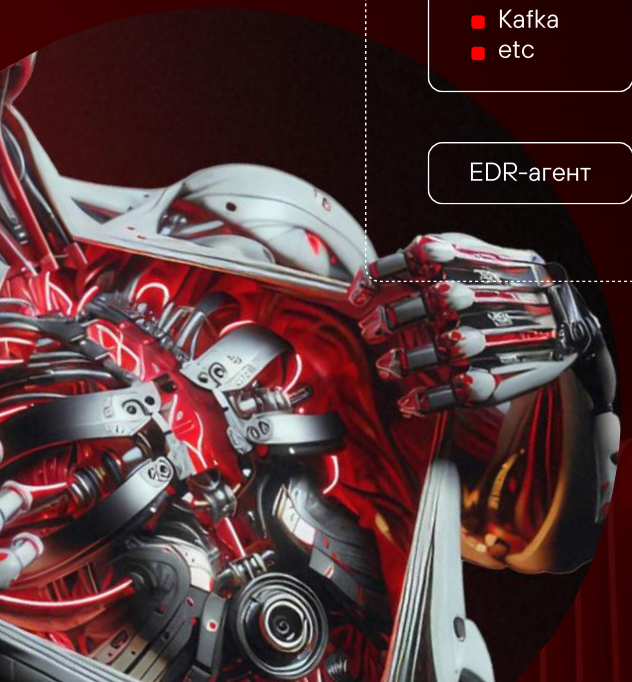
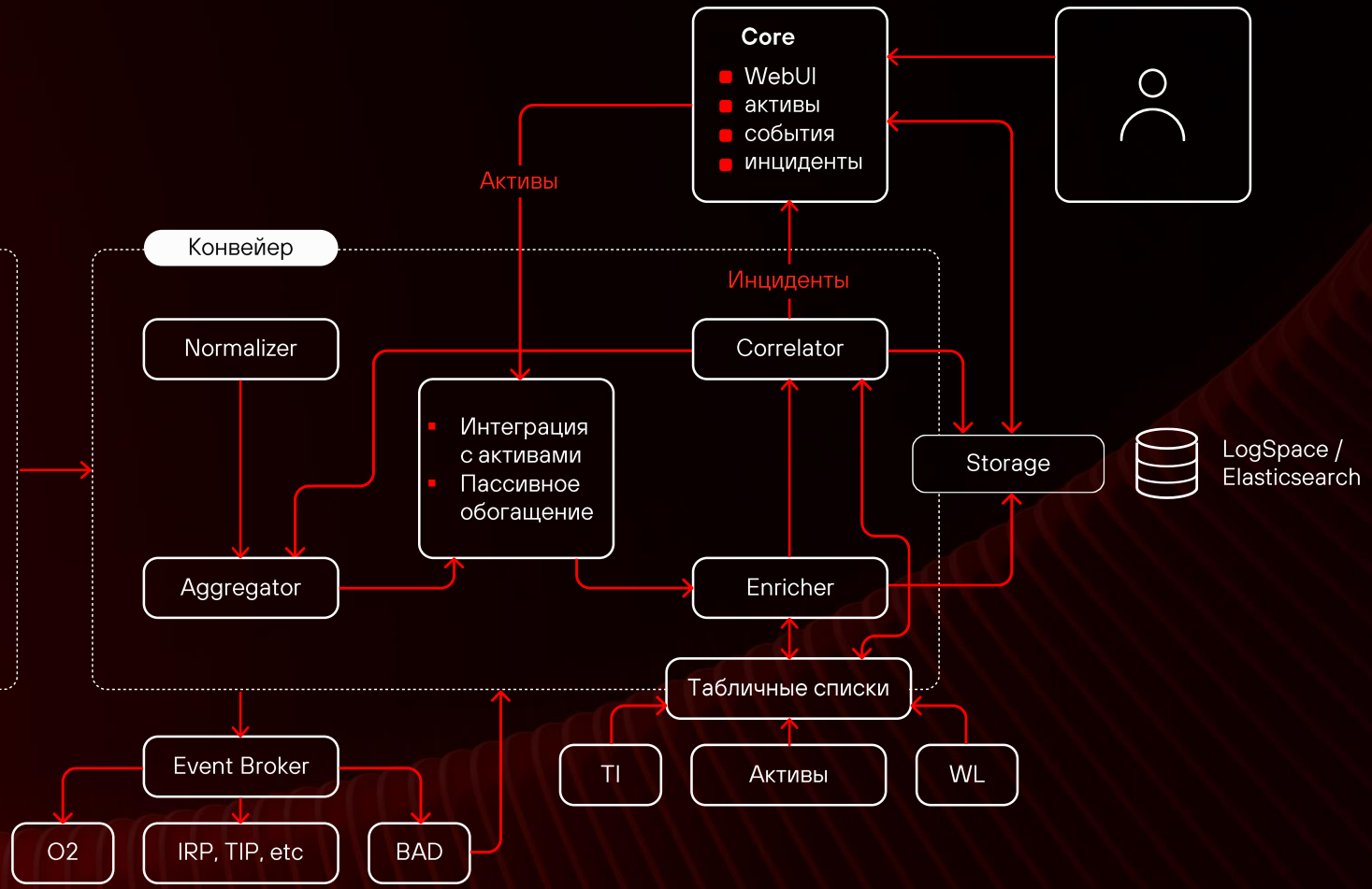
MaxPatrol SIEM:

оголяемся технологически.
Активоцентричность и хранение данных

Что под капотом у MaxPatrol SIEM



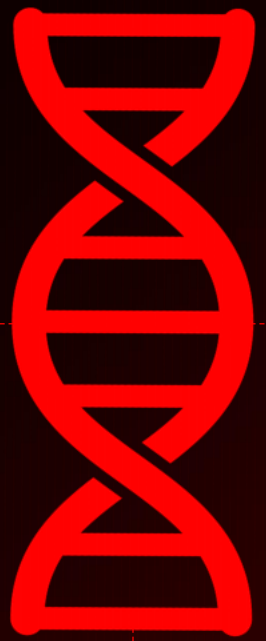


01

Работа с активами в MaxPatrol SIEM

Активы — часть ДНК продуктов семейства MaxPatrol

MaxPatrol VM



MaxPatrol SIEM

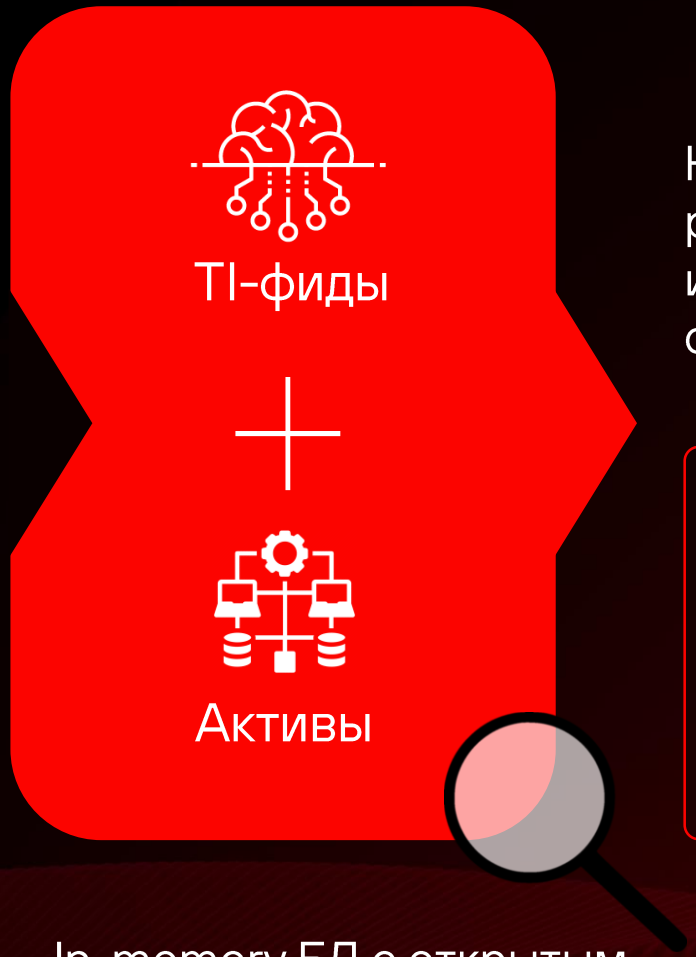
Активы

Зачем активы в SIEM-системе

- Определение контекста при обнаружении и работе с инцидентами (критичность, роли узлов, принадлежность к периметру, пользовательские атрибуты актива)
- Удобное управление доступом к событиям на базе групп активов
- Мониторинг источников

