***Архитектура решения для проекта по мониторингу продаж и выручки и автоматизации закупок компании СтройТорг***

Исполнитель: Норов Парвиз Мироджевич

Должность: Стажер-разработчик

Дата: 30.07.2023

Команда: 10

Версия документа: 2

ГК «Корус Консалтинг»

**Оглавление**

РАЗДЕЛ 1. СХЕМА АРХИТЕКТУРНОГО РЕШЕНИЯ…………………………………………3

РАЗДЕЛ 2. ОПИСАНИЕ КОМПОНЕНТОВ……………………………………………………….5

РАЗДЕЛ 3. Описание слоев SOURCES, DDS, DATAMARTS…………………………….6

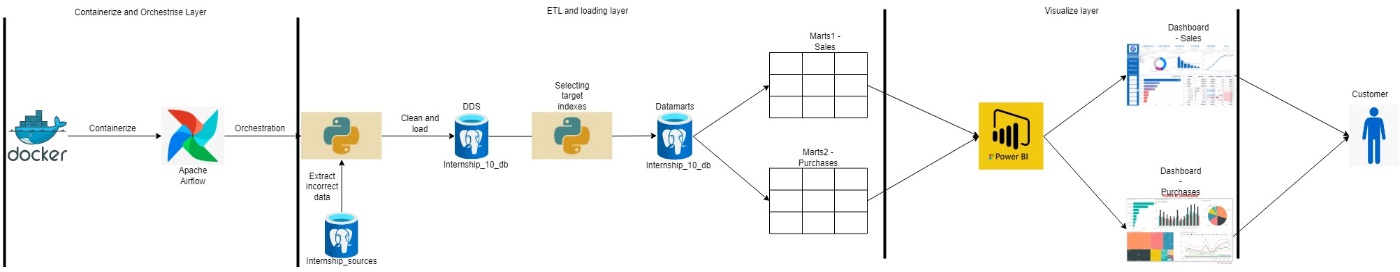


Рис 1

Описание схемы

Схема представленная на рисунке 1 представляет собой архитектуру решения по проекту «Мониторинг продаж и выручки и автоматизация закупок» для компании «СтройТорг»

Входе реализации проекта будет взаимодействие двух ролей: Команда состоящая из аналитика и разработчика, Заказчик. Для каждой из ролей определены компоненты к которым они имеют доступ:

1. **Аналитик и разработчик**

- Docker: Инструмент контейнеризации. Он позволяет упаковывать приложение в контейнер, который включает в себя все необходимые зависимости, такие как код, системные библиотеки, файлы конфигурации и другие ресурсы. Эти контейнеры могут быть легко перемещены и запущены на любой системе, поддерживающей Docker, с минимальными изменениями. В контейнере Docker разворачивается…

- Apache Airflow: Инструмент оркестрации и планирования выполнения задач. Оркестрация с помощью Apache Airflow - это процесс управления и автоматизации выполнения задач, процессов и рабочих нагрузок в распределенной среде. Airflow предоставляет инфраструктуру и инструменты для планирования, выполнения, мониторинга и управления ходом выполнения различных задач и рабочих процессов.

- Python скрипты: Разработчик использует Python скрипты для очистки и выгрузки данных в Корпоративное Хранилище Данных (далее КХД) internship\_10\_db в специальный слой DDS. А затем ведет расчет целевых показателей, которые формируют слой Datamarts

- Postgresql: В рамках проекта использует Система управления Реляционными Базами данных Postgresql, Базы данных (далее БД) Postgresql являются КХД

**2. Заказчик**

- Power BI. Это бизнес-аналитический инструмент, разработанный компанией Microsoft, который позволяет пользователям получать детализированный и интерактивный анализ данных. Он позволяет создавать дашборды, отчеты и визуализации, основанные на данных из различных источников, включая базы данных, электронные таблицы, онлайн-сервисы и другие.

Схема содержит в себе 3 логических блока:

1. **Слой контейнеризации и оркестрации –** В контейнере Docker разворачивается и запускается Apache Airflow
2. **ETL слой и хранения данных –** Очистка и выгрузка данных из БД заказчика в КХД в слой DDS, и расчет целевых показателей и выгрузка из в слой Datamarts
3. **Слой визуализации данных –** Построение интерактивных отчетов при помощи Power BI

**Раздел 2. Версии компонентов**

Таблица 1 – Версии используемого ПО (программное обеспечение)

|  |  |
| --- | --- |
| **Название ПО** | **Версия** |
| Docker | 24.0.2 |
| Apache Airflow | 2.6.3 |
| Python | 3.10 |
| Postgresql | 15.10 |
| Power BI | 2.119.986.0 |

Параметры подключения к компонентам

Таблица 2.1 Apache Airflow

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Url веб интерфейса | логин | пароль | Путь к Дагам | Путь к скриптам |
| http://localhost:8080 | airflow | airflow | /opt/airflow/dags | /opt/airflow/dags/scripts\_and\_files/script.py> |

Таблица 2.2 БД internship\_sources

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| host | port | БД | user | schema |
| 10.1.108.29 | 5432 | internship\_sources | Interns\_10 | sources |

Таблица 2.3 БД internship\_10\_db

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| host | port | БД | user | schemas |
| 10.1.108.29 | 5432 | internship\_10\_db | Interns\_10 | DDS, Datamarts |

**Раздел 3. Описание слое sources, dds, datamarts**

Таблица 3 Описание слоев

|  |  |  |
| --- | --- | --- |
| **sources** | **dds** | **datamarts** |
| Brand, category, product, stock, transaction | Brand, error\_brand, category, error\_category, product, error\_product, stores, error\_stores, stock, error\_stock, pos, error\_pos, transaction, error\_transaction | Marts1  Marts2 |

В ходе выполнения проекта, мы попросили заказчика выслать нам дополнительные таблицы stores и pos, однако их нет в БД заказчика internship\_sources. Разработчик вставил данные из csv файлов заказчика в КХД в слой dds