

## **Архитектура решения**

### **Команда № 5**

**Ответственные за заполнение:**

Соколова Полина

Унучек Ксения

Департамент аналитических решений

Корус Консалтинг

г. Санкт-Петербург

2024

## 1. Схема архитектуры решения

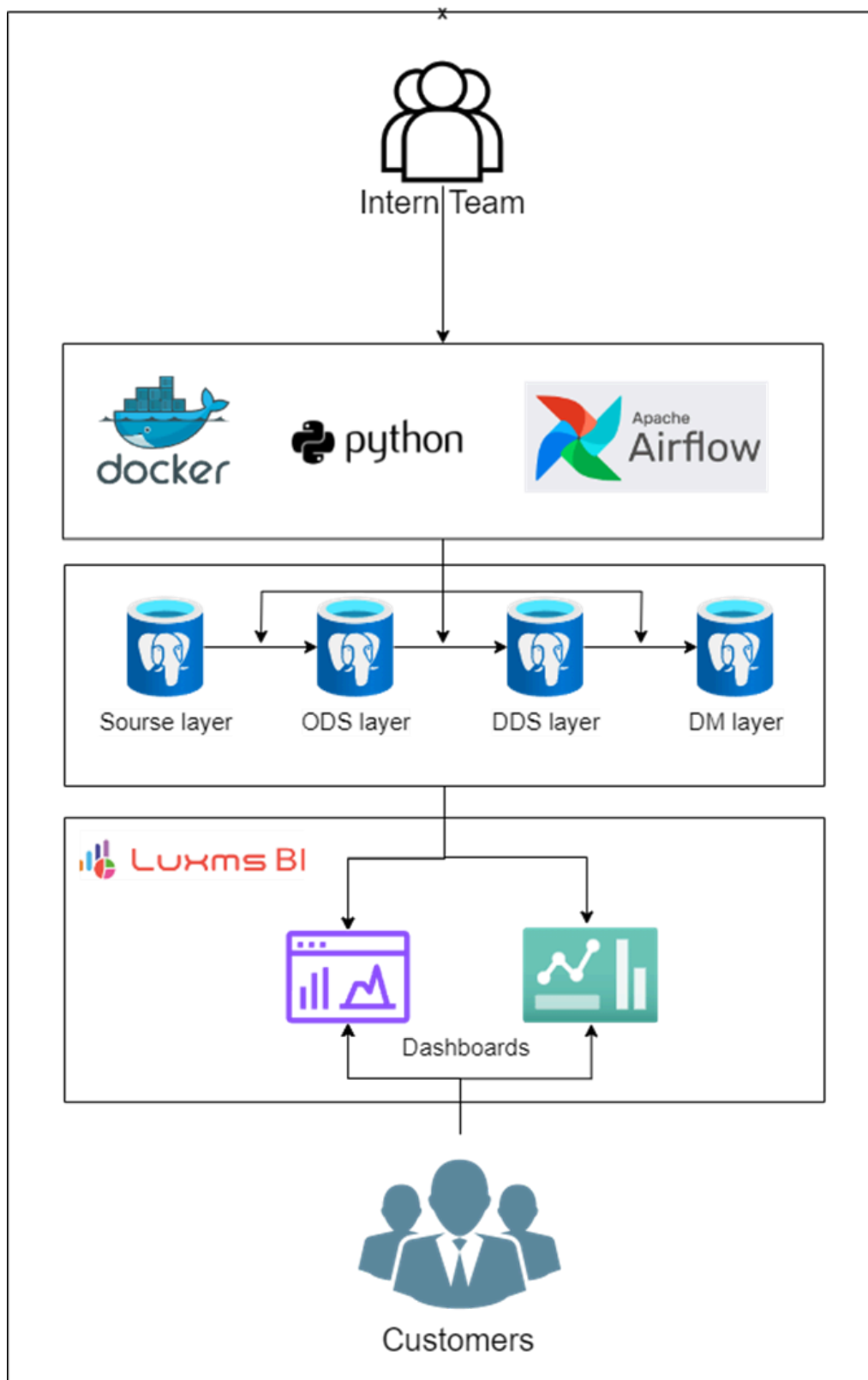


Рисунок 1. Схема архитектуры решения

## 2. Описание компонентов

### 2.1. Описание слоя управления

В Docker Desktop поднят контейнер Airflow. Airflow оркестрирует ETL-поток на базе SQL-скриптов.

