Projet ETL Dagster – S7 ISIBD ENSAB

Réalisé par : Hamza Elmourabit

Encadré par : Pr .Lamia Karim

15 décembre 2025

Table des matières

1 Présentation du projet

Ce projet consiste à construire un pipeline ETL (Extract, Transform, Load) avec **Dagster**. L'objectif est de lire un fichier CSV (input.csv), effectuer des transformations et sauvegarder le résultat dans output.csv. Le job peut être exécuté manuellement ou via un schedule automatique.

2 Structure du projet

```
dagster_etl_scheduled/
 .venv/                  # Environnement virtuel
 assets/
    etl_assets.py       # Code du pipeline
 workspace.yaml         # Config Dagster
 input.csv              # Données sources
 output.csv             #  Résultat transformé
```

3 Code Python (assets/etl_assets.py)

```python
import pandas as pd
from dagster import asset, job, materialize

# Extract : charger les donn es depuis input.csv
@asset
def load_data():
    df = pd.read_csv('input.csv')
    print('Data loaded:\n', df.head())
    return df

# Transform : ajouter colonne 'total'
@asset
def transform_data(load_data):
    df = load_data.copy()
    df['total'] = df['quantity'] * df['price']
    print('Data transformed:\n', df.head())
    return df

# Load : sauvegarder les donn es  transform es  dans  output.csv
@asset
def save_data(transform_data):
    df = transform_data
    df.to_csv('output.csv', index=False)
    print('Data saved to output.csv')
    return 'Success'
```

4 Configuration (workspace.yaml)

```
load_from :
  - python_file :
      relative_path :  assets/etl_assets.py
```

5 Instructions d'exécution

1. Activer l'environnement virtuel :

```
& .\.venv\Scripts\Activate.ps1
```

2. Définir DAGSTER_HOME :

```
$env:DAGSTER_HOME = "$HOME\dagster_home"
```

3. Lancer le Webserver Dagster :

```
dagster-webserver -w workspace.yaml -p 3001
```

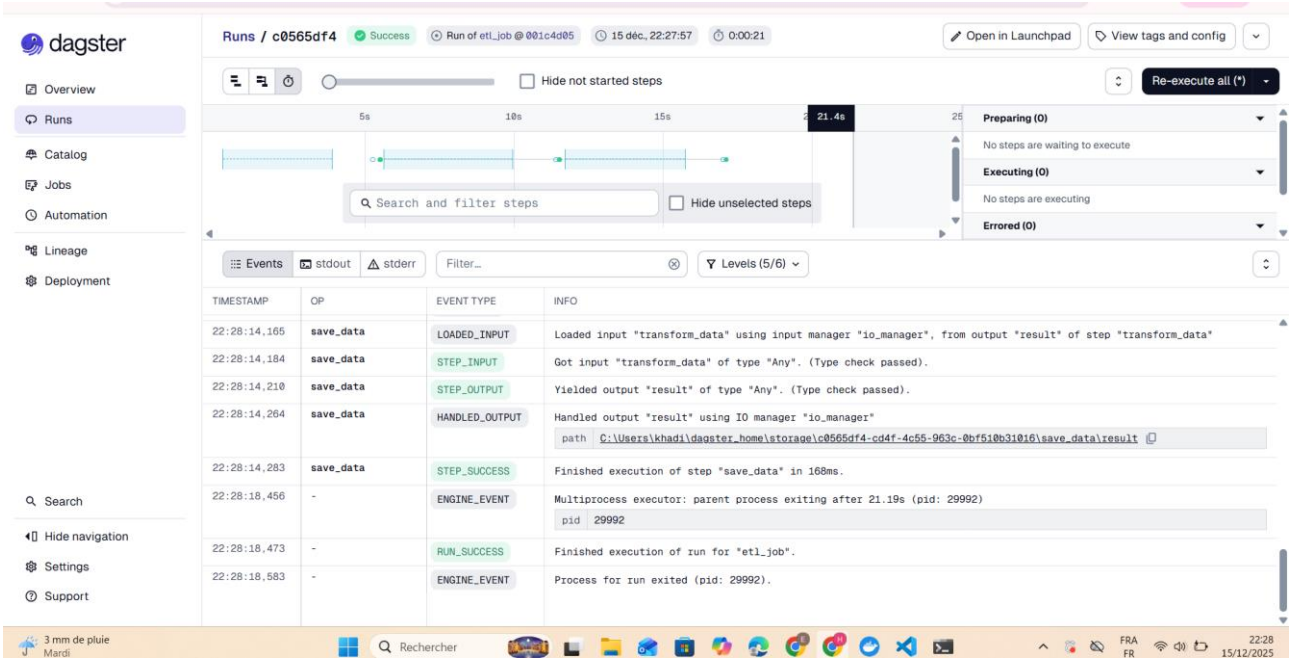
4. Lancer le daemon pour le schedule :

