ANALYZING THE SOCIAL MEDIA TRENDS USING THE TWITTER HASHTAGS

[Project – Mapreduce and Flume combined]

- Prepared by Vignesh.R (15CSE107)

Step 1:

Download the data from twitter. The procedures are already explained in the earlier flume documentation. There the keywords are entered before compiling the .jar file and then the mentioned tweet data is downloaded to the hadoop file system.

Step 2:

Then a new jar file named TopTrendsFinder is created with the necessary external jar files added to it. The jar files contains the following java programs.

- ReverseComparator.java
- TopTrendsFinder.java
- TrendMapper1.java
- TrendMapper2.java
- TrendReducer1.java
- TrendReducer2.java

The configuration is set in such a way that only the hashtags are taken into account.

Two mapper classes and two reducer classes are written in which the first mapper phase collects the hashtags and its corresponding reducer phase combines it. On the other hand, the second mapper phase re-orders it to shows the top trending hashtag in its reducer phase.

The TopTrendsFinder.jar can be downloaded from the following link:-

https://drive.google.com/file/d/0B_ip9omU6d3yeVdUTnlWdF9INXM/view?usp=sharing

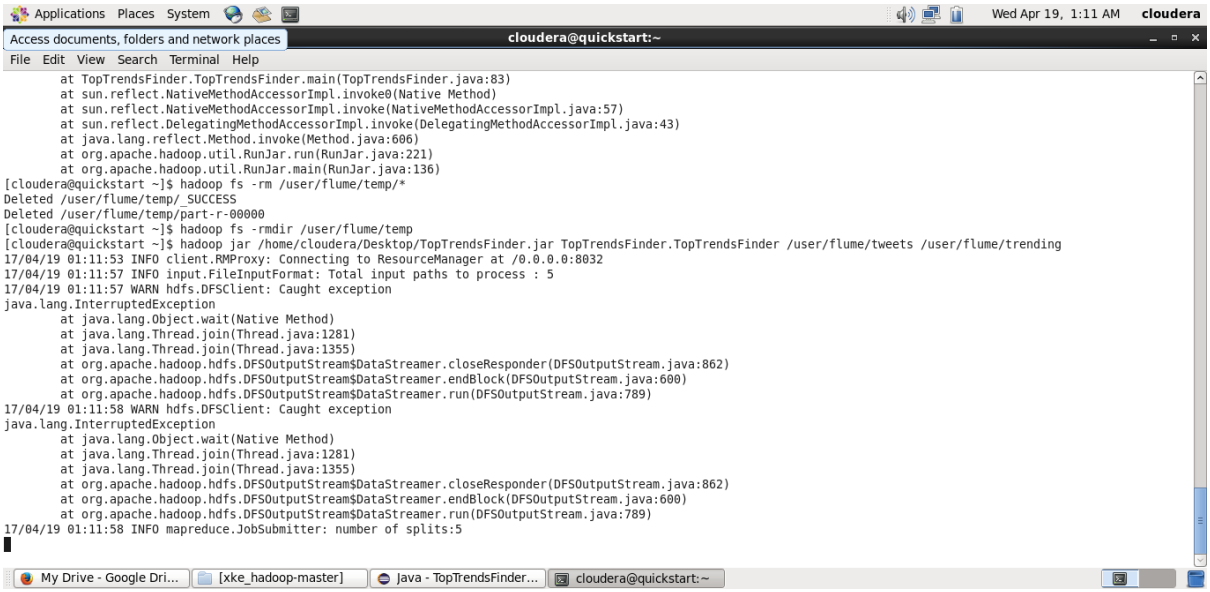
Step 3:

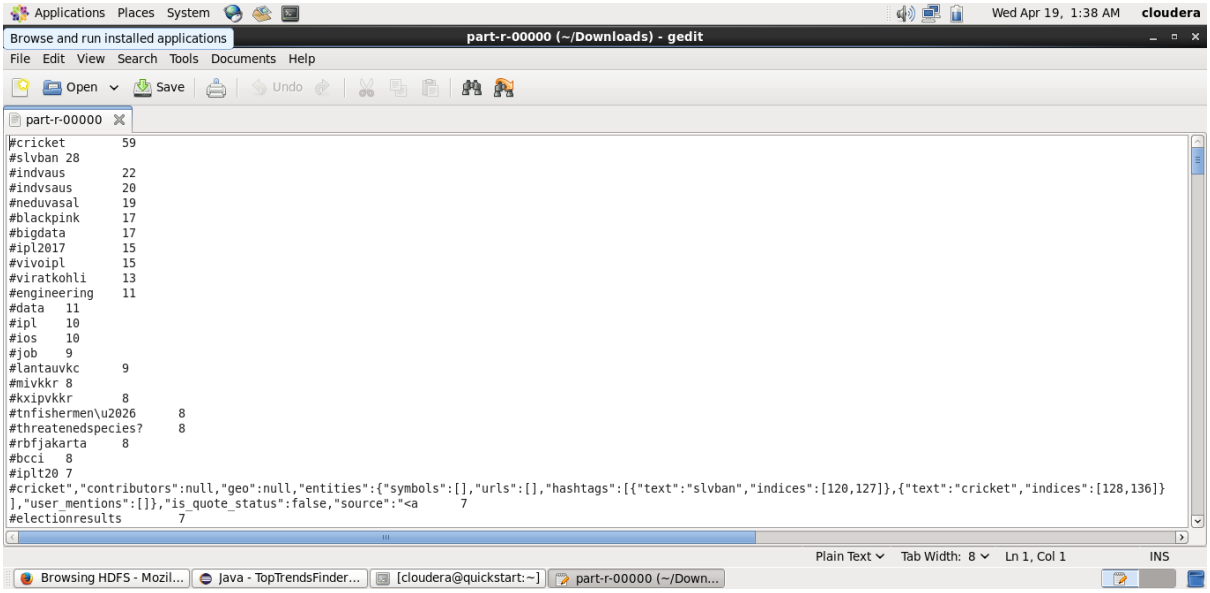
Command to execute the jar file:

hadoop jar /home/cloudera/Desktop/TopTrendsFinder.jar TopTrendsFinder.TopTrendsFinder /user/flume/tweets /user/flume/trending

If the input path is to be changed, then the input,output and temporary path are also needed to be changed in the main TopTrendsFinder.java program.

Result:


Applications Places System
Wed Apr 19, 1:11 AM cloudera
Access documents, folders and network places
cloudera@quickstart:~
File Edit View Search Terminal Help
at TopTrendsFinder.TopTrendsFinder.main(TopTrendsFinder.java:83)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:606)
at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
[cloudera@quickstart ~]$ hadoop fs -rm /user/flume/temp/*
Deleted /user/flume/temp/_SUCCESS
Deleted /user/flume/temp/part-r-00000
[cloudera@quickstart ~]$ hadoop fs -rmdir /user/flume/temp
[cloudera@quickstart ~]$ hadoop jar /home/cloudera/Desktop/TopTrendsFinder.jar TopTrendsFinder.TopTrendsFinder /user/flume/tweets /user/flume/trending
17/04/19 01:11:53 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
17/04/19 01:11:57 INFO input.FileInputFormat: Total input paths to process : 5
17/04/19 01:11:57 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:862)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:600)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:789)
17/04/19 01:11:58 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:862)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:600)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:789)
17/04/19 01:11:58 INFO mapreduce.JobSubmitter: number of splits:5
My Drive - Google Dri...
[xke_hadoop-master]
Java - TopTrendsFinder...
cloudera@quickstart:~


Applications Places System
Wed Apr 19, 1:38 AM cloudera
Browse and run installed applications
part-r-00000 (~/Downloads) - gedit
File Edit View Search Tools Documents Help
Open Save Undo
part-r-00000
#cricket 59
#slvban 28
#indvaus 22
#indvsaus 20
#neduvasal 19
#blackpink 17
#bigdata 17
#ipl2017 15
#vivoipl 15
#viratkohli 13
#engineering 11
#data 11
#ipl 10
#ios 10
#job 9
#lantauvkc 9
#mivkkr 8
#kxipvkkr 8
#tnfishermen\u2026 8
#threatenedspecies? 8
#rbfjakarta 8
#bcci 8
#iplt20 7
#cricket","contributors":null,"geo":null,"entities":{"symbols":[],"urls":[],"hashtags":[{"text":"slvban","indices":[120,127]},{"text":"cricket","indices":[128,136]}
],"user_mentions":[]},"is_quote_status":false,"source":"<a 7
#electionresults 7
Plain Text Tab Width: 8 Ln 1, Col 1 INS
Browsing HDFS - Mozil...
Java - TopTrendsFinder...
[cloudera@quickstart:~]
part-r-00000 (~/Down...