

???????? ???? ? ???? ???? ???? (Data Mining)

В отличие от потоковой корреляции, работающей в режиме реального времени, Data Mining правила позволяют с помощью языка SQL и функций ClickHouse ([примеры](#) запросов, почти все возможно использовать) распознавать и анализировать события, сохраненных в хранилище KUMA (можно указать и конкретный спейс хранилища).

???????? ???? ?

Выполнение **SQL-запросов к ClickHouse** и приведение результатов к формату нормализованных событий KUMA происходит на уровне **Core** с помощью новой сущности **DataMiningRule** и встроенного механизма **Scheduler**. Полученные результаты преобразуются в события и распространяются по корреляторам через стандартный API, который также используется коллекторами.

Важно, что:

- **запрос к ClickHouse выполняется строго один раз** за указанный период расписания;
- сформированный результат может быть направлен **одному или нескольким корреляторам**, а также в другие подсистемы в будущем;
- корреляторы, получающие данные, **могут принадлежать различным тенантам**, что обеспечивает гибкость и масштабируемость архитектуры.

Таким образом, Data Mining правила открывают возможность выявлять долгие и сложные цепочки активности - те, которые невозможно или крайне трудно обнаружить только средствами потоковой корреляции.

???????????? ? ???? ?

Плюсы	Минусы
Снижение нагрузки на корреляторы — отсутствует необходимость хранить большие объёмы временных данных в оперативной памяти	Увеличение нагрузки на хранилище данных , для которого постоянные сложные запросы не являются целевой нагрузкой.
Кросстенантное обнаружение — правило может передавать результаты нескольким корреляторам разных тенантов.	Риск тяжелых запросов — неэффективный SQL может существенно нагрузить кластер.

<p>Гибкость создания правил — возможность строить корреляцию напрямую на основе SQL-запросов.</p>	<p>Отложенное обнаружение — аналитика работает постфактум, поэтому алерт приходит позже, чем при потоковой корреляции.</p>
<p>Распределённое выполнение запросов — нагрузка обрабатывается кластером хранилища, а не одним сервером корреляции.</p>	
<p>Поддержка поиска аномалий и долгих сценариев атак — отклонения от нормы, тренды, девиации, редкие последовательности.</p>	
<p>Устойчивость к задержкам и несинхронности событий — если события приходят с опозданием или в неправильном порядке (например, правила по Golden Ticket), анализ всё равно будет корректным.</p>	
<p>Сохранность состояния при рестарте — бакеты и промежуточные данные не сбрасываются при перезагрузке коррелятора.</p>	

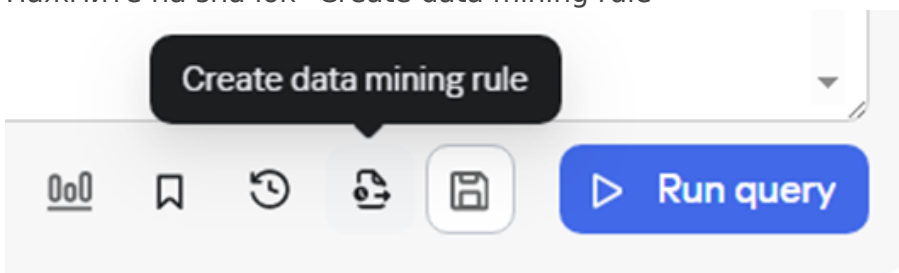
????????? ? ??????????? ??????????

Процесс создания правила можно разделить на три этапа:

1 ??????. ?????????? ?????????????????????? ??????? Data Mining ??????????

Создать правило можно двумя способами (из SQL-запроса в разделе Поиск по событиям):

- Формируем SQL-запрос в разделе Поиска по событиям (тестируем гипотезы, проводим атаку на полигоне, наполняем БД синтетическими событиями)
- Необходимо проверить что запрос выполняется, не вешает базу, возвращает осмысленный результат, который можно далее анализировать с помощью коррелятора
- Нажмите на значок "Create data mining rule"



- Далее открывается окно Создания правила, автоматически заполнится сам запрос, глубина и частота запуска. Заполнится маппинг полей из запроса в поля KUMA

Create data mining rule

Name*

Tenant*

Tags

SQL query execution frequency*

Depth ⓘ

SQL ⓘ

```
1 SELECT
2   DeviceHostName,
3   DeviceProcessName,
4   min(if(DeviceCustomString1 = 'red', Timestamp,
5   NULL)) AS red_event_time,
6   'red' AS new_status,
7   'green' AS old_status
8 FROM 'events'
9 WHERE DeviceFacility = 'agent'
   AND DeviceAction = 'service status changed'
```

Description ⓘ

Mapping

<input type="checkbox"/> Source field	<input type="checkbox"/> Event field
<input type="checkbox"/> <input type="text" value="DeviceHostName"/>	<input type="checkbox"/> <input type="text" value="DeviceHostName"/>

- Здесь необходимо вписать название, выбрать тенант, дозаполнить поля и создать правило.

Второй способ (создать правило как ресурс):

- В **Ресурсах - Правила сбора** и анализа данных Создать правило

Правила сбора и анализа данных

Все Мои

☆ Избранное

☰ Coverage Test

+ Добавить
📄 Дублировать
🗑 Удалить
🏷 Теги
📄 Показа

<input type="checkbox"/>	Название	Путь до ресурса	Последн...
<input type="checkbox"/>	Demo_Data_Mining	Shared	10.01.2025 11:2...

