

VINÍCIUS HENRIQUE DOS SANTOS

POSTECH

MACHINE LEARNING ENGINEERING

BIG DATA PIPELINES

AULA 03

SUMÁRIO

O QUE VEM POR AÍ?	3
HANDS ON	4
SAIBA MAIS	5
O QUE VOCÊ VIU NESTA AULA?	25
REFERÊNCIAS.....	26

EMSE

O QUE VEM POR AÍ?

Já parou para pensar no volume de dados que você gera a todo instante? Todas as vezes em que manda uma mensagem via WhatsApp, acessa o Instagram e dá um like em uma foto de uma amiga ou faz um pix no seu banco? Todas essas informações têm algo em comum: elas são geradas e processadas instantaneamente!

Nessa aula vamos estudar e entender juntos o cenário em que precisamos processar dados em tempo real, à medida que eles são criados. Para isso aperte os cintos, pois usaremos ferramentas que recebem e enviam informações de forma muito veloz.

Embarque comigo e vamos pilotar juntos esses foguetes chamados Apache Kafka, Apache Zookeeper e Apache Flink: com eles, nossos dados chegam em qualquer lugar do mundo em um piscar de olhos!

HANDS ON

Te convido a realizar a instalação das ferramentas Kafka e Apache Flink para que possa interagir com elas através do Python ou qualquer outra linguagem de programação da sua preferência, buscando os seguintes objetivos:

- Enviar e receber mensagens no próprio Apache Kafka via console.
- Desenvolver um código para fazer o papel de “data producer” de forma que você enxergue o Kafka consumindo essas informações em tempo real.
- Usar o Apache Flink para consumir informações em tempo real através da “table API”.

SAIBA MAIS

Configurando o ambiente para o Apache Kafka

Faça o download do virtualizador de sua preferência. Aqui eu seguirei com o [Virtual Box](#), visto que não tem custos e é compatível com a grande maioria dos sistemas operacionais.



Figura 1 – VirtualBox
Fonte: elaborado pelo autor (2024)

A versão do Linux que usaremos é o [Ubuntu](#) 22.04.3 LTS (a mais recente na data em que esse curso está sendo desenvolvido); já realize seu download.

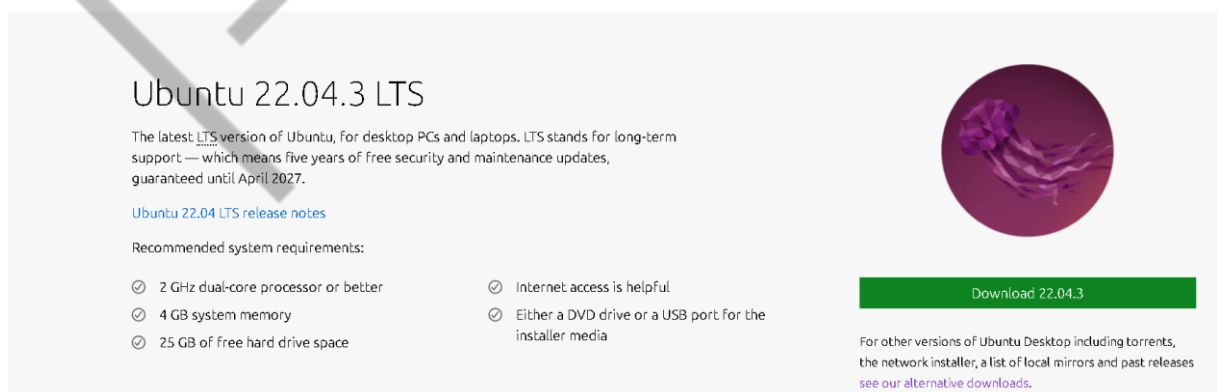


Figura 2 – Ubuntu 22.04.3 LTS
Fonte: elaborado pelo autor (2024)

Realize a instalação padrão do Virtual Box, abra o aplicativo e clique em “novo”.

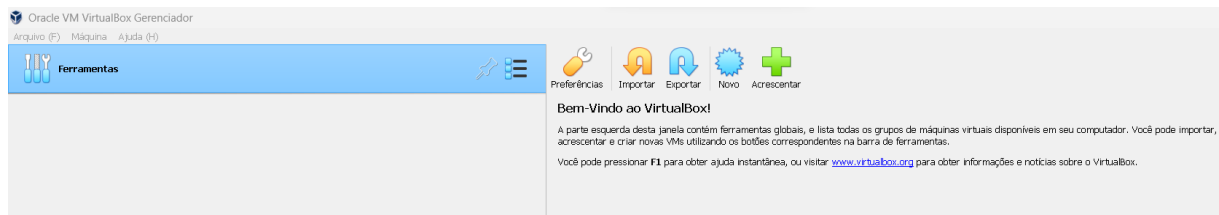


Figura 3 – Tela inicial do VirtualBox
Fonte: elaborado pelo autor (2024)

Preencha as informações necessárias apontando para a “.iso” que você baixou do Ubuntu.

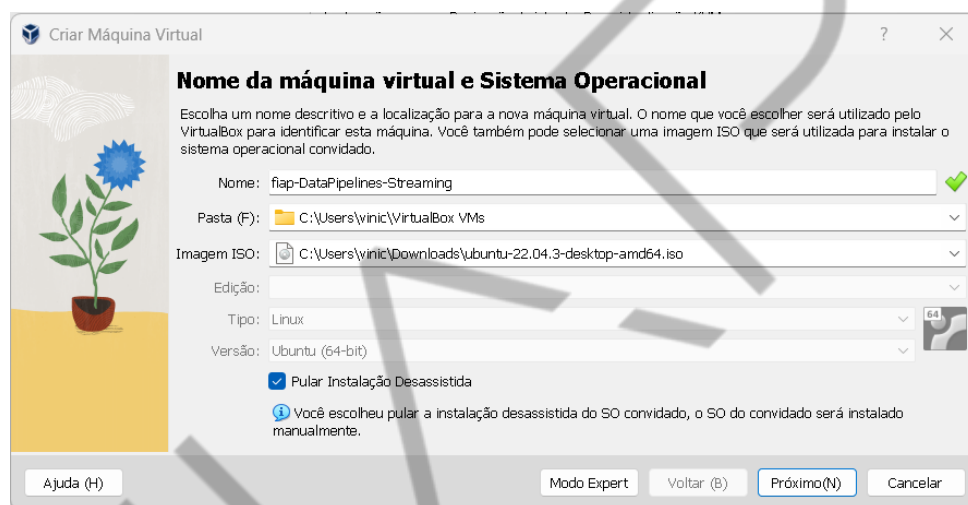


Figura 4 – Nome da máquina virtual e Sistema Operacional
Fonte: elaborado pelo autor (2024)

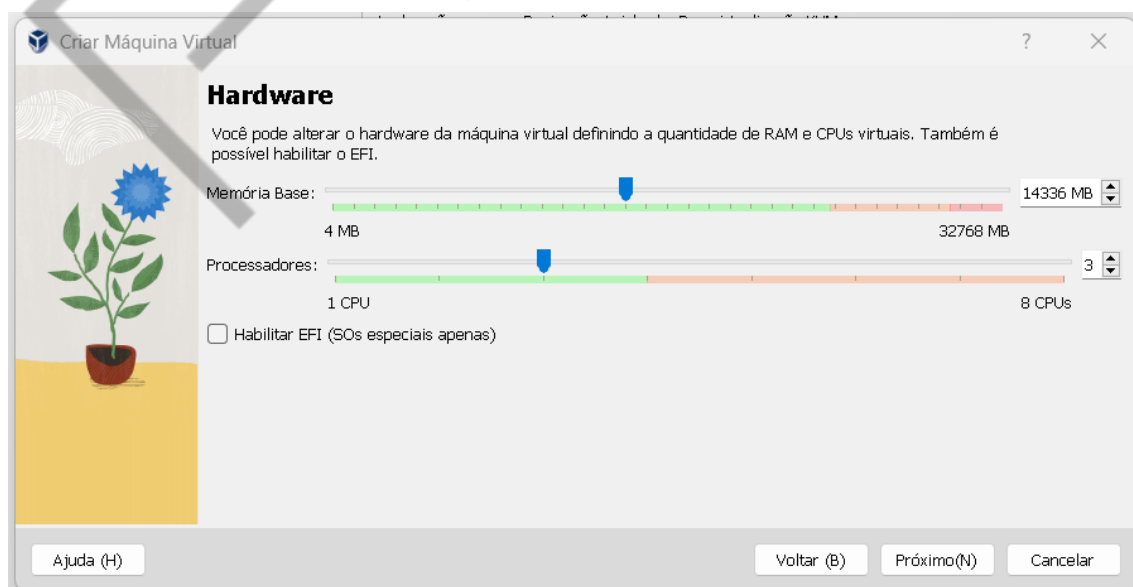


Figura 5 – Hardware
Fonte: elaborado pelo autor (2024)



