SOC В ЭПОХУ BIGDATA

Сергей Рублев, руководитель SOC, CISSP
rublev@infosecservice.ru

О нас

- SOC для компаний «Открытия»
- Реагирование 24/7
- 50 000 хостов
- Участники Spark Community
- Используем bigdata-решения в «бою»

Постановка задачи (в разрезе данных)

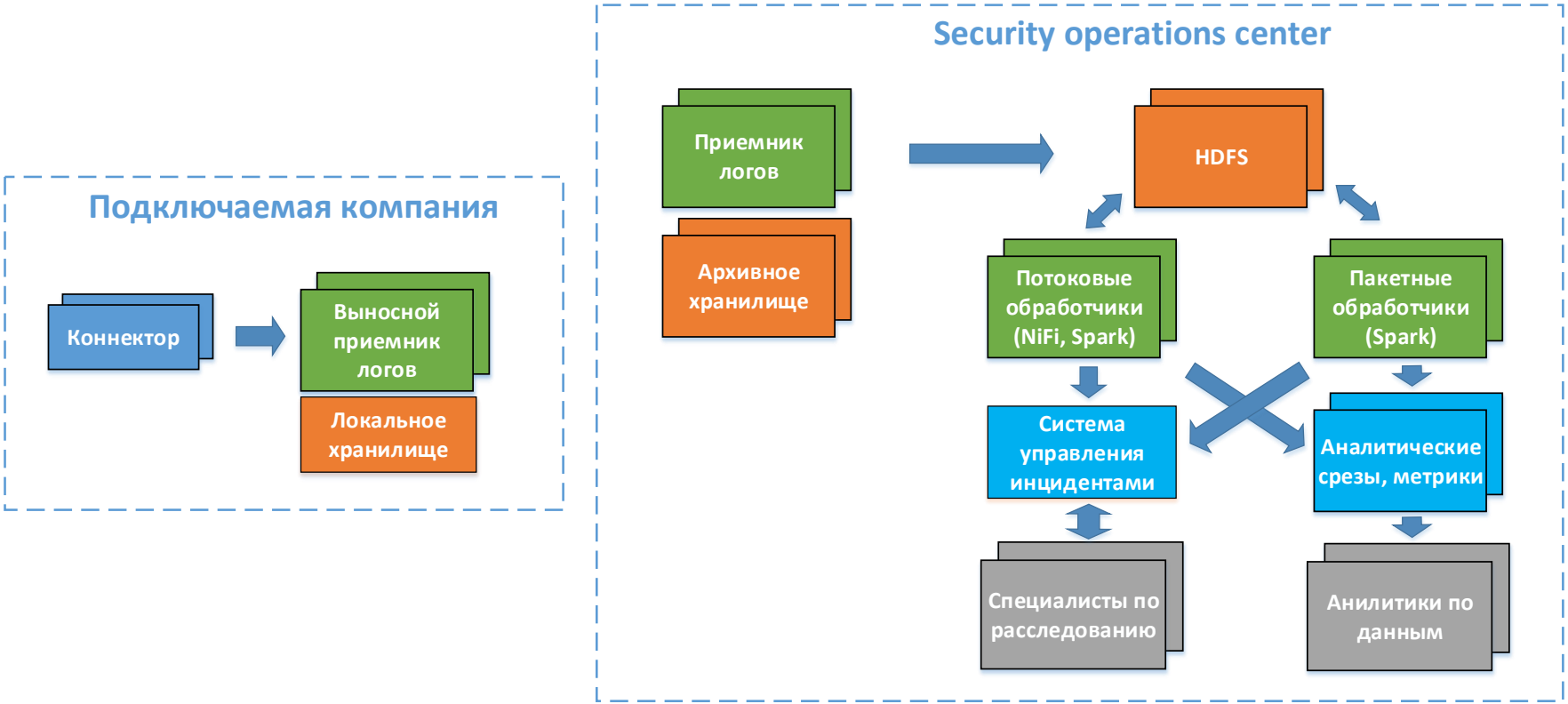
От «Открытия»

- Поток событий 4 ТБ/сутки
- Контроль потока (пропадание, задержки)
- Работа на исторических данных (старше 1 года)

От «Инфосекьюрити»

- Высокая доступность и отказоустойчивость
- Горизонтальная масштабируемость
- Построение аналитической платформы для широкого спектра задач
- «Быстрые» ретроспективные запросы
- Поиск по «сырым» данным
- Изоляция подключенных компаний (multitenancy)

