

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

???? (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).

Для работы необходимо указать, рассмотрим на примере:

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

Ресурсы и сервисы / Правила сбора и анализа данных

Правила сбора и анализа данных

Поиск...

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

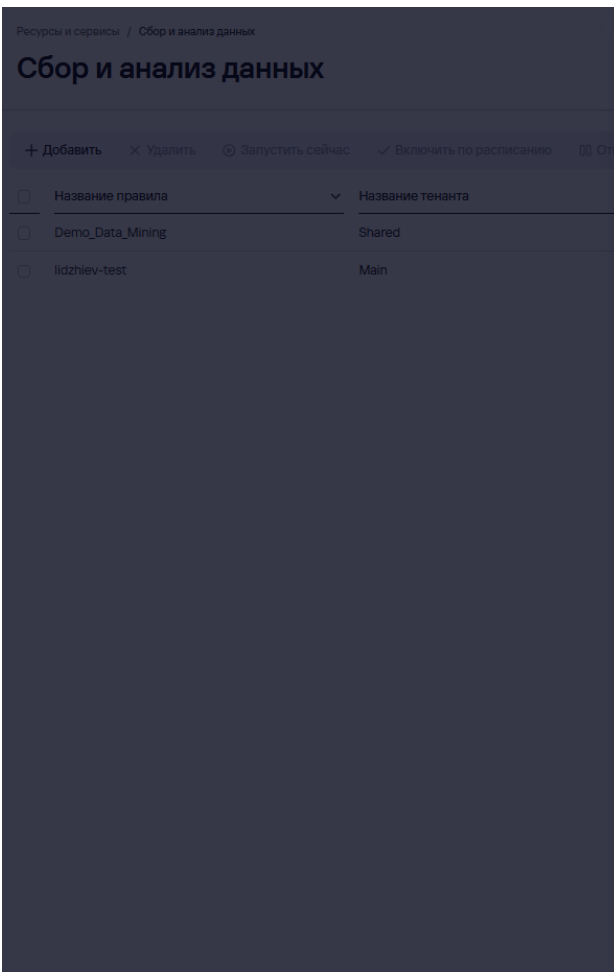
WHERE logout_events.logout_time >= login_events.login_time
GROUP BY login_events.DestinationUserName
ORDER BY avg_time_diff_s DESC
```

LIMIT 100

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

<input type="checkbox"/> 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

- В **Ресурсах - Сбор и анализ данных** добавить ранее созданное правило
- 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**



Редактирование правила сбора и анализа данных

Общие | Привязанные хранилища | Привязанные корреляторы

Планировщик

Состояние: Включено
Планировщик завершил работу в: 10.01.2025 12:30:41
Статус выполнения правила: ✔ OK

Настройки правила

Название*: Demo_Data_Mining
Тенант*: Shared
Теги:
Частота запуска запроса*: 1 Час
Глубина: now-24h
Sql:

```
1 SELECT
2 ...login_events.DestinationUserName AS
destination_user_name,
3 ...round(AVG(logout_events.logout_time --
login_events.login_time)/1000) AS avg_time_diff_s,
4 ...COUNT(DISTINCT login_events.login_time) AS
total_logins,
5 ...COUNT(DISTINCT logout_events.logout_time) AS
total_logouts,
6 ...concat(
```

Добавить сопоставление из SQL

Описание:

- По результатам запроса на выходе будут какие-то данные, которые не будут нигде сохраняться, но на них можно настроить правило корреляции. 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 е: DeviceCustomNumber1Label = avg_time_diff_s
- Если не е: Type = 3
- Если е: 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-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	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	[OOTB] Correlator
BaseEventCount	1
Priority	Низкий
Type	Correlated

Поля расширенной схемы событий

NKL_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	

Редактирование правила корреляции



Общие Селекторы Действия Корреляторы Исключения

Параметры Локальные переменные

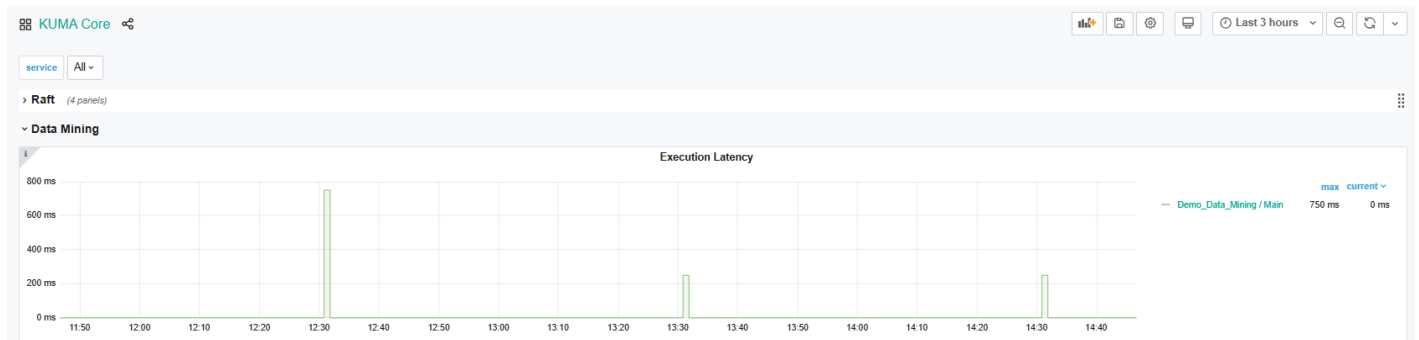
Параметры фильтра

Фильтр*

Сохранить фильтр

Если e: DeviceExternalID = D20

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



Revision #6

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

Updated 2025-02-03 11:49:01 UTC by Koala