Запросы в KUMA (примеры)

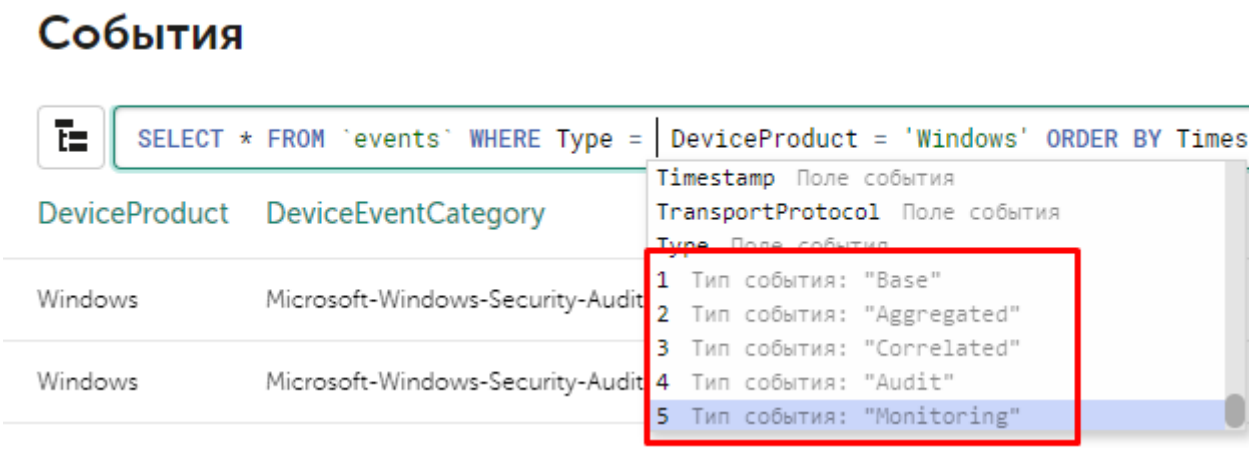
> Описание функций ClickHouse для работы с запросами:
> https://clickhouse.com/docs/ru/sql-reference/functions/

> Запрос из интерфейса пробрасывается в БД с добавлением границ временного промежутка и выбранных тенантов, пример: `... AND (Timestamp >= 1715689595208 AND Timestamp <= 1715689895208) AND (TenantID IN ('a1fbde7a-76d3-4bbc-a769-82126b41b56f')) ORDER BY ...`

https://www.youtube.com/embed/FnlT2hh_ACk?si=ktOOgrB_Q7rBL5wm

Базовые запросы

Типы событий в KUMA



Подсчет событий по полю

```
SELECT count(ID) as count_num, DeviceVendor
FROM `events`
```

```
GROUP BY DeviceVendor
ORDER BY count_num DESC LIMIT 250
```

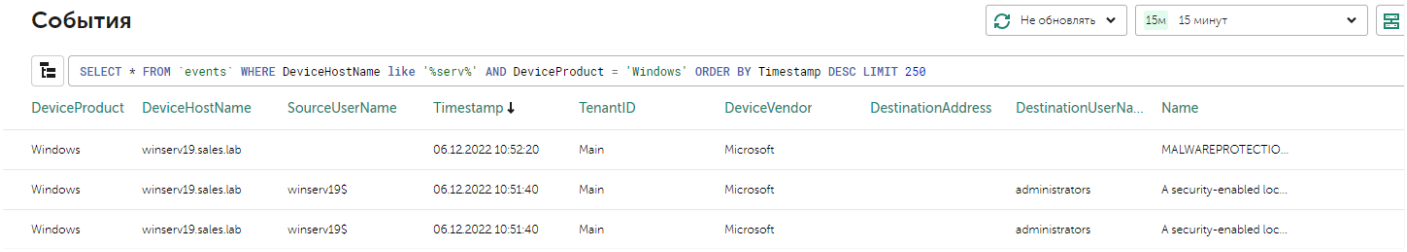
Выполнение математических операций и сравнений

```
SELECT DeviceProduct, SourceUserName, round(sum(BytesIn)/1024, 2) as KiloBytes
FROM `events`
WHERE BytesIn > '0' OR BytesOut > '0'
GROUP by SourceUserName, DeviceProduct
ORDER BY KiloBytes DESC LIMIT 250
```

Иногда бОльшую производительность дает условие со скобками `WHERE (BytesIn > '0' OR BytesOut > '0')`

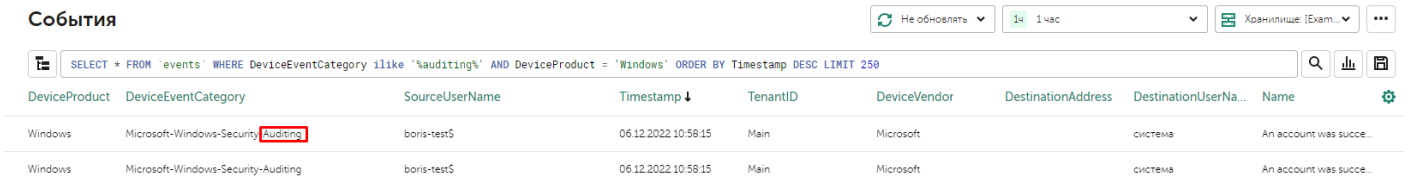
Поиск событий по подстроке (регистрозависимый) с условием И

```
SELECT *
FROM `events`
WHERE DeviceHostName like '%serv%' AND DeviceProduct = 'Windows'
ORDER BY Timestamp DESC LIMIT 250
```



