

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

Макарова Виктория Сергеевна

### **ЛАБОРАТОРНАЯ РАБОТА 6.1**

Проектный практикум по разработке ETL-решений

Направление подготовки

38.03.05 Бизнес-информатика

Профиль подготовки

Аналитика данных и эффективное управление

Курс обучения: 4

Форма обучения: очная

Преподаватель: кандидат технических наук,

доцент Босенко Тимур Муртазович

Москва

2025

### **Постановка задачи:**

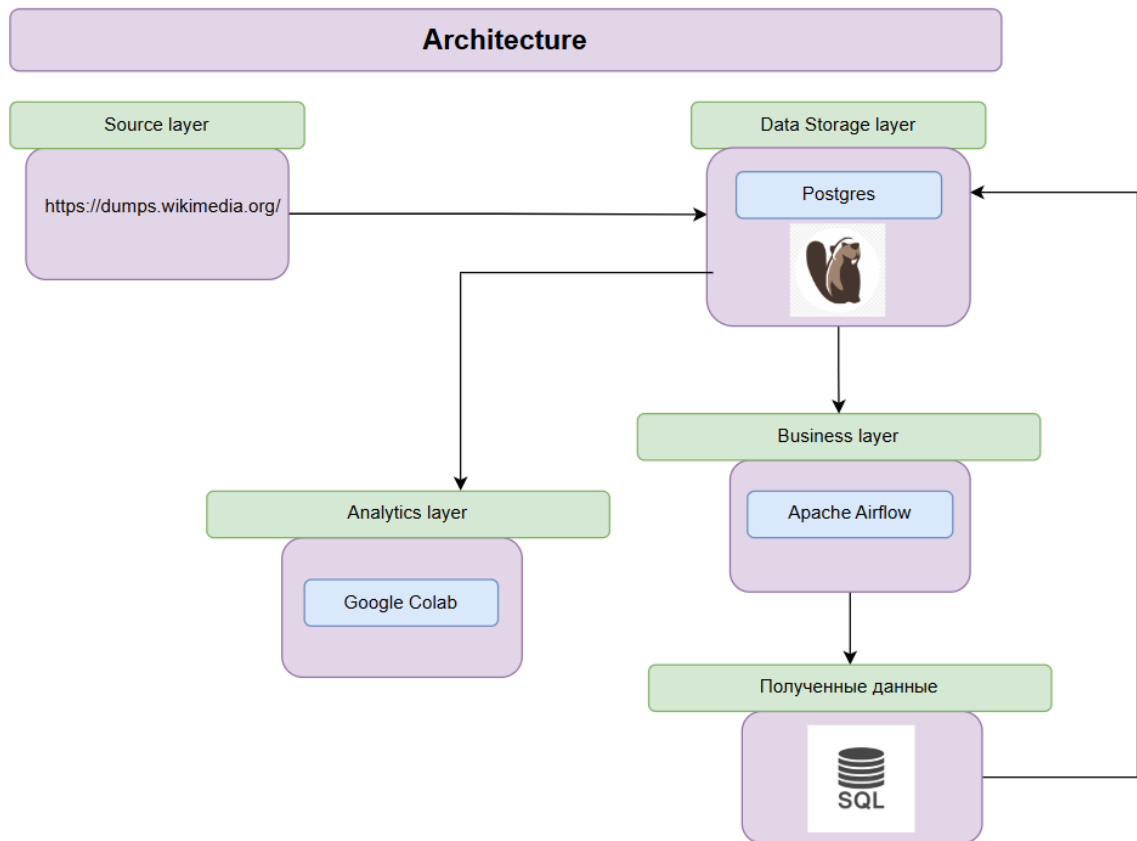
- исходный код всех DAGs, которые требовались для решения задачи, а также представить граф с DAG в Apache Airflow;
- верхнеуровневая архитектура задания Бизнес-кейса «StockSense», выполненная в draw.io;
- архитектура DAG Бизнес-кейса «StockSense», выполненная в draw.io;
- диаграмма Ганта DAG в Apache Airflow;
- ERD-схема базы данных Postgre SQL;
- SQL-запросы, позволяющие проверить наличие выгруженных агрегированных данных бизнес-задачи.

