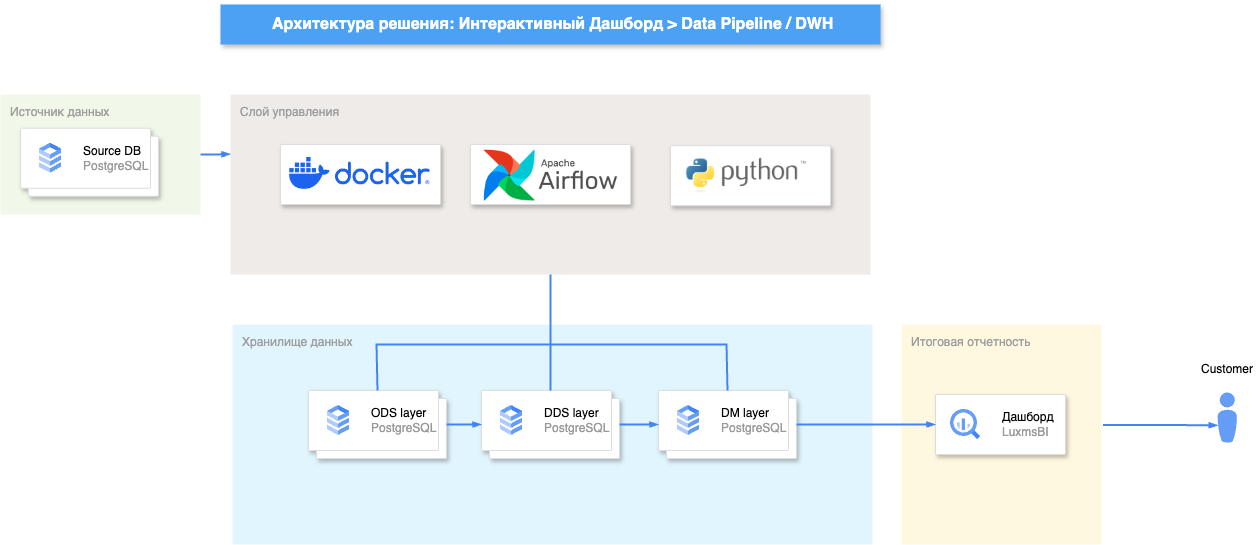
**Архитектура решения**

**«Интерактивный дашборд: статистика роста профессиональных компетенций сотрудников»**

1. **Схема архитектуры**



1. **Описание компонентов**

**Описание слоя управления:**

Слой управления реализован с использованием Apache Airflow, развернутого в контейнере на платформе Docker Desktop. Apache Airflow оркестрирует ETL-процессы, которые выполняются посредством Python-скриптов.

* **Веб-интерфейс:** доступ к интерфейсу управления предоставляется по адресу <http://localhost:8080>. Для авторизации используется учетная запись с именем пользователя airflow.
* **Хранилище DAG'ов и скриптов:** для хранения Directed Acyclic Graphs (DAG'ов) и исполняемых скриптов в контейнере примонтирована локальная папка /dags.

**Структура хранения кода DAGов и исполняемых скриптов:**

Код DAGа хранится в dags/refresh\_data.py

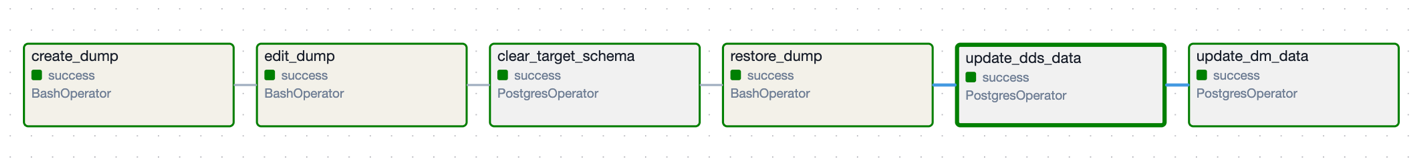
Код DDL- и DML-скриптов хранятся в dags/scripts/sql-scripts/

**Описание инициализации схем:**

Инициализация схем слоев ODS, DDS и DM производится посредством выполнения DDL-и DML-скриптов из директории sql-scripts/ для создания необходимых схем и таблиц, а также заполнения их данными. Исполняемые python-скрипты хранятся в папке init директории scripts/.