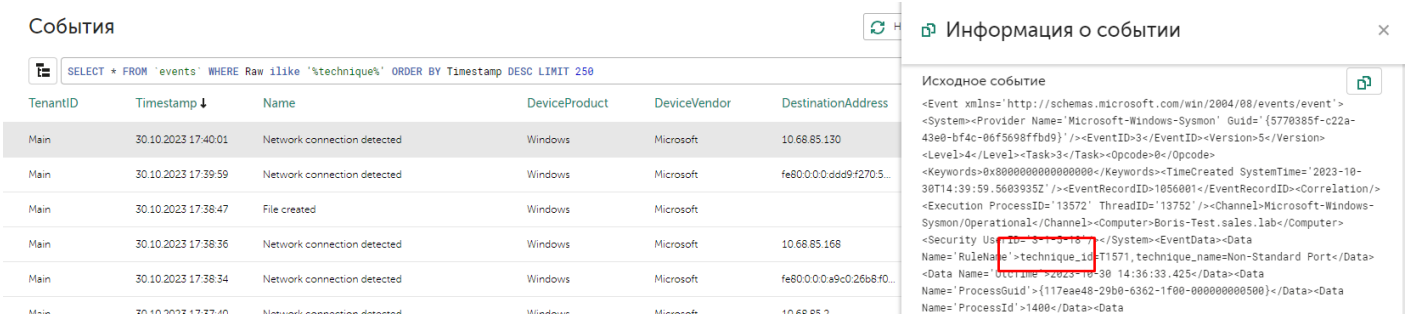
Поиск событий по подстроке (регистроНЕзависимый)

```
SELECT *
FROM `events`
WHERE DeviceEventCategory ilike '%auditing%' AND DeviceProduct = 'Windows'
ORDER BY Timestamp DESC LIMIT 250
```



Поиск по исходному событию

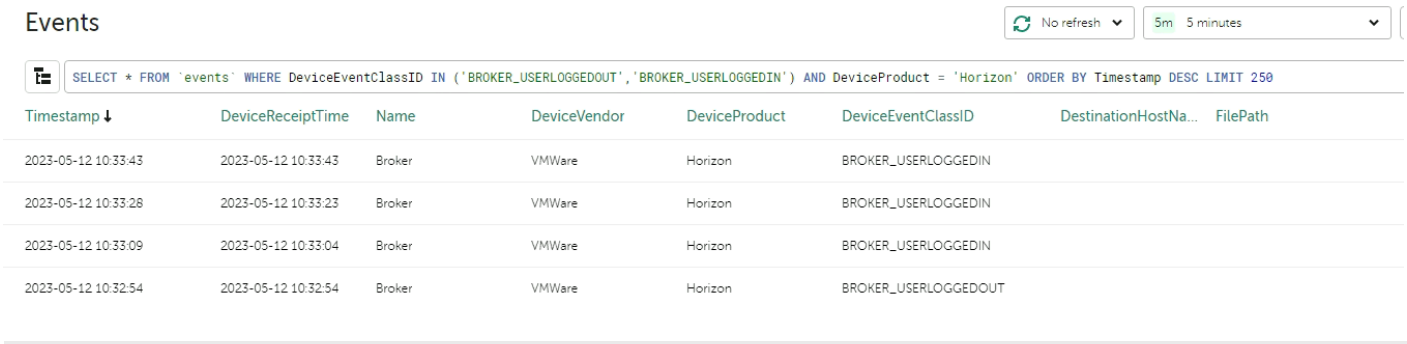
```
SELECT *
FROM `events`
WHERE Raw ilike '%technique%'
ORDER BY Timestamp DESC LIMIT 250
```



Поиск событий по нескольким значениям, вместо OR можно использовать IN

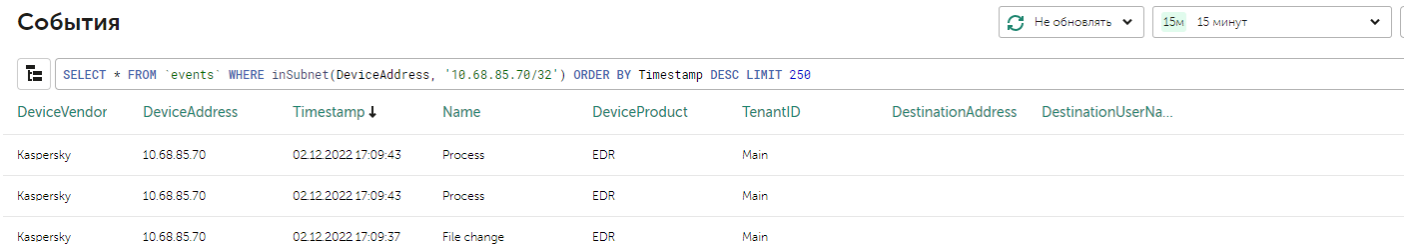
```
SELECT *
FROM `events`
```

```
WHERE DeviceEventClassID IN ('BROKER_USERLOGGEDOUT', 'BROKER_USERLOGGEDIN')
AND DeviceProduct = 'Horizon'
ORDER BY Timestamp DESC LIMIT 250
```