Контекст

Чтобы выявить атаки типа DCSync DCShadow

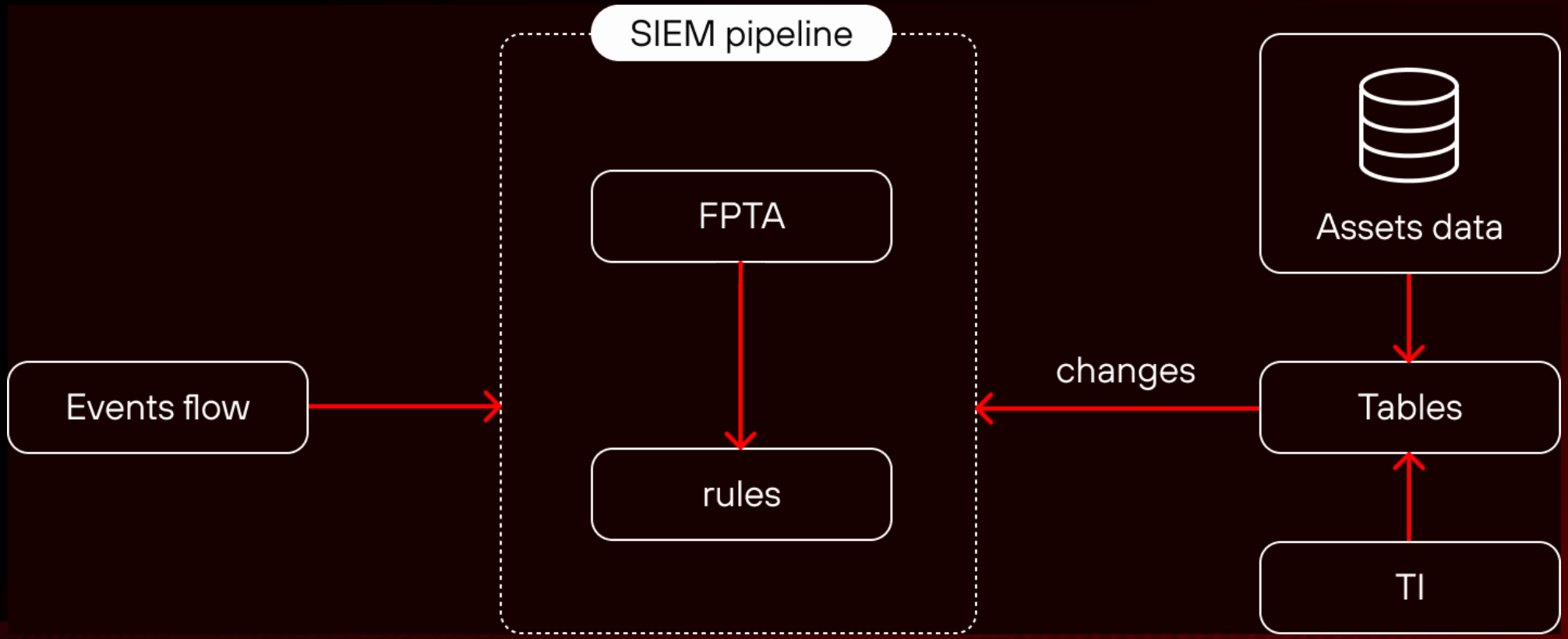


Необходимо знать, какие узлы выполняют роль контроллера домена в конкретной инфраструктуре. Контроллеры домена определяются PDQL-запросом:

```
filter(Host.HostRoles.Role = 'Domain
Controller') |
select(Host.Fqdn as fqdn_temp,
Host.@IpAddresses as ip, Host.@Id as id) |
filter(not ip in [127.0.0.1/8, ::1/128])
```

In-memory БД с открытым кодом FPTA (Fast Positive Tables)

Данные об активах в MaxPatrol SIEM



Привязка событий к активам

В событиях — множество адресов,
которые могут меняться.