Таблица 1. Интерфейс Airflow

<b>Вэб-интерфейс Airflow:</b>	http://localhost:8080
<b>User:</b>	airflow

К docker-контейнеру присоединена папка volumes/airflow/dags, где хранятся ДАГи и исполняемые скрипты.

Структура хранения ДАГов и исполняемых скриптов:

- Код основного ДАГа, который последовательно запускает все ДАГи проекта хранится в volumes/airflow/dags/entrypoint\_dag.py
- Коды ДАГов для формирования всех слоев хранятся:
  - volumes/airflow/dags/ods\_layer/ods\_layer\_transfer.py
  - volumes/airflow/dags/dds\_layer/dds\_layer\_transfer.py
  - volumes/airflow/dags/dm\_empl\_layer/dm\_empl\_layer\_transfer.py
  - volumes/airflow/dm\_dep\_layer/dm\_dep\_layer\_transfer.py
- Исполняемые скрипты хранятся в следующих папках:
  - volumes/airflow/dags/ods\_layer/sql
  - volumes/airflow/dags/dds\_layer/sql
  - volumes/airflow/dags/dm\_empl\_layer/sql
  - volumes/airflow/dags/dm\_dep\_layer/sql

## 2.2. ETL-поток

ETL-поток реализован с помощью ДАГа "entrypoint", который последовательно запускает ДАГи для формирования слоев потока.

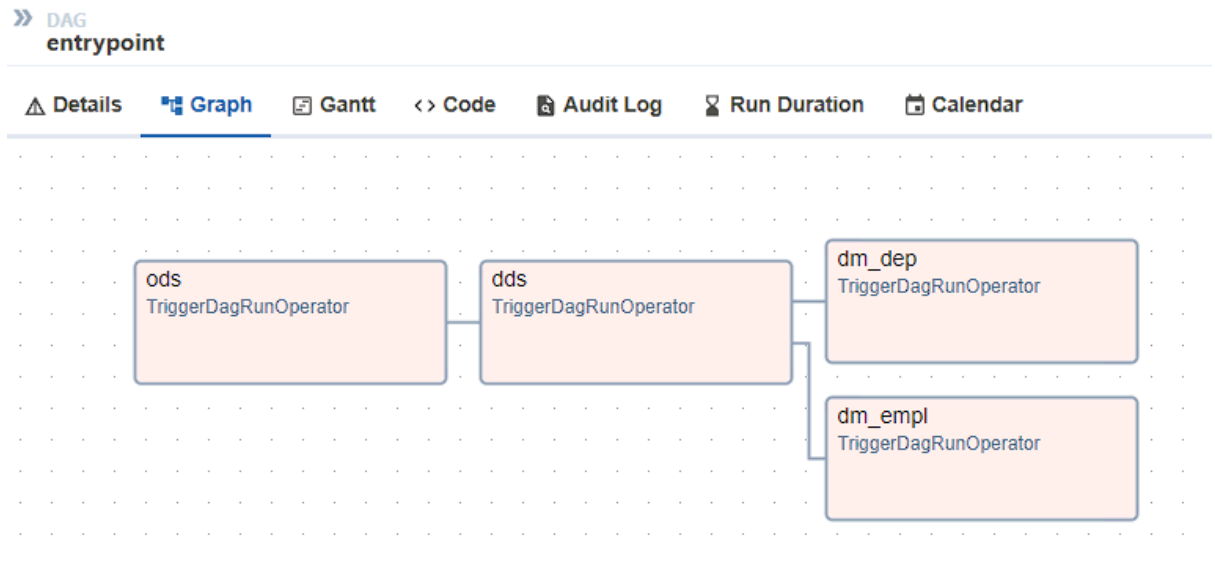


Рисунок 2. Граф ДАГа "entrypoint"

The screenshot shows the Airflow web interface with a table of DAGs. The table has columns for DAG name, Owner, Runs, Schedule, Last Run, Next Run, Recent Tasks, Actions, and Links. The DAGs listed are 'dds\_layer\_transfer', 'dm\_dep\_layer\_transfer', 'dm\_empl\_layer\_transfer', 'entrypoint', and 'ods\_layer\_transfer'. The 'entrypoint' DAG is highlighted, showing it is a daily scheduled task.

DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks	Actions	Links
dds_layer_transfer DDS   korus5	airflow	5	@once	2024-07-30, 21:05:00		1	[Play] [Stop]	...
dm_dep_layer_transfer Departments   DM   korus5	airflow	5	@once	2024-07-30, 21:05:11		2	[Play] [Stop]	...
dm_empl_layer_transfer DM   Employees   korus5	airflow	5	@once	2024-07-30, 21:05:11		3	[Play] [Stop]	...
entrypoint korus5	airflow	1	@daily	2024-07-29, 00:00:00	2024-07-30, 00:00:00	4	[Play] [Stop]	...
ods_layer_transfer korus5   ODS	airflow	4	@once	2024-07-30, 21:03:29		10	[Play] [Stop]	...

Рисунок 3. Вэб-интерфейс Airflow

ETL-поток реализован ДАГом Airflow "entrypoint", который последовательно с помощью TriggerDagRunOperator запускает следующие ДАГи:

- **ods\_layer\_transfer**, который состоит из следующих тасок:
  - **create\_schema** - создает схему "ods\_polina" в БД "etl\_db\_5", если её не существует (SQLExecuteQueryOperator)
  - **transfer\_data** - параллельно для каждой таблицы (GenericTransfer):
    - создает таблицу в БД "etl\_db\_5", если её не существует и очищает в ней данные
    - достает записи из таблицы в БД "source.source\_data" и добавляет их в БД "etl\_db\_5.ods\_polina"

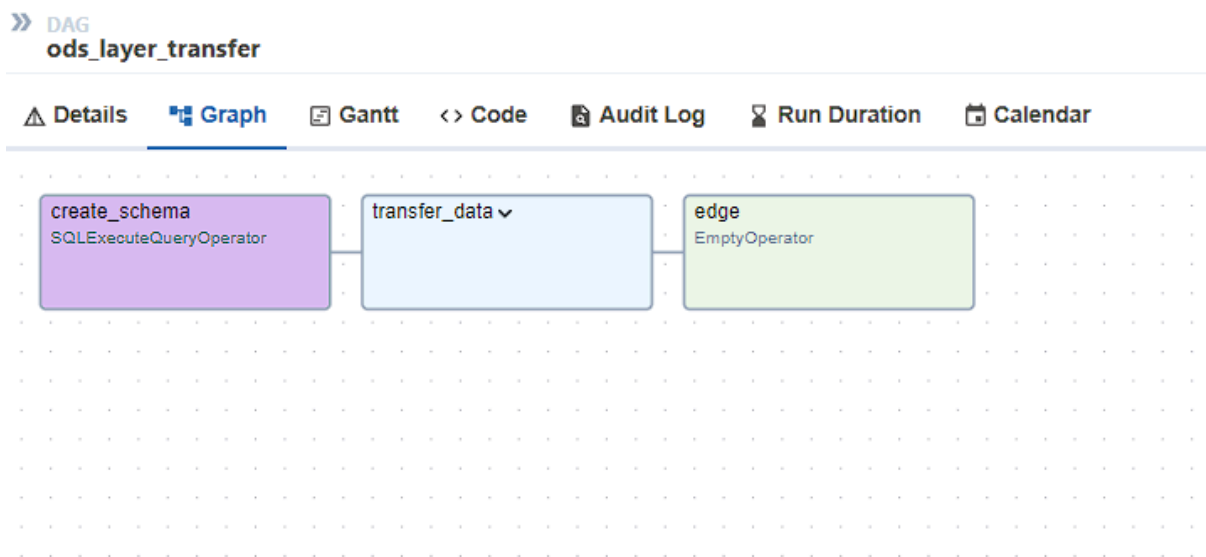


Рисунок 4. Граф ДАГа "ods\_layer\_transfer"

- **dds\_layer\_transfer**:
  - **create\_schema** - создает схему "ods\_polina" в БД "etl\_db\_5", если её не существует (SQLExecuteQueryOperator)
  - **transfer\_data** - последовательно (SQLExecuteQueryOperator):
    - создает таблицу в БД "etl\_db\_5", если её не существует и очищает в ней данные
    - достает записи из таблицы в БД "etl\_db\_5.ods\_polina", проводит необходимую подготовку (маппинг, очистку, фильтрацию) данных и добавляет их в БД "etl\_db\_5.dds\_polina"

