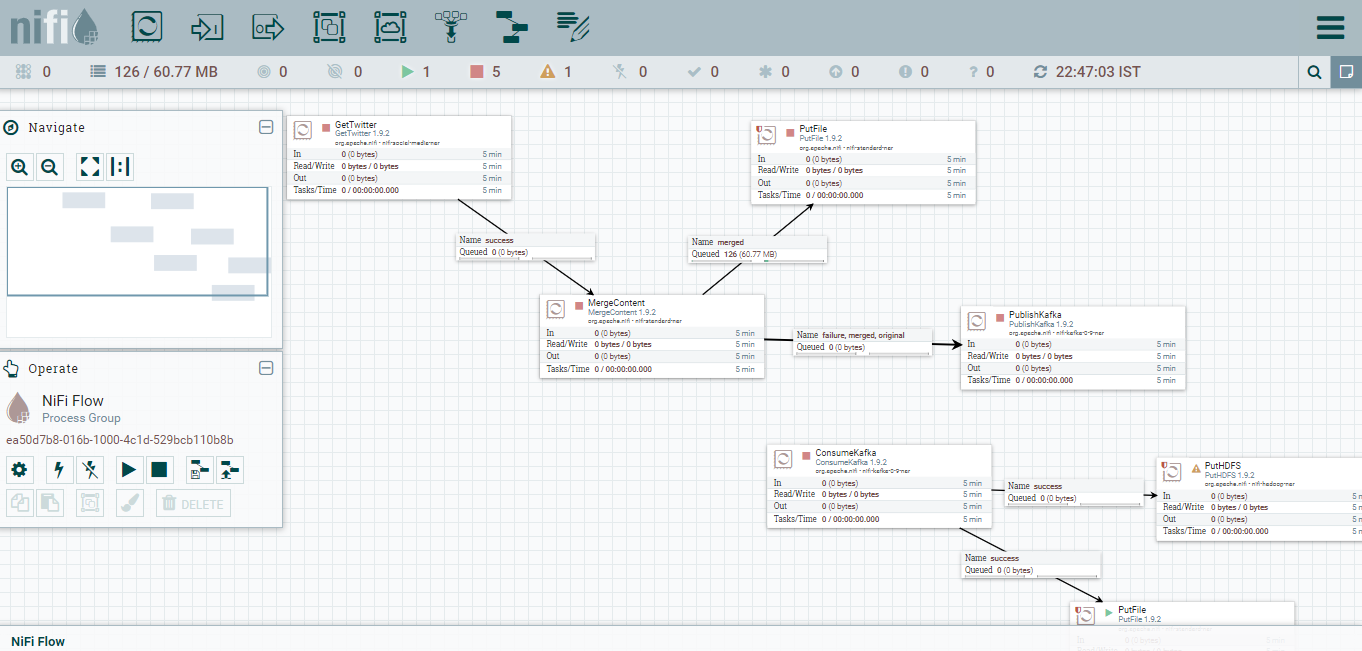
Fetching data from Twitter using Apache Nifi:

