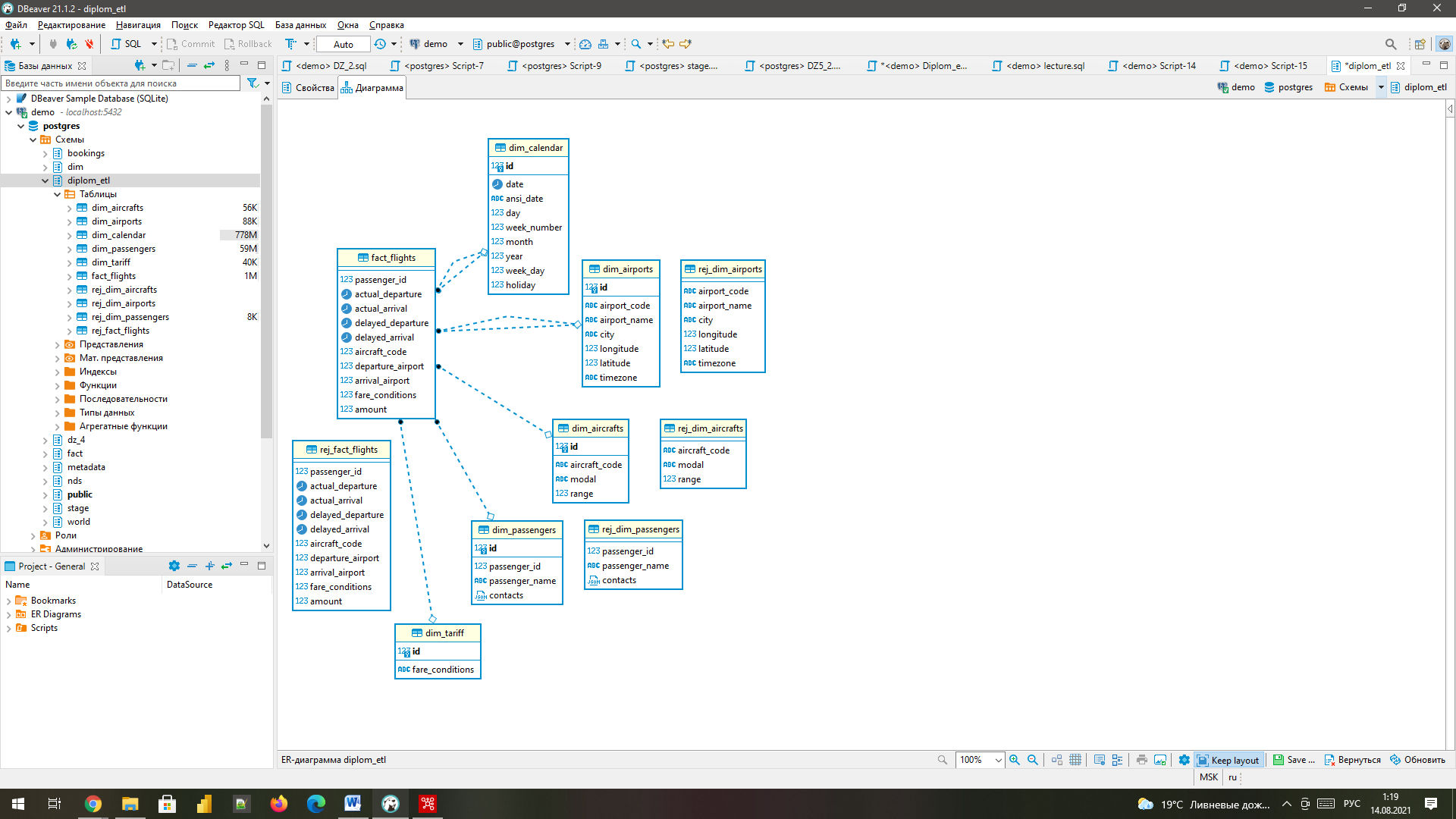
**Проектная работа по модулю “DWH”**

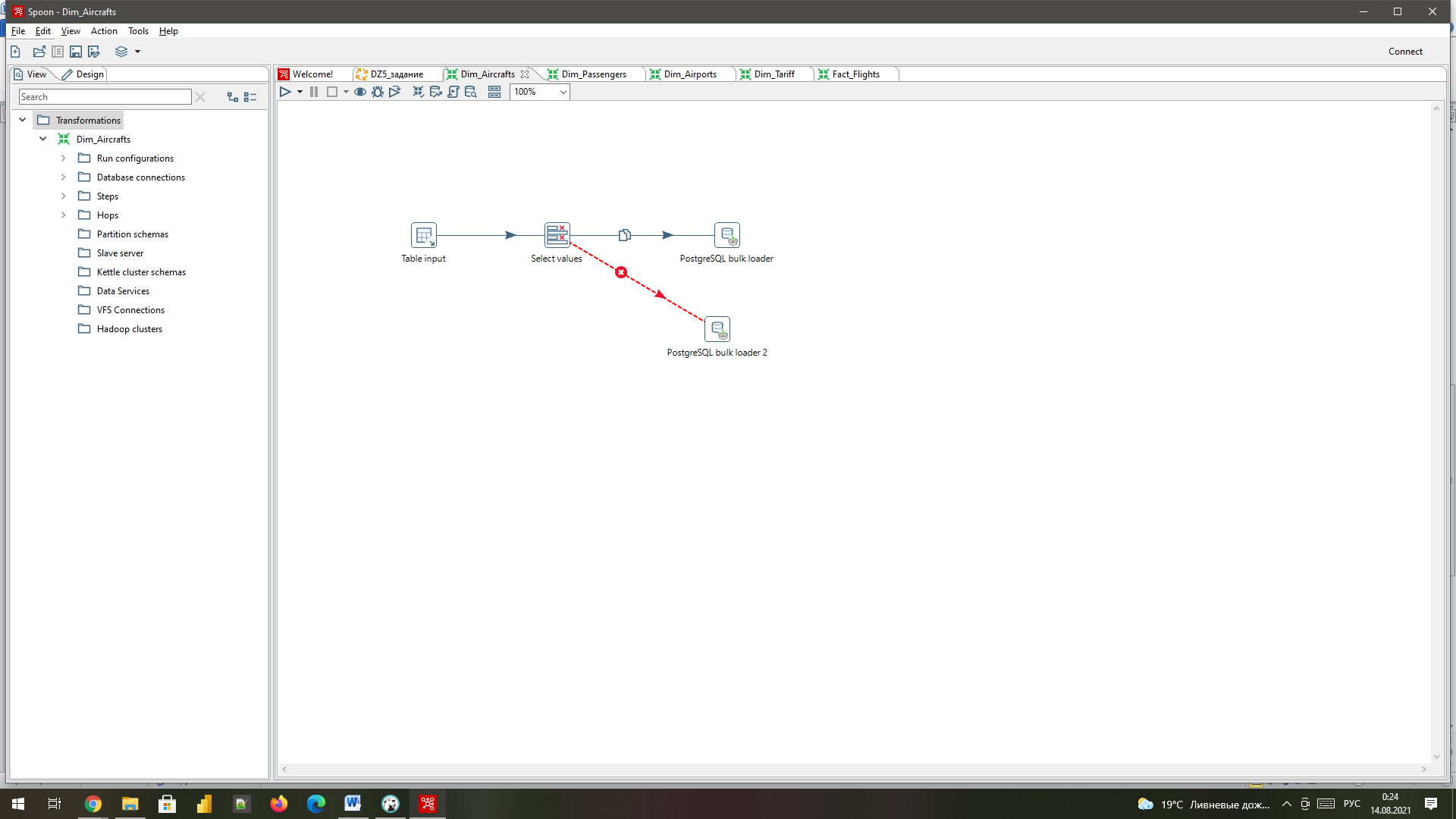
1. ER-диаграмма.



1. Dim\_Aircrafts – справочник самолетов.

Загружаем справочник из таблицы bookings.aircrafts, далее преобразовываем форматы данных и загружаем в таблицу Dim\_Aircrafts, некачественные строки записываются в таблицу Rej\_Dim\_Aircrafts.