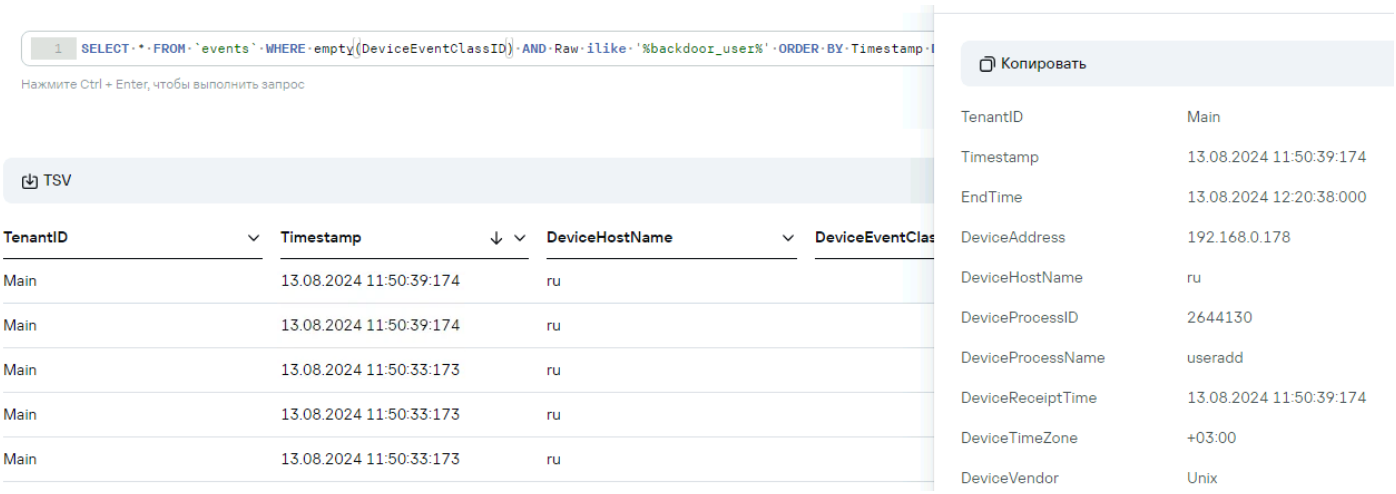
Поиск событий по подсети

```
SELECT *
FROM `events`
WHERE inSubnet(DeviceAddress, '10.68.85.70/32')
ORDER BY Timestamp DESC LIMIT 250
```



Поиск событий по отсутствующему полю

```
SELECT *
FROM `events`
WHERE empty(DeviceEventClassID) AND Raw ilike '%backdoor_user%'
ORDER BY Timestamp DESC LIMIT 250
```



Проверка работы обогащения

Обогащение событий Активами

```
SELECT *
FROM `events`
WHERE (DeviceAssetID != '' OR SourceAssetID != '' OR DestinationAssetID != '')
ORDER BY Timestamp DESC LIMIT 250
```

Обогащение событий с LDAP

```
SELECT *
FROM `events`
WHERE (SourceAccountID != '' OR DestinationAccountID != '')
ORDER BY Timestamp DESC LIMIT 250
```

Обогащение событий данными из TI

```
SELECT *
FROM `events`
WHERE NOT TI=""
```

```
ORDER BY Timestamp DESC LIMIT 250
```

Работа с Extra полем

Поиск событий по полю Extra содержащие ключ

```
SELECT *
FROM `events`
WHERE visitParamHas(Extra, 'memUsage')
ORDER BY Timestamp DESC LIMIT 250
```

Raw	{"cpu1":0,"cpu2":0,"cpu3":0,"cpu4":0,"diskBlock":4096,"diskSize":26056703,"diskUsage":8194425,"memBlock":65536,"memSize":65519,"memSizeKB":4193260,"memUsage":61367,"sysName":"dc-01.sales.lab","sysUpTime":571326915}\n
Extra	diskBlock: 4096 diskSize: 26056703 diskUsage: 8194425 memBlock: 65536 memSize: 65519 memSizeKB: 4193260 memUsage: 61367

Поиск событий по полю Extra содержащие ключ с определенным значением

```
SELECT *
FROM `events`
WHERE visitParamExtractString(Extra, 'memUsage') = '61367'
ORDER BY Timestamp DESC LIMIT 250
```

```
SELECT *
FROM `events`
WHERE JSONExtractString(Extra, 'memUsage') = '61367'
ORDER BY Timestamp DESC LIMIT 250
```

Type	Base
Raw	{"cpu1":0,"cpu2":1,"cpu3":0,"cpu4":0,"diskBlock":4096,"diskSize":26056703,"diskUsage":8194425,"memBlock":65536,"memSize":65519,"memSizeKB":4193260,"memUsage":61367,"sysName":"dc-01.sales.lab","sysUpTime":571333918}\n
Extra	diskBlock: 4096 diskSize: 26056703 diskUsage: 8194425 memBlock: 65536 memSize: 65519 memSizeKB: 4193260 memUsage: 61367