Слой ODS представляет собой временное хранилище данных перед их обработкой и трансформацией и является полной копией source. В схеме DDS создаются нормализованные и структурированные таблицы слоя ODS. Слой DM (Data Mart) предназначен для хранения агрегированных и подготовленных для анализа данных для последующего использования в отчетности.

**Описание ETL потока:**



ETL-поток реализован DAGом Airflow refresh\_data, который запускается ежедневно и отвечает за обновление данных в слоях базы данных.

Даг состоит из следующих taskов:

Обновление данных в ODS слое:

* create\_dump — с помощью bash-оператора создает дамп source\_data – схемы источника данных.
* edit\_dump — с помощью bash-оператора меняет название схемы в файле дампа с source\_data на ods\_data.
* clear\_target\_schema — очищает данные в схеме ods\_data.
* restore\_dump — выполняет выгрузку данных в слой ODS.

Обновление данных в DDS и DM слоях:

* update\_dds\_data — запускает DML-скрипт dds\_refresh.sql, который выполняет инкрементальное обновление данных в слое DDS на основе обновленных данных в слое ODS, используя дату изменения.
* update\_dm\_data — запускает DML-скрипт dm\_refresh.sql, который выполняет инкрементальное обновление данных в слое DM на основе обновленных данных в слое DDS, используя дату изменения.

**Описание Хранилища Данных:**

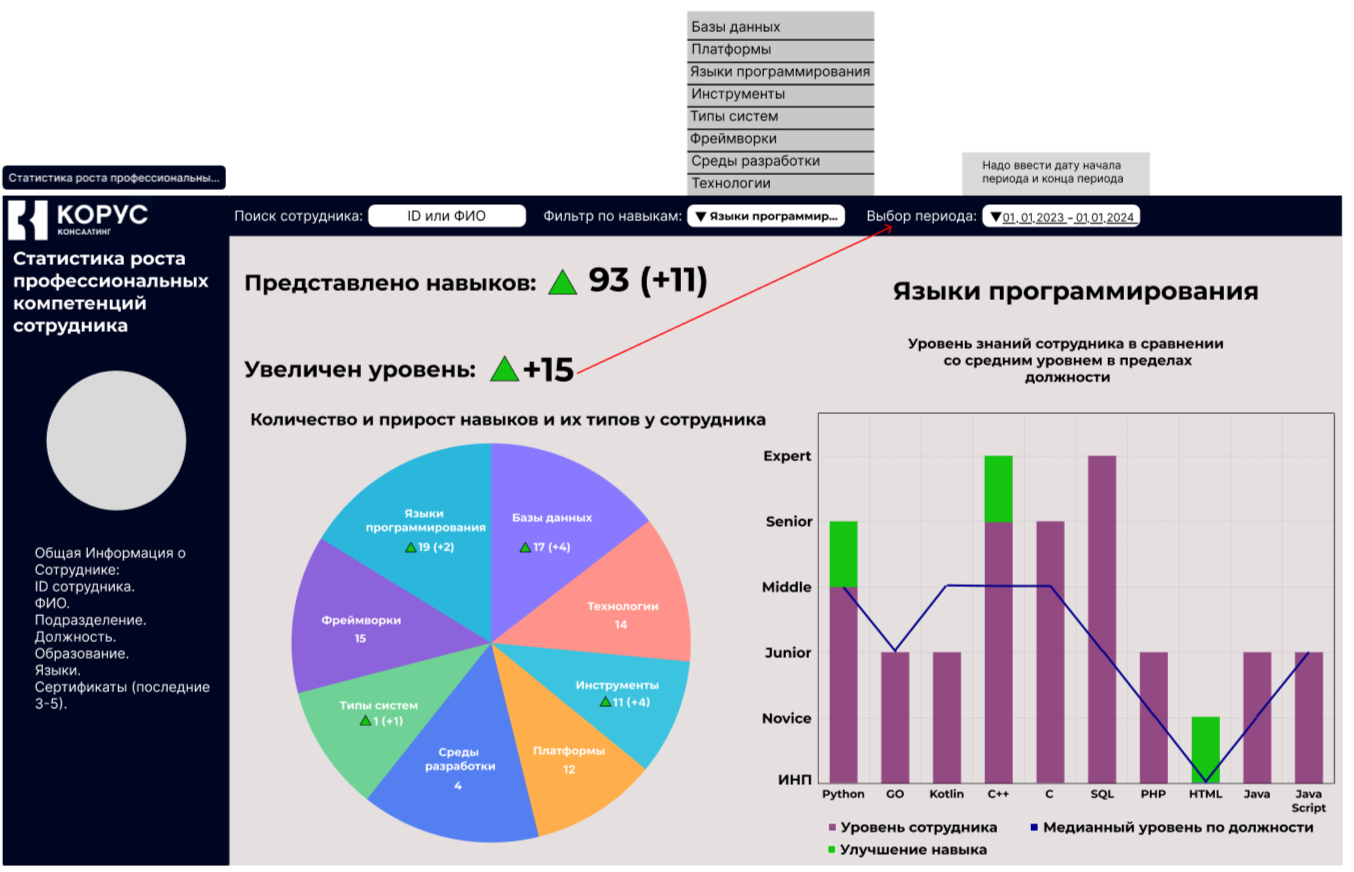
Хранилище данных реализовано в СУБД PostgreSQL со следующей структурой:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Слой | Схема | База Данных | Параметры подкюения к БД | Имя учетной записи |
| Source layer | sourse\_data | source | host: 10.82.0.4 port: 5432 | etl\_user\_3 |
| ODS layer | ods\_data | etl\_db\_3 |
| DDS layer | dds\_data |
| DM layer | dm\_data |