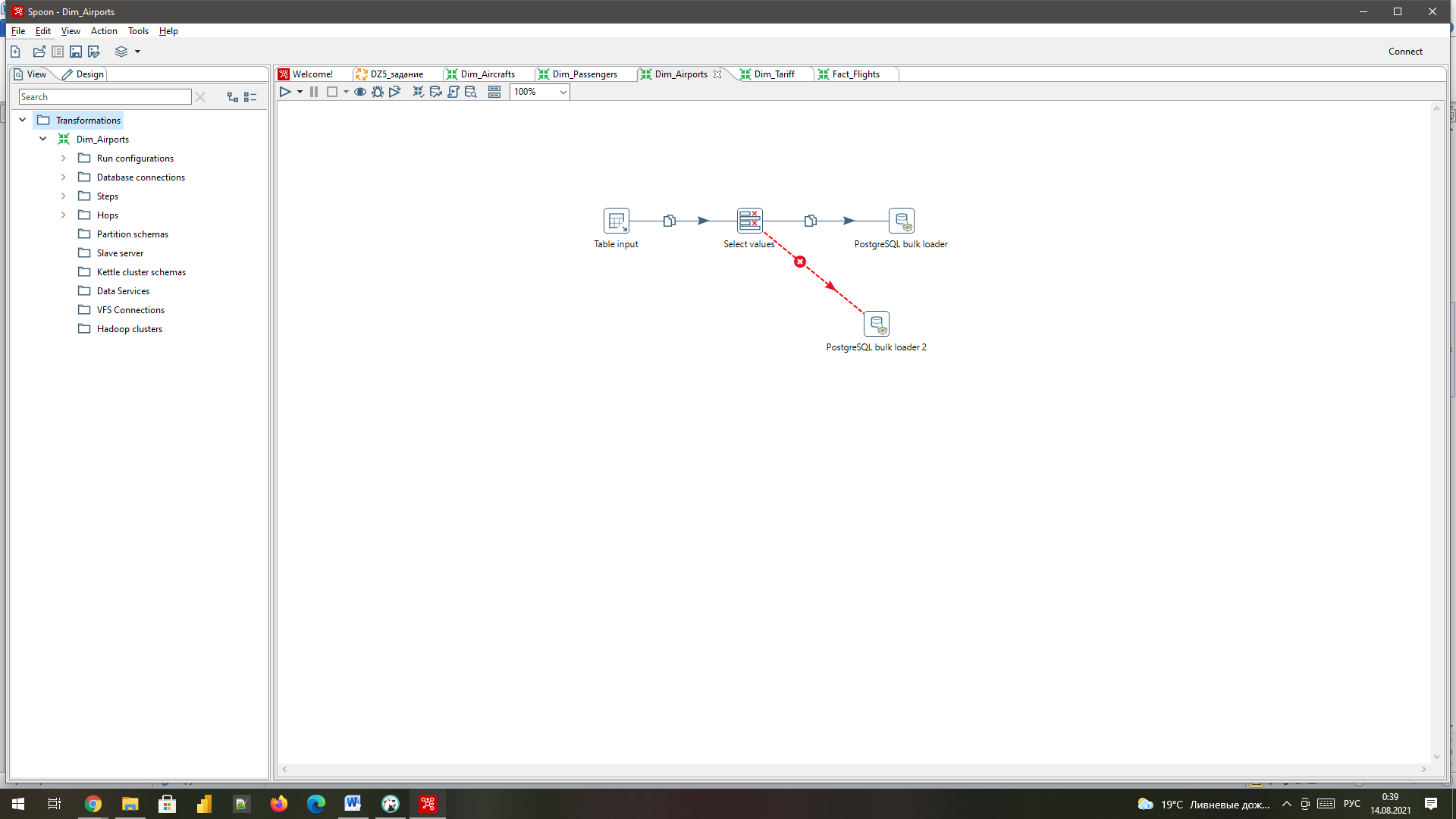
В таблице реализована проверка: код и модель самолета должны быть уникальными, дальность полета больше 0, все поля должны содержать данные.



1. Dim\_Airports – справочник аэропортов.

Загружаем справочник из таблицы bookings.airports, далее преобразовываем форматы данных и загружаем в таблицу Dim\_ Airports, некачественные строки записываются в таблицу Rej\_Dim\_ Airports.

