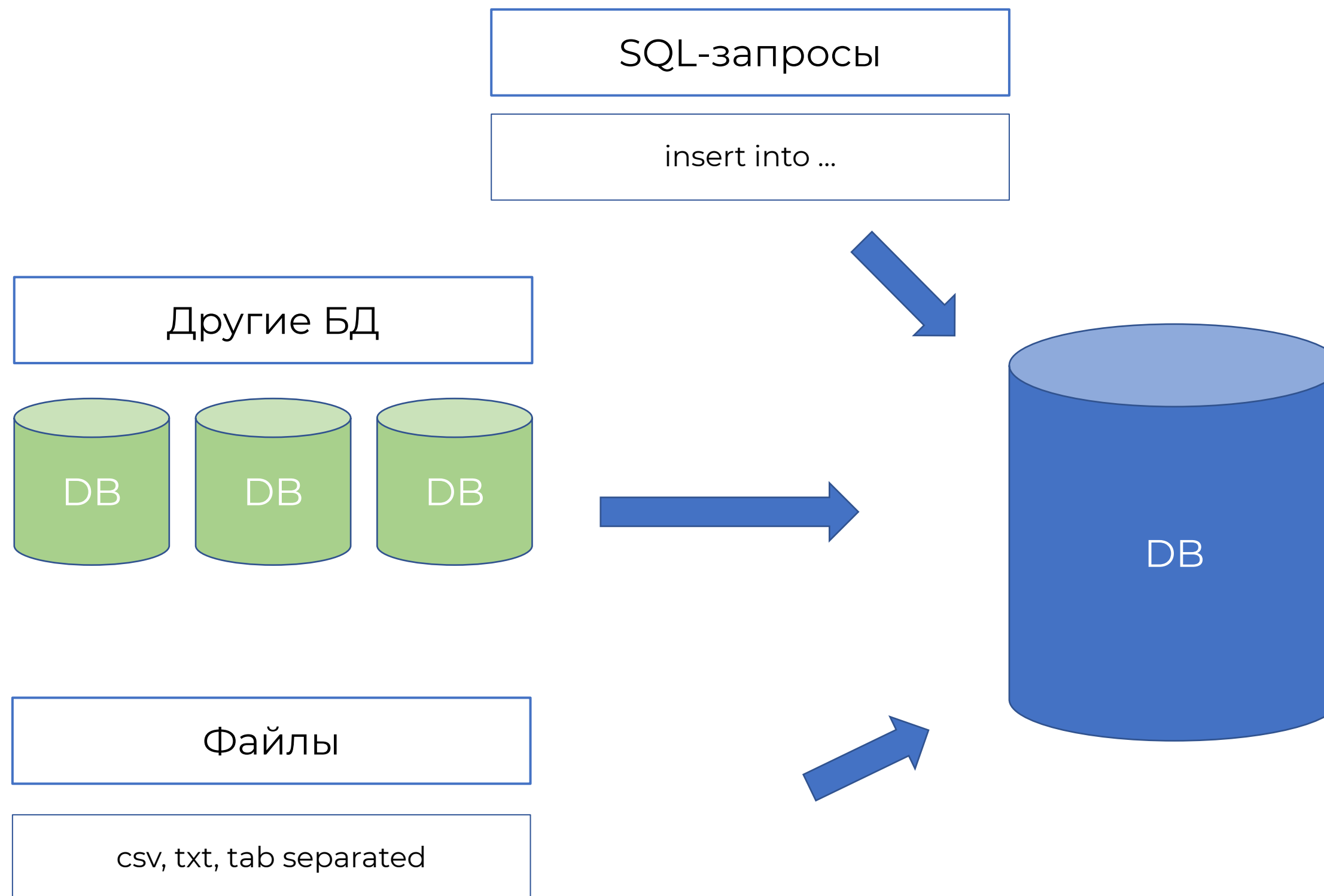


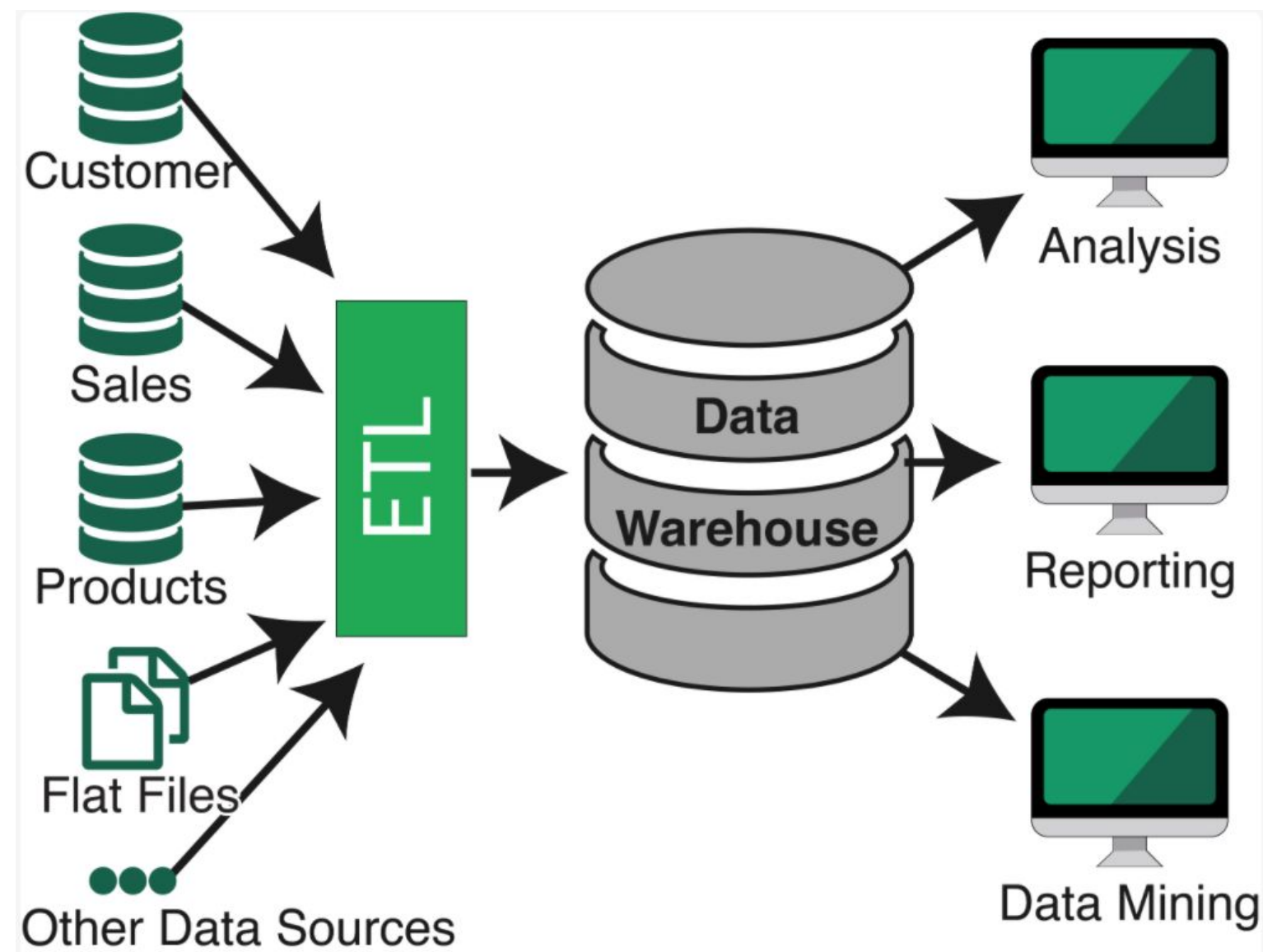


Импорт данных





ETL — «извлечение, преобразование, загрузка»



Импорт данных из текстового файла

CSV

```
0,2019-07-08,Mobile phones,Android,office B,79
1,2019-07-08,Cameras,camera,office A,86
2,2019-07-09,Cameras,webCam,office C,57
3,2019-07-09,Computers,notebook,office B,18
4,2019-07-09,Mobile phones,iPhone,office B,30
5,2019-07-09,Computers,server,office C,83
6,2019-07-09,Mobile phones,Android,office A,41
7,2019-07-09,Mobile phones,Android,office B,37
8,2019-07-09,Computers,server,office C,95
9,2019-07-09,Cameras,camera,office B,14
10,2019-07-09,Cameras,webCam,office C,18
11,2019-07-09,Cameras,camera,office B,20
```

