

Projeto Real de

# Data Engineering



# Etapas do Projeto

## 1 Criação do ambiente:

- Criação de instância Linux na AWS
- Instalação do Docker
- Instalação do Airflow no Docker



## 2 Criação do DataWarehouse:

- Criação de DW no Snowflake
- Criação de database, schema de stage/prod



## 3 Preparação dos dados:

- Dag e Task no airflow para consumir dados do Postgres e carregar no Snowflake
- Transformação de dados com DBT



## 4 Criação da análise:

- Desenvolvimento de análise no Looker Studio

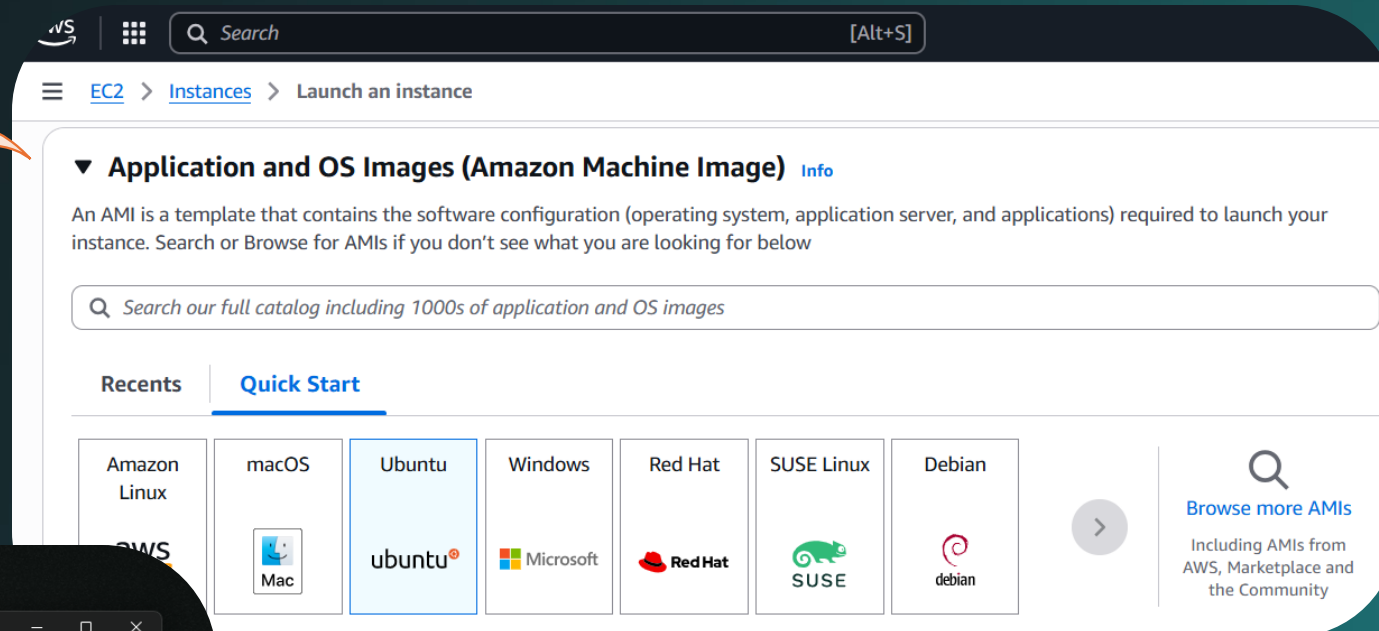


1

# Criação do ambiente:

## • Criação de instância Linux na AWS

- Escolha da AMI (Amazon Machine Image) baseada em Ubuntu.
- Configuração de segurança com grupos de segurança (SSH, portas necessárias).
- Atribuição de chave SSH para acesso remoto à instância.



```
ubuntu@ip-172-31-13-232: ~  
ubuntu@ip-172-31-13-232:~$ curl -Lfo 'https://airflow.apache.org/docs/apache-airflow/stable/docker-compose.yaml'  
% Total % Received % Xferd Average Speed Time Time Time Current  
Dload Upload Total Spent Left Speed  
100 12766 100 12766 0 0 83680 0 --:--:-- --:--:-- --:--:-- 83986  
ubuntu@ip-172-31-13-232:~$ ls  
docker-compose.yaml  
ubuntu@ip-172-31-13-232:~$
```

## • Instalação do Docker

- Atualização dos pacotes da instância e instalação de dependências.
- Adição do repositório oficial e instalação do Docker Engine.
- Verificação do serviço e inclusão do usuário no grupo docker.

## 2 Criação do DataWarehouse :

### • Criação de DW no Snowflake

- Abertura de conta no Snowflake e definição do warehouse inicial.
- Configuração de roles e permissões de acesso ao ambiente.
- Escolha do tamanho e tipo de warehouse (ex: X-Small para testes).

The screenshot displays the Snowflake web interface. On the left, a sidebar contains navigation options: Home, Search, Projects, Data, Databases (selected), Add Data, Migrations, Openflow, Data Products, AI & ML, Monitoring, and Admin. At the bottom of the sidebar, it shows '\$344 credits left' and 'Trial ends in 16 days' with an 'Upgrade' button. The main content area is titled 'Databases' and shows a list of 4 databases. A red box highlights the 'NOVADRIVE' database and its schemas: ANALYTIC, DBT\_ITORVES, INFORMATION\_SCHEMA, PUBLIC, and STAGE. The 'Databases' table lists the following databases:

NAME ↑	SOURCE	OWNER	CREATED
NOVADRIVE	Local	ACCOUNTADM...	2 weeks ago
SNOWFLAKE	Share	—	2 weeks ago
SNOWFLAKE_LEA...	Local	ACCOUNTADM...	2 weeks ago
SNOWFLAKE_SA...	Share	ACCOUNTADM...	2 weeks ago

### • Criação de database, schema de stage/prod

- Criação do banco principal para organização dos dados.
- Estruturação dos schemas stage (dados brutos) e prod (dados tratados).
- Aplicação de boas práticas de nomenclatura e versionamento.



- **Desenvolvimento de análise no Looker Studio**

- Conexão do Looker Studio ao Snowflake via conector de dados.
- Criação de dashboards interativos com gráficos, filtros e tabelas.
- Aplicação de fórmulas e campos calculados para insights personalizados.

