

Описание данных

База данных о поездках такси в Чикаго:

Таблица `neighborhoods` — информация о районах города:

`neighborhood_id` — код района;
`name` — название района.

Таблица `cabs` — информация о такси:

`cab_id` — идентификатор такси;
`vehicle_id` — уникальный идентификатор автомобиля;
`company_name` — компания, которой принадлежит автомобиль.

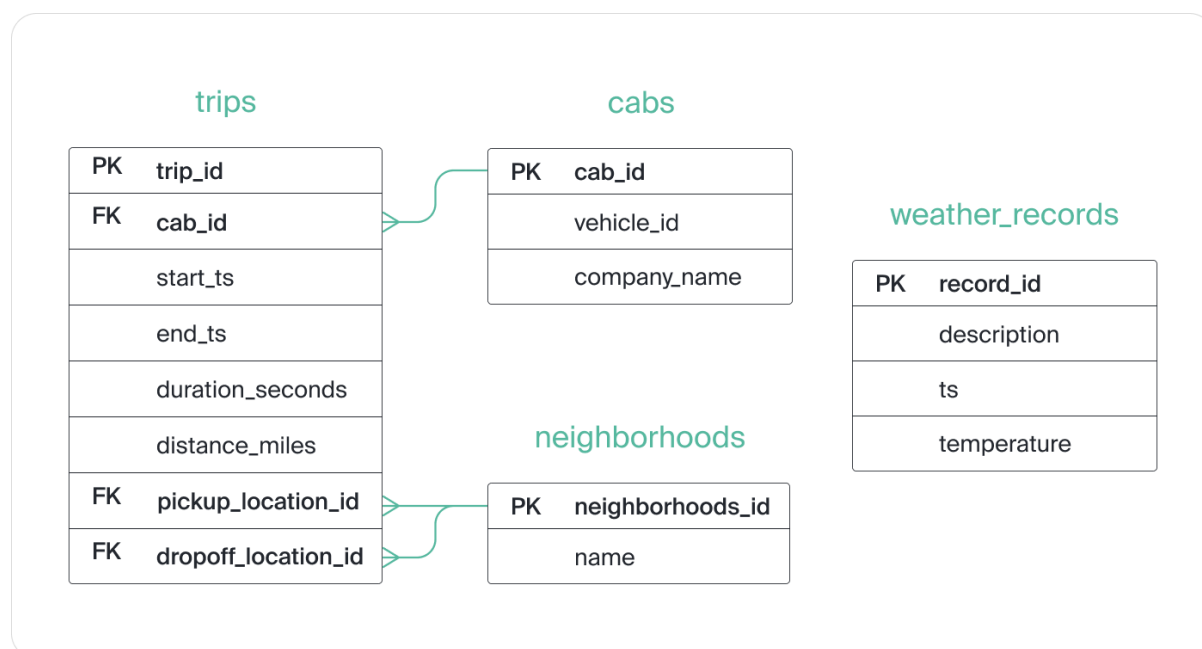
Таблица `trips` — информация о поездках:

`trip_id` — код поездки;
`cab_id` — идентификатор такси, на котором была совершена поездка;
`start_ts` — дата и время начала поездки (время округлено до часа);
`end_ts` — дата и время окончания поездки (время округлено до часа);
`duration_seconds` — длительность поездки в секундах;
`distance_miles` — дальность поездки в милях;
`pickup_location_id` — код района города, в котором была начата поездка;
`dropoff_location_id` — код района города, в котором завершилась поездка.

Таблица `weather_records` — информация о погоде:

`record_id` — код записи погодных наблюдений;
`ts` — дата и время наблюдения (время округлено до часа);
`temperature` — температура на момент наблюдения;
`description` — краткое описание погодных условий. Например, `light rain` или `scattered clouds`.