Всего 1 / Выбрано 0

- В правиле указать:
 - **Интервал (частота) выполнения SQL-запроса** можно указать в минутах, часах и днях (минимум 1 минута)
 - **SQL-запрос** должен содержать функцию агрегации ([примеры](#)) и/или группировку (GROUP BY) данных с обязательным указанием ограничения LIMIT (от 1 до 10 000)

Каждое выполнение такого правила происходит в виде запроса в Хранилище, а это значит неосторожным движением в виде частого или тяжелого правила можно нагрузить базу больше чем хотелось бы

В примере рассматривается запрос на основе событий Windows по пользователям (DestinationUserName) событиям входа (EventID 4624) и выхода (EventID 4634) с расчетом среднего времени сессии пользователя за последние 24 часа.

Посмотреть SQL запрос (пример)

```
SELECT
  login_events.DestinationUserName AS destination_user_name,
  round(AVG(logout_events.logout_time - login_events.login_time)/1000) AS
avg_time_diff_s,
  COUNT(DISTINCT login_events.login_time) AS total_logins,
  COUNT(DISTINCT logout_events.logout_time) AS total_logouts,
  concat(
    toString(floor(avg_time_diff_s / 86400)), ' days, ',
    toString(floor((avg_time_diff_s % 86400) / 3600)), ' hours, ',
    toString(floor((avg_time_diff_s % 3600) / 60)), ' minutes, ',
    toString(avg_time_diff_s % 60), ' seconds'
  ) AS human_readable_diff
```

```

FROM
  (SELECT
    DestinationUserName,
    toUnixTimestamp(EndTime) AS login_time,
    FlexString1 AS logon_id
  FROM `events`
  WHERE DeviceEventClassID = '4624'
  AND EndTime >= now() - INTERVAL 24 HOUR
  AND DestinationUserName NOT LIKE '%$%') AS login_events
INNER JOIN
  (SELECT
    DestinationUserName,
    toUnixTimestamp(EndTime) AS logout_time,
    FlexString1 AS logon_id
  FROM `events`
  WHERE DeviceEventClassID = '4634'
  AND EndTime >= now() - INTERVAL 24 HOUR
  AND DestinationUserName NOT LIKE '%$%') AS logout_events

ON login_events.DestinationUserName = logout_events.DestinationUserName
AND logout_events.logon_id = login_events.logon_id

WHERE logout_events.logout_time >= login_events.login_time
GROUP BY login_events.DestinationUserName
ORDER BY avg_time_diff_s DESC
LIMIT 100

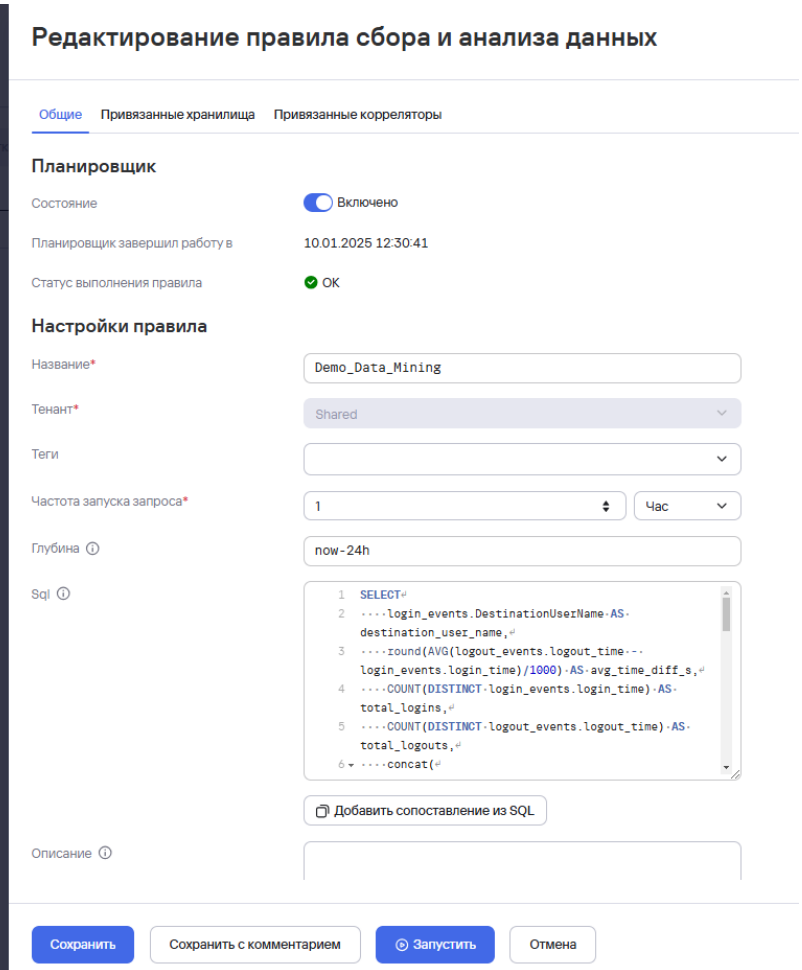
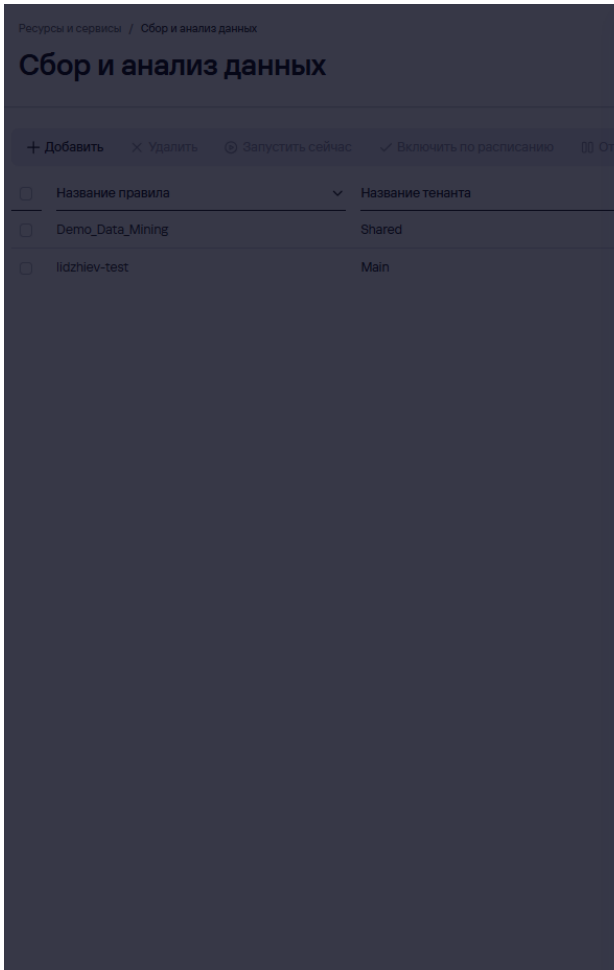
```

