

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

Лабораторная работа 6.1

Разработка полного ETL-процесса. Оркестровка конвейера данных

Выполнила: Сачкова Г.Г., группа: АДЭУ-211

Преподаватель: Босенко Т.М.

Москва

2025

Задачи

- верхнеуровневая архитектура задания Бизнес-кейса «StockSense», выполненная в draw.io;
- архитектура DAG Бизнес-кейса «StockSense», выполненная в draw.io;
- диаграмма Ганта DAG в Apache Airflow;
- ERD-схема базы данных Postgre SQL;
- SQL-запросы, позволяющие проверить наличие выгруженных агрегированных данных бизнес-задачи.

Сначала надо нарисовать верхнеуровневую архитектуру аналитического решения (рисунок 1)

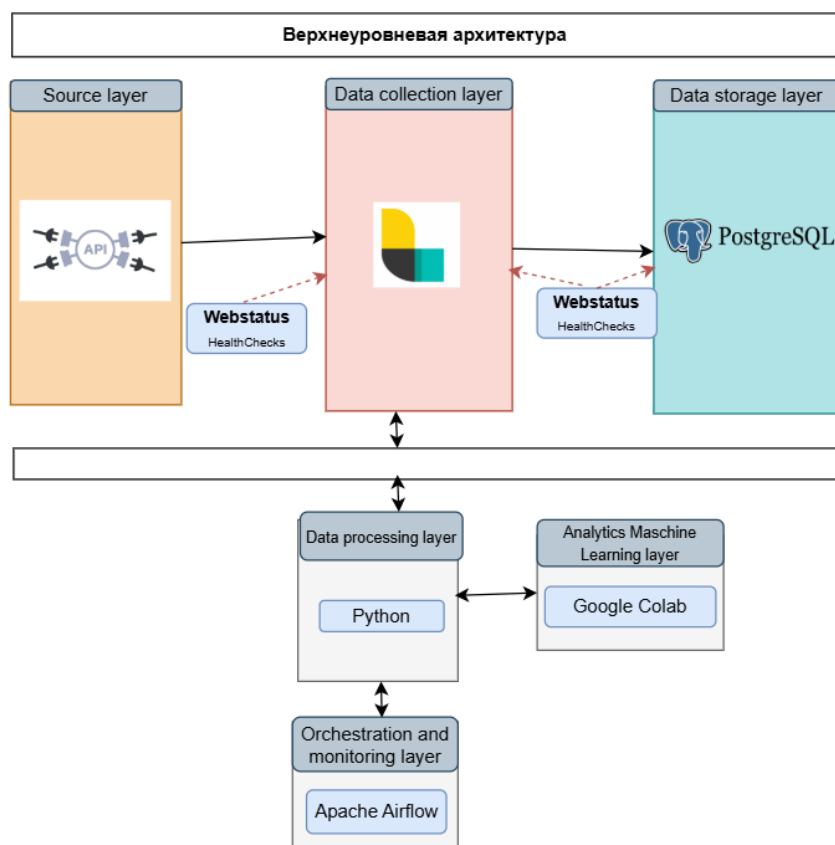


Рисунок 1 – Верхнеуровневая архитектура аналитического решения

На рисунке 2 показана архитектура дага

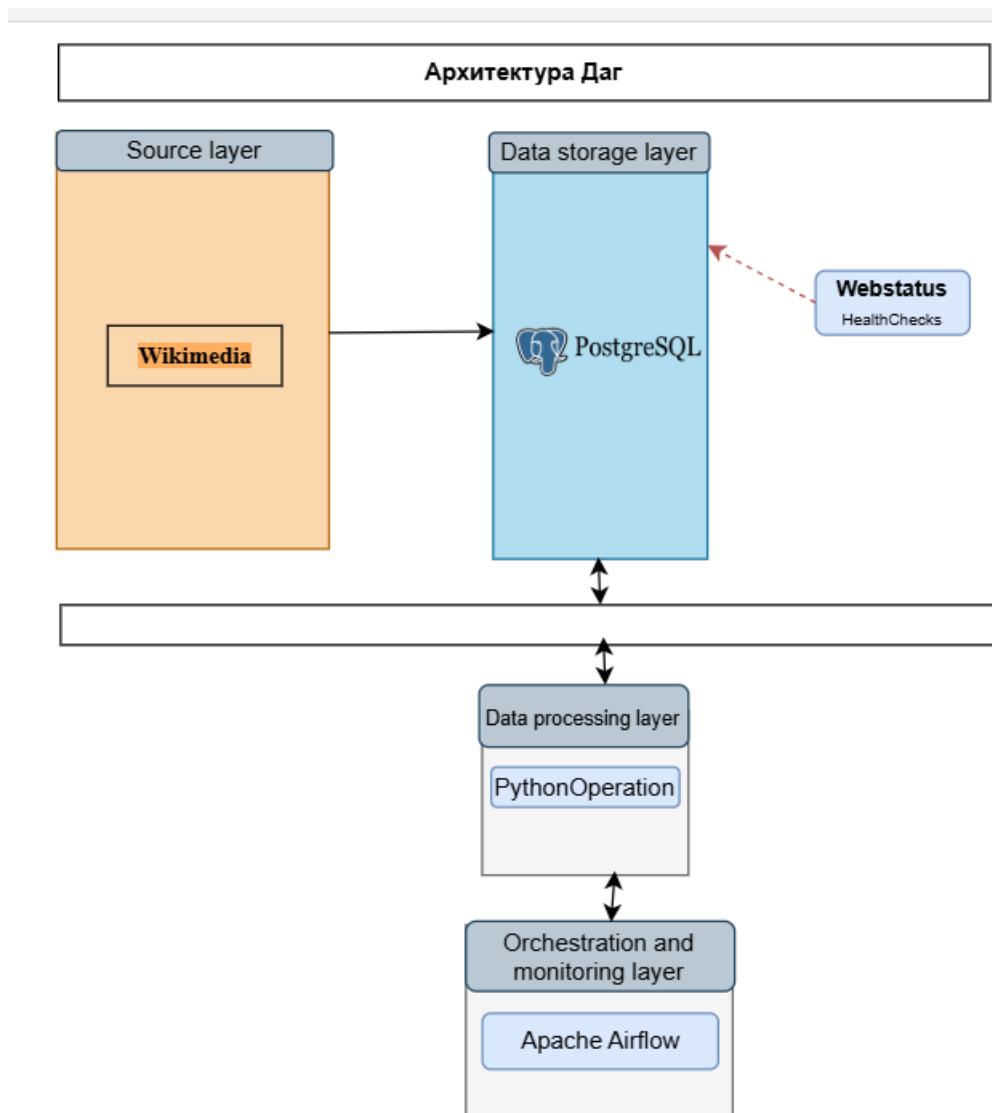


Рисунок 2 – Архитектура DAG

На рисунке 3 показан запуск сервисов

```
dev@dev-vm:~/workshop-on-ETL/business_case_stocksense_25$ sudo docker compose up --build
[+] Running 8/8
 ✓ Network business_case_stocksense_25_default          Created
 ✓ Volume "business_case_stocksense_25_postgres_data"   Created
 ✓ Volume "business_case_stocksense_25_logs"            Created
 ✓ Container business_case_stocksense_25-wiki_results-1 Created
 ✓ Container business_case_stocksense_25-postgres-1     Created
 ✓ Container business_case_stocksense_25-init-1         Created
 ✓ Container business_case_stocksense_25-webserver-1    Created
 ✓ Container business_case_stocksense_25-scheduler-1    Created
Attaching to init-1, postgres-1, scheduler-1, webserver-1, wiki_results-1
wiki_results-1 | The files belonging to this database system will be owned by user "postgres".
postgres-1     | The files belonging to this database system will be owned by user "postgres".
wiki_results-1 | This user must also own the server process.
postgres-1     | This user must also own the server process.
wiki_results-1 |
postgres-1     |
wiki_results-1 | The database cluster will be initialized with locale "en_US.utf8".
wiki_results-1 | The default database encoding has accordingly been set to "UTF8".
wiki_results-1 | The default text search configuration will be set to "english".
wiki_results-1 |
```

