Отчет Рябышева Евгения по итоговому проекту SQL для аналитика

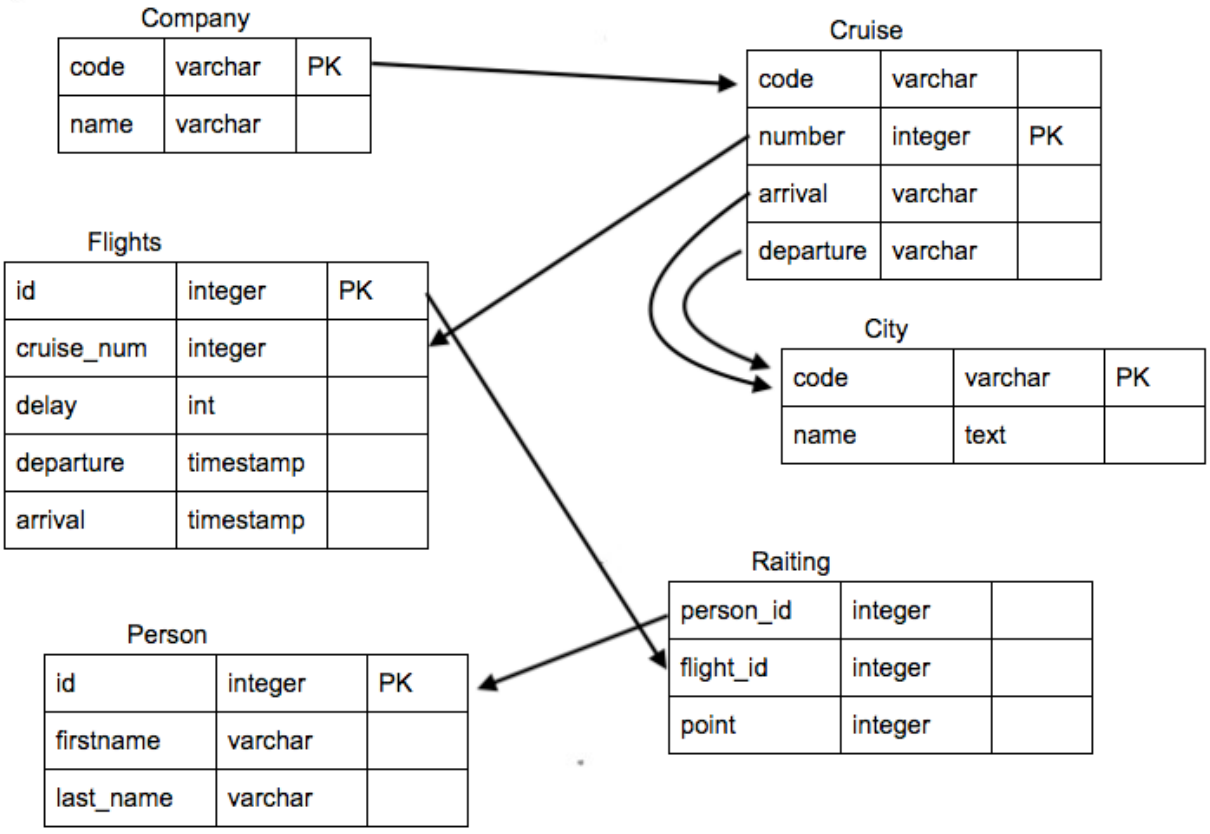
1. Генерация данных для проекта.

Для выполнения проекта был написан скрипт на языке Groovy. В директории table находятся классы отвечающие за структуру таблиц. В директории generator находятся классы отвечающие за генерацию данных и запись этих данных в БД.

Данный скрипт генерирует случайные данные, которые представляют статистические данные для нескольких авиакомпаний.

Итог работы скриптов представлен в директориях csv и dumps.