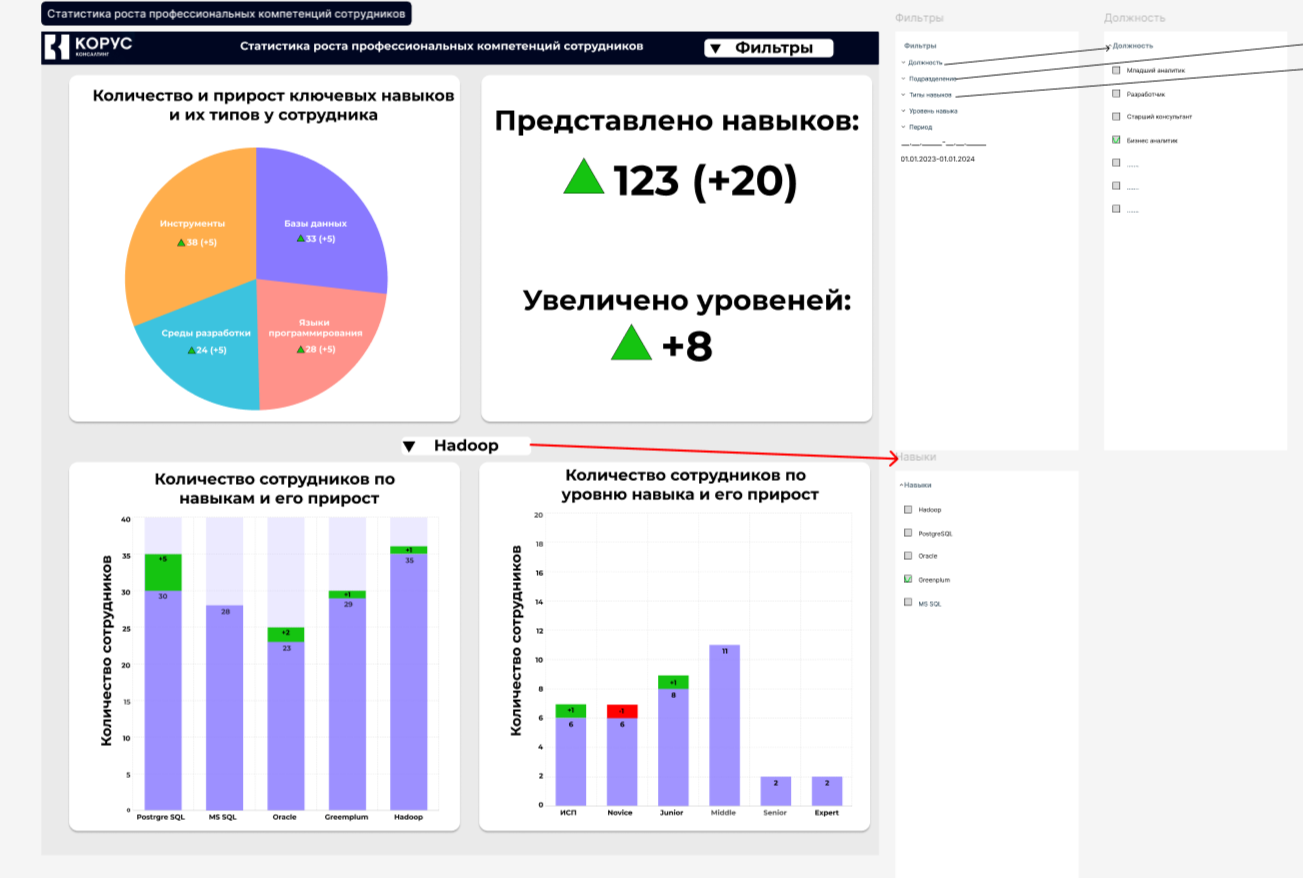
**Описание итоговой отчетности:**

Итоговая отчетность реализована с помощью дашбордов в LuxmsBI, построенных на основе данных из слоя «DM layer».

Страница статистики по сотруднику:



Страница статистики по подразделению/должности:



1. **Описание используемых сущностей с данными:**

|  |  |
| --- | --- |
| **Слой SOURCE/ODS** | **Слой DDS** |
| **Справочные таблицы** | |
| базы\_данных | d\_databases |
| инструменты | d\_tools |
| платформы | платформы d\_platforms |
| среды\_разработки | d\_ide |
| технологии | d\_technologies |
| фреймворки | d\_frameworks |
| типы\_систем | d\_system\_types |
| отрасли | d\_industries |
| предметная\_область | d\_subjects |
| языки | d\_languages |
| языки\_программирования | d\_programming\_languages |
| уровень\_образования | d\_education\_levels |
| уровни\_владения\_языками | d\_language\_proficiency\_levels |
| уровни\_знаний | d\_knowledge\_levels |
| уровни\_знаний\_в\_отрасли | d\_industry\_knowledge\_levels |
| уровни\_знаний\_в\_предметной\_области | d\_subject\_knowledge\_levels |
| **Таблицы сотруднических связей** | |
| базы\_данных\_и\_уровень\_знаний\_сотрудника | employee\_db\_skills |
| инструменты\_и\_уровень\_знаний\_сотрудника | employee\_tool\_skills |
| образование\_пользователей | employee \_education |
| опыт\_сотрудника\_в\_отраслях | employee\_industry\_experience |