Figura 6 – Disco Rígido Virtual
Fonte: elaborado pelo autor (2024)

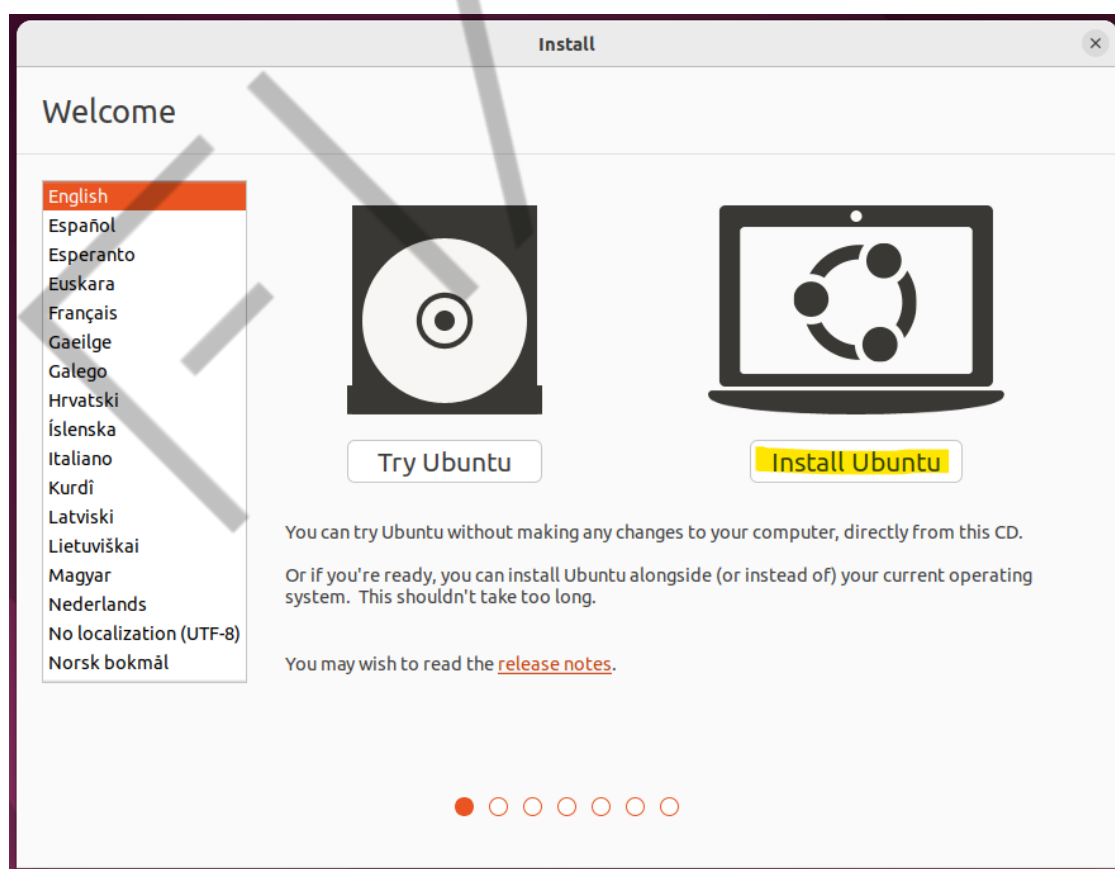


Figura 7 – Install Ubuntu
Fonte: elaborado pelo autor (2024)

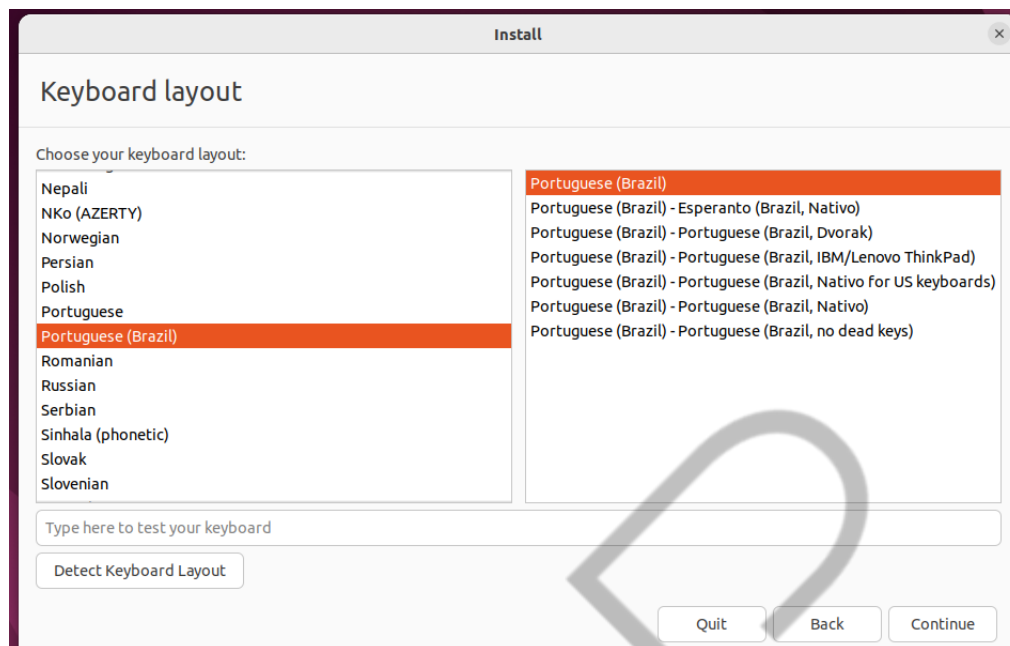


Figura 8 – Escolha da linguagem
Fonte: elaborado pelo autor (2024)

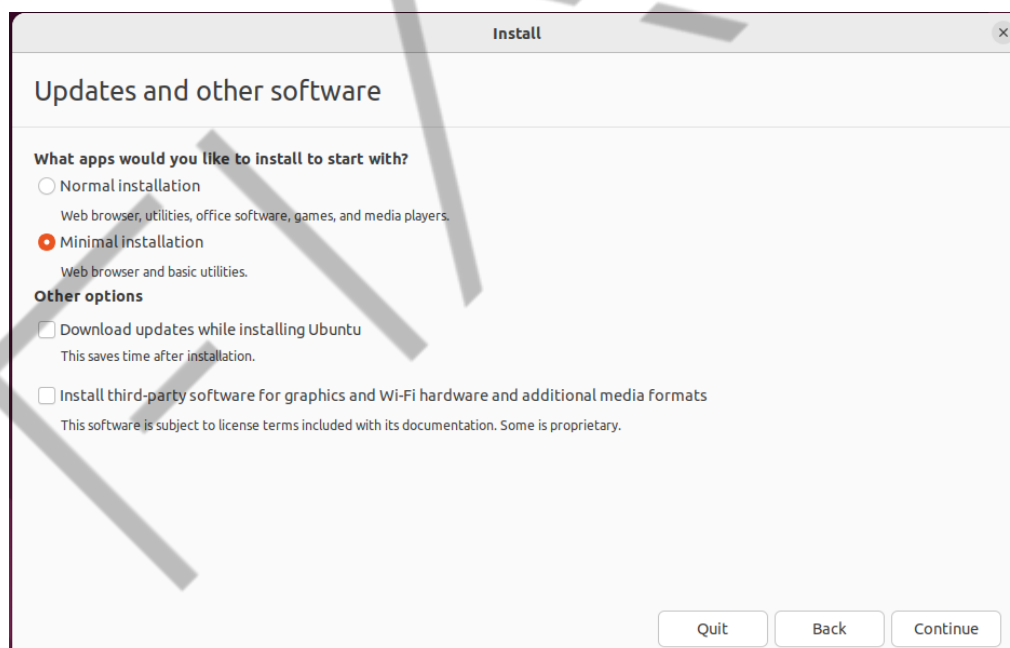


Figura 9 – Instalação mínima
Fonte: elaborado pelo autor (2024)