<http://localhost:8080/nifi/>

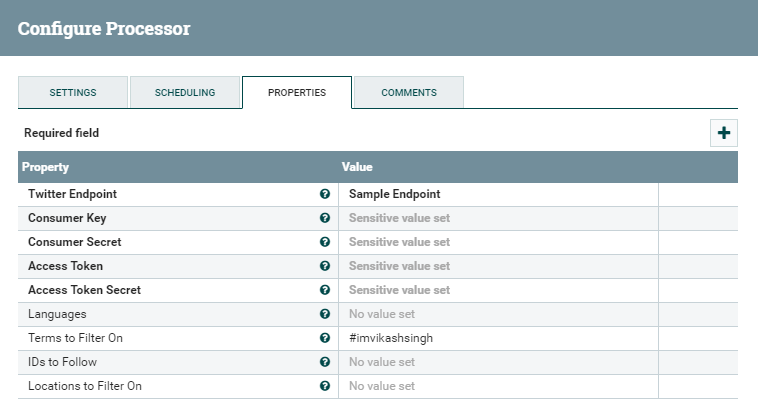
Nifi Version: 1.9.2



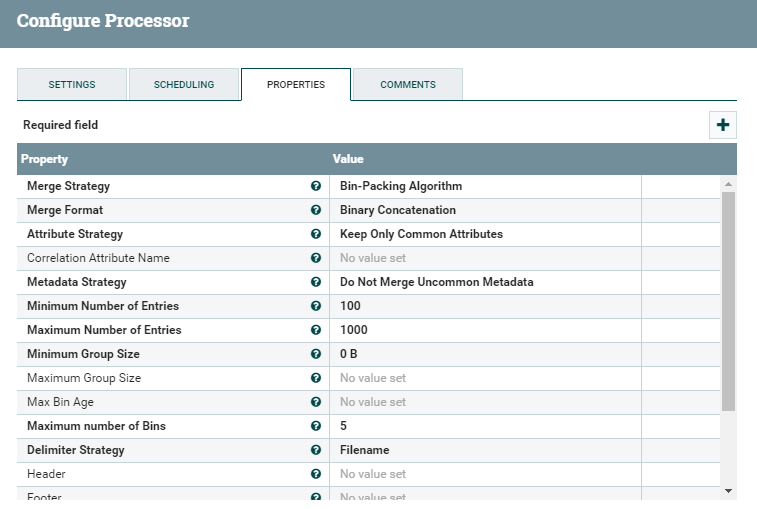
Apache Nifi is continuously loading the data into Kafka Topics (mytopics) .

Processor used to fetch the data into kafka

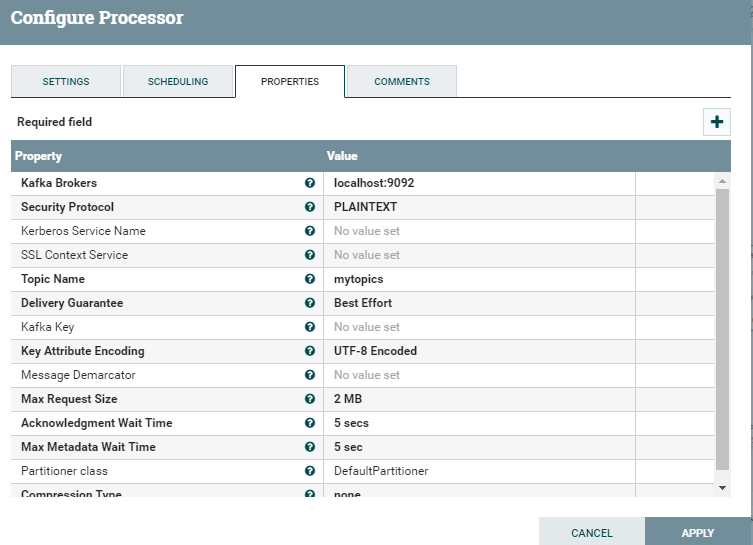
1. GetTwitter processor to get data from Twitter website. Currently I am using the raw data



