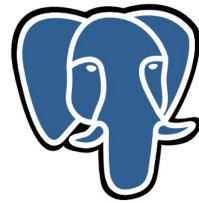


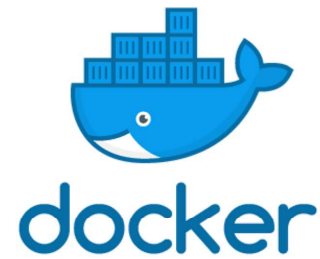
Ferramentas empregadas no teste



PostgreSQL



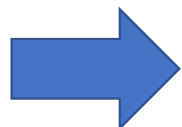
Apache
Airflow



T1

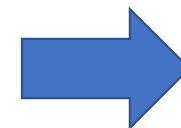
Transformação dos dados
e criação do arquivo csv

LimpezaDados.py



T2

Checagem dos dados



T3

Criação do banco de
dados em Postgres

ExportaDados.py



T4

Realizar query a cada 10
segundos

query.py



Softwares necessários para abrir o código

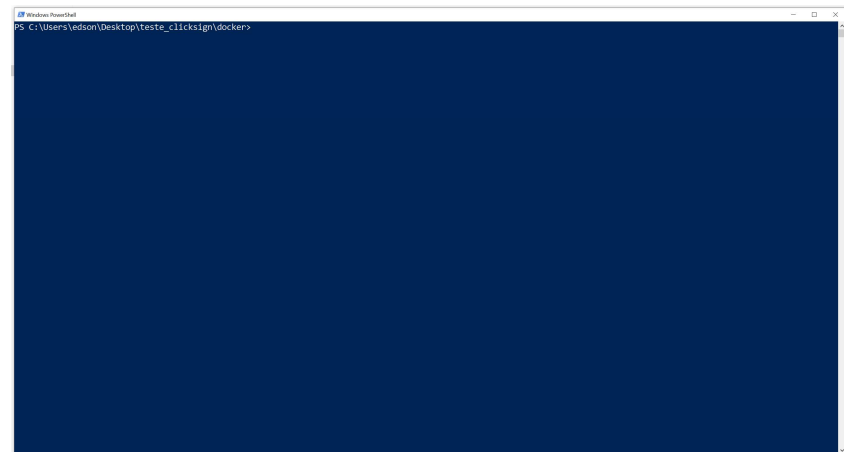
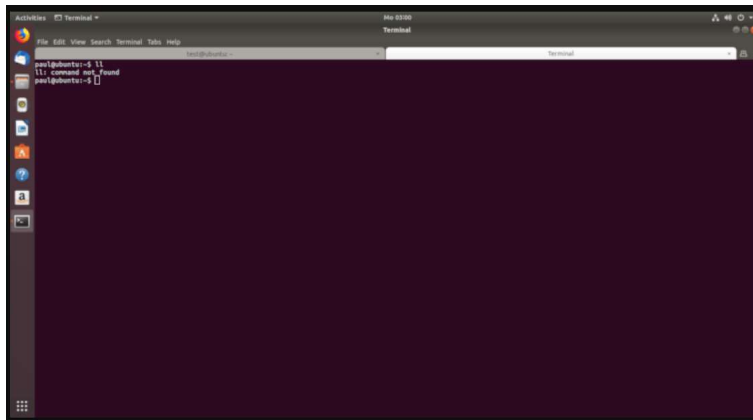
Docker Desktop para o pipeline