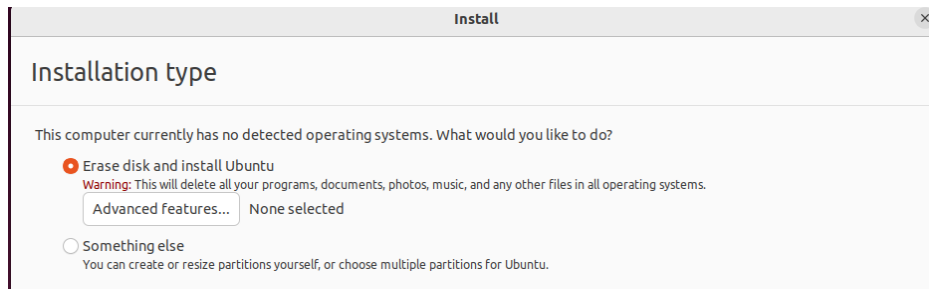


Figura 10 – Tipo de instalação
Fonte: elaborado pelo autor (2024)

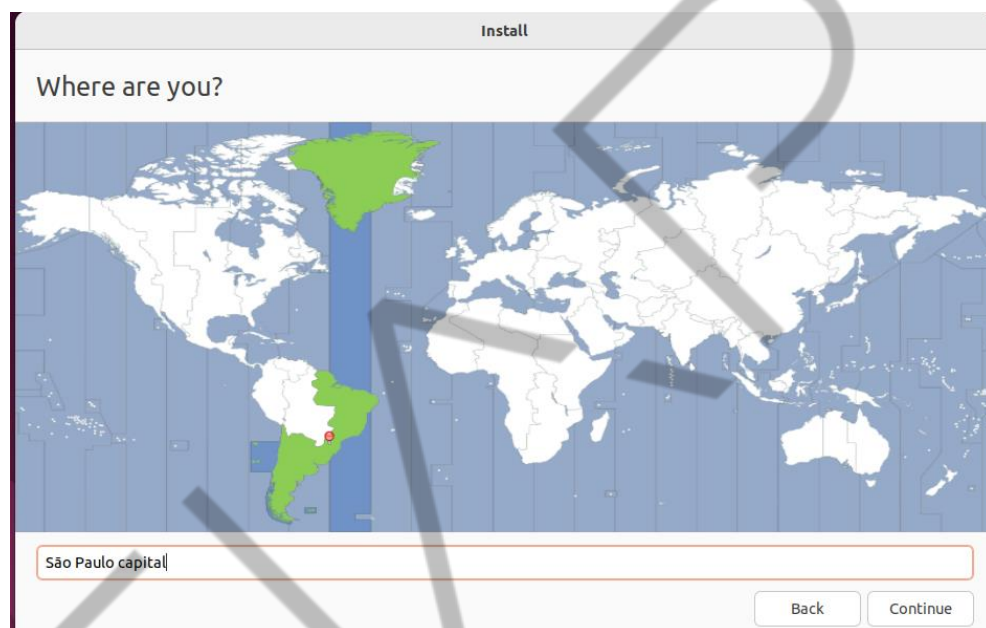


Figura 11 – Localização
Fonte: elaborado pelo autor (2024)

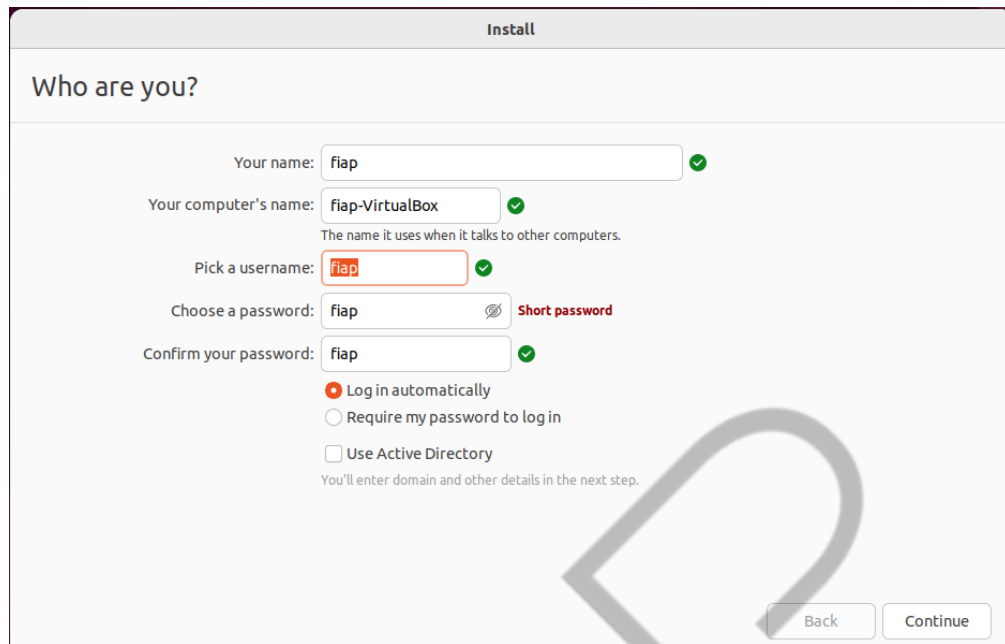


Figura 12 – Escolhendo um usuário
Fonte: elaborado pelo autor (2024)

Agora que o Ubuntu iniciou, vamos atualizar os pacotes e garantir que tudo que precisamos está com a versão atualizada:

```
sudo apt update
```

```
sudo apt upgrade
```

Depois que os arquivos do sistema operacional forem copiados e instalados, vá em “Dispositivos” no Virtual box e selecione a opção destacada:

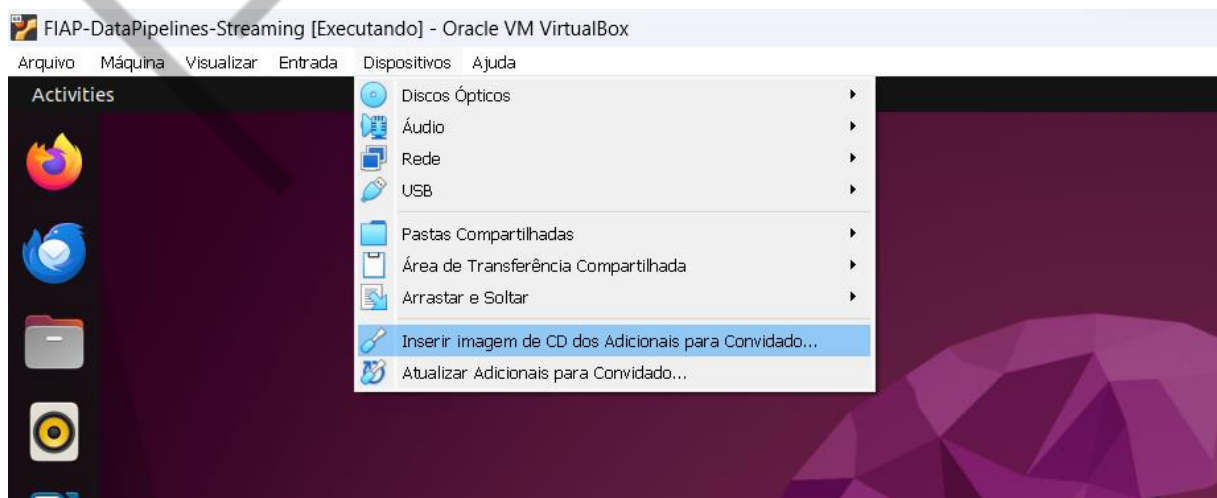


Figura 13 – Inserir Imagem de CD dos Adicionais para Convidado
Fonte: elaborado pelo autor (2024)

Após isso, desligue a máquina virtual.