### **Индивидуальное задание Вариант 7**

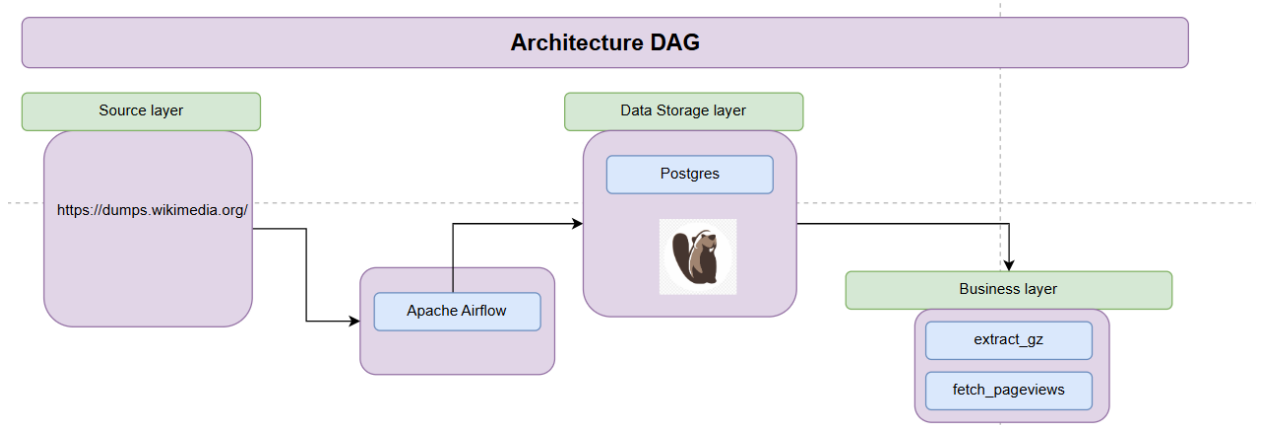
1. Получить данные за месяц для сайта **Sberbank**
2. Скачайте данные за месяц для страницы Sberbank и сохраните их в базе данных. Напишите SQL-запрос для подсчета максимальных просмотров по часам и отобразите эти данные на графике.

## ХОД РАБОТЫ

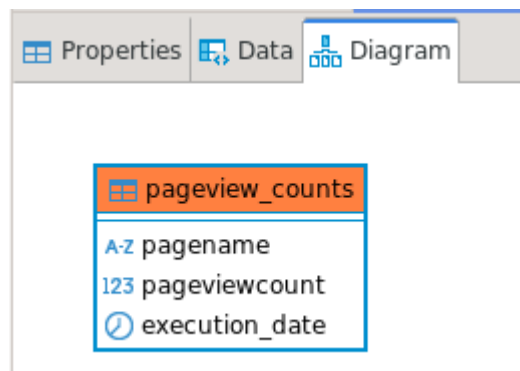
Верхнеуровневая архитектура задания Бизнес-кейса «StockSense».



Архитектура DAG Бизнес-кейса «StockSense», выполненная в draw.io.



ERD-схема базы данных Postgre SQL



# Остановка и удаление всех контейнеров Docker. Удаление всех образов Docker.

```
● dev@dev-vm: ~/Downloads/business_case_stockssense_25$ sudo docker stop $(sudo docker ps -a -q)
[sudo] password for dev:
fc007051f0e1
6df627e8ff13
2753belf29ae
bccd94b81758
69e5d6b73077
a60dc47fa9a1
148713ab18ab
2653d61b65e8
● dev@dev-vm: ~/Downloads/business_case_stockssense_25$ sudo docker rm -f $(sudo docker ps -a -q)
fc007051f0e1
6df627e8ff13
2753belf29ae
bccd94b81758
69e5d6b73077
a60dc47fa9a1
148713ab18ab
2653d61b65e8
● dev@dev-vm: ~/Downloads/business_case_stockssense_25$ sudo docker rmi -f $(sudo docker images -q)
Untagged: custom-airflow:slim-2.8.1-python3.11
Deleted: sha256:25c169463df682454db3c5005bcf82383414df8fa005e523ea63391978bb157d
Untagged: postgres:12-alpine
Untagged: postgres@sha256:7c8f4870583184ebadf7f17a6513620aac5f365a7938dc6a6911c1d5df2f481a
Deleted: sha256:097d7e0eba6e7120560ac2d3a7e4b2a110d219b7dcbf43c0318dce1849c9761a
Deleted: sha256:1c6cab8231a7049122e7c4b7acb67e78559a8bafef7199b66e81255f4ec813b1
Deleted: sha256:9e186a4f0e28f0d123152f535c55a9f089587e4e0f675b5d02f1c7a6a73fbald
Deleted: sha256:109288deabff2a135d2949de5ab4c731caf7368baf2136bb31b59bd40f4a3335
Deleted: sha256:90246f39d5df9f7f2bd057a29e4b07da6d4041224fb8a177aca0563547d6f396
Deleted: sha256:8556313fe0327040ae65ace97253e7f5a5374b961c64369fe7f6d33b77e4bb04
Deleted: sha256:353b991678409b24c6e5eb46fc4c721f920bf5e9de787b1e426380a6ec8eb502
Deleted: sha256:7bd8729de8f37344919f3c646f21391f70514c8bb331da3cc9ce138b81f38c58
Deleted: sha256:693138fbc86453ae9c7a5a97569c2bf319454e905231a026c7807142442f7b4a
Deleted: sha256:6bc5f5702e2f900cf1005d4497b022e16764fd68960e0d065d68baaf1d31dfbe
Deleted: sha256:56c9620121b8b9d688cf6bbddb14d7c2253972e9aae854f41fe170f6a7972882
Deleted: sha256:a0904247e36a7726c03c71ee48f3e64462021c88dafeb13f37fda613b27f11c
○ dev@dev-vm: ~/Downloads/business_case_stockssense_25$
```

