**Процедура ETL**

Для выполнения итоговой работы блока DWH был произведен процесс ETL, в котором из облачной базы «bookings» (host: 85.201.153.170; port: 19001; db: demo; user: netology; password: NetoSQL2019) в локальную базу (host: localhost; port: 5432; db: bookings\_dwh; user: postgres; password: \*\*\*\*\*\*\*) были загружены данные. Для этого на локальной базе были созданы таблицы справочников и фактов (скрипт прилагается).

**Справочник дат**. Данная таблица была создана и наполнена данными (даты сгенерированы) с помощью языка SQL в самой базе.

**Справочник пассажиров**. Данная таблица содержит информацию о пассажирах и их контактах. В скрипте создания данной таблицы была обеспечена проверка на соответствие вводимых значений в полях email и phone с помощью команды CHECK. Данной командой производится проверка на обязательное ввод данных в одно из полей (email и phone) в рамках одного кортежа.

**Справочник самолетов.** Данная таблица содержит информацию о всех самолетах, используемых для выполнения рейсов. С помощью команды CHECK производится проверка на превышение максимальной дальности полета всех самолетов.

**Справочник аэропортов**. Данная таблица содержит информацию о всех аэропортах, их координатах и временной зоне.

**Справочник тарифов**. Данная таблица содержит информацию о стоимости и классе обслуживания на каждый рейс.

**Таблица фактов**. Данная таблица содержит информацию о факте перелета каждого пассажира, месте и времени отправления, месте и времени прибытия, задержке, классе обслуживания и стоимости перелета. В таблице фактов помощью команды CHECK произведено три проверки качества данных, а именно: время отправления меньше времени прибытия, стоимость перелета должна быть больше значения «0» и класс обслуживания должен соответствовать одному из трех значений (Economy, Business или Comfort).

Для выгрузки данных из облачной базы и загрузки их в необходимом виде и формате был использован инструмент Pentaho Data Integration и шаги трансформации Table Input и Table Output.