

Instalar KAFKA

# Primeiros passos

1. Acessar máquina virtual
2. Acessar terminal
3. Atualizar Ubuntu 16.04
  - `sudo apt-get update -y`
  - `sudo apt-get upgrade -y`
4. Esperar alguns minutos

# Instalar o Java

```
sudo add-apt-repository -y ppa:webupd8team/java
```

```
gpg: keyring `/tmp/tmpkjr4mnm/secring.gpg' created
gpg: keyring `/tmp/tmpkjr4mnm/pubring.gpg' created
gpg: requesting key EEA14886 from hkp server keyserver.ubuntu.com
gpg: /tmp/tmpkjr4mnm/trustdb.gpg: trustdb created
gpg: key EEA14886: public key "Launchpad VLC" imported
gpg: no ultimately trusted keys found
gpg: Total number processed: 1
gpg:      imported: 1 (RSA: 1)
OK
```

```
sudo apt-get update
```

```
sudo apt-get install oracle-java8-installer -y
```

```
sudo java -version
```

```
java version "1.8.0_66"
Java(TM) SE Runtime Environment (build 1.8.0_66-b17)
Java HotSpot(TM) 64-Bit Server VM (build 25.66-b17, mixed mode)
```

# Instalar o ZooKeeper

- `sudo apt-get install zookeeperd`

Para testar a instalação:

- `netstat -ant | grep :2181`

```
tcp6      0      0 :::2181          :::*              LISTEN
```

# Instalar o Kafka

## Baixar o kafka:

- `wget https://archive.apache.org/dist/kafka/0.10.0.1/kafka_2.10-0.10.0.1.tgz`

## Criar um diretório para a instalação do kafka:

- `sudo mkdir kafka`

## Extrair os dados para o diretório:

- `sudo tar -xvf kafka_2.10-0.10.0.1.tgz -C kafka/`

# Inicializar o Servidor Kafka

- `cd kafka/kafka_2.10-0.10.0.1/bin`
- `sudo ./kafka-server-start.sh`  
`~/kafka/kafka_2.10-0.10.0.1/config/server.properties`

```
[2016-08-22 21:43:48,279] WARN No meta.properties file under dir /tmp/kafka-logs/meta.properties
```

```
(kafka.server.BrokerMetadataCheckpoint)
```

```
[2016-08-22 21:43:48,516] INFO Kafka version : 0.10.0.1 (org.apache.kafka.common.utils.AppInfoParser)
```

```
[2016-08-22 21:43:48,525] INFO Kafka commitId : a7a17cdec9eaa6c5 (org.apache.kafka.common.utils.AppInfoParser)
```

```
[2016-08-22 21:43:48,527] INFO [Kafka Server 0], started (kafka.server.KafkaServer)
```

```
[2016-08-22 21:43:48,555] INFO New leader is 0 (kafka.server.ZookeeperLeaderElector$LeaderChangeListener)
```

# Criar novo tópico

7. Inicializar outra janela de terminal

8. Acessar diretório KAFKA

- `cd kafka/kafka_2.10-0.10.0.1/bin`

9. Criar e listar um novo tópico

- `./kafka-topics.sh --create`  
`--zookeeper localhost:2181`  
`--replication-factor 1 --partitions 1`  
`--topic any`
- `./kafka-topics.sh --list --zookeeper`  
`localhost:2181`

# Gerar dados de um tópico

## 8. Executar o *producer*

- `./kafka-console-producer.sh --broker-list localhost:9092 --topic any`



# Consumir dados de um tópico

9. Inicializar outra janela de terminal

10. Acessar diretório KAFKA

- `cd kafka/kafka_2.10-0.10.0.1/bin`

11. Obter os dados de um determinado tópico

```
./kafka-console-consumer.sh --zookeeper  
localhost:2181 --bootstrap-server localhost:9092  
--topic any --from-beginning
```

# Produtor x Consumidor de dados

```

maria_dev@sandbox:/usr/hdp/current/kafka-broker/bin
Using username "maria_dev".
maria_dev@127.0.0.1's password:
Last login: Mon Feb 13 19:31:16 2017 from 10.0.2.2
[maria_dev@sandbox ~]$ cd /usr/hdp/current/kafka-broker/bin
[maria_dev@sandbox bin]$ ./kafka-console-consumer.sh --bootstrap-server sandbox.hortonworks.com:6667 --zookeeper localhost:2181 --topic fred --from-beginning
{metadata.broker.list=sandbox.hortonworks.com:6667, request.timeout.ms=30000,
ient.id=console-consumer-89447, security.protocol=PLAINTEXT}
This is a line of data
I am sending this on the fred topic
Here is yet another line
[]

kafka-consumer-perf-test.sh      kafka-zookeeper-run-class.sh
kafka-mirror-maker.sh            windows
kafka-preferred-replica-election.sh  zookeeper-security-migration.sh
kafka-producer-perf-test.sh        zookeeper-server-start.sh
kafka-reassign-partitions.sh        zookeeper-server-stop.sh
kafka-replay-log-producer.sh        zookeeper-shell.sh
[maria_dev@sandbox bin]$ ./kafka-topics.sh --create --zookeeper sandbox.hortonwo
rks.com:2181 --replication-factor 1 --partitions 1 --topic fred
Created topic "fred".
[maria_dev@sandbox bin]$ ./kafka-topics.sh --list --zookeeper sandbox.hortonwork
s.com:2181
ATLAS_ENTITIES
ATLAS_HOOK
__consumer_offsets
fred
log-topic
log-topic2
test - marked for deletion
[maria_dev@sandbox bin]$ ./kafka-console-producer.sh --broker-list sandbox.horto
nworks.com:6667 --topic fred
This is a line of data
I am sending this on the fred topic
Here is yet another line
```

Para sair: ctrl+C

# KAFKA - Atividade 1

# Atividade 1

1. Crie um tópico com o nome de vocês;
2. Liste os tópicos e verifique se o seu foi criado;
3. Gere dados para o tópico criado.



# KAFKA-PYTHON - Atividade 2

# Produtor/consumidor

## 1. Baixar código python produtor/consumidor

Terminal 2

```
wget  
https://raw.githubusercontent.com/antoniomralmeida/streaming/master/producer_consumer.py
```

## 2. Instalar dependência do kafka no python

- `sudo apt-get install python-pip`
- `pip install kafka-python`

## 2. Rodar script

- `python producer_consumer.py`

## Atividade 2 (vale pt)



1. Imprima apenas o conteúdo da tupla;
2. Gere dados de quatro producers simultâneos;
3. Aumente a frequência de geração das tuplas (geração mais rápida);
4. Filtre e imprima apenas por tuplas que possuem valores de peso maiores que 80;
5. Filtre e imprima apenas por tuplas que possuem valores de IMC acima de 35 ( $IMC = peso/altura^2$ ).

Fim

