**Проектная работа по модулю**

**“DWH”**

1. База booking была развернута на локальном сервере Postgeres ( в PDI connection booking)
2. В докере был скачан и запуще официальный контейнер Postgeres ( параметры запуска 12 версия Postgeres порт 49154 ), ( в PDI connection dds)
3. В базе докера были развернуты 3 схемы: dds, stage & rejected

Общее описание процессов загрузок измерений:

Измерение dim\_calendar было создано в SQL.

**Общее описание процессов загрузок измерений**

Алгоритм абсолютно одинаковый за исключением проверок – они чуть отличаются.

Хопы:

1. Чтение данных из сорсовой базы (booking)
2. Шаг Data Validation c проверками ( в основном на is not null)
3. При нахождении полей , которые не проходили проверку, строка писалась в таблицу схемы rejected.
4. Данные прошедшие валидацию писались в stage ( без ключей и записимостей) а затем грузились сразу в dds.

**Загрузка фактовой таблицы**

Для загрузки был сформирован запрос SQL с несколько избыточными данными для последующей подмены ключей.

Хопы:

1. Чтение данных из сорсовой таблицы
2. Поиск по потоку по измерению самолетов
3. Поиск по потоку по измерению аэропортов ( справочник использовался два раза : аэропорт вылета и прилета)
4. Поиск по потоку по измерению календарь ( использовался два раза: для прилета и отлета)
5. Поиск по потоку по измерению пассажиры
6. Поиск по потоку по измерению тарифы
7. Маппинг полей, все названия полей были приведены к таковым в таблицах базы dds
8. Были реализованы 4 проверки (насколько хватило фантазии) :  
   1) отфильтрованы полеты без пассажиров   
   2) проверка что дата вылета меньше даты прилета  
   3) задержка вылета больше или рана нулю - ни разу не встречал в реально жизни вылет раньше расписания  
   4) цена полета больше нуля и не пуста
9. Строки не прошедшие проверки уходили в rejected
10. Все что прошло проверку - > Stage -> dds