

Stack Bootcamp de Data Science

Preparação do ambiente

1. Faça download do Anaconda no site:
<https://www.anaconda.com/products/individual#Downloads>
2. Instalar o Docker Desktop no Windows ou no Linux
<https://www.docker.com/get-started>
3. Faça download do Visual Studio code
<https://code.visualstudio.com/download>

Crie um diretório na sua máquina para armazenar scripts e outros artefatos, exemplo:

C:\bootcampds

/home/<seunome>/bootcampds

Instalação e Configuração do Mysql Server

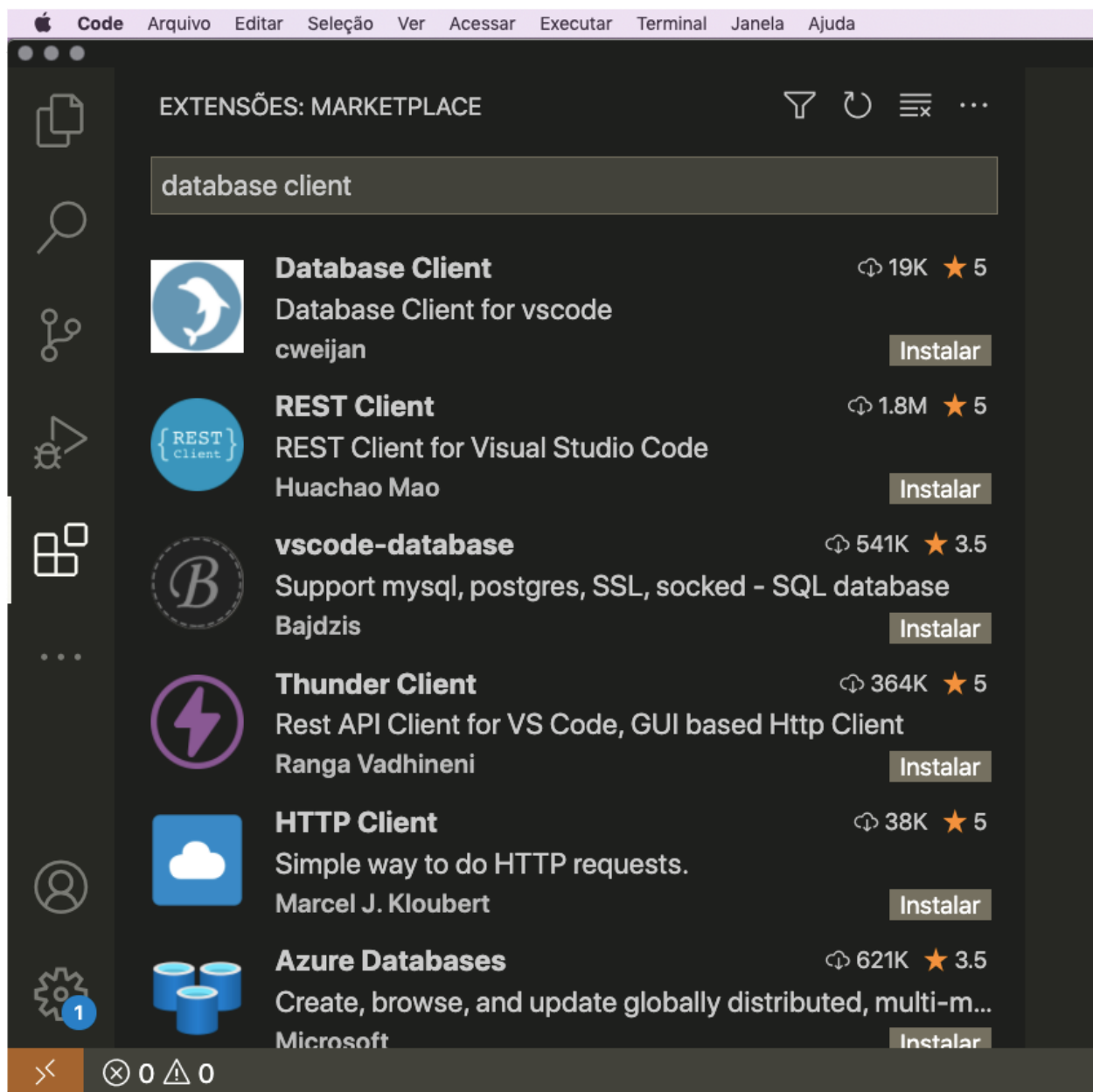
Se estiver usando Windows abra o **Powershell** e digite:

Crie o container do mysql habilitando a porta 3307:

```
docker run --name mysqlbd1 -e MYSQL_ROOT_PASSWORD=bootcamp -p "3307:3306" -d mysql
```

Teste o acesso ao banco de dados usando o Visual Studio Code:

Abra o Visual Studio Code e instale a extensão: **Database Client**



Teste o acesso ao banco de dados Mysql:

