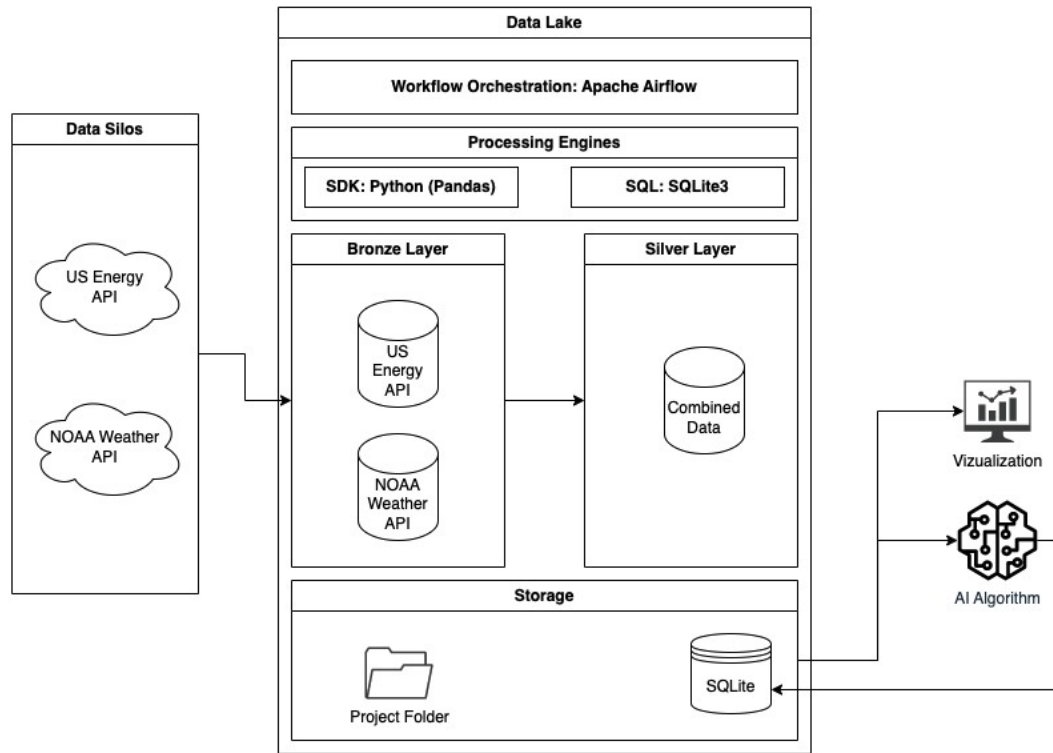


Name: Onassis Nottage
Course: Big Data Infrastructure

For explanation purposes, see README.md file in deliverable, and Flask App after running.

ARCHITECTURE:



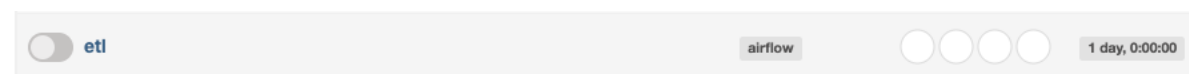
AIRFLOW ORCHESTRATION:

```
Scripts > ETL.py > ...
1  from airflow import DAG
2  from airflow.operators.python_operator import PythonOperator
3  from datetime import datetime, timedelta
4  import sys
5  sys.path.insert(0, '/Users/oanottage/Desktop/BTS/BDI/Final_Project/Scripts')
6  import bronze_layer
7  import silver_layer
8
9
10 # Default arguments for the DAG starting April 8th, 2023 at 5:00 AM
11 default_args = {
12     'owner': 'airflow',
13     'depends_on_past': False,
14     'start_date': datetime(2023, 4, 8, 5, 0),
15     'retries': 1,
16     'retry_delay': timedelta(minutes=5)
17 }
18
19 # Create ETL DAG
20 etl_dag = DAG(
21     'etl',
22     default_args=default_args,
23     description='ETL API data into bronze and silver layers',
24     schedule_interval=timedelta(days=1)
25 )
26
27 #Task to Collect Energy and Weather API data
28 bronze_task = PythonOperator(
29     task_id='bronze',
30     python_callable=bronze_layer.bronze,
31     dag=etl_dag)
32
33
34 #Task to Combine data into one table and store to database
35 silver_task = PythonOperator(
36     task_id='silver',
37     python_callable=silver_layer.silver,
38     dag=etl_dag
39 )
40
41 bronze_task >> silver_task
```