Поиск событий по полю Extra НЕ содержащие ключ с определенным значением

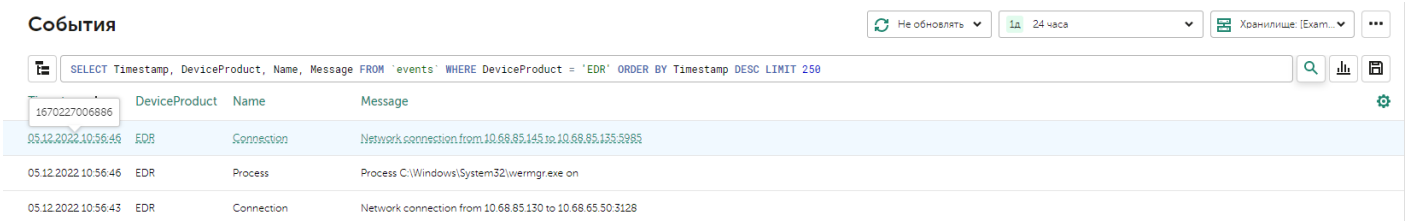
```
SELECT *
FROM `events`
WHERE visitParamHas(Extra, 'memUsage')
AND NOT visitParamExtractString(Extra, 'memUsage') = '61367'
ORDER BY Timestamp DESC LIMIT 250
```

Type	Base
Raw	{"cpu1":1,"cpu2":1,"cpu3":1,"cpu4":0,"diskBlock":4096,"diskSize":26056703,"diskUsage":8194885,"memBlock":65536,"memSize":65519,"memSizeKB":4193260,"memUsage":61693,"sysName":"dc-01.sales.lab","sysUpTime":571362433}
Extra	diskBlock: 4096 diskSize: 26056703 diskUsage: 8194885 memBlock: 65536 memSize: 65519 memSizeKB: 4193260 memUsage: 61693

Работа со временем

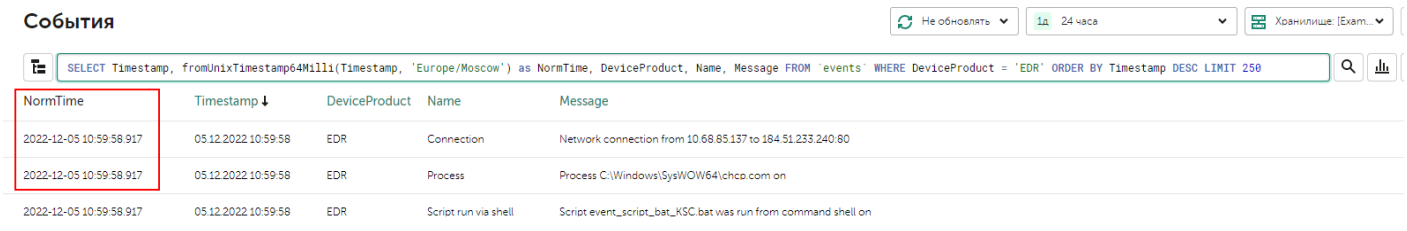
Timestamp по умолчанию в формате epoch time числа с милисекундами (UnixTimestamp)

> Список тайм зон: https://en.wikipedia.org/wiki/List_of_tz_database_time_zones



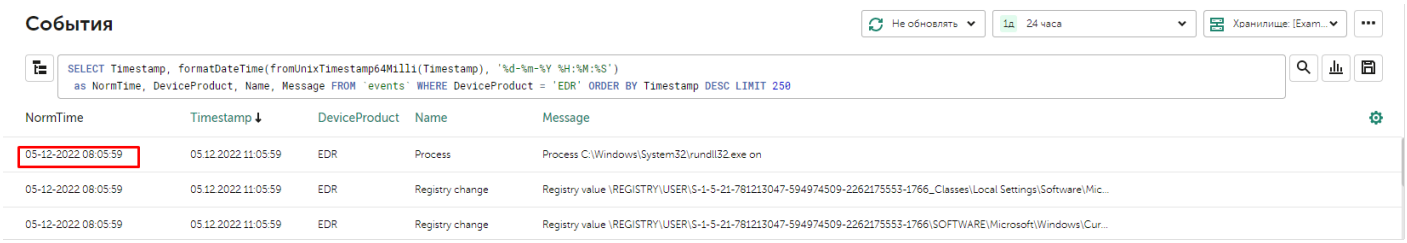
Задание таймзоны

```
SELECT Timestamp, fromUnixTimestamp64Milli(Timestamp, 'Europe/Moscow') as NormTime,
DeviceProduct, Name, Message
FROM `events`
WHERE DeviceProduct = 'EDR'
ORDER BY Timestamp DESC LIMIT 250
```



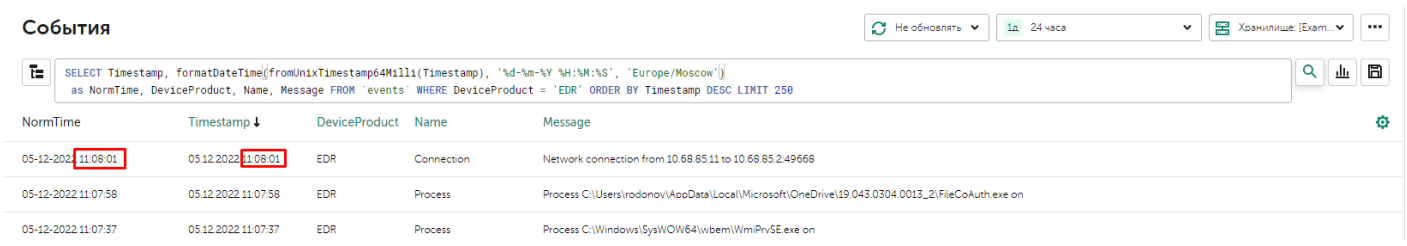
Указание своего формата времени

```
SELECT Timestamp, formatDateTime(fromUnixTimestamp64Milli(Timestamp), '%d-%m-%Y %H:%i:%S') as
NormTime,
DeviceProduct, Name, Message
FROM `events`
WHERE DeviceProduct = 'EDR'
ORDER BY Timestamp DESC LIMIT 250
```