semicolon separated

```
0;2019-07-08;Computers;server;office B;85
1;2019-07-08;Computers;notebook;office B;1
2;2019-07-08;Computers;desktop;office C;96
3;2019-07-08;Mobile phones;Android;office C;53
4;2019-07-08;Cameras;actionCam;office A;62
5;2019-07-08;Cameras;webCam;office C;46
6;2019-07-08;Mobile phones;iPhone;office A;29
7;2019-07-08;Computers;server;office B;95
8;2019-07-08;Cameras;webCam;office C;12
9;2019-07-08;Mobile phones;Android;office C;33
10;2019-07-08;Computers;server;office A;19
11;2019-07-08;Mobile phones;Android;office A;8
12;2019-07-08;Computers;server;office C;88
```

tab separated

```
0 2019-07-08 Computers desktop office B 36
1 2019-07-08 Computers notebook office A 59
2 2019-07-08 Computers notebook office B 13
3 2019-07-08 Computers notebook office A 86
4 2019-07-08 Mobile phones iPhone office C 71
5 2019-07-08 Cameras camera office A 43
6 2019-07-08 Mobile phones iPhone office B 65
7 2019-07-08 Cameras camera office B 84
8 2019-07-08 Computers server office B 26
9 2019-07-08 Mobile phones iPhone office B 32
10 2019-07-08 Computers notebook office C 13
11 2019-07-08 Mobile phones simplePhone office B
12 2019-07-08 Mobile phones iPhone office B 11
13 2019-07-08 Mobile phones Android office C 19
```

Импорт данных из текстового файла

Заголовок

Разделитель между столбцами

id	dt	productCategory	product	office	cnt
0	2019-07-08	Computers	notebook	office A	8
1	2019-07-08	Computers	notebook	office C	77
2	2019-07-08	Mobile phones	Android	office C	60
3	2019-07-08	Cameras	camera	office B	39
4	2019-07-08	Mobile phones	Android	office B	22
5	2019-07-08	Computers	desktop	office B	62
6	2019-07-08	Cameras	camera	office C	2
7	2019-07-08	Mobile phones	simplePhone	office A	3
8	2019-07-08	Cameras	actionCam	office B	20
9	2019-07-08	Computers	notebook	office A	47
10	2019-07-08	Cameras	webCam	office C	84
11	2019-07-08	Cameras	camera	office C	60
12	2019-07-08	Cameras	camera	office C	17

Разделитель между строками

Выделение строк

Преобразование типов

Подбор нужной кодировки

Типы полей

Количество допустимых ошибок импорта

Импорт данных из текстового файла

```
id;dt;tm;productCategory;product;office;cnt;price
0;2019-07-10;10:15;Cameras;camera;office C;47;370
1;2019-07-10;10:28;Mobile phones;simplePhone;office C;44;320
2;2019-07-10;11:12;Mobile phones;Android;office B;78;1100
3;2019-07-10;11:25;Cameras;webCam;office C;56;250
4;2019-07-10;12:06;Mobile phones;Android;office A;27;1100
5;2019-07-10;12:49;Cameras;actionCam;office A;93;400
6;2019-07-10;13:11;Computers;notebook;office B;51;1024
7;2019-07-10;13:16;Cameras;webCam;office C;1;250
8;2019-07-10;13:44;Cameras;actionCam;office C;9;400
9;2019-07-10;14:40;Cameras;actionCam;office B;49;400
10;2019-07-10;14:51;Computers;notebook;office A;92;1024
11;2019-07-11;10:08;Computers;desktop;office A;31;850
12;2019-07-11;10:12;Cameras;actionCam;office C;7;400
13;2019-07-11;10:20;Computers;notebook;office B;9;1024
14;2019-07-11;11:10;Cameras;webCam;office A;43;250
```



Импорт данных из текстового файла

csv

```
0,2019-07-08,Mobile phones,Android,office B,79
1,2019-07-08,Cameras,camera,office A,86
2,2019-07-09,Cameras,webCam,office C,57
3,2019-07-09,Computers,notebook,office B,18
4,2019-07-09,Mobile phones,iPhone,office B,30
5,2019-07-09,Computers,server,office C,83
6,2019-07-09,Mobile phones,Android,office A,41
7,2019-07-09,Mobile phones,Android,office B,37
8,2019-07-09,Computers,server,office C,95
9,2019-07-09,Cameras,camera,office B,14
10,2019-07-09,Cameras,webCam,office C,18
11,2019-07-09,Cameras,camera,office B,20
```