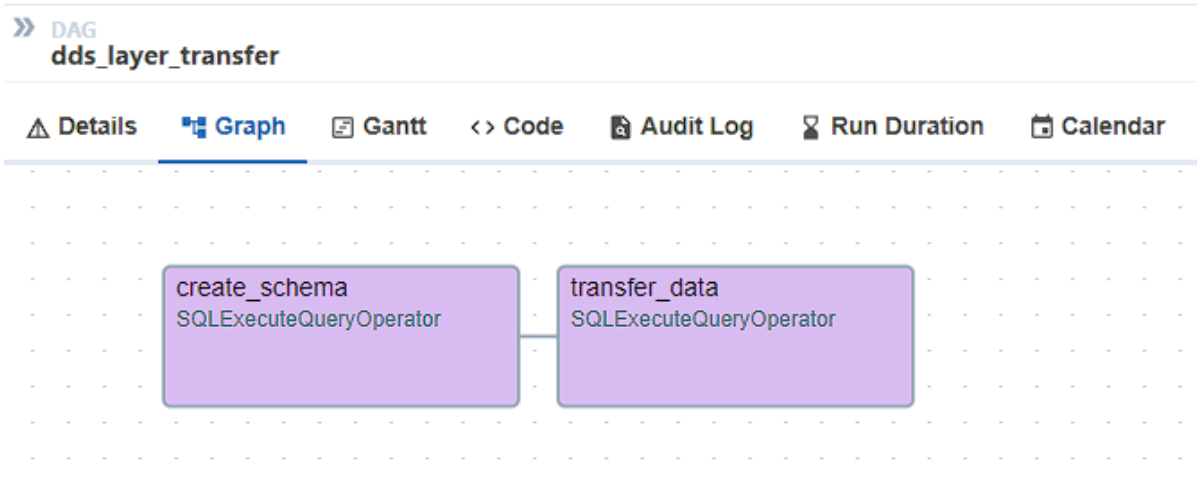


Рисунок 5. Граф ДАГа "dds\_layer\_transfer"

- **dm\_empl\_layer\_transfer:**

- **create\_schema** - создает схему "dm\_polina" в БД "etl\_db\_5", если её не существует (SQLExecuteQueryOperator):
- **transfer\_data** - последовательно (SQLExecuteQueryOperator):
  - создает таблицу в БД "etl\_db\_5", если её не существует и очищает в ней данные
  - достает записи из таблицы в БД "etl\_db\_5.dds\_polina", производит необходимые расчеты и агрегации и сохраняет результат в целевые таблицы слоя в БД "etl\_db\_5.dm\_polina"

- **dm\_dep\_layer\_transfer:**

- **create\_schema** - создает схему "ods\_polina" в БД "etl\_db\_5", если её не существует (SQLExecuteQueryOperator)
- **transfer\_data** - последовательно:
  - создает таблицу в БД "etl\_db\_5", если её не существует и очищает в ней данные
  - достает записи из таблицы в БД "etl\_db\_5.dds\_polina", производит необходимые расчеты и агрегации и сохраняет результат в целевые таблицы слоя в БД "etl\_db\_5.dm\_dep\_polina"

Графы ДАГов "dm\_empl\_layer\_transfer" и "dm\_dep\_layer\_transfer" идентичны графу "dds\_layer\_transfer" (рис. 5).

## 2.3. Хранилище данных

Хранилище данных реализовано с помощью СУБД PostgreSQL со следующей структурой:

Таблица 2. Структура Хранилища данных

Слой	Схема	БД	Параметры подключения к БД	Имя учетной записи
Source layer	source-data	source	host: 10.82.0.4  port: 5432	etl_user_5
ODS layer	ods_polina	etl_db_5		
DDS layer	dds_polina			
DM layer	dm_polina			
	dm dep polina			

## 2.4. Построение итоговой отчетности

Итоговая отчетность реализована с помощью Luxms BI с дашбордами, построенными на основе данных из схем “dm\_polina” и “dm\_dep\_polina”.

