*Mastere 2 Data Engineer*
*Expert informatique et système d'information - Titre certifié niveau 7*

# ENERGIEWATCH

**ADIB Melissa**
**DIALLO Alimou M.**
**GBE Keagnon Grâce H.**

# SOMMAIRE



1.

# UNE PLONGÉE DANS ENERGIEWATCH

# UN PEU D'HISTOIRE…

**3000 av. J.-C. à 500 apr. J.-C.**

**500 à 1500 apr. J.-C.**

**1750 à 1850**

**2000 à nos jours**

# CONTEXTE

Hausse des Prix de l'Énergie

Complexité Croissante des Réseaux

L'Ère des Données

Urgence Environnementale
Transition Énergétique

# QUELS ENJEUX ?

Détection des Anomalies

Optimisation de la Consommation

Prévision de la Demande

Sécurité des Données

2.

# ENERGIEWATCH, QUELLE RÉPONSE ?

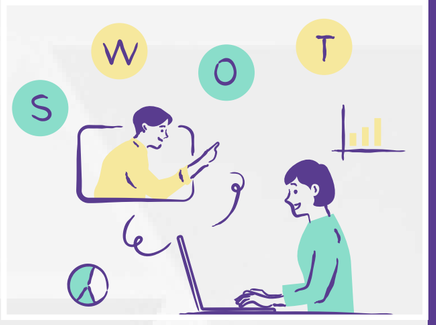
# QUELLE **RÉPONSE ?**

Construire un avenir énergétique plus durable


## COMMENT ?

- Optimisation gestion énergétique
- Surveillance continue
- Détection d'anomalies en temps réel
- Anticipation des besoins énergétiques

# ANALYSE




## S

- Innovation et Flexibilité
- Facilité d'Utilisation
- Adaptabilité
- Orientation Durable

## W

- Manque d'Isolation des Données
- Complexité de Gestion des Mises à Jour

## O

- Marché en Croissance
- Partenariats Stratégiques
- Éducation du Marché
- Évolution Technologique

## T

- Dépendance aux Données
- Vulnérabilité aux Cyberattaques
- Absence de Système de Prévention
- Scalabilité Limitée

# LA CONCURRENCE

Quelques solutions existantes dans le domaine
de la détection d'anomalies



3.

# OBJECTIFS ET LIVRABLES

# OBJECTIFS

Mise en place d’un outil qui permet :

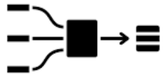
📈 La prédiction de la consommation

🔍 La détection des Anomalies

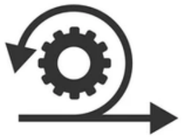
🌍 Le calcul et le suivi de l’empreinte carbone

# LIVRABLES

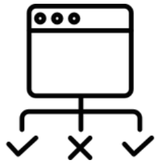
## Pour nos équipes


Pipeline automatisé


Modèles ML


Scripts pour le ré-entraînement

Logs empreinte carbone


Tests unitaires


Documentation Technique

## Pour les clients

Tableau de bord interactif

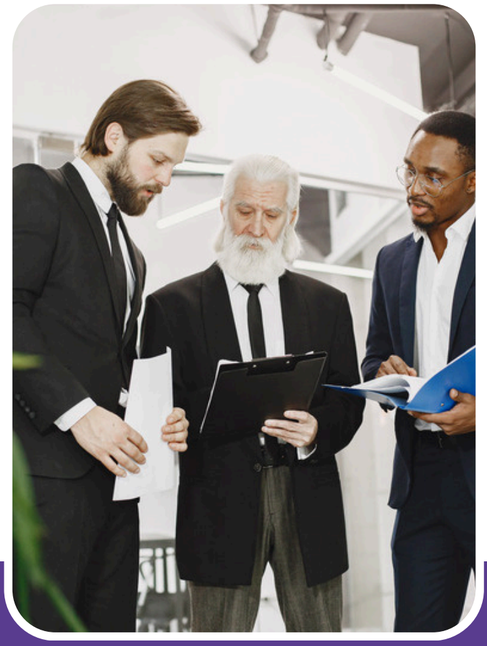

Rapports de prédiction

Rapport d'empreinte carbone


Modèles prédictifs

# CONTRAINTES

Plusieurs contraintes peuvent être rencontrées à différents niveaux :




% Accuracy


Disponibilité

Environnementales

Stockage des logs

Anomalie critique

4.

# NOTRE BUSINESS MODEL

# MODÈLE ÉCONOMIQUE
# OBJECTIF DE RENTABILITÉ

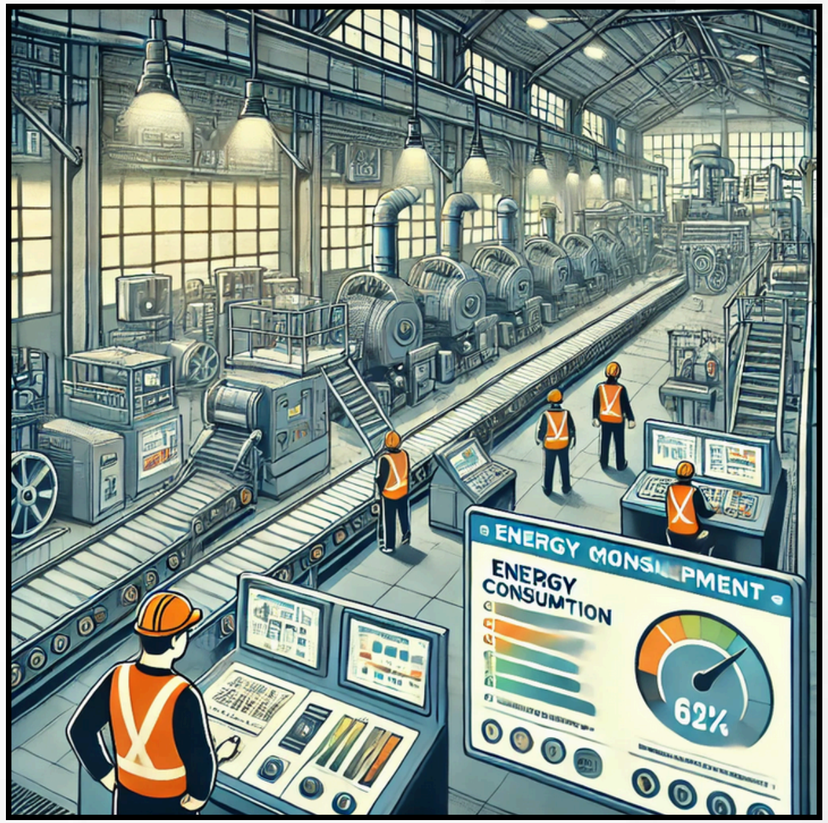
- **Abonnements Mensuels/Annuels :**
  - **Basique : 700** EUR/mois
  - **Avancé : 1500** EUR/mois
  - **Entreprise : 3000** EUR/mois
- **Services Additionnels :**
  - Consulting
  - Formations
- **Premium**

- **Objectif de Rentabilité :**
  - **Abonnement basique :** 8 400 EUR/an
  - **Abonnement Niveau Avancé :** 18 000 EUR/an
  - **Abonnement Niveau Entreprise** : 36 000 EUR/an

# CIBLE MARKET

**Entreprises d'Électricité**


**Industries à Forte Consommation**



# PERSONA

## HELENA 40 ans

**Profession** : Directrice des Opérations d'une Entreprise Énergétique

**Contexte** : Gère la production et la distribution d'énergie dans une grande entreprise d'électricité, avec la nécessité de s'adapter à des fluctuations imprévisibles de la demande causées par des événements sociaux et des changements météorologiques.

**Objectifs :**

- Optimiser l'ajustement de la production énergétique en fonction de la demande tout en réduisant les coûts et en améliorant l'efficacité.
- Répondre rapidement aux perturbations causées par les conditions météorologiques extrêmes ou les mouvements sociaux.
- Suivre et réduire l'empreinte carbone associée aux opérations énergétiques.

**Besoins** :

- Une solution qui intègre des variables sociales et météorologiques pour mieux prédire la demande énergétique.
- Un outil qui permet de détecter rapidement des anomalies dans les données de consommation et de production.
- Un suivi précis de l'empreinte carbone des traitements et des infrastructures IT pour mieux aligner les opérations avec les objectifs environnementaux.

**Ce que EnergieWatch peut apporter :**

- Prédictions fiables basées sur des facteurs sociaux et climatiques pour mieux planifier la production.
- Détection des anomalies dans les données en temps réel pour prévenir les problèmes avant qu'ils n'impactent les opérations.
- Calcul et suivi de l'empreinte carbone des traitements énergétiques, aidant l'entreprise à réduire son impact environnemental.

5.

# GESTION DE PROJET

# APPROCHE

**Identification du besoin**

**Etablir les exigences**

**Planification et suivi**

Trello

- KANBAN
- Planification des tâches

- Canal de discussion
- Planification des sprints
- Partage de documents


# RESSOURCES
# LOGICIELLES

# BUDGETS ALLOUÉ AUX RESSOURCES

- **Coût mensuel** : 500 ---- 1 000 EUR/mois
- **Coût annuel** : 6 000 et 12 000 EUR.

- **Coût mensuel :** 21 EUR/utilisateur**.**
- **Coût annuel (pour 5 utilisateurs) :** 1 260 EUR**.**

- **Coût mensuel :** 200 --- 500 EUR/mois
- **Coût annuel :** 2 400 et 6 000 EUR.

- **Coût mensuel :** Entre 200 et 1 000 EUR/mois en fonction du niveau de ressources alloué (espace, nombre de clusters**).**
- **Coût annuel :** Entre 2 400 et 12 000 EUR

- **Coût mensuel :** Entre 200 et 1 000 EUR/mois en fonction du niveau de ressources alloué (espace, nombre de clusters**).**
- **Coût annuel :** Entre 2 400 et 12 000 EUR

- **Coût mensuel :** Environ 100 à 500 EUR/mois pour une utilisation standard**.**
- **Coût annuel :** Entre 1 200 et 6 000 EUR**.**

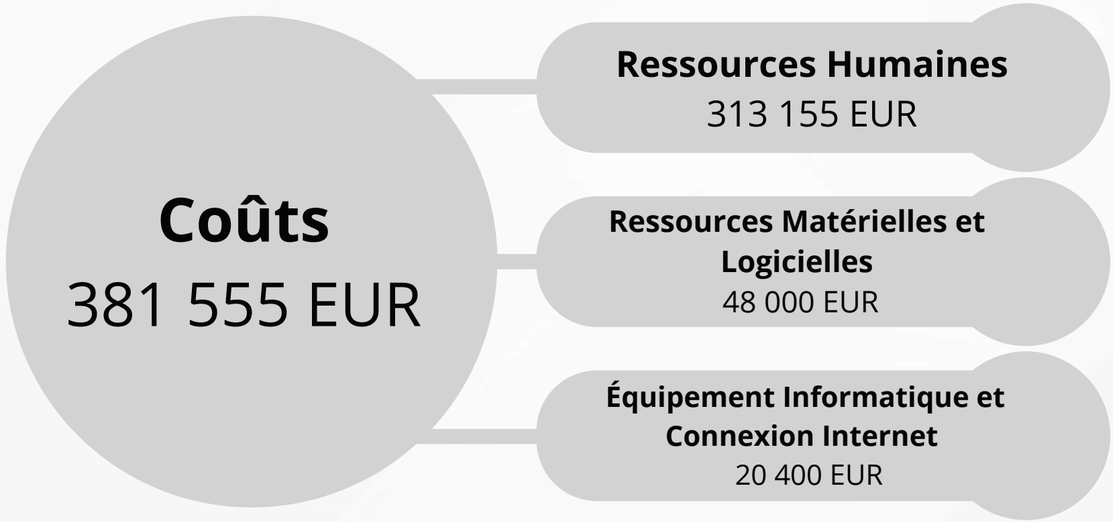
- **Coût mensuel :** Varie grandement selon les ressources utilisées. On peut estimer environ 500 à 1 500 EUR/mois**.**
- **Coût annuel :** Entre 6 000 et 18 000 EUR.