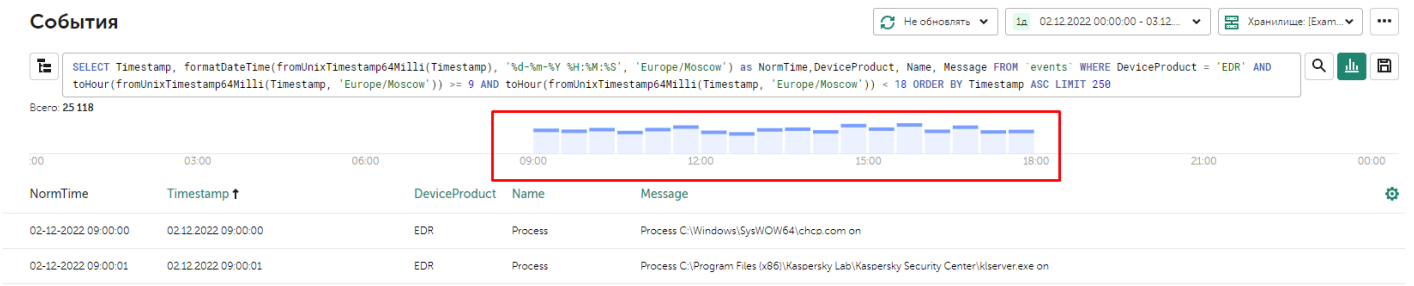
Свой формат времени с таймзоной

```
SELECT Timestamp,
formatDateTime(fromUnixTimestamp64Milli(Timestamp), '%d-%m-%Y %H:%i:%S', 'Europe/Moscow') as
NormTime,
DeviceProduct, Name, Message
FROM `events`
WHERE DeviceProduct = 'EDR'
ORDER BY Timestamp DESC LIMIT 250
```



События в "рабочее время" с 9 до 18

```
SELECT Timestamp,
formatDateTime(fromUnixTimestamp64Milli(Timestamp), '%d-%m-%Y %H:%i:%S', 'Europe/Moscow') as NormTime,
DeviceProduct, Name, Message
FROM `events`
WHERE DeviceProduct = 'EDR' AND toHour(fromUnixTimestamp64Milli(Timestamp, 'Europe/Moscow')) >= 9
AND toHour(fromUnixTimestamp64Milli(Timestamp, 'Europe/Moscow')) < 18
ORDER BY Timestamp ASC LIMIT 250
```

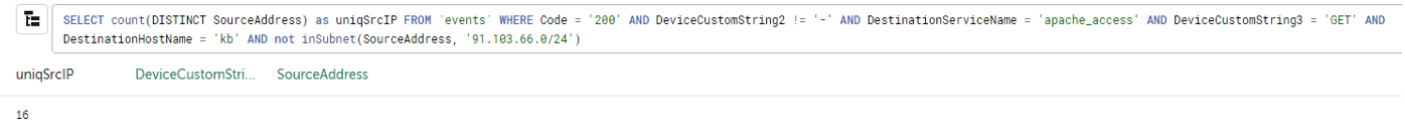


Подсчет события по фильтру за период вчерашнего дня с 00 до 04 часов с группировкой по SourceAddress

```
SELECT count(ID) as Count_Num, SourceAddress, groupArray(fromUnixTimestamp64Milli(Timestamp)) as time_stm
FROM `events`
WHERE DeviceEventClassID = '4624' AND DeviceProduct = 'Windows'
AND Timestamp >= toUnixTimestamp(toStartOfDay((now() - INTERVAL 1 DAY)))*1000
AND Timestamp <= toUnixTimestamp(toStartOfDay((now() - INTERVAL 1 DAY)) + INTERVAL 4 HOUR)*1000
GROUP BY SourceAddress LIMIT 250
```

Подсчет уникальных адресов посетителей kb.kuma-community.ru

```
SELECT count(DISTINCT SourceAddress) as uniqSrcIP
FROM `events`
WHERE Code = '200' AND DeviceCustomString2 != '-' AND DestinationServiceName = 'apache_access'
AND DeviceCustomString3 = 'GET' AND DestinationHostName = 'kb'
AND not inSubnet(SourceAddress, '91.103.66.0/24')
```



Можно использовать аналогичную более быструю функцию:

```
SELECT uniq(SourceAddress) as uniqSrcIP
FROM `events`
WHERE Code = '200' AND DeviceCustomString2 != '-' AND DestinationServiceName = 'apache_access'
AND DeviceCustomString3 = 'GET' AND DestinationHostName = 'kb'
AND not inSubnet(SourceAddress, '91.103.66.0/24')
```

