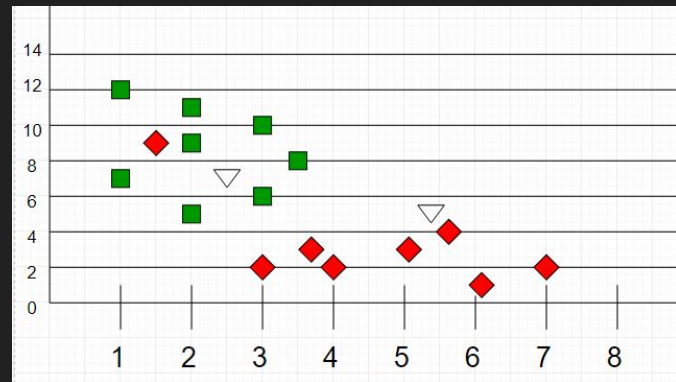


# Programação concorrente

Implementação do algoritmo Knn(k-nearest neighbors algorithm) Utilizando:

- Apache Kafka
- Apache Spark



# Kafka

Versão utilizada:  
kafka\_2.12-2.7.0

## Inicialização zookeeper.

```
marcos@marcos-340XAA-350XAA-550XAA: ~/Documentos/kafka_2.12-2.7.0
marcos@marcos-340XAA-350XAA-550XAA: ~/Documentos/kafka_2.12-2.7.0$ bin/zookeeper-server-start.sh config/zookeeper.properties
```

```
1 import findspark
2 import math
3 from pyspark import SparkContext
4 from pyspark.streaming import StreamingContext
```

Atividades Terminal 22 de abr 12:39

```
marcos@marcos-340XAA-350XAA-550XAA: ~/Documentos/kafka_2.12-2.7.0
marcos@marcos-340XAA-350XAA-550XAA: ~/Documentos/kafka_2.12-2.7.0$ bin/zookeeper-server-start.sh config/zookeeper.properties
```

```
2021-04-22 12:39:34,507 INFO Server environment:java.io.tmpdir=/tmp (org.apache.zookeeper.server.ZooKeeperServer)
2021-04-22 12:39:34,507 INFO Server environment:java.compiler=NA (org.apache.zookeeper.server.ZooKeeperServer)
2021-04-22 12:39:34,508 INFO Server environment:os.name=Linux (org.apache.zookeeper.server.ZooKeeperServer)
2021-04-22 12:39:34,508 INFO Server environment:os.arch=amd64 (org.apache.zookeeper.server.ZooKeeperServer)
2021-04-22 12:39:34,508 INFO Server environment:os.version=5.8.0-generic (org.apache.zookeeper.server.ZooKeeperServer)
2021-04-22 12:39:34,508 INFO Server environment:user.name=marcos (org.apache.zookeeper.server.ZooKeeperServer)
2021-04-22 12:39:34,508 INFO Server environment:user.home=/home/marcos (org.apache.zookeeper.server.ZooKeeperServer)
2021-04-22 12:39:34,508 INFO Server environment:user.dir=/home/marcos/Documentos/kafka_2.12-2.7.0 (org.apache.zookeeper.server.ZooKeeperServer)
2021-04-22 12:39:34,508 INFO Server environment:os.memory.free=496MB (org.apache.zookeeper.server.ZooKeeperServer)
2021-04-22 12:39:34,508 INFO Server environment:os.memory.max=512MB (org.apache.zookeeper.server.ZooKeeperServer)
2021-04-22 12:39:34,508 INFO Server environment:os.memory.total=512MB (org.apache.zookeeper.server.ZooKeeperServer)
2021-04-22 12:39:34,510 INFO minSessionTimeout set to 6000 (org.apache.zookeeper.server.ZooKeeperServer)
2021-04-22 12:39:34,510 INFO maxSessionTimeout set to 60000 (org.apache.zookeeper.server.ZooKeeperServer)
2021-04-22 12:39:34,511 INFO Created server with tickTime 3000 minSessionTimeout 6000 maxSessionTimeout 60000 dataDir /tmp/zookeeper/version-2 snapshot /tmp/zookeeper/version-2 (org.apache.zookeeper.server.ZooKeeperServer)
2021-04-22 12:39:34,525 INFO Using org.apache.zookeeper.server.NIOServerCnxnFactory as server connection factory (org.apache.zookeeper.server.ServerCnxnFactory)
2021-04-22 12:39:34,530 INFO Configuring NIO connection handler with 105 sessionless connection timeout, 2 selector thread(s), 16 worker threads, and 64 kb direct buffers. (org.apache.zookeeper.server.NIOServerCnxnFactory)
2021-04-22 12:39:34,537 INFO Binding to port 0.0.0.0/0.0.0.0:2181 (org.apache.zookeeper.server.NIOServerCnxnFactory)
2021-04-22 12:39:34,559 INFO zookeeper.snapshotSizeFactor = 0.33 (org.apache.zookeeper.server.ZKDatabase)
2021-04-22 12:39:34,560 INFO Reading snapshot /tmp/zookeeper/version-2/snapshot.b5 (org.apache.zookeeper.server.persistence.FileSnap)
2021-04-22 12:39:34,575 WARN Exception caught close of session 0x0: ZooKeeperServer not running (org.apache.zookeeper.server.persistence.FileSnap)
2021-04-22 12:39:34,593 INFO Snapshotting: oxfcr to /tmp/zookeeper/version-2/snapshot.cf (org.apache.zookeeper.server.persistence.FileTxnSnaplog)
2021-04-22 12:39:34,638 INFO Using checkIntervalMs=60000 maxPerMinute=10000 (org.apache.zookeeper.server.ContainerManager)
```