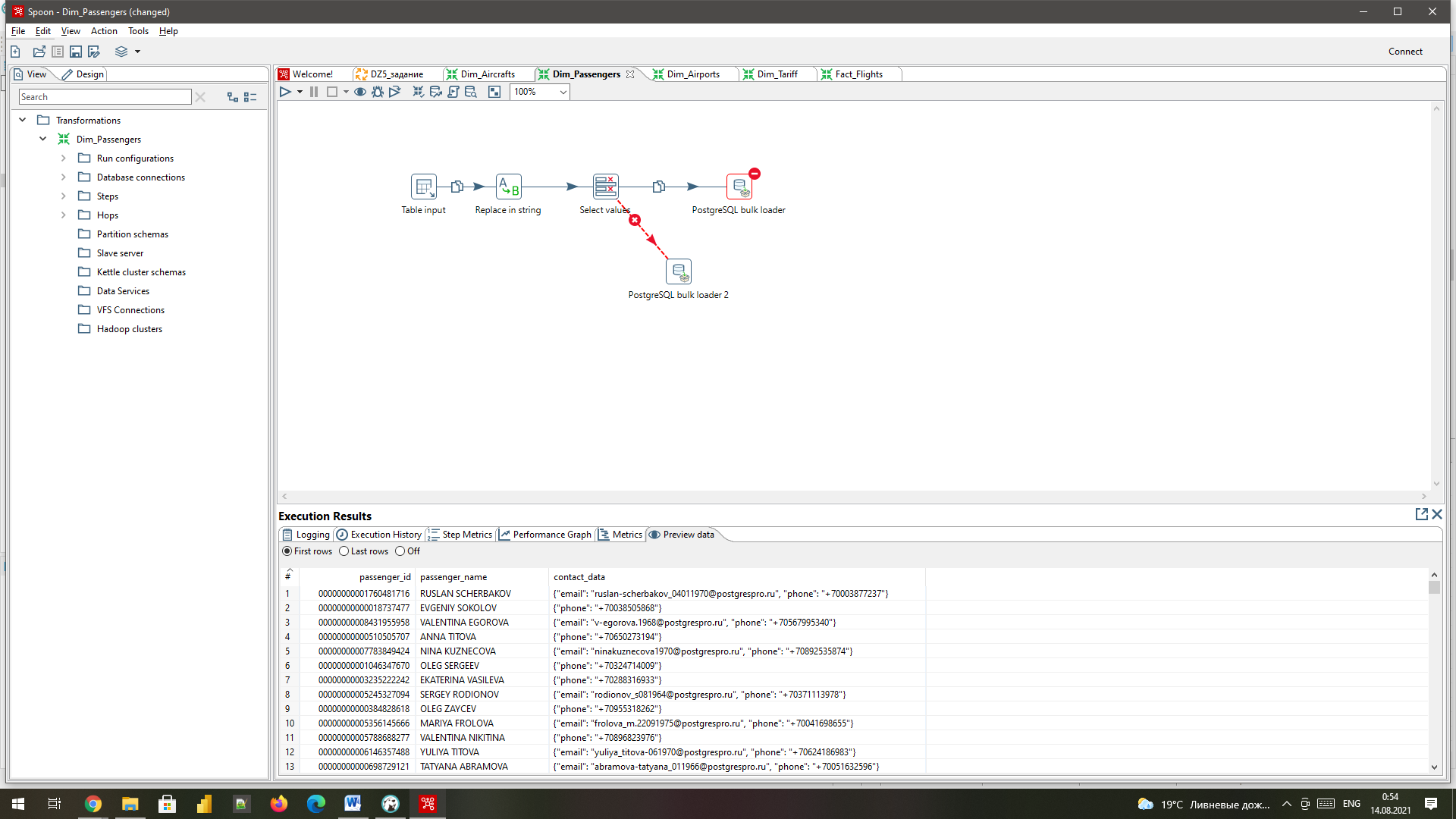
В таблице реализована проверка: код и название аэропорта должны быть уникальными, координаты аэропорта должны быть больше 0, все поля должны содержать данные.



1. Dim\_Passengers – справочник пассажиров.

Загружаем справочник из таблицы bookings.tickets, далее убираем в passenger\_id лишний пробел, преобразуем в число и загружаем в таблицу Dim\_ Passengers, некачественные строки записываются в таблицу Rej\_Dim\_ Passengers.

