

Консоль

Задание 1

От разработчиков поступила задача: нужно выяснить, какие запросы шли с IP-адреса. IP-адрес состоит из четырёх чисел, они разделены точками. Тебе нужны адреса, которые начинаются с «233.201.».

Логи лежат на удалённом сервере по адресу logs/2019/12. День, когда случилась ошибка, неизвестен.

Твоя задача — узнать, какие запросы были отправлены.

В ответе приложи:

1. команду, которой тебе удалось получить нужные логи;
2. подходящие строки, например: 184.79.247.161 - -
[30/12/2019:21:38:13 +0000] "PUT /alerts HTTP/1.1" 400 3557

Ответ:

Последовательность команд, которой удалось получить нужные логи:

```
cd logs/2019/12
```

```
grep -R "^233.201"
```

Логи:

```
apache_2019-12-18.txt:233.201.188.154 - - [18/12/2019:21:46:01 +0000] "DELETE /events  
HTTP/1.1" 403 3971
```

```
apache_2019-12-21.txt:233.201.182.9 - - [21/12/2019:21:56:20 +0000] "PATCH /users  
HTTP/1.1" 400 4118
```

Задание 2

В системе обнаружен баг. Он проявлялся 30.12.2019 и 31.12.2019 с 21:30:00 до 21:39:59. При этом появлялись ошибки с номерами 400 и 500. Твоя задача — сохранить в отдельный файл логи, которые были записаны в этот период. Затем эти логи надо разложить по отдельным файлам: логи с одинаковой ошибкой положи в один файл. Как это сделать:

1. В домашней директории на удалённом сервере создай директорию bug1.
2. Все запросы, которые произошли в указанный период, положи в файл main.txt в директорию bug1.
3. Внутри директории bug1 создай директорию events.

4. Внутри директории events создай файлы для ошибок с номерами 400 и 500. Назови эти файлы 400.txt и 500.txt соответственно. В них выдели логи с соответствующей ошибкой из файла main.txt.

В ответе приложи:

1. команды, которые создают директории bug1 и events;
2. команду, которой ты выбираешь запросы за указанный период. Это те запросы, которыми ты отбираешь логи в файл main.txt;
3. команды, которыми ты кладёшь логи в файлы 400.txt и 500.txt из main.txt;
4. тексты файлов 400.txt и 500.txt.

Ответ:

Создание директорий bug1 и events:

```
mkdir bug1
```

```
mkdir ~/bug1/events
```

Выборка запросов за указанный период. Выборка логов в файл main.txt:

```
cd logs/2019/12
```

```
grep -R "3./12/2019:21:3:..." > ~/bug1/main.txt
```

Копирование логов в файлы 400.txt и 500.txt из main.txt:

```
grep " 400" ~/bug1/main.txt > ~/bug1/events/400.txt
```

```
grep " 500" ~/bug1/main.txt > ~/bug1/events/500.txt
```

400.txt

```
apache_2019-12-31.txt:86.34.86.182 - - [31/12/2019:21:35:10 +0000] "POST /auth  
HTTP/1.1" 400 3626
```

```
apache_2019-12-31.txt:167.37.16.117 - - [31/12/2019:21:35:17 +0000] "PATCH /customers  
HTTP/1.1" 400 3294
```

```
apache_2019-12-31.txt:199.128.92.19 - - [31/12/2019:21:35:43 +0000] "PUT /users  
HTTP/1.1" 400 4180
```

500.txt.

apache_2019-12-31.txt:208.205.133.127 - - [31/12/2019:21:35:17 +0000] "DELETE /alerts HTTP/1.1" 500 4561

apache_2019-12-31.txt:20.145.255.91 - - [31/12/2019:21:35:30 +0000] "GET /parsers HTTP/1.1" 500 3051

apache_2019-12-31.txt:91.66.134.13 - - [31/12/2019:21:35:53 +0000] "POST /lists HTTP/1.1" 500 3319

База данных

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:

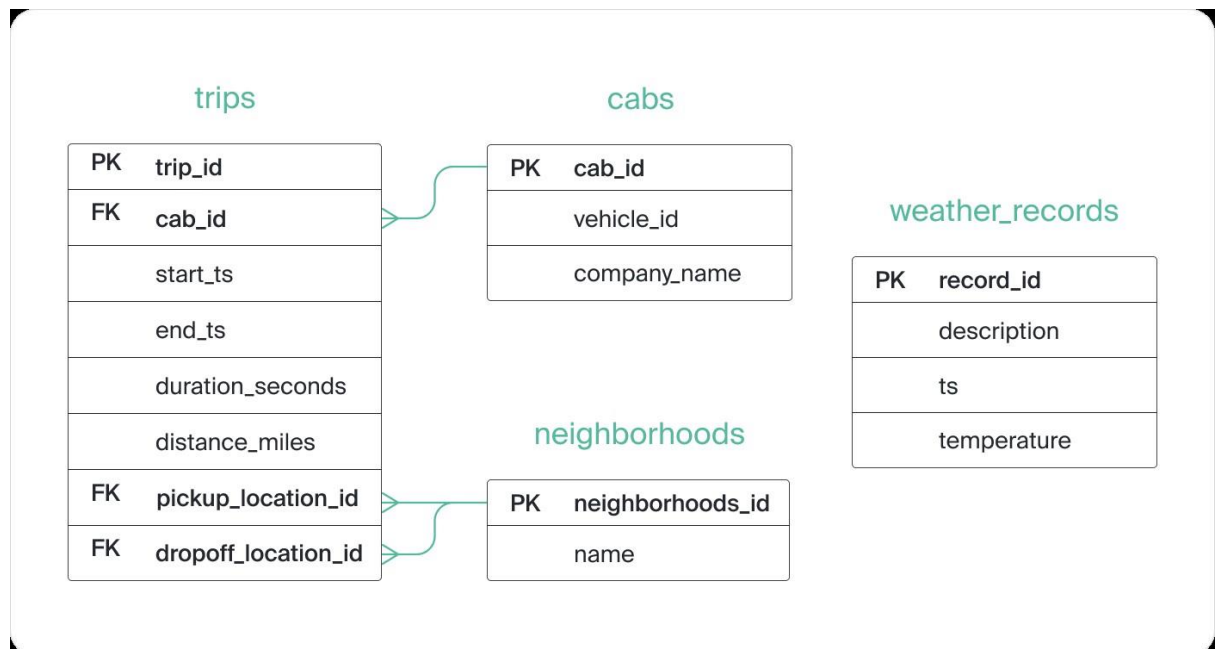
- cab_id — идентификатор такси;
- vehicle_id — уникальный идентификатор автомобиля;
- company_name — компания, которой принадлежит автомобиль.

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`.



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`.

В ответе приложи:

1. число автомобилей;
2. запрос, которым тебе удалось решить задачу

Ответ:

Число автомобилей:

5500

Запрос, которым удалось решить задачу:

```
select count(distinct vehicle_id) as uniq_vehicle from cabs;
```