# Inicialização Kafka server:

```
marcos@marcos-340XAA-350XAA-550XAA:~/Documentos/kafka_2.12-2.7.0$ bin/kafka-server-start.sh config/server.properties
```

```
1 import findspark
2 import math
```

Atividades Terminal

22 de abr 12:43

pt

marcos@marcos-340XAA-350XAA-550XAA: ~/Documentos/kafka\_2.12-2.7.0

Q

≡

-

□

×

marcos@marcos-340XAA-350XAA-550XAA: ~/Documentos/kafka\_2.12-2.7.0

×

marcos@marcos-340XAA-350XAA-550XAA: ~/Documentos/kafka\_2.12-2.7.0

×

▼

```
afka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:07,994] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-20 in 74 milliseconds, of which 74 milliseconds was spent in the scheduler. (k
afka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:07,994] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-23 in 75 milliseconds, of which 75 milliseconds was spent in the scheduler. (k
afka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:07,994] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-26 in 75 milliseconds, of which 75 milliseconds was spent in the scheduler. (k
afka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:07,995] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-29 in 75 milliseconds, of which 75 milliseconds was spent in the scheduler. (k
afka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:07,999] INFO Static member MemberMetadata(memberId=consumer-console-consumer-53841-1-86378c15-1842-4fb9-871c-e2e7979d39c9, groupInstanceId=Some(null), clientId=consumer-console-consumer-53841-1
-client-host-1/127.0.0.1, sessionTimeoutMs=10000, supportedProtocols=List(range), ).groupInstanceId of group console-consumer-53841 loaded with member id consumer-console-consumer-53841
-1-86378c15-1842-4fb9-871c-e2e7979d39c9 at generation 1. (kafka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:08,000] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-32 in 81 milliseconds, of which 76 milliseconds was spent in the scheduler. (k
afka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:08,000] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-35 in 80 milliseconds, of which 80 milliseconds was spent in the scheduler. (k
afka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:08,001] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-38 in 81 milliseconds, of which 80 milliseconds was spent in the scheduler. (k
afka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:08,001] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-0 in 81 milliseconds, of which 81 milliseconds was spent in the scheduler. (ka
fka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:08,001] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-3 in 81 milliseconds, of which 81 milliseconds was spent in the scheduler. (ka
fka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:08,002] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-6 in 82 milliseconds, of which 81 milliseconds was spent in the scheduler. (ka
fka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:08,002] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-9 in 82 milliseconds, of which 82 milliseconds was spent in the scheduler. (ka
fka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:08,002] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-12 in 82 milliseconds, of which 82 milliseconds was spent in the scheduler. (k
afka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:08,003] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-15 in 83 milliseconds, of which 82 milliseconds was spent in the scheduler. (k
afka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:08,003] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-18 in 83 milliseconds, of which 83 milliseconds was spent in the scheduler. (k
afka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:08,004] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-21 in 84 milliseconds, of which 83 milliseconds was spent in the scheduler. (k
afka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:08,004] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-24 in 84 milliseconds, of which 84 milliseconds was spent in the scheduler. (k
afka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:08,004] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-27 in 84 milliseconds, of which 84 milliseconds was spent in the scheduler. (k
afka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:08,005] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-30 in 85 milliseconds, of which 84 milliseconds was spent in the scheduler. (k
afka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:08,005] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-33 in 85 milliseconds, of which 85 milliseconds was spent in the scheduler. (k
afka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:08,005] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-36 in 85 milliseconds, of which 85 milliseconds was spent in the scheduler. (k
afka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:08,006] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-39 in 86 milliseconds, of which 85 milliseconds was spent in the scheduler. (k
afka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:08,006] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-42 in 85 milliseconds, of which 85 milliseconds was spent in the scheduler. (k
afka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:08,006] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-45 in 85 milliseconds, of which 85 milliseconds was spent in the scheduler. (k
afka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:08,006] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-48 in 85 milliseconds, of which 85 milliseconds was spent in the scheduler. (k
afka.coordinator.group.GroupMetadataManager)
```