| опыт\_сотрудника\_в\_предметных\_областях | employee\_subject\_experience |
| платформы\_и\_уровень\_знаний\_сотрудника | employee\_platform\_skills |
| сертификаты\_пользователей | employee\_certificates |
| среды\_разработки\_и\_уровень\_знаний\_сотрудника | employee\_ide\_skills |
| технологии\_и\_уровень\_знаний\_сотрудника | employee\_tech\_skills |
| типы\_систем\_и\_уровень\_знаний\_сотрудника | employee\_system\_skills |
| фреймворки\_и\_уровень\_знаний\_сотрудника | employee\_framework\_skills |
| языки\_пользователей | employee\_languages |
| языки\_программирования\_и\_уровень\_знаний\_сотрудника | employee\_programming\_skills |
| **Таблицы сотрудников и сводного резюме** | |
| сотрудники\_дар | employees |
| резюме | resumes |
| **Слой DM** | |
| employees (сотрудники департамента ДАР) | |
| skills (навыки) | |
| skills\_types (типы навыков) | |
| skills\_levels (уровни знаний навыков) | |
| departments (подразделения) | |
| depart\_key\_skills (ключевые навыки в подразделении) | |
| depart\_key\_skill\_types (ключевые типы навыков в подразделении) | |
| job\_titles (должности) | |
| job\_key\_skills (ключевые навыки в должности) | |
| job\_key\_skill\_types (ключевые типы навыков в должности) | |
| job\_key\_skill\_type\_skills (ключевые навыки в типах навыков в должности) | |
| resume (навыки сотрудников с датой присвоения) | |