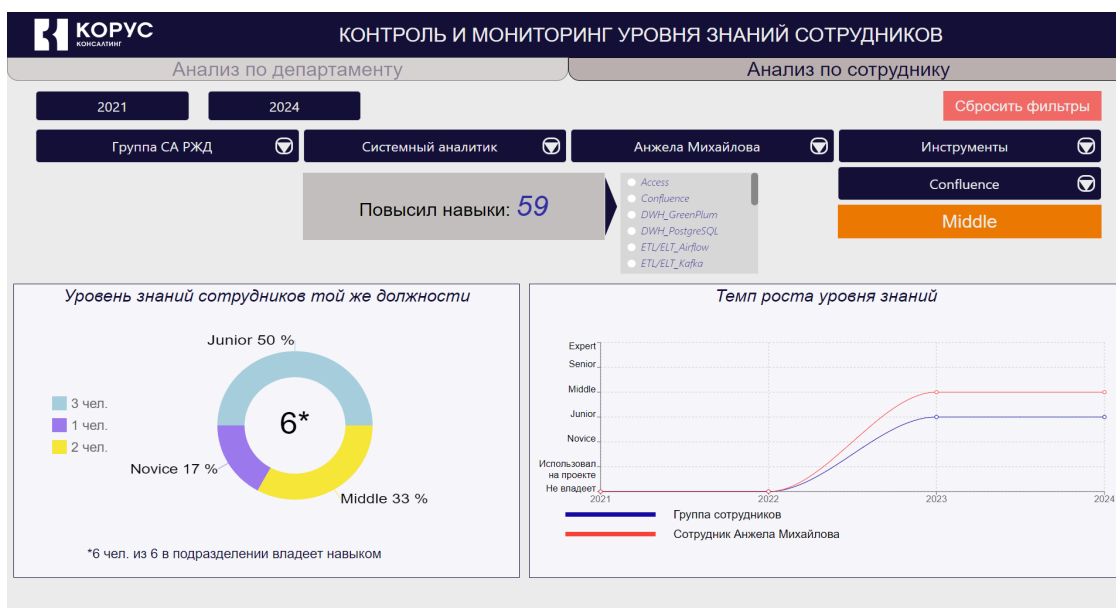


Рисунок 6. Анализ по сотруднику

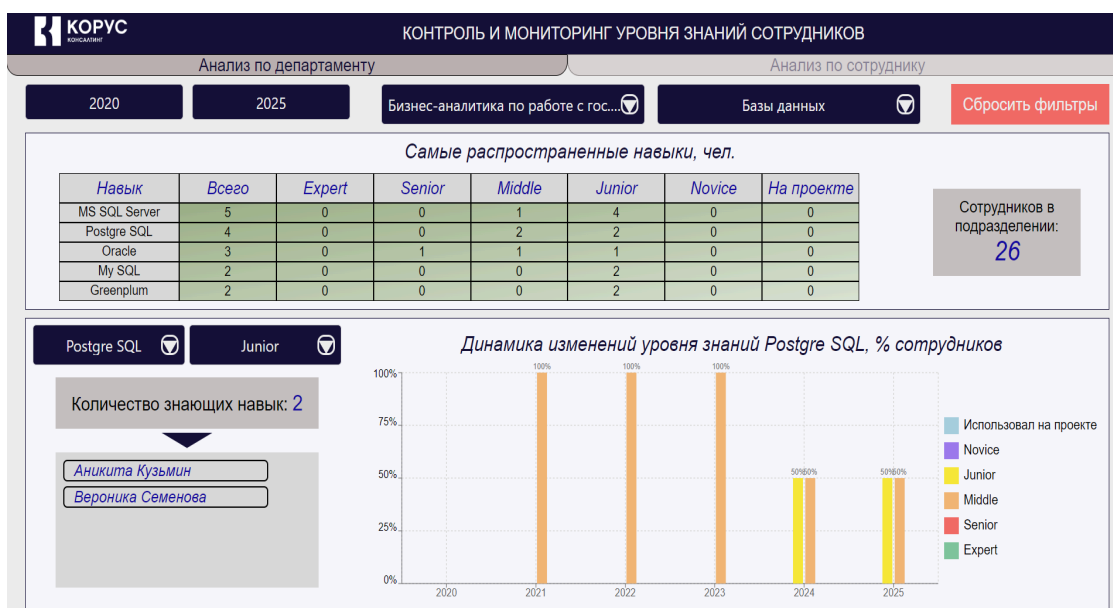


Рисунок 7. Анализ по департаменту

### 3. Описание используемых сущностей с данными

Таблица 2. Структура Хранилища данных

Слой	Сущность	Описание
Source layer, ODS layer	базы_данных	Справочник для используемых БД
Source layer, ODS layer	базы_данных_и_уровень_знаний_с_отру	Связь между БД и уровнями их знаний сотрудника
Source layer, ODS layer	инструменты	Справочник для используемых инструментов
Source layer, ODS layer	инструменты_и_уровень_знаний_с_отр	Связь между инструментам и уровнями их знаний сотрудника
Source layer, ODS layer	образование_пользователей	Информация о уровне образования, учебном заведении, факультете, кафедре, специальности, квалификации и году окончания
Source layer, ODS layer	опыт_сотрудника_в_отраслях	Информация об отрасли и о уровне знаний в отрасли сотрудника



