

Ответы на вопросы отборочного тура в треке Data Engineering

ВОПРОС 1

(56% верных ответов)

Для какого значения А истинен предикат?

A like '%_x*'

Варианты ответа

✗ 'wxy'

✓ '_x*'

✗ 'a_xy'

✗ 'a\|x*'

ВОПРОС 2

(50% верных ответов)

Для произвольной колонки 'a' какое утверждение верно?

Варианты ответа

✗ count(1) > count(*) >= count(a)

✗ count(*) > count(1) = count(a)

✓ count(1) = count(*) >= count(a)

✗ ни одно из перечисленных

Примечание:

count() всегда равен count(*), но может быть больше чем count(a), если в поле "a" встречаются NULL значения.*

ВОПРОС 3

(41% верных ответов)

Две таблицы А и В имеют одинаковую структуру.

Приведенный запрос выдал пустой результат.

Описание таблицы

```
create table A(  
  a numeric,  
  b numeric,  
  c numeric,  
  d numeric  
);  
  
create table B(  
  a numeric,  
  b numeric,  
  c numeric,  
  d numeric  
);
```

Запрос

```
(  
  select * from A  
  except  
  select * from B  
)  
  
union all  
  
(  
  select * from B  
  except  
  select * from A  
)
```

Какое утверждение верно?

Варианты ответа

- ☒ Количество строк в таблицах одинаковое
- ☒ Количество строк в таблице А больше или равно количеству строк в таблице В
- ☒ Количество строк в таблице В больше или равно количеству строк в таблице А
- ☒ Нельзя точно утверждать о соотношении количества строк в таблицах

Примечание:

В данном примере нет ограничений на уникальность значений. Поэтому если в одной из таблиц окажутся две или более полностью одинаковых строк, которые есть и в другой таблице, то запрос всё так же будет выдавать пустой результат.

ВОПРОС 4

(57% верных ответов)

Сколько строк выведет запрос?

Описание таблицы

```
create table test_table (  
  a numeric,  
  b numeric,  
  c numeric,  
  d numeric  
);  
  
insert into test_table values(2, 12, 12, 2);  
insert into test_table values(3, 12, 8, 2);  
insert into test_table values(4, 12, 6, 2);  
insert into test_table values(6, 12, 4, 2);
```

Запрос

```
select a, b, c, d from test_table  
where  
  b > c or a > d and a > c
```

Варианты ответа

✗ 1

✗ 2

✓ 3

✗ 4

ВОПРОС 5

(43% верных ответов)

Сколько строк выведет запрос?

Описание таблицы

```
create table test_table (  
  a numeric,  
  b numeric,  
  c numeric,  
  d numeric  
);  
  
insert into test_table values(2, 12, 12, 2);  
insert into test_table values(3, 12, 8, 2);  
insert into test_table values(4, 12, 6, 2);  
insert into test_table values(6, 12, 4, 2);
```

Запрос

```
select t1.a  
from  
  test_table t1  
  left join test_table t2  
    on t1.b = t2.b  
where  
  t2.a > 0  
  
union all  
  
select t1.a  
from  
  test_table t1  
  left join test_table t2  
    on t1.a = t2.c  
where  
  t2.a > 0
```

Варианты ответа

- ☒ 6
- ☒ 8
- ☒ 18
- ☒ будет ошибка

ВОПРОС 6

(45% верных ответов)

Сколько строк выведет запрос?

Описание таблицы

```
create table test_table (  
  a numeric,  
  b numeric,  
  c numeric,  
  d numeric  
);  
  
insert into test_table values(2, 12, 12, 2);  
insert into test_table values(3, 12, 8, 2);  
insert into test_table values(4, 12, 6, 2);  
insert into test_table values(6, 12, 4, 2);
```

Запрос

```
select a, sum(b) from test_table  
where  
  a > 3  
group by a  
having max(c) > d
```

Варианты ответа

- ☒ 1
- ☒ 2
- ☒ 4
- ☒ Будет ошибка

Примечание:

Ошибка будет вызвана тем, что в блоке *having* используется колонка "d", которая отсутствует в группировке и берётся сразу без какой-либо агрегатной функции.

ВОПРОС 7

(51% верных ответов)

Сколько строк выведет запрос?

