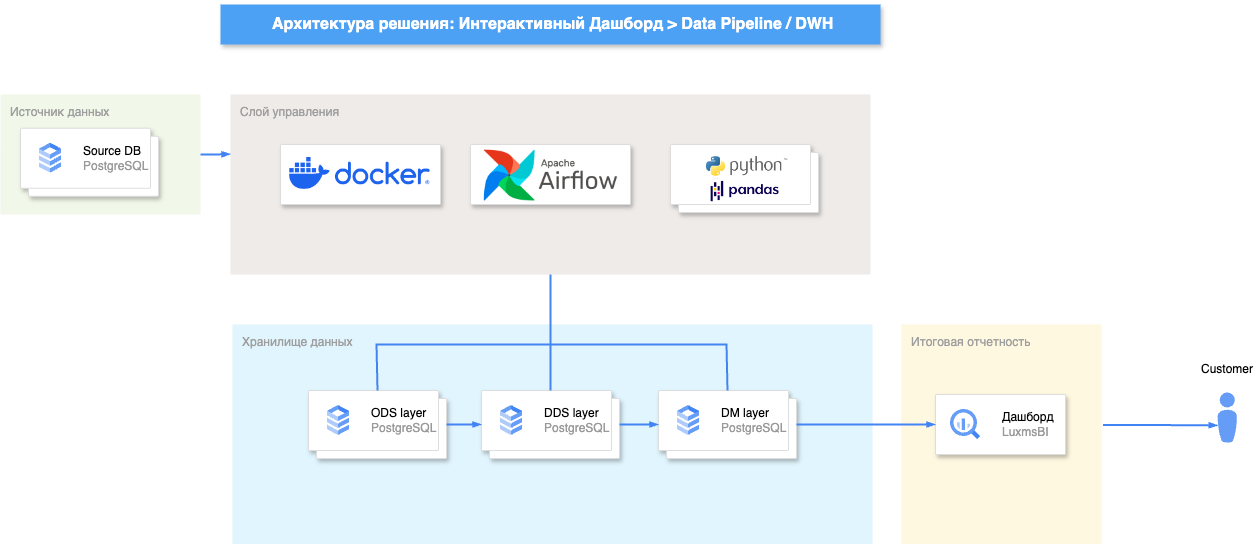
**Архитектура решения**

**«Интерактивный дашборд: анализ 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/data\_pipeline.py

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

**Описание инициализации схем:**

Инициализация схем слоев ODS, DDS и DM производится посредством выполнения DDL-скриптов из директории sql-scripts для создания необходимых схем и таблиц баз данных.

Слой ODS представляет собой временное хранилище данных перед их обработкой и трансформацией. 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:**

*# после написания DAGа будет добавлен граф из Airflow UI*

ETL-поток реализован DAGом Airflow data\_pipeline, который запускается с заданной частотностью (в зависимости от частоты обновления данных в source\_data) и состоящий из следующих taskов:

* check\_data\_changes — вызывает функцию отправки запроса к целевой базе данных для проверки наличия записей в таблице слоя ODS. Возвращает True, если таблица пуста, и False, если содержит данные. При отсутствии, выполнение переходит к задаче load\_ods\_data, в противном случае — к update\_ods\_data.

Загрузка данных (при их отсутствии в схемах):

* load\_ods\_data — запускает DML-скрипт load\_ods\_data.sql, который выполняет «as is» копирование данных из source\_data в слой «ODS layer» целевой базы данных.
* load\_dds\_data — запускает DML-скрипт load\_dds\_data.sql, который выполняет загрузку и обработку (нормализацию) данных из слоя ODS в слой DDS.
* load\_dm\_data — запускает DML-скрипт load\_dm\_data.sql, который производит необходимые агрегации данных из DDS слоя и сохраняет результирующие данные в DM слой.

Обновление данных:

* update\_ods\_data — запускает DML-скрипт update\_ods\_data.sql, который выполняет инкрементальное обновление данных (по столбцу last\_update) в слое ODS целевой базы данных.
* update\_dds\_data — запускает DML-скрипт update\_dds\_data.sql, который выполняет инкрементальное обновление данных в «DDS layer» целевой базы данных в соответствии с обновленными данными в слое ODS.
* update\_dds\_data — запускает DML-скрипты clear\_dm\_data.sql, который выполняет очистку данных из таблиц слоя DM, и load\_dm\_data.sql для повторной загрузки уже обновленных данных.

**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 | etl\_db\_3 |
| DDS layer | dds |
| DM layer | dm |

**Описание итоговой отчетности:**

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

*# позже будет добавлен макет в Figma*

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

|  |  |
| --- | --- |
| **Слой** | **Нейминг** |
| **Справочные таблицы** | |
| source/ODS | базы\_данных (d\_databases) |
| source/ODS | инструменты (d\_tools) |
| source/ODS | платформы (d\_platforms) |
| source/ODS | среды\_разработки (d\_development\_environments) |
| source/ODS | технологии (d\_technologies) |
| source/ODS | фреймворки (d\_frameworks) |
| source/ODS | типы\_систем (d\_system\_types) |
| source/ODS | отрасли (d\_industries) |
| source/ODS | предметная\_область (d\_subjects) |
| source/ODS | языки (d\_languages) |
| source/ODS | языки\_программирования (d\_programming\_languages) |
| source/ODS | уровень\_образования (d\_education\_levels) |
| source/ODS | уровни\_владения\_языками (d\_language\_proficiency\_levels) |
| source/ODS | уровни\_знаний (d\_knowledge\_levels) |
| source/ODS | уровни\_знаний\_в\_отрасли (d\_industry\_knowledge\_levels) |
| source/ODS | уровни\_знаний\_в\_предметной\_области (d\_subject\_knowledge\_levels) |
| **Таблицы сотруднических связей** | |
| source/ODS | базы\_данных\_и\_уровень\_знаний\_сотрудника (employee\_db\_skills) |
| source/ODS | инструменты\_и\_уровень\_знаний\_сотрудника (employee\_tool\_skills) |
| source/ODS | образование\_пользователей (employee \_education) |
| source/ODS | опыт\_сотрудника\_в\_отраслях (employee\_industry\_experience) |
| source/ODS | опыт\_сотрудника\_в\_предметных\_областях (employee\_subject\_experience) |
| source/ODS | платформы\_и\_уровень\_знаний\_сотрудника (employee\_platform\_skills) |
| source/ODS | сертификаты\_пользователей (employee \_certificates) |
| source/ODS | среды\_разработки\_и\_уровень\_знаний\_сотрудника (employee\_dev\_env\_skills) |
| source/ODS | технологии\_и\_уровень\_знаний\_сотрудника (employee\_tech\_skills) |
| source/ODS | типы\_систем\_и\_уровень\_знаний\_сотрудника (employee\_system\_skills) |
| source/ODS | фреймворки\_и\_уровень\_знаний\_сотрудника (employee\_framework\_skills) |
| source/ODS | языки\_пользователей (employee \_languages) |
| source/ODS | языки\_программирования\_и\_уровень\_знаний\_сотрудника (employee\_programming\_skills) |
| **Таблицы сотрудников и сводного резюме** | |
| source/ODS | сотрудники (employees) |
| source/ODS | резюме (resumes) |

*# после добавления DM слоя таблица будет обновлена соответствующими сущностями.*