Слой	Сущность	Описание
Source layer, ODS layer	опыт_сотрудника_в_предметных_обла	Информация о предметной области и о уровне знаний в предметной области сотрудника
Source layer, ODS layer	отрасли	Справочник для отраслей
Source layer, ODS layer	платформы	Справочник для платформ
Source layer, ODS layer	платформы_и_уровень_знаний_сотруд	Связь между платформами и уровням их знаний сотрудника
Source layer, ODS layer	предметная_область	Справочник для предметных областей
Source layer, ODS layer	резюмедар	Сводная таблица с информацией о сотрудниках и их образовании, сертификатах, языках, базах данных, инструментах, отраслях, платформах, предметных областях, средах разработки, типах систем, фреймворках, языках программирования, технологиях.
Source layer, ODS layer	сертификаты_пользователей	Информация о сертификатах сотрудников
Source layer, ODS layer	сотрудники_дар	Информация о сотрудниках ДАР, должности, цфо, подразделении
Source layer, ODS layer	среды_разработки	Справочник для сред разработки
Source layer, ODS layer	среды_разработки_и_уровень_знаний_	Связь между средами разработки и уровнями их знаний сотрудника

Слой	Сущность	Описание
Source layer, ODS layer	технологии	Справочник для технологий
Source layer, ODS layer	технологии_и_уровень_знаний_сот ру	Связь между технологиями и уровнями их знаний сотрудника
Source layer, ODS layer	типы_систем	Справочник для типов систем
Source layer, ODS layer	типы_систем_и_уровень_знаний_с отру	Связь между типами систем и уровнями их знаний сотрудника
Source layer, ODS layer	уровень_образования	Справочник для уровней образования
Source layer, ODS layer	уровни_владения_ин	Справочник для уровней владения иностранными языками
Source layer, ODS layer	уровни_знаний	Справочник для уровней знаний
Source layer, ODS layer	уровни_знаний_в_отрасли	Справочник для уровней знаний в отрасли
Source layer, ODS layer	уровни_знаний_в_предметной_обл аст	Справочник для уровней знаний в предметной области
Source layer, ODS layer	фреймворки	Справочник для фреймворков
Source layer, ODS layer	фреймворки_и_уровень_знаний_со тру	Связь между фреймворками и уровнями их знаний сотрудника
Source layer, ODS layer	языки	Справочник для иностраных языков
Source layer, ODS layer	языки_пользователей	Информация о языках и уровнях их знаний сотрудниками
Source layer, ODS layer	языки_программирования	Справочник для языков программирования

Слой	Сущность	Описание
Source layer, ODS layer	языки_программирования_и_уровень	Информация о языках программирования и уровнях их знаний сотрудниками
DDS layer	cv_dar	Информация об активных сотрудниках
DDS layer	departments	Справочник для отделов
DDS layer	employees	Информация о сотрудниках, их должностях и отделах
DDS layer	failed_entities	Таблица для строк с ошибками
DDS layer	levels	Справочник для уровней знаний
DDS layer	position	Справочник для должностей
DDS layer	skills	Справочник для всех навыков
DDS layer	skills_group	Справочник для объектов анализа
DDS layer	skills_levels	Информация о изменениях уровня навыков сотрудниками
DM layer (dm_polina)	departments	Справочник для департаментов
DM layer (dm_polina)	employees	Информация о сотрудниках, их должностях и отделах
DM layer (dm_polina)	employees_statistic	Витрина №1
DM layer (dm_polina)	levels	Справочник для уровней знаний
DM layer (dm_polina)	positions	Справочник для должностей
DM layer (dm_polina)	skill_groups	Справочник для объектов анализа

Слой	Сущность	Описание
DM layer (dm_polina)	skills	Справочник для всех навыков
DM layer (dm_dep_polina)	departments	Справочник для департаментов
DM layer (dm_dep_polina)	departments_statistic	Витрина №2
DM layer (dm_dep_polina)	employees	Информация о сотрудниках, их должностях и отделах
DM layer (dm_dep_polina)	levels	Справочник для уровней знаний
DM layer (dm_dep_polina)	skill_groups	Справочник для объектов анализа
DM layer (dm_dep_polina)	skills	Справочник для всех навыков