

Защита SAMSUNG IT HACK 2025

Задание от VK Tech



tech



VK Cloud

gopher go data

Файтельсон Антон - Teamlead/Data Engineer

Гришанов Егор - Analyst

Матанцев Иван - Developer

Создание/ ОПТИМИЗАЦИЯ таблиц

- Q1, Q2, Q3, Q6, Q7, Q8, Q10, Q11, Q12, Q14, Q17, Q19, Q22, Q5
- Создание временных/ оптимизированных таблиц (часто с партицированием) для предварительной фильтрации данных.

partitioning = ARRAY('key')

```
CREATE TABLE  
data.temp.filtered_data AS SELECT ...  
FROM lineitem WHERE ...
```

ИСПОЛЬЗОВАЛИ В:

- Q1, Q2, Q3, Q6, Q7, Q8, Q10, Q11, Q12, Q14, Q17, Q19, Q22, Q5

SQL

Пример запроса

```
CREATE TABLE filtered_parts_q2
WITH (
    format = 'PARQUET',
    partitioning = ARRAY['size']
) AS ...
```

Агрегация с группировкой

- Q1, Q2, Q3, Q7, Q8, Q10, Q11, Q14, Q17, Q19, Q5
- Использование **GROUP BY** для сумм, средних, счетчиков.

GROUP BY

```
SELECT nation, SUM(revenue) FROM  
... GROUP BY nation
```

ИСПОЛЬЗОВАЛИ В:

- Q1, Q2, Q3, Q7, Q8, Q10, Q11, Q14, Q17, Q19, Q5

SQL

Пример запроса

```
SELECT
    returnflag,
    linestatus,
    SUM(quantity) AS sum_qty,
    AVG(extendedprice) AS
avg_price
FROM data.temp.filtered_data
GROUP BY returnflag, linestatus;
```

Множественные JOIN таблиц

- Q2, Q7, Q8, Q12, Q19, Q5
- Соединение 3+ таблиц (например, `lineitem`, `supplier`, `nation`).

LEFT JOIN

```
FROM lineitem  
JOIN supplier ON ...  
JOIN nation ON ...
```

ИСПОЛЬЗОВАЛИ В:

- Q2, Q7, Q8, Q12, Q19, Q5

SQL

Пример запроса

```
SELECT
    fs.acctbal,
    fp.partkey,
    fp.mfgr
FROM filtered_parts_q2 fp
JOIN partsupp ps ON fp.partkey =
ps.partkey
JOIN min_supplycost_q2 msc ON
fp.partkey = msc.partkey AND
ps.supplycost = msc.min_cost
JOIN filtered_supplier_q2 fs ON
ps.supkey = fs.supkey;
```

Использование CTE (WITH)

- Q3, Q6, Q7, Q8, Q12, Q19, Q5
- Поэтапная фильтрация или подготовка данных через CTE.

WITH

```
WITH filtered_parts AS (...),  
     filtered_lineitems AS (...) SELECT ...
```

ИСПОЛЬЗОВАЛИ В:

- Q3, Q6, Q7, Q8, Q12, Q19, Q5

SQL

Пример запроса

```
WITH filtered_customers AS (  
    SELECT custkey  
    FROM customer  
    WHERE mktsegment = 'BUILDING'  
),  
filtered_orders AS (  
    SELECT orderkey, custkey  
    FROM orders  
    WHERE orderdate < DATE  
    '1995-03-11'  
)  
SELECT fo.orderkey,  
SUM(fl.extendedprice * (1 -  
fl.discount)) AS revenue  
FROM filtered_orders fo  
JOIN lineitem fl ON fo.orderkey =  
fl.orderkey;
```

Фильтрация по датам и диапазонам

- Q1, Q6, Q7, Q8, Q10, Q12, Q14, Q19, Q5
- Условия вида `shipdate BETWEEN ... , size BETWEEN 1 AND 5`.

BETWEEN

```
WHERE shipdate >= DATE  
'1995-01-01' AND shipdate < DATE  
'1997-01-01'
```

ИСПОЛЬЗОВАЛИ В:

- Q1, Q6, Q7, Q8, Q10, Q12, Q14, Q19, Q5

SQL

Пример запроса

```
SELECT *  
FROM lineitem  
WHERE shipdate ≥ DATE  
    '1995-09-01'  
    AND shipdate < DATE '1995-09-01'  
+ INTERVAL '1' MONTH;
```

УСЛОВНЫЕ ВЫЧИСЛЕНИЯ (CASE, SUM с УСЛОВИЯМИ)

- Q8, Q12, Q14, Q19, Q5
- Использование **CASE** или сложных условий в **WHERE**.

**CASE
WHERE**

```
SUM(CASE WHEN brand =  
'Brand#23' THEN revenue ELSE 0  
END)
```

ИСПОЛЬЗОВАЛИ В:

- Q8, Q12, Q14, Q19, Q5

SQL

Пример запроса

```
SELECT
    o_year,
    SUM(CASE WHEN nation =
        'GERMANY' THEN volume ELSE 0 END) /
    SUM(volume) AS mkt_share
FROM filtered_predcalc
GROUP BY o_year;
```

Подзапросы и пороговые значения

- Q8, Q11, Q5
- Расчет агрегированных значений (например, `SUM(...)` * `0.00000003`) и их использование в `HAVING`.

HAVING

`HAVING SUM(value) > (SELECT threshold FROM ...)`

ИСПОЛЬЗОВАЛИ В:

- Q8, Q11, Q5

SQL

Пример запроса

```
HAVING SUM(supplycost * availqty) >  
(SELECT threshold FROM  
total_germany);
```

Отчеты с сортировкой и ограничением

- Q2, Q3, Q7, Q10, Q19, Q5
- Итоговые **SELECT** с **ORDER BY** и **LIMIT**.

