## Общая структура

ETL процедуры разработаны с использованием Pentaho Data Integration. Преобразования включают несколько трансформаций (transformation), выполнение которых прописано в соответствующем задании (job).

Трансформации можно разделить на две группы:

* Создание/обновление измерений
* Заполнение таблиц данных

Создание таблиц измерений и таблицы фактов вынесено в отдельный SQL-скрипт.

Процедура выполняется на локальном инстансе PostgreSQL в базе данных «avia», схеме «bookings».

## Основное задание (job)

Основное задание представляет собой задачи по последовательному выполнению трансформаций в рамках ETL-процедуры

Первые 4 трансформации – заполнение и обновление таблиц измерений. При этом каждое измерение вынесено в отдельную трансформацию. Это упрощает мониторинг выполнения трансформаций и их редактирование.

Оставшиеся трансформации – формирование таблицы фактов. При этом формирование таблицы фактов осуществляется в 2 шага:

* Формирование данных о выполненных перелетах
* Формирование данных о пассажирах и билетах

## Трансформации измерений (dimension)

В рамках ETL-процесса используются следующие таблицы измерений:

* dim\_calendar
* dim\_passenger
* dim\_aircrafts
* dim\_airports
* dim\_tariff

### Calendar

### Passenger

### Aircrafts

### Airports

### Tariff

## Трансформации таблицы фактов (fact)

В связи с большим количеством записей о пассажирах и билетах, принято решение разделить формирование таблицы фактов на два шага.

На первом шаге формируется таблица с данными о фактически выполненных перелетах (с установкой связей на измерения).

На втором шаге формируется массив данных о пассажирах, классе обслуживания и стоимости перелета. Также на этом шаге производится объединение данных о перелетах и данных о пассажирах и билетах.

### Flights

### Tickets+Flights