## Criação do topic:

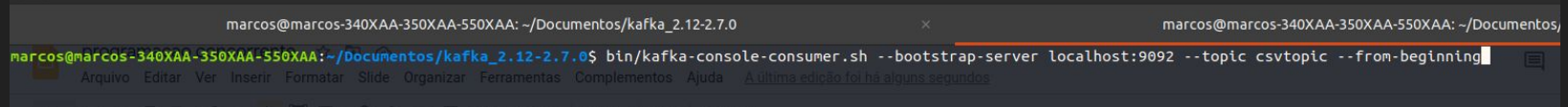
```
marcos@marcos-340XAA-350XAA-550XAA: ~/Documentos/kafka_2.12-2.7.0 x marcos@marcos-340XAA-350XAA-550XAA: ~/Documentos/kafka_2.12-2.7.0
marcos@marcos-340XAA-350XAA-550XAA:~/Documentos/kafka_2.12-2.7.0$ bin/kafka-topics.sh --create --bootstrap-server localhost:9092 --replication-factor 1 --partitions 1 --topic csvtopic2
[2021-04-22 12:43:07,994] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-20 in 74 milliseconds, of which 74 milliseconds was spent
afka.coordinator.group.GroupMetadataManager)
```

## Criação do Producer:

```
marcos@marcos-340XAA-350XAA-550XAA: ~/Documentos/kafka_2.12-2.7.0 x marcos@marcos-340XAA-350XAA-550XAA: ~/Documentos/kafka_2.12-2.7.0 x
marcos@marcos-340XAA-350XAA-550XAA:~/Documentos/kafka_2.12-2.7.0$ bin/kafka-console-producer.sh --broker-list localhost:9092 --topic csvtopic < /home/marcos/Documentos/prog_concorrente/yellow_tripdata_2016-01.csv
[2021-04-22 12:43:07,994] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-20 in 74 milliseconds, of which 74 milliseconds was spent in the scheduler. (k
afka.coordinator.group.GroupMetadataManager)
[2021-04-22 12:43:07,994] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-23 in 75 milliseconds, of which 75 milliseconds was spent in the scheduler. (k
```



# Inicialização do consumer:



The image shows a terminal window with a dark background. The title bar at the top reads "marcos@marcos-340XAA-350XAA-550XAA: ~/Documentos/kafka\_2.12-2.7.0". The terminal content shows a prompt "marcos@marcos-340XAA-350XAA-550XAA: ~/Documentos/kafka\_2.12-2.7.0\$" followed by the command "bin/kafka-console-consumer.sh --bootstrap-server localhost:9092 --topic csvtopic --from-beginning". Below the command line, a menu bar is visible with options: "Arquivo", "Editar", "Ver", "Inserir", "Formatar", "Slide", "Organizar", "Ferramentas", "Complementos", "Ajuda", and a link "A última edição foi há alguns segundos".

```
marcos@marcos-340XAA-350XAA-550XAA: ~/Documentos/kafka_2.12-2.7.0$ bin/kafka-console-consumer.sh --bootstrap-server localhost:9092 --topic csvtopic --from-beginning
```

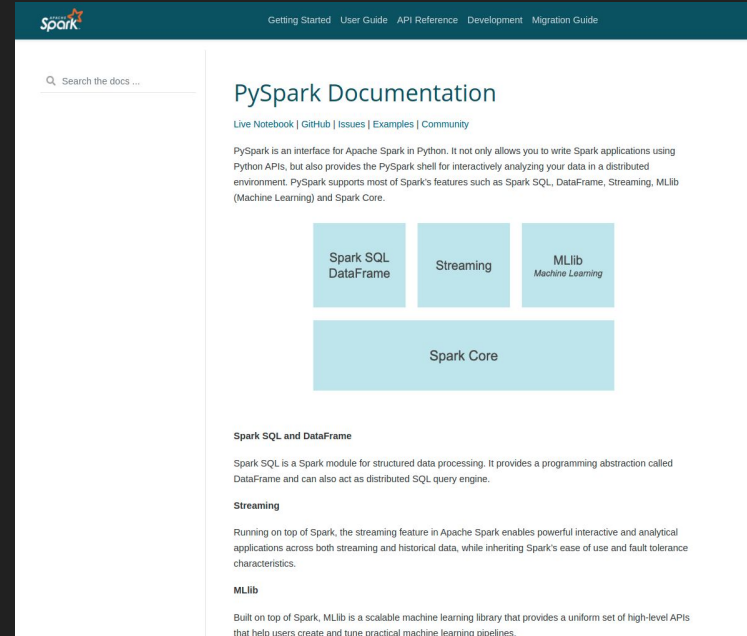
# Spark

Linguagem utilizada:

