

```

root@workstation:/home/aherman/Documents# mkdir TP-archi
root@workstation:/home/aherman/Documents# cd !$
cd TP-archi
root@workstation:/home/aherman/Documents/TP-archi# git clone https://github.com/AlexandreHERMANepsi/ATL-Datamart.git
Cloning into 'ATL-Datamart'...
remote: Enumerating objects: 67, done.
remote: Counting objects: 100% (34/34), done.
remote: Compressing objects: 100% (23/23), done.
remote: Total 67 (delta 22), reused 11 (delta 11), pack-reused 33
Receiving objects: 100% (67/67), 462.42 KiB | 3.45 MiB/s, done.
Resolving deltas: 100% (23/23), done.
root@workstation:/home/aherman/Documents/TP-archi# ll
total 4
drwxr-xr-x 12 root root 4096 avril 23 10:38 ATL-Datamart
root@workstation:/home/aherman/Documents/TP-archi# cd ATL-Datamart/
root@workstation:/home/aherman/Documents/TP-archi/ATL-Datamart# ll
total 84
drwxr-xr-x 3 root root 4096 avril 23 10:38 airflow
drwxr-xr-x 7 root root 4096 avril 23 10:38 data
-rw-r--r-- 1 root root 10062 avril 23 10:38 docker-compose.yml
drwxr-xr-x 2 root root 4096 avril 23 10:38 docs
-rw-r--r-- 1 root root 1484 avril 23 10:38 LICENSE
-rw-r--r-- 1 root root 4680 avril 23 10:38 Makefile
drwxr-xr-x 2 root root 4096 avril 23 10:38 minio
drwxr-xr-x 2 root root 4096 avril 23 10:38 models
drwxr-xr-x 2 root root 4096 avril 23 10:38 notebooks
-rw-r--r-- 1 root root 6567 avril 23 10:38 README.md
drwxr-xr-x 2 root root 4096 avril 23 10:38 references
drwxr-xr-x 3 root root 4096 avril 23 10:38 reports
-rw-r--r-- 1 root root 156 avril 23 10:38 requirements.txt
-rw-r--r-- 1 root root 262 avril 23 10:38 setup.py
drwxr-xr-x 4 root root 4096 avril 23 10:38 src
-rw-r--r-- 1 root root 632 avril 23 10:38 test_environment.py
-rw-r--r-- 1 root root 50 avril 23 10:38 tox.ini

```

```

Digest: sha256:72e367209f7d75fd4c9a93129f3df083feef5923b2690a52b5db205f7cd7890b
Status: Downloaded newer image for apache/airflow:latest-python3.11
Creating atl-datamart_redis_1 ... done
Creating atl-datamart_postgres-airflow_1 ... done
Creating minio ... done
Creating atl-datamart_db_1 ... done
Creating atl-datamart_airflow-init_1 ... done
Creating atl-datamart_airflow-triggerer_1 ... done
Creating atl-datamart_airflow-webserver_1 ... done
Creating atl-datamart_airflow-worker_1 ... done
Creating atl-datamart_airflow-scheduler_1 ... done
root@workstation:/home/aherman/Documents/TP-archi/ATL-Datamart# vim docker-compose.yml
root@workstation:/home/aherman/Documents/TP-archi/ATL-Datamart# █

```

TP1

Création bucket

création access

```
import os
import requests
from minio import Minio
from minio.error import S3Error
from datetime import datetime

# Fonction pour télécharger un fichier et le stocker dans Minio
def download_and_upload_to_minio(url, minio_client, bucket_name):
    filename = url.split("/")[-1]
    response = requests.get(url)
    if response.status_code == 200:
        with open(filename, 'wb') as f:
            f.write(response.content)
        try:
            minio_client.fput_object(bucket_name, filename,
filename)
            os.remove(filename) # Supprimer le fichier après
l'avoir téléchargé
            print(f"Le fichier {filename} a été téléchargé et
stocké dans Minio.")
        except S3Error as e:
            print(f"Une erreur s'est produite lors du stockage dans
Minio : {e}")

# Fonction pour récupérer les datasets de janvier 2023 à août 2023
def retrieve_datasets_jan_to_aug_2023(minio_client, bucket_name):
    base_url = "https://d37ci6vzurychx.cloudfront.net/trip-data"
    for month in range(1, 9): # De janvier à août
        year_month = f"2023-{month:02d}"
        url = f"{base_url}/yellow_tripdata_{year_month}.parquet"
        download_and_upload_to_minio(url, minio_client,
bucket_name)

# Fonction pour récupérer le dernier mois disponible
def retrieve_latest_dataset(minio_client, bucket_name):
    current_date = datetime.now()
    year_month = current_date.strftime("%Y-%m")
    url =
```

```
f"https://d37ci6vzurychx.cloudfront.net/trip-data/yellow_tripdata_
{year_month}.parquet"
download_and_upload_to_minio(url, minio_client, bucket_name)

# Paramètres Minio
minio_endpoint = "127.0.0.1:9000"
minio_access_key = "jyQQ3BcRl073rtbb9GnA"
minio_secret_key = "2kVRJbQUIVFfn0VTBZFsFEW09XUyFs9rDW6vHj6nB"
minio_bucket_name = "tp1"

# Connexion au client Minio
minio_client = Minio(minio_endpoint,
                      access_key=minio_access_key,
                      secret_key=minio_secret_key,
                      secure=False)

# Appels aux fonctions pour récupérer et stocker les données
retrieve_datasets_jan_to_aug_2023(minio_client, minio_bucket_name)
retrieve_latest_dataset(minio_client, minio_bucket_name)
```

```
root@workstation:/home/aherman/Documents/TP-archi/ATL-Datamart/src/data# python3 tp1-get-date.py
Le fichier yellow_tripdata_2023-01.parquet a été téléchargé et stocké dans Minio.
Le fichier yellow_tripdata_2023-02.parquet a été téléchargé et stocké dans Minio.
Le fichier yellow_tripdata_2023-03.parquet a été téléchargé et stocké dans Minio.
Le fichier yellow_tripdata_2023-04.parquet a été téléchargé et stocké dans Minio.
Le fichier yellow_tripdata_2023-05.parquet a été téléchargé et stocké dans Minio.
Le fichier yellow_tripdata_2023-06.parquet a été téléchargé et stocké dans Minio.
Le fichier yellow_tripdata_2023-07.parquet a été téléchargé et stocké dans Minio.
Le fichier yellow_tripdata_2023-08.parquet a été téléchargé et stocké dans Minio.
```

tp1

Created: Mon May 13 2024 13:41:30 GMT+0200 (Central European Summer Time)

Access: R/W



Usage

396.7 MiB

Objects

8

TP2

```
CREATE SCHEMA IF NOT EXISTS datamart_taxi;

CREATE TABLE datamart_taxi.dimension_temps (
    temps_id SERIAL PRIMARY KEY,
    date_heure TIMESTAMP NOT NULL,
    heure INT,
    jour INT,
    mois INT,
    annee INT
);

CREATE TABLE datamart_taxi.dimension_lieu (
    lieu_id SERIAL PRIMARY KEY,
    location_id INT NOT NULL
);

CREATE TABLE datamart_taxi.dimension_paiement (
    paiement_id SERIAL PRIMARY KEY,
    type_paiement INT
);
```

TP3