***Coût minimum :***
*21 660 EUR*
***Coût maximum :***
*67 260 EUR*

# RESSOURCES HUMAINES

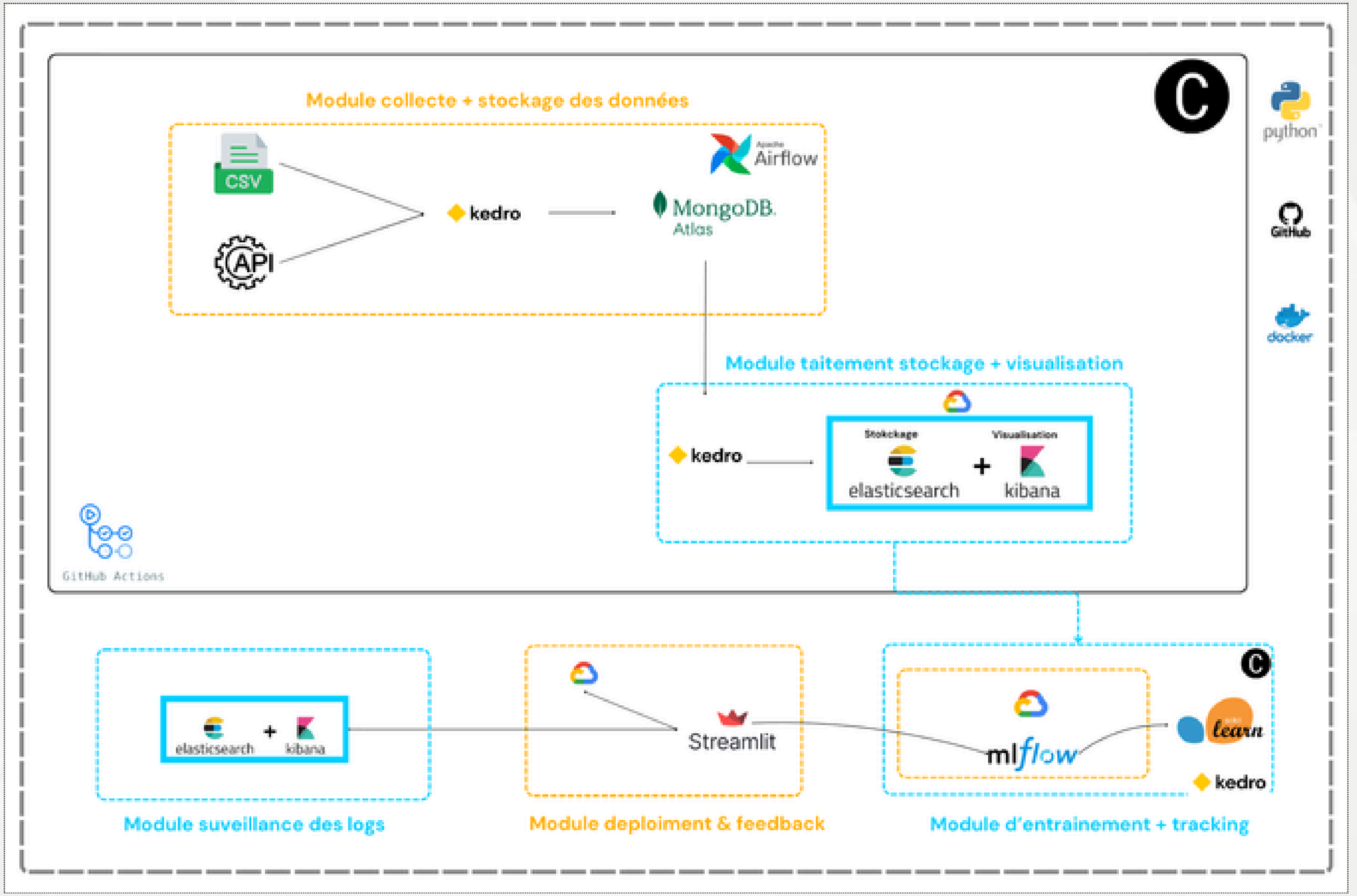
| RÔLE | RESPONSABILITÉS | COMPÉTENCES | COÛT ANNUEL (EUR) |
|---|---|---|---|
| Data Analyst | • Collecter et analyser données<br>• Préparer des rapports et des visualisations | • Python (pandas, NumPy)<br>• Tableau, Power BI ou Kibana | 60,269 |
| Data Engineer | • Concevoir, développer et maintenir les pipelines de données<br>• Assurer la qualité et l'intégrité des données | • MongoDB, SQL<br>• Kedro, Airflow, oozie<br>• GCP, AWS<br>• Python, Java, Docker | 60,269 |
| Ingénieur DevOps | • Automatisation des Processus de Déploiement<br>• Gestion des Infrastructures | • Jenkins, GitHub action<br>• AWS, GCP, Prometheus<br>• Docker, ELK, Grafana | 67,519 |
| Data scientist | • Développer, entraîner, et surveiller les modèles ML<br>• Gérer les expérimentations<br>• Mettre en production les modèles | • scikit-learn, TensorFlow<br>• Docker, MIflow, Airflow<br>• Deep learning | 67,519 |
| Coordinateur / PO | • Gestion de projet, coordination des équipes<br>• Interaction avec les parties prenantes | • Méthode Agiles | 60.269 |

# STRUCTURE DES COÛTS

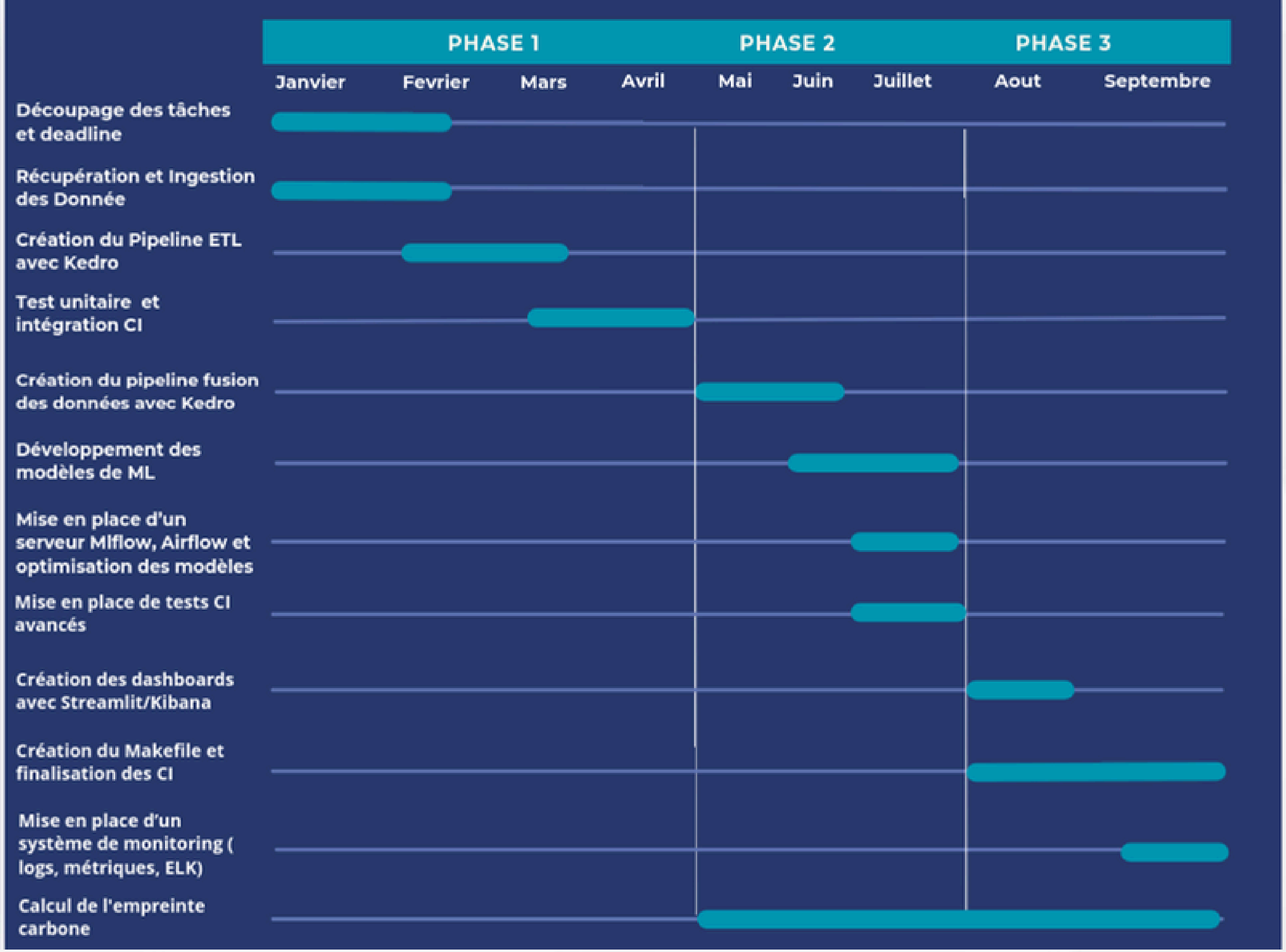


6.

# ARCHITECTURE


Module collecte + stockage des données
CSV
API
kedro
Airflow
MongoDB.
Atlas
Module taitement stockage + visualisation
kedro
Stockage
Visualisation
elasticsearch
+
kibana
GitHub Actions
python
GitHub
docker
elasticsearch
+
kibana
Streamlit
mlflow
learn
kedro
Module suveillance des logs
Module deploiment & feedback
Module d'entrainement + tracking

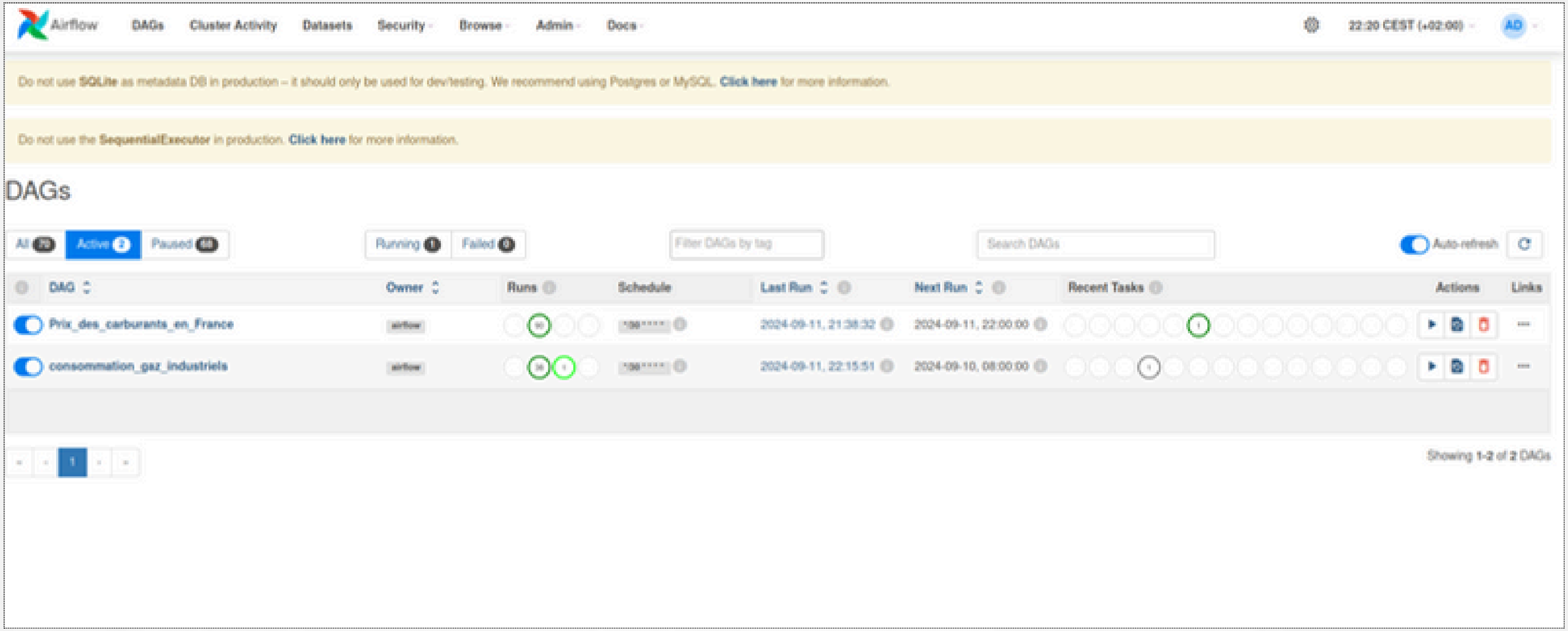
# ROADMAP


PHASE 1
PHASE 2
PHASE 3
Janvier
Fevrier
Mars
Avril
Mai
Juin
Juillet
Aout
Septembre
Découpage des tâches et deadline
Récupération et Ingestion des Donnée
Création du Pipeline ETL avec Kedro
Test unitaire et intégration CI
Création du pipeline fusion des données avec Kedro
Développement des modèles de ML
Mise en place d'un serveur Mlflow, Airflow et optimisation des modèles
Mise en place de tests CI avancés
Création des dashboards avec Streamlit/Kibana
Création du Makefile et finalisation des CI
Mise en place d'un système de monitoring ( logs, métriques, ELK)
Calcul de l'empreinte carbone