Задание 2

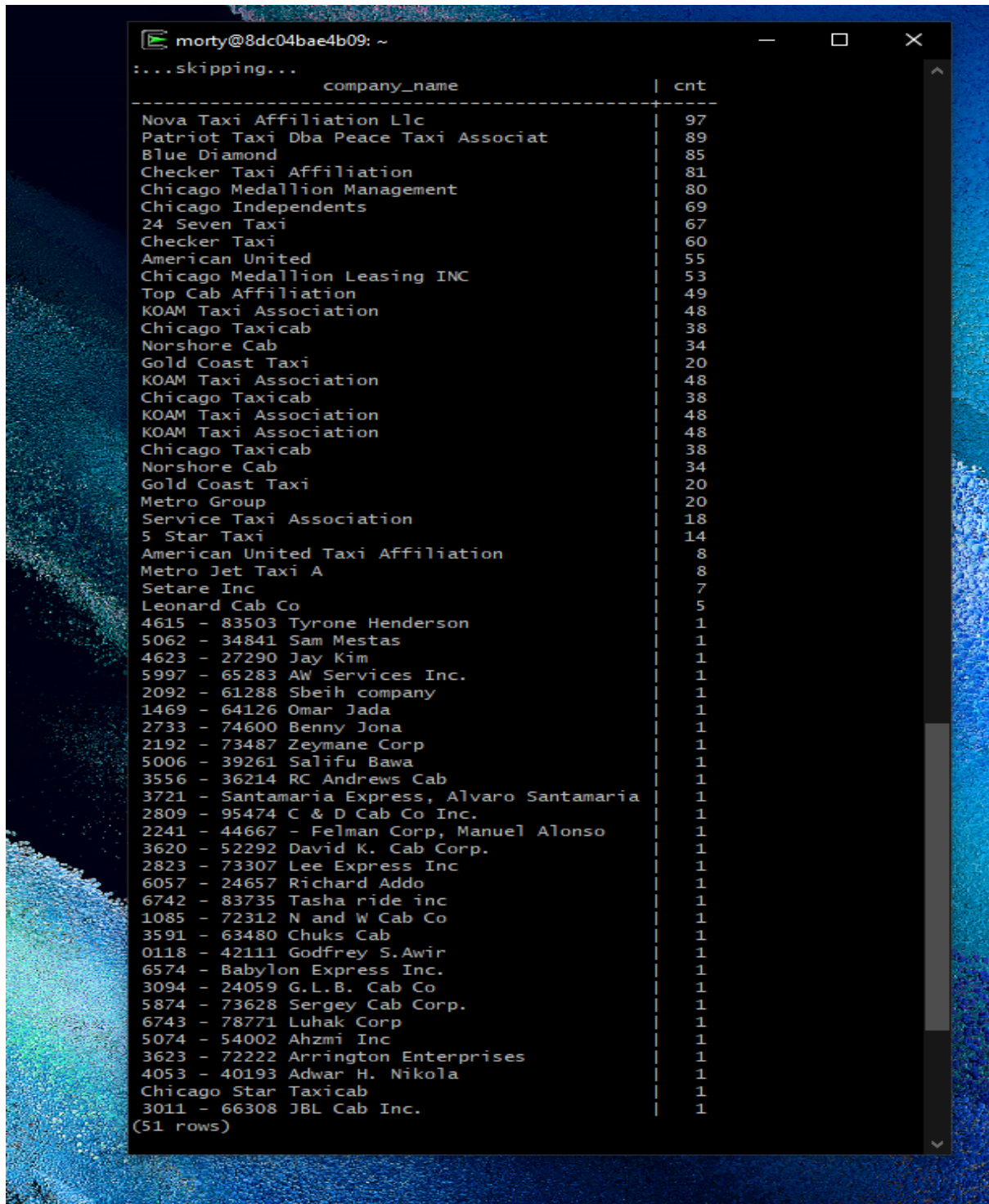
Посчитай количество автомобилей в каждой компании из таблицы `sabs`. Отсортируй значения по убыванию. Команда предполагает, что некоторые компании не вывели достаточно автомобилей на линию. Выведи те компании, в которых меньше 100 автомобилей. Поле с числом автомобилей назови `cnt`, поле с названием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

В ответе приложи:

1. список компаний с числом автомобилей меньше 100;
2. запрос, которым тебе удалось решить задачу.

Ответ:

Список компаний с числом автомобилей меньше 100.



```
morty@8dc04bae4b09: ~
:....skipping...
company_name | cnt
-----|-----
Nova Taxi Affiliation Llc | 97
Patriot Taxi DbA Peace Taxi Associat | 89
Blue Diamond | 85
Checker Taxi Affiliation | 81
Chicago Medallion Management | 80
Chicago Independents | 69
24 Seven Taxi | 67
Checker Taxi | 60
American United | 55
Chicago Medallion Leasing INC | 53
Top Cab Affiliation | 49
KOAM Taxi Association | 48
Chicago Taxicab | 38
Norshore Cab | 34
Gold Coast Taxi | 20
KOAM Taxi Association | 48
Chicago Taxicab | 38
KOAM Taxi Association | 48
KOAM Taxi Association | 48
Chicago Taxicab | 38
Norshore Cab | 34
Gold Coast Taxi | 20
Metro Group | 20
Service Taxi Association | 18
5 Star Taxi | 14
American United Taxi Affiliation | 8
Metro Jet Taxi A | 8
Setare Inc | 7
Leonard Cab Co | 5
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. | 1
(51 rows)
```

Запрос, которым удалось решить задачу:

```
SELECT company_name AS company_name, COUNT(cab_id) AS cnt 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:

1. Получи описание погодных условий из таблицы `weather_records` для каждого часа.
2. Раздели все часы на две группы оператором `CASE`: `Bad`, если поле `description` содержит слова `rain` или `storm`; `Good` для всех остальных.
3. 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.

В ответе приложи:

1. полученную таблицу с данными за указанный период;
2. запрос, которым удалось решить задачу.

Ответ:

Таблица с данными за указанный период.