ORDER BY

ORDER BY revenue DESC LIMIT 200

ИСПОЛЬЗОВАЛИ В:

- Q2, Q3, Q7, Q10, Q19, Q5

SQL

Пример запроса

```
ORDER BY revenue DESC LIMIT 200;
```

Обработка отсутствующих данных (LEFT JOIN + NULL)

- Q13, Q22
- Поиск записей без соответствий (например, клиенты без заказов).

NULL

LEFT JOIN orders ON
customer.custkey = orders.custkey
WHERE orders.custkey IS NULL

ИСПОЛЬЗОВАЛИ В:

- Q13, Q22

SQL

Пример запроса

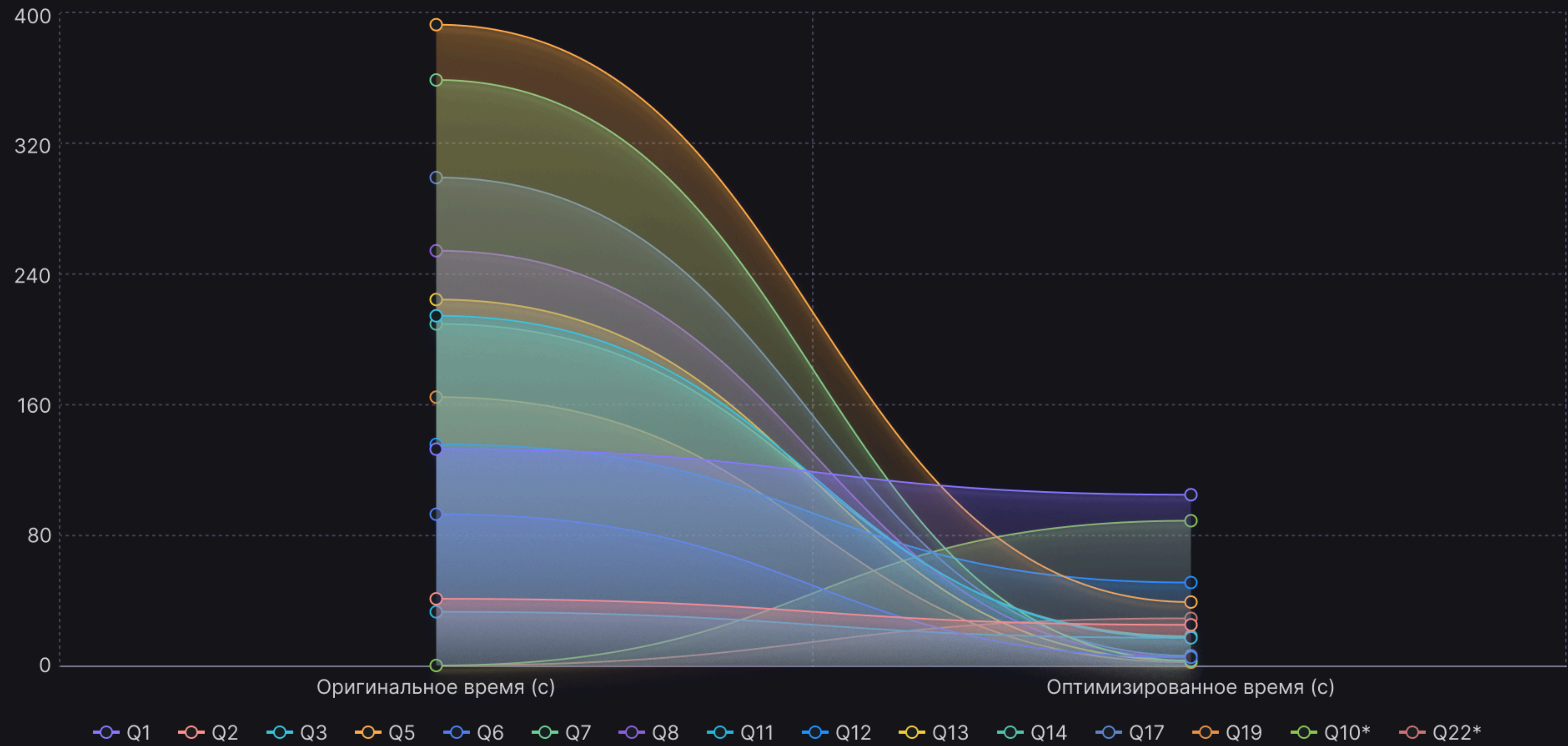
```
SELECT c.custkey  
FROM customer c  
LEFT JOIN orders o ON c.custkey =  
o.custkey  
WHERE o.custkey IS NULL;
```

Результаты оптимизации (Часть 1)

	Оригинальное время	Оптимизированное время	Оптимизация в %
Q1	2m13s	1m45s	21
Q2	41s	25s	39
Q3	3m35s	17s	92
Q5	6m34s	39s	90
Q6	1m33s	5s	95
Q7	6m	3s	99
Q8	4m15s	5s	98

Результаты оптимизации (Часть 2)

	Оригинальное время	Оптимизированное время	Оптимизация в %
Q11	33s	17s	49
Q12	2m16s	51s	62
Q13	3m45s	3s	99
Q14	2m30s	18s	91
Q17	5m	6s	98
Q19	2m45s	2s	99
Q10*	-	Выполняется	100
Q22*	-	Выполняется	100



Курск, КГУ



SAMSUNG

gopher go
data



Спасибо
преподавателям!

- Кривонос Алексей Владимирович,
Овсянников Александр Владимирович
- Организаторы хакатона кафедры ПОиАИС

Спасибо за внимание!