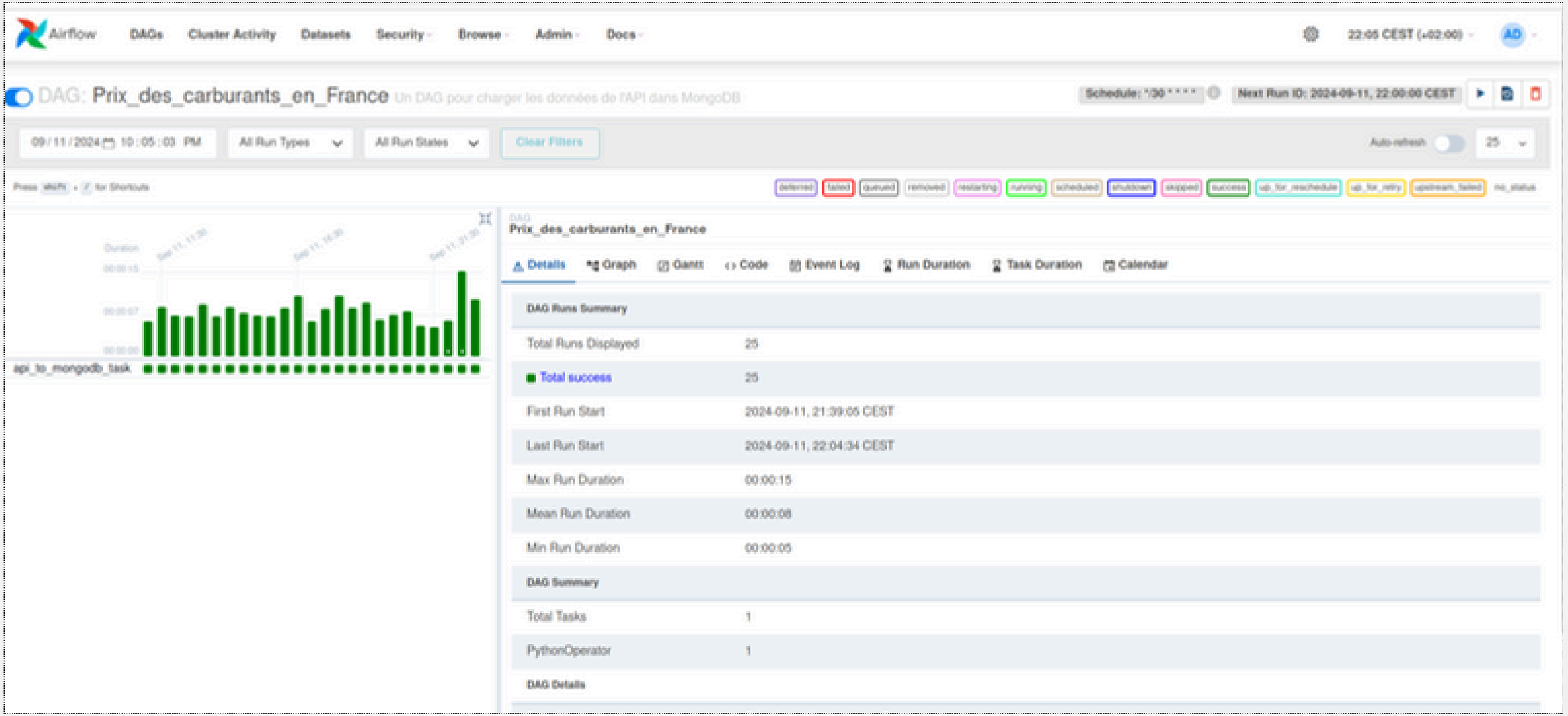
7.

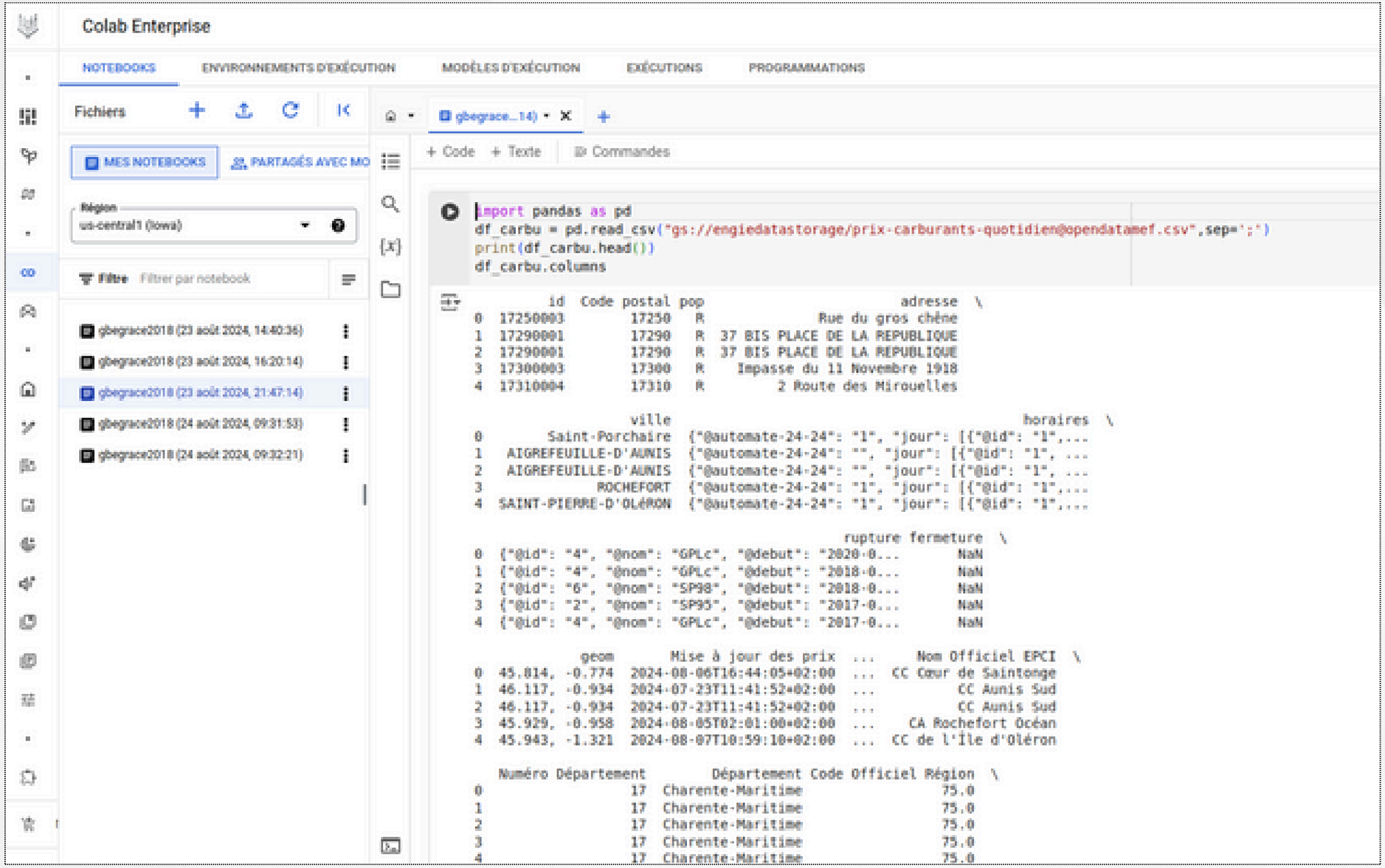
# FLUX DE TRAVAIL : DE L'ETL AU DÉPLOIMENT CLOUD

# AIRFLOW


Airflow
DAGs
Cluster Activity
Datasets
Security
Browse
Admin
Docs
22:20 CEST (+02:00)
AD
Do not use SQLite as metadata DB in production – it should only be used for dev/testing. We recommend using Postgres or MySQL. Click here for more information.
Do not use the SequentialExecutor in production. Click here for more information.
DAGs
All 2
Active 2
Paused 0
Running 1
Failed 0
Filter DAGs by tag
Search DAGs
Auto-refresh
DAG
Owner
Runs
Schedule
Last Run
Next Run
Recent Tasks
Actions
Links
Prix_des_carburants_en_France
airflow
2024-09-11, 21:38:32
2024-09-11, 22:00:00
consommation_gaz_industriels
airflow
2024-09-11, 22:15:51
2024-09-10, 08:00:00
1
Showing 1-2 of 2 DAGs

# AIRFLOW


Airflow
DAGs
Cluster Activity
Datasets
Security
Browse
Admin
Docs
22:05 CEST (+02:00)
AD
DAG: Prix_des_carburants_en_France
Un DAG pour charger les données de l'API dans MongoDB
Schedule: '*/30 * * * *'
Next Run ID: 2024-09-11, 22:00:00 CEST
09/11/2024 10:05:03 PM
All Run Types
All Run States
Clear Filters
Auto-refresh
25
Duration
api_to_mongodb_task
DAG
Prix_des_carburants_en_France
Details
Graph
Gantt
Code
Event Log
Run Duration
Task Duration
Calendar
DAG Runs Summary
Total Runs Displayed 25
Total success 25
First Run Start 2024-09-11, 21:39:05 CEST
Last Run Start 2024-09-11, 22:04:34 CEST
Max Run Duration 00:00:15
Mean Run Duration 00:00:08
Min Run Duration 00:00:05
DAG Summary
Total Tasks 1
PythonOperator 1
DAG Details




# GOOGLE COLAB

# PIPELINES

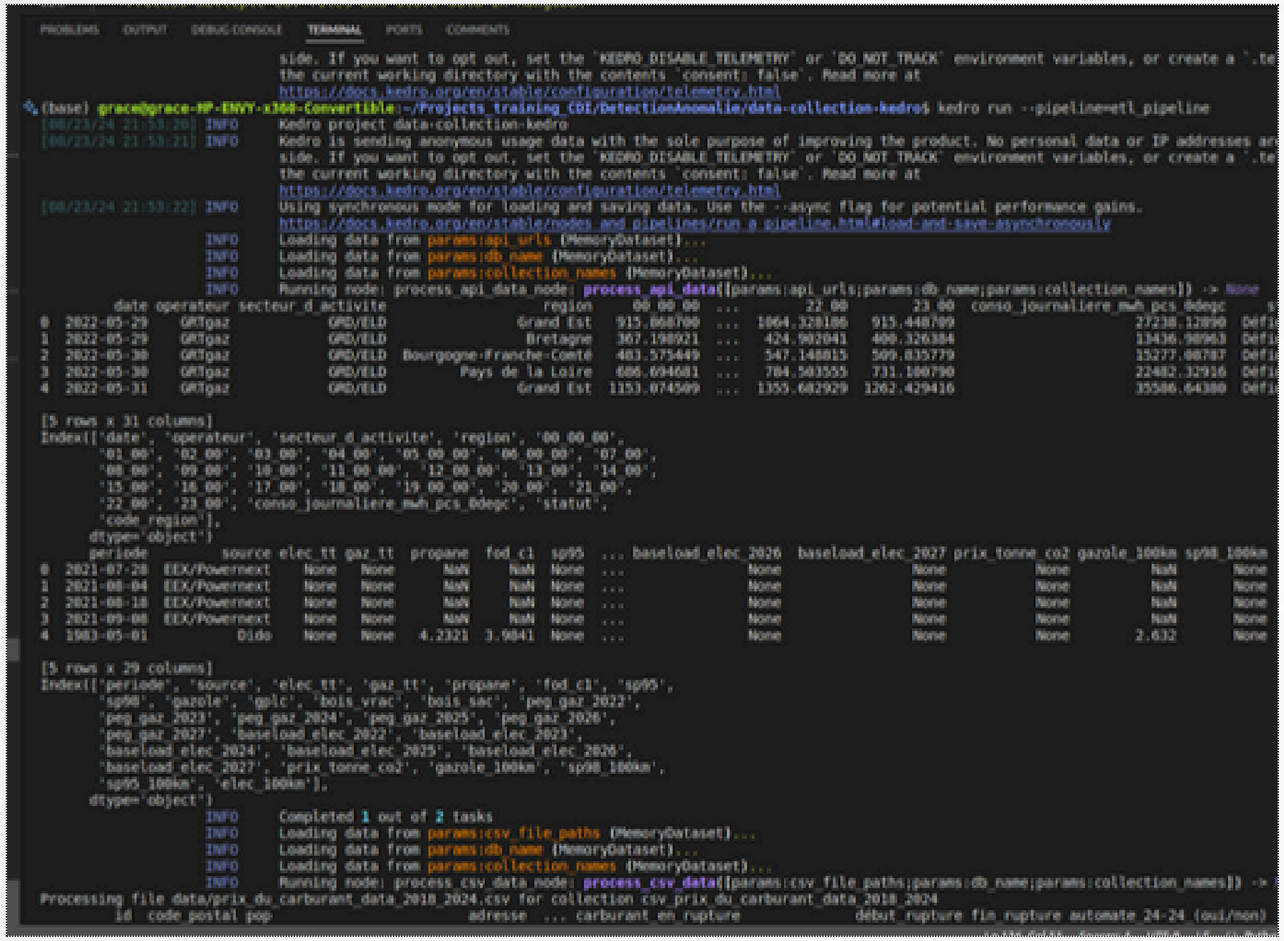
data-collection-kedro > src > data_collection_kedro > pipelines > etl_pipeline > pipeline.py > create_pipeline