Com a VM desligada, vá em configurações e faça as alterações para que você tenha os mesmos parâmetros da figura 14 e inicie a máquina novamente.

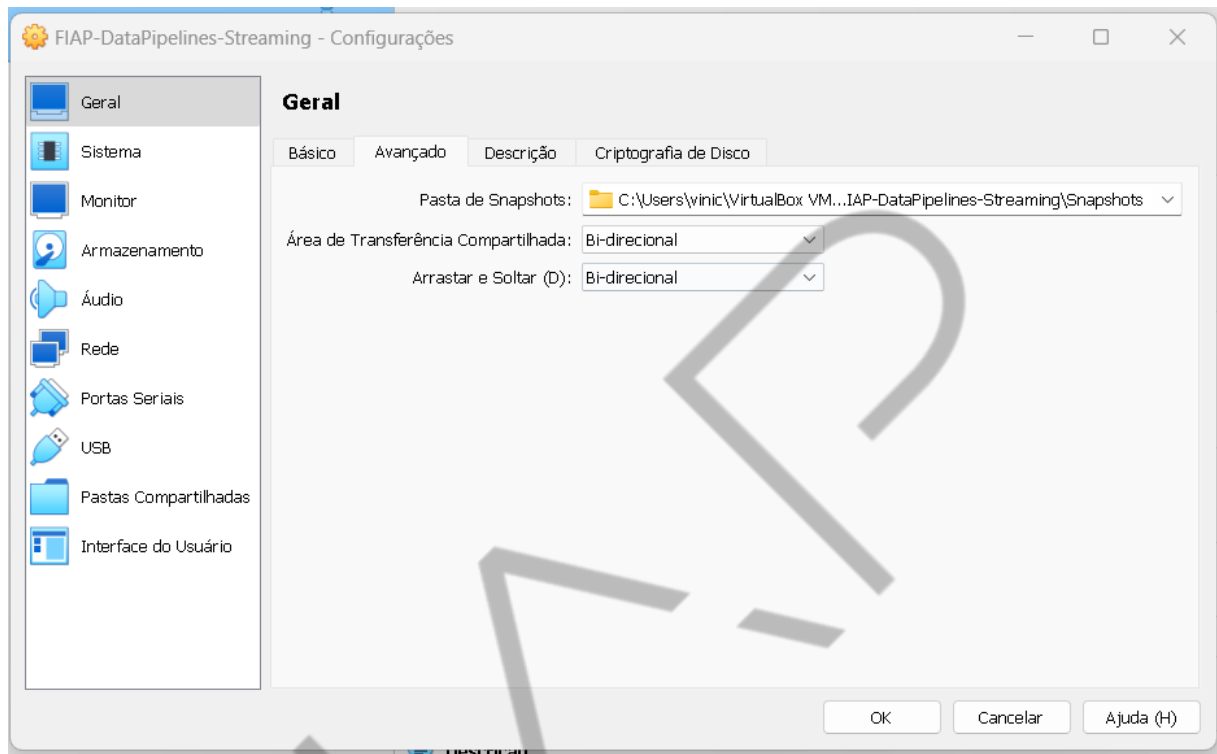


Figura 14 – Parâmetros indicados
Fonte: elaborado pelo autor (2024)

Abra o terminal (atalho Control + Alt + T) do Linux para executarmos os comandos necessários:

```
sudo apt update
```

```
sudo apt upgrade
```

```
sudo apt install build-essential gcc make perl dkms curl tcl
```

Abra o catálogo de software do Ubuntu e instale o VSCode.

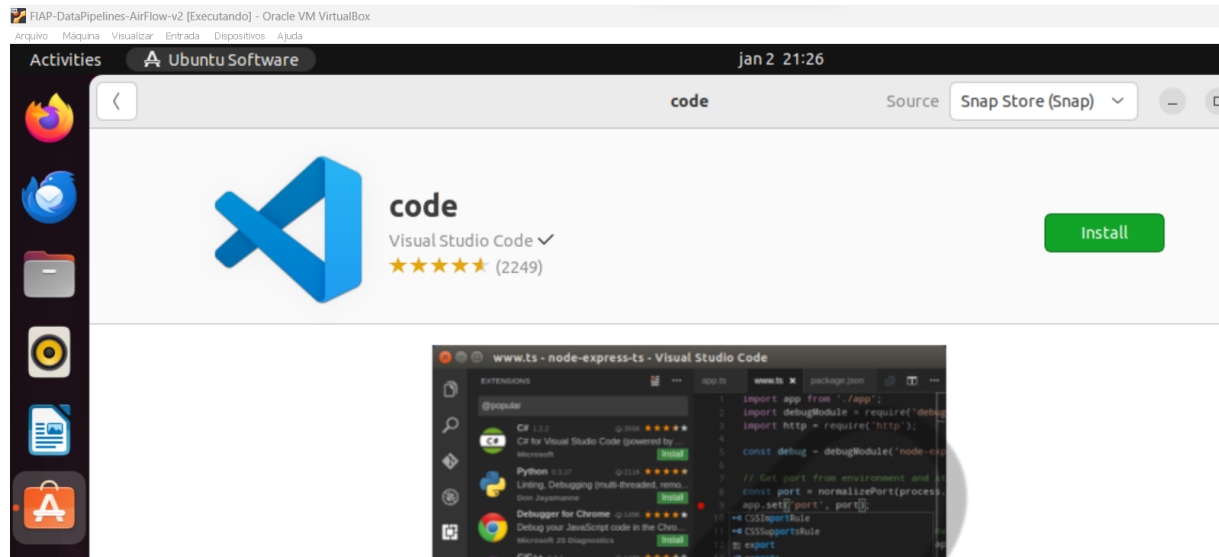


Figura 15 – Instalação do VSCode
Fonte: elaborado pelo autor (2024)

Com o VS Code aberto, procure pela extensão do Python:

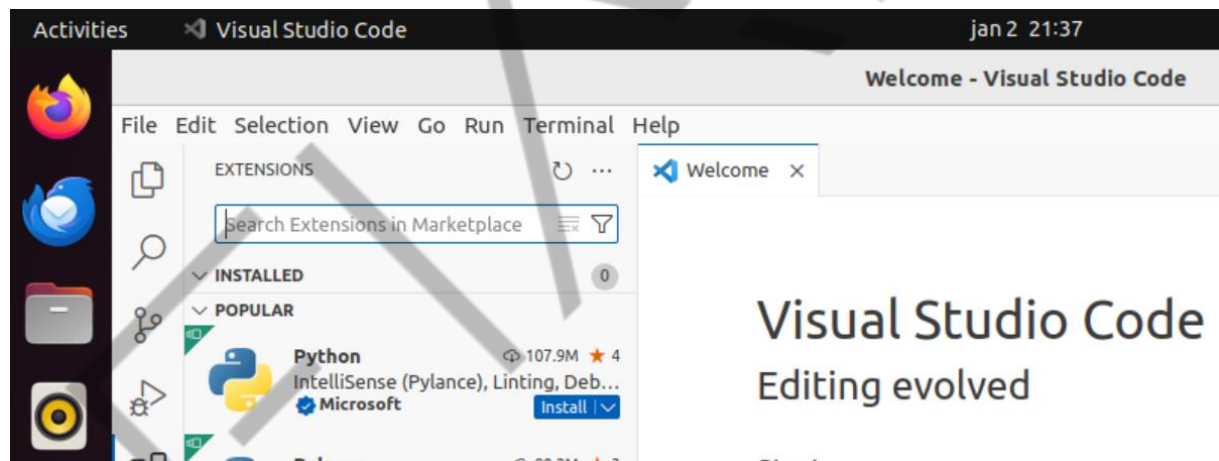


Figura 16 – Instalação do Python
Fonte: elaborado pelo autor (2024)