- **Добавить маппинг** (сопоставление) по полям запроса и модели KUMA

+ Добавить сопоставление		Удалить
Исходное поле	Поле события	Подпись
<input type="checkbox"/> destination_user_name	DestinationUserName	
<input type="checkbox"/> avg_time_diff_s	DeviceCustomNumber1	avg_time_diff_s
<input type="checkbox"/> total_logins	DeviceCustomString1	total_logins
<input type="checkbox"/> total_logouts	DeviceCustomString2	total_logouts
<input type="checkbox"/> human_readable_diff	DeviceCustomString3	human_readable_diff

2 ?????. ?????????? ????????????????

- Перейти в раздел **Ресурсы - Сбор и анализ данных** добавить планировщик по ранее созданному правилу
- Открыть правило и установить связи:
 - **Привязать хранилище** по которому будет осуществляться поиск на вкладке **Привязанные хранилища**
 - **Привязать коррелятор** с соответствующим правилом корреляции для сработки на вкладке **Привязанные корреляторы**
- Для ручного запуска нажмите кнопку **Запустить**



- По результатам запроса на выходе будут сформированы базовые события, которые не будут сохранены. Далее необходимо создать простое правило корреляции, чтобы создать корреляционное событие и алерт на данное событие и привязать правило к нужным корреляторам

3 ???? ????????? simple ????????? ?? ???????????

В нашем случае правило ловит события, где время сессии меньше 5 секунд:

Редактирование правила корреляции



Общие **Селекторы** Действия Корреляторы Исключения

Параметры Локальные переменные

Параметры фильтра

Фильтр* ▼

Сохранить фильтр

- Если e: DeviceCustomNumber1Label = avg_time_diff_s
- Если не e: Type = 3
- Если e: DeviceCustomNumber1 < 5

Корреляционное событие выглядит следующим образом:

Алерты > DataMiningBoris- Win AVG Session less 5 sec Новый

Уровень важности: Назначить:

Информация об алерте

Уровень важности правила корреляции	Первое появление	Тенант
Низкий	10.01.2025 11:30:43	Main
Наивысшая важность категории активов	Последнее появление	Правило корреляции
Нет значения	10.01.2025 11:30:43	DataMiningBoris- Win AVG Session less 5 sec

Идентификатор алерта
e7a32232-c6ae-4f6a-bf83-1d66e9636915

Связанные события

Время ↓	Информация о событии	ТЕНАНТ
10.01.2025 11:30:43	DestinationUserName: kmg-idap, DeviceCustomString1: 193, DeviceCustomString2: 193, DeviceCustomString3: 0 days, 0 hours, 0 minutes, 0 seconds	Main
10.01.2025 11:30:43	DestinationUserName: myznikov, DeviceCustomString1: 1111, DeviceCustomString2: 1109, DeviceCustomString3: 0 days, 0 hours, 0 minutes, 3 seconds, DeviceCustomNumber1: 3	Main
10.01.2025 11:30:40	EndTime: 10.01.2025 11:30:40, DeviceTimeZone: +03:00, DestinationUserName: myznikov, DeviceCustomNumber1: 3, DeviceCustomNumber1Label: avg_time_diff_s, DeviceCustomString1: 1111, DeviceCustomString1Label: totalLogins, DeviceCustomString2: 1109, DeviceCustomString2Label: total_logouts, DeviceCustomString3: 0 days, 0 hours, 0 minutes, 3 seconds	Main

[Найти в событиях 1](#)

Информация о корреляционном событии

TenantName	Main
Timestamp	10.01.2025 11:30:43:274
Name	DataMiningBoris- Win AVG Session less 5 sec
StartTime	10.01.2025 11:30:40:180
EndTime	10.01.2025 11:30:40:180
DeviceProduct	KUMA
DeviceTimeZone	+03:00
DeviceVendor	Kaspersky
DestinationUserName	myznikov
DeviceCustomNumber1	3
DeviceCustomString1	1111
DeviceCustomString2	1109
DeviceCustomString3	0 days, 0 hours, 0 minutes, 3 seconds
CorrelationRule	DataMiningBoris- Win AVG Session less 5 sec
Service	[QOTB] Correlator
BaseEventCount	1
Priority	Низкий
Type	Correlated

Поля расширенной схемы событий

N.KL_CorrelationRulePriority	1
------------------------------	---

А событие на основе которого произошла сработка:

Уровень важности: Низкий Назначить: Не назначено Закрыть алерт Создать инцидент Привязать

Информация об алерте

Уровень важности правила корреляции: **Низкий**
Первое появление: 10.01.2025 11:30:43
Тенант: Main
Наивысшая важность категории активов: Нет значения
Последнее появление: 10.01.2025 11:30:43
Правило корреляции: [DataMiningBoris- Win AVG Session less 5 sec](#)
Идентификатор алерта: e7a32232-c6ae-4f6a-bf83-1d66e9636915

Связанные события

Время ↓	Информация о событии	Тенант
10.01.2025 11:30:43	DestinationUserName: ksmg-ldap, DeviceCustomString1: 193, DeviceCustomString2: 193, DeviceCustomString3: 0 days, 0 hours, 0 minutes, 0 seconds	Main
10.01.2025 11:30:43	DestinationUserName: myznikov, DeviceCustomString1: 1111, DeviceCustomString2: 1109, DeviceCustomString3: 0 days, 0 hours, 0 minutes, 3 seconds, DeviceCustomNumber1: 3	Main
10.01.2025 11:30:40	EndTime: 10.01.2025 11:30:40, DeviceTimeZone: +03:00, DestinationUserName: myznikov, DeviceCustomNumber1: 3, DeviceCustomNumber1Label: avg_time_diff_s, DeviceCustomString1: 1111, DeviceCustomString1Label: total_logins, DeviceCustomString2: 1109, DeviceCustomString2Label: total_logouts, DeviceCustomString3: 0 days, 0 hours, 0 minutes, 3 seconds	

Информация о событии

Копировать

Timestamp	10.01.2025 11:30:40:180
EndTime	10.01.2025 11:30:40:180
DeviceTimeZone	+03:00
DestinationUserName	myznikov
DeviceCustomNumber1	3
DeviceCustomNumber1Label	avg_time_diff_s
DeviceCustomString1	1111
DeviceCustomString1Label	total_logins
DeviceCustomString2	1109
DeviceCustomString2Label	total_logouts
DeviceCustomString3	0 days, 0 hours, 0 minutes, 3 seconds
DeviceCustomString3Label	human_readable_diff
Service	core
Type	Base

Еще пример:

Редактирование правила сбора и анализа данных

[Общие](#) [Привязанные хранилища](#) [Привязанные корреляторы](#)

Глубина 1ч now-1h

Sql

```
1 SELECT count(CASE WHEN DeviceEventClassID = '4625' THEN 1 END) as F, count(CASE WHEN DeviceEventClassID = '4624' THEN 1 END) as S, F/S as A, 'D20' as DeviceExternalID
2 FROM 'events'
3 HAVING A > 0.03
4 LIMIT 250
```

Добавить сопоставление из SQL

Описание

Rule detects when failed/success logins ratio exceeds 3%

+ Добавить сопоставление Удалить

Исходное поле	Поле события	Подпись
<input type="checkbox"/> F	DeviceCustomNumber1	Failed logins
<input type="checkbox"/> S	DeviceCustomNumber2	Success logins
<input type="checkbox"/> A	DeviceCustomFloatingPoint1	Ratio
<input type="checkbox"/> DeviceExternalID	DeviceExternalID	

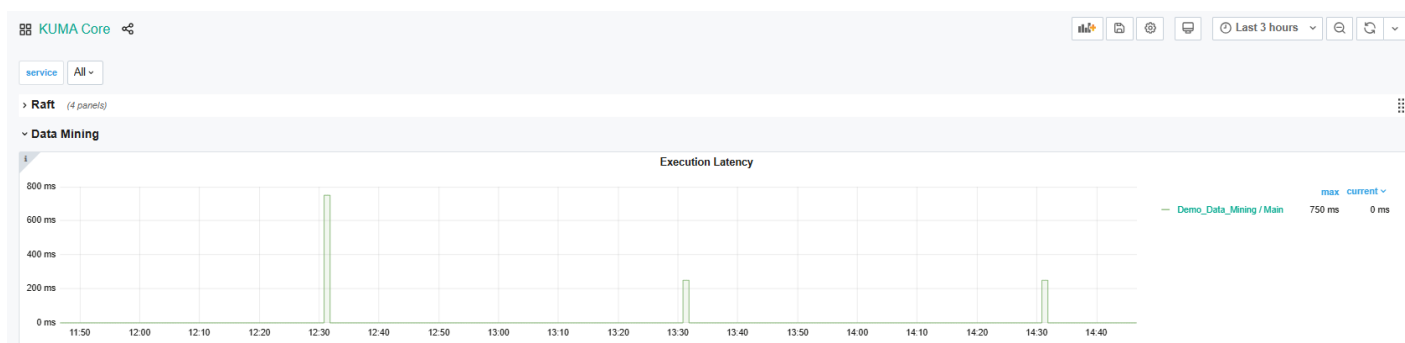
Параметры Локальные переменные

Параметры фильтра

Фильтр* Сохранить фильтр

Если e: =

Работу правил можно отслеживать с помощью метрик в разделе KUMA Core:



????? ?????????????? ???????

Использование Data Mining правил особенно актуально в ситуациях, когда классическая потоковая корреляция либо неэффективна, либо слишком ресурсоёмка. Рассмотрим основные практические сценарии:

- 1. Когда нужно обработать много событий за период**
Например, большое число неуспешных логинов за 5 минут или массовое сканирование портов.
Data Mining позволяет считать такие вещи в ClickHouse, не загружая корреляторы.
- 2. Когда нужно суммировать или усреднять значения**
Можно использовать агрегирующие функции SQL: `SUM()`, `AVG()` и т.д.
Пример: средний объём исходящего трафика или количество DNS-запросов.
Алерт срабатывает при превышении порога.
- 3. Когда нужен сравнительный анализ**
Например, сравнить количество событий за последний час с таким же периодом сутки назад.
- 4. Когда нужно работать со “скользящим” окном времени**
Анализировать события за период, независимо от того, с какой задержкой они

пришли.

