

## Lecture du flux de tweet générer en utilisant Kafka

Prérequis :

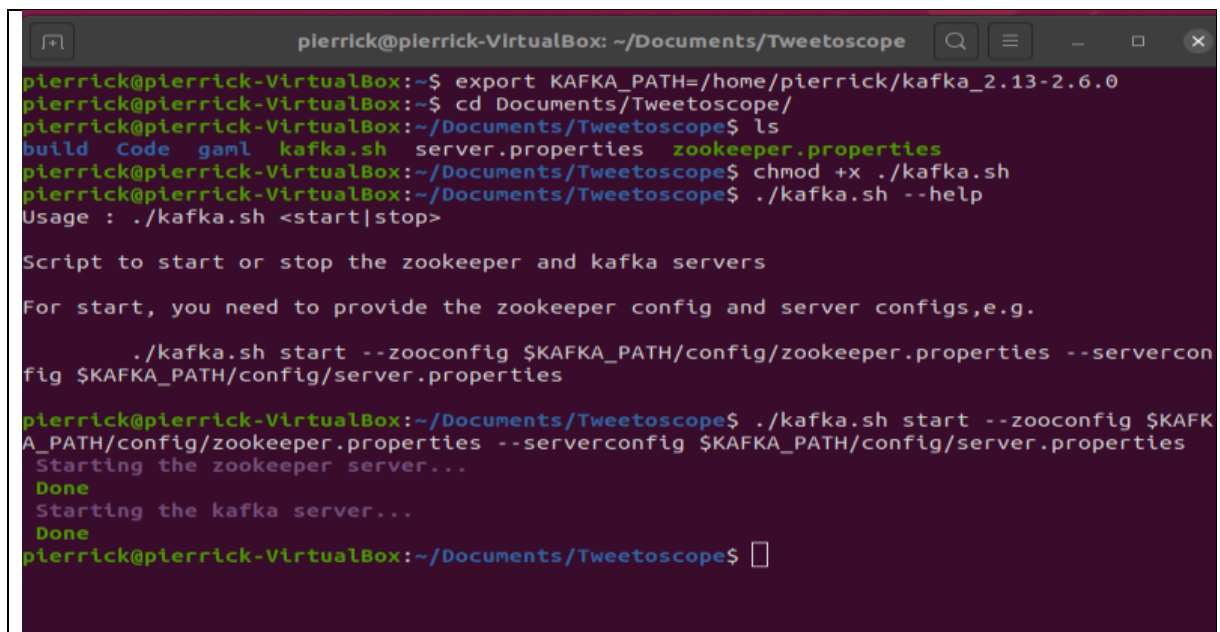
- Avoir installer les différents modules [Lien](#)
- Avoir télécharger les différents fichiers

### Lancer le tweet generator

Dans un terminal, on créer une variable « KAFKA\_PATH » qui indique le chemin de l'installation de Kafka.

Ensuite, il faut se rendre dans le dossier où est posé le fichier « kafka.sh » et créer un exécutable de ce fichier « `chmod +x kafka.sh` ».

Puis en utilisant la commande « `./kafka.sh --help` », on obtient la ligne de commande pour démarrer les serveurs



```
pierrick@pierrick-VirtualBox: ~/Documents/Tweetoscope
pierrick@pierrick-VirtualBox:~$ export KAFKA_PATH=/home/pierrick/kafka_2.13-2.6.0
pierrick@pierrick-VirtualBox:~$ cd Documents/Tweetoscope/
pierrick@pierrick-VirtualBox:~/Documents/Tweetoscope$ ls
build Code gaml kafka.sh server.properties zookeeper.properties
pierrick@pierrick-VirtualBox:~/Documents/Tweetoscope$ chmod +x ./kafka.sh
pierrick@pierrick-VirtualBox:~/Documents/Tweetoscope$ ./kafka.sh --help
Usage : ./kafka.sh <start|stop>

Script to start or stop the zookeeper and kafka servers

For start, you need to provide the zookeeper config and server configs,e.g.

    ./kafka.sh start --zooconfig $KAFKA_PATH/config/zookeeper.properties --servercon
fig $KAFKA_PATH/config/server.properties

pierrick@pierrick-VirtualBox:~/Documents/Tweetoscope$ ./kafka.sh start --zooconfig $KAFK
A_PATH/config/zookeeper.properties --serverconfig $KAFKA_PATH/config/server.properties
Starting the zookeeper server...
Done
Starting the kafka server...
Done
pierrick@pierrick-VirtualBox:~/Documents/Tweetoscope$
```