1. MergeContent processor to merge the files received based on the file size



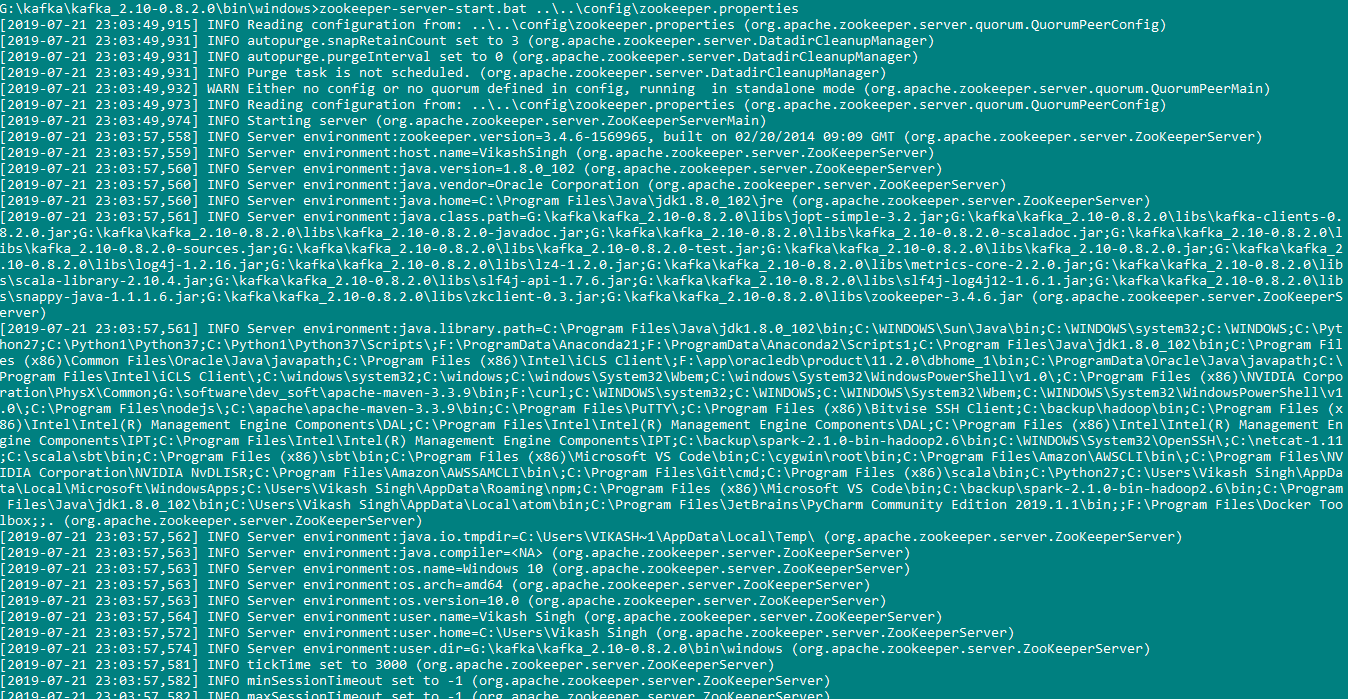
1. PublishKafka processor : This processor is publishing the data to kafka topics(mytopics)



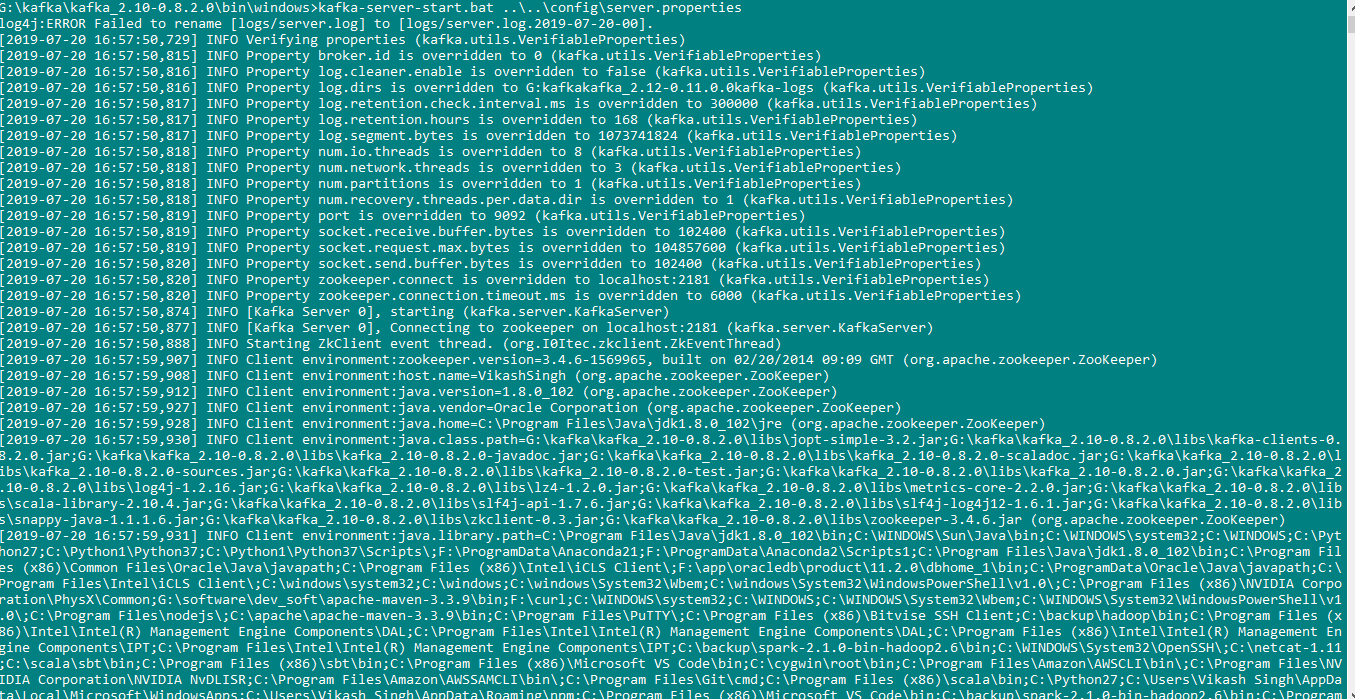
Apache Kafka:

Version: 2.10-0.8.2.0

1. Zookeeper server Start



1. Kafka Server Start



1. Kafka Topics: For Demo purpose I have created my topics with one partition with one replication facor.

<kafka\_dir>\bin\windows\kafka-topics.bat --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic mytopics

Data Received from twitter to kafka Topics

{"created\_at":"Sat Jul 13 10:07:40 +0000 2019","id":1149983726245482496,"id\_str":"1149983726245482496","text":"@realFFK Hmmmmm","display\_text\_range":[9,15],"source":"\u003ca href=\"http:\/\/twitter.com\/download\/iphone\" rel=\"nofollow\"\u003eTwitter for iPhone\u003c\/a\u003e","truncated":false,"in\_reply\_to\_status\_id":1149748835104804869,"in\_reply\_to\_status\_id\_str":"1149748835104804869","in\_reply\_to\_user\_id":597224882,"in\_reply\_to\_user\_id\_str":"597224882","in\_reply\_to\_screen\_name":"realFFK","user":{"id":848964040361725953,"id\_str":"848964040361725953","name":"Rashidabkd","screen\_name":"cuterashhhh","location":"Kaduna, Nigeria","url":null,"description":null,"translator\_type":"none","protected":false,"verified":false,"followers\_count":283,"friends\_count":152,"listed\_count":0,"favourites\_count":3486,"statuses\_count":2182,"created\_at":"Mon Apr 03 18:22:55 +0000 2017","utc\_offset":null,"time\_zone":null,"geo\_enabled":false,"lang":null,"contributors\_enabled":false,"is\_translator":false,"profile\_background\_color":"F5F8FA","profile\_background\_image\_url":"","profile\_background\_image\_url\_https":"","profile\_background\_tile":false,"profile\_link\_color":"1DA1F2","profile\_sidebar\_border\_color":"C0DEED","profile\_sidebar\_fill\_color":"DDEEF6","profile\_text\_color":"333333","profile\_use\_background\_image":true,"profile\_image\_url":"http:\/\/pbs.twimg.com\/profile\_images\/1149979796027453440\/MGmorHI-\_normal.jpg","profile\_image\_url\_https":"https:\/\/pbs.twimg.com\/profile\_images\/1149979796027453440\/MGmorHI-\_normal.jpg","profile\_banner\_url":"https:\/\/pbs.twimg.com\/profile\_banners\/848964040361725953\/1561452032","default\_profile":true,"default\_profile\_image":false,"following":null,"follow\_request\_sent":null,"notifications":null},"geo":null,"coordinates":null,"place":null,"contributors":null,"is\_quote\_status":false,"quote\_count":0,"reply\_count":0,"retweet\_count":0,"favorite\_count":0,"entities":{"hashtags":[],"urls":[],"user\_mentions":[{"screen\_name":"realFFK","name":"Femi Fani-Kayode","id":597224882,"id\_str":"597224882","indices":[0,8]}],"symbols":[]},"favorited":false,"retweeted":false,"filter\_level":"low","lang":"und","timestamp\_ms":"1563012460658"}

Apache Spark:

Version: 2.10

1. **First Transformation** (SparkStreamingFromKafka.scala): Retrieve the all twitter data based on the condition apply on the data in json ( used spark sql and apply query)

2> **Second Transformation** (TwitterDataAnalysis.scala): Search the provided hash code from twitter and save to Cassandra on every window time frame of 10 second.

Cassandra Version: -3.11.4

Query Need to create in Cassandra:

1. For keySpace

CREATE KEYSPACE vkspace WITH REPLICATION = {'class': 'SimpleStrategy', 'replication\_factor': 1 };

1. For table 1

CREATE TABLE vkspace.twitterdata (id VARCHAR PRIMARY KEY, created\_at VARCHAR , favorite\_count VARCHAR );

1. For table 2

CREATE TABLE vkspace.toptentag (count int PRIMARY KEY, tags VARCHAR);

