## Разработка архитектуры решения

## Андронов Кирилл, команда 4

# 1 Схема архитектурного решения

**Источник Данных (Source Data)**:

Данные о сотрудниках хранятся в базе данных PostgreSQL. Этот источник содержит информацию о текущем уровне компетенций и изменениях в навыках сотрудников.

**Пайплайн Переноса Данных (Airflow Transfer Pipeline)**:

Используется Apache Airflow для оркестрации и автоматизации процессов переноса данных. Airflow управляет потоками данных из исходного хранилища в систему ETL.

Docker используется для контейнеризации процессов, что обеспечивает гибкость и простоту развертывания пайплайна.

**ETL Процесс (ETL\_DB\_4)**:

**ODS (Operational Data Store)**: Оперативное хранилище данных. Здесь данные очищаются и нормализуются перед дальнейшей обработкой.

**DDS (Data Distribution Service)**: Сервис распределения данных. В этой части данные подвергаются более глубокой трансформации, включающей агрегацию и обогащение.

**DM (Data Mart)**: Хранилище данных для конечного анализа. Обработанные данные загружаются сюда для использования в аналитике и визуализации.

**Визуализация**:

Инструмент Luxms BI используется для создания визуализаций и дашбордов. Он подключается к Data Mart и предоставляет визуальное представление прогресса навыков и технологий сотрудников, позволяя анализировать данные и принимать обоснованные решения.

Данные из исходного хранилища PostgreSQL проходят через пайплайн, оркестрируемый Apache Airflow. Пайплайн включает этапы очистки и трансформации данных в ETL (ETL\_DB\_4), состоящий из ODS, DDS и DM. После обработки данные поступают в Luxms BI, где они визуализируются на дашборде, позволяя отслеживать и анализировать изменения в навыках и технологиях сотрудников.

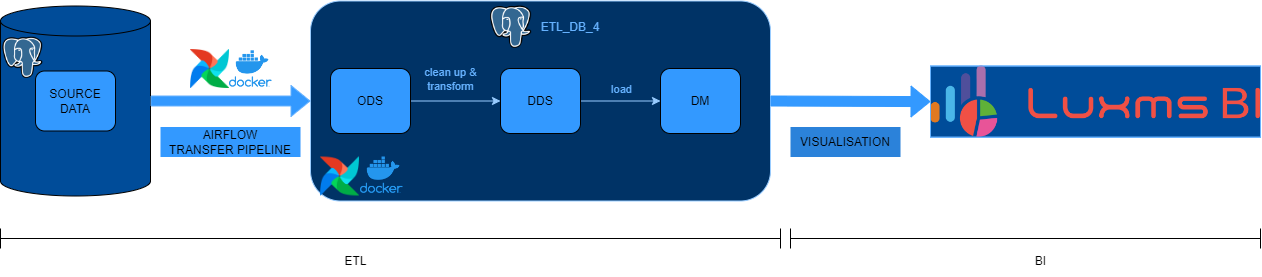


Рисунок 1 – Архитектура проекта

# 2 Описание компонентов архитектуры

1. **Источник Данных (Source Data)**

**PostgreSQL (версия 13.0)**:

Реляционная база данных для хранения исходных данных.

Обеспечивает надежное хранение и доступ к данным для последующего анализа.

2. **Пайплайн Переноса Данных (Airflow Transfer Pipeline)**

**Apache Airflow (версия 2.6.2)**:

Инструмент для оркестрации рабочих процессов и автоматизации ETL пайплайнов.

Управляет процессами извлечения данных из исходного хранилища и их передачи в ETL.

Airflow DAGs (Directed Acyclic Graphs) используются для определения и планирования задач.

**Docker (версия 23.0.3)**:

Платформа для контейнеризации, обеспечивающая изоляцию и легкость развертывания рабочих процессов Airflow.

Контейнеры позволяют масштабировать и легко управлять компонентами пайплайна.

3. **ETL Процесс (ETL\_DB\_4)**: **PostgreSQL (версия 13.0):**

**ODS (Operational Data Store)**:

Оперативное хранилище данных для первичной очистки и нормализации.

Данные из исходного хранилища передаются сюда для начальной обработки.

**DDS (Data Distribution Service)**:

Сервис распределения данных для более глубокой трансформации.

Включает агрегацию, обогащение и подготовку данных для конечного хранилища.

