Департамент образования города Москвы

Государственное автономное образовательное учреждение

высшего образования города Москвы

«Московский городской педагогический университет»

Институт цифрового образования

Департамент информатики, управления и технологий

ОТЧЕТ

по дисциплине «Интеграция и развертывание программного обеспечения с помощью контейнеров»

Направление подготовки 38.03.05 – бизнес-информатика

Профиль подготовки «Аналитика данных и эффективное управление»

(очная форма обучения)

Лабораторная работа 6.1 Разработка полного ETL-процесса. Оркестровка конвейера данных

Выполнил:

St\_89

Руководитель:

Москва  
2025

**Цель:** Изучить методы и разработку полного ETL-процесса и оркестровка конвейера данных (Вариант 2)

Задачи:

1. Спроектировать верхнеуровневую архитектуру аналитического решения Бизнес-кейса «StockSense»
2. Спроектировать архитектуру DAG **Бизнес-кейса «StockSense»**
3. Скачайте данные за день для страницы VK и сохраните их в базе данных. Напишите SQL-запрос для подсчета максимального числа просмотров по часам, а затем визуализируйте данные для анализа.

**Ход работы:**

1. Для начала измени DAG в связи с заданием и запустим его.

Рис. 1 – Запуск DAG

Из-за того, что иногда доступ к сайту отваливается, было решено добавить в DAG delay и повторный запуск



Рис. 2 – Добавление delay

После этого видим данные, которые записались в postgres

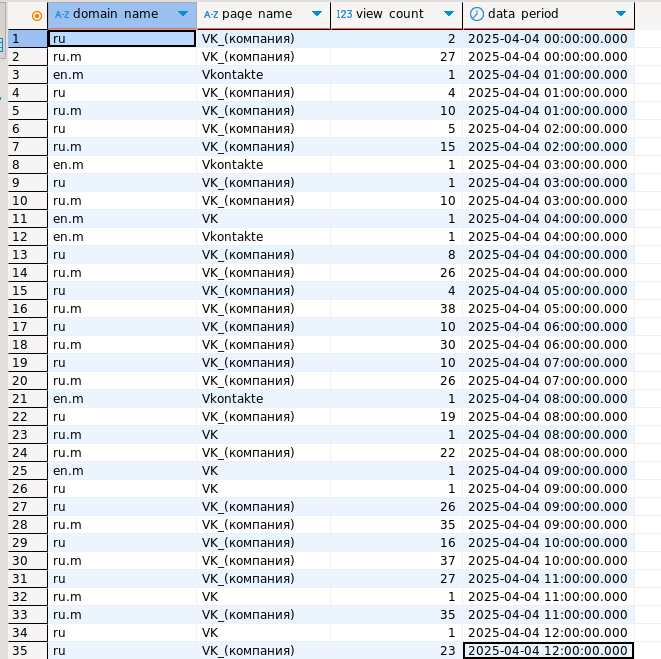


Рис. 3 – Данные из Postgres

Чтобы узнать, сколько максимум заходило на страницу, напишем SQL script

*SELECT x.page\_name, x.hr AS "hour", x.average AS "max pageviews"*

*FROM (*

*SELECT*

*page\_name,*

*date\_part('hour', data\_period) AS hr,*

*max(view\_count) AS average,*

*ROW\_NUMBER() OVER (PARTITION BY page\_name ORDER BY max(view\_count) DESC) AS row\_number*

*FROM pageview\_counts*

*GROUP BY page\_name, hr*

*) AS x*

*WHERE x.row\_number = 1;*

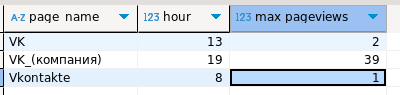


Рис. 4 – Полученные данные

После запроса, видим, что страница VK\_(компания) имеет 39 просмотров за 13 часов

1. Загрузим основные данные в Yandex Datalence и визуализируем их. Также добавлен селектор, чтобы смотреть детальную статистику по странице или домену

<https://datalens.yandex/ovwre8wqn0i4a>



Рис. 5 – Дашборд данных

1. Верхнеуровневаю архитектуру аналитического решения Бизнес-кейса «StockSense»

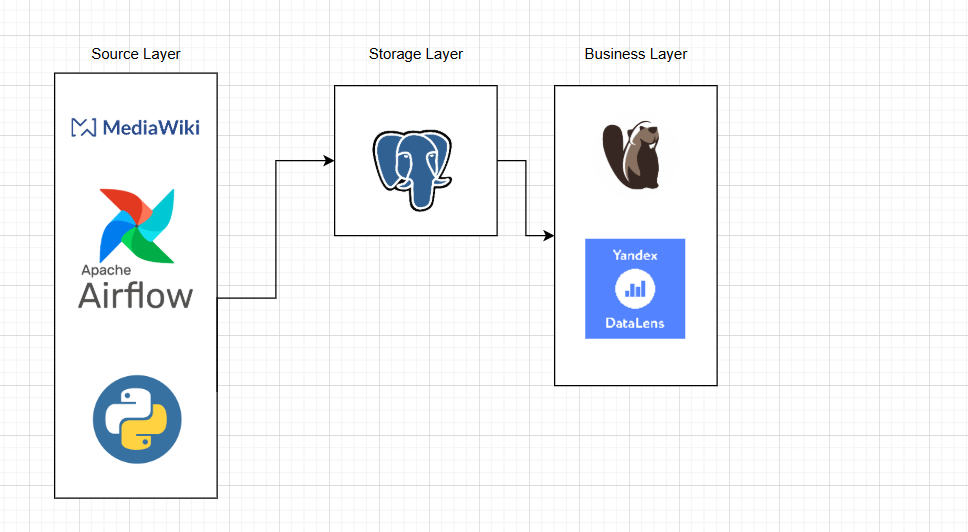


Рис. 6 - Верхнеуровневая архитектура аналитического решения

1. Спроектировать архитектуру DAG **Бизнес-кейса «StockSense»**

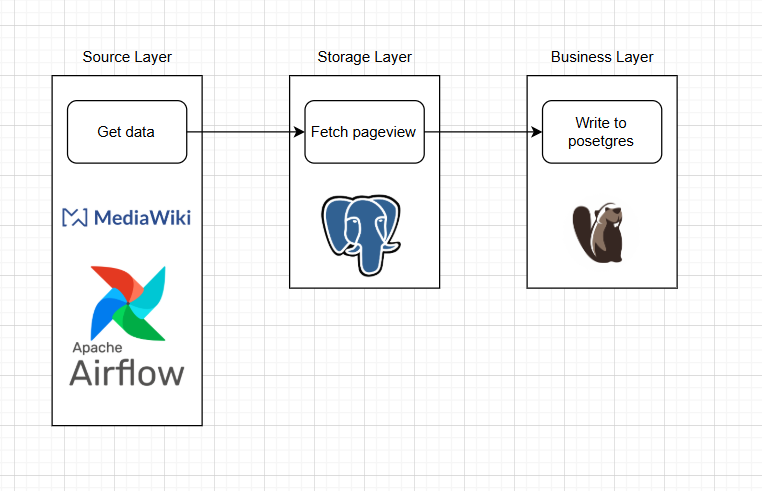


Рис. 7 – Архитектура DAG

**Вывод:** Была проведена выгрузка данных посещаемости страницы VK из Wikimedia. Данные загружены в Postgres и визуализированы