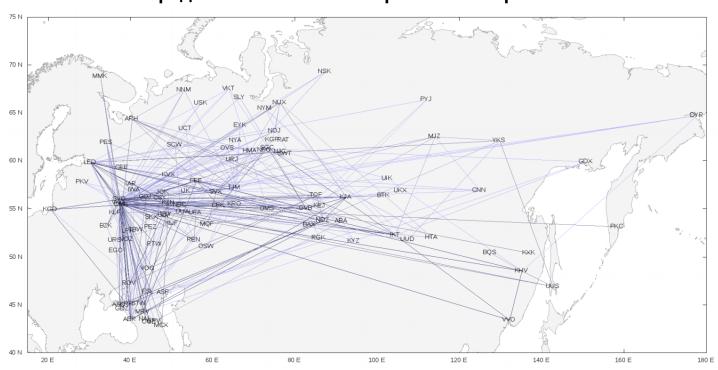


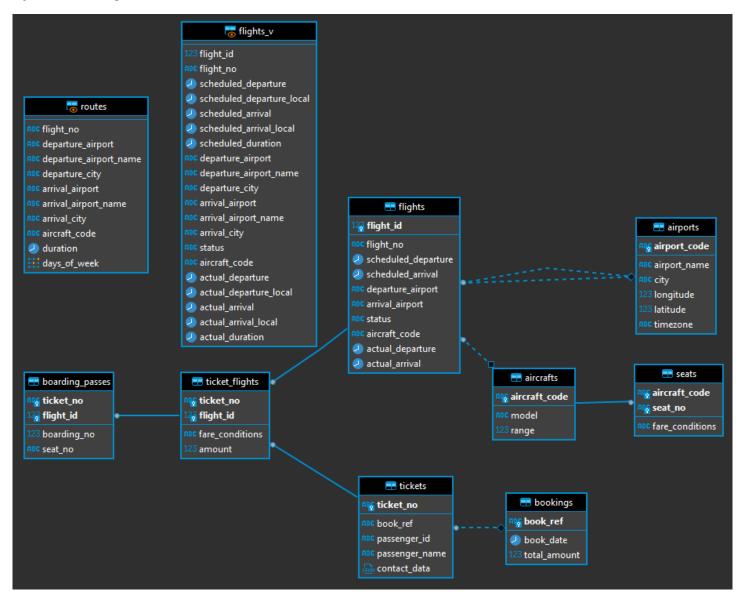
# SQL и получение данных

## В качестве предметной области выбраны авиаперевозки по России



## 1)База была развернута из \*.backup файла

## 2)ER-диаграмма



#### Описание схемы

Основной сущностью является бронирование (bookings).

В одно бронирование можно включить несколько пассажиров, каждому из которых выписывается отдельный билет (tickets).

Билет имеет уникальный номер и содержит информацию о пассажире. Как таковой пассажир не является отдельной сущностью. Как имя, так и номер документа пассажира могут меняться с течением времени, так что невозможно однозначно найти все билеты одного человека.

Билет включает один или несколько перелетов (ticket\_flights). Несколько перелетов могут включаться в билет в случаях, когда нет нет прямого рейса, соединяющего пункты отправления и назначения (полет с пересадками), либо когда билет взят «туда и обратно».

Каждый рейс (flights) следует из одного аэропорта (airports) в другой. Рейсы с одним

номером имеют одинаковые пункты вылета и назначения, но будут отличаться датой отправления.

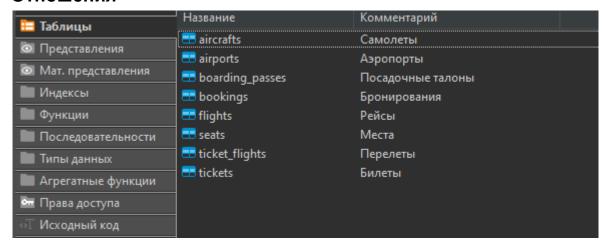
При регистрации на рейс пассажиру выдается посадочный талон (boarding\_passes), в котором указано место в самолете. Пассажир может зарегистрироваться только на тот рейс, который есть у него в билете. Комбинация рейса и места в самолете должна быть уникальной, чтобы не допустить выдачу двух посадочных талонов на одно место.