Подсчет уникальных адресов посетителей kb.kuma-community.ru по дням

```
SELECT count(DISTINCT SourceAddress) as `metric`,
formatDateTime(fromUnixTimestamp64Milli(Timestamp), '%d-%m-%Y', 'Europe/Moscow') as value
FROM `events`
WHERE Code = '200' AND DeviceCustomString2 != '-' AND DestinationServiceName = 'apache_access'
AND DeviceCustomString3 = 'GET' AND DestinationHostName = 'kb'
AND not inSubnet(SourceAddress, '91.103.66.0/24')
GROUP BY value
```

Events

No refresh | 11d 2023-10-22 00:00:00 - 2023... | Storage: [OOTB] St...

```
SELECT count(DISTINCT SourceAddress) as `metric`, formatDateTime(fromUnixTimestamp64Milli(Timestamp), '%d-%m-%Y', 'Europe/Moscow') as value FROM `events` WHERE Code = '200' AND DeviceCustomString2
!= '-' AND DestinationServiceName = 'apache_access' AND DeviceCustomString3 = 'GET' AND DestinationHostName = 'kb' AND not inSubnet(SourceAddress, '91.103.66.0/24') GROUP BY value
```

metric	value
147	26-10-2023
140	27-10-2023
115	24-10-2023
103	25-10-2023
120	31-10-2023
166	30-10-2023
39	28-10-2023
33	29-10-2023
145	23-10-2023
22	22-10-2023
115	01-11-2023

Подсчет среднего количесва переданных байт посетителей kb.kuma-community.ru по источнику IP с 00:00 по 08:00

```
SELECT avg(toInt32(BytesOut)), SourceAddress
FROM `events`
WHERE BytesOut!= 0 AND Timestamp >= toUnixTimestamp(toStartOfDay((now() - INTERVAL 1 DAY)))*1000
AND Timestamp <= toUnixTimestamp(toStartOfDay((now() - INTERVAL 1 DAY)) + INTERVAL 8 HOUR)*1000
GROUP BY SourceAddress DESC LIMIT 250
```

Events

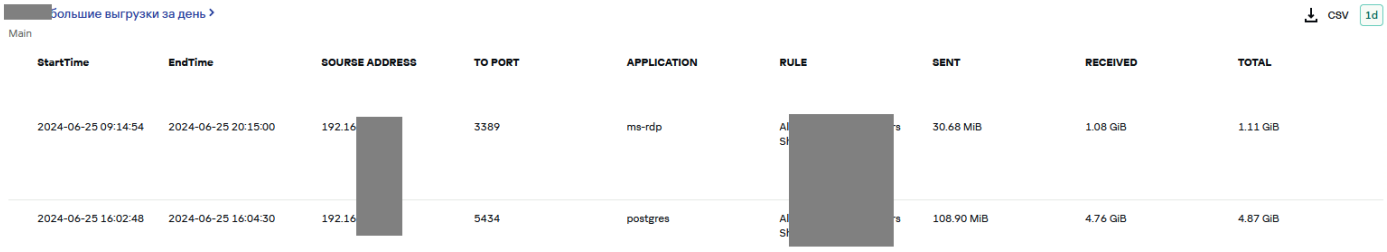
No refresh | 2d 2023-11-22 00:00:00 - 2023-... | Storage: STOR

```
SELECT avg(toInt32(BytesOut)), SourceAddress FROM `events` WHERE BytesOut!= 0 AND Timestamp >= toUnixTimestamp(toStartOfDay((now() - INTERVAL 1 DAY)))*1000 AND Timestamp <= toUnixTimestamp
(toStartOfDay((now() - INTERVAL 1 DAY)) + INTERVAL 8 HOUR)*1000 GROUP BY SourceAddress DESC LIMIT 250
```

avg(toInt32(BytesOut))	SourceAddress
126	::1
14365.5	207.[redacted]153
9499.666666666666	185.[redacted]3.50
58497.565217391304	89.1[redacted]42
43983.75	46.1[redacted].148
91397.61538461539	176.[redacted].33
5874	213.[redacted]3.221

Отображение количества переданных и принятых байт по внутренним адресам более 1 Гб

```
SELECT StartTime , EndTime , SourceAddress AS `SOURSE ADDRESS`, DestinationPort AS `TO PORT`, ApplicationProtocol AS `APPLICATION`, DeviceCustomString1 AS `RULE`, formatReadableSize(BytesIn) AS `SENT`, formatReadableSize(BytesOut) AS `RECEIVED`, formatReadableSize(FlexNumber1) AS `TOTAL`
FROM `events`
WHERE FlexNumber1 > 1000000000
GROUP BY StartTime, EndTime, SourceAddress, DestinationPort, ApplicationProtocol, DeviceCustomString1, BytesIn, BytesOut, FlexNumber1
ORDER BY `EndTime` DESC LIMIT 250
```



StartTime	EndTime	SOURSE ADDRESS	TO PORT	APPLICATION	RULE	SENT	RECEIVED	TOTAL
2024-06-25 09:14:54	2024-06-25 20:15:00	192.16[redacted]	3389	ms-rdp	[redacted]	30.68 MiB	1.08 GiB	1.11 GiB
2024-06-25 16:02:48	2024-06-25 16:04:30	192.16[redacted]	5434	postgres	[redacted]	108.90 MiB	4.76 GiB	4.87 GiB

