Новый аналитический движок

Smart Monitor Engine

VB-Trend 2019

plan>b

VolgaBlob

Что такое Smart Monitor Engine?

- Новый аналитический движок для обработки данных
- Поддерживает pipes в командах
- Обратная совместимость с командами Splunk
- Хранение данных в Elasticsearch и не только
- Многое другое

История развития Smart Monitor Engine

Март ’19
Начало разработки

Май ’19
SME Core

Октябрь ’19
70+ команд, функций

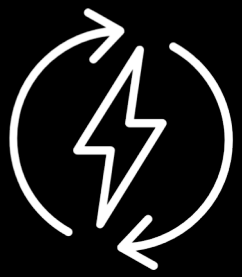
Июль ’19
Первые команды

Для чего Smart Monitor Engine?

VB-Trend 2019 plan>b VolgaBlob

Проще

Разрабатывается с учетом специфики работы профильными данными (технические логи, логи безопасности, аудита и т.д.)



Быстрее

За счет оптимизации и понимания специфики данных, обрабатываем данные быстрее.

Интеллектуальнее

Разработка ведется с участием профильных специалистов, мы понимаем что нужно Заказчику и сразу это ищем

Дорога, полная боли

VB-Trend 2019 plan>b VolgaBlob

Стандартные механизмы обработки запросов в ELK сложные → упростили в SME

Не всегда данные из ELK точные → реализовали собственные проверки

Низкая скорость загрузки данных в ELK → оптимизировали

Все, что нельзя реализовать командами, можно сделать скриптами →

собственный механизм работы с внешними данными

Smart Monitor Engine®

Как устроен SME?