В сканах активов — информация об узле с мегабайтами подробностей, среди которых есть эти самые адреса.

Основная задача — идентифицировать активы
по набору информации с адресами.

AssetId — это просто UUID
AssetId = Resolve(addreses[], props[])

Один из основных сервисов продукта — Assets Identity — суть которого сводится к имплементации функции Resolve

```
<?xml version="1.0"?>
<Event
    xmlns="http://schemas.microsoft.com/win/2004/08/events/event">
    <System>
        <Provider Name="Microsoft-Windows-Security-Auditing" Guid="{54849625-5478-4994-A5BA-3E3B0328C30D}"/>
        <EventID>4624</EventID>
        <TimeCreated SystemTime="2015-11-12T00:24:35.079785200Z"/>
        <Channel>Security</Channel>
        <Computer>dc-05.company.com</Computer>
        <Security/>
    </System>
    <EventData>
        <Data Name="IpAddress">10.0.43.217</Data>
        <Data Name="IpPort">50090</Data>
    </EventData>
</Event>
```

Привязка событий к активам

Задачи идентификации активов для потока сканов и потока событий имеют принципиально разную технологическую сложность

Сканы

~1000 в сутки

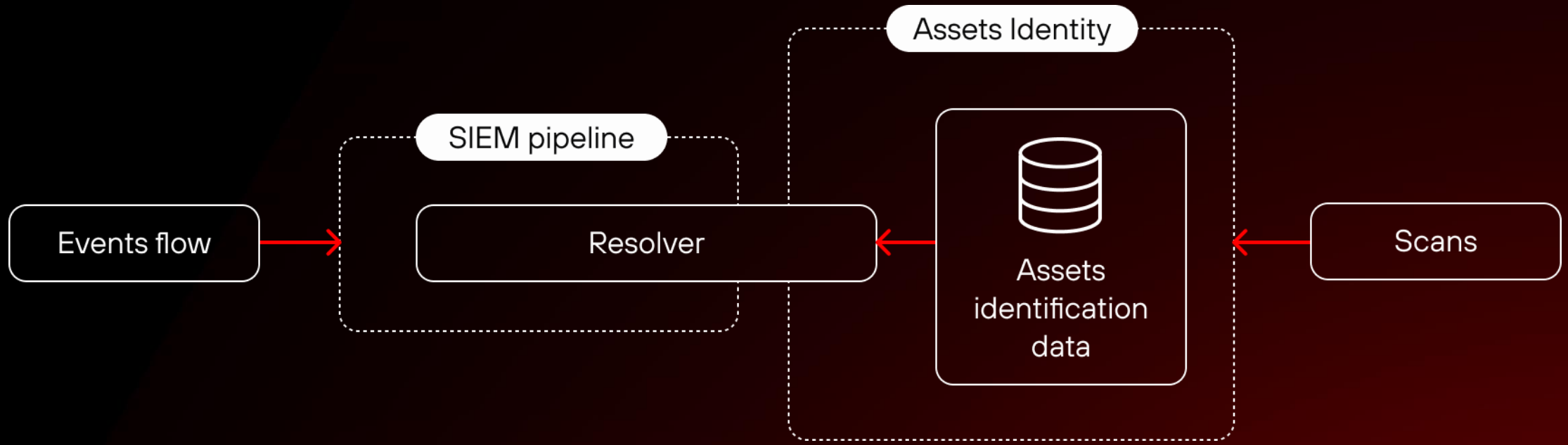
Подробные, несут много разной информации

События

~10 000 в секунду

Привязаны к нескольким активам, проще по структуре и менее насыщены информацией

Привязка событий к активам



Что делать с DHCP

«В компьютерных науках есть только две сложные проблемы – инвалидация кэша и придумывание названий»

Фил Карлтон

С DHCP IP-адрес актива может поменяться в непредсказуемый момент времени

Варианты решений:

- Логи записывает DHCP-сервер
- Логи может записывать DHCP-клиент
- Информация об изменении leases может быть извлечена из трафика при помощи DPI

