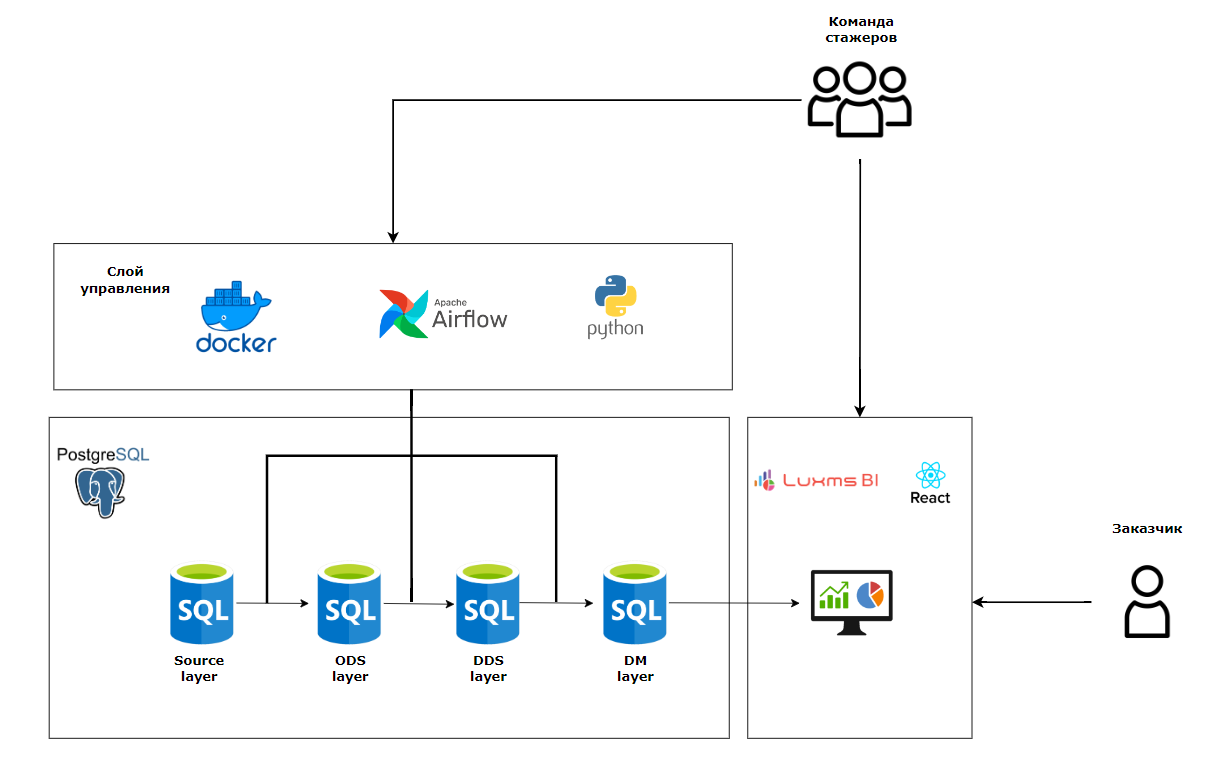
**АРХИТЕКТУРА РЕШЕНИЯ**

**Исполнитель:** Годзун Егор Андреевич

**Должность:** стажер

**Дата:** 21.07.2024

1. **Схема архитектуры**



1. **Описание компонентов**

**Описание слоя управления**

В Docker Dekstop поднят контейнер с Airflow. Airflow оркестрирует ETL-поток на база python-скриптов (ДАГов)

Веб-интерфейс Airflow: <http://localhost:8080>

User: airflow

Структура проекта:

* dags: директория dags содержит файлы DAG, предназначенные для определения рабочих процессов, путем устанавки задач и их зависимостей, расписания и параметров выполнения

scripts: поддиректория для хранения вспомогательных скриптов;

sql\_scripts: поддиректория для хранения вспомогательных sql-скриптов

test\_dag.py: Пример DAG, который читает данные из базы данных.

dag\_ods\_layer.py: DAG, который загружает исходные данные в слой ODS.

dag\_dds\_layer.py: DAG, который загружает данные из слоя ODS в слой DDS.

dag\_dm\_layer.py: DAG, который загружает данные из слоя DDS в слой DM.