Количество мест (seats) в самолете и их распределение по классам обслуживания зависит от модели самолета (aircrafts), выполняющего рейс. Предполагается, что каждая модель самолета имеет только одну компоновку салона.

### 3)Краткое описание БД

Схема данных состоит из восьми таблиц и нескольких представлений.

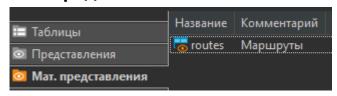
#### Отношения



#### Представления



#### Мат. представления



## 4)Развернутый анализ БД:

#### Описание таблиц

#### Самолеты bookings.aircrafts

Каждая модель воздушного судна идентифицируется своим трехзначным кодом (aircraft\_code). Указывается также название модели (model) и максимальная дальность полета в километрах (range).

```
Столбец
              | Тип | Модификаторы |
                                                         0писание
 aircraft_code | char(3) | NOT NULL | Код самолета, IATA
               | text | NOT NULL | Модель самолета
| integer | NOT NULL | Максимальная дал
model
 range
                                         Максимальная дальность полета, км
Индексы:
    PRIMARY KEY, btree (aircraft_code)
Ограничения-проверки:
    CHECK (range > 0)
Ссылки извне:
    TABLE "flights" FOREIGN KEY (aircraft_code)
        REFERENCES aircrafts(aircraft_code)
    TABLE "seats" FOREIGN KEY (aircraft_code)
        REFERENCES aircrafts(aircraft_code) ON DELETE CASCADE
```

#### Аэропорты bookings.airports

Аэропорт идентифицируется трехбуквенным кодом (airport\_code) и имеет свое имя (airport\_name).

Для города не предусмотрено отдельной сущности, но название (city) указывается и может служить для того, чтобы определить аэропорты одного города. Также указывается широта (longitude), долгота (latitude) и часовой пояс (timezone).

Столбец	Тип	Модификаторы	Описание	
airport_code airport_name city longitude latitude	char(3)   text   text   float   float	NOT NULL   NOT NULL   NOT NULL   NOT NULL   NOT NULL	Код аэропорта   Название аэропорта   Город   Координаты аэропорта: долгота   Координаты аэропорта: широта	
timezone	text	NOT NULL	Временная зона аэропорта	
Индексы:				
PRIMARY KEY, btree (airport_code)				
Ссылки извне:				
TABLE "flights" FOREIGN KEY (arrival_airport)				
REFERENCES airports(airport_code)				
TABLE "flights" FOREIGN KEY (departure_airport)				
REFERENCES airports(airport_code)				

#### Посадочные талоны bookings.boarding\_passes

При регистрации на рейс, которая возможна за сутки до плановой даты отправления, пассажиру выдается посадочный талон. Он идентифицируется также, как и перелет — номером билета и номером рейса. Посадочным талонам присваиваются последовательные номера (boarding\_no) в порядке регистрации пассажиров на рейс (этот номер будет уникальным только в пределах данного рейса). В посадочном талоне указывается номер места (seat\_no).

Столбец	Тип	Модификаторы	0писание	
ticket_no flight_id boarding_no seat_no	char(13)     integer     integer     varchar(4)	NOT NULL NOT NULL NOT NULL NOT NULL	Номер билета   Идентификатор рейса   Номер посадочного талона   Номер места	
Индексы:				
PRIMARY KEY, btree (ticket_no, flight_id)				
UNIQUE CONSTRAINT, btree (flight_id, boarding_no)				
<pre>UNIQUE CONSTRAINT, btree (flight_id, seat_no)</pre>				
Ограничения внешнего ключа:				
FOREIGN KEY (ticket_no, flight_id)				
<pre>REFERENCES ticket_flights(ticket_no, flight_id)</pre>				

#### Бронирования bookings.bookings

Пассажир заранее (book\_date, максимум за месяц до рейса) бронирует билет себе и, возможно, нескольким другим пассажирам. Бронирование идентифицируется номером (book\_ref, шестизначная комбинация букв и цифр).

Поле total\_amount хранит общую стоимость включенных в бронирование перелетов всех пассажиров.