Instale o Java 11 pelo comando “sudo apt install openjdk-11-jdk” no terminal.

```
fiap@fiap-VirtualBox: ~  
fiap@fiap-VirtualBox:~$ sudo apt install openjdk-11-jdk  
[sudo] password for fiap:  
Reading package lists... Done  
Building dependency tree... Done  
Reading state information... Done  
The following additional packages will be installed:  
  ca-certificates-java fonts-dejavu-extra java-common libatk-wrapper-java  
  libatk-wrapper-java-jni libice-dev libpthread-stubs0-dev libsm-dev  
  libx11-dev libxau-dev libxcb1-dev libxdmcp-dev libxt-dev  
  openjdk-11-jdk-headless openjdk-11-jre openjdk-11-jre-headless x11proto-dev  
  xorg-sgml-doctools xtrans-dev  
Suggested packages:  
  default-jre libice-doc libsm-doc libx11-doc libxcb-doc libxt-doc  
  openjdk-11-demo openjdk-11-source visualvm fonts-ipafont-gothic  
  fonts-ipafont-mincho fonts-wqy-microhei | fonts-wqy-zenhei  
The following NEW packages will be installed:  
  ca-certificates-java fonts-dejavu-extra java-common libatk-wrapper-java  
  libatk-wrapper-java-jni libice-dev libpthread-stubs0-dev libsm-dev
```

Figura 17 – Instalando o Java 11
Fonte: elaborado pelo autor (2024)

Verifique sua versão:

```
fiap@fiap-VirtualBox: ~  
fiap@fiap-VirtualBox:~$ java --version  
openjdk 11.0.21 2023-10-17  
OpenJDK Runtime Environment (build 11.0.21+9-post-Ubuntu-0ubuntu122.04)  
OpenJDK 64-Bit Server VM (build 11.0.21+9-post-Ubuntu-0ubuntu122.04, mixed mode,  
sharing)  
fiap@fiap-VirtualBox:~$
```

Figura 18 – Verificando a versão
Fonte: elaborado pelo autor (2024)

Clique na imagem do “CD” que apareceu na barra de ferramentas no Ubuntu.

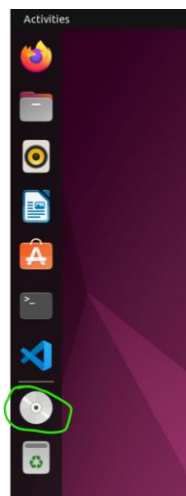


Figura 19 – Ícone de CD
Fonte: elaborado pelo autor (2024)

Em seguida, na janela, abra um terminal dentro dessa pasta.

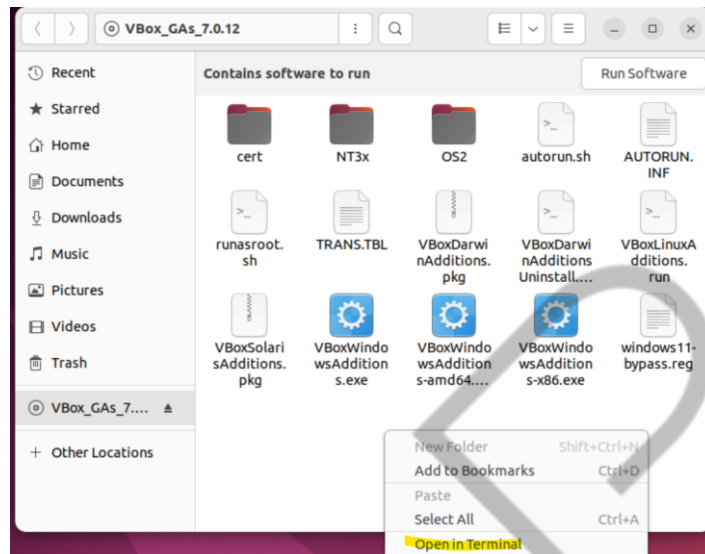


Figura 20 – Abrindo um terminal
Fonte: elaborado pelo autor (2024)

Execute os seguintes comandos no terminal:

```
sudo ./VboxLinuxAdditions.run
```

```
flap@FIAP-DataPipelines-Streaming: /media/flap/VBox_GAs_7.0.12
flap@FIAP-DataPipelines-Streaming:/media/flap/VBox_GAs_7.0.12$ sudo ./VboxLinuxAdditions.run
[sudo] password for flap:
sudo: ./VboxLinuxAdditions.run: command not found
flap@FIAP-DataPipelines-Streaming:/media/flap/VBox_GAs_7.0.12$ sudo ./VBoxLinuxAdditions.run
Verifying archive integrity... 100% MD5 checksums are OK. All good.
Uncompressing VirtualBox 7.0.12 Guest Additions for Linux 100%
VirtualBox Guest Additions installer
Removing installed version 7.0.12 of VirtualBox Guest Additions...
update-initramfs: Generating /boot/initrd.img-6.2.0-26-generic
Copying additional installer modules ...
Installing additional modules ...
VirtualBox Guest Additions: Starting.
VirtualBox Guest Additions: Setting up modules
VirtualBox Guest Additions: Building the VirtualBox Guest Additions kernel
modules. This may take a while.
VirtualBox Guest Additions: To build modules for other installed kernels, run
VirtualBox Guest Additions: /sbin/rcvboxadd quicksetup <version>
VirtualBox Guest Additions: or
VirtualBox Guest Additions: /sbin/rcvboxadd quicksetup all
VirtualBox Guest Additions: Building the modules for kernel 6.5.0-17-generic.
```

Figura 21 – Execução de comandos no terminal
Fonte: elaborado pelo autor (2024)

Feche o terminal e reinicie a máquina virtual. Quando ela religar, você conseguirá copiar e colar texto ou arquivos entre máquina hospedeira e convidado, o que facilitará bastante as coisas daqui para frente.

Configuração do ambiente dentro da Máquina Virtual

Pela sua máquina virtual, acesse o link de download do [Apache Kafka](#).

Até a presente data estamos na versão 3.6.1; recomendo que você use a mesma.

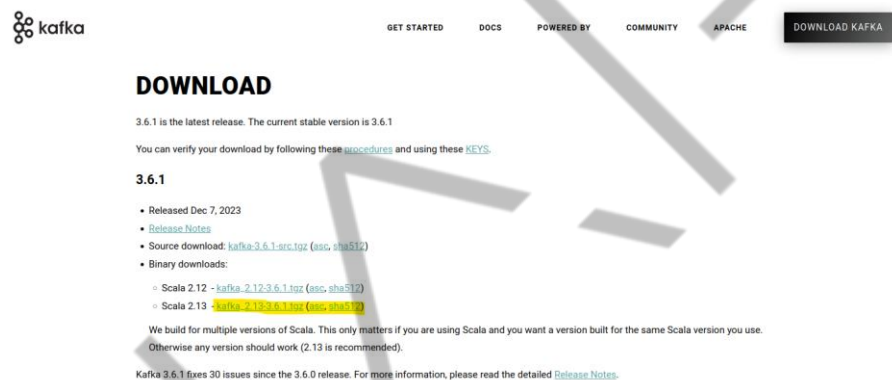


Figura 22 – [Apache Kafka](#).
Fonte: elaborado pelo autor (2024)

Clique no link do Scala 3.13 e aguarde o download.

