

avito.tech

# SQL И БАЗЫ ДАННЫХ

## Оконные функции, CTE и рекурсия

# Повторим пройденное

- ▶ Как сделать пересечение/объединение/разность множеств?
- ▶ Чем **rank()** отличается от **dense\_rank()** ?
- ▶ Как посчитать накопительный итог? (два способа)
- ▶ Зачем нужен оператор LIKE?

# ЧТО БУДЕМ ДЕЛАТЬ СЕГОДНЯ

- ▶ Продолжение разговора об оконных функциях
  - lag / lead
  - first\_value / last\_value
- ▶ Рекурсия в SQL
  - CTE / with
  - with recursive



# Зачем это нужно?



Обсуждение



5 минут

# Зачем это нужно?

- Нужно посмотреть динамику изменения атрибута в витрине
- Посмотреть пользовательский путь по платформе
- Группировка пользователей по первому действию



Обсуждение

# Вопросы по прериду

- Что выдаёт функция, если не может найти предыдущее значение?
- Что делает **lag / lead** ? Являются ли они взаимозаменяемыми?
- Что делает **first\_value / last\_value** ? Являются ли они взаимозаменяемыми?
- Можно ли использовать комбинацию lead + distinct?
- Когда функция lag может выдавать разные значения для одного источника?



Обсуждение



5 минут



# Задача 1

## Условие:

-- d8\_item\_price (item, price, actual\_date);

## Вопрос:

Выведите все айтемы, в которых продавец увеличил цену и разницу между первой и последней ценой айтема.

123 item_id 	123 delta 
1	30
4	19



Работаем вместе



5 мин



# Задача 2

## Условие:

-- s\_item\_title (item\_id, title, actual\_date) (таблица, в которой лежат все

– d9\_item\_title\_source (item\_id, price, title, log\_date)

## Вопрос:

Есть источник данных, в котором лежат данные об изменении шапки объявления (d9\_item\_title\_source) и таблица в хранилище данных, в которую пишется инкремент обновления (s\_item\_title). Необходимо вывести (item\_id, price, title, log\_date) тех объявлений, информацию о которых нужно обновить в s\_item\_title.



Работаем вместе



10 мин

# Задача 3

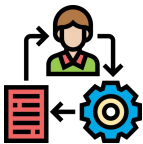
## Условие:

-- d9\_credit (lod\_id, acc\_id, date, amt)

## Вопрос:

Найти аккаунт, у которого разница между первой и последней записью максимальна.

123 acc_id	⌵⌶
	10



Практика



5 мин

# Задача 4

## Условие:

d9\_h\_item (item\_id)

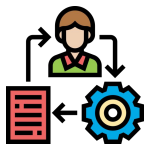
d9\_s\_item\_discount (item\_id, discount, actual\_date)

## Вопрос:

На каждую дату с 2021-01-06 по 2021-01-10

найдите актуальное значение скидки на каждый айтем

item_id	dt	discount
1	2021-01-06	0.16
1	2021-01-07	0.16
1	2021-01-08	0.31
1	2021-01-09	0.31
1	2021-01-10	0.31
2	2021-01-06	<null>
2	2021-01-07	0.13
2	2021-01-08	0.13
2	2021-01-09	0.13
2	2021-01-10	0.43
3	2021-01-06	<null>
3	2021-01-07	<null>
3	2021-01-08	<null>
3	2021-01-09	<null>
3	2021-01-10	0.04



Практика



10 мин

# Задача 5

## Условие:

clickstream (event\_date, user\_id, eventtype)

## Вопрос:

Дан лог событий пользователя

Цепочка событий заканчивается

событием D, потом начинается

следующая

Найдите дату начала и дату конца

каждой цепочки каждого пользователя

123 user_id	123 eventchain	🕒 min	🕒 max
1	0	2021-01-01	2021-01-04
1	1	2021-01-05	2021-01-09
1	2	2021-01-10	2021-01-14
1	3	2021-01-15	2021-01-16
2	0	2021-01-01	2021-01-04
2	1	2021-01-05	2021-01-09
2	2	2021-01-10	2021-01-14



Практика



10 мин



# Перерыв

10 минут



Академия  
Аналитиков  
Авито

# Рекурсия



# Зачем это нужно?



Обсуждение



5 минут

# Зачем это нужно?

- ИНОГДА некоторые штуки удобнее сделать с помощью рекурсии
- Обход древовидной структуры



Обсуждение



# Зачем это нужно?

*Как понимать рекурсию*      *что выведет запрос? Он завершится?*

```
with recursive t(n) as (  
    select 1  
    union  
    select mod(n + 1, 5) from t  
)  
select * from t
```

# Задача 6

## Вопрос:

Реализуйте функцию вычисления чисел Фиббоначи (до 21)



Работаем вместе

123 n	↑↓
	1
	1
	2
	3
	5
	8
	13
	21



5 мин

# Задача 7\*

## Условие:

– d9\_category (id, parent\_id, level, category, parent\_category);

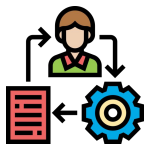
## Вопрос:

для подкатегорий категории A соберите полный путь от корня в формате

'A -> A:X -> A:X:1'

(конкатенацию строк выполняет оператор  
column\_1 || column\_2)

123 id	ABC category	ABC p
1	A	A
2	A:X	A -> A:X
9	A:Y	A -> A:Y
16	A:Z	A -> A:Z
3	A:X:1	A -> A:X -> A:X:1
4	A:X:2	A -> A:X -> A:X:2
5	A:X:3	A -> A:X -> A:X:3
6	A:X:4	A -> A:X -> A:X:4
7	A:X:5	A -> A:X -> A:X:5



Практика



10 мин

# Задача 8\*\*

## Условие:

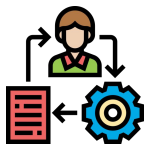
d9\_category (id, parent\_id, level, category, parent\_category);

d9\_sales (dt, category\_id, value)

## Вопрос:

Посчитайте суммарное значение value с подытогом в разрезе первых двух уровней категорий.

ABC l1_cat	ABC l2_cat	123 sum_value
A	A:X	9,066
A	A:Y	16,698
A	A:Z	16,599
A	N/A	1,160
A	[NULL]	43,523
B	B:X	11,684
B	B:Y	10,584
B	B:Z	16,699
B	N/A	2,210



Практика



15 мин

# ДОМАШНЕЕ ЗАДАНИЕ



**Мягкий дедлайн:** 10:00 BC

**Жесткий дедлайн:** 10:00 BT

# Вопросы

Что осталось непонятным?

# Фидбек

Что нового вы узнали на этом занятии?

Что показалось самым важным?

Что будете применять и в каких ситуациях?

Что хочется изучить подробнее?

Ссылка на Обратную связь.

# Что делаем в следующий раз?

- ▶ Узнаем, как изменять данные, а не просто вытаскивать их из БД
- ▶ Посмотрим, зачем нам так нужны были связанные подзапросы

**Обязательно ознакомьтесь с преридом!**