	2017-05-15 21:49:51.671
Trump	4	2017-05-15 21:49:31.703
Trump	4	2017-05-15 21:53:21.751
Trump	5	2017-05-15 21:51:01.77
Trump	4	2017-05-15 21:48:22.098
Trump	2	2017-05-15 21:53:11.67
Trump	1	2017-05-15 21:53:31.615
Trump	3	2017-05-16 15:38:02.601
Trump	2	2017-05-15 21:54:11.693
Trump	4	2017-05-15 21:54:31.769
Trump	3	2017-05-16 15:36:28.792
Trump	1	2017-05-16 15:41:32.424
Impeach	1	2017-05-16 15:41:32.424
Trump	4	2017-05-15 21:50:11.665
Trump	2	2017-05-16 15:38:42.791
Trump	4	2017-05-15 21:49:21.705
Trump	2	2017-05-16 15:42:52.537
Trump	5	2017-05-16 15:39:03.318
Trump	6	2017-05-16 15:42:22.22
Trump	8	2017-05-15 21:52:21.588
Trump	3	2017-05-15 21:55:01.773
Trump	2	2017-05-16 15:42:32.301

Gráfica usando Pandas


Applications
Figure 1
Terminal - root@clien...
20:35
jessica_cotrina
Terminal - root@clientnode: /home/jessica_cotrina
File Edit View Terminal Tabs Help
17/05/16 20:35:41 INFO scheduler.TaskSetManager: Finished task 196.0 in stage 1.0 (TID 200) in 16
ms on 10.128.0.4 (executor 0) (197/200)
17/05/16 20:35:41 INFO scheduler.TaskSetManager: Starting task 198.0 in stage 1.0 (TID 202, 10.12
8.0.4, executor 0, partition 198, PROCESS_LOCAL, 5997 bytes)
17/05/16 20:35:41 INFO scheduler.TaskSetManager: Finished task 197.0 in stage 1.0 (TID 201) in 13
ms on 10.128.0.4 (executor 0) (198/200)
17/05/16 20:35:41 INFO scheduler.TaskSetManager: Starting task
8.0.4, executor 0, partition 199, PROCESS_LOCAL, 5997 bytes)
17/05/16 20:35:41 INFO scheduler.TaskSetManager: Finished task
ms on 10.128.0.4 (executor 0) (199/200)
17/05/16 20:35:41 INFO scheduler.TaskSetManager: Finished task
ms on 10.128.0.4 (executor 0) (200/200)
17/05/16 20:35:41 INFO scheduler.TaskSchedulerImpl: Removed Tas
pleted, from pool
17/05/16 20:35:41 INFO scheduler.DAGScheduler: ResultStage 1 (t
inished in 4.209 s
17/05/16 20:35:41 INFO scheduler.DAGScheduler: Job 0 finished:
took 8.272985 s
17/05/16 20:35:41 INFO storage.BlockManagerInfo: Removed broadc
n memory (size: 11.9 KB, free: 1007.8 MB)
17/05/16 20:35:41 INFO storage.BlockManagerInfo: Removed broadc
n memory (size: 11.9 KB, free: 366.3 MB)
97
Figure 1
Impeach
2
95
Trump