Subscription Box Analytics – Курсовая документация

1. Описание бизнеса:  
Сервис предоставляет пользователям возможность оформить подписку на коробки, содержащие товары в соответствии с их интересами (например, книги, еда, косметика). Возможна одноразовая покупка или подписка на определённый период. Каждый пользователь может указывать свои интересы с указанием значемости этого интереса от 1 до 10. Коробки комплектуются товарами, доставляются пользователям, и сопровождаются оплатами, отзывами и статусами доставки.

2. Структура базы данных:

2.1. OLTP:  
Содержит следующие сущности:  
- Users (пользователи)  
- Interests (интересы)  
- Users\_Interests (мост между пользователями и интересами)  
- Boxes (коробки)  
- Items (товары)  
- Box\_Items (товары в коробке)  
- Subscriptions (подписки)  
- Payments (платежи)  
- Delivers (доставки)  
- Reviews (отзывы)

2.2. OLAP:  
- Измерения:  
 - dim\_Date  
 - dim\_Interest  
 - dim\_Customer (с SCD Type 2)  
 - dim\_Box  
- Факты:  
 - fact\_Subscription  
 - fact\_Delivery\_Performance  
- Связующая таблица:  
 - bridge\_Customer\_Interest

3. Для чего нужны OLTP и OLAP:

- OLTP - собирает все данные от бизнеса и используется для повседневных запросов

- OLAP – С помощью таблиц фактов можно узнать какие коробки самые популярные, как меняется активность во времени, насколько эффективно работает доставка.

5. Архитектура работы всех скриптов:  
- Используется промежуточный staging таблицы  
- Данные загружаются через команду COPY из CSV-файлов, находящихся в системной папке PostgreSQL.  
- После загрузки осуществляется пересылка из stg\_\* в основные OLTP-таблицы.  
- Из OLTP данные ETL-процессом перемещаются в OLAP-структуру.

7. Как запустить скрипты:

7.1. Как запустить OLTP

- Переносим все CSV файлы в системную папку Postgress(import) с помощью терминала

-Копируем весь скрипты из файлов 1.1-1.5 в PgAdmin запускаем последовательно каждый скрипт

Note:

- Файл 1.5, это queries для OLTP

7.2 .Как запустить OLAP:

-Копируем весь скрипты из файлов 1.1-1.4 в PgAdmin запускаем последовательно каждый скрипт

Note:

- Файл 1.4, это queries для OLTP

6. Инструменты:  
- PostgreSQL (pgAdmin)  
- Power BI Web  
- Excel (для хранения экспортированных CSV)

-Python(чтобы соединить все файлы CSV в один excel)  
- Terminal

7. Визуализация (Power BI):  
Созданы визуальные отчёты по:  
- Подпискам по времени  
- Объёму доставок и задержкам  
- Интересам пользователей  
- Рейтингам коробок  
Реализована настройка моделей данных, связи между таблицами и добавлены слайсеры для фильтрации по дате, интересу, полу.