Описание таблицы

```
create table test_table (  
  a numeric,  
  b numeric,  
  c numeric,  
  d numeric  
);  
  
insert into test_table values(2, 12, 12, 2);  
insert into test_table values(3, 12, 8, 2);  
insert into test_table values(4, 12, 6, 2);  
insert into test_table values(6, 12, 4, 2);
```

Запрос

```
select a from (  
  select  
    a,  
    lead(a, 2, 5) over(order by a) as a_lead  
  from test_table  
) t  
where  
  a_lead > a
```

Варианты ответа

✗ 1

✗ 2

✓ 3

✗ 4

ВОПРОС 8

(59% верных ответов)

Какое утверждение верно, если в следующем запросе изменить inner на left?

Описание таблицы

```
create table test_table (  
  a numeric,  
  b numeric,  
  c numeric,  
  d numeric  
);  
  
insert into test_table values(2, 12, 12, 2);  
insert into test_table values(3, 12, 8, 2);  
insert into test_table values(4, 12, 6, 2);  
insert into test_table values(6, 12, 4, 2);
```

Запрос

```
select t1.a, t2.a  
from test_table t1  
  inner join test_table t2  
    on t1.a = t2.d
```

Варианты ответа

- ☒ Запрос в обоих вариантах упадёт с ошибкой column reference "a" is ambiguous
- ☒ Число строк не изменится
- ☒ Число строк увеличится
- ☒ Число строк уменьшится

ВОПРОС 9

(44% верных ответов)

Какое значение вернет запрос?

Описание таблицы

```
create table test_table (  
    a numeric,  
    b numeric,  
    c numeric,  
    d numeric  
);  
  
insert into test_table values(2, 12, 12, 2);  
insert into test_table values(3, 12, 8, 2);  
insert into test_table values(4, 12, 6, 2);  
insert into test_table values(6, 12, 4, 2);
```

Запрос

```
select sum_a from (  
    select  
        a,  
        sum(a) over(order by a) as sum_a  
    from test_table  
    ) t  
where  
    a = 4
```

Варианты ответа

- ☒ 4
- ☒ 9
- ☒ 16
- ☒ Будет ошибка

Примечание:

Будет произведен расчёт нарастающего итога с помощью оконной функции sum, после чего будет выбрана по фильтру строка, где поле "a" было равно 4. К тому моменту это 2+3+4=9

ВОПРОС 10

(64% верных ответов)

Какой результат выведет запрос?

Описание таблицы

```
create table test_table (  
  a numeric,  
  b numeric,  
  c numeric,  
  d numeric  
);  
  
insert into test_table values(2, 12, 12, 2);  
insert into test_table values(3, 12, 8, 2);  
insert into test_table values(4, 12, 6, 2);  
insert into test_table values(6, 12, 4, 2);
```

Запрос

```
select  
  distinct count(a) over(partition by d)  
from test_table
```

Варианты ответа

- ☒ 1
- ☒ 2
- ☒ 4
- ☒ Нет верного ответа

Примечание:

В данном случае, сначала будет произведён расчёт count для каждой строки в окне по полю "d". Так как в этом поле все значения одинаковые, то для каждой строки мы получим число строк в таблице. После будет произведено удаление дублей с помощью distinct и останется одна строка со значением 4.

ВОПРОС 11

(28% верных ответов)

Какое утверждение верно для полей col1 и col2?

Описание таблицы

```
create table test_table (  
  a numeric,  
  b numeric,  
  c numeric,  
  d numeric  
);  
  
insert into test_table values(2, 12, 12, 2);  
insert into test_table values(3, 12, 8, 2);  
insert into test_table values(4, 12, 6, 2);  
insert into test_table values(6, 12, 4, 2);
```

Запрос

```
select a,  
       lead(a) over(order by c) as col1,  
       lag(a) over(order by c desc) as col2  
from test_table  
order by c
```

Варианты ответа

- ☒ Всегда равны
- ☒ Всегда различаются
- ☒ Равны для всех строк, кроме первой и последней
- ☒ Равны для всех строк, кроме одной

ВОПРОС 12

(37% верных ответов)

Какое утверждение верно для полей col1 и col2?