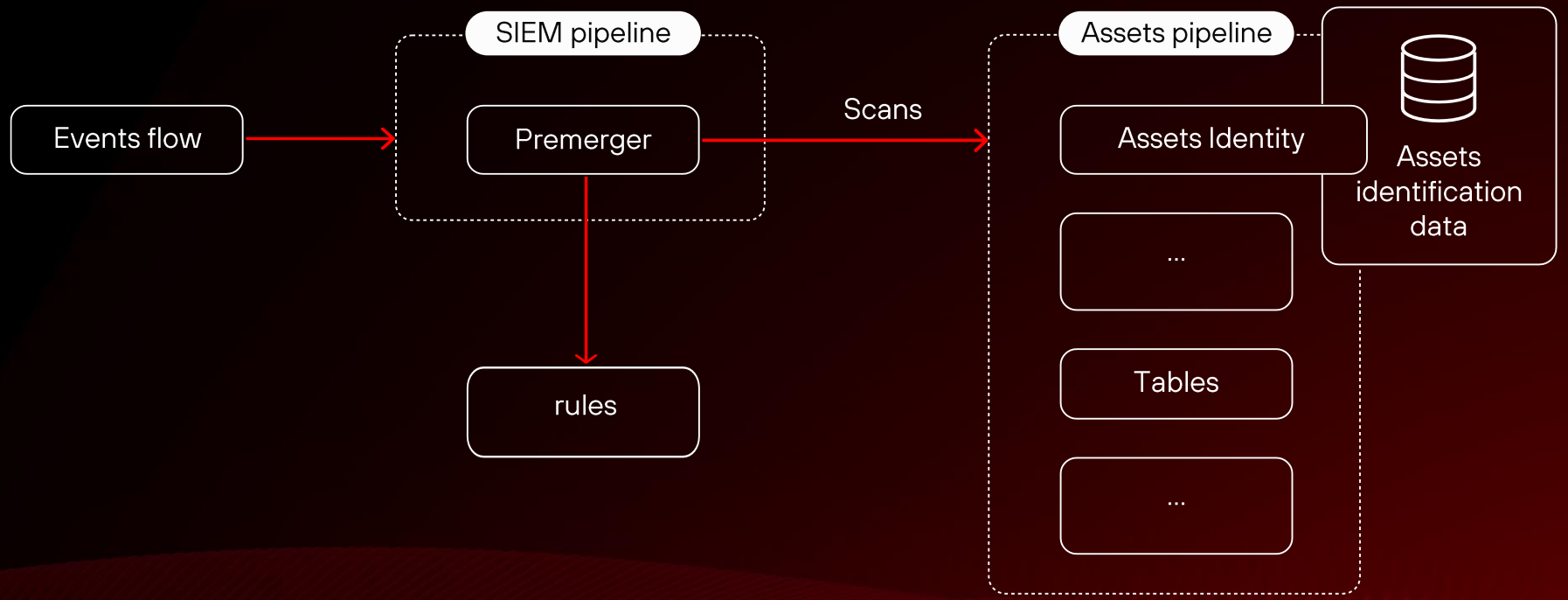
SIEM-система — источник данных об активах

Сканировать актив целиком — дорого
(хотя агентское сканирование частично решает и эту проблему)

Самая важная информация об активе меняется достаточно редко, а эти изменения обычно логируются

Отсюда возможности scanless scan:
пассивные детекты, etc

События меняют активы



Примеры пассивных детектов

DHCP приносит новый маппинг MAC-IP — формируем скан с двумя адресами и перестраиваем ключи идентификации

В wineventlog залогировалось событие 19 об установке KB на узел — меняем список KB на узле и закрываем соответствующие уязвимости

В журналах wineventlog есть определенные события — считаем узел контроллером домена Active Directory