Экстра запросы

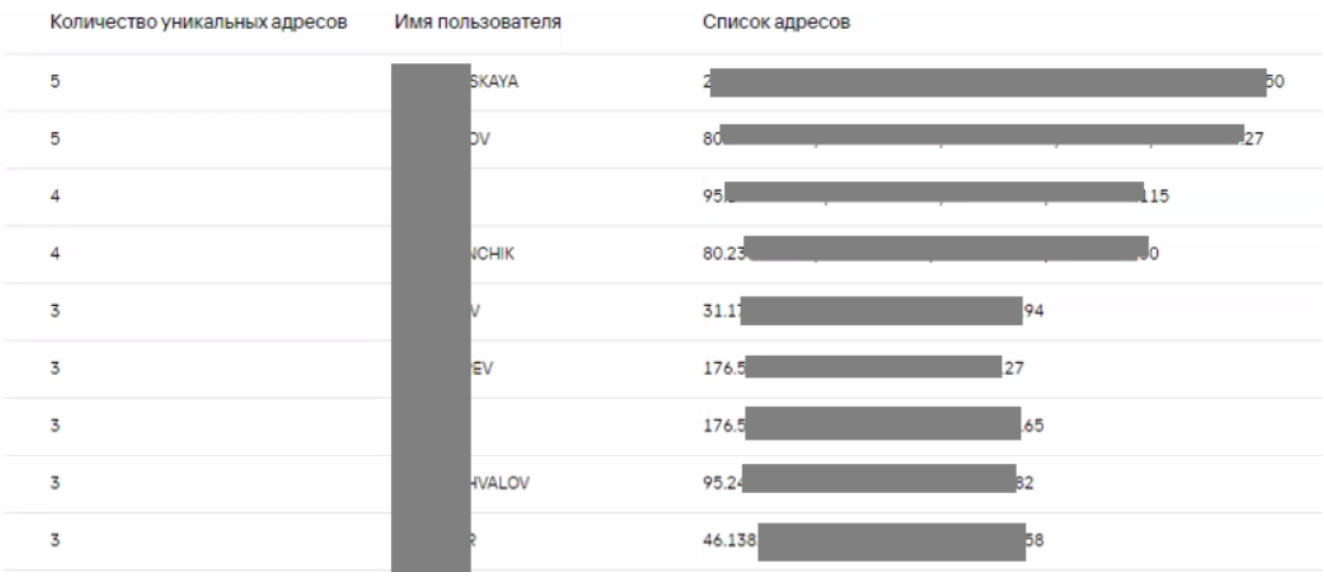
Склейка по времени подключений пользователей с одного адреса и на один VPN сервер

```
SELECT SourceUserName, SourceAddress, SourceHostName, groupArray(FlexString1) as time
FROM `events`
WHERE SourceProcessName = 'Create session'
GROUP BY SourceUserName, SourceAddress, SourceHostName LIMIT 10
```

Склейка различных адресов подключений от одного пользователя на VPN сервере

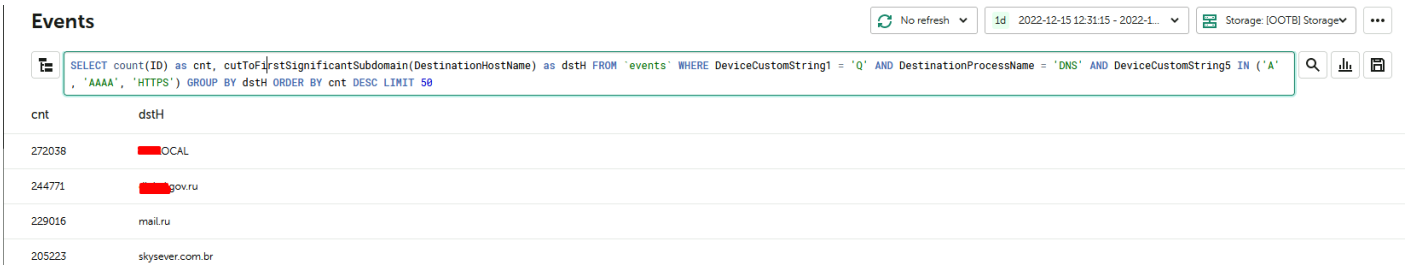
> Оператор HAVING доступен с версии KUMA 3.0

```
SELECT uniq(SourceAddress) as "Количество уникальных адресов", DestinationUserName  as "Имя пользователя", groupUniqArray(SourceAddress) as  "Список адресов"
FROM `events`
WHERE DeviceProduct = 'Ngate' AND Name = 'Create session'
GROUP BY "Имя пользователя"
HAVING "Количество уникальных адресов" > 1
ORDER BY "Количество уникальных адресов" DESC
LIMIT 10
```