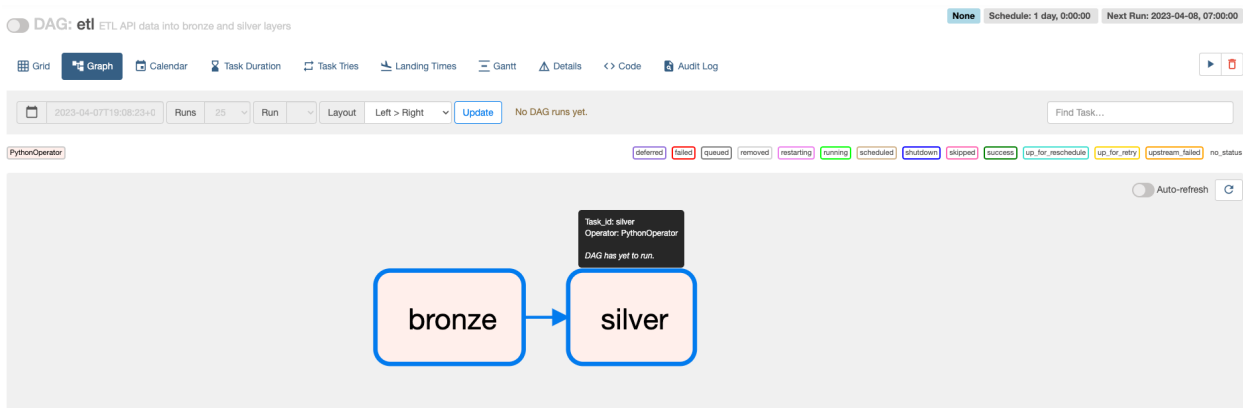
PROBLEMS 3 OUTPUT DEBUG CONSOLE TERMINAL JUPYTER

```
py:49 MovedIn20Warning: [31mDeprecated API features detected! These feature(s) are not compatible with SQLAlchemy 2.0. [32mTo prevent incompatible upgrades prior to updating applications, ensure requirements files are pinned to "sqlalchemy<2.0". [36mSet environment variable SQLALCHEMY_WARN_20=1 to show all deprecation warnings. Set environment variable SQLALCHEMY_SILENCE_UBER_WARNING=1 to silence this message.[0m (Background on SQLAlchemy 2.0 at: https://sqlalche.me/e/b8d9)
/Users/oanottage/opt/anaconda3/envs/energyvenv/lib/python3.9/site-packages/airflow/models/base.py:49 MovedIn20Warning: [31mDeprecated API features detected! These feature(s) are not compatible with SQLAlchemy 2.0. [32mTo prevent incompatible upgrades prior to updating applications, ensure requirements files are pinned to "sqlalchemy<2.0". [36mSet environment variable SQLALCHEMY_WARN_20=1 to show all deprecation warnings. Set environment variable SQLALCHEMY_SILENCE_UBER_WARNING=1 to silence this message.[0m (Background on SQLAlchemy 2.0 at: https://sqlalche.me/e/b8d9)
```