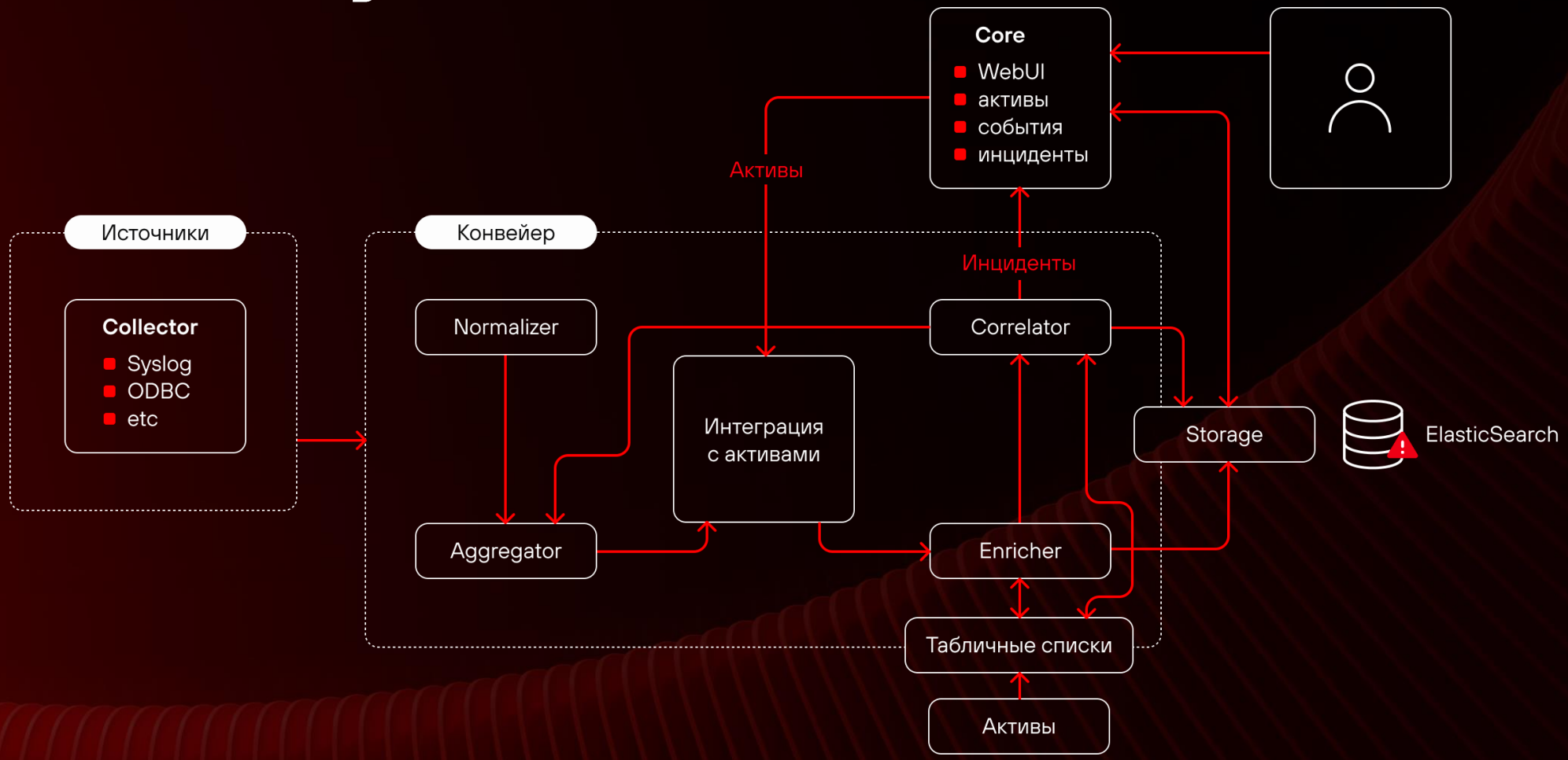
Эксперты пишут правила для имплементации пассивных детектов на специальном DSL, который построен на базе NET-рантайма

Один из основных вызовов — не перегрузить активы повторяющимися запросами на модификацию от однотипных событий

02

Хранение данных в MaxPatrol SIEM

Что было под капотом MaxPatrol SIEM в 2017 году



Хранение данных

Вызовы в 2017 году

Традиционные СУБД не эффективны для обработки разнообразных по структуре событий и действительно больших объемов данных

1

Нет влияния на развитие open-source-разработок и возможности гарантировать пользователю качество хранения данных

