

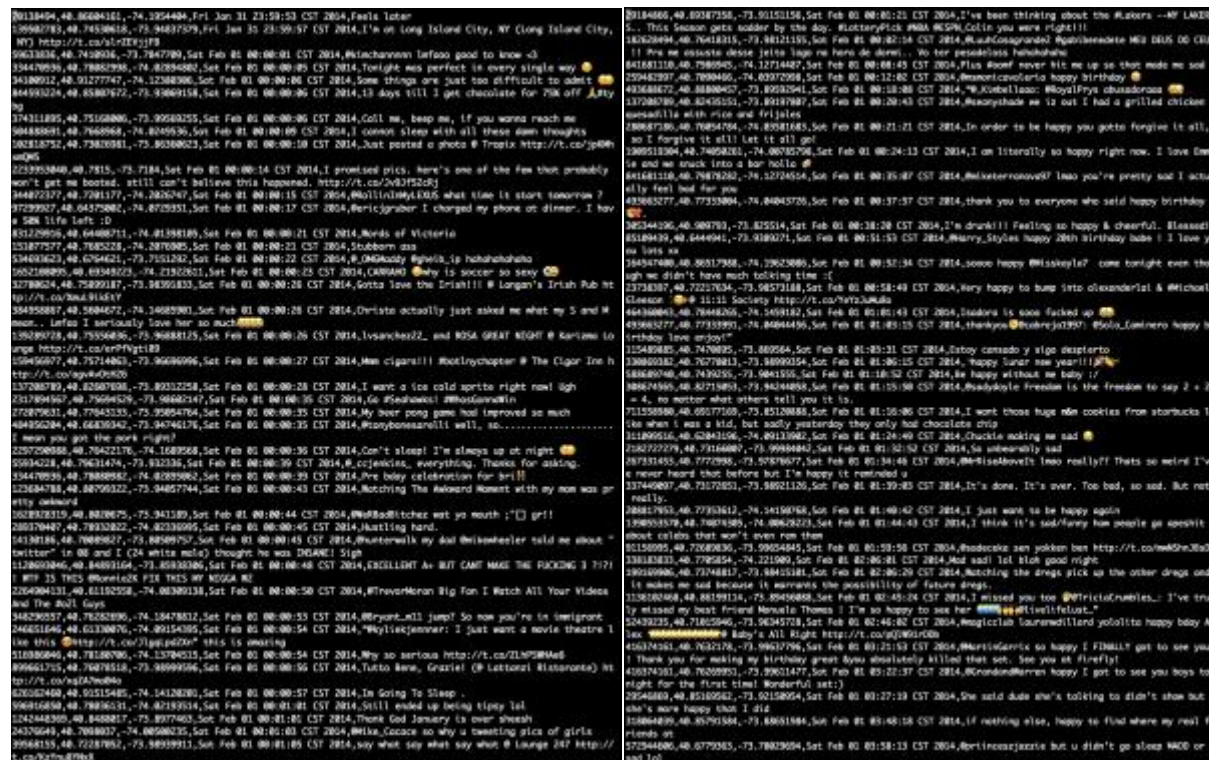
Karl Roth  
GEOG 479  
Lab 11

In this lab we explored the use of PIG grunt shell for processing twitter data and learned how to run several pig scripts on HDFS. Figure 1 demonstrates the files created in our HDFS for this laboratory. The lab 11 folder was imported in from ROGER and the result.txt and result\_keywords.txt file were created. The result.txt file was created by loading the data into the grunt shell “cleaning” the data and filtering out only the tweets originating in New York. The result\_keywords.txt file was created by limiting to tweets that contain the word keywords “happy” or “sad”. Fig 2 demonstrates the contents of results.txt and result\_keywords.txt respectively.

drwx-----	-	karoth4	hdfs	0	2017-11-15	00:00	.Trash
drwx-----	-	karoth4	hdfs	0	2017-11-24	11:17	.staging
drwxr-xr-x	-	karoth4	hdfs	0	2017-11-24	10:50	lab11
-rw-r--r--	3	karoth4	hdfs	2781761974	2017-11-14	15:49	ny_taxi_1.csv
drwxr-xr-x	-	karoth4	hdfs	0	2017-11-24	11:08	result.txt
drwxr-xr-x	-	karoth4	hdfs	0	2017-11-24	11:17	result_keywords.txt

Fig 1: The contents of the HDFS after lab 11 was completed.