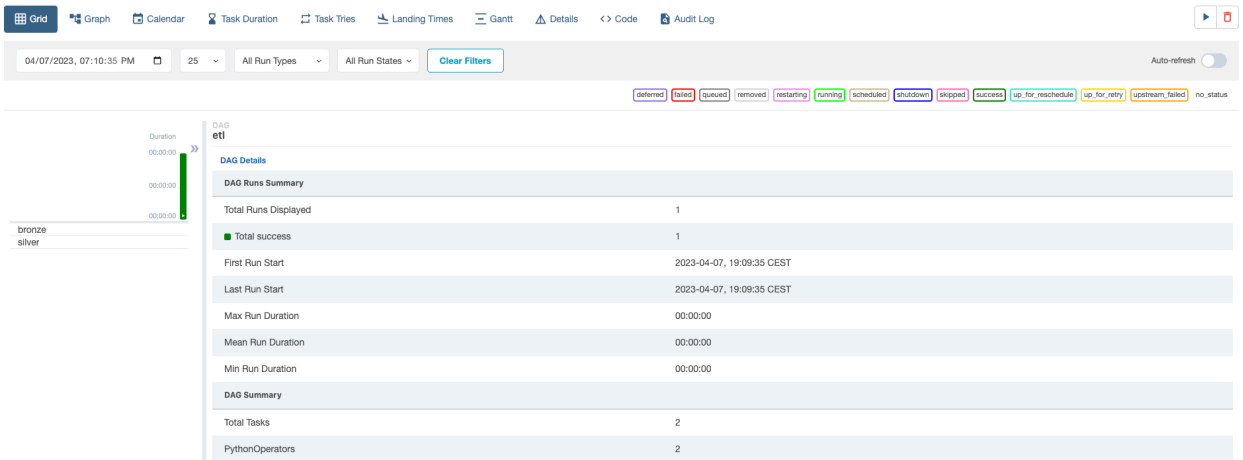
Airflow file



etl DAG



DAG Graph



Successful Run

WEBSITE:

← → ↺ 127.0.0.1:5000

Home About AI Prediction

Energy Consumption Prediction - By Onassis Nottage

Course: Big Data Infrastructure

The goal of this project is to predict energy consumption based on various features such as temperature, humidity, time of day, etc. The two sources of data for this project could be energy consumption data and weather data. You can get the energy consumption data from the US Energy Information Administration (<https://www.eia.gov/opa/data/qb.php?category=2122628>) and the weather data from NOAA (<https://www.ncdc.noaa.gov/cdo-web/webservices/v2>).

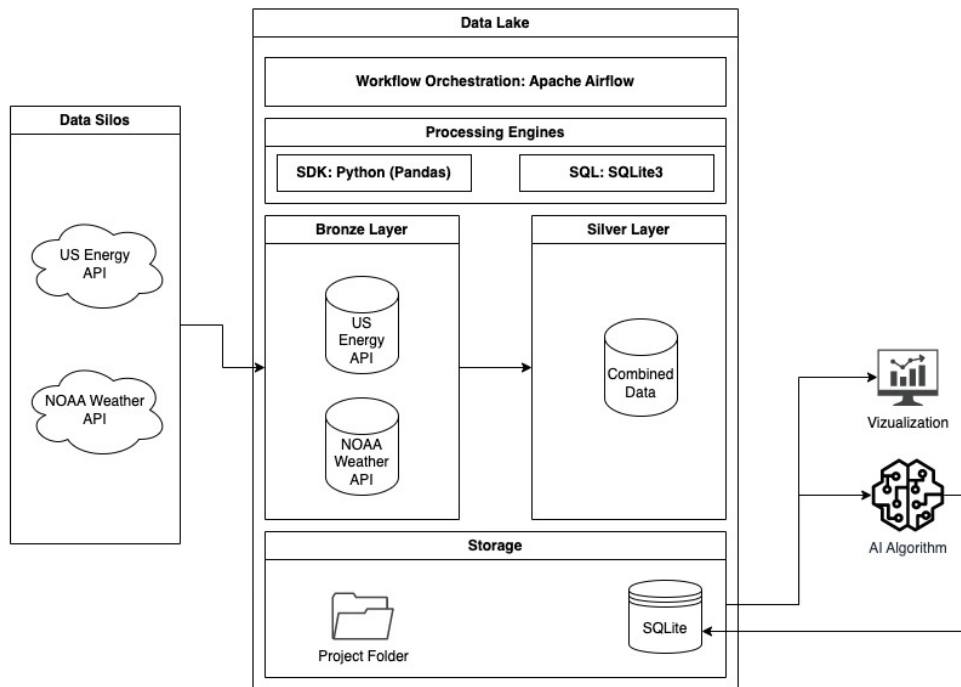
```
graph LR
    subgraph DataStos [Data Stos]
        A[US Energy API]
        B[NOAA Weather API]
    end
    subgraph DataLake [Data Lake]
        subgraph WorkflowOrchestration [Workflow Orchestration: Apache Airflow]
            C[SDK: Python (Pandas)]
            D[SQL: SQLite3]
        end
        subgraph ProcessingEngines [Processing Engines]
            E[Bronze Layer]
            F[Silver Layer]
        end
        subgraph Storage [Storage]
            G[Project Folder]
            H[SQLite]
        end
        E --> F
        F --> H
    end
    A --> E
    B --> E
    H --> I[AI Algorithm]
    I --> J[Visualization]
```

Dashboard Features

- Energy Demand Forecasts Over Time
- Energy Demand Forecasts by Region and Weather Conditions

Main Page

two sources of data for this project could be energy consumption data and weather data. You can get the energy consumption data from the US Energy Information Administration (<https://www.eia.gov/opa/data/qb.php?category=2122628>) and the weather data from NOAA (<https://www.ncdc.noaa.gov/cdo-web/webservices/v2>).



