

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**

Выполнила: Шведова С.С., группа: АДЭУ-211

Преподаватель: Босенко Т.М.

Москва

2025

## Задания

- 1.1. Развернуть VM ubuntu\_mgpu.ova в VirtualBox.
  - 1.2. Клонировать на ПК задание бизнес-кейс Umbrella в домашний каталог VM.
  - 1.3. Запустить контейнер 01\_umbrella.py с кейсом, изучить и описать основные элементы интерфейса Apache Airflow.
  - 1.4. Спроектировать верхнеуровневую архитектуру для real\_umbrella.py
- 13 вариант.** Получить прогноз в Праге на 3 дня. Добавить столбец: разница с предыдущим днём. Построить таблицу изменений температуры.

На рисунке 1 показана архитектура для real\_umbrella.py

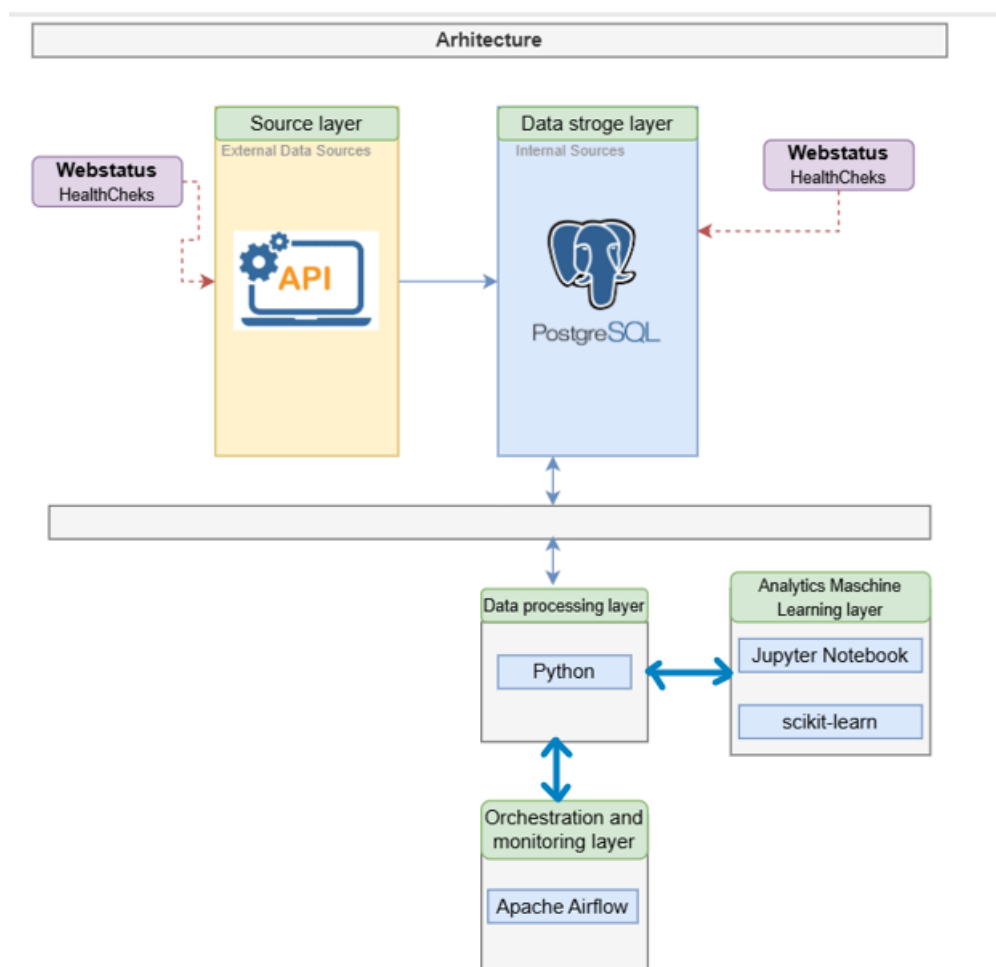


Рисунок 1. Архитектура

## Основные элементы интерфейса Apache Airflow:

1. Веб-сервер: это основной компонент, который предоставляет веб-интерфейс для использования Airflow. Он позволяет пользователям управлять, мониторить и планировать рабочие процессы (DAGs). Веб-сервер можно запустить с помощью команды `airflow webserver`.