Описание таблицы

```
create table test_table (  
  a numeric,  
  b numeric,  
  c numeric,  
  d numeric  
);  
  
insert into test_table values(2, 12, 12, 2);  
insert into test_table values(3, 12, 8, 2);  
insert into test_table values(4, 12, 6, 2);  
insert into test_table values(6, 12, 4, 2);
```

Запрос

```
select a,  
       max(a) over(order by a) as col1,  
       max(a) over() as col2  
from test_table
```

Варианты ответа

- ☐ Всегда равны
- ☐ Всегда различаются
- ☐ Запрос упадет с ошибкой пустого окна
- ☒ Равны только для одной строки

ВОПРОС 13

(45% верных ответов)

Какие поля из таблицы ниже могут служить первичным ключом?
(несколько вариантов ответа)

Запрос

```
CREATE TABLE t (  
  a int NOT NULL,  
  b date NOT NULL,  
  c varchar(10),  
  d varchar(10) NOT NULL,  
  e float(10) NOT NULL  
)
```

Варианты ответа

- ✓ Поле a, если оно уникально
- ✓ Поле b, если оно уникально
- ✗ Поле c, если оно уникально (может быть NULL только в одной строке)
- ✓ Поля d и e, если их связка уникальна

ВОПРОС 14

(55% верных ответов)

Выберите лишнее:

Варианты ответа

- ✗ hash join
- ✓ cross join
- ✗ merge join

☒ nested loop

Примечание:

В данном случае, все варианты, кроме CROSS JOIN, являются частями плана запроса, в то время как CROSS JOIN является частью самого запроса.

ВОПРОС 15

(64% верных ответов)

Какое описание справедливо для SCD2?

Варианты ответа

- ☒ Добавление колонки с предыдущим значением
- ☒ Сохранение истории в отдельной таблице
- ☒ Изменение той же строки
- ☒ Добавление строки с новой версией

ВОПРОС 16

(71% верных ответов)

Какое утверждение верно про merge join?

Варианты ответа

- ☒ квадратичная асимптотическая сложность
- ☒ требует объем оперативной памяти, сопоставимый с объемом одной из таблиц
- ☒ обе таблицы должны быть отсортированы
- ☒ возможен только для таблиц из числовых полей

ВОПРОС 17

(35% верных ответов)

Есть две таблицы, которые учитывают вход пассажиров в метро. Какая здесь форма нормализации?

```
CREATE TABLE entry (  
  entry_id int NOT NULL, -- идентификатор прохода  
  passenger_id int NOT NULL, -- идентификатор пассажира  
  student_flag boolean, -- флаг того, является ли пассажир студентом  
  entry_dt timestamp, -- время прохода  
  PRIMARY KEY (entry_id)  
)  
  
CREATE TABLE passenger (  
  passenger_id int NOT NULL, -- идентификатор пассажира  
  last_name varchar(100), -- фамилия пассажира  
  first_name varchar(100), -- имя пассажира  
  place_of_birth varchar(100), -- место рождения пассажира  
  PRIMARY KEY (passenger_id)  
)
```

Варианты ответа

- ☒ 1нф
- ☒ 2нф
- ☒ 3нф
- ☒ Нормальная форма Бойса-Кодда

Примечание:

В данном случае это 2нф, так как в таблице entry имеется поле student_flag. Если бы данное поле было в таблице passenger, то была бы 3нф.

ВОПРОС 18

(48% верных ответов)

Что из перечисленного не подходит для организации корпоративного хранилища?

Варианты ответа

- ☐ SCD
- ☒ OLTP
- ☐ MPP СУБД
- ☐ Data Vault

Примечание:

DWH, MPP СУБД, Data Vault могут использоваться в хранилище

OLTP – не может

ВОПРОС 19

(69% верных ответов)

Выберите утверждение, не характерное для суррогатного ключа:

Варианты ответа

- ☐ Уникальность
- ☐ Генерация системой
- ☒ Бизнес-информативность
- ☐ Неизменность

ВОПРОС 20

(71% верных ответов)

Выберите лишнее:

Варианты ответа

- ☐ Link
- ☐ Hub
- ☐ Satellite
- ☒ Anchor

Примечание:

Anchor – часть якорной модели хранилищ данных. Всё остальное – часть Data Vault методологии.

ВОПРОС 21

(23% верных ответов)

Выберите верные утверждения, характерные для Data Vault:
(несколько вариантов ответа)

Варианты ответа

- ☐ Записи в Hub могут изменяться и могут иметь версии
- ☐ Hub может хранить неключевые атрибуты сущности
- ☒ Один Hub может иметь несколько Satellite, хранящих разные наборы атрибутов
- ☒ Satellite могут быть связаны как с Hub, так и с Link