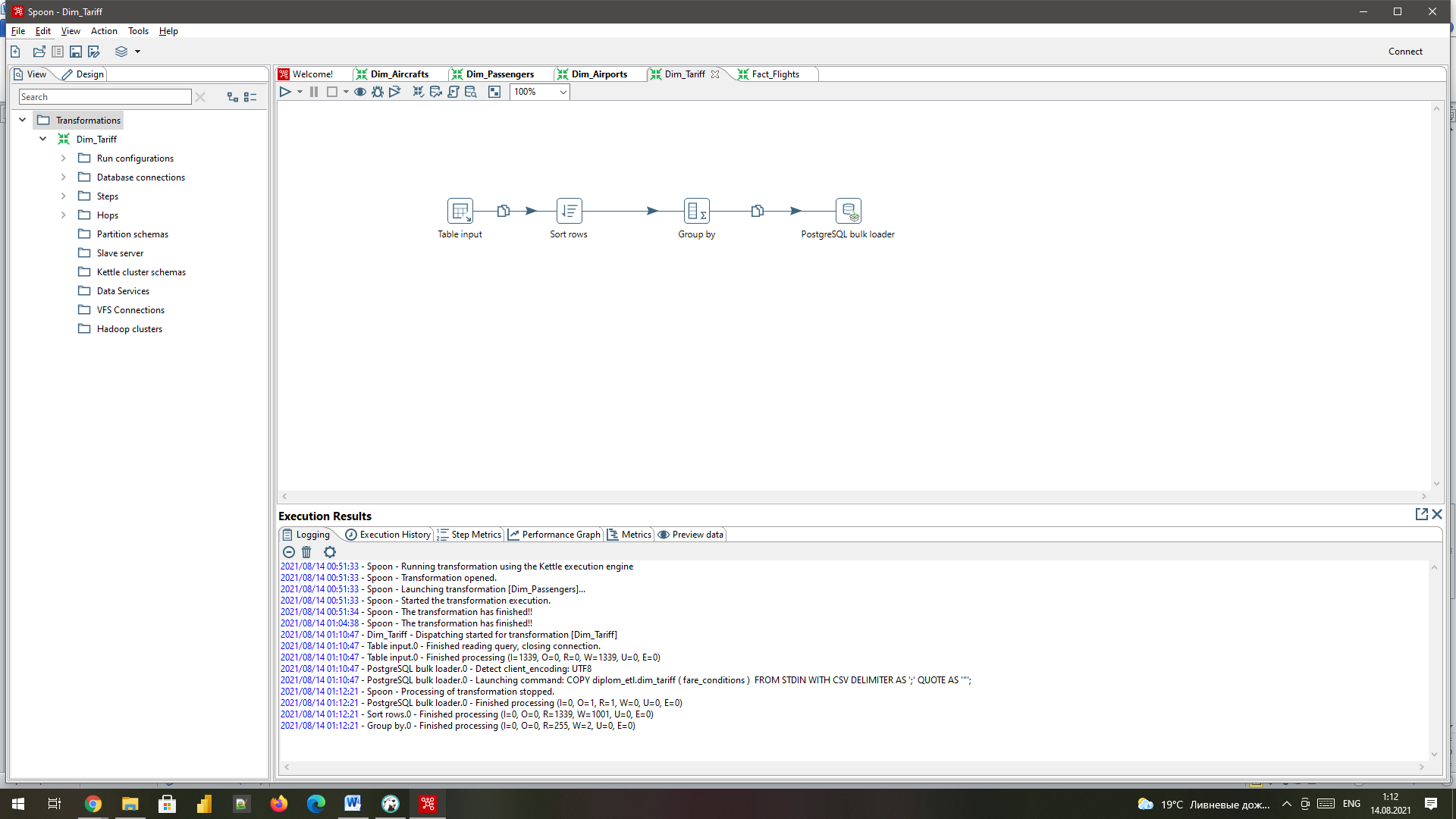
В таблице реализована проверка: идентификатор пассажира должен быть уникальным, а так же идентификатор и имя пассажира не должны быть пустыми.



1. Dim\_Tariff – справочник тарифов.

Загружаем справочник из таблицы bookings.seats, сортируем строки, группируем и загружаем в таблицу Dim\_ Tariff.

В таблице реализована проверка: класс обслуживания не должен быть пустым, должен быть уникальным и должен содержать одно из значений 'Economy', 'Comfort', 'Business'.



1. Fact\_Flights – совершенные перелеты.

Все, необходимы поля, достаем из таблиц Tickets, Ticket\_flights и Flights, объединяем таблицы по очереди с помощью шагов Stream loockup, далее фильтруем строки по полю «Status» - «Arrived», т.к. нужны только совершенные перелеты и убираем строки, в которых не указан пассажир, затем убираем лишние пробелы в passenger\_id и fare\_conditions. Далее нам нужно найти задержку вылета и прилета в секундах, для этого приводим форматы к Дате и с помощью шага калькулятор вычисляем задержку. Затем меняем форматы полей passenger\_id, delayed\_arrival, delayed\_departure на необходимые для загрузки в БД.

Далее для загрузки в БД нашу таблицу перелетов нужно привести к нормализованному виду, c помощью Stream loockup подтягиваем id атрибутов из ранее загруженных справочников, переименовываем столбцы и загружаем в таблицу Fact\_Flights.