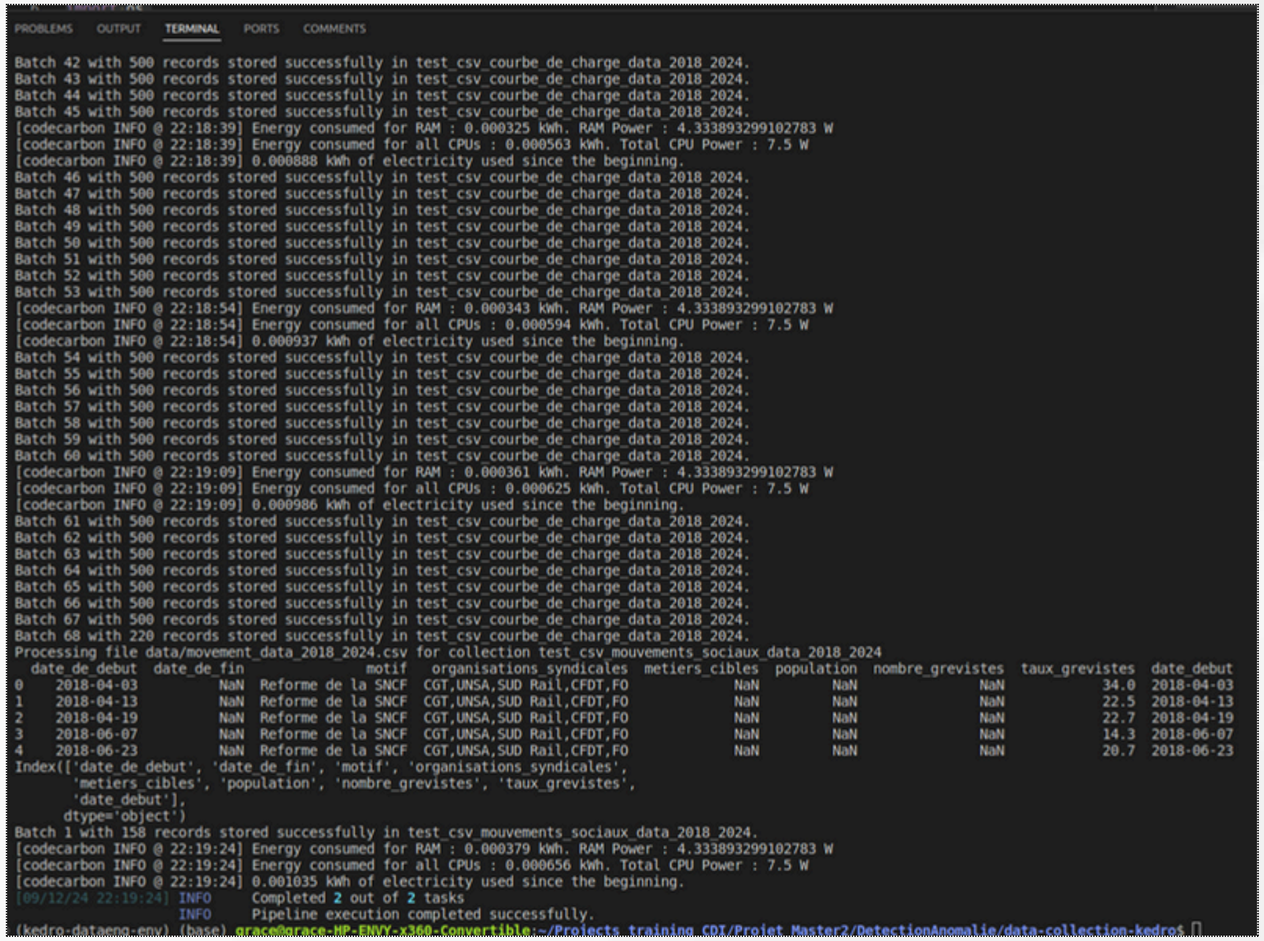
```python
"""
pipeline 'etl_pipeline'
"""
from kedro.pipeline import Pipeline, node

from .nodes import etl_api_data, etl_csv_data

def create_pipeline(**kwargs) -> Pipeline:
    """
    Create a Kedro pipeline for the ETL process with multiple API sources.
    Returns::
        Pipeline: The constructed Kedro pipeline.
    """
    return Pipeline(
        [
            node(
                func=etl_api_data,
                inputs=["params:api_urls", "params:db_name", "params:collection_names"],
                outputs=None,
                name="process_api_data_node",
            ),
            node(
                func=etl_csv_data,
                inputs=[
                    "params:csv_file_paths",
                    "params:db_name",
                    "params:collection_names",
                ],
                outputs=None,
                name="process_csv_data_node",
            ),
        ]
    )

```

data-collection-kedro > src > data_collection_kedro > pipelines > data_fusion_pipeline > pipeline.py > ...

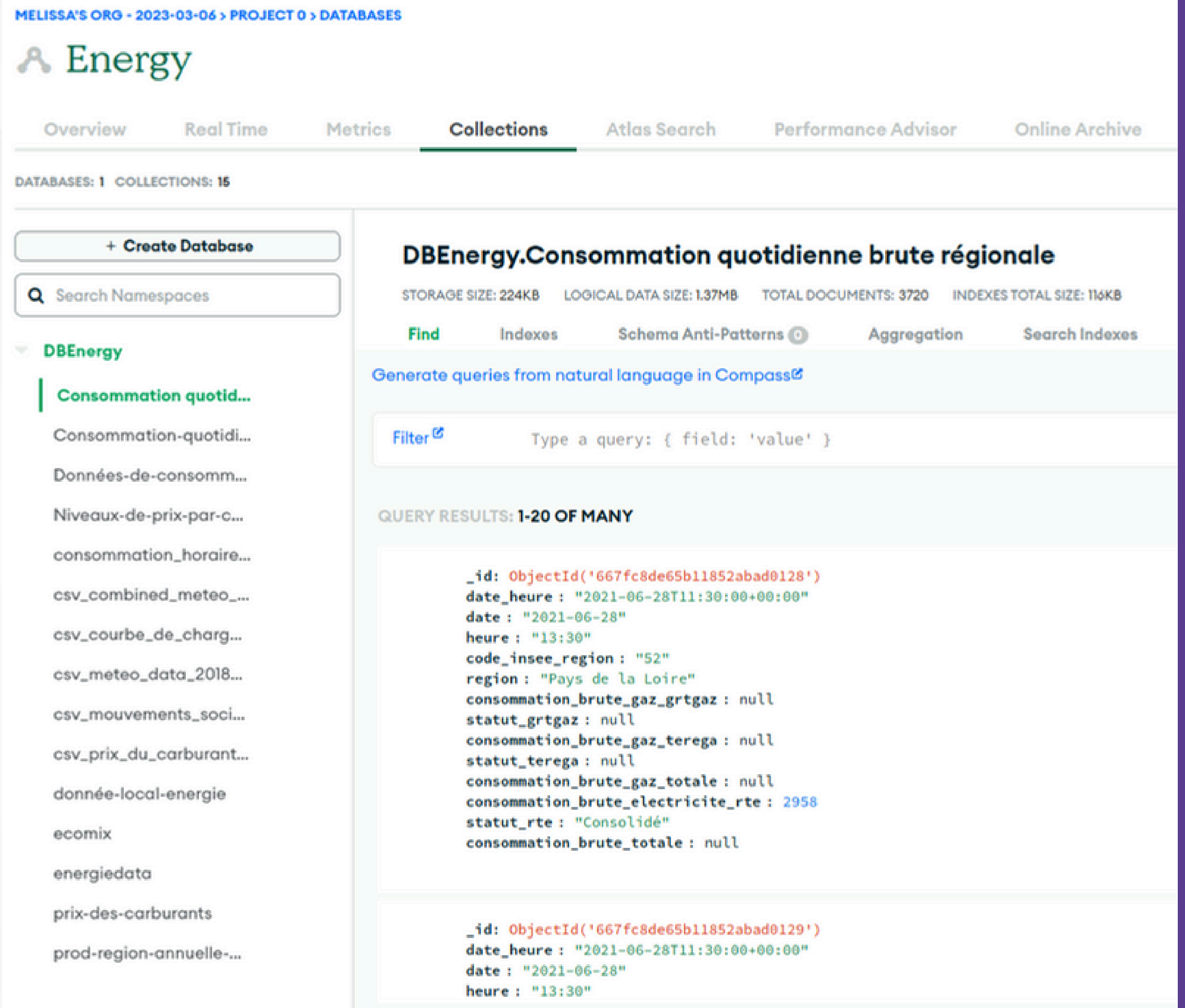
```python
"""Data fusion pipeline module."""
from kedro.pipeline import Pipeline, node

from .nodes import (
    display_dataframes,
    load_collections,
    merge_data_store_in_elastic,
    normalize_columns,
    select_columns,
)

def create_pipeline(**kwargs) -> Pipeline:
    return Pipeline(
        [
            node(
                func=load_collections,
                inputs=dict(
                    collection_names="params:collection_names1",
                    db_name="params:db_name1",
                ),
                outputs="loaded_dataframes",
                name="load_collections_node",
            ),
            node(
                func=select_columns,
                inputs=dict(
                    dataframes="loaded_dataframes",
                    columns_to_select="params:columns_to_select",
                ),
                outputs="selected_dataframes",
                name="select_columns_node",
            ),
            node(
                func=display_dataframes,
                inputs="selected_dataframes",
                outputs="displayed_selected_data",
                name="display_selected_data_node",
            ),
            node(
                func=normalize_columns,
                inputs="selected_dataframes",
                outputs="normalized_dataframes",
                name="normalize_columns_node",
            ),
            node(
                func=merge_data_store_in_elastic,
                inputs=["normalized_dataframes"],
                outputs=["merged_meteo_courbe", "merged_courbe_mouvement"],
                name="merge_data_node",
            ),
        ]
    )