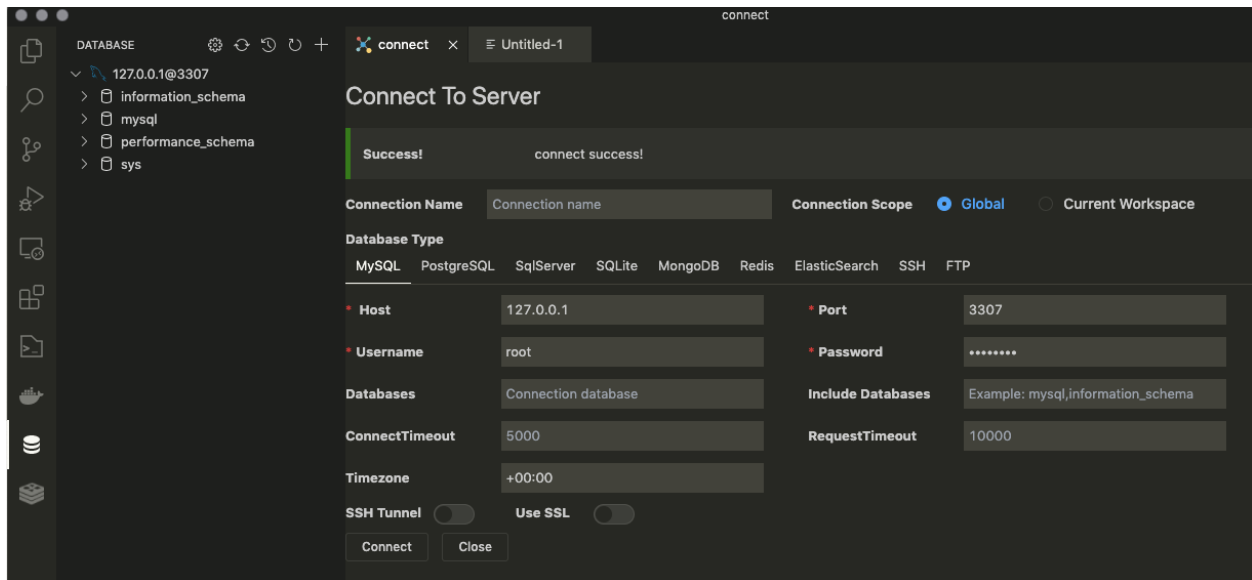
Coloque as configurações:

Host: 127.0.0.1

Username: root

Port: 3307

Password: bootcamp



Instalação e Configuração do Data Lake com Minio Server

Dentro do diretório **bootcampds** crie o diretório **datalake**.

Se estiver usando Windows abra o terminal do Powershell e execute o comando:

```
docker run --name minio -d -p 9000:9000 -p 9001:9001 -v "$PWD/datalake:/data" minio/minio server /data --console-address ":9001"
```

Teste o acesso ao Minio:

Abra o browser e digite: <http://localhost:9001/login>

username: minioadmin

password: minioadmin

Instalação e Configuração do Airflow.

1. Dentro do **diretório bootcampds** crie o diretório **airflow**.
2. Navegue até o diretório **airflow** e crie o diretório **dags**.
3. Faça download da imagem e execute o container do Apache Airflow

3a. Se estiver usando Windows abra o terminal do Powershell e execute o comando:

```
docker run -d -p 8080:8080 -v "$PWD/airflow/dags:/opt/airflow/dags/" --  
entrypoint=/bin/bash --name airflow apache/airflow:2.1.1-python3.8 -c '(airflow  
db init && airflow users create --username admin --password stack --firstname  
Felipe --lastname Lastname --role Admin --email admin@example.org); airflow  
webserver & airflow scheduler'
```

3b. Instale as bibliotecas necessárias para o ambiente:

Execute o comando abaixo para se conectar ao container do airflow:

```
docker container exec -it airflow bash
```

Em seguida instale as bibliotecas:

```
pip install pymysql xlrd openpyxl minio
```

3c. Se não der nenhum erro, acesse a interface web do Apache Airflow com o endereço (*Aguarde uns 5 minutos antes de abrir o terminal*):

<https://localhost:8080>

Faça o login de acesso ao Apache Airflow

Login: admin

Senha: bootcamp

Clique em Admin >> Variables

Crie as seguintes variáveis:

```
data_lake_server = 172.17.0.4:9000
```

```
data_lake_login = minioadmin
```


```
data_lake_password = minioadmin
```

```
database_server = 172.17.0.3 ( Use o comando inspect para descobrir o ip do  
container: docker container inspect mysqlbd1 - localizar o atributo IPAddress)
```

```
database_login = root
```

```
database_password = bootcamp
```

```
database_name = employees
```

 Airflow **DAGs** Security ▾ Browse ▾ Admin ▾ Docs ▾

Add Variable

Key *


data_lake_login


Val

minioadmin

















Description

Login de acesso ao Data Lake

Save 



As variáveis criadas devem ficar como:

List Variable			
Search ▾			
<div><div></div><div>Actions ▾</div><div></div></div>			
<input type="checkbox"/>	Key ▾	Val ▾	Description ▾
<input type="checkbox"/>  	data_lake_login	minioadmin	Login de conexão com o Data Lake
<input type="checkbox"/>  	data_lake_password	*****	Senha do usuário para acesso ao Data Lake
<input type="checkbox"/>  	data_lake_server	172.17.0.4:9000	Endereço do servidor Data Lake
<input type="checkbox"/>  	database_login	root	login do usuário do banco de dados
<input type="checkbox"/>  	database_name	employees	senha do usuário do banco de dados
<input type="checkbox"/>  	database_password	*****	Senha do usuário de banco de dados
<input type="checkbox"/>  	database_server	172.17.0.2	Ip e porta do servidor de banco de dados