Схема таблиц



Задание 1

У тебя есть база данных с поездками на такси. По плану на линию обслуживания должно было выйти 10550 автомобилей — эта цифра покрывает спрос пользователей. Команде поступило много жалоб: свободных автомобилей оказалось недостаточно. Сколько такси вышло на линии на самом деле? Информация о всех машинах на линии есть в таблице `cabs`.

Число автомобилей:

```
5500
```

Запрос, которым тебе удалось решить задачу.

```
SELECT COUNT(DISTINCT vehicle_id) FROM cabs;
```

Задание 2

Посчитай количество автомобилей в каждой компании из таблицы `cabs`. Отсортируй значения по убыванию. Команда предполагает, что некоторые компании не вывели достаточно автомобилей на линию. Выведи те компании, в которых меньше 100 автомобилей. Поле с числом автомобилей назови `cnt`, поле с названием компании — `company_name`.

Список компаний с числом автомобилей меньше 100.

```
cnt | company_name
-----+-----
97 | Nova Taxi Affiliation Llc
89 | Patriot Taxi Dbа Peace Taxi Associat
85 | Blue Diamond
81 | Checker Taxi Affiliation
80 | Chicago Medallion Management
69 | Chicago Independents
67 | 24 Seven Taxi
60 | Checker Taxi
55 | American United
53 | Chicago Medallion Leasing INC
49 | Top Cab Affiliation
48 | KOAM Taxi Association
38 | Chicago Taxicab
34 | Norshore Cab
20 | Gold Coast Taxi
20 | Metro Group
18 | Service Taxi Association
14 | 5 Star Taxi
8 | American United Taxi Affiliation
8 | Metro Jet Taxi A
7 | Setare Inc
5 | Leonard Cab Co
1 | 4615 - 83503 Tyrone Henderson
1 | 5062 - 34841 Sam Mestas
1 | 4623 - 27290 Jay Kim
1 | 5997 - 65283 AW Services Inc.
```

```
1 | 2092 - 61288 Sbeih company
1 | 1469 - 64126 Omar Jada
1 | 2733 - 74600 Benny Jona
1 | 2192 - 73487 Zeymane Corp
1 | 5006 - 39261 Salifu Bawa
1 | 3556 - 36214 RC Andrews Cab
1 | 3721 - Santamaria Express, Alvaro Santamaria
1 | 2809 - 95474 C & D Cab Co Inc.
1 | 2241 - 44667 - Felman Corp, Manuel Alonso
1 | 3620 - 52292 David K. Cab Corp.
1 | 2823 - 73307 Lee Express Inc
1 | 6057 - 24657 Richard Addo
1 | 6742 - 83735 Tasha ride inc
1 | 1085 - 72312 N and W Cab Co
1 | 3591 - 63480 Chuks Cab
1 | 0118 - 42111 Godfrey S.Awir
1 | 6574 - Babylon Express Inc.
1 | 3094 - 24059 G.L.B. Cab Co
1 | 5874 - 73628 Sergey Cab Corp.
1 | 6743 - 78771 Luhak Corp
1 | 5074 - 54002 Ahzmi Inc
1 | 3623 - 72222 Arrington Enterprises
1 | 4053 - 40193 Adwar H. Nikola
1 | Chicago Star Taxicab
1 | 3011 - 66308 JBL Cab Inc.
```

(51 rows)

Запрос, которым тебе удалось решить задачу.

```
SELECT
    COUNT(cab_id) AS cnt,
    company_name
FROM
    cabs
GROUP BY
    company_name
HAVING
    COUNT(cab_id) < 100
ORDER BY
    cnt DESC;
```

Задание 3

В приложении такси рассчитывается коэффициент стоимости поездки. Если погода хорошая, значение коэффициента равно 1. Если на улице дождь или шторм, коэффициент повышается до 2. У команды есть гипотеза, что в расчётах коэффициента ошибка. Чтобы проверить расчёт коэффициента, команде нужна выборка данных: разработчик может сверить коэффициент с данными в логах и исправить баг. Твоя задача — получить выборку.