Ensuite, il faut compiler « tweet-generator », pour cela il faut taper la commande :

« `g++ -o tweet-generator -O3 -Wall -std=c++17 `pkg-config --libs --cflags gaml cppkafka` -lpthread tweet-generator.cpp` »

Puis, on lance le flux de tweet :

```
pierrick@pierrick-VirtualBox: ~/Documents/Tweetoscope/Code
[sudo] Mot de passe de pierrick :
Parameters :
  tweets_path           = /home/pierrick/Documents/Tweetoscope/Code/news-data.csv
  cascades_path         = /home/pierrick/Documents/Tweetoscope/Code/news-index.csv
  tweets_index_path     = ./tweets.idx
  cascades_index_path   = ./cascades.idx
  tweets_cache_page_size = 1000
  tweets_cache_nb_pages = 10
  cascades_cache_page_size = 100
  cascades_cache_nb_pages = 10
  cascade_random_start_time_range = 10
  seconds_per_time_unit = 1
  time_acceleration_factor = 1000
  max_number_of_cascades = 5
  number_of_threads     = 5
  kafka_topic           = tweets
  kafka_brokers         = localhost:9092

Dataset files read
```

Créer un consumer permettant de lire les infos du tweet generator

Pour récupérer le flux, il faut utiliser le programme python suivant :

```
from kafka import KafkaConsumer

consumerProperties = { "bootstrap_servers":['localhost:9092'],
                      "auto_offset_reset":"earliest",
                      "group_id":"myOwnPrivatePythonGroup" }

consumer = KafkaConsumer(**consumerProperties)
consumer.subscribe("tweets")

for record in consumer:
    print("%s" % (record.value.decode()))
```

Et dans un terminal, lancer le script python :

```
pierrick@pierrick-VirtualBox: ~/Documents/Tweetoscope/Code/KAFKA
pierrick@pierrick-VirtualBox:~/Documents/Tweetoscope/Code/KAFKA$ python3 consumer.py
{"type": "retweet", "msg": "None", "t": 229005, "m": 10386, "source": 2, "info": "cascade=4472"}
{"type": "retweet", "msg": "None", "t": 229128, "m": 277, "source": 2, "info": "cascade=4472"}
{"type": "retweet", "msg": "None", "t": 229169, "m": 208, "source": 2, "info": "cascade=4472"}
{"type": "retweet", "msg": "None", "t": 228515, "m": 3287, "source": 0, "info": "cascade=8088"}
{"type": "retweet", "msg": "None", "t": 229184, "m": 539, "source": 2, "info": "cascade=4472"}
{"type": "retweet", "msg": "None", "t": 229836, "m": 17, "source": 3, "info": "cascade=14273"}
{"type": "retweet", "msg": "None", "t": 228592, "m": 341, "source": 0, "info": "cascade=8088"}
{"type": "retweet", "msg": "None", "t": 229281, "m": 38, "source": 2, "info": "cascade=4472"}
{"type": "retweet", "msg": "None", "t": 229296, "m": 17, "source": 2, "info": "cascade=4472"}
{"type": "retweet", "msg": "None", "t": 229486, "m": 1156, "source": 4, "info": "cascade=1320"}
{"type": "retweet", "msg": "None", "t": 229323, "m": 309, "source": 2, "info": "cascade=4472"}
{"type": "retweet", "msg": "None", "t": 229358, "m": 404, "source": 2, "info": "cascade=4472"}
{"type": "retweet", "msg": "None", "t": 229513, "m": 455, "source": 2, "info": "cascade=4472"}
{"type": "retweet", "msg": "None", "t": 229012, "m": 1063, "source": 0, "info": "cascade=8088"}
{"type": "retweet", "msg": "None", "t": 229700, "m": 60, "source": 2, "info": "cascade=4472"}
{"type": "retweet", "msg": "None", "t": 229781, "m": 218, "source": 2, "info": "cascade=4472"}
{"type": "retweet", "msg": "None", "t": 229799, "m": 46, "source": 2, "info": "cascade=4472"}
{"type": "retweet", "msg": "None", "t": 229843, "m": 235, "source": 2, "info": "cascade=4472"}
{"type": "retweet", "msg": "None", "t": 229876, "m": 23, "source": 2, "info": "cascade=4472"}
{"type": "retweet", "msg": "None", "t": 229930, "m": 431, "source": 2, "info": "cascade=4498"}
{"type": "retweet", "msg": "None", "t": 229934, "m": 23, "source": 2, "info": "cascade=4472"}
```