Что построили



Текущие показатели

2ТБ /сутки
поток событий

1:50
коэффициент сжатия

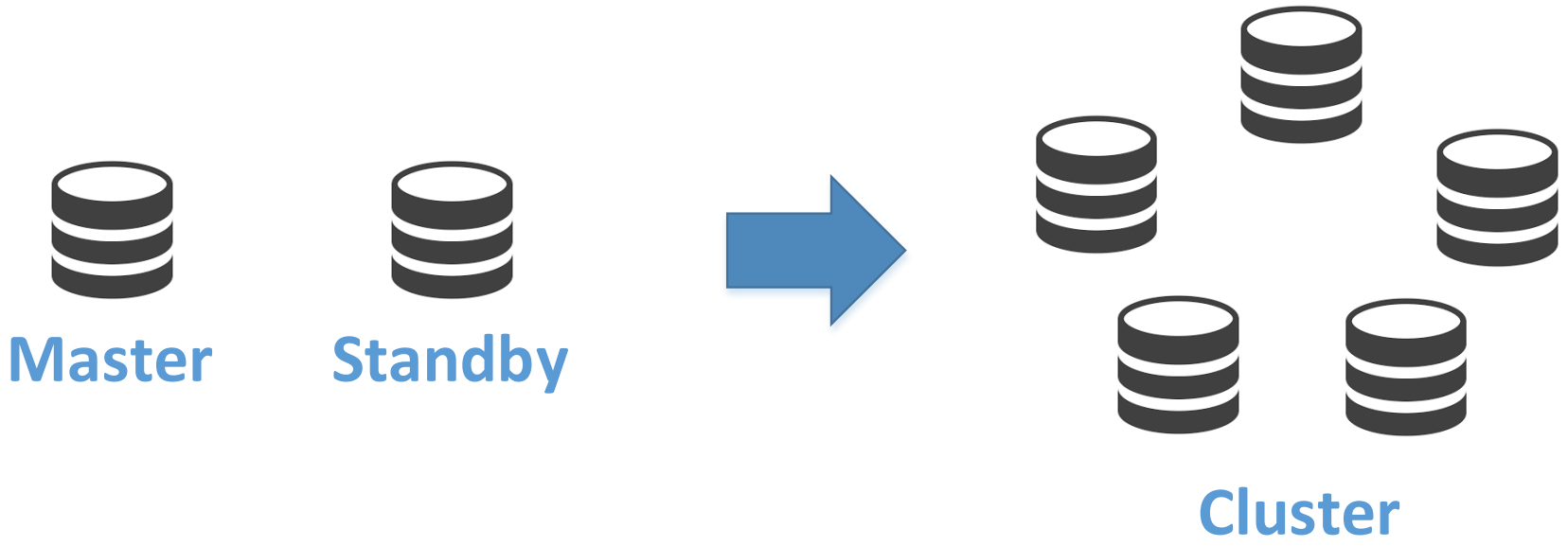
2 мин
поиск
по IP-адресу

8 мин
GroupBy + Sort

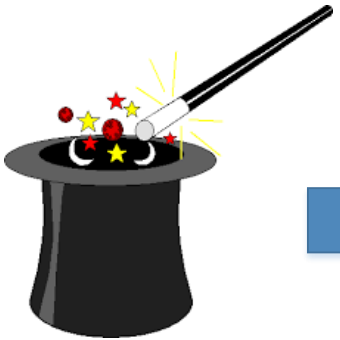
Производительность кластера на тестовой выборке 3 мес. Netflow:

17млрд событий
24тб данных

Отказоустойчивость всех компонент



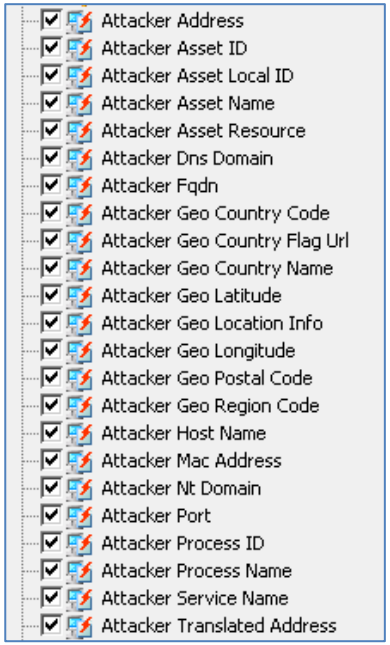
Возможности работы с данными

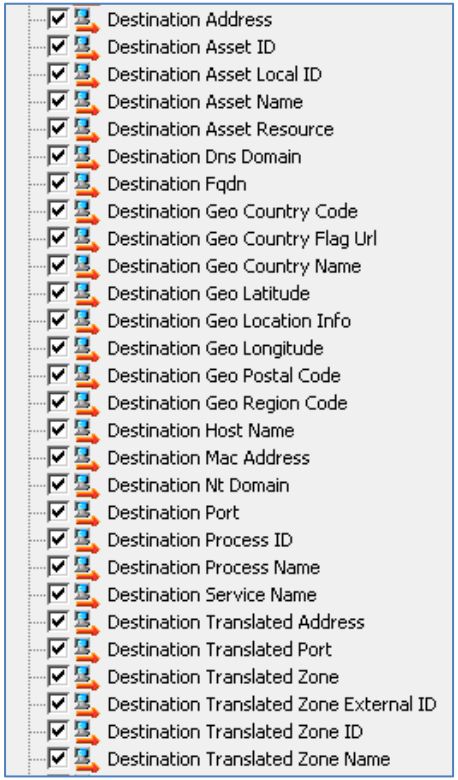


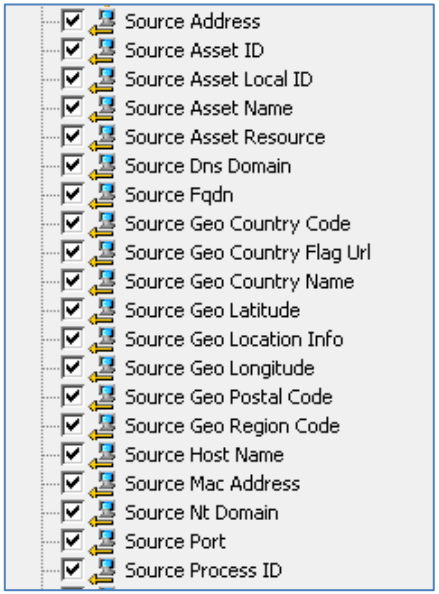
```
// Чтение HIVE-таблицы
val df = sqlContext.read.table("windows_data").
        where('sys_year === 2017
                and 'sys_month === 9
                and 'sys_day === 29)

// Чтение файлов
val json = sqlContext.read.json("jsondata.json")

// Чтение хранилища в Cassandra
val sql_data_count = sqlContext.read.format("org.apache.spark.sql.cassandra")
            .options(Map("table" -> "table", "keyspace" -> "keyspace"))
            .filter('year === 2017)
            .groupBy('dest_ip, 'dest_port)
            .count

// Аналитические запросы
val result = df.select('dvchost, 'eventid, 'sys_timestamp, 'targetusername)
                .groupBy('dvchost, 'eventid)
                .count
                .orderBy('count.desc)
```

SIEM fail

Неудобная таксономия событий



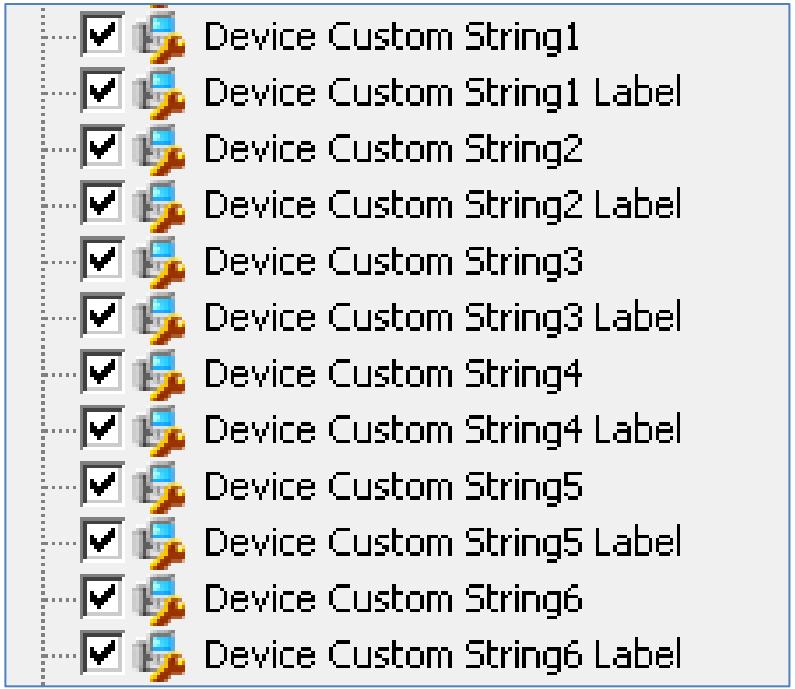




SIEM fail

Неудобная таксономия событий

Данные для самых «стреляющих» корреляций здесь ☺



SIEM fail

Загрузка и выгрузка исторических данных

Неожиданные победы

- Shell — это круто ☺
- Поисковые запросы –> правила выявления инцидентов
- На Scala легко писать обычным программистам
- DevOps для правил корреляции

Что дальше

- Больше данных!
- Применение алгоритмов машинного обучения для поиска аномалий
- UEBA
- Анализ графов (достижимость, связность)

gk-is.ru

+7 (499) 677-10-00