Navegue até sua pasta de downloads e copie o arquivo baixado para a pasta “home”:

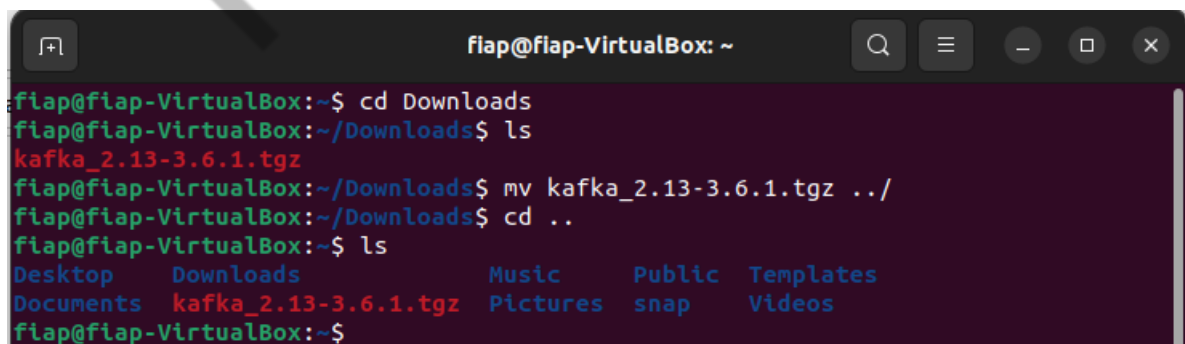
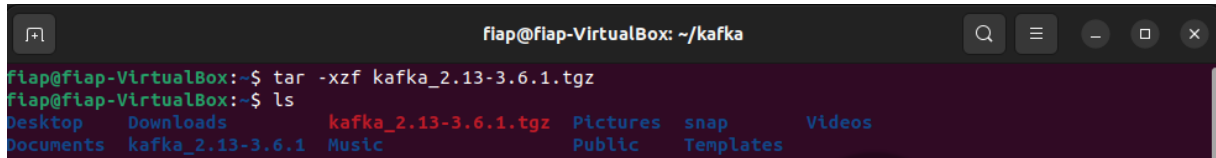


Figura 23 – Copiando o arquivo baixado para a pasta “home”
Fonte: elaborado pelo autor (2024)

Agora, nessa mesma pasta descompacte o arquivo usando o programa tar. Note que ele criará uma pasta com o nome da versão que iremos usar, mas, para facilitar os comandos, renomeie essa pasta para manter apenas “kafka”.

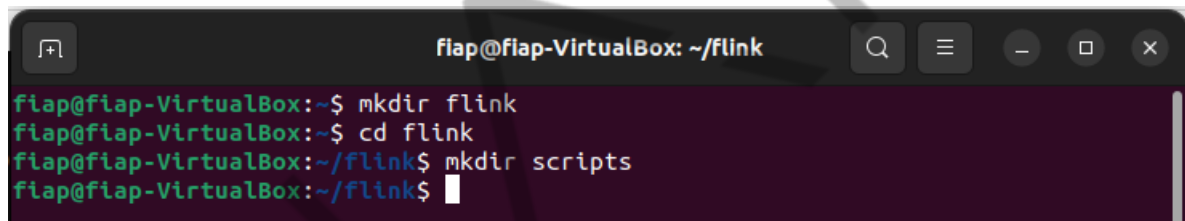
Para facilitar, crie uma pasta de scripts dentro da pasta “kafka”:



```
fiap@fiap-VirtualBox: ~/kafka
fiap@fiap-VirtualBox:~$ tar -xzf kafka_2.13-3.6.1.tgz
fiap@fiap-VirtualBox:~$ ls
Desktop  Downloads  kafka_2.13-3.6.1.tgz  Pictures  snap  Videos
Documents  kafka_2.13-3.6.1  Music  Public  Templates
```

Figura 24 – Criando uma pasta de scripts dentro da pasta “kafka”
Fonte: elaborado pelo autor (2024)

Renomeie a pasta kafka_2.13-3.6.1 para “kafka”. Dentro dela, crie uma pasta chamada scripts. Por fim, crie uma pasta chamada “flink” e dentro dela uma pasta chamada “scripts”.



```
fiap@fiap-VirtualBox: ~/flink
fiap@fiap-VirtualBox:~$ mkdir flink
fiap@fiap-VirtualBox:~$ cd flink
fiap@fiap-VirtualBox:~/flink$ mkdir scripts
fiap@fiap-VirtualBox:~/flink$
```

Figura 25 – Pasta “scripts”
Fonte: elaborado pelo autor (2024)

Instale o java: `sudo apt install openjdk-11-jre-headless`


```
fiap@fiap-VirtualBox:~/kafka/bin$ sudo apt install openjdk-11-jre-headless
[sudo] password for fiap:
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following additional packages will be installed:
  ca-certificates-java java-common
Suggested packages:
  default-jre fonts-dejavu-extra fonts-ipafont-gothic fonts-ipafont-mincho
  fonts-wqy-microhei | fonts-wqy-zenhei
The following NEW packages will be installed:
  ca-certificates-java java-common openjdk-11-jre-headless
0 upgraded, 3 newly installed, 0 to remove and 5 not upgraded.
Need to get 42,5 MB of archives.
After this operation, 176 MB of additional disk space will be used.
Do you want to continue? [Y/n] y
Get:1 http://br.archive.ubuntu.com/ubuntu jammy/main amd64 java-common all 0.72b
uild2 [6.782 B]
Get:2 http://br.archive.ubuntu.com/ubuntu jammy-updates/main amd64 openjdk-11-jr
e-headless amd64 11.0.21+9-0ubuntu1~22.04 [42,5 MB]
Get:3 http://br.archive.ubuntu.com/ubuntu jammy-updates/main amd64 ca-certificat
es-java all 20190909ubuntu1.2 [12,1 kB]
Fetched 42,5 MB in 3s (16,1 MB/s)
Selecting previously unselected package java-common.
```

Figura 26 – Instalando o Java
Fonte: elaborado pelo autor (2024)

Instale o Python 3.8:

```
sudo apt update
```

```
sudo apt install python3.8
```

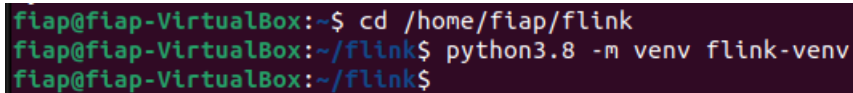
```
sudo apt install python3.8-venv
```

```
fiap@fiap-VirtualBox: ~$ sudo apt install python3.8 python3.8-venv
Do you want to continue? [Y/n] y
Get:1 https://ppa.launchpadcontent.net/deadsnakes/ppa/ubuntu jammy/main amd64 py
thon3.8-lib2to3 all 3.8.18-1+jammy1 [126 kB]
Get:2 https://ppa.launchpadcontent.net/deadsnakes/ppa/ubuntu jammy/main amd64 py
thon3.8-distutils all 3.8.18-1+jammy1 [193 kB]
Get:3 https://ppa.launchpadcontent.net/deadsnakes/ppa/ubuntu jammy/main amd64 py
thon3.8-venv amd64 3.8.18-1+jammy1 [4.311 kB]
Fetched 4.630 kB in 3s (1.429 kB/s)
Selecting previously unselected package python3.8-lib2to3.
(Reading database ... 187445 files and directories currently installed.)
Preparing to unpack .../python3.8-lib2to3_3.8.18-1+jammy1_all.deb ...
Unpacking python3.8-lib2to3 (3.8.18-1+jammy1) ...
Selecting previously unselected package python3.8-distutils.
Preparing to unpack .../python3.8-distutils_3.8.18-1+jammy1_all.deb ...
Unpacking python3.8-distutils (3.8.18-1+jammy1) ...
Selecting previously unselected package python3.8-venv.
Preparing to unpack .../python3.8-venv_3.8.18-1+jammy1_amd64.deb ...
Unpacking python3.8-venv (3.8.18-1+jammy1) ...
Setting up python3.8-lib2to3 (3.8.18-1+jammy1) ...
Setting up python3.8-distutils (3.8.18-1+jammy1) ...
Setting up python3.8-venv (3.8.18-1+jammy1) ...
fiap@fiap-VirtualBox:~$ python3.8 --version
Python 3.8.18
fiap@fiap-VirtualBox:~$
```

Figura 27 – Instalando o Python 3.8
Fonte: elaborado pelo autor (2024)

Navegue até a pasta “flink” e crie um ambiente virtual com o Python 3.8:

```
python3.8 -m venv flink-venv
```

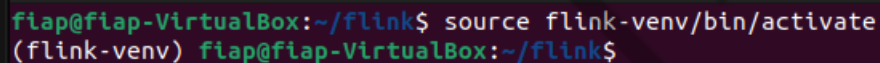


```
fiap@fiap-VirtualBox:~$ cd /home/fiap/flink
fiap@fiap-VirtualBox:~/flink$ python3.8 -m venv flink-venv
fiap@fiap-VirtualBox:~/flink$
```