Столбец	Тип	Модификаторы	Описание
total_amount Индексы:	char(6)   timestamptz   numeric(10,2)	NOT NULL	Номер бронирования   Дата бронирования   Полная сумма бронирования
Ссылки извне:	, , –	,	FERENCES bookings(book_ref)

#### Рейсы bookings.flights

Естественный ключ таблицы рейсов состоит из двух полей — номера рейса (flight\_no) и даты отправления (scheduled departure).

Чтобы сделать внешние ключи на эту таблицу компактнее, в качестве первичного используется суррогатный ключ (flight\_id). Рейс всегда соединяет две точки — аэропорты вылета (departure\_airport) и прибытия (arrival\_airport).

Такое понятие, как «рейс с пересадками» отсутствует: если из одного аэропорта до другого нет прямого рейса, в билет просто включаются несколько необходимых рейсов. У каждого рейса есть запланированные дата и время вылета (scheduled\_departure) и прибытия (scheduled\_arrival). Реальные время вылета (actual\_departure) и прибытия (actual\_arrival) могут отличаться: обычно не сильно, но иногда и на несколько часов, если рейс задержан.

Столбец	Тип	Модификаторы	Описание
scheduled_departure scheduled_arrival departure_airport arrival_airport status aircraft_code actual_departure actual_arrival Индексы: PRIMARY KEY, btree UNIQUE CONSTRAINT, Oграничения-проверки: CHECK (scheduled_a CHECK ((actual_arri OR ((actual_de	l   timestamptz   NOT NULL   Время прилёта по расписанию t   char(3)   NOT NULL   Аэропорт отправления   char(3)   NOT NULL   Аэропорт прибытия   varchar(20)   NOT NULL   Статус рейса   char(3)   NOT NULL   Код самолета, ІАТА   timestamptz   Фактическое время вылета   timestamptz   Фактическое время прилёта   оtree (flight_id)   AINT, btree (flight_no, scheduled_departure)		
Ограничения внешнего ключа:     FOREIGN KEY (aircraft_code)         REFERENCES aircrafts(aircraft_code)     FOREIGN KEY (arrival_airport)         REFERENCES airports(airport_code)     FOREIGN KEY (departure_airport)         REFERENCES airports(airport_code)  Ссылки извне:     TABLE "ticket_flights" FOREIGN KEY (flight_id)         REFERENCES flights(flight_id)			

#### Mеста bookings.seats

Места определяют схему салона каждой модели. Каждое место определяется своим номером (seat\_no) и имеет закрепленный за ним класс обслуживания (fare\_conditions) — Economy, Comfort или Business.

```
Столбец | Тип | Модификаторы | Описание

aircraft_code | char(3) | NOT NULL | Код самолета, IATA
seat_no | varchar(4) | NOT NULL | Номер места
fare_conditions | varchar(10) | NOT NULL | Класс обслуживания
Индексы:
    PRIMARY KEY, btree (aircraft_code, seat_no)
Ограничения-проверки:
    CHECK (fare_conditions IN ('Economy', 'Comfort', 'Business'))
Ограничения внешнего ключа:
    FOREIGN KEY (aircraft_code)
    REFERENCES aircrafts(aircraft_code) ON DELETE CASCADE
```

#### Таблица связей bookings.ticket\_flights

Перелет соединяет билет с рейсом и идентифицируется их номерами. Для каждого перелета указываются его стоимость (amount) и класс обслуживания (fare conditions).

```
Столбец
                         Тип
                               | Модификаторы |
                                                          0писание
                | char(13)
| integer
 ticket_no
                                NOT NULL
                                                 | Номер билета
flight_id
                                 NOT NULL
                                                  | Идентификатор рейса
fare_conditions | varchar(10) | NOT NULL | Класс обслуживания amount | numeric(10,2) | NOT NULL | Стоимость перелета
Индексы:
    PRIMARY KEY, btree (ticket_no, flight_id)
Ограничения-проверки:
    CHECK (amount \geq 0)
    CHECK (fare_conditions IN ('Economy', 'Comfort', 'Business'))
Ограничения внешнего ключа:
    FOREIGN KEY (flight_id) REFERENCES flights(flight_id)
    FOREIGN KEY (ticket_no) REFERENCES tickets(ticket_no)
Ссылки извне:
    TABLE "boarding_passes" FOREIGN KEY (ticket_no, flight_id)
        REFERENCES ticket_flights(ticket_no, flight_id)
```

#### Билеты bookings.tickets

Билет имеет уникальный номер (ticket no), состоящий из 13 цифр.