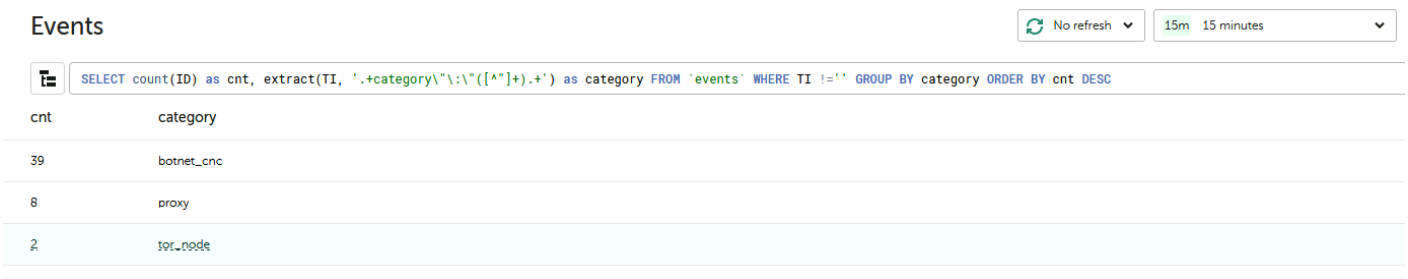
Обрезка доменов до второго уровня и условие с несколькими запросами

```
SELECT count(ID) as cnt, cutToFirstSignificantSubdomain(DestinationHostName) as dstH
FROM `events`
WHERE DeviceCustomString1 = 'Q' AND DestinationProcessName = 'DNS'
AND DeviceCustomString5 IN ('A', 'AAAA', 'HTTPS')
GROUP BY dstH
ORDER BY cnt DESC LIMIT 50
```



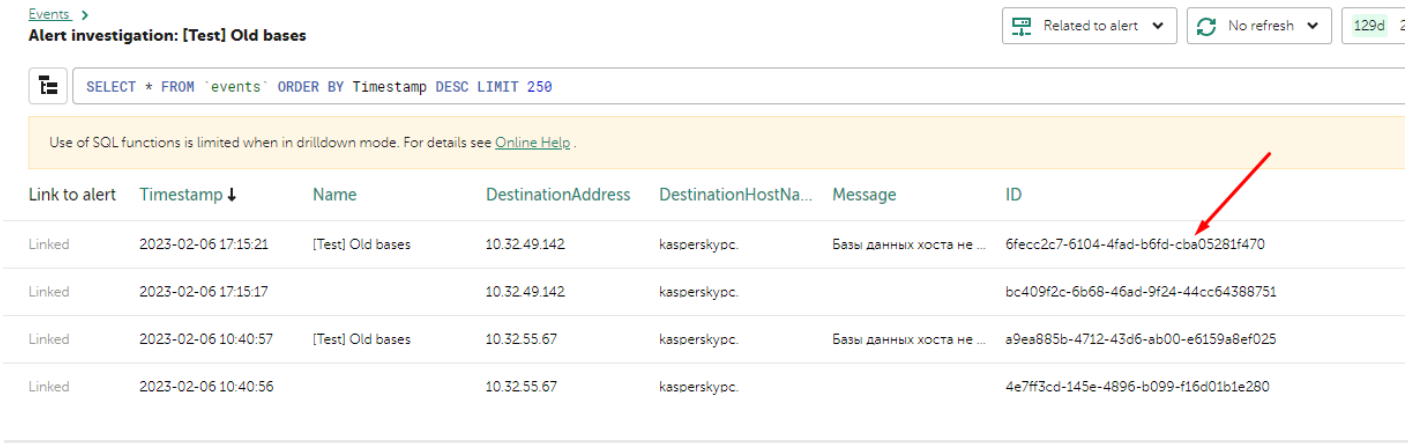
Группировка событий по категории из TI (Regex)

```
SELECT count(ID) as cnt, extract(TI, '.+category\"\:\"([^"]+).+') as category
FROM `events`
WHERE TI !=''
GROUP BY category
ORDER BY cnt DESC
```



Найти и вывести все базовые события конкретного корреляционного события по его ID

```
SELECT *
FROM `events`
WHERE ID IN (
  SELECT arrayJoin(splitByChar(',', replaceRegexpAll(JSON_QUERY(BaseEvents, '$[*].ID'), '\\[|\\]|"| ', ''))) as TestID
  FROM `events`
  WHERE (Type = 3) AND (ID = '25e4eae3-ad6d-4114-b99e-019de3574d16')
)
```



Запросы к системным таблицам

У ClickHouse есть системные таблицы (https://clickhouse.com/docs/en/operations/system-tables).

Одна из них - query_log (https://clickhouse.com/docs/en/operations/system-tables/query_log), содержит интересную информацию о запросах.

Чтобы обратиться к данным таблицами можно использовать клиент ClickHouse, который располагается по пути `/opt/kaspersky/kuma/clickhouse/bin/client.sh`

Можно заходить в консоль ClickHouse с помощью данного клиента, а также выполнять запросы через аргументы клиента, а также формировать запросы с помощью curl. В примерах ниже будет рассмотрен именно первый способ.

Вывод запросов с сортировкой по длительности

В примере ниже представлен запрос с сортировкой по длительности выполнения запросов за конкретный день

```
SELECT *
FROM system.query_log
WHERE query_kind = 'Select' AND current_database = 'kuma' AND event_date = '2024-01-22'
AND query_duration_ms != 0
ORDER BY query_duration_ms DESC LIMIT 10
```

Вывод запросов, завершившихся ошибкой

```
SELECT *
FROM system.query_log
WHERE type = 'ExceptionWhileProcessing' AND current_database = 'kuma'
LIMIT 10
```

Revision #24
Created 10 August 2023 11:21:22 by Boris Rzr
Updated 16 September 2024 13:17:21 by Boris Rzr