```

# ETL PIPELINE



# MONGODB

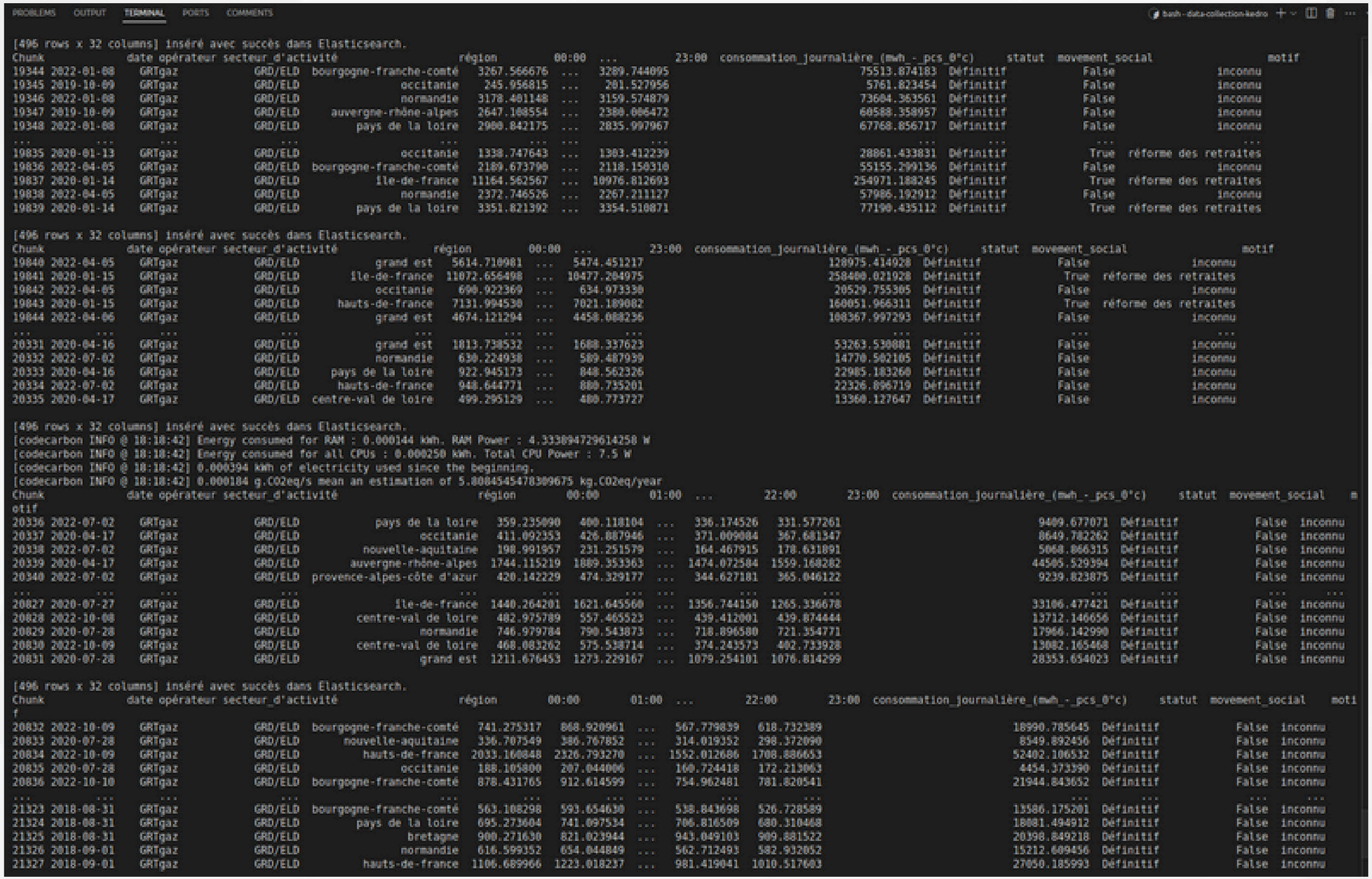

MELISSA'S ORG - 2023-03-06 > PROJECT 0 > DATABASES
Energy
Overview
Real Time
Metrics
Collections
Atlas Search
Performance Advisor
Online Archive
DATABASES: 1 COLLECTIONS: 15
+ Create Database
Search Namespaces
DBEnergy
Consommation quotid...
Consommation-quotidi...
Données-de-consomm...
Niveaux-de-prix-par-c...
consommation_horaire...
csv_combined_meteo_...
csv_courbe_de_charg...
csv_meteo_data_2018...
csv_mouvements_soci...
csv_prix_du_carburant...
donnée-local-energie
ecomix
energiedata
prix-des-carburants
prod-region-annuelle-...
DBEnergy.Consommation quotidienne brute régionale
STORAGE SIZE: 224KB LOGICAL DATA SIZE: 1.37MB TOTAL DOCUMENTS: 3720 INDEXES TOTAL SIZE: 116KB
Find
Indexes
Schema Anti-Patterns
Aggregation
Search Indexes
Generate queries from natural language in Compass
Filter
Type a query: { field: 'value' }
QUERY RESULTS: 1-20 OF MANY
_id: ObjectId('667fc8de65b11852abad0128')
date_heure : "2021-06-28T11:30:00+00:00"
date : "2021-06-28"
heure : "13:30"
code_insee_region : "52"
region : "Pays de la Loire"
consommation_brute_gaz_grtgaz : null
statut_grtgaz : null
consommation_brute_gaz_terega : null
statut_terega : null
consommation_brute_gaz_totale : null
consommation_brute_electricite_rte : 2958
statut_rte : "Consolidé"
consommation_brute_totale : null
_id: ObjectId('667fc8de65b11852abad0129')
date_heure : "2021-06-28T11:30:00+00:00"
date : "2021-06-28"
heure : "13:30"

# DATA FUSION PIPELINE

```
PROBLEMS  OUTPUT  TERMINAL  PORTS  COMMENTS                                                                    python3.11 - data-collection-kedro

      dtype='object')
[09/10/24 12:53:06] INFO     Saving data to loaded_dataframes (MemoryDataset)...                                                  data_catalo
g.py:583
                    INFO     Completed 1 out of 5 tasks                                                                           sequential_runn
er.py:93
                    INFO     Loading data from loaded_dataframes (MemoryDataset)...                                               data_catalo
g.py:539
                    INFO     Loading data from params:columns_to_select (MemoryDataset)...                                        data_catalo
g.py:539
                    INFO     Running node: select_columns_node: select_columns([loaded_dataframes;params:columns_to_select]) -> [selected_dataframes]  nod
e.py:364
Colonnes disponibles dans le DataFrame: Index(['_id', 'Unnamed: 0', 'TempMax_Deg', 'TempMin_Deg', 'Wind_kmh',
       'Wet_percent', 'Visibility_km', 'CloudCoverage_percent',
       'Dayduration_hour', 'region', 'day'],
      dtype='object')
Colonnes disponibles dans le DataFrame: Index(['_id', 'date', 'opérateur', 'secteur_d'activité', 'région', '00:00',
       '01:00', '02:00', '03:00', '04:00', '05:00', '06:00', '07:00', '08:00',
       '09:00', '10:00', '11:00', '12:00', '13:00', '14:00', '15:00', '16:00',
       '17:00', '18:00', '19:00', '20:00', '21:00', '22:00', '23:00',
       'consommation_journalière_(mwh_-_pcs_0°c)', 'statut', 'code_région'],
      dtype='object')
Colonnes disponibles dans le DataFrame: Index(['_id', 'date_de_debut', 'date_de_fin', 'motif',
       'organisations_syndicales', 'metiers_cibles', 'population',
       'nombre_grevistes', 'taux_grevistes', 'date_debut'],
      dtype='object')
                    INFO     Saving data to selected_dataframes (MemoryDataset)...                                                data_catalo
g.py:583
                    INFO     Completed 2 out of 5 tasks                                                                           sequential_runn
er.py:93
                    INFO     Loading data from selected_dataframes (MemoryDataset)...                                             data_catalo
g.py:539
                    INFO     Running node: display_selected_data_node: display_dataframes([selected_dataframes]) -> [displayed_selected_data]     nod
e.py:364
DataFrame 1:
   TempMax_Deg  TempMin_Deg  Wind_kmh  Wet_percent  Visibility_km  CloudCoverage_percent Dayduration_hour     region         day
0          9.0          6.0        25           74          8.250                   70.0          8:23:0     alsace  2018/01/01
1         11.0          7.0        47           78          9.750                   42.0          8:27:0   bretagne  2018/01/01
2          8.0          6.0        38           83          9.500                   91.0          8:18:0   lorraine  2018/01/01
3         14.0         11.0        23           95          7.625                   96.0          8:53:0  aquitaine  2018/01/02
4         12.0          5.0        20           93          8.000                   78.0          8:29:0     centre  2018/01/02


DataFrame 2:
         date opérateur secteur_d'activité                  région       00:00  ...       21:00       22:00       23:00  consommation_journalière_(mwh_-_pcs_0°c)     stat
ut
0  2018-11-15    GRTgaz           GRD/ELD           Hauts-de-France  6880.095246  ...  6378.954872  6741.885649  6930.526476                                 148113.988853  Définit
if
```

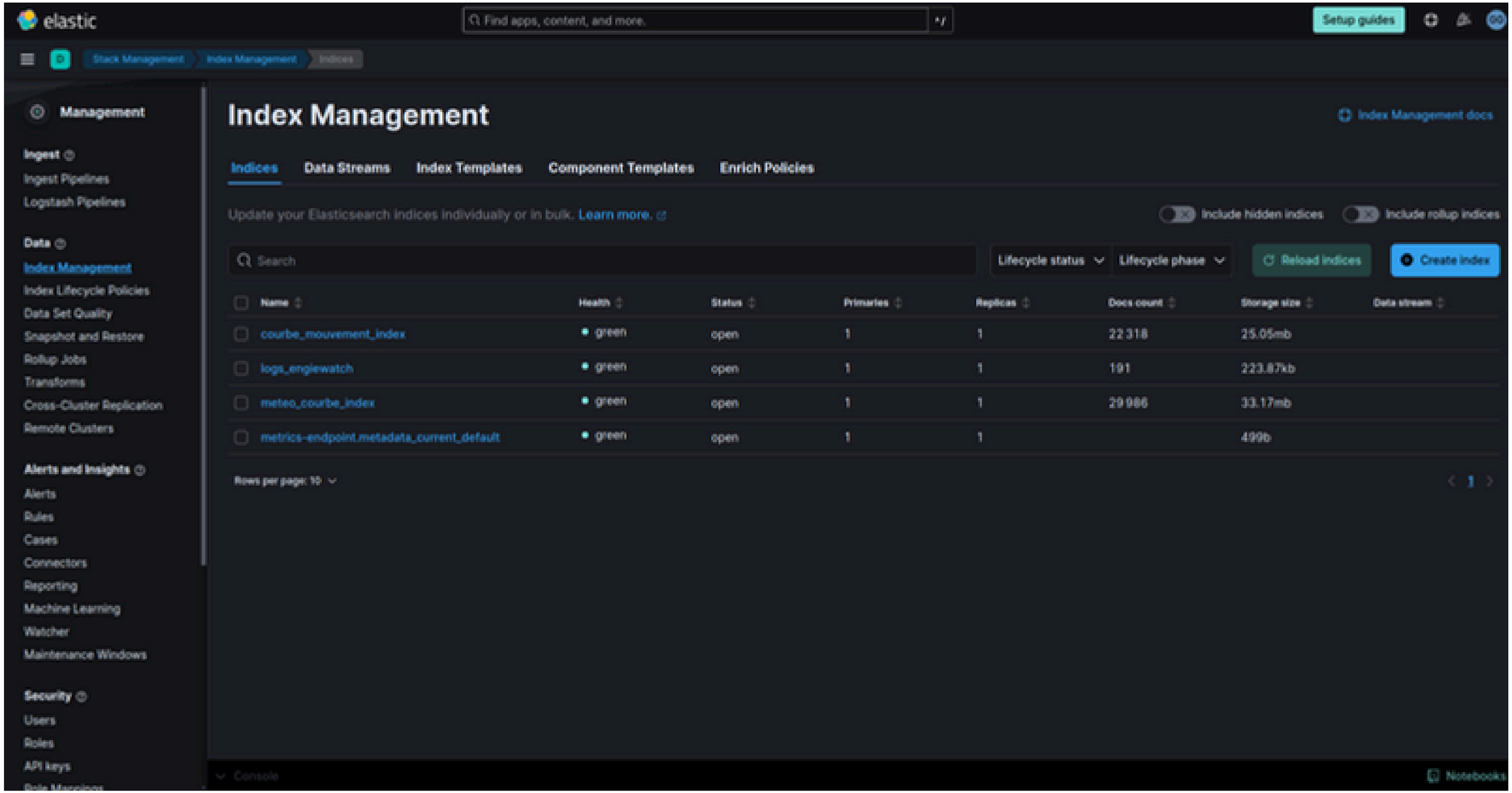
```
Total carbon emissions for this run: 1.3437832202069785e-05 kgCO2eq
                    INFO     Saving data to merged_meteo_courbe (MemoryDataset)...                                                data_c
                    INFO     Saving data to merged_courbe_mouvement (MemoryDataset)...                                            data_c
                    INFO     Completed 5 out of 5 tasks                                                                           sequential
                    INFO     Pipeline execution completed successfully.
                    INFO     Loading data from merged_meteo_courbe (MemoryDataset)...                                             data_c
                    INFO     Loading data from merged_courbe_mouvement (MemoryDataset)...                                         data_c
                    INFO     Loading data from displayed_selected_data (MemoryDataset)...                                         data_c
(kedro-dataeng-env) (base) grace@grace-HP-ENVY-x360-Convertible:~/Projects training CDI/Projet Master2/DetectionAnomalie/data-collection-kedro$
```