Билет содержит идентификатор пассажира (passenger\_id) — номер документа, удостоверяющего личность, — его фамилию и имя (passenger\_name) и контактную информацию (contact\_date).

Ни идентификатор пассажира, ни имя не являются постоянными (можно поменять паспорт, можно сменить фамилию), поэтому однозначно найти все билеты одного и того же пассажира невозможно.

Столбец	Тип	Модификаторы	Описание	
ticket_no book_ref passenger_id passenger_name contact_data	char(13)   char(6)   varchar(20)   text   jsonb	NOT NULL   NOT NULL   NOT NULL   NOT NULL 	Номер билета   Номер бронирования   Идентификатор пассажира   Имя пассажира   Контактные данные пассажира	
Индексы: PRIMARY KEY, btree (ticket_no)				
Ограничения внешнего ключа: FOREIGN KEY (book_ref) REFERENCES bookings(book_ref)				
Ссылки извне:  TABLE "ticket_flights" FOREIGN KEY (ticket_no) REFERENCES tickets(ticket_no)				

#### Представление "bookings.flights\_v"

Над таблицей flights создано представление flights\_v, содержащее дополнительную информацию:

Столбец	Тип	Описание
flight_id	integer	Идентификатор рейса
flight_no	char(6)	Номер рейса
scheduled_departure	timestamptz	Время вылета по расписанию
scheduled_departure_local	timestamp	Время вылета по расписанию,
		местное время в пункте отправления
scheduled_arrival	timestamptz	Время прилёта по расписанию
scheduled_arrival_local	timestamp	Время прилёта по расписанию,
		местное время в пункте прибытия
scheduled_duration	interval	Планируемая продолжительность полета
departure_airport	char(3)	Код аэропорта отправления
departure_airport_name	text	Название аэропорта отправления
departure_city	text	Город отправления
arrival_airport	char(3)	Код аэропорта прибытия
arrival_airport_name	text	Название аэропорта прибытия
arrival_city	text	Город прибытия
status	varchar(20)	Статус рейса
aircraft_code	char(3)	Код самолета, ІАТА
actual_departure	timestamptz	Фактическое время вылета
actual_departure_local	timestamp	Фактическое время вылета,
		местное время в пункте отправления
actual_arrival	timestamptz	Фактическое время прилёта
actual_arrival_local	timestamp	Фактическое время прилёта,
		местное время в пункте прибытия
actual_duration	interval	Фактическая продолжительность полета

#### Материализованное представление bookings.routes

Таблица рейсов содержит избыточность: из нее можно было бы выделить информацию о маршруте (номер рейса, аэропорты отправления и назначения), которая не зависит от конкретных дат рейсов.

Столбец	Тип	Описание
flight_no departure_airport departure_airport_name departure_city arrival_airport arrival_airport_name arrival_city aircraft_code duration	char(6) char(3) text text char(3) text char(3) text text	Номер рейса Код аэропорта отправления Название аэропорта отправления Город отправления Код аэропорта прибытия Название аэропорта прибытия Город прибытия Код самолета, IATA
days_of_week	integer[]	Дни недели, когда выполняются рейсы

#### Примеры бизнес задач, которые можно решить используя БД

#### 1)Определять % заполняемости самолетов по рейсам.

Можно использовать, чтобы подобрать оптимальный самолет для рейса, увеличить % заполняемости, управлять ценой на билеты и предлагать акции на свободные места.

#### 2)Определять выручку за рейс

Помогает управлять экономикой перевозок

#### 3)Искать ближайшие рейсы, вылетающие по нужному направлению.

Информацию по конкретным рейсам и свободных местах можно использовать для продажи билетов через рекламу.

#### 4)Определять какие самолеты можно ставить на какие рейсы

Самолеты не всегда летают по одному рейсу, иногда есть необходимость послать самолет из конкретного города в другой, и надо быть уверенным что максимальная дальность полета позволяет это сделать.

#### 5)Определение самого выгодного предложения по перелету в рамках рейса по классам.

Иногда продать бизнес класс дешевле эконом класса выгоднее. Можно использовать в рекламе или как способ повышения лояльности для постоянных клиентов.