Figura 28 – Criando um ambiente virtual com o Python 3.8
Fonte: elaborado pelo autor (2024)

Ative o ambiente virtual:

```
source flink-venv/bin/activate
```



```
fiap@fiap-VirtualBox:~/flink$ source flink-venv/bin/activate
(flink-venv) fiap@fiap-VirtualBox:~/flink$
```

Figura 29 – Ativando o ambiente virtual
Fonte: elaborado pelo autor (2024)

Instale as bibliotecas no ambiente virtual:

```
pip install apache-flink==1.13.6
```

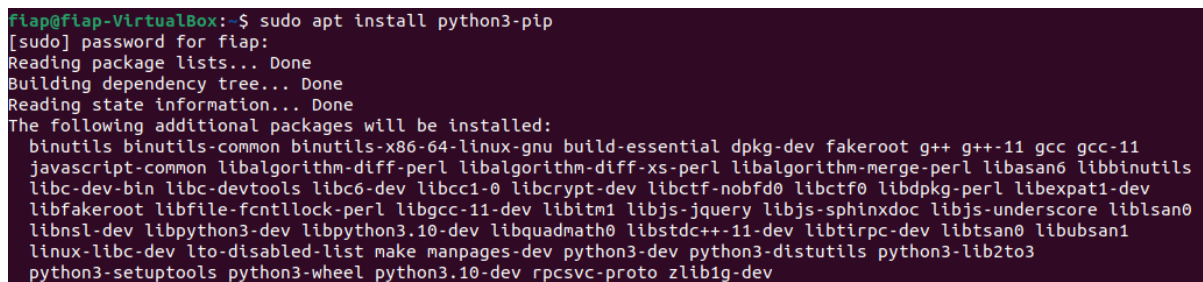
Abra um novo terminal e instale o pip:

```
sudo apt install python3-pip
```

```
pip3 install psycpg2-binary
```

```
sudo apt-get install libpq-dev python3-dev
```

```
pip3 install psycpg2
```



```
fiap@fiap-VirtualBox:~$ sudo apt install python3-pip
[sudo] password for fiap:
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following additional packages will be installed:
  binutils binutils-common binutils-x86-64-linux-gnu build-essential dpkg-dev fakeroot g++ g++-11 gcc gcc-11
  javascript-common libalgorithm-diff-perl libalgorithm-diff-xs-perl libalgorithm-merge-perl libasan6 libbinutils
  libc-dev-bin libc-devtools libc6-dev libcc1-0 libcrypt-dev libctf-nobfd0 libctf0 libdpkg-perl libexpat1-dev
  libfakeroot libfile-fcntllock-perl libgcc-11-dev libitm1 libjs-jquery libjs-sphinxdoc libjs-underscore liblsan0
  libnsl-dev libpython3-dev libpython3.10-dev libquadmath0 libstdc++-11-dev libtirpc-dev libtsan0 libubsan1
  linux-libc-dev lto-disabled-list make manpages-dev python3-dev python3-distutils python3-lib2to3
  python3-setuptools python3-wheel python3.10-dev rpcsvc-proto zlib1g-dev
```

Figura 30 – Instalando o pip
Fonte: elaborado pelo autor (2024)

Aqui você está instalando as bibliotecas no seu Python3.10, local. Usaremos ambientes diferentes para as tarefas da aula.

```
pip install confluent-kafka
```

```
pip install kafka-python
```

```
pip install mysql-connector-python
```

```
pip install fake
```

```
pip install faker
```

```
pip3 install psycopg2-binary
```

```
sudo apt-get install libpq-dev python3-dev
```

```
pip3 install psycopg2
```

Agora, abra este [link](#) e faça o download do conector SQL do Flink.

org/apache/flink/flink-sql-connector-kafka_2.11/1.13.6

...		
flink-sql-connector-kafka_2.11-1.13.6-sources...	2022-02-04 09:59	5370
flink-sql-connector-kafka_2.11-1.13.6-sources...	2022-02-04 09:59	833
flink-sql-connector-kafka_2.11-1.13.6-sources...	2022-02-04 09:59	32
flink-sql-connector-kafka_2.11-1.13.6-sources...	2022-02-04 09:59	40
flink-sql-connector-kafka_2.11-1.13.6.jar	2022-02-04 09:59	3674190
flink-sql-connector-kafka_2.11-1.13.6.jar.asc	2022-02-04 09:59	833
flink-sql-connector-kafka_2.11-1.13.6.jar.md5	2022-02-04 09:59	32
flink-sql-connector-kafka_2.11-1.13.6.jar.sha...	2022-02-04 09:59	40
flink-sql-connector-kafka_2.11-1.13.6.pom	2022-02-04 09:59	6328
flink-sql-connector-kafka_2.11-1.13.6.pom.asc	2022-02-04 09:59	833
flink-sql-connector-kafka_2.11-1.13.6.pom.md5	2022-02-04 09:59	32
flink-sql-connector-kafka_2.11-1.13.6.pom.sha...	2022-02-04 09:59	40

Figura 31 – Conector SQL do Flink
Fonte: elaborado pelo autor (2024)

Recorte o arquivo da pasta de downloads e cole na pasta flink/scripts.

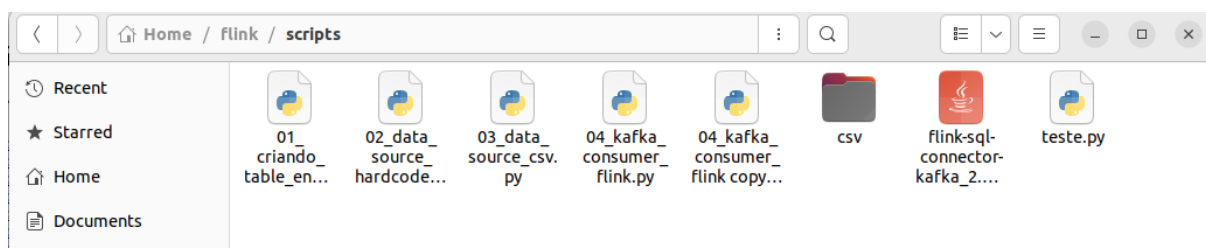


Figura 32 – Pasta flink/scripts
Fonte: elaborado pelo autor (2024)

Baixe [aqui](#) o conector jdbc para o Flink e cole na pasta “flink”.

Instalando o banco MySQL

Parte do laboratório será desenvolver um consumidor que lê os dados do tópico Kafka e armazena em um banco de dados. Para isso, usaremos o MySQL da Oracle.

Abra o terminal e digite o comando “sudo apt install mysql-server”

```
kafka@kafka-VirtualBox:~$ sudo apt install mysql-server
[sudo] password for kafka:
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following additional packages will be installed:
  libaio1 libbcgi-fast-perl libbcgi-pm-perl libevent-core-2.1-7
  libevent-pthreads-2.1-7 libfcgi-bin libfcgi-perl libfcgi0ldbl
  libhtml-template-perl libmecab2 libprotobuf-lite23 mecab-ipadic
  mecab-ipadic-utf8 mecab-utils mysql-client-8.0 mysql-client-core-8.0
  mysql-common mysql-server-8.0 mysql-server-core-8.0
Suggested packages:
  libipc-sharedcache-perl mailx tinycd
The following NEW packages will be installed:
  libaio1 libbcgi-fast-perl libbcgi-pm-perl libevent-core-2.1-7
  libevent-pthreads-2.1-7 libfcgi-bin libfcgi-perl libfcgi0ldbl
  libhtml-template-perl libmecab2 libprotobuf-lite23 mecab-ipadic
  mecab-ipadic-utf8 mecab-utils mysql-client-8.0 mysql-client-core-8.0
  mysql-common mysql-server mysql-server-8.0 mysql-server-core-8.0
0 upgraded, 20 newly installed, 0 to remove and 2 not upgraded.
Need to get 29,3 MB of archives.
After this operation, 242 MB of additional disk space will be used.
Do you want to continue? [Y/n] y
```

Figura 33 – Comando “sudo apt install mysql-server”

Fonte: elaborado pelo autor (2024)