Python 2.7

- Pyspark

\* `pyspark.streaming.kafka`



Versão utilizada: spark-2.4.5

# Funcionamento:

```

kafka-0.8.2.11;2.4.5 testspark2.py /home/marcos/Documents/trab_final_programacao_terceira_unid5 /spark-2.4.5-bin-hadoop2.7/bin/spark-submit --packages org.apache.spark:spark-streaming-
Ivy Default Cache set to: /home/marcos/.ivy2/cache
The jars for the packages stored in: /home/marcos/.ivy2/jars
:: loading settings :: url = jar:file:/home/marcos/Documents/spark-2.4.5-bin-hadoop2.7/jars/ivy-2.4.0.jar!/org/apache/ivy/core/settings/ivysettings.xml
org.apache.spark#spark-streaming-kafka-0.8.2.11 added as a dependency
:: resolving dependencies :: org.apache.spark#spark-submit-parent-89fe4a74-bb1a-4e3e-aef0-dbaa6ffbe654;1.0
  confs: [default]
    found org.apache.spark#spark-streaming-kafka-0.8.2.11;2.4.5 in central
    found org.apache.kafka#kafka_2.11;0.8.2.1 in central
    found org.scala-lang.modules#scala-xml_2.11;1.0.2 in central
    found com.yammer.metrics#metrics-core;2.2.0 in central
    found org.slf4j#slf4j-api;1.7.16 in central
    found org.scala-lang.modules#scala-parser-combinators_2.11;1.1.0 in central
    found com.101tec#zkclient;0.3 in central
    found log4j#log4j;1.2.17 in central
    found org.apache.kafka#kafka-clients;0.8.2.1 in central
    found net.jpountz.lz4#lz4;1.2.0 in central
    found org.xerial.snappy#snappy-java;1.1.7.3 in central
    found org.spark-project.spark#unused;1.0.0 in central
:: resolution report :: resolve 692ms :: artifacts dl 12ms
  :: modules in use:
    com.101tec#zkclient;0.3 from central in [default]
    com.yammer.metrics#metrics-core;2.2.0 from central in [default]
    log4j#log4j;1.2.17 from central in [default]
    net.jpountz.lz4#lz4;1.2.0 from central in [default]
    org.apache.kafka#kafka-clients;0.8.2.1 from central in [default]
    org.apache.kafka#kafka_2.11;0.8.2.1 from central in [default]
    org.apache.spark#spark-streaming-kafka-0.8.2.11;2.4.5 from central in [default]
    org.scala-lang.modules#scala-parser-combinators_2.11;1.1.0 from central in [default]
    org.scala-lang.modules#scala-xml_2.11;1.0.2 from central in [default]
    org.slf4j#slf4j-api;1.7.16 from central in [default]
    org.spark-project.spark#unused;1.0.0 from central in [default]
    org.xerial.snappy#snappy-java;1.1.7.3 from central in [default]
  -----
  | conf | number | modules | search | downloaded | evicted | | artifacts | |
  |-----|-----|-----|-----|-----|-----| |-----|
  | default | 12 | 0 | 0 | 0 | 0 | | 12 | 0 |
  -----
  :: format(x.x, x.y, x.value, x.distance))
:: retrieving :: org.apache.spark#spark-submit-parent-89fe4a74-bb1a-4e3e-aef0-dbaa6ffbe654
  confs: [default]
  0 artifacts copied, 12 already retrieved (0kB/12ms)
21/04/22 12:30:39 WARN Utils: Your hostname, marcos-340XAA-350XAA-550XAA resolves to a loopback address: 127.0.1.1; using 192.168.25.203 instead (on interface wlp3s0)
21/04/22 12:30:39 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
21/04/22 12:30:39 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
>>>>>>> CONSUMINDO KAFKA <<<<<<<<
[26.0, 24.0, 27.0]
The value classified to unknown point is: 25.6666666667
>>>>>>> FIM DO CONSUMO <<<<<<<
>>>>>>> B09W60-3F0V66-EC0V66- /Documents/tra... /spark-2.4.5-bin-hadoop2.7/bin/spark-submit --packages org.apache.spark:spark-streaming-

```