Dashboard Features

- Energy Demand Forecasts Over Time
- Energy Demand Forecasts by Region and Weather Conditions
- Distribution of Energy Demand Forecasts
- User input form for predicting energy consumption

I will be using Python and popular data science libraries such as Pandas, Numpy, and Scikit-learn to build the prediction model and Flask to create the web application.

INITIALIZE DOCUMENT

VIEW STORY

Initialize document button (to create graphs and charts for the story)

Document Successfully Initialized!

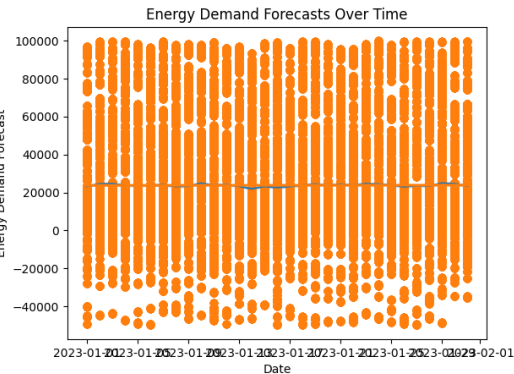
This document has been successfully initialized. You can now view the story.

VIEW STORY

Resulting Page after Initialization

Energy Demand Forecasts Over Time

This line chart shows how energy demand forecasts have changed over time. The x-axis represents date and the y-axis represents energy demand forecast.



The trend line on this chart shows that energy demand forecasts have been steadily increasing over time. However, there are some fluctuations in the data, suggesting that there are other factors influencing energy demand as well.

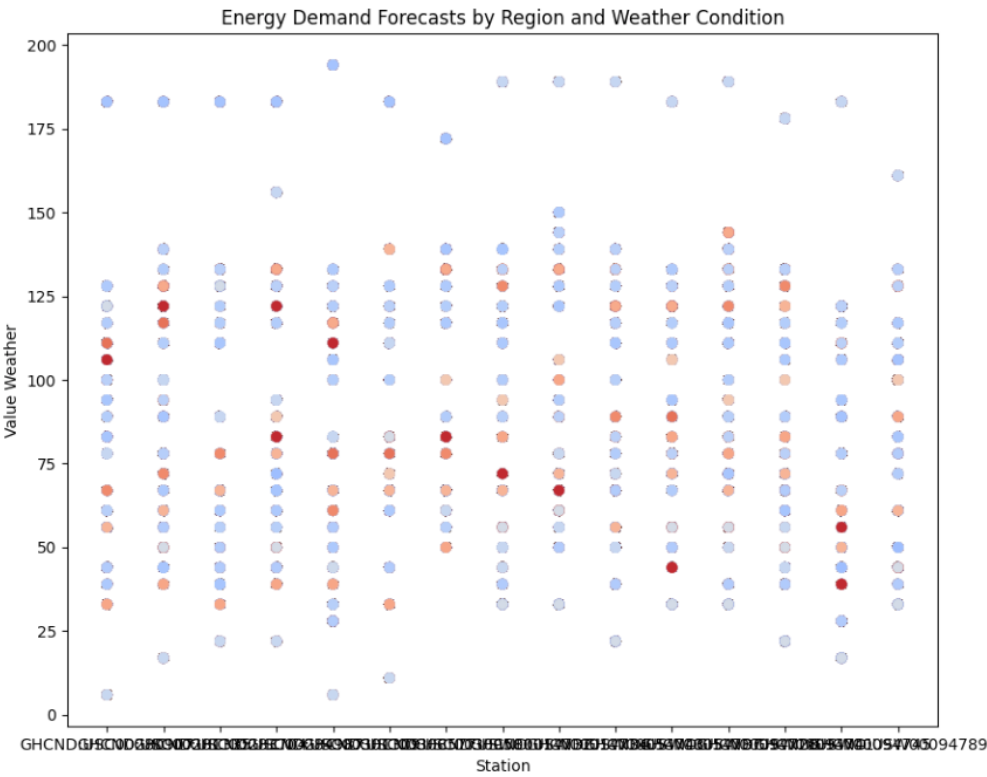
By using filters, we could explore the trends by respondent, type, or region to gain more insight into what factors are driving changes in energy demand over time.