5. **Операции которые невозможны на уровне цепочки событий** (Количественный прирост, анализ на схожесть, а не на одинаковость)

6. **Подсчет энтропии**

Определение при помощи энтропии, какие хосты генерируют одни и те же сработки для формирования исключений (рандомность, неожиданность цепочки)

????????? ???????

Рассмотрим более подробно на парочке примеров:

1. ??????? ?????????????? ?????????? ???????

Корреляционная логика, при которой требуется длительное накопление событий. Например:

- более **10 неуспешных попыток аутентификации под пользователем** `root`
- Берём **окно поиска 15 минут** и запускаем правило каждые **14 минут**.
- Так мы анализируем накопившиеся события и получаем результат без необходимости хранить все данные в памяти коррелятора.

```
1 SELECT
2 count(ID) AS cnt,
3 SourceHostName
4 FROM `events`
5 WHERE DeviceCustomString2 = 'ssh' AND DeviceEventClassID = 'USER_AUTH' AND DeviceVendor = 'Unix'
6 AND EventOutcome = 'failed' AND DestinationUserName = 'root'
7 GROUP BY SourceHostName
8 HAVING cnt > 10
```

Press Ctrl + Enter to run query

Groups (1)

Cards Table >>

TSV Search in group fields... 🔍

cnt: 161 —
SourceHostName: 192.168.100.47

Group events

TSV

Ti...	Device...	DeviceAction	Source...	Desti...
2025-11-12...	USER_AUTH	PAM:authenticati...	192.168.100.47	root
2025-11-12...	USER_AUTH	PAM:authenticati...	192.168.100.47	root

Частые неуспешные попытки входа под УЗ root

```

SELECT
  []SourceAddress,
  []SourceHostName,
  []DestinationUserName,
  min(Timestamp) as StartTime,
  max(Timestamp) as EndTime,
  []count(Distinct(DestinationServiceName)) as cnt_spn_names,
  []arrayCompact(groupUniqArray(DestinationServiceName)) as spn_names,
  arrayStringConcat(
    arrayMap(x -> '\' || x || '\'', groupUniqArray(Distinct(ID))),
    ', '
  ) AS BaseEventIDs,
  Count(*) as EventIDsCount,
  'SOCSH_Kerberoasting' as exID
FROM 'events'
WHERE
DeviceEventClassID = '4769' AND SourceUserID != '' AND NOT ( SourceAddress in ('', ':::1')
or SourceAddress like '127.0.0.%') AND NOT endsWith(DestinationUserName, '$')
GROUP BY SourceAddress, SourceHostName, DestinationUserName
HAVING cnt_spn_names > 10
LIMIT 100

```

2. ??????? ????????????? ????????? ?????????

Простой пример, где нужно суммировать данные и обнаруживать превышение порога.

- Суммируем исходящий трафик (`SUM(bytes_out)`) и группируем по адресу источника.
- В поле «**Глубина**» оставляем пусто — тогда нижняя граница интервала определяется автоматически как конец предыдущего запроса + 1.
- В поле «**Частота запуска**» ставим минимальное значение — **1 минута**.

Create data mining rule

Name*

Tenant*

Tags

SQL query execution frequency

Depth ⓘ

SQL ⓘ

```
1 SELECT sum(BytesOut) as Traffic, SourceAddress
2 FROM `events`
3 WHERE BytesOut!= 0 GROUP BY SourceAddress ORDER BY
SourceAddress DESC LIMIT 10
```

Description ⓘ

|

<input type="checkbox"/> Source field	Event field
<input type="checkbox"/> Traffic	<input type="text"/>
<input type="checkbox"/> SourceAddress	SourceAddress

В результате Scheduler каждую минуту запускает SQL-запрос с небольшим окном данных. Получается **скользящее окно**, которое постоянно обновляется и позволяет корректно работать даже при задержках в доставке событий и нарушении их порядка. Это как раз тот случай, когда **Data Mining правила способны сделать то, с чем потоковая корреляция справиться не может.**

???????? ?

1. ????????? TGS ????????? (Kerberoasting)

Kerberoasting

```
SELECT
[SourceAddress,
[SourceHostName,
```

```

[]DestinationUserName,
min(Timestamp) as StartTime,
max(Timestamp) as EndTime,
[]count(Distinct(DestinationServiceName)) as cnt_spn_names,
[]arrayCompact(groupUniqArray(DestinationServiceName)) as spn_names,
arrayStringConcat(
  arrayMap(x -> '\' || x || '\'', groupUniqArray(Distinct(ID))),
  ', '
) AS BaseEventIDs,
Count(*) as EventIDsCount,
'SOCSh_Kerberoasting' as exID
FROM 'events'
WHERE
DeviceEventClassID = '4769' AND SourceUserID != '' AND NOT ( SourceAddress in ('',':::1')
or SourceAddress like '127.0.0.%') AND NOT endsWith(DestinationUserName,'$')
GROUP BY SourceAddress, SourceHostName, DestinationUserName
HAVING cnt_spn_names > 10
LIMIT 100

```

2. ?????????????? ??????? ? ????????????????? ??????? (????????? ??????????)