#### 6)Узнавать информацию о текущих бронях на будущие рейсы

Помогает заранее заполнить самолеты

## 5)Список SQL запросов

#### Вопросы на которые надо ответить используя SQL запросы

1. В каких городах больше одного аэропорта?

2. В каких аэропортах есть рейсы, выполняемые самолетом с максимальной дальностью перелета?

3. Вывести 10 рейсов с максимальным временем задержки вылета

```
-- 3) Вывести 10 рейсов с максимальным временем задержки вылета

SELECT
-- f.flight_id AS "ID Полета"
    f.flight_no AS "Homep peйca"
-- ,f.scheduled_departure AS "Запланированный отъезд"
-- ,f.actual_departure AS "Фактический отъезд"
    ,age(f.actual_departure, f.scheduled_departure) AS "Время задержки вылета"

FROM flights f

WHERE f.actual_departure IS NOT NULL

ORDER BY age(f.actual_departure, f.scheduled_departure) DESC

LIMIT 10
```

4. Были ли брони, по которым не были получены посадочные талоны?

```
--- 4)Были ли брони, по которым не были получены посадочные талоны:

SELECT

t.book_ref
,count(bp.boarding_no) AS count_boarding_no

FROM tickets t

LEFT JOIN boarding_passes bp ON bp.ticket_no = t.ticket_no

GROUP BY book_ref

HAVING count(bp.boarding_no) = 0
```

5. Найдите свободные места для каждого рейса, их % отношение к общему количеству мест в самолете. Добавьте столбец с накопительным итогом - суммарное накопление количества вывезенных пассажиров из каждого аэропорта на каждый день. Т.е. в этом столбце должна отражаться накопительная сумма - сколько человек уже вылетело из данного аэропорта на этом или более ранних рейсах за день.

```
Найдите свободные места для каждого рейса, их % отношение к общему количеству мест в самолете.
SELECT
   f.flight_no
   ,f.aircraft_code
   ,f.actual_departure
   ,f.departure_airport
   ,t.all_seat_aircraft - t2.booked_seat AS free_seat
   ,round((t.all_seat_aircraft - t2.booked_seat)::float / t.all_seat_aircraft * 100) AS "%_free_seat"
    ,sum(t2.booked seat) OVER (PARTITION BY f.departure airport, f.actual departure::date
   ORDER BY f.actual_departure, f.aircraft_code) AS sum_pas
FROM flights f
LEFT JOIN (SELECT
   aircraft_code
    ,count(seat_no) AS all_seat_aircraft
   FROM seats s
   GROUP BY aircraft_code) AS t ON t.aircraft_code = f.aircraft_code
LEFT JOIN (SELECT
   flight_id
    ,count(seat_no) AS booked_seat
   FROM boarding_passes bp
   GROUP BY flight_id ) AS t2 ON t2.flight_id = f.flight_id
WHERE t2.booked_seat IS NOT NULL AND f.actual_departure IS NOT NULL
```

6. Найдите процентное соотношение перелетов по типам самолетов от общего количества.

7. Были ли города, в которые можно добраться бизнес - классом дешевле, чем эконом-классом в рамках перелета?

```
7) Были ли города, в которые можно добраться бизнес - классом дешевле,
-- 7.1) Создаем 2 CTE с группировкой по flight id, fare conditions
-- и агрегацией max(amount) для Economy и min(amount) для Business
-- 7.3) Добавляем СТЕ business по flight id
-- 7.5) Фильтруем строки по условию что бизнес - классом дешевле, чем эконом-классом
WITH economy AS (
SELECT
    flight id
    ,fare conditions
    ,max(amount) AS max economy flight
FROM ticket flights tf
WHERE fare conditions = 'Economy'
GROUP BY
    flight_id
    ,fare_conditions
ORDER BY flight id
SELECT
    flight id
    ,fare conditions
    ,min(amount) AS min business flight
FROM ticket flights tf
WHERE fare conditions = 'Business'
GROUP BY
    flight id
    ,fare_conditions
ORDER BY flight_id
SELECT
    economy.flight id
    ,economy.max economy flight
    ,bu.min business flight
    ,bu.min_business_flight < economy.max_economy_flight AS business_cheaper</pre>
    ,ai.city AS city_from
    ,fl.arrival airport
    ,ai2.city AS city to
FROM economy
INNER JOIN business bu ON bu.flight id = economy.flight id
LEFT JOIN flights fl ON fl.flight id = economy.flight id
LEFT JOIN airports ai ON ai.airport code = fl.departure airport
LEFT JOIN airports ai2 ON ai2.airport code = fl.arrival airport
WHERE bu.min_business_flight < economy.max_economy_flight
```