Python

- Pandas
- Matplotlib
- Seaborn

Powershell ou prompt de comando no caso do Windows

Terminal no caso do Linux

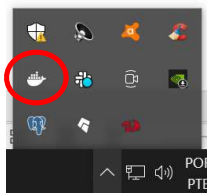


Após clonar o repositório no github as seguintes pastas irão compor o diretório

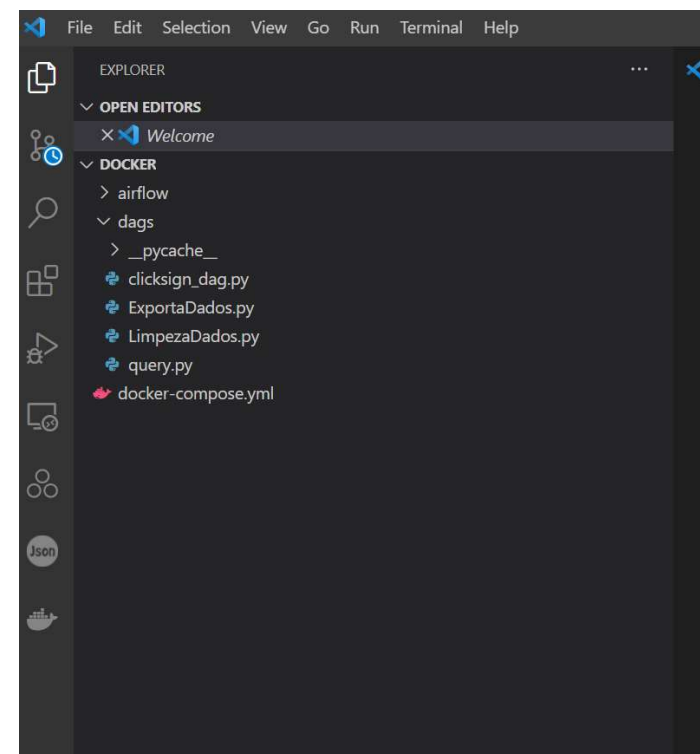
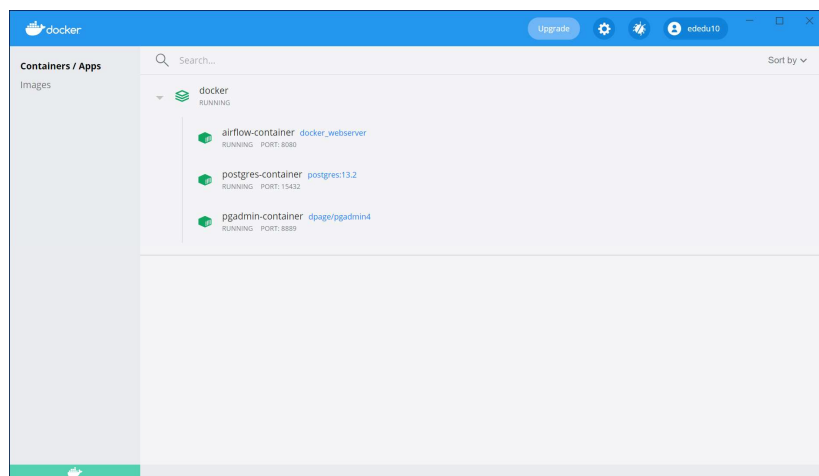
Nome	Data de modificação	Tipo	Tamanho
airflow	18/03/2021 11:22	Pasta de arquivos	
dags	19/03/2021 18:28	Pasta de arquivos	
docker-compose.yml	19/03/2021 16:18	Arquivo Fonte Yaml	2 KB

O arquivo clicksign_dag.py contém o código do dag que o Airflow utilizará para orquestrar as tarefas

Dentro do diretório do projeto e com tendo iniciado o Docker Desktop



Digite no powershell ou no prompt de comando o seguinte comando `docker-compose up -d`
As imagens serão baixadas e os containeres serão criados baseado no arquivo `docker-compose.yml`
3 containeres serão gerados



docker ps para checar os containeres que estão rodando

```
Windows PowerShell
PS C:\Users\edson\Desktop\teste_clicksign\docker> docker ps
CONTAINER ID   IMAGE          COMMAND                  CREATED        STATUS                    PORTS                    NAMES
e0b383970039   docker_webserver  "/entrypoint.sh"        3 hours ago   Up Less than a second (health: starting)  0.0.0.0:8080->8080/tcp   airflow-container
a964e693f4c6   postgres:13.2    "docker-entrypoint.s..." 3 hours ago   Up Less than a second                    0.0.0.0:15432->5432/tcp   postgres-container

PS C:\Users\edson\Desktop\teste_clicksign\docker>
```