# DATA FUSION PIPELINE

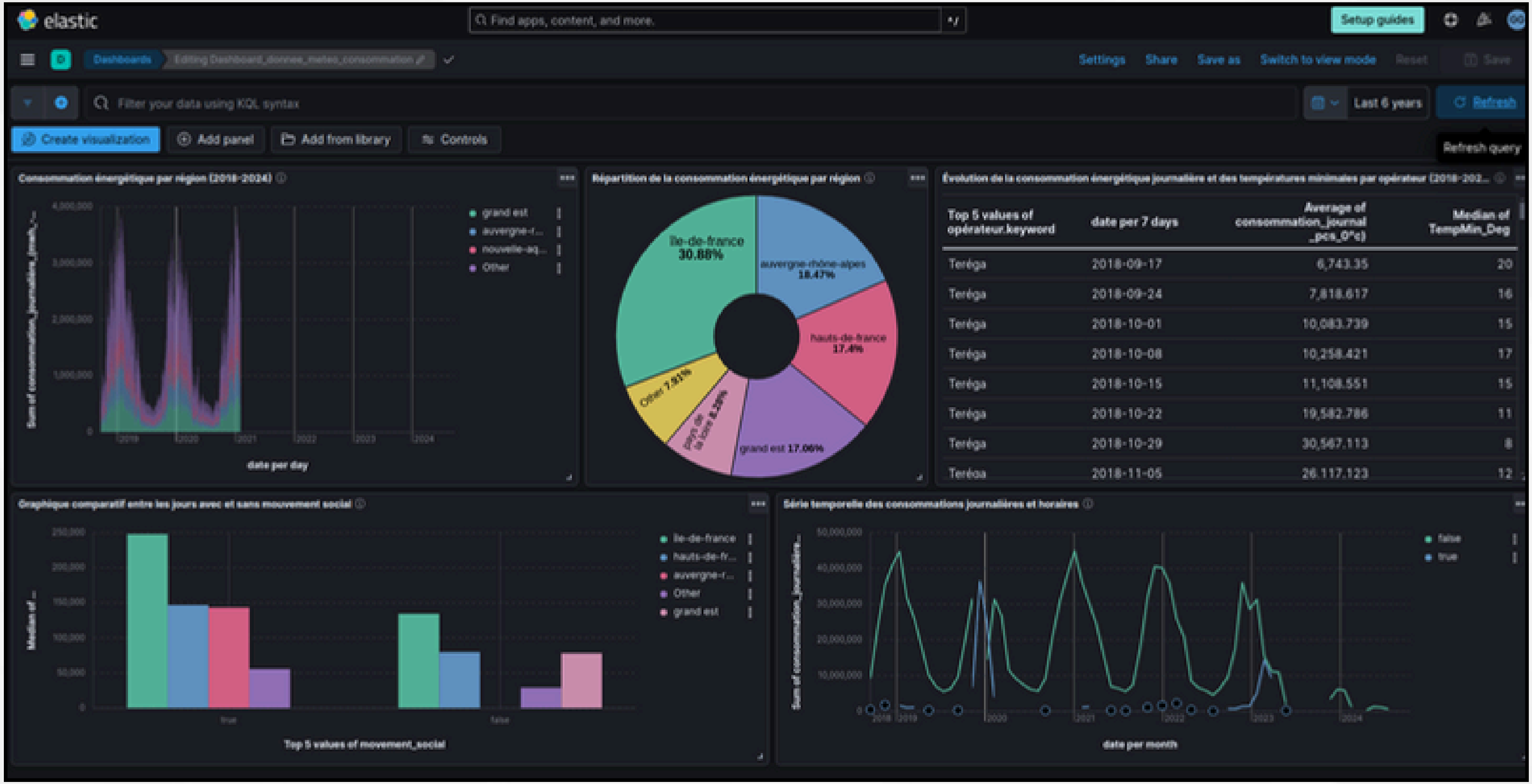


```
PROBLEMS  OUTPUT  TERMINAL  PORTS  COMMENTS                                                    bash - data-collection-kedro

[496 rows x 32 columns] inséré avec succès dans Elasticsearch.
Chunk            date opérateur secteur_d'activité              région        00:00  ...        23:00  consommation_journalière_(mwh_-_pcs_0°c)     statut  movement_social               motif
19344 2022-01-08    GRTgaz           GRD/ELD  bourgogne-franche-comté   3267.566676  ...   3289.744095                                  75513.874183  Définitif            False               inconnu
19345 2019-10-09    GRTgaz           GRD/ELD               occitanie    245.956815  ...    281.527956                                   5761.823454  Définitif            False               inconnu
19346 2022-01-08    GRTgaz           GRD/ELD              normandie   3178.401148  ...   3159.574879                                  73604.363561  Définitif            False               inconnu
19347 2019-10-09    GRTgaz           GRD/ELD  auvergne-rhône-alpes   2647.108554  ...   2380.006472                                  60588.358057  Définitif            False               inconnu
19348 2022-01-08    GRTgaz           GRD/ELD       pays de la loire   2900.842175  ...   2835.997967                                  67768.856717  Définitif            False               inconnu
...              ...      ...              ...                ...          ...  ...          ...                                           ...        ...              ...                   ...
19835 2020-01-13    GRTgaz           GRD/ELD               occitanie   1338.747643  ...   1303.412239                                  28861.433831  Définitif             True  réforme des retraites
19836 2022-04-05    GRTgaz           GRD/ELD  bourgogne-franche-comté   2189.673790  ...   2118.158310                                  55155.299136  Définitif            False               inconnu
19837 2020-01-14    GRTgaz           GRD/ELD          ile-de-france  11164.562567  ...  10976.812603                                 254071.188245  Définitif             True  réforme des retraites
19838 2022-04-05    GRTgaz           GRD/ELD              normandie   2372.746526  ...   2267.211127                                  57986.192912  Définitif            False               inconnu
19839 2020-01-14    GRTgaz           GRD/ELD       pays de la loire   3351.821392  ...   3354.510871                                  77190.435112  Définitif             True  réforme des retraites

[496 rows x 32 columns] inséré avec succès dans Elasticsearch.
Chunk            date opérateur secteur_d'activité              région        00:00  ...        23:00  consommation_journalière_(mwh_-_pcs_0°c)     statut  movement_social               motif
19840 2022-04-05    GRTgaz           GRD/ELD             grand est   5614.710981  ...   5474.451217                                 128975.414928  Définitif            False               inconnu
19841 2020-01-15    GRTgaz           GRD/ELD          ile-de-france  11072.656498  ...  10477.204975                                 258400.021928  Définitif             True  réforme des retraites
19842 2022-04-05    GRTgaz           GRD/ELD               occitanie    690.922369  ...    634.973330                                  20529.755305  Définitif            False               inconnu
19843 2020-01-15    GRTgaz           GRD/ELD        hauts-de-france   7131.994530  ...   7021.189082                                 160051.966311  Définitif             True  réforme des retraites
19844 2022-04-06    GRTgaz           GRD/ELD             grand est   4674.121294  ...   4458.088236                                 108367.997293  Définitif            False               inconnu
...              ...      ...              ...                ...          ...  ...          ...                                           ...        ...              ...                   ...
20331 2020-04-16    GRTgaz           GRD/ELD             grand est   1013.738532  ...   1608.337623                                  53263.530881  Définitif            False               inconnu
20332 2022-07-02    GRTgaz           GRD/ELD              normandie    630.224938  ...    589.487939                                  14770.582105  Définitif            False               inconnu
20333 2020-04-16    GRTgaz           GRD/ELD       pays de la loire    922.945173  ...    848.562326                                  22985.183260  Définitif            False               inconnu
20334 2022-07-02    GRTgaz           GRD/ELD        hauts-de-france    948.644771  ...    880.735201                                  22326.896719  Définitif            False               inconnu
20335 2020-04-17    GRTgaz           GRD/ELD  centre-val de loire    499.295129  ...    480.773727                                  13360.127647  Définitif            False               inconnu

[496 rows x 32 columns] inséré avec succès dans Elasticsearch.
[codecarbon INFO @ 18:18:42] Energy consumed for RAM : 0.000144 kWh. RAM Power : 4.333894729614258 W
[codecarbon INFO @ 18:18:42] Energy consumed for all CPUs : 0.000250 kWh. Total CPU Power : 7.5 W
[codecarbon INFO @ 18:18:42] 0.000394 kWh of electricity used since the beginning.
[codecarbon INFO @ 18:18:42] 0.000184 g.CO2eq/s mean an estimation of 5.8084545478309675 kg.CO2eq/year
Chunk            date opérateur secteur_d'activité                    région       00:00       01:00  ...       22:00       23:00  consommation_journalière_(mwh_-_pcs_0°c)     statut  movement_social    m
otif
20326 2022-07-02    GRTgaz           GRD/ELD             pays de la loire   359.235890   400.118104  ...   336.174526   331.577261                                 9409.677071  Définitif            False  inconnu
20327 2020-04-17    GRTgaz           GRD/ELD                  occitanie   411.092353   426.887946  ...   371.009084   367.681347                                 8649.782262  Définitif            False  inconnu
20328 2022-07-02    GRTgaz           GRD/ELD           nouvelle-aquitaine   198.991957   231.251579  ...   164.467915   178.631891                                 5068.866315  Définitif            False  inconnu
20329 2020-04-17    GRTgaz           GRD/ELD       auvergne-rhône-alpes  1744.115219  1889.353363  ...  1474.072584  1559.160282                                44505.529394  Définitif            False  inconnu
20340 2022-07-02    GRTgaz           GRD/ELD  provence-alpes-côte d'azur   420.142229   474.329177  ...   344.627181   365.046122                                 9239.823875  Définitif            False  inconnu
...              ...      ...              ...                    ...          ...          ...  ...          ...          ...                                           ...        ...              ...    ...
20827 2020-07-27    GRTgaz           GRD/ELD                 ile-de-france  1440.264201  1621.645560  ...  1356.744150  1265.336678                                33186.477421  Définitif            False  inconnu
20828 2022-10-08    GRTgaz           GRD/ELD           centre-val de loire   482.975789   557.465523  ...   439.412001   439.874444                                13712.146656  Définitif            False  inconnu
20829 2020-07-28    GRTgaz           GRD/ELD                  normandie   746.979784   790.543873  ...   718.896580   721.354771                                17966.142990  Définitif            False  inconnu
20830 2022-10-09    GRTgaz           GRD/ELD           centre-val de loire   468.083262   575.538714  ...   374.243573   402.733928                                13082.165468  Définitif            False  inconnu
20831 2020-07-28    GRTgaz           GRD/ELD                    grand est  1211.676453  1273.229167  ...  1079.254101  1076.814299                                28353.654023  Définitif            False  inconnu

[496 rows x 32 columns] inséré avec succès dans Elasticsearch.
Chunk            date opérateur secteur_d'activité                    région       00:00       01:00  ...       22:00       23:00  consommation_journalière_(mwh_-_pcs_0°c)     statut  movement_social    moti
f
20832 2022-10-09    GRTgaz           GRD/ELD  bourgogne-franche-comté   741.275317   868.920961  ...   567.779839   618.732389                                18990.785645  Définitif            False  inconnu
20833 2020-07-28    GRTgaz           GRD/ELD           nouvelle-aquitaine   336.707549   386.767852  ...   314.019352   298.372090                                 8549.892456  Définitif            False  inconnu
20834 2022-10-09    GRTgaz           GRD/ELD        hauts-de-france  2033.160848  2326.793270  ...  1552.012686  1708.886653                                52402.106532  Définitif            False  inconnu
20835 2020-07-28    GRTgaz           GRD/ELD               occitanie   188.105800   207.044006  ...   160.724418   172.213063                                 4454.373390  Définitif            False  inconnu
20836 2022-10-10    GRTgaz           GRD/ELD  bourgogne-franche-comté   878.431765   912.614599  ...   754.962481   781.820541                                21944.843652  Définitif            False  inconnu
...              ...      ...              ...                    ...          ...          ...  ...          ...          ...                                           ...        ...              ...    ...
21323 2018-08-31    GRTgaz           GRD/ELD  bourgogne-franche-comté   563.108298   593.654630  ...   538.843698   526.728589                                13586.175201  Définitif            False  inconnu
21324 2018-08-31    GRTgaz           GRD/ELD       pays de la loire   695.273604   741.097534  ...   706.816509   680.310468                                18081.494912  Définitif            False  inconnu
21325 2018-08-31    GRTgaz           GRD/ELD               bretagne   900.271630   821.023944  ...   943.049103   909.881522                                20398.849218  Définitif            False  inconnu
21326 2018-09-01    GRTgaz           GRD/ELD              normandie   616.599352   654.044849  ...   562.712493   582.932052                                15212.609456  Définitif            False  inconnu
21327 2018-09-01    GRTgaz           GRD/ELD        hauts-de-france  1106.689966  1223.018237  ...   981.419041  1010.517603                                27050.185993  Définitif            False  inconnu
```

