

Документация

к проектной работе по модулю

“DWH”

Нина Сперанская DWH-14

Для подключения к базе данных используются переменные окружения:

DB_NAME_AVIA	- название базы данных,
DB_HOST	- хост,
DB_PORT	- порт,
DB_USER	- имя пользователя,
DB_PASSWORD	- пароль.

dimensions.ktr

В трансформации dimensions.ktr данные из базы bookings переносятся в таблицы измерений passengers, airports, aircrafts и tariffs. Поскольку данные переносятся практически без изменений (за исключением dim.passengers, где контактные данные пассажира из json-формата разнесены по столбцам), то никаких проверок не проводится.

fact_flights.ktr

В трансформации fact_flights.ktr первым шагом объединяются данные из таблиц flights, ticket_flights и tickets базы данных bookings.

Затем к ним добавляется сегодняшняя дата.

После этого проверяется корректность актуальных даты и времени вылета и прибытия: они должны быть непустыми, вылет раньше прибытия и прибытие не позже чем сегодня. Некорректные строки записываются в таблицу public.rejected_flights с описанием ошибки в поле error.

Затем добавляется суррогатный ключ класса обслуживания из измерения tariffs (если тариф не найден, такая строка тоже записывается в public.rejected_flights с описанием ошибки в поле error).

Теперь к проверенным данным добавляются поля с задержкой вылета и прибытия в секундах, а также поле date_key для связи с измерением dim.calendar.

Наконец, данные записываются в таблицу fact.flights.