docker inspect + as primeiras letras do ID do container

```
PS C:\Users\edson\Desktop\teste_clicksign\docker> docker inspect a9
```

```
Windows PowerShell
"docker_clicksign_network": {
  "IPAMConfig": null,
  "Links": null,
  "Aliases": [
    "db",
    "a964e693f4c6"
  ],
  "NetworkID": "a3c117983d09a8719f0196464dfb1fed15516c5ec6557bd7b3b7ac28d142a58b",
  "EndpointID": "8e830971fa3005423c0891929b96dd5c74c3cbdacc0c18982b8468060884fc90",
  "Gateway": "172.23.0.1",
  "IPAddress": "172.23.0.3",
  "IPPrefixLen": 16,
  "IPv6Gateway": "",
  "GlobalIPv6Address": "",
  "GlobalIPv6PrefixLen": 0,
  "MacAddress": "02:42:ac:17:00:03",
  "DriverOpts": null
}
}
```

O IPAddress é utilizado como Host para ter acesso ao banco de dados

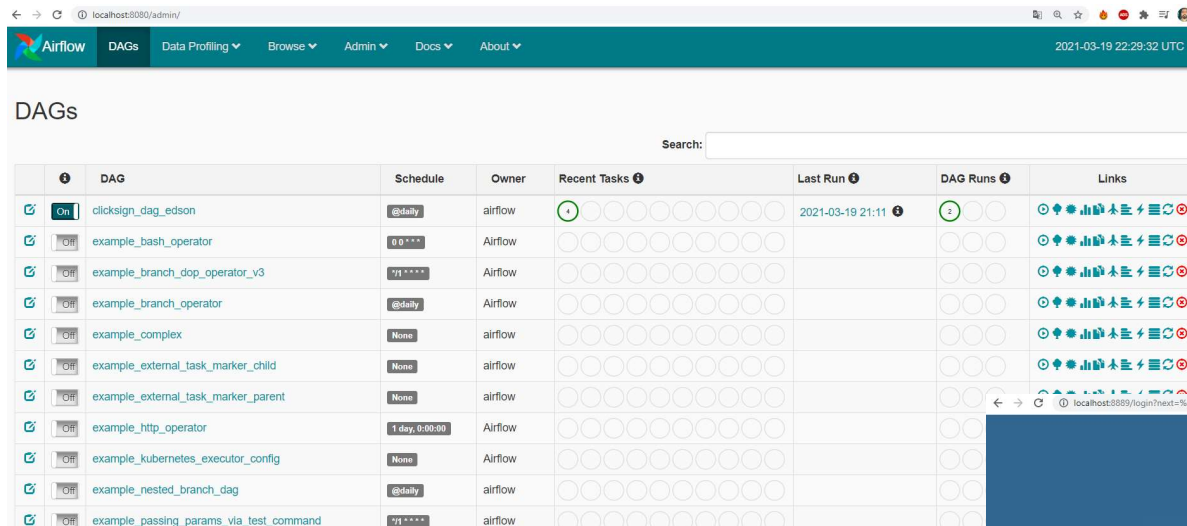
Trocando Host de acordo com o valor checado no docker inspect

```
1 def exporta_dados():
2
3     import psycopg2
4
5     host = '172.23.0.3'
6     port = 5432
7     db_name = 'postgres'
8     username = 'edson'
9     password = 'clicksign'
10
11     # Create connection
12     conn = psycopg2.connect(host = host, port = port, dbname = db_name, user = username, password = password)
13
14     # Create a cursor object
15     cur = conn.cursor()
16
17     # executar query (criando tabela)
18
19     cur.execute("""DROP TABLE IF EXISTS clicksign;
20     CREATE TABLE clicksign(
21         age            INTEGER,
22         workclass       VARCHAR(20),
23         fnlwtg         DECIMAL,
24         education       VARCHAR(20),
```

```
1 def query():
2
3     import psycopg2
4     import math
5     import pandas as pd
6     import time
7
8     host = '172.23.0.3'
9     port = 5432
10    db_name = 'postgres'
11    username = 'edson'
12    password = 'clicksign'
13
14    # Create connection
15    conn = psycopg2.connect(host = host, port = port, dbname = db_name, user = username, password = password)
16
17    # Create a cursor object
18    cur = conn.cursor()
19
20    # executar query (criando tabela)
21
22    df = pd.read_csv('/usr/local/airflow/dags/clicksign_df.csv')
23    linhas = df.shape[0]
24    limite = 1630
25    repeticoes = math.ceil(linhas/limite)
26
27
28 def query_limite(limite_inferior, limite_superior):
29
30
```