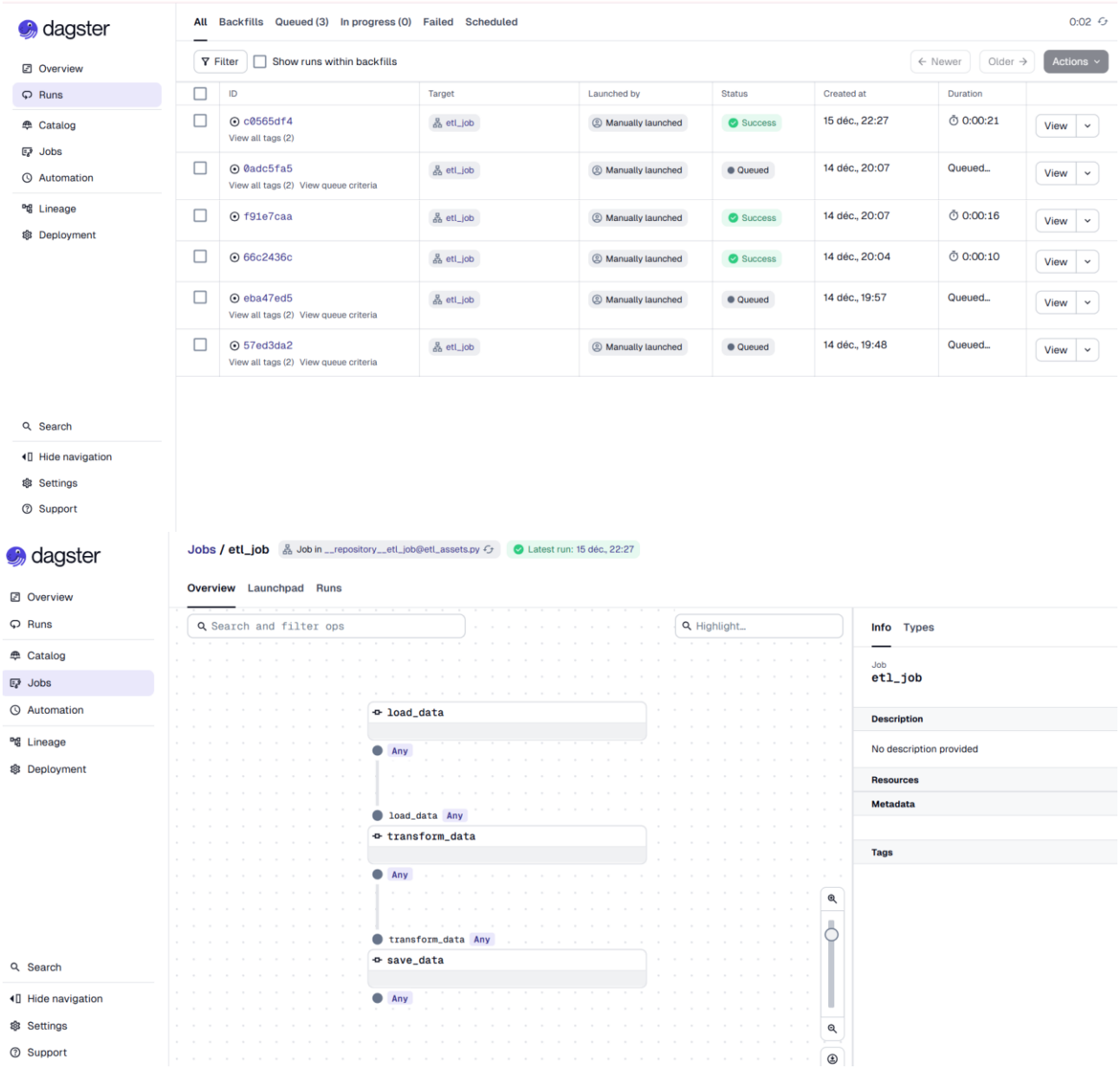
```
dagster-daemon run -w workspace.yaml
```

5. Vérifier le fichier output.csv après exécution.

6 Résultats

Après exécution du pipeline, le fichier output.csv contient les données transformées avec la nouvelle colonne total.




dagster
Overview
Runs
Catalog
Jobs
Automation
Lineage
Deployment
Search
Hide navigation
Settings
Support
All Backfills Queued (3) In progress (0) Failed Scheduled
0:02
Filter
Show runs within backfills
Newer
Older
Actions
ID
Target
Launched by
Status
Created at
Duration
c0565df4
View all tags (2)
etl_job
Manually launched
Success
15 déc., 22:27
0:00:21
View
0adc5fa5
View all tags (2) View queue criteria
etl_job
Manually launched
Queued
14 déc., 20:07
Queued...
View
f91e7caa
etl_job
Manually launched
Success
14 déc., 20:07
0:00:16
View
66c2436c
etl_job
Manually launched
Success
14 déc., 20:04
0:00:10
View
eba47ed5
View all tags (2) View queue criteria
etl_job
Manually launched
Queued
14 déc., 19:57
Queued...
View
57ed3da2
View all tags (2) View queue criteria
etl_job
Manually launched
Queued
14 déc., 19:48
Queued...
View

dagster
Overview
Runs
Catalog
Jobs
Automation
Lineage
Deployment
Search
Hide navigation
Settings
Support
Jobs / etl_job
Job in __repository__etl_job@etl_assets.py
Latest run: 15 déc., 22:27
Overview
Launchpad
Runs
Search and filter ops
Highlight...
load_data
Any
load_data Any
transform_data
Any
transform_data Any
save_data
Any
Info
Types
Job
etl_job
Description
No description provided
Resources
Metadata
Tags