2

3

Open-source-СУБД с поддержкой колоночного формата хранения теряют эффективность при работе с исходным текстовым представлением событий

В поисках альтернативы

для MaxPatrol SIEM

- работа с неструктурированными данными
- колоночный формат хранения
- адекватные требования к железу
- безопасный бэкап
- и т.д.

elasticsearch

ClickHouse

MICRO FOCUS

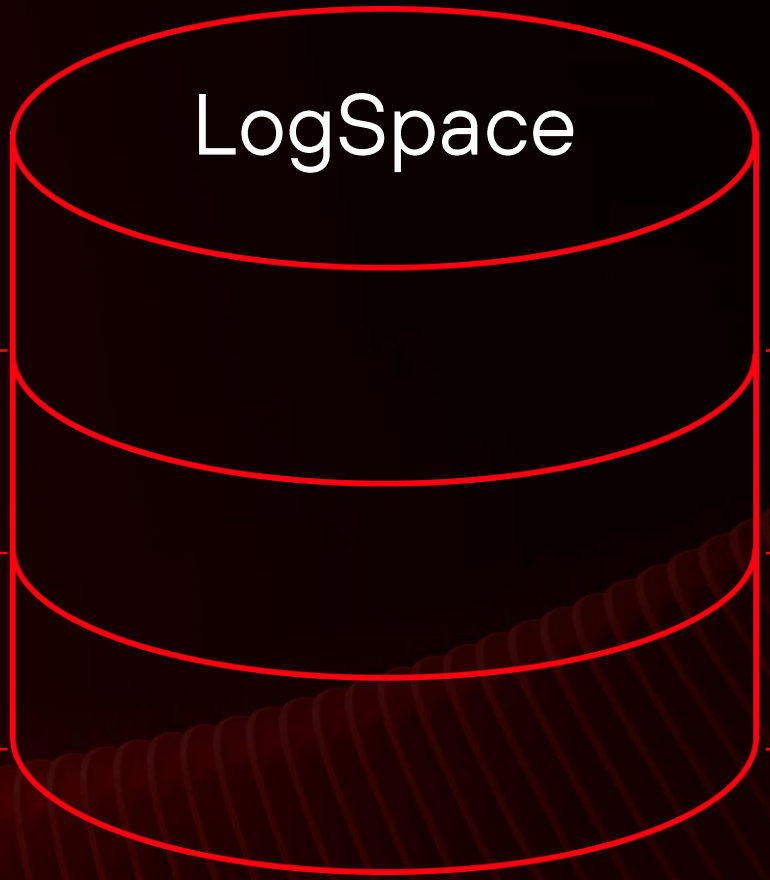
IBM DB2

cassandra

...и другие

СУБД от Positive Technologies

учитывающая специфику SIEM-систем



Колоночный формат хранения

SQL для работы с данными

Автоматическая адаптируемость

Высокая плотность хранения данных

Безопасное резервное копирование

Разработка ведется на C++

LogSpace

Хранение данных

На входе — данные в виде кортежей (CSV, JSON, etc)

time	uuid	subject	msgid	subject.account.name	body
2024.04.01 12:35:45.3456	0xA343-DCD5-...	account	4624	vpupkin	<eventid = 4624> <data .... />

Прозрачное хранение данных Можно изучить при помощи ls/find

Внутренний формат хранения данных

folder 20240401----056

time.dat	uuid.dat
2024.04.01 12:35:45.3456	0xA343-DCD5-...

subject_dat	msgid.dat
account	4624

subject_account_name.dat
vpupkin

body.dat
<eventid = 4624> <data .... />

Каждая папка — это сегмент данных, который хранит в себе несколько строк таблицы (от 1000 до 1 000 000 000 000)

Мы используем стратегию вставки со слиянием — вставить данные с минимальной задержкой, а на фоне заняться оптимизацией