A troca do Host deve ser feita nos arquivos
ExportaDados.py e query.py

No navegador abra duas abas
Uma para o airflow e digite localhost:8080
E outra para o pgadmin, digite localhost:8889



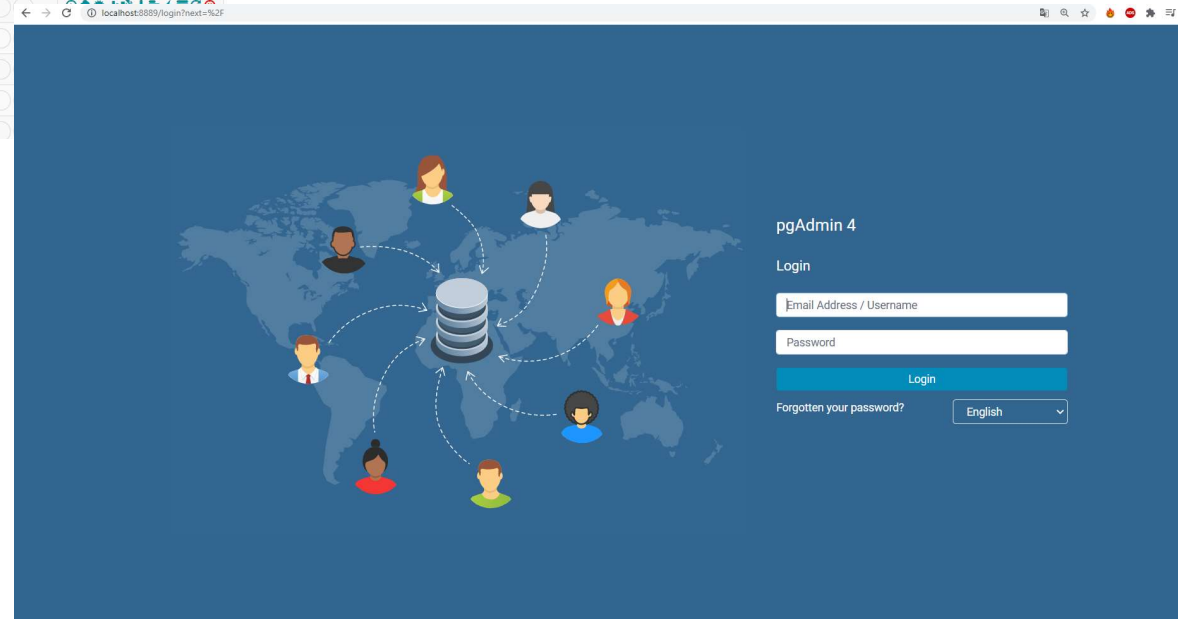
The screenshot shows the Apache Airflow web interface. The top navigation bar includes links for DAGs, Data Profiling, Browse, Admin, Docs, and About. The main content area is titled 'DAGs' and features a search bar. Below the search bar is a table listing various DAGs. The table has columns for DAG name, Schedule, Owner, Recent Tasks, Last Run, DAG Runs, and Links. The first DAG listed is 'clicksign_dag_edson', which is currently 'On' and has a daily schedule.