Чтобы это сделать:

Получи описание погодных условий из таблицы `weather_records` для каждого часа.

Раздели все часы на две группы оператором `CASE`: `Bad`, если поле `description` содержит слова `rain` или `storm`; `Good` для всех остальных.

Полученное поле назови `weather_conditions`.

В результирующей таблице должно быть два поля — дата и час (`ts`) и `weather_conditions`.
Сделай выборку за период с 2017-11-05 00:00 по 2017-11-06 00:00.

Таблица с данными за указанный период.

ts	weather_conditions
2017-11-05 00:00:00	Good
2017-11-05 01:00:00	Bad
2017-11-05 02:00:00	Good
2017-11-05 03:00:00	Good
2017-11-05 04:00:00	Bad
2017-11-05 05:00:00	Bad
2017-11-05 06:00:00	Good
2017-11-05 07:00:00	Good
2017-11-05 08:00:00	Good
2017-11-05 09:00:00	Good
2017-11-05 10:00:00	Good
2017-11-05 11:00:00	Good
2017-11-05 12:00:00	Good
2017-11-05 13:00:00	Good
2017-11-05 14:00:00	Bad
2017-11-05 15:00:00	Good
2017-11-05 16:00:00	Bad
2017-11-05 17:00:00	Good
2017-11-05 18:00:00	Bad
2017-11-05 19:00:00	Bad
2017-11-05 20:00:00	Bad
2017-11-05 21:00:00	Good
2017-11-05 22:00:00	Good
2017-11-05 23:00:00	Good
2017-11-06 00:00:00	Good

(25 rows)

Запрос, которым удалось решить задачу.

```
SELECT
  ts,
  CASE
    WHEN description LIKE '%rain%' THEN 'Bad'
    WHEN description LIKE '%storm%' THEN 'Bad'
    ELSE 'Good' END AS weather_conditions
FROM
  weather_records
```

WHERE

ts BETWEEN '2017-11-05 00:00' AND '2017-11-06 00:00';

Задание 4

После обновления ПО таксопарки стали сообщать, что прибыль, которую они получают, не сходится с данными, которые отдаёт приложение. Разработка предполагает, что проблема может быть в данных о количестве поездок.

Чтобы определить, есть ли баг, нужно получить выборку с количеством поездок каждого таксопарка за 15 и 16 ноября 2017 года.

Выведи поле company_name. Поле с числом поездок назови trips_amount и выведи его.

Результаты, полученные в поле trips_amount, отсортируй по убыванию.

Таблица с данными за указанный период.

trips_amount	company_name
19558	Flash Cab
11422	Taxi Affiliation Services
10367	Medallion Leasin
9888	Yellow Cab
9299	Taxi Affiliation Service Yellow
9181	Chicago Carriage Cab Corp
8448	City Service
7701	Sun Taxi
7455	Star North Management LLC
5953	Blue Ribbon Taxi Association Inc.
5015	Choice Taxi Association
4383	Globe Taxi
3355	Dispatch Taxi Affiliation
3175	Nova Taxi Affiliation Llc
2235	Patriot Taxi Dbа Peace Taxi Associat
2216	Checker Taxi Affiliation
2070	Blue Diamond
1955	Chicago Medallion Management
1775	24 Seven Taxi
1607	Chicago Medallion Leasing INC
1486	Checker Taxi
1404	American United
1296	Chicago Independents
1259	KOAM Taxi Association
1014	Chicago Taxicab
978	Top Cab Affiliation
428	Gold Coast Taxi
402	Service Taxi Association
310	5 Star Taxi
250	303 Taxi
230	Setare Inc