2. Описание таблиц



3. SQL запросы в БД

```
-
-
1
```

```sql
-- количество полетов для каждого рейс
select company.name, cruise.code, cruise.number, count(flights.id)
from company
        join cruise on company.code = cruise.code
        join flights on cruise.number = flights.cruise_number
group by company.name, cruise.code, cruise.number;
```

```sql
-- 2
-- среднее время задержки рейсов для каждой компании
select company.name, round(avg((flights.delay) / 3600000), 2) as avg_delay
from company
        join cruise on company.code = cruise.code
        join flights on cruise.number = flights.cruise_number
group by company.name
order by avg_delay
```

```sql
-- 3
-- топ 10 самых популярных городов
select DISTINCT city.name from city
```

```
join (
    with quantity_flight as (
            select company.name as name, cruise.code as code, cruise.number as number,
count(flights.id) as count
        from company
            join cruise on company.code = cruise.code
            join flights on cruise.number = flights.cruise_number
        group by company.name, cruise.code, cruise.number
        order by count DESC
    )
    select cruise.arrival from cruise
                        join quantity_flight on quantity_flight.code = cruise.code


  ) as temp on temp.arrival = city.code
LIMIT 10;
```

```
-- 4
-- отклонение рейтинга пользователя от среднего рейтинга, который он склонен
выставлять
select id, cruise_number, point,
     point - AVG(point) OVER (PARTITION BY id) rating_deviance_simplex,
      point - SUM(point) OVER (PARTITION BY id) /COUNT(point) OVER (PARTITION BY id) as
rating_deviance_complex
from (
    select distinct person.id, flights.cruise_number, rating.point from person
                                        join rating on rating.person_id = person.id
```

```
                                    join flights on rating.flight_id = flights.id
                                        join cruise on flights.cruise_number =
cruise.number
  ) as temp
ORDER BY id, point DESC;




-- 5
-- отклонение оценки от максимальной
select id, cruise_number, point,
           round((id - min(point) over (partition by  id))/(max(point) over (partition by
id))::numeric, 2) as rating_deviance_simplex
from (
   select distinct person.id, flights.cruise_number, rating.point from person
                                        join rating on rating.person_id = person.id
                                        join flights on rating.flight_id = flights.id
                                            join cruise on flights.cruise_number =
cruise.number
   ) as temp
ORDER BY id;




-- 6
-- рейтинг компаний
select name, avg, row_number() over (partition by name) as number
```

```
from (
    select company.name, avg(rating.point) from company
                                        join cruise on company.code = cruise.code
                                        join flights on cruise.number = flights.cruise_number
                                        join rating on flights.id = rating.flight_id
    group by company.name
   ) as temp;
```

```
-- 7
-- рейсы выполненные за апрель и которые были задержаны
with april as (
   select id, arrival, departure
   from flights
   where arrival between '2018-04-01' :: timestamp and '2018-05-01' :: timestamp
   order by arrival, departure
)
select company.code, flights.id, round(flights.delay / 3600000, 2) as delay
from company
      join cruise on company.code = cruise.code
      join flights on flights.id = cruise.number
      join april on april.id = flights.id
where delay <> 0
order by company.code, delay DESC
```

```
--8
-- пассажиры совершившие полет между Indianapolis, IN: Indianapolis International и Chattanooga, TN: Lovell Field
select person.id, person.firstname, person.lastname, cruise.arrival, cruise.departure
from person
    join rating on person_id = id
    join flights on flights.id = rating.flight_id
    join cruise on flights.cruise_number = cruise.number
where cruise.departure LIKE 'IND' and cruise.arrival LIKE 'CHA'
order by cruise.arrival, cruise.departure;
```

```
-- 9
-- топ 50 лояльных пользователей к компании American Airlines Inc.
select rating.person_id, cruise.code, temp.avg, temp.count from cruise
                    join flights on cruise_number = number
                    join rating on id = flight_id
join (
        select id, avg(rating.point) as avg, count(rating.point) as count from person
join rating on rating.person_id = person.id
   group by id
  ) as temp on temp.id = rating.person_id
where cruise.code like 'AA'
order by avg DESC, count DESC, rating.person_id DESC
limit 50;
```

```sql
-- 10
-- рейтинг компании относительно задержек рейса
with delay as (
  select company.name, round(avg((flights.delay) / 3600000), 2) as avg_delay
           from company
                join cruise on company.code = cruise.code
                join flights on cruise.number = flights.cruise_number
           group by company.name
           order by avg_delay
),
  raiting as (
    select company.name, avg(rating.point) as avg_raiting from company
                                       join cruise on company.code = cruise.code
                                       join flights on cruise.number = flights.cruise_number
                                       join rating on flights.id = rating.flight_id
    group by company.name
 )

select delay.name, delay.avg_delay, raiting.avg_raiting from delay
                                              join raiting on raiting.name = delay.name
order by delay.avg_delay desc, raiting.avg_raiting desc;
```