

Итоговая работа

Выполнил: Чипчагов М.С.

Проект № 4

Клиенты и счета

- Таблица клиентов
- Таблица счетов
- Операции по счетам
- Курсы валют по отношению к рублю

Необходимо сформировать три витрины на следующие даты:
2020-11-01, 2020-11-02, 2020-11-03, 2020-11-04.

Алгоритм обработки данных

1. Данных из таблиц импортируются в БД PostgreSQL и там хранятся.
2. Spark — приложение для создания витрин данных.
3. Для сокращения работы и объемов обрабатываемых данных в Spark, часть задач выполняется на уровне СУБД в виде материализованных представлений.

План реализации

1. Развернуть контейнеры для PostgreSQL и Spark (1 master и 2 worker`а).
2. Импортировать данные из csv-файлов в соответствующие таблицы PostgreSQL.
3. Создать материализованные представления, хранящие агрегированные данные по таблицам.
4. Создать приложение Spark, которое создаст необходимые витрины, извлекая данные из PostgreSQL.

Что сделано

- Созданы docker-compose.yml для развертывания контейнеров PostgreSQL и Spark.
- В PostgreSQL загружены данные из scv-файлов.
- Настроено взаимодействие Spark и PostgreSQL.

Что не успел :(

- Создать материализованные представления.
- Дописать приложение Spark для создания витрин данных.
-