#### 8. Между какими городами нет прямых рейсов?

```
--Логика решения:
-- 2)Надо получить все возможные сценарии полетов для каждого города
-- 3)Вычесть из списка № 2 список № 1 текущие полеты для этого города
SELECT
    a.city AS departure_city
    ,a2.city AS arrival_city
FROM flights f
LEFT JOIN airports a ON a.airport_code = f.departure_airport
LEFT JOIN airports a2 ON a2.airport_code = f.arrival_airport
GROUP BY a.city, a2.city
ORDER BY departure_city, arrival_city
SELECT DISTINCT
FROM airports a
SELECT
FROM unique city uc, unique city uc2
WHERE uc.city != uc2.city
ORDER BY uc.city, uc2.city
WITH unique city AS (
SELECT DISTINCT a.city
FROM airports a)
SELECT
EXCEPT
SELECT
    a.city AS departure_city
FROM flights f
LEFT JOIN airports a ON a.airport_code = f.departure_airport
LEFT JOIN airports a2 ON a2.airport_code = f.arrival airport
GROUP BY a.city, a2.city
ORDER BY 1, 2
```

9. Вычислите расстояние между аэропортами, связанными прямыми рейсами, сравните с допустимой максимальной дальностью перелетов в самолетах, обслуживающих эти рейсы

```
-- 9) Вычислите расстояние между аэропортами, связанными прямыми рейсами, сравните
-- с допустимой максимальной дальностью перелетов в самолетах, обслуживающих эти рейсы
-- Логика решения:
-- 9.1) Берем таблицу flights и обогащаем городами, также в таблице airports находятся координаты аэропортов
-- 9.2) Расчитываем d — расстояние между пунктами, измеряется в радианах длиной дуги большого круга земного шара.
-- Кратчайшее расстояние между двумя точками A и B на земной поверхности (если принять ее за сферу) определяется зависимостью:
-- d = arccos {sin(latitude_a)·sin(latitude_b) + cos(latitude_a)·cos(latitude_b)·cos(longitude_a - longitude_b)}
-- latitude_a и latitude_b — широты,
-- longitude_a, longitude_b — долготы данных пунктов
-- d — расстояние между пунктами измеряется в радианах длиной дуги большого круга земного шара.
-- 9.3) Определяем расстояние в км
-- Расстояние между пунктами, измеряемое в километрах, определяется по формуле:
-- L = d·R, где R = 6371 км — средний радиус земного шара.
-- 9.4) Сравниваем с допустимой максимальной дальностью перелетов в самолетах через САЅЕ
```

```
SELECT DISTINCT
    f.departure_airport
    ,f.arrival_airport
    ,f.aircraft_code
    ,a3."range"
    ,round(acos(sind(a.latitude) * sind(a2.latitude) + cosd(a.latitude) *
    cosd(a2.latitude) * cosd(a.longitude - a2.longitude)) * 6371) AS distance_km
        WHEN a3."range" < round(acos(sind(a.latitude) * sind(a2.latitude) +</pre>
        cosd(a.latitude) * cosd(a2.latitude) * cosd(a.longitude - a2.longitude)) * 6371) THEN 'We have a problem'
WHEN a3."range" > round(acos(sind(a.latitude) * sind(a2.latitude)
         + cosd(a.latitude) * cosd(a2.latitude) * cosd(a.longitude - a2.longitude)) * 6371) THEN 'Good'
    END AS will_fly_or_not
FROM flights f
LEFT JOIN airports a ON a.airport_code = f.departure_airport
LEFT JOIN airports a2 ON a2.airport_code = f.arrival_airport
LEFT JOIN aircrafts a3 ON a3.aircraft_code = f.aircraft_code
ORDER BY
    f.departure_airport
    ,f.arrival_airport
```