

Департамент образования и науки города Москвы
Государственное автономное образовательное учреждение высшего
образования города Москвы
«Московский городской педагогический университет»
Институт цифрового образования
Департамент информатики, управления и технологий

ДИСЦИПЛИНА:

Проектный практикум по разработке ETL-решений

Вебинар 21-03-2025

Практическая работа на вебинаре

Выполнил: Егоров А.С., группа: АДЭУ-211

Преподаватель: Босенко Т.М.

Москва

2025

Задания:

Вариант 4

1. Получить прогноз в Берлине на 3 дня
2. Рассчитать среднюю температуру
3. Вывести дату с самой высокой температурой

Сначала необходимо построить верхнеуровневую архитектуру аналитического решения (рисунок 1).

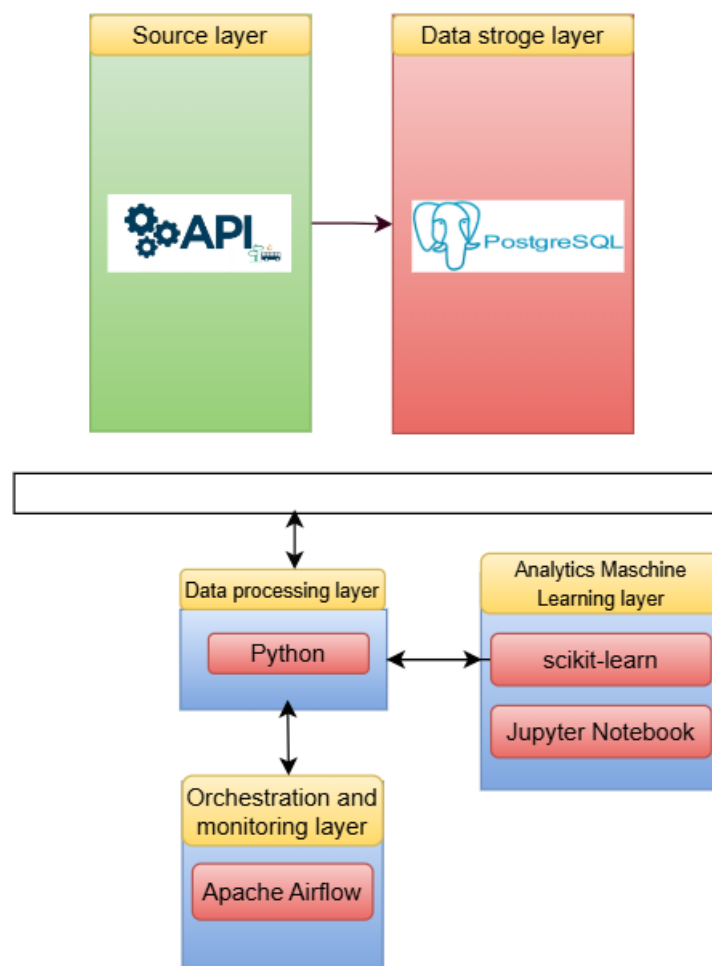


Рисунок 1. Верхнеуровневая архитектура аналитического решения

Затем согласно варианту надо изменить, чтобы выводил прогноз Берлина на 3 дня. И далее запустить контейнер (рисунок 2).

```
7 weatherapi.com/v1/forecast.json?key={api_key}&q=Berlin&days=3"
8 get(url)
9 ()
0 y['date'], day['day']]['avgtemp_c']) for day in data['forecast']
1 recast_data, columns=['date', 'temperature'])
2 flow/data'
```

Terminal output:

```
ella_25_default R
] Running 5/business_case_umbrella_25-init-1 R
Container business_case_umbrella_25-webserver-1 Removed7.9s r business_case_
brella_25-postgres-1 R
Container business_case_umbrella_25-scheduler-1 Removed4.7s business_case_um
ella_25_default R
Container business_case_umbrella_25-init-1 Removed0.0s
Container business_case_umbrella_25-postgres-1 Removed0.4s
Network business_case_umbrella_25_default Removed0.3s
pu@mgpu-VirtualBox:~/workshop-on-ETL/business_case_umbrella_25$
```

Рисунок 2. Изменение на Берлин, прогноз 3 дня

На рисунке 3 видно, что dag сработал



Рисунок 3. dag сработал

На рисунке 4 показан блокнот с выполнением индивидуального задания, где применяется модель машинного обучения, вычисляется средняя температура и выводит дату с самой высокой температурой

```
[68] from google.colab import files
      uploaded = files.upload()

      Выберите файлы Число файлов: 2
      • joined_data.csv(text/csv) - 79 bytes, last modified: 21.03.2025 - 100% done
      • ml_model.pkl(n/a) - 926 bytes, last modified: 21.03.2025 - 100% done
      Saving joined_data.csv to joined_data (4).csv
      Saving ml_model.pkl to ml_model (3).pkl

[22] !pip install dill

      Requirement already satisfied: dill in /usr/local/lib/python3.11/dist-packages (0.3.9)

[73] import joblib
      model = joblib.load("ml_model.pkl")
      import pandas as pd
      print(model.predict(pd.DataFrame({'temperature': [5, 10, 15, 20]})))

      [100. 100. 100. 100.]
      /usr/local/lib/python3.11/dist-packages/sklearn/base.py:380: InconsistentVersionWarning: Trying to unpickle estimator Linear
      https://scikit-learn.org/stable/model\_persistence.html#security-maintainability-limitations
      warnings.warn(

[74] df = pd.read_csv('joined_data.csv')

[75] df

      date temperature sales
0  2025-03-21         13.2   10
1  2025-03-22         11.9   15
2  2025-03-23         11.9   20

      Далее: Посмотреть рекомендованные графики New interactive sheet

[76] average_temperature = df['temperature'].mean()
      print(f"Средняя температура: {average_temperature:.2f} °C")

      Средняя температура: 12.33 °C

      max_temp_row = df.loc[df['temperature'].idxmax()]
      print(f"Дата с самой высокой температурой: {max_temp_row['date']} ({max_temp_row['temperature']} °C)")

      Дата с самой высокой температурой: 2025-03-21 (13.2 °C)
```

Рисунок 4. Выполнение индивидуального задания

Выводы:

1. Получен прогноз в Берлине на 3 дня
2. Рассчитана средняя температура
3. Выведена дата с самой высокой температурой