2. Панель управления (UI): панель управления предоставляет различные функции, включая:

- Списки DAG: отображает все доступные DAGs с информацией о их состоянии.
- График выполнения (Graph View): визуализирует зависимость между задачами (операми) в виде графа.
- Дерево задач (Tree View): Показано состояние выполнения каждой задачи в рамках DAG.
- Измерение времени выполнения задач (Task Duration): график времени выполнения задач и DAG.
- Журнал выполнения задач (Logs): позволяет просматривать логи выполнения для каждой задачи

3. Проблемы и отделы (XComs): Эти механизмы позволяют передавать данные между задачами в DAG.

4. Пользовательские настройки: Веб-интерфейс включает разделы для управления переменными, подключениями к внешним сервисам, а также плагинами, что дает возможность интегрировать Airflow с другими сервисами и APIs.

5. Мониторинг и уведомления: Airflow поддерживает мониторинг состояния выполнения задач и отправку уведомлений при возникновении ошибок или при завершении задач

На рисунке 2 показано изменение `real_umbrella.py`, а именно api ключа и предсказание именно Праги на 3 дня.

```
real_umbrella.py x joined_data.csv
dags > real_umbrella.py
17
22
23
24 а погоды
25 cast():
26 fc9a9486d87683630252103" # замените на ваш API ключ
27 i.weatherapi.com/v1/forecast.json?key={api_key}&q=Praga&days=3
28 s.get(url)
29 on()
30 day['date'], day['day']['avgtemp_c'] for day in data['forecast
31 forecast data.columns=['date', 'temperature']

er/css/ab.css HTTP/1.1" 304 0 "http://localhost:8080/dags/real_umbrella_containe
rized/grid?tab=graph&dag_run_id=manual_2025-03-21T12%3A11%3A05.566394%2B00%3A00
" Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:123.0) Gecko/20100101 Firefox/123.0"
webserver-1 | 172.20.0.1 - - [21/Mar/2025:12:16:36 +0000] "GET /static/dist/mat
erialicons.ce0f77d10d4dc51f5f07.css HTTP/1.1" 304 0 "http://localhost:8080/dags/
real_umbrella_containerized/grid?tab=graph&dag_run_id=manual_2025-03-21T12%3A11
%3A05.566394%2B00%3A00" Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:123.0) Gecko
/20100101 Firefox/123.0"
webserver-1 | 172.20.0.1 - - [21/Mar/2025:12:16:37 +0000] "GET /static/dist/mai
n.0645e1a08ff7a660a6ff7.css HTTP/1.1" 304 0 "http://localhost:8080/dags/real_umb
```

Рисунок 2. Изменение файла real\_umbrella.py

На рисунке 3 видно, что даг отработал успешно

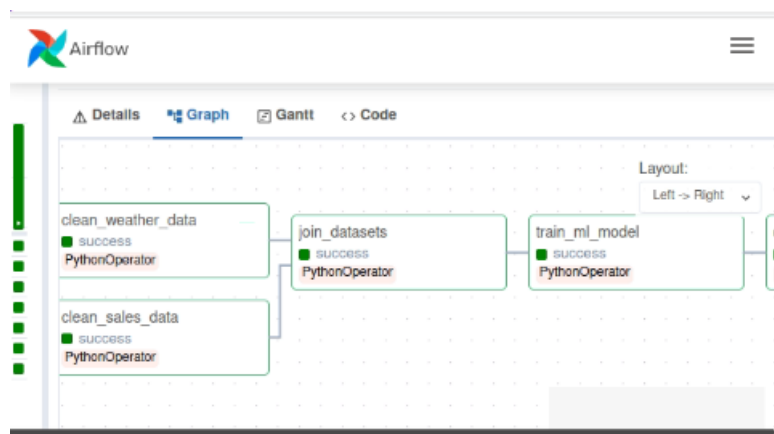


Рисунок 3. Даг успешно отработал

На рисунке 4 показано выполнение индивидуального задания, а именно загрузка модели машинного обучения, просмотр таблицы с температурой и продажами, а также представлена разница между предыдущим днем.

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

[12] !pip install dill

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

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

df = pd.read_csv('joined_data.csv')
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](#)

```
[20] df['temp_difference'] = df['temperature'].diff()
      df['sales_difference'] = df['sales'].diff()
      df
```

	date	temperature	sales	temp_difference	sales_difference
0	2025-03-21	13.2	10	NaN	NaN
1	2025-03-22	11.9	15	-1.3	5.0
2	2025-03-23	11.9	20	0.0	5.0

Рисунок 4. Выполнение индивидуального задания

## Выводы

1. Спроектирована верхнеуровневая архитектура
2. Получен прогноз в Праге на 3 дня.
3. Добавлен столбец: разница с предыдущим днём
4. Построена таблица изменений температуры.