Рисунок 3 – Запуск сервисов

На рисунке 4 показаны даги airflow

DAGs

All 1 Active 4 Paused 4

Running 2 Failed 3

Filter DAGs by tag

Search DAGs

Auto-refresh

DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks	Actions	Links
listing_4_01	airflow	10	@hourly	2025-04-02, 01:00:00	2025-04-02, 02:00:00	...	▶ 🛑	...
listing_4_03	airflow	5	@daily	2025-04-03, 00:00:00	2025-04-04, 00:00:00	...	▶ 🛑	...
listing_4_05	airflow	5	@hourly	2025-04-04, 11:00:00	2025-04-04, 12:00:00	...	▶ 🛑	...
listing_4_07	airflow	5	@daily	2025-04-04, 12:49:28	2025-04-04, 00:00:00	...	▶ 🛑	...
listing_4_08	airflow	5	@daily	2025-04-03, 00:00:00	2025-04-04, 00:00:00	...	▶ 🛑	...
listing_4_13	airflow	5	@hourly	2025-04-04, 11:00:00	2025-04-04, 12:00:00	...	▶ 🛑	...
listing_4_15	airflow	10	@hourly	2025-04-04, 12:50:55	2025-04-04, 09:00:00	...	▶ 🛑	...
listing_4_18	airflow	10	@hourly	2025-04-04, 13:00:00	2025-04-04, 14:00:00	...	▶ 🛑	...

Рисунок 4 – DAG Airflow

На рисунке 5 показан работающий даг.