```
morty=# \c chicago_taxi
You are now connected to database "chicago_taxi" as user "morty".
chicago_taxi=# SELECT ts AS ts, CASE WHEN description LIKE '%rain%' OR description LIKE '%storm%' THEN 'bad' ELSE 'good' END AS weather_conditions FROM weather_records WHERE ts BETWEEN '2017-11-05' AND '2017-11-06';
```

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)

chicago_taxi=#
```

Запрос, которым удалось решить задачу:

```
SELECT ts AS ts, CASE WHEN description LIKE '%rain%' OR description LIKE '%storm%' THEN 'bad' ELSE 'good' END AS weather_conditions FROM weather_records WHERE ts BETWEEN '2017-11-05' AND '2017-11-06';
```


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.

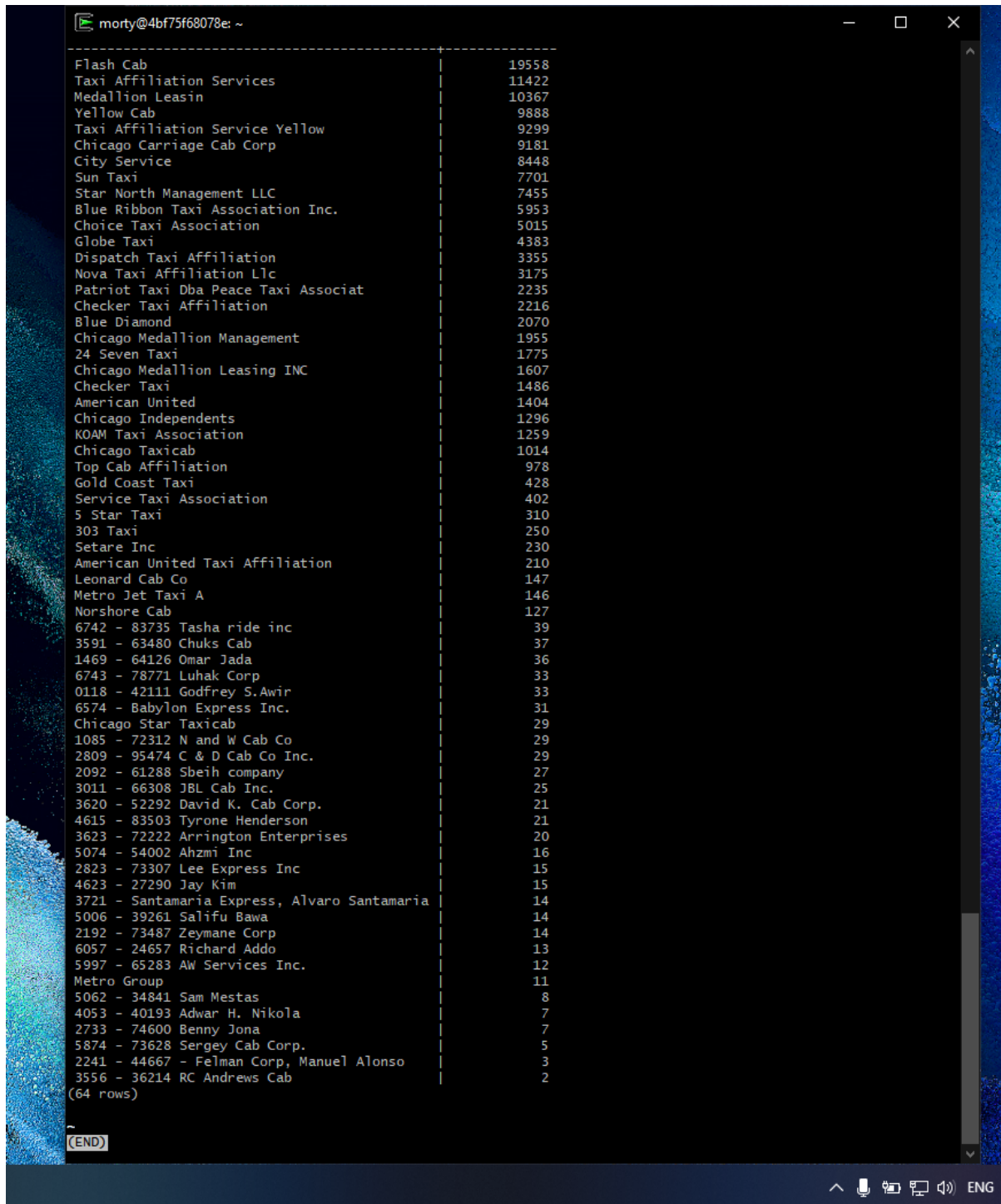
1. Выведи поле `company_name`. Поле с числом поездок назови `trips_amount` и выведи его.
2. Результаты, полученные в поле `trips_amount`, отсортируй по убыванию.