	DAG	Schedule	Owner	Recent Tasks	Last Run	DAG Runs	Links
<input checked="" type="checkbox"/>	clicksign_dag_edson	@daily	airflow		2021-03-19 21:11		
<input checked="" type="checkbox"/>	example_bash_operator	0 0 * * *	Airflow				
<input checked="" type="checkbox"/>	example_branch_dop_operator_v3	* * * * *	Airflow				
<input checked="" type="checkbox"/>	example_branch_operator	@daily	Airflow				
<input checked="" type="checkbox"/>	example_complex	None	airflow				
<input checked="" type="checkbox"/>	example_external_task_marker_child	None	airflow				
<input checked="" type="checkbox"/>	example_external_task_marker_parent	None	airflow				
<input checked="" type="checkbox"/>	example_http_operator	1 day, 0:00:00	Airflow				
<input checked="" type="checkbox"/>	example_kubernetes_executor_config	None	Airflow				
<input checked="" type="checkbox"/>	example_nested_branch_dag	@daily	airflow				
<input checked="" type="checkbox"/>	example_passing_params_via_test_command	* * * * *	airflow				

O DAG responsável por organizar as tarefas é o
click_dag_edson

As duas aplicações estão rodando nos containeres que
foram criados na mesma rede































Para acessar o pgadmin
edson.costa@hotmail.com
Senha: clicksign




Clique no botão ON e posteriormente no trigger dag

DAGs

Search:

	DAG	Schedule	Owner	Recent Tasks	Last Run	DAG Runs	Links
	clicksign_dag_edson	@daily	airflow	                            			


Tarefas


 Airflow


[DAGs](#) [Data Profiling ▾](#) [Browse ▾](#) [Admin ▾](#) [Docs ▾](#) [About ▾](#)


On


DAG: clicksign_dag_edson


 Graph View


 Tree View

 Task Duration

 Task Tries

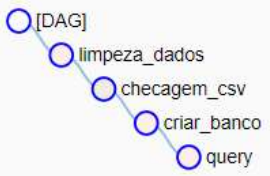
 Landing Times

 Gantt

 Data

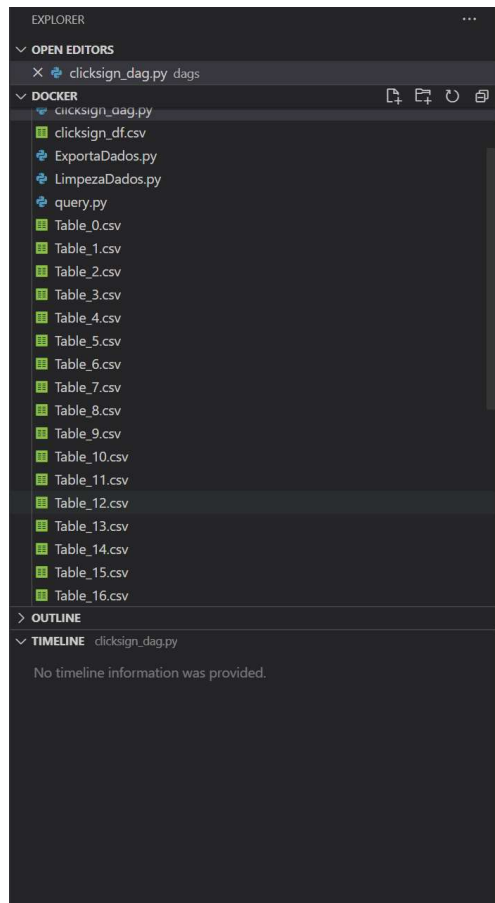
Base date: Number of runs:

☐ BashOperator ☐ PythonOperator

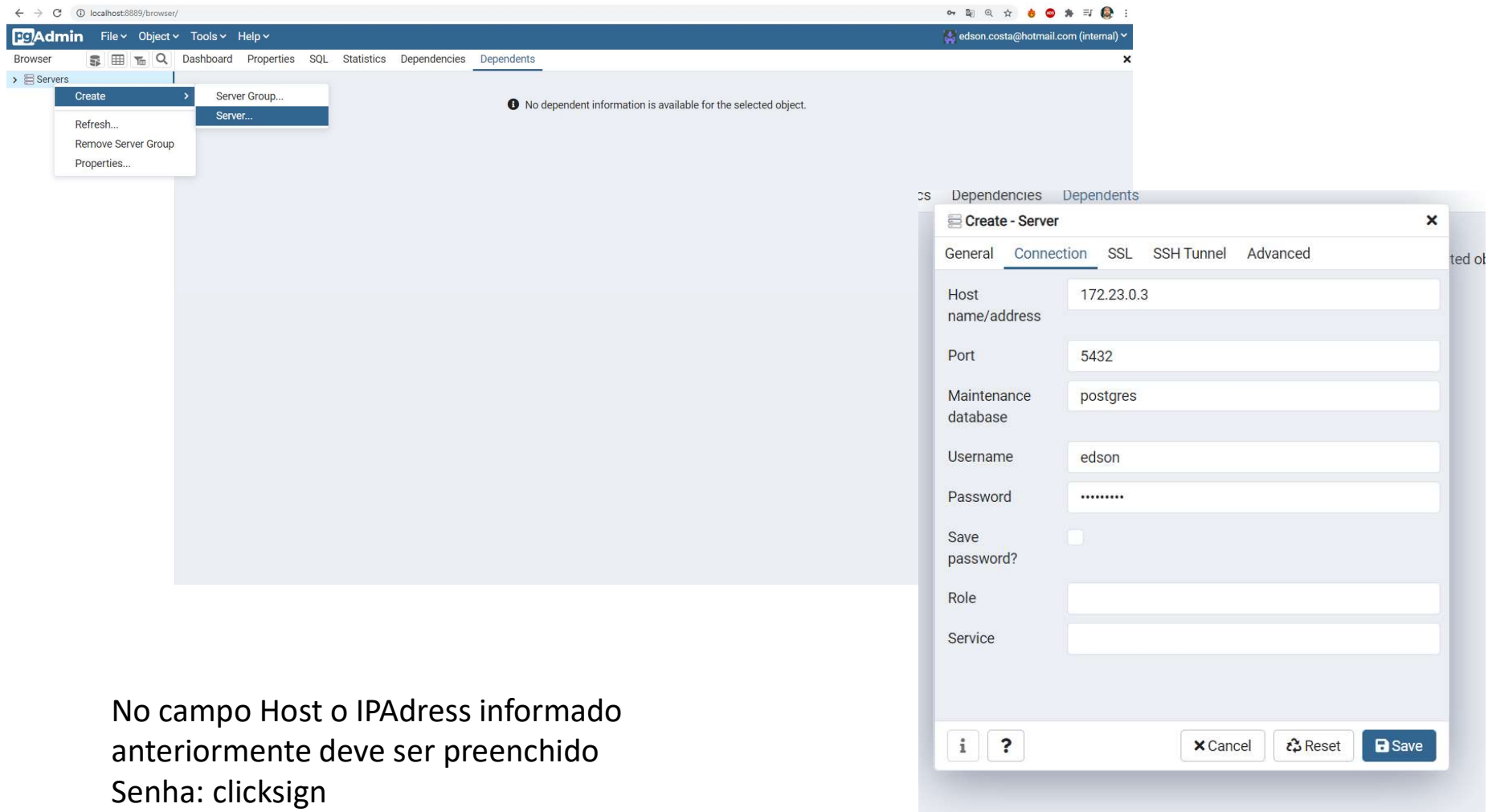


09 PM

A cada 10 segundos é processado 1630 registros e são salvos em um csv que fica armazenado na pasta dags