Сканирование портов

```

SELECT
arrayStringConcat(arraySort(groupUniqArray(Distinct(DeviceProduct))), ', ' ) AS
DeviceProduct,
arrayStringConcat(arraySort(groupUniqArray(Distinct(DeviceAddress))), ', ' ) AS
DeviceAddresses,
SourceAddress,
SourceHostName,
SourceNtDomain,
DestinationAddress,
min(Timestamp) as StartTime,
max(Timestamp) as EndTime,
arrayStringConcat(arraySort(groupUniqArray(Distinct(DestinationPort))), ', ' ) AS
DeviceCustomString1,
arrayStringConcat(
  arrayMap(x -> '\' || x || '\'', groupUniqArray(Distinct(ID))),

```

```

', '
) AS DeviceCustomString2,
Count(*) as DeviceCustomNumber2,
Count(Distinct(DestinationPort)) as DeviceCustomNumber1,
'SOCSh_ScanPort' as exID
FROM `events`
WHERE
Type=1 and
(DestinationPort < 1024 or DestinationPort in
(1434,1521,3306,3389,5432,8080,9200,1352,1540,1541)) AND
SourcePort>1024 and DestinationPort!=0 and SourceAddress!='' and DestinationAddress!=''
AND
(isIPAddressInRange(SourceAddress, '192.168.0.0/16') or isIPAddressInRange(SourceAddress,
'10.0.0.0/8') or isIPAddressInRange(SourceAddress, '172.16.0.0/12')) AND
(isIPAddressInRange(DestinationAddress, '192.168.0.0/16') or
isIPAddressInRange(DestinationAddress, '10.0.0.0/8') or
isIPAddressInRange(DestinationAddress, '172.16.0.0/12'))
GROUP BY SourceAddress,SourceHostName,SourceNtDomain,DestinationAddress
HAVING DeviceCustomNumber1>=10
LIMIT 100

```

3. ??????? ?????????????????? ??????? (????????????? ??????????????)

Прирост корреляционных событий более 20% за сутки

```

SELECT
'CorrelationSplash' as ExternalId,
TenantID,
CorrelationRuleID,
CorrelationRuleName,
countIf(Timestamp between toUnixTimestamp64Milli(now64()) - 1*3600000 and
toUnixTimestamp64Milli(now64())) as today,
countIf(Timestamp between toUnixTimestamp64Milli(now64()) - 25*3600000 and
toUnixTimestamp64Milli(now64())-24*3600000) as yesterday,
round(today/yesterday,2) as k
FROM `events`
WHERE Type=3 and toDayOfWeek(now64())!=1
GROUP BY TenantID,CorrelationRuleID,CorrelationRuleName

```

```
HAVING yesterday > 20 and k>1.2
LIMIT 250
```

4. ???????????/?????? ????????

SQL запрос правила Password Spraying

```
SELECT
  [SourceAddress, SourceHostName,
  [min(Timestamp) as StartTime, max(Timestamp) as EndTime,
    count(Distinct(DestinationUserName)) as cnt_usernames, /*кол-во уникальных УЗ*/
    arrayCompact(groupUniqArray(DestinationUserName)) as spray_usernames, /*уникальные
сортированные имена УЗ, склеенные в строку*/
    arrayStringConcat(arrayMap(x -> '\' || x || '\', groupUniqArray(Distinct(ID))),', ')
  as BaseEventIDs, /*уникальные сортированные ID базовых событий, склеенные в строку*/
  [Count(*) as EventIDsCount,
  ['SOCSH_PasswordSpray' as exID
FROM
  'events'
WHERE
  DeviceEventClassID = '4625'
  AND DestinationNtDomain != ''
  AND NOT endsWith(DestinationUserName,'$')
GROUP BY SourceAddress, SourceHostName
HAVING cnt_usernames > 10
LIMIT 100
```

Severity: Medium Assign to: Unassigned Close alert Create incident Link

Details on alert

Correlation rule severity	First seen	Tenant
Medium	2025-11-11 14:31:09	Main
Max asset category severity	Last seen	Correlation rule
None	2025-11-13 13:52:02	[SOCSh] Retro-Araka Password Spraying
Alert ID	bcc7eaad-405e-4ad8-acd9- eff7f543016d	

Related events

Timestamp ↓	Event details
2025-11-13 13:52:02	ExternalID: SOCSH_PasswordSpray, FlexString2: d62d0ada-9989-4314-94d8-a4e4e678f338, FlexString2Label: TenantID, SourceAddress: 192.168.100.20, SourceHostName: dc1, StartTime: 2025-11-13 13:51:59, EndTime: 2025-11-13 13:51:59, DeviceCustomString1: 18, DeviceCustomString1Label: Count UserNames, DeviceCustomString2: ['test_svc2','kuma','test321','kdrills-test','test123','nsh','testcve','test','test123','nsh','testcve','test','repl_account','eav','devil666','testuser','devil777','администратор','test', DeviceCustomString4: 18, DeviceCustomString3: '02c42b71-d096-49fb-bf89-falcc1059274','20be62106b1-4dcb-8e9c-ea5a21187a80','dab4564f-740c-430e-995a-c8b6d8ad7dcb','e61fb620-0ac9-4a73-9969ec639e022','276e6ef3-6ef6-488d-b2db-098cd9ab68f4','e3e95d3c-ca16-42e9-8835-a061327751a8ca45132-c0cf-4218-acdd-945e7547496e','3b2bf822-741a-41de-bed2-e05197fe77e9','c3f4cb2a-b0c9-47b7-9f32-6a5feddccc26','a549fba7-37d7-4550-9590-9c2e6faf5b36','d3b6d261-2a20-4c75-b2ca-d878e909e1f1','ed74a2ca-1038-4df6-8d34-efaf2falbccd','576b5da7-93d6-4be4-9519-46d667cb87d5','8aada710-61b8-4372-8023-caa3edc0467a','8c873b27-bf1d-45fe-a00d-bd2ea13d2154','9d3f6ba9-3b04fbc-81f6-ab0347f5725c','54835774-4a60-4266-9fa6-300e1339ec7f','761f02a9-debd-4dd4-b144-6708bf4144c5', DeviceCustomString3Label: BaseEventIDs
2025-11-13 13:51:59	StartTime: 2025-11-13 13:48:05, EndTime: 2025-11-13 13:48:36, DeviceTimeZone: +03:00, SourceAddress: 192.168.100.20, SourceHostName: dc1, DeviceCustomString1: 18, DeviceCustomString1Label: Count UserNames, DeviceCustomString2: ['test_svc2','kuma','test321','kdrills-