Quando finalizar, digite o comando “sudo mysql_secure_installation”.

```
kafka@kafka-VirtualBox:~$ sudo mysql_secure_installation

Securing the MySQL server deployment.

Connecting to MySQL using a blank password.

VALIDATE PASSWORD COMPONENT can be used to test passwords
and improve security. It checks the strength of password
and allows the users to set only those passwords which are
secure enough. Would you like to setup VALIDATE PASSWORD component?

Press y|Y for Yes, any other key for No: █
```

Figura 34 – Comando “sudo mysql_secure_installation”

Fonte: elaborado pelo autor (2024)

Responda negativamente a todas as perguntas no terminal.

```
fiap@FIAP-DataPipelines-Streaming:~$ sudo mysql_secure_installation

Securing the MySQL server deployment.

Connecting to MySQL using a blank password.

VALIDATE PASSWORD COMPONENT can be used to test passwords
and improve security. It checks the strength of password
and allows the users to set only those passwords which are
secure enough. Would you like to setup VALIDATE PASSWORD component?

Press y|Y for Yes, any other key for No: n

Skipping password set for root as authentication with auth_socket is used by default.
If you would like to use password authentication instead, this can be done with the "ALTER_USER" command.
See https://dev.mysql.com/doc/refman/8.0/en/alter-user.html#alter-user-password-management for more information.

By default, a MySQL installation has an anonymous user,
allowing anyone to log into MySQL without having to have
a user account created for them. This is intended only for
testing, and to make the installation go a bit smoother.
You should remove them before moving into a production
environment.

Remove anonymous users? (Press y|Y for Yes, any other key for No) : n

... skipping.

Normally, root should only be allowed to connect from
'localhost'. This ensures that someone cannot guess at
the root password from the network.

Disallow root login remotely? (Press y|Y for Yes, any other key for No) : n

... skipping.

By default, MySQL comes with a database named 'test' that
anyone can access. This is also intended only for testing,
and should be removed before moving into a production
environment.

Remove test database and access to it? (Press y|Y for Yes, any other key for No) : n

... skipping.
Reloading the privilege tables will ensure that all changes
made so far will take effect immediately.
```

Figura 35 – Respostas negativas no terminal
Fonte: elaborado pelo autor (2024)

Conecte no banco MySQL através do terminal: “sudo mysql”. A partir disso, crie um usuário chamado Kafka com uma senha que seja aceita.

```
create user 'fiap'@'localhost' identified by 'fiap';

grant all privileges on *.* to 'fiap'@'localhost' with grant option;

FLUSH PRIVILEGES;
```

```
fiap@FIAP-DataPipelines-Streaming:~$ sudo mysql
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 10
Server version: 8.0.36-0ubuntu0.22.04.1 (Ubuntu)

Copyright (c) 2000, 2024, Oracle and/or its affiliates.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> create user 'fiap'@'localhost' identified by 'fiap';
Query OK, 0 rows affected (0,03 sec)

mysql> grant all privileges on *.* to 'fiap'@'localhost' with grant option;
Query OK, 0 rows affected (0,01 sec)

mysql> FLUSH PRIVILEGES;
Query OK, 0 rows affected (0,00 sec)

mysql> █
```

Figura 36 – Criando um usuário
Fonte: elaborado pelo autor (2024)

Saia do serviço do MySQL através do comando “exit” e logue com o novo usuário para se certificar de que deu certo.

```
fiap@FIAP-DataPipelines-Streaming:~$ mysql -u fiap -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 11
Server version: 8.0.36-0ubuntu0.22.04.1 (Ubuntu)

Copyright (c) 2000, 2024, Oracle and/or its affiliates.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

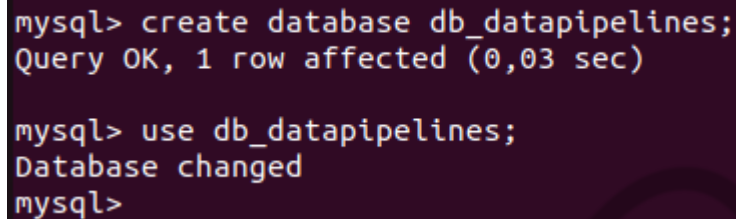
Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> select count(1) as qd from dual;
+----+
| qd |
+----+
|  1 |
+----+
1 row in set (0,00 sec)

mysql> █
```

Figura 37 – Logando com o novo usuário
Fonte: elaborado pelo autor (2024)

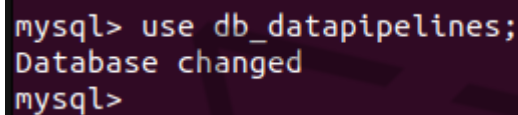
Ainda no serviço do MySQL com o usuário fiap, crie um banco de dados com o comando a seguir: create database db_datapipelines;



```
mysql> create database db_datapipelines;  
Query OK, 1 row affected (0,03 sec)  
  
mysql> use db_datapipelines;  
Database changed  
mysql>
```

Figura 38 – Criando um banco de dados
Fonte: elaborado pelo autor (2024)

Atribua essa conexão ao banco de dados que acabou de criar:



```
mysql> use db_datapipelines;  
Database changed  
mysql>
```

Figura 39 – Atribuindo essa conexão ao banco de dados
Fonte: elaborado pelo autor (2024)

Execute esse script para criar a tabela:

```
CREATE TABLE tb_log_sensor (  
  id_log_sensor VARCHAR(36) PRIMARY KEY,  
  id_sensor INT,  
  umidade FLOAT,  
  pressao FLOAT,  
  data DATETIME(6),  
  data_carga TIMESTAMP DEFAULT CURRENT_TIMESTAMP  
);
```

```
mysql> CREATE TABLE tb_log_sensor (  
-> id_log_sensor VARCHAR(36) PRIMARY KEY,  
-> id_sensor INT,  
-> umidade FLOAT,  
-> pressao FLOAT,  
-> data DATETIME(6),  
-> data_carga TIMESTAMP DEFAULT CURRENT_TIMESTAMP  
-> );  
Query OK, 0 rows affected (0,16 sec)  
mysql> 
```

Figura 40 – Executando esse script para criar a tabela
Fonte: elaborado pelo autor (2024)

Certifique-se de que sua tabela foi criada:

```
mysql> select count(1) from tb_log_sensor;  
+-----+  
| count(1) |  
+-----+  
|      0 |  
+-----+  
1 row in set (0,00 sec)
```

Figura 40 – Tabela criada
Fonte: elaborado pelo autor (2024)

O QUE VOCÊ VIU NESTA AULA?

Nessa aula você entendeu os fundamentos teóricos de como funcionam e quais são os cenários de uso para o streaming de dados e aprendeu a configurar e desenvolver em ferramentas “top of mind” do mercado atual, como Apache Kafka e Apache Flink.

Você conseguiu criar tópicos no Kafka através de formas variadas, enviou e recebeu mensagens pela ferramenta e compreendeu sua integração com o Apache Zookeeper.

Além disso, realizou a configuração e o desenvolvimento dentro do Apache Flink que possibilitaram conectar fontes de dados em tempo real com fontes estáticas, analisando os logs de sensores simulados em aula.

REFERÊNCIAS

FLINK. **Apache Kafka Connector.** 2023. Disponível em: <<https://nightlies.apache.org/flink/flink-docs-master/docs/connectors/datastream/kafka/>>. Acesso em: 28 mai. 2024.

FLINK. **CSV Format.** 2023. Disponível em: <<https://nightlies.apache.org/flink/flink-docs-master/docs/connectors/table/formats/csv/>>. Acesso em: 28 mai. 2024.

KAFKA. **Introduction.** 2023. Disponível em: <<https://kafka.apache.org/intro>>. Acesso em: 28 mai. 2024.

KAFKA. **Configuration.** 2023. Disponível em: <<https://kafka.apache.org/documentation/#configuration>>. Acesso em: 28 mai. 2024.

KAFKA. **The Producer.** 2023. Disponível em: <<https://kafka.apache.org/documentation/#theproducer>>. Acesso em: 28 mai. 2024.

PALAVRAS-CHAVE

Palavras-chave: Streaming. Kafka. Flink. Apache. Real time. Pyflink. Confluent. Stream.

EMSE



POSTECH