Data Pipeline Project - Full History

This document summarizes the entire development history of the automated data pipeline project using Dagster, dbt, DuckDB, and Streamlit. It includes configuration, setup, debugging, and final working implementations.

# 1. Project Initialization

• Created environment and installed Dagster, dbt, DuckDB, Streamlit.

• Configured API extraction from Météo Concept.

# 2. Code Files

• main\_pipeline.py: Contains Dagster job with fetch/store/run\_dbt.

• dashboard.py: Interactive Streamlit dashboard to visualize weather data.

• assets.py: Dagster assets for raw\_weather\_data, stored\_weather\_data, run\_dbt\_asset.

• repository.py: Dagster Definitions with assets and schedule.

• test\_fetch\_weather.py: Pytest tests for asset logic.

# 3. Pipeline Execution and Debugging

• Faced multiple Unicode, subprocess, and dependency issues.

• Fixed subprocess paths, encoding issues, scheduling conflicts.

• Final Dagster pipeline executes with assets and dbt integration.

# 4. Final Test

• Successfully ran Pytest tests for raw and stored data.

• Confirmed successful dbt transformations and visual output.