В ответе приложи:

1. полученную таблицу с данными за указанный период;
2. запрос, которым удалось решить задачу.

Ответ:

Таблица с данными за указанный период.



Flash Cab	19558
Taxi Affiliation Services	11422
Medallion Leasin	10367
Yellow Cab	9888
Taxi Affiliation Service Yellow	9299
Chicago Carriage Cab Corp	9181
City Service	8448
Sun Taxi	7701
Star North Management LLC	7455
Blue Ribbon Taxi Association Inc.	5953
Choice Taxi Association	5015
Globe Taxi	4383
Dispatch Taxi Affiliation	3355
Nova Taxi Affiliation Llc	3175
Patriot Taxi DbA Peace Taxi Associat	2235
Checker Taxi Affiliation	2216
Blue Diamond	2070
Chicago Medallion Management	1955
24 Seven Taxi	1775
Chicago Medallion Leasing INC	1607
Checker Taxi	1486
American United	1404
Chicago Independents	1296
KOAM Taxi Association	1259
Chicago Taxicab	1014
Top Cab Affiliation	978
Gold Coast Taxi	428
Service Taxi Association	402
5 Star Taxi	310
303 Taxi	250
Setare Inc	230
American United Taxi Affiliation	210
Leonard Cab Co	147
Metro Jet Taxi A	146
Norshore Cab	127
6742 - 83735 Tasha ride inc	39
3591 - 63480 Chuks Cab	37
1469 - 64126 Omar Jada	36
6743 - 78771 Luhak Corp	33
0118 - 42111 Godfrey S.Awir	33
6574 - Babylon Express Inc.	31
Chicago Star Taxicab	29
1085 - 72312 N and W Cab Co	29
2809 - 95474 C & D Cab Co Inc.	29
2092 - 61288 Sbeih company	27
3011 - 66308 JBL Cab Inc.	25
3620 - 52292 David K. Cab Corp.	21
4615 - 83503 Tyrone Henderson	21
3623 - 72222 Arrington Enterprises	20
5074 - 54002 Ahzmi Inc	16
2823 - 73307 Lee Express Inc	15
4623 - 27290 Jay Kim	15
3721 - Santamaria Express, Alvaro Santamaria	14
5006 - 39261 Salifu Bawa	14
2192 - 73487 Zeymane Corp	14
6057 - 24657 Richard Addo	13
5997 - 65283 AW Services Inc.	12
Metro Group	11
5062 - 34841 Sam Mestas	8
4053 - 40193 Adwar H. Nikola	7
2733 - 74600 Benny Jona	7
5874 - 73628 Sergey Cab Corp.	5
2241 - 44667 - Felman Corp, Manuel Alonso	3
3556 - 36214 RC Andrews Cab	2

(64 rows)

(END)

Запрос, которым удалось решить задачу:

```
SELECT cabs.company_name AS company_name, COUNT(trip_id) AS trips_amount
FROM trips RIGHT JOIN cabs ON cabs.cab_id = trips.cab_id WHERE start_ts BETWEEN
'2017-11-15 00:00' AND '2017-11-16 23:59' GROUP BY company_name ORDER BY
trips_amount DESC;
```