semicolon separated

```
0;2019-07-08;Computers;server;office B;85
1;2019-07-08;Computers;notebook;office B;1
2;2019-07-08;Computers;desktop;office C;96
3;2019-07-08;Mobile phones;Android;office C;53
4;2019-07-08;Cameras;actionCam;office A;62
5;2019-07-08;Cameras;webCam;office C;46
6;2019-07-08;Mobile phones;iPhone;office A;29
7;2019-07-08;Computers;server;office B;95
8;2019-07-08;Cameras;webCam;office C;12
9;2019-07-08;Mobile phones;Android;office C;33
10;2019-07-08;Computers;server;office A;19
11;2019-07-08;Mobile phones;Android;office A;8
12;2019-07-08;Computers;server;office C;88
```

tab separated

```
0 2019-07-08 Computers desktop office B 36
1 2019-07-08 Computers notebook office A 59
2 2019-07-08 Computers notebook office B 13
3 2019-07-08 Computers notebook office A 86
4 2019-07-08 Mobile phones iPhone office C 71
5 2019-07-08 Cameras camera office A 43
6 2019-07-08 Mobile phones iPhone office B 65
7 2019-07-08 Cameras camera office B 84
8 2019-07-08 Computers server office B 26
9 2019-07-08 Mobile phones iPhone office B 32
10 2019-07-08 Computers notebook office C 13
11 2019-07-08 Mobile phones simplePhone office B
12 2019-07-08 Mobile phones iPhone office B 11
13 2019-07-08 Mobile phones Android office C 19
```

	id	dt	tm	productCategory	product	office	cnt	price
1	0	2019-07-09	10:31:00	Computers	desktop	office C	27	850
2	1	2019-07-09	11:24:00	Mobile phones	iPhone	office A	65	1200
3	2	2019-07-09	11:40:00	Computers	desktop	office C	19	850
4	3	2019-07-09	12:36:00	Cameras	actionCam	office B	74	400
5	4	2019-07-09	13:10:00	Computers	server	office A	87	2766
6	5	2019-07-09	13:34:00	Computers	server	office B	29	2766
7	6	2019-07-09	13:43:00	Mobile phones	Android	office C	60	1100
8	7	2019-07-09	13:55:00	Cameras	actionCam	office A	4	400
9	8	2019-07-09	14:27:00	Mobile phones	Android	office A	83	1100
10	9	2019-07-09	15:26:00	Computers	server	office C	48	2766
11	10	2019-07-09	16:09:00	Computers	server	office B	46	2766
12	11	2019-07-09	16:18:00	Computers	notebook	office B	31	1024

Возможные проблемы и ошибки при импорте данных

0	2019-07-08	Mobile phones	Android	office B	79
1	2019-07-08	Cameras	camera	office A	86
2	2019-07-09	Cameras	webCam	office C	57
3	2019-07-09	Computers	notebook	office B	18
4	2019-07-09	Mobile phones	iPhone	office B	30
5	2019-07-09	Computers	server	office C	83
6	2019-07-08	Mobile phones	iPhone	office A	29
7	2019-07-08	Computers	server	office B	95
8	2019-07-08	Cameras	webCam	office C	12
9	2019-07-08	Mobile phones	Android	office C	33
10	2019-07-08	Computers	server	office A	19
11	2019-07-08	Cameras	camera	office C	60
12	2019-07-08	Cameras	camera	office C	17
13	2019-07-08	Cameras_camera	office	A	15
14	2019-07-08	Computers	notebook	office A	29
15	2019-07-08	Mobile phones	Android	office A	36

- Неправильные разделители
- Специальные символы внутри строковых значений
- Разное количество столбцов
- Национальные кодировки
- Большой размер файла
- Ошибки при конвертации значений

Практика к уроку

Ссылка № 2 — тестовая среда MySQL — в описании к модулю.

В среде реализована функция **split_str(str, splitter_char, position)**, которая реализует выделение подстроки из строки с использованием определённого разделителя между колонками. В качестве примера использования выполнен запрос, разделяющий таблицу **skill_raw_data_csv** на колонки:

```
select split_str(data,',',1) id, split_str(data,',',2) dt, split_str(data,',',3) tm, split_str(data,',',4)  
product_category,  
split_str(data,',',5) product, split_str(data,',',6) office, split_str(data,',',7) cnt, split_str(data,',',8) price  
from skill_raw_data_csv
```

Используя функцию **split_str**, напишите запрос (запросы), разделяющий (разделяющие) данные из таблицы **skill_raw_data_mix** на отдельные колонки. При написании запроса необходимо учесть, что часть строк таблицы использует как разделитель между колонками символ **запятой**, а другая часть — символ **точку с запятой**.



**Спасибо
за внимание!**