**DM (Data Mart)**:

Хранилище данных для конечного анализа.

Готовые к использованию данные загружаются сюда и используются для создания визуализаций.

4. **Визуализация**

**Luxms BI (версия 10.0)**:

Инструмент для создания аналитических дашбордов и визуализаций.

Подключается к Data Mart для извлечения обработанных данных.

Предоставляет интерфейс для визуального представления прогресса и изменений в навыках сотрудников.

# 3 Описание подключений

STG: база данных source, схема source\_data:

* Хост: 10.82.0.4
* Порт: 5432
* Пользователь: etl\_user\_4

DWH: База данных etl\_db\_4, схемы: ods, dds, dm:

* Хост: 10.82.0.4
* Порт: 5432
* Пользователь: etl\_user\_4

Airflow: оркестратор для пайплайна данных:

* Пользователь: airflow
* Пароль: airflow
* URL: localhost:8080
* Dags: ./dags, используется mount

Luxmsbi: \*будет дополняться\*

# 4 Описание сущностей

Слой ODS

|  |  |
| --- | --- |
| Сущность | Атрибуты |
| базы\_данных | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| базы\_данных\_и\_уровень\_знаний\_сотру | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| Базы данных |
| дата |
| Уровень знаний |
| инструменты | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| инструменты\_и\_уровень\_знаний\_сотр | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| дата |
| инструменты |
| Уровень знаний |
| образование\_пользователей | User ID |
| активность |
| Сорт. |
| Дата изм. |
| id |
| Уровень образование |
| Название учебного заведения |
| Фиктивное название |
| Факультет, кафедра |
| специальность |
| квалификация |
| Год окончания |
| опыт\_сотрудника\_в\_отраслях | User ID |
| активность |
| Сорт. |
| Дата изм. |
| id |
| дата |
| отрасли |
| Уровень знаний в отрасли |
| опыт\_сотрудника\_в\_предметных\_обла | User ID |
| активность |
| Сорт. |
| Дата изм. |
| id |
| дата |
| Предметные области |
| Уровень знаний в предметной облас |
| отрасли | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| платформы | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| платформы\_и\_уровень\_знаний\_сотруд | User ID |
| активность |
| Сорт. |
| Дата изм. |
| id |
| дата |
| платформы |
| Уровень знаний |
| предметная\_область | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| резюмедар | UserID |
| ResumeID |
| Активность |
| Образование |
| Сертификаты/Курсы |
| Языки |
| Базы данных |
| Инструменты |
| Отрасли |
| Платформы |
| Предметные области |
| Среды разработки |
| Типы систем |
| Фреймворки |
| Языки программирования |
| Технологии |
| сертификаты\_пользователей | User ID |
| активность |
| Сорт. |
| Дата изм. |
| id |
| Год сертификата |
| Наименование сертификата |
| Организация, выдавшая сертификат |
| сотрудники\_дар | id |
| Дата рождения |
| активность |
| пол |
| фамилия |
| имя |
| Последняя авторизация |
| должность |
| цфо |
| Дата регистрации |
| Дата изменения |
| подразделения |
| E-Mail |
| логин |
| компания |
| Город проживания |
| среды\_разработки | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| среды\_разработки\_и\_уровень\_знаний\_ | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| дата |
| Среды разработки |
| Уровень знаний |
| технологии | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| технологии\_и\_уровень\_знаний\_сотру | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| дата |
| технологии |
| Уровень знаний |
| типы\_систем | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| типы\_систем\_и\_уровень\_знаний\_сотру | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| дата |
| Типы систем |
| Уровень знаний |
| уровень\_образования | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| уровни\_владения\_ин | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| уровни\_знаний | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| название |
| уровни\_знаний\_в\_отрасли | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| уровни\_знаний\_в\_предметной\_област | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| фреймворки | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| фреймворки\_и\_уровень\_знаний\_сотру | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| дата |
| Уровень знаний |
| фреймворки |
| языки | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| языки\_пользователей | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| язык |
| Уровень знаний ин. языка |
| языки\_программирования | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| языки\_программирования\_и\_уровень | название |
| активность |
| Сорт. |
| Дата изм. |
| id |
| дата |
| Уровень знаний |
| Языки программирования |

Слой DDS:

\*Будет дополняться\*

Слой DM:

\*Будет дополняться\*