The contents o



70138494,40.86844161,-74.1954444,Fri Jan 31 23:59:53 CST 2014,Feels later  
139580793,40.74530616,-73.94837579,Fri Jan 31 23:59:57 CST 2014,I'm on Long Island City, NY (Long Island City, NY) <http://t.co/v1EEX3j9>  
59433335,40.74530616,-73.94837579,Sat Feb 01 00:00:01 CST 2014,Mitchellsonn infaso good to know :D  
334470936,40.78857096,-74.82584802,Sat Feb 01 00:00:05 CST 2014,Tonight was perfect in every single way 🍷  
34180012,40.91277747,-74.12386306,Sat Feb 01 00:00:08 CST 2014,Some things are just too difficult to admit  
84459324,40.65807672,-73.93691158,Sat Feb 01 00:00:09 CST 2014,13 days till I get chocolate for 7th off :D  
37431189,40.75150886,-73.95502355,Sat Feb 01 00:00:09 CST 2014,Cali me, keep me, if you wanna reach me  
944885891,40.7646966,-74.8245636,Sat Feb 01 00:00:09 CST 2014,I cannot sleep with all these damn thoughts  
102813732,40.73803981,-73.90380623,Sat Feb 01 00:00:10 CST 2014,Just posted a photo # Tootix <http://t.co/gpmh>  
w0G5  
2239353840,40.7815,-73.7184,Sat Feb 01 00:00:14 CST 2014,I promised pics, here's one of the few that probably  
won't get me booted, still can't believe this happened. <http://t.co/2d3j25d3>  
34482337,40.7261177,-74.2625747,Sat Feb 01 00:00:15 CST 2014,ollin(n)ny1303 what time is start tomorrow?  
37299327,40.64575982,-74.8729351,Sat Feb 01 00:00:17 CST 2014,Berlicjgnber I changed my phone or dinner. I lov  
a 5th life left :D  
83127995,40.6448711,-74.8158180,Sat Feb 01 00:00:21 CST 2014,Needs of Victoria  
13287577,40.7482323,-74.3870585,Sat Feb 01 00:00:23 CST 2014,Indebm oia  
53403621,40.6794322,-73.7511282,Sat Feb 01 00:00:22 CST 2014,0.DAGADADY RYH4S,lp hahahaha  
162218885,40.69348523,-74.2132831,Sat Feb 01 00:00:23 CST 2014,CARHARD 🍷why is soccer so sexy  
37280624,40.7589187,-73.9359383,Sat Feb 01 00:00:28 CST 2014,Gotta love the Irish!! @ Langan's Irish Pub b  
ut <http://t.co/9M4A1H4EY>  
358358817,40.5844572,-74.34853801,Sat Feb 01 00:00:28 CST 2014,Christie actually just asked me what my 5 and 6  
meas... Lefee I seriously love her so much 🍷  
138280728,40.73556836,-73.96881125,Sat Feb 01 00:00:26 CST 2014,Ivanchev22... and ROSA GREAT NGAR @ Karimou  
to jump <http://t.co/9r7Ngt1B>  
10945697,40.75714803,-73.9609096,Sat Feb 01 00:00:27 CST 2014,Was cigar!!! #ootychapter @ The Elgar Inn h  
ttp://t.co/9o9y4d08D0  
137280782,40.82587038,-73.8312258,Sat Feb 01 00:00:28 CST 2014,I want a ice cold apette right now! igh  
231789467,40.79669359,-73.98862147,Sat Feb 01 00:00:35 CST 2014,Go @Sawadai #Moscowdin  
270879631,40.77845333,-73.35854704,Sat Feb 01 00:00:35 CST 2014,My beer pong game had improved so much  
44435204,40.6830342,-73.34746176,Sat Feb 01 00:00:35 CST 2014,Mybonybonesare!!! well, so...  
I mean you got the pork pig!!  
220750888,40.70421716,-74.1639558,Sat Feb 01 00:00:36 CST 2014,Can't sleep! I'm always up at night  
55934228,40.79631474,-73.832336,Sat Feb 01 00:00:39 CST 2014,c\_cjceinas, everything. Thanks for asking.  
334470936,40.78857096,-74.82584802,Sat Feb 01 00:00:39 CST 2014,Pre baby celebration for br!!  
132384781,40.68709322,-73.94857744,Sat Feb 01 00:00:43 CST 2014,Batching The Award Moment with my mom was gr  
atly ednary  
102813732,40.73803981,-73.90380623,Sat Feb 01 00:00:44 CST 2014,Me@BdRitcher wet ya mouth :D! G!!  
280379847,40.7893822,-74.82336955,Sat Feb 01 00:00:45 CST 2014,Kustling hard.  
14138188,40.78857096,-73.85895757,Sat Feb 01 00:00:45 CST 2014,#hunderwalk my dad @hunderwalk told me about "t  
witter" in 80 and I (aka white male) thought he was DEMENT! 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious <http://t.co/2L9M9M46>  
899661735,40.76878514,-73.9899936,Sat Feb 01 00:00:55 CST 2014,Tutto bene, Gracie! @ Lational Rianoranta) b  
ut <http://t.co/v1EEX3j9>  
34272957,40.76252896,-74.18478012,Sat Feb 01 00:00:55 CST 2014,Myport...all jump! So now you're in immigrant  
246551546,40.61338076,-74.80154306,Sat Feb 01 00:00:54 CST 2014, #Weylconner: I just want a movie theatre I  
love this <http://t.co/3jgpp6209> this is waiting  
51838046,40.78150886,-74.12386306,Sat Feb 01 00:00:54 CST 2014,My so serious

Then we ran the heatmap.pig script. This file loads radiation data from HDFS, cleans the data, and filters it for radiation and locations data. The area is separated into a grid with each block having a latitude and longitude "key". The radiation value is then averaged over each block on the grid using the user-defined function ".getAverage()" and the result is outputted into the heatmap\_output folder in HDFS.

Finally, the kmeans.pig script was run. This file loads the data from HDFS and cleans the data. Then the "request" field is filtered for entries regarding "Graffiti Removal", but other requests also exist. The data is then filtered by time and streamed through the kmeans.py script. The kmeans.py script uses the kmeans clustering method to "group" the instances of Graffiti Removal geographically. Each entry is assigned a group number from 0-4 inclusive. The output is then returned to the pig script where it is saved in HDFS. Streaming data through a python script allows us to use simple python code and industry standard packages to accomplish tasks not achievable by PIG alone.