**АРХИТЕКТУРА РЕШЕНИЯ.**

Исполнитель: Асет Иманкулов

Должность: разработчик

Дата: 31.07.2023

Версия документа: 3.0

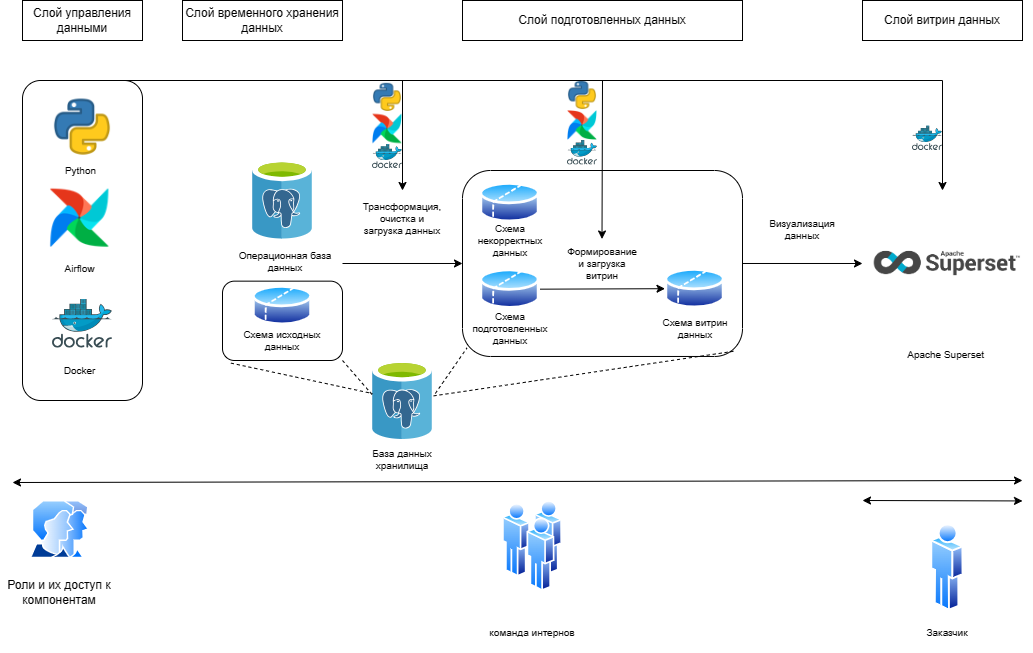
Оглавление

[1. СХЕМА АРХИТЕКТУРНОГО РЕШЕНИЯ. 3](#_Toc140263898)

[2. Описание компонентов. 4](#_Toc140263899)

[3. Описание слоев. 6](#_Toc140263900)

# 1. СХЕМА АРХИТЕКТУРНОГО РЕШЕНИЯ.



# 2. Описание компонентов.

2.1. Наименование баз данных и их элементов в информационной системе

|  |  |
| --- | --- |
| **База данных** | **Наименование** |
| Операционная база данных | internship\_sources |
| База данных хранилища | internship\_7\_db |

|  |  |  |
| --- | --- | --- |
| **База данных** | **Схема** | **Наименование** |
| Операционная база данных | Схема исходных данных | sources |
| База данных хранилища | Схема исходных данных | sources |
| Схема подготовленных данных | dds |
| Схема некорректных данных | data\_quality |
| Схема витрин данных | datamarts |

2.2. Версии используемого ПО.

|  |  |
| --- | --- |
| **Программное обеспечение** | **Версия программного обеспечения** |
| PostgreSQL | 12.10 |
| Python | 3.10 |
| Docker (клиент) | 24.0.2 |
| Docker (сервер) | 24.0.2 |
| Apache Airflow | 2.6.2 |
| Apache Superset | 2.1.0 |

2.3. Параметры подключения к компонентам.

Подключение к базам данных:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **База данных** | **Хост** | **Порт** | **Пользователь** | **Схема** |
| internship\_sources | 10.1.108.29 | 5432 | interns\_7 | sources |
| internship\_7\_db | 10.1.108.29 | 5432 | interns\_7 | sources |
| dds |
| data\_quality |
| datamarts |

Подключение к Apache Airflow:

|  |  |
| --- | --- |
| **Параметр** | **Значение** |
| URL вэб-интерфейса | http://localhost:8080/home |
| Логин | interns\_7 |
| Путь к ДАГам | ../docker/airflow/dags/ |
| Путь к исполняемому скрипту | ../docker/airflow/docker-compose.yml |

Подключение к Apache Superset:

|  |  |
| --- | --- |
| **Параметр** | **Значение** |
| URL вэб-интерфейса | http://localhost:9000/superset/welcome/ |
| Логин | interns\_7 |
| Путь к исполняемому скрипту | ../docker/superset/docker-compose-non-dev.yml |

2.4. Используемые сущности.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Слой** | **База данных** | **Схема** | **Сущность** | **Тип** |
| Слой временного хранения данных | internship\_sources | sources | brand | таблица |
| category | таблица |
| product | таблица |
| stock | таблица |
| transaction | таблица |
| internship\_7\_db | sources | product\_quantity | таблица |
| stores | таблица |
| stores\_emails | таблица |
| transaction\_stores | таблица |
| Слой подготовленных данных | internship\_7\_db | dds | brand | таблица |
| category | таблица |
| product | таблица |
| stock | таблица |
| transaction | таблица |
| product\_quantity | таблица |
| stores | таблица |
| stores\_emails | таблица |
| transaction\_stores | таблица |
| data\_quality | brand | таблица |
| category | таблица |
| product | таблица |
| stock | таблица |
| transaction | таблица |
| product\_quantity | таблица |
| stores | таблица |
| stores\_emails | таблица |
| transaction\_stores | таблица |
| upload\_tables\_tech | таблица |
| Слой витрин данных | internship\_7\_db | datamarts | orders\_data | таблица |
| stock\_data | таблица |
| stores\_data | таблица |
| orders\_data\_mart | представление |
| stock\_data\_mart | представление |
| stores\_data\_mart | представление |

# 3. Описание слоев.

Архитектурное решение представлено четырьмя слоями: слой управления данных, слой временного хранения данных, слой подготовленных данных, слой витрин данных.

В слое управления данным указаны технологии, осуществляющие обработку данных. Docker осуществляет контейнеризацию Airflow с Python. При этом, Airflow оркестрирует исполнение двух Python-скриптов, содержащих логику обработки данных, и двух python-скриптов, предназначенных для запуска DAG’ов. Первый скрипт предназначен для трансформации и очистки исходных данных и их последующую загрузку в схему подготовленных данных операционной базы данных. Данные, непрошедшие проверку качества попадают в таблицы логирования, размещенные на схеме некорректных данных.

Второй python-скрипт на основе полученных данных формирует и загружает витрины данных. Формирование витрин происходит в два этапа: сперва создаются таблицы соединения, а затем, на их основе, создаются представления. Образованные витрины данных визуализируются с помощью инструмента создания чартов и дашбордов Superset, который, в свою очередь, контейнеризирован с помощью Docker.

Слой временного хранения данных представлен операционной базой (internship\_sources), в которой хранятся исходные «сырые» данные и схемой «sources», размещенной на базе данных хранилища.

Слой подготовленных данных содержит базу данных хранилища (internship\_7\_db) с тремя схемами – схема подготовленных данных (dds), схема некорректных данных (data\_quality) и схема витрин данных (datamarts).