PREVIOUS NEXT

Story 1

Energy Demand Forecasts by Region and Weather Conditions

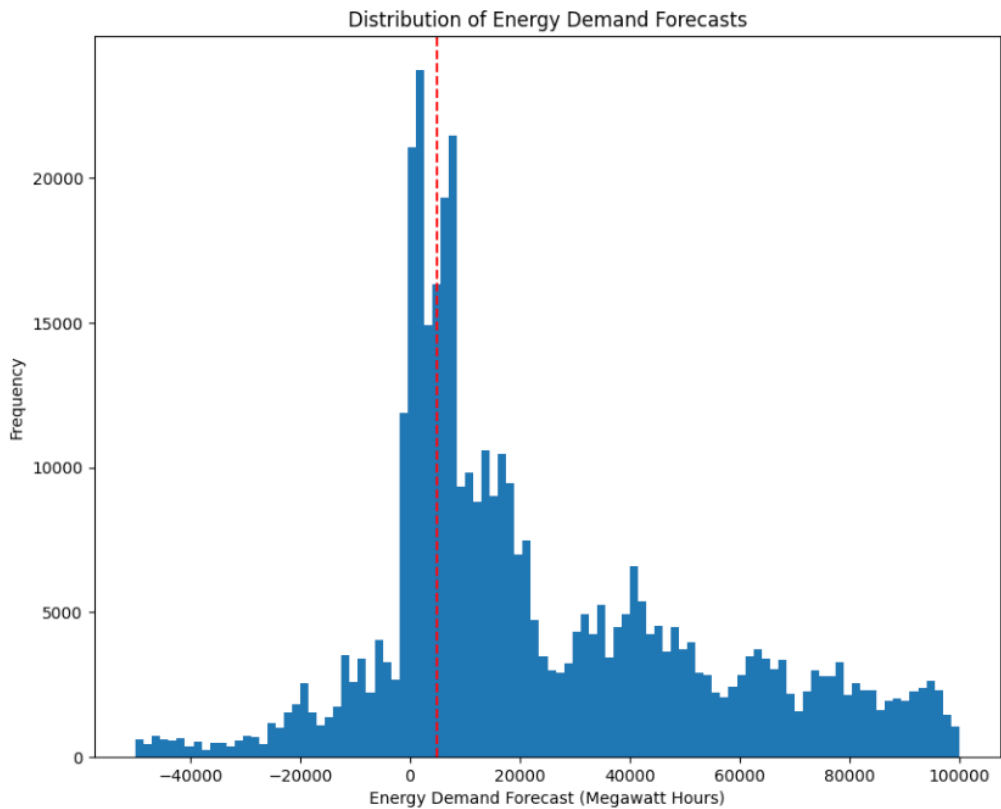
This heat map or scatterplot displays how energy demand forecasts vary by region and/or by weather conditions. The x-axis represents station and/or value_weather, while the y-axis represents forecasted energy demand. Color or shape is used to represent different levels of forecasted energy demand.



Story 2

Distribution of Energy Demand Forecasts

This histogram or frequency chart displays the distribution of energy demand forecasts, showing how they are spread across different values of energy demand. The x-axis represents value_energy, while the y-axis represents frequency (count). Binning is used to group the data into intervals, and any outliers or unusual patterns are highlighted using annotations or a reference line.



Story 3

Energy Consumption Prediction

This dashboard shows the energy consumption prediction for the next 24 hours.

Type name:

- ☐ Net generation
- ☐ Total interchange
- ☐ Demand
- ☐ Day-ahead demand forecast

Timezone:

Pacific

Station:

Attributes:

- ☐ „7,0630
- ☐ „7,0700
- ☐ „7,0800
- ☐ „W,
- ☐ „W,2400
- ☐ „7,

Value weather:

Predict

Energy Consumption Prediction Page