Correlation event details

Copy Detailed view	
TenantName	Main
Timestamp	2025-11-13 13:52:02 :513
Name	[SOCSh] Retro-Araka Password Spraying
StartTime	2025-11-13 13:51:59 :348
EndTime	2025-11-13 13:51:59 :348
DeviceProduct	KUMA
DeviceTimeZone	+03:00
DeviceVendor	Kaspersky
SourceAddress	192.168.100.20
SourceHostName	dc1
DeviceCustomString1	18
DeviceCustomString1Label	Count UserNames
DeviceCustomString2	['test_svc2','kuma','test321','kdrills-test','test123','nsh','testcve','test','repl_account','eav','devil666','testuser','devil777','администратор','test_svc','xdrisc','casual_user','safsafsa']
DeviceCustomString3	'02c42b71-d096-49fb-bf89-falcc1059274','20be6218-06b1-4dcb-8e9c-ea5a21187a80','dab4564f-740c-430e-995a-c8b6d8ad7dcb','e61fb620-0ac9-4a73-9969ec639e022','276e6ef3-6ef6-488d-b2db-098cd9ab68f4','e3e95d3c-ca16-42e9-8835-a061327751a8ca45132-c0cf-4218-acdd-945e7547496e','3b2bf822-741a-41de-bed2-e05197fe77e9','c3f4cb2a-b0c9-47b7-9f32-6a5feddccc26','a549fba7-37d7-4550-9590-9c2e6faf5b36','d3b6d261-2a20-4c75-b2ca-d878e909e1f1','ed74a2ca-1038-4df6-8d34-efaf2falbccd','576b5da7-93d6-4be4-9519-46d667cb87d5','8aada710-61b8-4372-8023-caa3edc0467a','8c873b27-bf1d-45fe-a00d-bd2ea13d2154','9d3f6ba9-3b04fbc-81f6-ab0347f5725c','54835774-4a60-4266-9fa6-300e1339ec7f','761f02a9-debd-4dd4-b144-6708bf4144c5'
DeviceCustomString3Label	BaseEventIDs
DeviceCustomString4	18

SQL запрос правила Password Spraying с сохранением имен пользователей успешного и неудачного логина

```

SELECT
  SourceAddress, SourceHostName, StartTime,
  EndTime, failure_logins, success_logins, failed_usernames, success_usernames, exID
FROM (
  SELECT
    SourceAddress, SourceHostName, min(Timestamp) as StartTime, max(Timestamp) as
  EndTime,
  countIf(DeviceEventClassID = '4625') AS failure_logins,
  countIf(DeviceEventClassID = '4624') AS success_logins,
  arrayCompact(groupUniqArrayIf(DestinationUserName, DeviceEventClassID = '4625')) AS
  failed_usernames,
  arrayCompact(groupUniqArrayIf(DestinationUserName, DeviceEventClassID = '4624')) AS
  success_usernames,
  'SOCSh_PasswordSpray' as exID
FROM events
WHERE
  DeviceEventClassID IN ('4625', '4624')

```

```

AND DestinationNtDomain != ''
AND NOT endsWith(DestinationUserName, '$')
GROUP BY SourceAddress, SourceHostName)
WHERE failure_logins > 10
LIMIT 100

```

Events

```

11 - FROM (
12   SELECT
13     SourceAddress,
14     SourceHostName,
15     min(Timestamp) as StartTime,
16     max(Timestamp) as EndTime,
17     countIf(DeviceEventClassID = '4625') AS failure_logins,
18     countIf(DeviceEventClassID = '4624') AS success_logins,
19     arrayCompact(groupUniqArrayIf(DestinationUserName, DeviceEventClassID = '4625')) AS failed_usernames,
20     arrayCompact(groupUniqArrayIf(DestinationUserName, DeviceEventClassID = '4624')) AS success_usernames,

```

Press Ctrl + Enter to run query

Groups (2)

TSV

Search in group fields...

```

SourceAddress: 192.168.100.20
SourceHostName: dc1
StartTime: 2025-11-13 16:42:41.402
EndTime: 2025-11-13 16:43:09.126
failure_logins: 12
success_logins: 5
failed_usernames: test_svc2,devil666,devil777,test123,test321,nsh,testcve,testuser,casual_user,reprl_account,kidrills-test,safsa
success_usernames: eav,kuma,администратор,test_svc,xdrsoc
exID: SOCSh_PasswordSpray

```

Event details

exID	SOCSh_PasswordSpray
failed_usernames	test_svc2,devil666,devil777,test123,test321,nsh,testcve,testuser,casual_user,reprl_account,kidrills-test,safsa
failure_logins	12
success_logins	5
success_usernames	eav,kuma,администратор,test_svc,xdrsoc
StartTime	2025-11-13 16:42:41.402
EndTime	2025-11-13 16:43:09.126
SourceAddress	192.168.100.20
SourceHostName	dc1

6. ??????? ?????????? ??? ?????????????? ??????????????