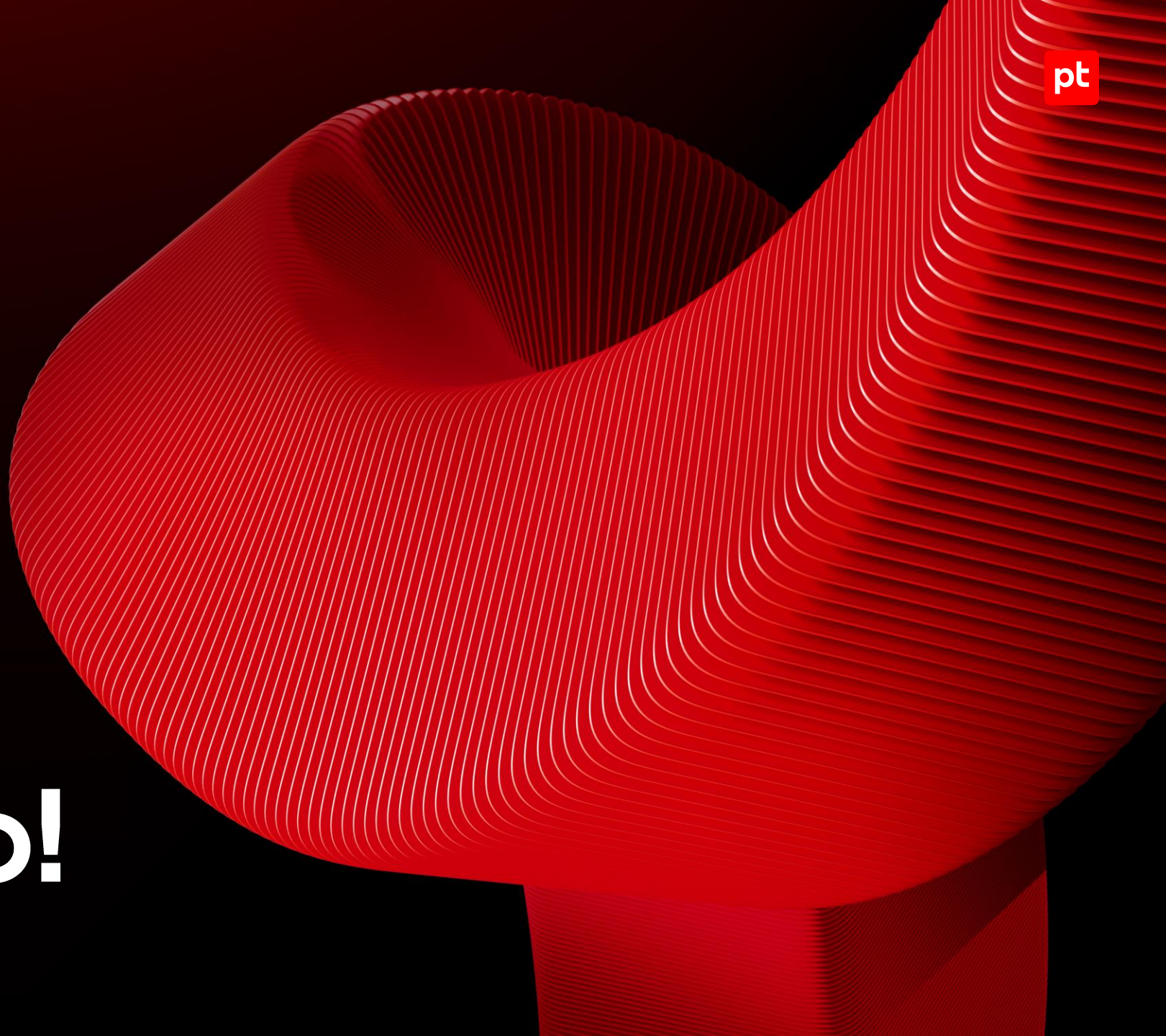
Эффективное хранение

time	delta
2024.04.01 12:35:45.3456	0
...	
2024.04.01 12:35:48.000	2.6544

subject orig	enum id
account	1
...	
process	15

body	dictionary	compressed blob
<eventid = 4624> <data .... />	eventid, data, ...	....
...		
<eventid = 4688> <data .... />		

Каждый файл для каждой колонки хранит в себе данные для множества событий. Можно использовать специфику данных — низкокардинальные множества, монотонно растущие последовательности «чисел», словарное кодирование на множестве коротких текстов



Присоединяйтесь к
нашему тг-сообществу

Спасибо!