

Instalando o Apache Airflow



Instalando o Apache Airflow

Usar o Docker é, sem dúvida, a maneira mais fácil de configurar e executar o Apache Airflow

Mas!

Saber como instalar o Apache Airflow manualmente também é importante

Por quê?

Como você pode personalizar sua imagem do Airflow Docker (por exemplo, instalando dependências) se não souber o que modificar?

Passos

Vamos descobrir as diferentes etapas para instalar e configurar o Airflow.

1/ Python

Primeiro, certifique-se de ter o Python 3.6+ instalado em seu computador.

Observe que você não pode mais executar o Apache Airflow 2.0 no Python 2.7. (o que é ótimo btw 😊)

2/ Pip

A maneira oficial de instalar o Airflow é com o Pip.

Pip é um gerenciador de pacotes python usado para instalar e gerenciar pacotes python.

Você usará o pip para instalar o Apache Airflow, bem como todas as dependências necessárias para executar.

3/ Dependências do sistema (somente usuários Linux)

No Linux, você precisa instalar certas dependências do sistema operacional, conforme mostrado abaixo

```
sudo apt-get install -y --no-install-recommends \
    freetds-bin \
    krb5-user \
    ldap-utils \
    libffi6 \
    libsasl2-2 \
    libsasl2-modules \
    libssl1.1 \
    locales \
    lsb-release \
    sasl2-bin \
    sqlite3 \
    unixodbc
```

4/ Instalando o Apache Airflow

Depois de concluir as etapas anteriores, você estará pronto para instalar o Apache Airflow.

O comando é bem simples, você só precisa executar:

```
pip install apache-airflow ...mas, espere um minuto!
```

Essa não é a melhor maneira. O Airflow vem com MUITAS dependências.

O problema é que, se uma dependência for atualizada, o comando acima instalará a versão mais recente. Isso pode causar incompatibilidades e erros de dependência.

É por isso que é sempre recomendável adicionar um **arquivo de restrição** .

Um arquivo de restrição contém todas as bibliotecas python necessárias ao Airflow junto com sua versão pinada.

Dessa forma, você pode ter certeza de que as dependências corretas estão instaladas.

Execute o seguinte comando:

```
pip install apache-airflow --constraint
```

```
https://raw.githubusercontent.com/apache/airflow/constraints-${AIRFLOW_VERSION}/constraints-${PYTHON_VERSION}.txt
```

(Onde `AIRFLOW_VERSION` é a versão do Airflow que você deseja instalar, 2.0.0 e `PYTHON_VERSION`, sua versão do python, 3.6, 3.7 ou 3.8)

Bem feito!

Neste ponto, você instalou o Airflow com sucesso, mas... não pode executá-lo.

Por quê?

Porque ele precisa ser inicializado primeiro.

Este processo consiste em duas etapas:

1. Gerando os arquivos e pastas necessários para o Airflow (logs/, airflow.cfg, unittests.cfg etc.)
2. Inicializando o banco de dados de metadados (SQLite por padrão)

AIRFLOW_HOME

Por padrão, os arquivos e pastas são gerados na pasta `~/airflow`

Você pode modificar esse comportamento configurando a variável de ambiente `AIRFLOW_HOME` .

Por exemplo, para inicializar o airflow na pasta /opt, exporte a seguinte variável de ambiente:

```
AIRFLOW_HOME=/opt/airflow
```

Inicialização do banco de dados

Quando estiver bem com sua casa do Airflow, inicie o Airflow com:

```
airflow db init
```

Parabéns!

O Airflow está pronto para ser executado com as configurações padrão!



PRODUTOS	RECURSOS	COMPANHIA
astro	Documentos	Sobre nós
Status	Registro	Carreiras
Apoiar	Guias	Contato
Preços	Webinars	Parceiros
Por que o	Academia	
fluxo de ar	Fórum	
	Blogue	
	Demonstração	
	semanal	