* docker: директория содержит файлы docker-compose.yaml, Dockerfile, README.md и requirements.txt
* files: директория для хранения документации

**Описание ETL потока**

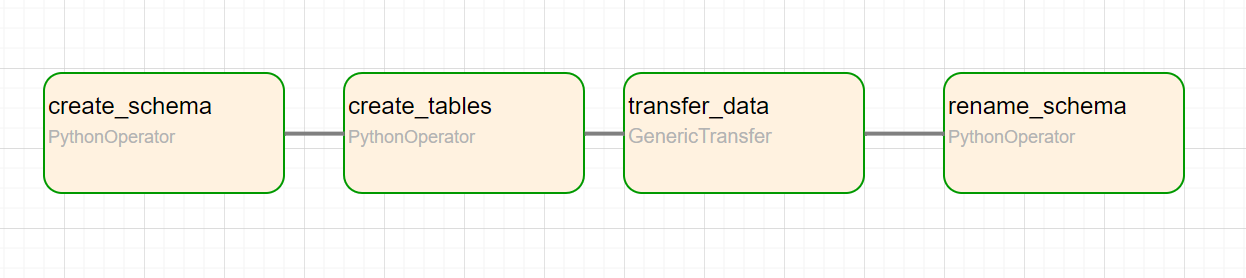


Схема для ДАГа dag\_ods\_layer

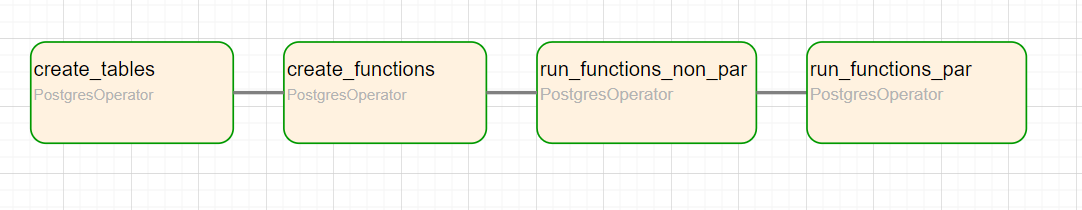


Схема для ДАГа dag\_dds\_layer

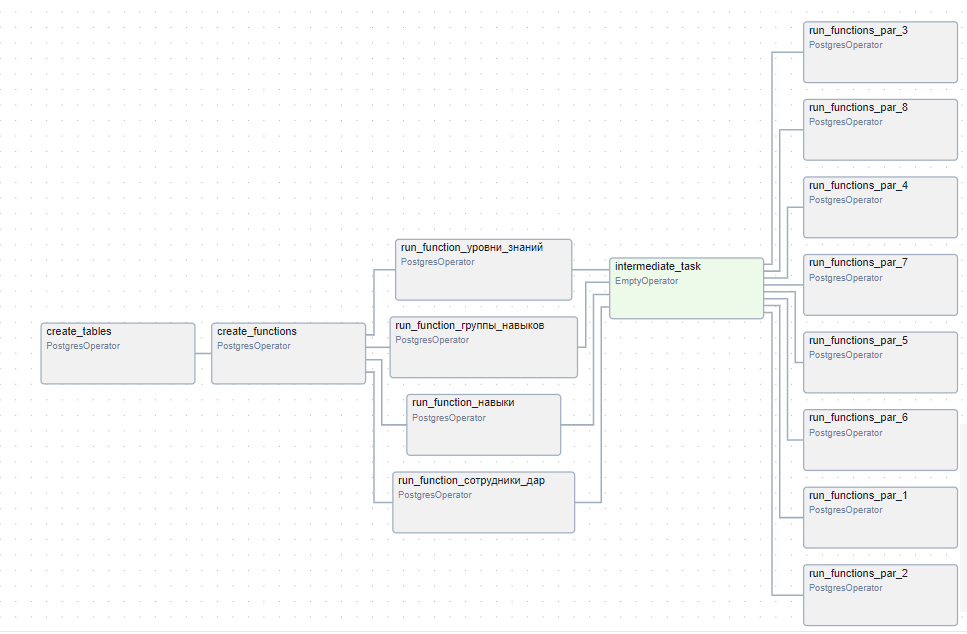


Схема для ДАГа dag\_dm\_layer

ETL-поток реализован поочередным запуском ДАГов:

* dag\_ods\_layer: загружает данные в слой ODS из базы данных source (Source layer);
* dag\_dds\_layer: загружает данные в слой DDS из слоя ODS, в процессе переноса данные очищаются, преобразуются и удаляются ненужные атрибуты
* dag\_dm\_layer: загружает данные в DM из слоя DDS, при этом выполняется фильтрация и необходимые вычисления

При передаче данных между слоями происходит распараллеливание процессов.