210		American United Taxi Affiliation
147		Leonard Cab Co
146		Metro Jet Taxi A
127		Norshore Cab
39		6742 - 83735 Tasha ride inc
37		3591 - 63480 Chuks Cab
36		1469 - 64126 Omar Jada
33		6743 - 78771 Luhak Corp
33		0118 - 42111 Godfrey S.Awir
31		6574 - Babylon Express Inc.
29		Chicago Star Taxicab
29		1085 - 72312 N and W Cab Co
29		2809 - 95474 C & D Cab Co Inc.
27		2092 - 61288 Sbeih company
25		3011 - 66308 JBL Cab Inc.
21		3620 - 52292 David K. Cab Corp.
21		4615 - 83503 Tyrone Henderson
20		3623 - 72222 Arrington Enterprises
16		5074 - 54002 Ahzmi Inc
15		2823 - 73307 Lee Express Inc
15		4623 - 27290 Jay Kim
14		3721 - Santamaria Express, Alvaro Santamaria
14		5006 - 39261 Salifu Bawa
14		2192 - 73487 Zeymane Corp
13		6057 - 24657 Richard Addo
12		5997 - 65283 AW Services Inc.
11		Metro Group
8		5062 - 34841 Sam Mestas
7		4053 - 40193 Adwar H. Nikola
7		2733 - 74600 Benny Jona
5		5874 - 73628 Sergey Cab Corp.
3		2241 - 44667 - Felman Corp, Manuel Alonso
2		3556 - 36214 RC Andrews Cab

(64 rows)

Запрос, которым удалось решить задачу.

```

SELECT
    COUNT (trips.trip_id) AS trips_amount,
    cabs.company_name AS company_name
FROM
    cabs
INNER JOIN trips ON cabs.cab_id = trips.cab_id
WHERE
    start_ts :: date BETWEEN '2017-11-15' AND '2017-11-16'
GROUP BY
    company_name
ORDER BY
    trips_amount DESC;

```

Исходные данные для заданий ниже

Ты тестируешь сервис, который доставляет еду за 30 минут. Пока это маленький стартап, поэтому ты работаешь всего с четырьмя таблицами:

Orders — все доставленные заказы;

ORDERS_ID — ID заказов, int;

USER_ID — ID пользователей, int;

EMPLOYEE_ID — ID сотрудников, int;

DELIVERY_TIME — время доставки в минутах, int;

ITEMS — список товаров, char;

Users — пользователи;

USER_ID — ID пользователей, int;

FULL_NAME — полное ФИО пользователя, char;

PHONE — номер телефона пользователя, char;

ADDRESS — адрес пользователя, char;

Employees — работники;

EMPLOYEE_ID — ID сотрудника, int;

FIRST_NAME — имя сотрудника, char;

LAST_NAME — фамилия сотрудника, char;

PHONE — телефон сотрудника, char;

JOB_ID — ID специализации, int;