Запрос на Подсчет энтропии, который может помочь для определения и внесения исключений в правила корреляции.

Если очень грубо, энтропия это показатель случайности и чем она ниже - тем ниже случайности попадания источника DeviceAddress в корреляционное событие

Иными словами, можно определить, какие одни и те же хосты попадают в одни и те же алерты на постоянной основе и после проверки внести их в исключения

Подсчет энтропии для определения исключений

```

SELECT
  CorrelationRuleID,
  CorrelationRuleName,
  entropy(DeviceAddress) as entr, /*"показатель случайности" попадающих DeviceAddress. Чем
ниже - тем меньше случайности*/
  count(*) as cnt, /*объем выборки энтропии (маленькая выборка не информативна)*/
  count(distinct(DeviceAddress)) as hosts, /*количество уникальных DeviceAddress*/
  'SOCSh_Entropy' as ExternalID

```

```

FROM events
WHERE Type=3 /*корр. события*/
GROUP BY CorrelationRuleID, CorrelationRuleName
HAVING
    hosts > 1 and /*хостов в результате больше 1 (иначе о энтропии речи быть не может)*/
    cnt > 10 /*количество событий достаточно для оптимальной оценки*/
ORDER BY entr ASC /*сортируем по принципу "наименее случайные последовательности"*/
LIMIT 250

```

Events

```

1 SELECT
2   CorrelationRuleID, CorrelationRuleName,
3   entropy(DeviceAddress) as entr, /*"показатель случайности" попадающих DeviceAddress. Чем ниже - тем меньше случайности*/
4   count(*) as cnt, /*объем выборки энтропии (маленькая выборка не информативна)*/
5   count(distinct(DeviceAddress)) as hosts /*количество уникальных DeviceAddress*/
6 FROM events
7 WHERE Type=3 /*корр. события*/
8 GROUP BY CorrelationRuleID, CorrelationRuleName
9 HAVING
10  hosts > 1 and /*хостов в результате больше 1 (иначе о энтропии речи быть не может)*/

```

Groups (1)

CorrelationRuleID: ec50f5cc-350e-4697-a96d-81944031f438

CorrelationRuleName: R224_Обор информации о системе

entr: 0.058647474848024954

cnt: 442

hosts: 2

Group events

Time	Device...	DeviceAction	Desti...
2025-11-13...			root
2025-11-13...			root
2025-11-13...			root
2025-11-13...			root

Correlation event details

TenantID	Main
SpaceID	KUMA Default
Timestamp	2025-11-13 00:05:15.414
Name	R224_Обор информации о системе
StartTime	2025-11-13 00:05:15.413
EndTime	2025-11-13 00:05:15.413
Message	Пользователь root на хосте alddc1.ald.soc-lab.local (192.168.100.36) попытался получить доступ к файлу с помощью команды /usr/sbin/CRON - f, где могут храниться данные по парольной политике.
DeviceAddress	192.168.100.36
DeviceEventCategory	openat_exit
DeviceHostName	alddc1.ald.soc-lab.local
DeviceProduct	KUMA
DeviceTimeZone	+03:00
DeviceVendor	Kaspersky
SourceProcessID	822

А также другие примеры:

1. Скачок событий/алертов со средств защиты
2. Большое количество DNS запросов с хоста

Пакет ресурсов Data Mining правил: [Shared 20251113 231513 DMRules](#)

Пароль к ресурсу:

????? ?????????????? ?????????? SQL ? ??????????

Здесь перечислены самые часто встречающиеся функции, которые используются в запросах:

1. `arrayStringConcat` - объединяет элементы массива в строку
2. `arrayCompact` - удаляет последовательные дублирующиеся элементы из массива
3. `distinct` - уникальные значения
4. `groupUniqArray` - собирает значения в массив
5. `arraySort` - сортирует массив

6. `arrayMap` - применяет выражение к каждому элементу массива и возвращает новый массив с результатами

Возможно использовать все функции, описанные в документации ClickHouse:

<https://clickhouse.com/docs/ru/sql-reference/functions>

А также набор специальных функций `enrich` и `lookup` в KUMA:

<https://support.kaspersky.com/help/KUMA/4.0/ru-RU/294927.htm>

Например:

1. Уникальные отсортированные имена пользователей, склеенные в строку.

```
arrayCompact(arraySort(groupUniqArray(DestinationUserName)))
```

```
spray_usernames: casual_user,devil666,devil777,eav,kldrills-test,kuma,nsh,repr_account,safsafsa,test,test123,test321,test_svc,t  
est_svc2,testcve,testuser,xdrsoc,администратор
```

2. Уникальные ID базовых событий, склеенные в строку

```
arrayStringConcat( arrayMap(x -> '\ ' || x || '\ ', groupUniqArray(Distinct(ID))), ', ') AS  
BaseEventIDs
```

```
BaseEventIDs: '02c42b71-d096-49fb-bf89-fa1cc1059274','20be6218-06b1-4dcb-8e9c-ea5a21187a80','dab4564f-740c-430e  
-995a-c8b6d8ad7dcb','e61fb620-0ac9-4a73-99eb-6e9ec6396022','276e6ef3-6ef6-488d-b2db-098cd9ab68f4','e3e95d3...
```

Revision #11

Created 2025-01-10 08:44:23 UTC by Boris RZR

Updated 2026-06-16 12:05:04 UTC by Boris RZR