Создание нового Docker образа.  
Эта команда строит новый Docker образ с именем custom-airflow:slim-2.8.1-python3.11 на основе текущего контекста:

```
dev@dev-vm: ~/Downloads/business_case_stockssense_25$ sudo docker build -t custom-airflow:slim-2.8.1-python3.11 .
[sudo] password for dev:
2025/04/04 19:00:13 in: {}string{}
2025/04/04 19:00:13 Parsed entitlements: {}
[+] Building 423.3s (7/7) FINISHED
=> [internal] load build definition from Dockerfile
=> => transferring dockerfile: 615B
=> [internal] load metadata for docker.io/apache/airflow:slim-2.8.1-python3.11
=> [internal] load .dockerignore
=> => transferring context: 2B
=> CACHED [1/5] FROM docker.io/apache/airflow:slim-2.8.1-python3.11@sha256:751babd58a83e44ae23393fe1552196c25f3e2683c97db1a6d98b7d15e7a0e8
=> [2/3] RUN pip install --no-cache-dir pandas scikit-learn joblib requests azure-storage-blob==12.8.1 psycopg2-binary "connexion[swagger-ui]" apache-airflow-providers-postgres==2
=> [3/3] RUN mkdir -p /opt/airflow/data /opt/airflow/logs && chown -R airflow: /opt/airflow/data /opt/airflow/logs
=> => exporting layers
=> => writing image sha256:266b0f93e85757cb42448929a93db25478786cd70d3e76b00f707a32ad8577b
=> => naming to docker.io/library/custom-airflow:slim-2.8.1-python3.11
dev@dev-vm: ~/Downloads/business_case_stockssense_25$
```

## Запуск среды с помощью Docker Compose.

```
○ dev@dev-vm: ~/Downloads/business_case_stockssense_25$ sudo docker compose up --build
[+] Running 13/13
✔ wiki_results Pulled
  ✔ 1f3e46996e29 Pull complete
  ✔ 47e20ba08731 Pull complete
  ✔ 101b82465a4f Pull complete
  ✔ 319529a7ccb0 Pull complete
  ✔ c2f9392cf04c Pull complete
  ✔ 4eb446cc095d Pull complete
  ✔ 47bfe778b869 Pull complete
  ✔ b1066b287aa8 Pull complete
  ✔ 7865e52a4759 Pull complete
  ✔ 7d7514147c2 Pull complete
  ✔ 11852a542467 Pull complete
✔ postgres Pulled
[+] Running 8/8
✔ Network business_case_stockssense_25_default Created
✔ Volume "business_case_stockssense_25_logs" Created
✔ Volume "business_case_stockssense_25_postgres_data" Created
✔ Container business_case_stockssense_25_postgres-1 Created
✔ Container business_case_stockssense_25-wiki_results-1 Created
✔ Container business_case_stockssense_25-init-1 Created
✔ Container business_case_stockssense_25-scheduler-1 Created
✔ Container business_case_stockssense_25-webserver-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".
wiki_results-1 | This user must also own the server process.
wiki_results-1 |
```

## Подключение к PostgreSQL через DBeaver для просмотра таблицы

### Создание нового подключения в DBeaver

Connection "airflow" configuration

Connection settings  
PostgreSQL connection settings

Connection settings  
Initialization  
Shell Commands  
Client identification  
Transactions  
General  
Metadata  
Errors and timeouts  
Data Transfer  
Data Editor  
SQL Editor

Main PostgreSQL Driver properties SSH SSL + Network configurations...