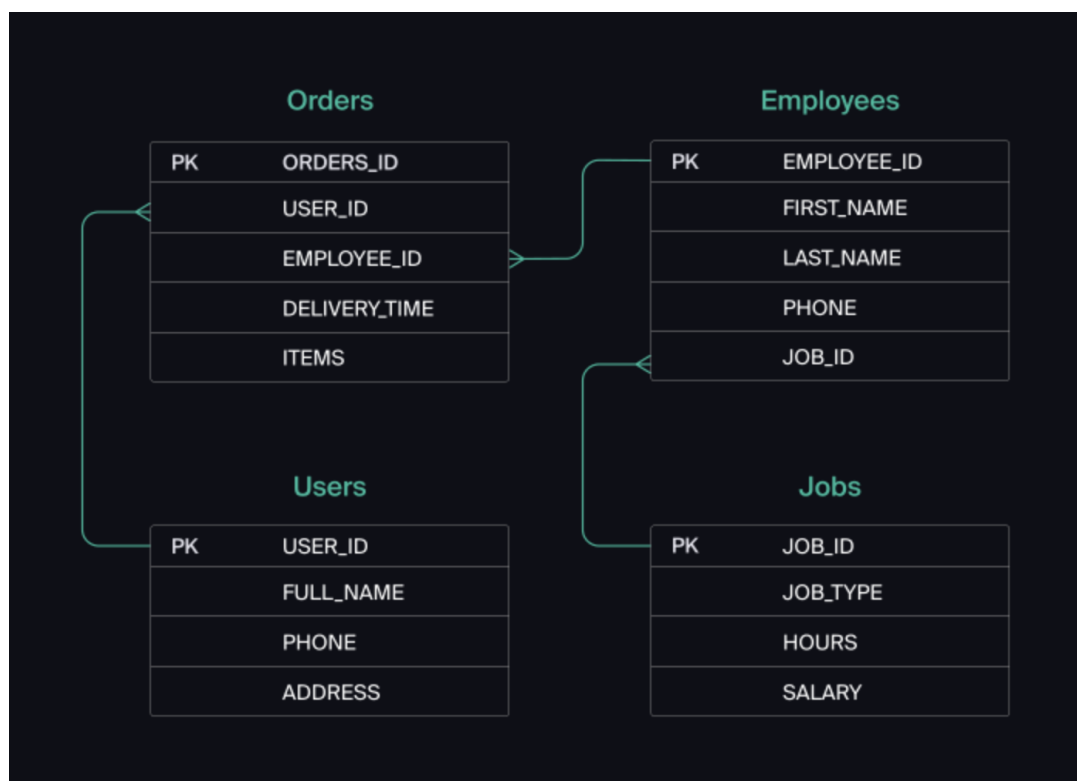
Jobs — типы работ в сервисе

JOB_ID — ID специализации, int;

JOB_TYPE — тип специализации, char;

HOURS — число рабочих часов в неделю, int;

SALARY — зарплата сотрудника с данной специализацией в рублях, int;



Задание 1

В службу поддержки пришло много жалоб: заказы, в которых есть гречка, доставляют почти час, хотя сервис обещает успеть в 30 минут.

Проверь, действительно ли курьеры опаздывают. Выбери все заказы, где есть хотя бы один товар - «гречка» и время доставки свыше 30 минут. В результирующей таблице должны быть ID заказов и ID курьеров.

```
SELECT
    ORDERS_ID,
    EMPLOYEE_ID
FROM
    Orders
WHERE
    DELIVERY_TIME > 30 AND ITEMS LIKE '%гречка%';
```

Задание 2

Менеджер предложил добавить новую функциональность в продукт: мониторинг, который показывает самых активных клиентов за всё время работы компании.

Проверь, что список пользователей корректно выводится на экран. На этом этапе разработки достаточно проверить только ID клиентов.

Выбери пять самых активных клиентов по количеству заказов.

В результирующую таблицу выведи ID каждого пользователя и число заказов.

Отсортируй данные по убыванию числа заказов, выбери пять самых активных клиентов.

```
SELECT
    USER_ID,
    COUNT(ORDERS_ID) as ORDERS_ID_CNT
FROM
    Orders
GROUP BY
    USER_ID
ORDER BY
    ORDERS_ID_CNT DESC
LIMIT
    5;
```

Задание 3

Из бухгалтерии пришёл баг-репорт: зарплаты сотрудников рассчитываются некорректно. Оказалось, что почти все ошибки в расчётах — в расчётных листах менеджеров.

Выведи список ID всех сотрудников, у которых в специализации содержится «менеджер», с зарплатой больше 70 000 рублей.

SELECT

Employees.EMPLOYEE_ID **AS** EMPLOYEE_ID,
Jobs.JOB_ID **AS** JOB_ID,
Jobs.JOB_TYPE **AS** JOB_TYPE,
Jobs.SALARY **AS** SALARY

FROM

Employees

LEFT JOIN Jobs **ON** Jobs.JOB_ID = Employees.JOB_ID

WHERE

Jobs.JOB_TYPE **LIKE** '%менеджер%' **AND** Jobs.SALARY > 70000;