clicksign_dag.py	19/03/2021 19:41	Arquivo Fonte Python	2 KB
clicksign_df.csv	19/03/2021 19:41	Microsoft Excel Com...	5.767 KB
ExportaDados.py	19/03/2021 16:31	Arquivo Fonte Python	2 KB
LimpezaDados.py	19/03/2021 13:35	Arquivo Fonte Python	4 KB
query.py	19/03/2021 18:08	Arquivo Fonte Python	2 KB
Table_0.csv	19/03/2021 19:44	Microsoft Excel Com...	190 KB
Table_1.csv	19/03/2021 19:44	Microsoft Excel Com...	190 KB
Table_2.csv	19/03/2021 19:44	Microsoft Excel Com...	190 KB
Table_3.csv	19/03/2021 19:45	Microsoft Excel Com...	191 KB
Table_4.csv	19/03/2021 19:45	Microsoft Excel Com...	191 KB
Table_5.csv	19/03/2021 19:45	Microsoft Excel Com...	191 KB
Table_6.csv	19/03/2021 19:45	Microsoft Excel Com...	193 KB
Table_7.csv	19/03/2021 19:45	Microsoft Excel Com...	192 KB
Table_8.csv	19/03/2021 19:45	Microsoft Excel Com...	192 KB
Table_9.csv	19/03/2021 19:46	Microsoft Excel Com...	192 KB
Table_10.csv	19/03/2021 19:46	Microsoft Excel Com...	192 KB
Table_11.csv	19/03/2021 19:46	Microsoft Excel Com...	193 KB
Table_12.csv	19/03/2021 19:46	Microsoft Excel Com...	193 KB
Table_13.csv	19/03/2021 19:46	Microsoft Excel Com...	192 KB
Table_14.csv	19/03/2021 19:46	Microsoft Excel Com...	193 KB
Table_15.csv	19/03/2021 19:47	Microsoft Excel Com...	192 KB
Table_16.csv	19/03/2021 19:47	Microsoft Excel Com...	192 KB
Table_17.csv	19/03/2021 19:47	Microsoft Excel Com...	193 KB
Table_18.csv	19/03/2021 19:47	Microsoft Excel Com...	192 KB
Table_19.csv	19/03/2021 19:47	Microsoft Excel Com...	193 KB
Table_20.csv	19/03/2021 19:47	Microsoft Excel Com...	196 KB
Table_21.csv	19/03/2021 19:48	Microsoft Excel Com...	195 KB
Table_22.csv	19/03/2021 19:48	Microsoft Excel Com...	195 KB
Table_23.csv	19/03/2021 19:48	Microsoft Excel Com...	196 KB
Table_24.csv	19/03/2021 19:48	Microsoft Excel Com...	195 KB
Table_25.csv	19/03/2021 19:48	Microsoft Excel Com...	196 KB
Table_26.csv	19/03/2021 19:49	Microsoft Excel Com...	196 KB
Table_27.csv	19/03/2021 19:49	Microsoft Excel Com...	195 KB
Table_28.csv	19/03/2021 19:49	Microsoft Excel Com...	196 KB
Table_29.csv	19/03/2021 19:49	Microsoft Excel Com...	188 KB



pgAdmin

FileObjectToolsHelp

Browser

Extensions

Foreign Data Wrappers

Languages (1)

Publications

Schemas (1)

public

Collations

Domains

FTS Configurations

FTS Dictionaries

FTS Parsers

FTS Templates

Foreign Tables

Functions

Materialized Views

Procedures

Sequences

Tables (1)

clicksign

Columns

Constraints

Indexes

RLS Policies

Rules

Triggers

Trigger Functions

Types

Views

Subscriptions

Login/Group Roles

Tablespaces (2)

DashboardPropertiesSQLStatisticsDependenciesDependentspostgres/edson@click*

postgres/edson@click

Query EditorQuery HistoryScratchpad

1SELECT * FROM clicksign;

Data OutputExplainMessagesNotifications

	age integer	workclass character varying (20)	fnlwgt numeric	education character varying (20)	education_num integer	marital_status character varying (30)	occupation character varying (30)
1	39	State-gov	77516.0	Bachelors	13	Never-married	Adm-clerical
2	50	Self-emp-not-inc	83311.0	Bachelors	13	Married-civ-spouse	Exec-managerial
3	38	Private	215646.0	HS-grad	9	Divorced	Handler-cleaner
4	53	Private	234721.0	11th	7	Married-civ-spouse	Handler-cleaner
5	28	Private	338409.0	Bachelors	13	Married-civ-spouse	Professional-specialty
6	37	Private	284582.0	Masters	14	Married-civ-spouse	Exec-managerial
7	49	Private	160187.0	9th	5	Married-spouse-absent	Other-service
8	52	Self-emp-not-inc	209642.0	HS-grad	9	Married-civ-spouse	Exec-managerial
9	31	Private	45781.0	Masters	14	Never-married	Professional-specialty