Server  
Connect by: ☒ Host ☐ URL  
URL: jdbc:postgresql://localhost:5433/airflow  
Host: localhost Port: 5433  
Database: airflow ☐ Show all databases

Authentication  
Authentication: Database Native  
Username: airflow  
Password:  Save password

Advanced  
Session role:  Local Client:

[Connection variables information](#) [Database documentation](#)

Driver name: PostgreSQL Driver Settings Driver license

Test Connection ... Cancel OK

Test Connection, чтобы убедиться, что DBeaver может подключиться к базе данных.

Connection test

**i** Connected (5596 ms)

Server: PostgreSQL 12.22  
PostgreSQL 12.22 on x86\_64-pc-linux-musl, compiled by gcc (Alpine 14.2.0) 14.2.0, 64-bit

Driver: PostgreSQL JDBC Driver 42.7.2

Details >> OK

Проверка доступности AirFlow <http://localhost:8080/>



**DBeaver 25.0.1 - pageview\_counts**

File Edit Navigate Search SQL Editor Database Window Help

airflow public@airflow

pageview\_counts

Enter a part of object name ...

localhost:5433

airflow

public

pageview\_counts

Columns

AZ pagename (varchar(50))

I23 pageviewcount (int4)

execution date (timestamp)

Project - General

Name DataS

Bookmarks

Dashboards

Diagrams

Scripts

Grid

	AZ pagename	I23 pageviewcount	execution date
1	Sberbank	0	2025-01-10 12:00:00.000
2	Sberbank	0	2025-01-11 12:00:00.000
3	Sberbank	0	2025-01-12 12:00:00.000
4	Sberbank	0	2025-01-13 12:00:00.000
5	Sberbank	0	2025-03-06 14:00:00.000
6	Sberbank	0	2025-04-04 22:39:05.746

Refresh Save Cancel Export data 200 6

6 row(s) fetched - 0.018s (0.001s fetch), on 2025-04-05 at 01:47:34

Выгружались данные за месяц март 2025 за 14:00

**DBeaver 25.0.1 - pageview\_counts**

File Edit Navigate Search SQL Editor Database Window Help

airflow public@airflow

pageview\_counts

Enter a SQL expression to filter results (use Ctrl+Space)

Grid

	AZ pagename	I23 pageviewcount	execution date
1	Sberbank	0	2025-01-10 12:00:00.000
2	Sberbank	0	2025-01-11 12:00:00.000
3	Sberbank	0	2025-01-12 12:00:00.000
4	Sberbank	0	2025-01-13 12:00:00.000
5	Sberbank	0	2025-03-06 14:00:00.000
6	Sberbank	0	2025-04-04 22:39:05.746
7	Sberbank.com	0	2025-03-06 14:00:00.000
8	Sberbank	0	2025-03-06 14:00:00.000
9	Sberbank.ru	0	2025-03-06 14:00:00.000
10	Sberbank.com	0	2025-03-07 14:00:00.000
11	Sberbank	0	2025-03-07 14:00:00.000
12	Sberbank.ru	0	2025-03-07 14:00:00.000
13	Sberbank.com	0	2025-03-08 14:00:00.000
14	Sberbank	0	2025-03-08 14:00:00.000
15	Sberbank.ru	0	2025-03-08 14:00:00.000
16	Sberbank.com	0	2025-03-09 14:00:00.000
17	Sberbank	0	2025-03-09 14:00:00.000
18	Sberbank.ru	0	2025-03-09 14:00:00.000
19	Sberbank.com	0	2025-03-10 14:00:00.000
20	Sberbank	0	2025-03-10 14:00:00.000
21	Sberbank.ru	0	2025-03-10 14:00:00.000
22	Sberbank.com	0	2025-03-11 14:00:00.000
23	Sberbank	0	2025-03-11 14:00:00.000
24	Sberbank.ru	0	2025-03-11 14:00:00.000