

## Описание процедуры ETL

### Dim\_Airports - справочник аэропортов

Считываем данные из таблицы bookings.airports и после проверки качества данных записываем в таблицу dwh.dim\_airports.

Проверки качества данных:

airport_code	длина строки не менее 3х символов допустимы только заглавные буквы латинского алфавита
longitude	допустимы значения в диапазоне -180 до +180
latitude	допустимы значения в диапазоне -90 до +90
timezone	обязательно наличие «/», но не первым и не последним символами

Проверки значений полей на null не требуются, так как в исходных данных такое ограничение уже заложено.

Данные не прошедшие проверку записываются в таблицу rejected.airports

### Dim\_Aircrafts - справочник самолетов

Считываем данные из таблиц bookings.aircrafts и bookings.seats.

Данные из таблицы bookings.seats сортируем, группируем по полю aircraft\_code и считаем количество мест для каждого самолета. Получившийся результат объединяем с данными из таблицы bookings.aircrafts и после проверки качества данных записываем в таблицу dwh.Dim\_Aircrafts

Проверки качества данных:

aircraft_code	длина строки не менее 3х символов допустимы только заглавные буквы латинского алфавита и цифры
range	максимальное значение не больше именованной переменной \${range_max} (значение по умолчанию 25000)
seats_count	значение должно быть больше 0 максимальное значение не больше именованной переменной \${seats_max} (значение по умолчанию 1500)

Проверки значений полей на null не требуются, так как в исходных данных такое ограничение уже заложено.

Данные не прошедшие проверку записываются в таблицу rejected.aircrafts

### **Dim\_Passengers - справочник пассажиров**

Считываем данные из таблицы bookings.tickets, отбрасываем ненужные поля, разбиваем поле contact\_data на два поля phone и email. После проверки качества данных записываем в таблицу dwh.dim\_Passengers.

Проверки качества данных:

passenger_name	не менее двух слов, разделенных пробелом.
phone	is not null 12 символов, начинается с «+», оставшиеся символы только цифры
email	либо null, либо должен содержать символ «@». На реальных данных лучше использовать шаг трансформации «Mail Validator»

Проверки значений полей на null не требуются, так как в исходных данных такое ограничение уже заложено.

Данные не прошедшие проверку записываются в таблицу rejected.passengers

### **Dim\_Tariff - справочник тарифов**

Считываем данные из таблицы bookings.ticket\_flights, отбрасываем ненужные поля, сортируем по полю fare\_conditions и оставляем только уникальные значения данного поля. Результат записываем в таблицу dwh.Dim\_Tariff.

Проще таблицу public.Dim\_Tariff сгенерировать в SQL, так как в исходной таблице bookings.ticket\_flights уже есть ограничение CHECK (fare\_conditions IN ('Economy', 'Comfort', 'Business'))

```
insert into public.Dim_Tariff (fare_conditions) values  
(unnest(array['Economy', 'Comfort', 'Business']));
```

### Fact\_Flights - совершенные перелеты

Считываем данные из таблицы bookings.flights у которых status «Arrived» (перелет совершен) совместно с подсчетом суррогатного ключа даты вылета, объединяем с таблицами bookings.ticket\_flights и bookings.tickets. Присоединяем к потоку суррогатные ключи пассажира, самолета, аэропорт вылета и прилета, класса обслуживания. Считаем значения задержек вылета и прилета. Переименовываем поля.

После проверки качества данных записываем в таблицу dwh.Fact\_Flights.

Проверки качества данных:

scheduled_arrival — scheduled_departure	должно быть больше 0
actual_departure	не должно быть больше текущей даты
actual_arrival	должно быть не больше текущей даты
actual_arrival — actual_departure	должно быть больше 0
delay_departure	находится в диапазоне именованных переменных $\{\text{delay\_departure\_min}\}$ и $\{\text{delay\_departure\_max}\}$
delay_arrival	находится в диапазоне именованных переменных $\{\text{delay\_arrival\_min}\}$ и $\{\text{delay\_arrival\_max}\}$
amount	максимальное значение не больше именованной переменной $\{\text{amount\_max}\}$ (значение по умолчанию 1000000)

Проверки значений полей на null не требуются, так как в исходных данных такое ограничение уже заложено.