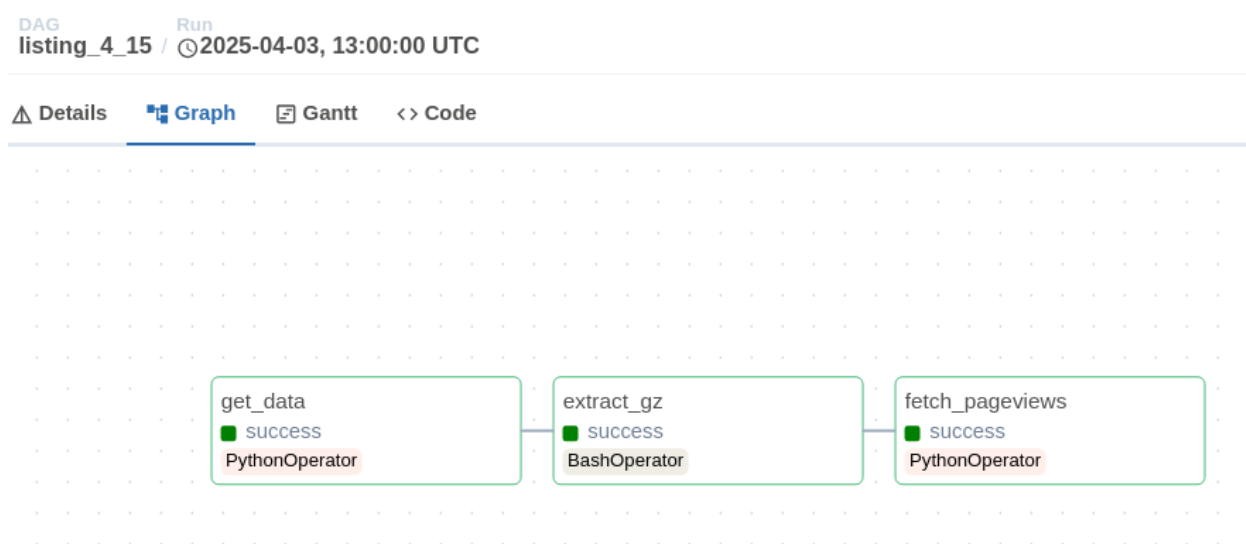


Рисунок 5 – Работающий DAG

Рисунок 6 демонстрирует диаграмму Ганта

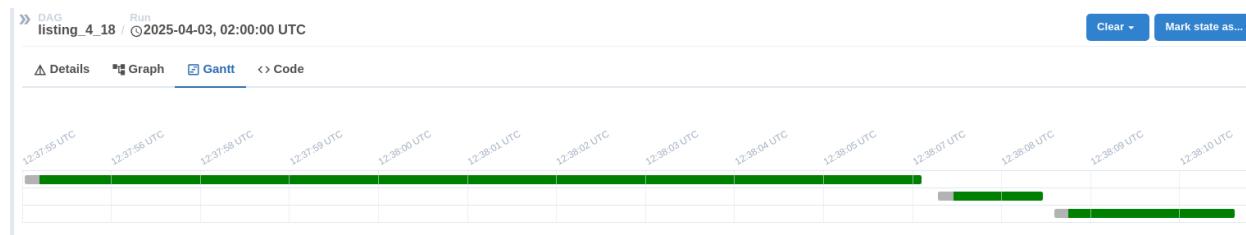


Рисунок 6 – Диаграмма Ганта

На рисунке 7 показана ERD диаграмма базы данных

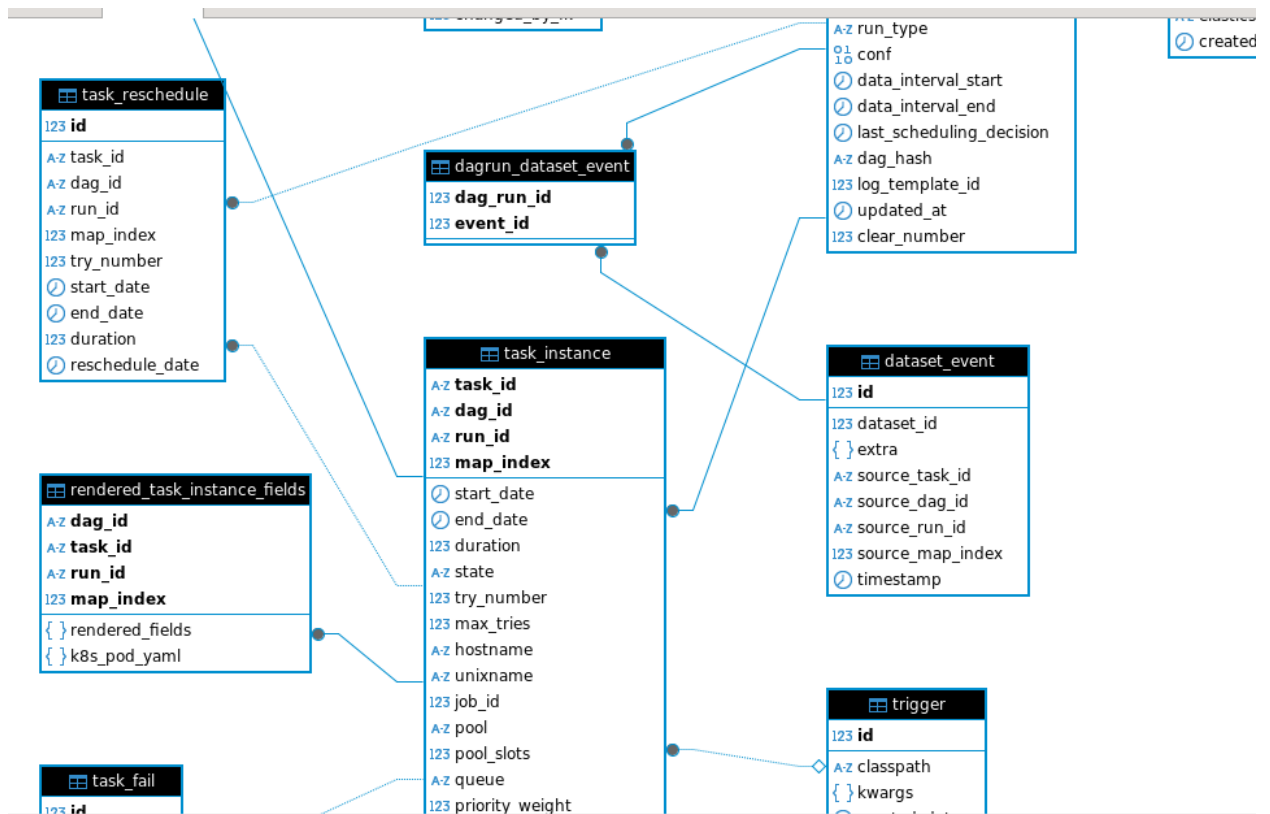


Рисунок 7 – ERD

На рисунке 8 показано выполнение запроса в таблице ab_role

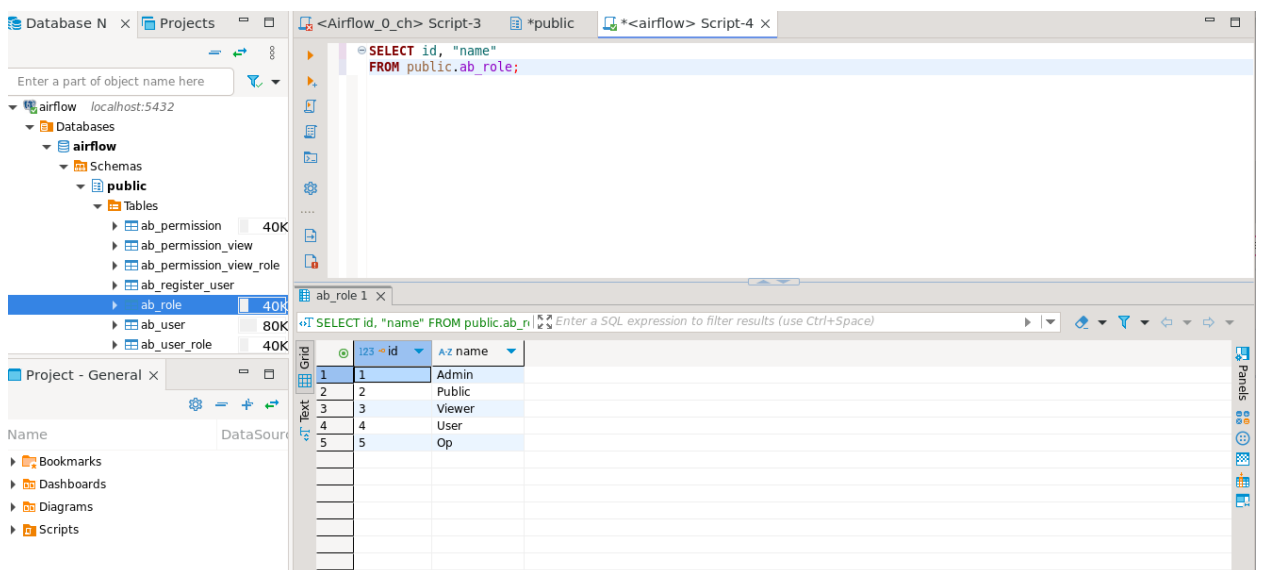


Рисунок 8 – Выполнение запроса

Выводы:

- Построена верхнеуровневая архитектура задания Бизнес-кейса «StockSense», выполненная в draw.io;
- Построена архитектура DAG Бизнес-кейса «StockSense», выполненная в draw.io;

- Построена диаграмма Ганта DAG в Apache Airflow;
- Построена ERD-схема базы данных Postgre SQL;
- Выполнены SQL-запросы, позволяющие проверить наличие выгруженных агрегированных данных бизнес-задачи.