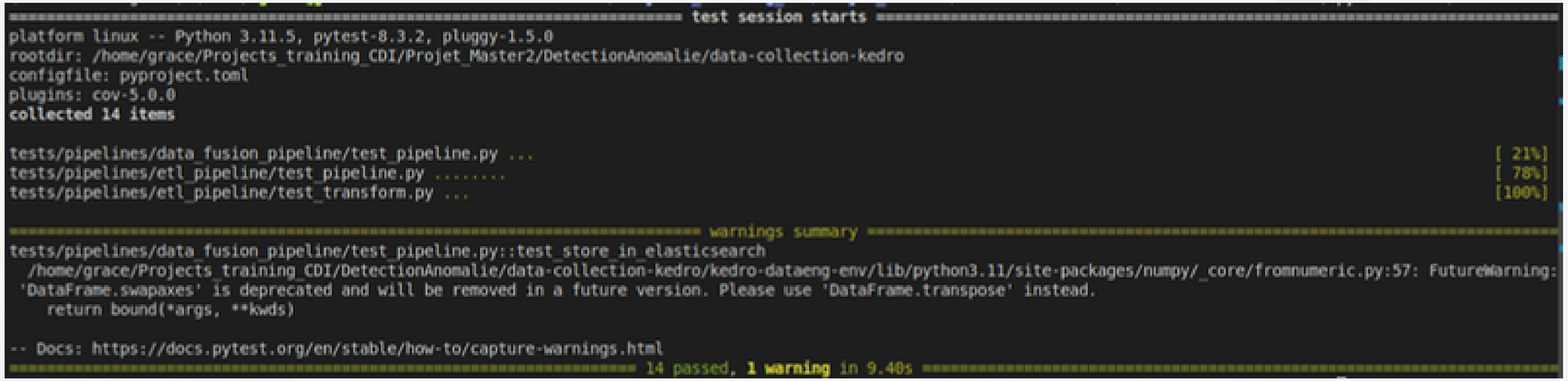
# ELASTIC SEARCH



# KIBANA


elastic
Find apps, content, and more.
Setup guides
Dashboards
Editing Dashboard donnee_meteo_consommation
Settings
Share
Save as
Switch to view mode
Reset
Save
Filter your data using KQL syntax
Last 6 years
Refresh
Refresh query
Create visualization
Add panel
Add from library
Controls
Consommation énergétique par région (2018-2024)
grand est
auvergne-r...
nouvelle-aq...
Other
date per day
Répartition de la consommation énergétique par région
île-de-france 30.88%
auvergne-rhône-alpes 18.47%
hauts-de-france 17.4%
grand est 17.06%
Other 7.31%
Évolution de la consommation énergétique journalière et des températures minimales par opérateur (2018-202...
Top 5 values of opérateur.keyword
date per 7 days
Average of consommation_journal _pcs_0°c)
Median of TempMin_Deg
Teréga 2018-09-17 6,743.35 20
Teréga 2018-09-24 7,818.617 16
Teréga 2018-10-01 10,083.739 15
Teréga 2018-10-08 10,258.421 17
Teréga 2018-10-15 11,108.551 15
Teréga 2018-10-22 19,582.786 11
Teréga 2018-10-29 30,567.113 8
Teréga 2018-11-05 26,117.123 12
Graphique comparatif entre les jours avec et sans mouvement social
île-de-france
hauts-de-fr...
auvergne-r...
Other
grand est
Top 5 values of movement_social
Série temporelle des consommations journalières et horaires
false
true
date per month

# RESULTATS
# TEST UNITAIRES



8.

# MODÈLE DE MACHINE LEARNING

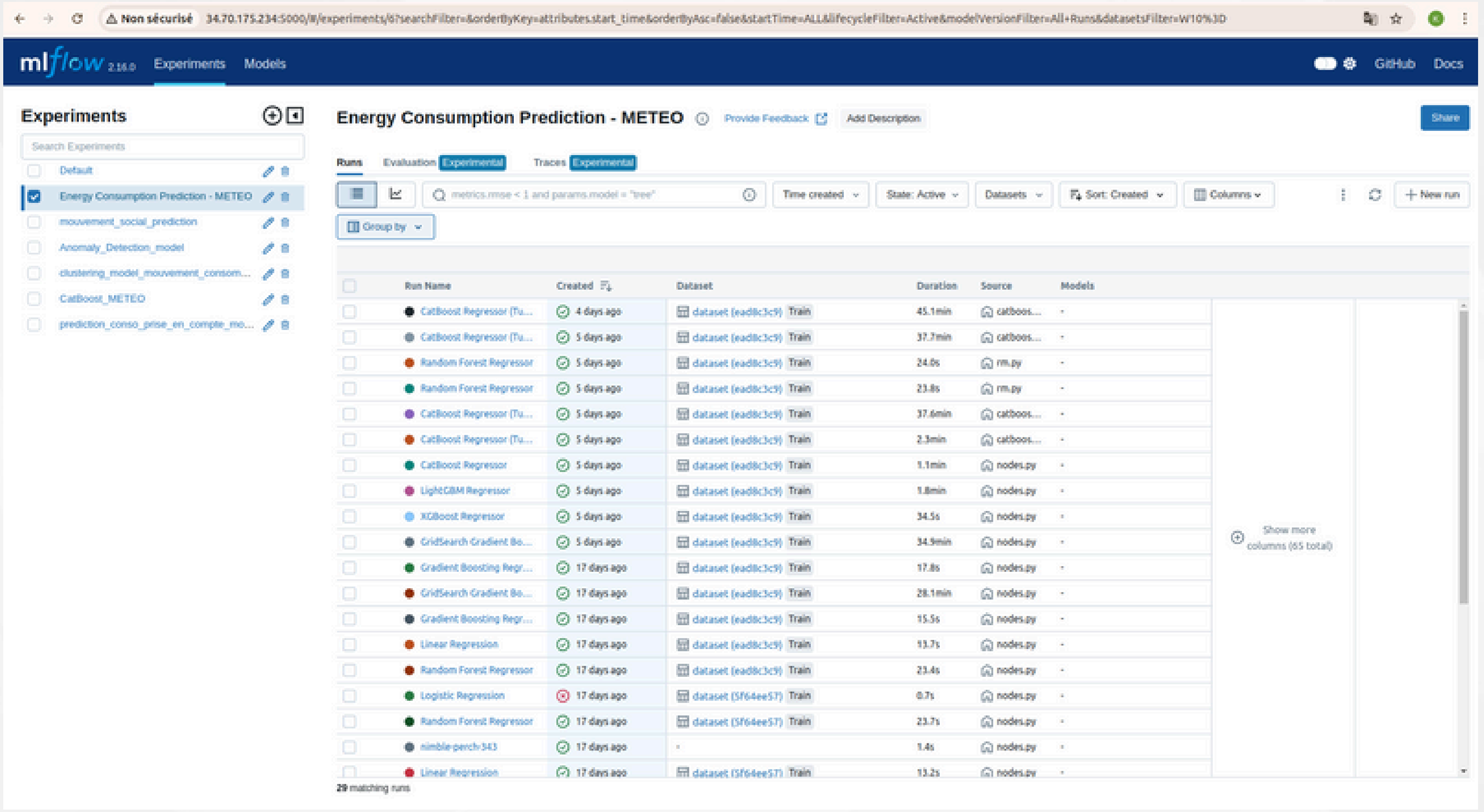
# NOTRE SELECTION

| Modèle | Algorithme | Objectif |
|---|---|---|
| Détection des Anomalies | Isolation Forest | Identifier les anomalies dans les habitudes de consommation |
| Clustering par Région | DBSCAN & KMeans | Regrouper les régions en fonction de profils de consommation |
| Prédiction de la Consommation avec Données de Mouvements | Random Forest | Prédire la consommation en fonction des mouvements humains |
| Prédiction de la Consommation avec Données Météorologiques | CatBoost | Prédire la consommation en fonction de la météo |

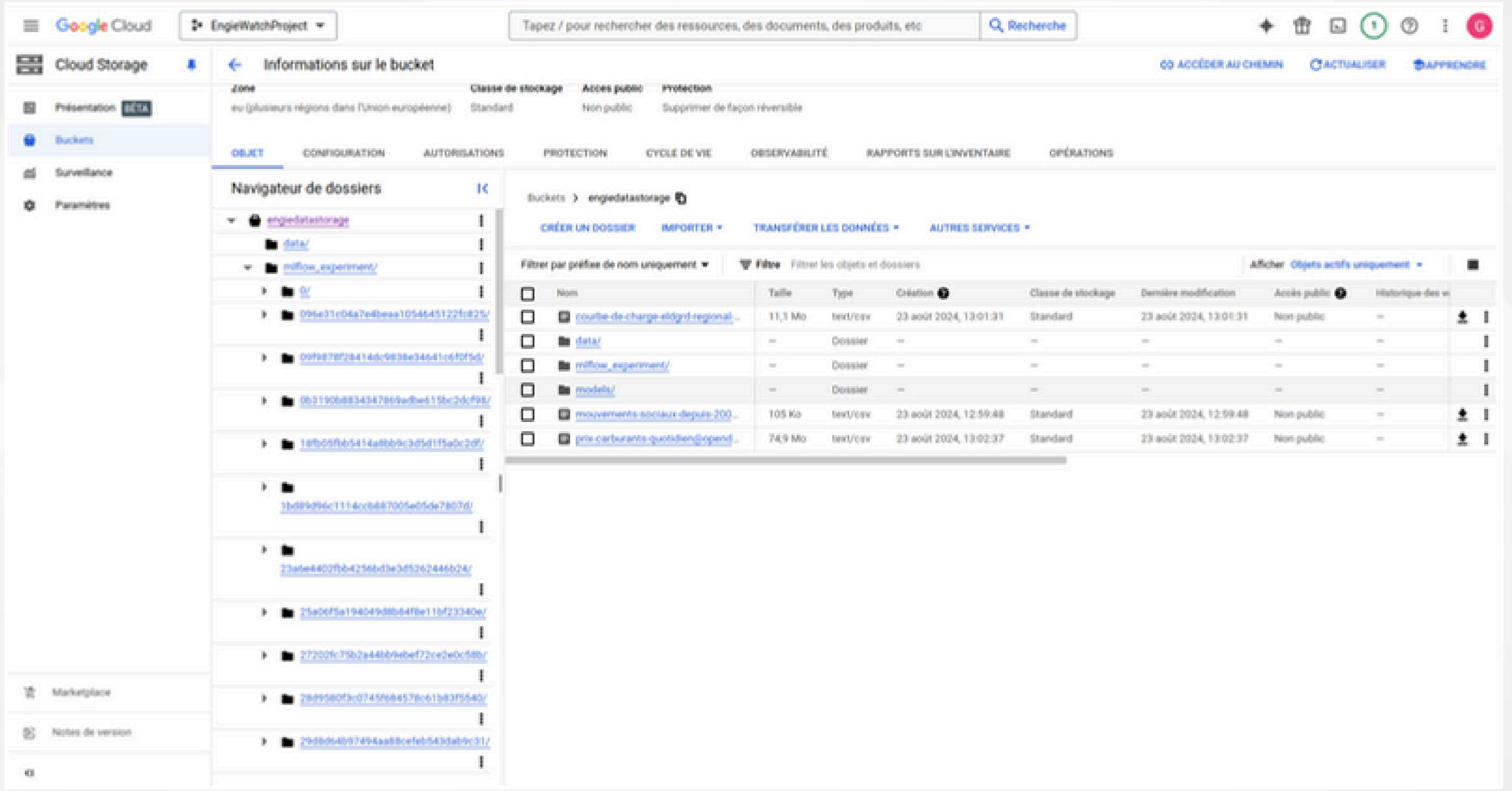
# DÉTECTION D'ANOMALIES


| | date | région | consommation_moyenne_journalière | statut | anomaly |
|---|---|---|---|---|---|
| 67 | 2018-12-26 00:00:00 | bourgogne-franche-comté | 3739.191901 | Définitif | 1 |
| 68 | 2018-12-29 00:00:00 | bourgogne-franche-comté | 3488.272348 | Définitif | 1 |
| 69 | 2018-12-30 00:00:00 | île-de-france | 11126.350450 | Définitif | 1 |
| 70 | 2019-01-02 00:00:00 | île-de-france | 30958.687564 | Définitif | 1 |
| 71 | 2019-01-03 00:00:00 | nouvelle-aquitaine | 2829.525728 | Définitif | 1 |
| 72 | 2019-01-05 00:00:00 | occitanie | 1214.573892 | Définitif | 1 |
| 73 | 2019-01-06 00:00:00 | provence-alpes-côte d'azur | 3505.729834 | Définitif | 1 |
| 74 | 2019-01-07 00:00:00 | île-de-france | 12547.795176 | Définitif | 1 |
| 75 | 2019-01-07 00:00:00 | hauts-de-france | 6963.340698 | Définitif | 1 |
| 76 | 2019-01-07 00:00:00 | provence-alpes-côte d'azur | 3739.080995 | Définitif | 1 |