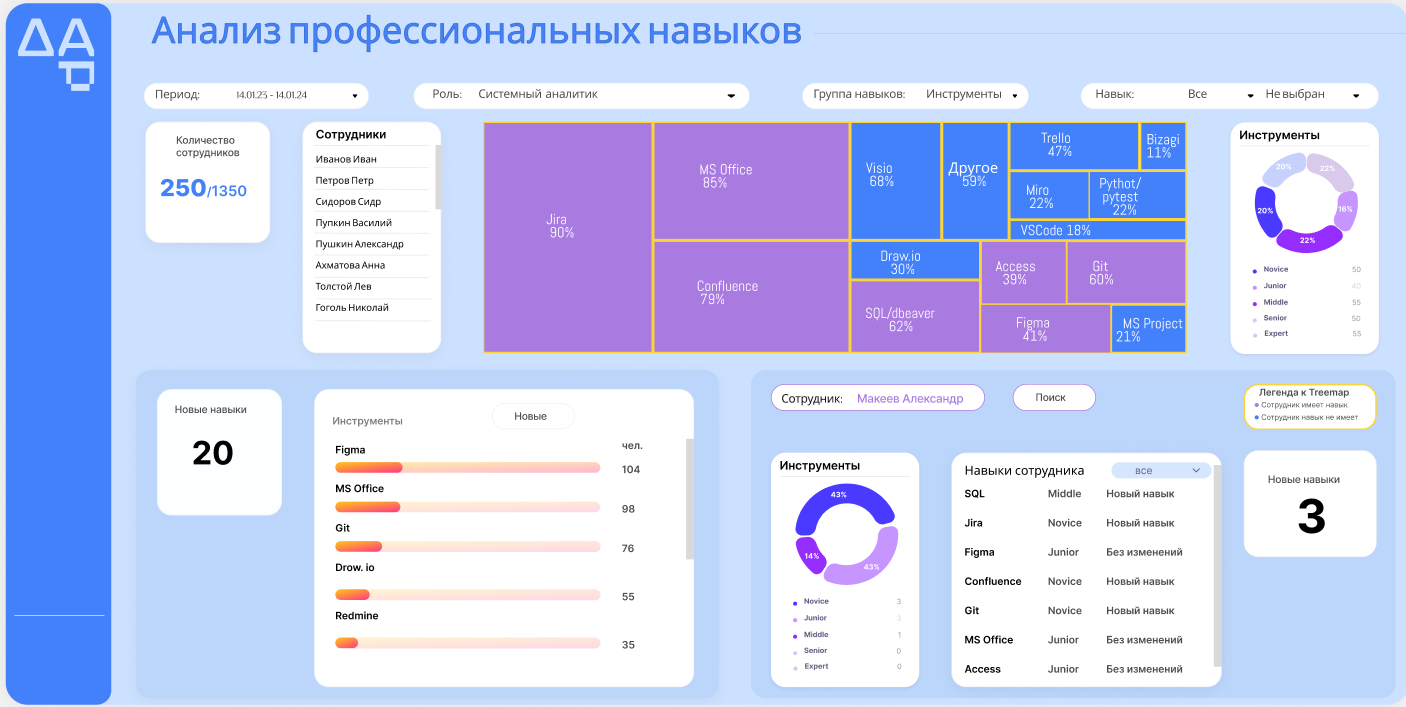
**Описание хранилища данных**

Хранилища данных реализовано в СУБД PostgreSQL со следующей структурой:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Слой** | **Схема** | **База данных** | **Параметры подключения** | **Имя учетной записи** |
| Source layer | source\_data | source | хост: 10.82.0.4  порт: 5432 | etl\_user\_8 |
| ODS layer | ods | etl\_db\_8 |
| DDS layer | dds |
| DM layer | dm |

**Описание построения итоговой отчетности**

Итоговая отчетность реализована с помощью Luxms BI и React с дашбордами, реализованными на основе данных из слоя «DM layer».



Итоговый дашборд

1. **Описание используемых сущностей с данными**

Используемые сущности:

|  |  |
| --- | --- |
| **Схема** | **Сущности** |
| source\_data | «базы\_данных», «базы\_данных\_и\_уровень\_знаний\_сотру», «инструменты», «инструменты\_и\_уровень\_знаний\_сотр», «образование\_пользователей», «опыт\_сотрудника\_в\_отраслях», «опыт\_сотрудника\_в\_предметных\_обла», «отрасли», «платформы», «платформы\_и\_уровень\_знаний\_сотруд», «предметная\_область», «резюмедар», «сертификаты\_пользователей», «сотрудники\_дар», «среды\_разработки», «среды\_разработки\_и\_уровень\_знаний\_», «технологии», «технологии\_и\_уровень\_знаний\_сотру», «типы\_систем», «типы\_систем\_и\_уровень\_знаний\_сотру», «уровень\_образования», «уровни\_владения\_ин», «уровни\_знаний», «уровни\_знаний\_в\_отрасли», «уровни\_знаний\_в\_предметной\_област», «фреймворки», «фреймворки\_и\_уровень\_знаний\_сотру», «языки», «языки\_пользователей», «языки\_программирования», «языки\_программирования\_и\_уровень» |
| ods | те же, что и в source\_data кроме «резюмедар» |
| dds | те же, что и в ods |
| dds\_er | те же, что и в dds кроме «сотрудники\_дар» |
| dm | «сотрудники\_дар», «группы\_навыков», «навыки», «уровни\_знаний», «группы\_навыков\_и\_уровень\_знаний\_сотруд» |