# CYCLE DE VIE DES MODÈLES



# STOCKAGE ARTEFACTS MLFLOW



9.

# EMPREINTE CARBONE

# CALCUL
# EMPREINTE CARBONE

- **Globalement faible** : Empreinte carbone totale de 0.000593 kgCO2eq
- **Modèle le plus énergivore** : Modèle CatBoost.

```
scripts/empreinte_carbone_projet/emissions/emissions_catboosting.csv: Dernier run: 0.0004381058554062 kgCO2eq
scripts/empreinte_carbone_projet/emissions/emissions_data_fusion_pipeline.csv: Dernier run: 1.3437832202069784e-05 kgCO2eq
scripts/empreinte_carbone_projet/emissions/emissions_dbscan.csv: Dernier run: 4.407894442148034e-06 kgCO2eq
scripts/empreinte_carbone_projet/emissions/emissions_etl_pipeline.csv: Dernier run: 5.8006733084674365e-05 kgCO2eq
scripts/empreinte_carbone_projet/emissions/emissions_isolationforest.csv: Dernier run: 6.892161395051047e-06 kgCO2eq
scripts/empreinte_carbone_projet/emissions/emissions_random_forest.csv: Dernier run: 1.603075068442143e-05 kgCO2eq
scripts/empreinte_carbone_projet/emissions/emissions_ridge_model.csv: Dernier run: 2.3823826691643307e-06 kgCO2eq

Empreinte carbone totale du projet: 0.0005392636098837 29 kgCO2eq
```

10.

# INTEGRATION CI

# MAKEFILE

ci.yml | Makefile M

Makefile

```
MYPY_PATHS = data-collection-kedro/src/data_collection_kedro/pipelines/etl_pipeline \
             data-collection-kedro/src/data_collection_kedro/pipelines/data_fusion_pipeline \

PYLINT_PATHS = data-collection-kedro/src/data_collection_kedro/pipelines/etl_pipeline \
               data-collection-kedro/src/data_collection_kedro/pipelines/data_fusion_pipeline \
               dashboard_ui/

ISORT_PATHS = data-collection-kedro/src/data_collection_kedro/pipelines/etl_pipeline \
              data-collection-kedro/src/data_collection_kedro/pipelines/data_fusion_pipeline \
              dashboard_ui/

TEST_PATH = tests/

mypy:
    mypy $(MYPY_PATHS)
pylint:
    pylint $(PYLINT_PATHS)
isort:
    isort $(ISORT_PATHS)
black:
    black $(PYLINT_PATHS)
test:
    pytest $(TEST_PATH) --disable-warnings

# Commande pour exécuter tous les outils (mypy, pylint, isort, black, pytest)
check: mypy pylint isort black test

# Nettoyage des fichiers temporaires
clean:
    rm -rf __pycache__
    rm -rf .mypy_cache
    rm -rf .pytest_cache
```

# GITHUB ACTION

page_anomalie_detection.py | ci.yml M

.github > workflows > ci.yml

```yaml
name: Python Lint, Type Check, and Tests
on: [push, pull_request]
jobs:
  build:
    runs-on: ubuntu-latest
    steps:
    - uses: actions/checkout@v2
    - name: Set up Python
      uses: actions/setup-python@v2
      with:
        python-version: '3.11.5'
    - name: Install dependencies
      run: |
        python -m pip install --upgrade pip
        pip install -r requirements_test.txt
        pip install -r data-collection-kedro/requirements.txt
        pip install boto3 botocore
        python3 -m pip install types-requests
        pip install pytest pytest-cov  # Ajout de pytest et pytest-cov
        pip install pytest-mock
    - name: Check import formatting with isort
      run: isort .
    - name: Check code formatting with black
      run: black .
    - name: Lint with pylint
      run: pylint --rcfile=.pylintrc .
    - name: Run tests with pytest
      run: |
        pytest --disable-warnings --maxfail=5
    - name: Generate full coverage report
      run: |
        pytest --cov=. data-collection-kedro/tests/ --cov-report=xml
        pytest --cov=. data-collection-kedro/tests/ --cov-report=html
    - name: Upload coverage report artifact
      uses: actions/upload-artifact@v3
      with:
        name: coverage-report
        path: htmlcov/
```

11.

# INTERFACE UTILISATEUR

# IHM


Prédiction Météo
Anomalie Détection
Prédiction Consommation
Clustering
Feedback
Tracking
Prédiction d'énergie consommée
Dans cette section, nous faisons la prédiction en prenant en compte des conditions météorologiques.
Température Max (°C)
25,00
Température Min (°C)
15,00
Vitesse du vent (km/h)
10
Humidité (%)
50
Visibilité (km)
10
Couverture Nuageuse (%)
50
0
100
Prédire la Consommation Énergétique

12.

# DIFFICULTES RENCONTRÉES ET EVOLUTIONS

# DIFFICULTES RENCONTREES

- Qualité du dataset
- Technologies diverses et variées à prendre en main
- Gestion des Données Massives
- Contraintes Budgétaires
- Données sur l’Empreinte Carbone

# EVOLUTIONS

**Prédiction des risques d’incidents**

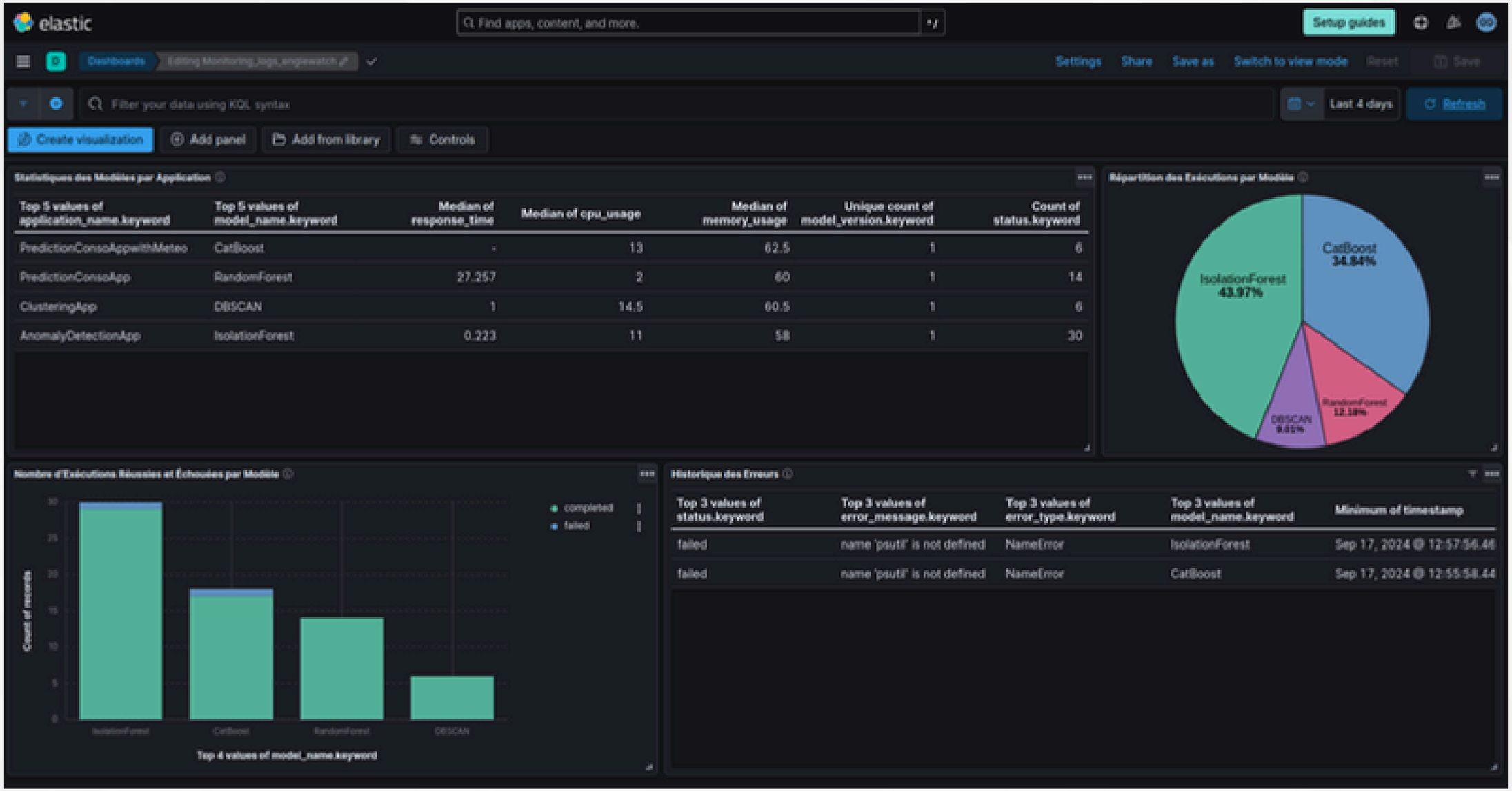
**Anticipation des Impacts**

**Optimisation de la Planification**

13.

# MAITENANCE DE L’APPLICATION

# MONITORING


elastic
Find apps, content, and more.
Setup guides
Dashboards
Editing Monitoring_logs_englewatch
Settings
Share
Save as
Switch to view mode
Reset
Save
Filter your data using KQL syntax
Last 4 days
Refresh
Create visualization
Add panel
Add from library
Controls
Statistiques des Modèles par Application
Top 5 values of application_name.keyword
Top 5 values of model_name.keyword
Median of response_time
Median of cpu_usage
Median of memory_usage
Unique count of model_version.keyword
Count of status.keyword
PredictionConsoAppwithMeteo CatBoost - 13 62.5 1 6
PredictionConsoApp RandomForest 27.257 2 60 1 14
ClusteringApp DBSCAN 1 14.5 60.5 1 6
AnomalyDetectionApp IsolationForest 0.223 11 58 1 30
Répartition des Exécutions par Modèle
CatBoost 34.84%
IsolationForest 43.97%
RandomForest 12.18%
DBSCAN 9.01%
Nombre d'Exécutions Réussies et Échouées par Modèle
completed
failed
Count of records
IsolationForest
CatBoost
RandomForest
DBSCAN
Top 4 values of model_name.keyword
Historique des Erreurs
Top 3 values of status.keyword
Top 3 values of error_message.keyword
Top 3 values of error_type.keyword
Top 3 values of model_name.keyword
Minimum of timestamp
failed name 'psutil' is not defined NameError IsolationForest Sep 17, 2024 @ 12:57:56.46
failed name 'psutil' is not defined NameError CatBoost Sep 17, 2024 @ 12:55:58.44

14.

# DEMO

# MERCI