Rest-интерфейс

Web-интерфейс

Ядро SME

Команда1

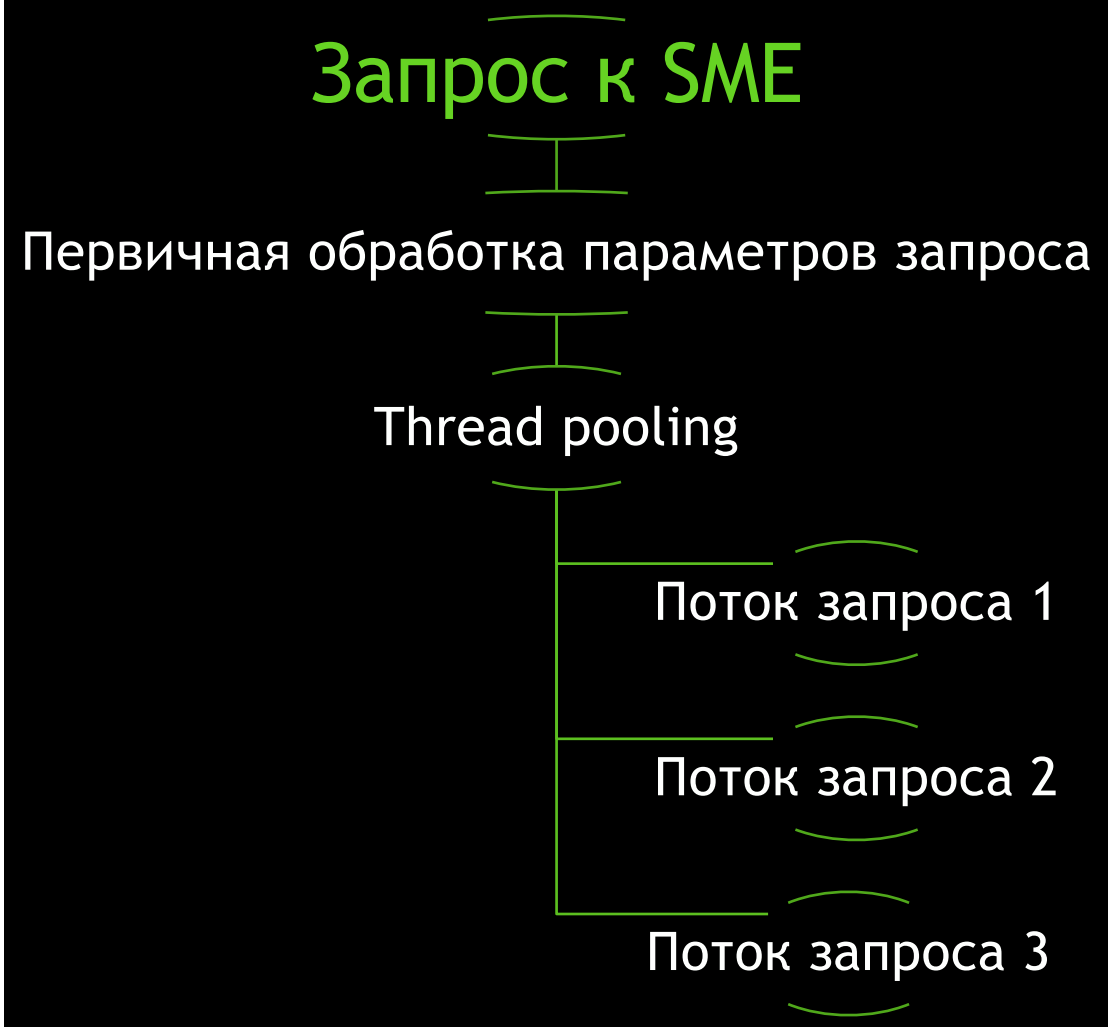
Команда2

In Memory Store

Внешние обработчики команд

Elasticsearch данные

Мультипоточность SME



Неблокирующая схема обработки запросов. Запускайте столько запросов, сколько нужно!

Режимы поиска

Smart Search

Собственный движок оптимизации запроса

Полнотекстовый поиск

Ищет все и по всем полям

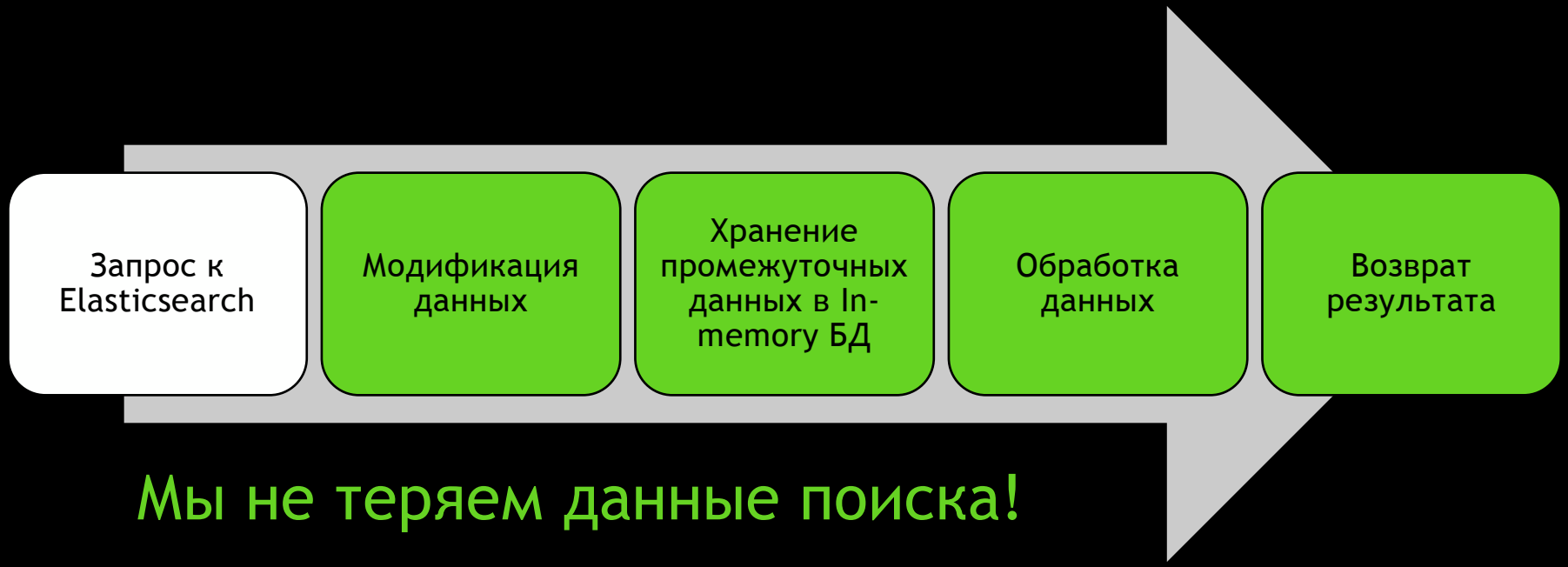
Статистика по запросу

Всегда понятно сколько данных обработано и какой результат

Динамическая подгрузка результатов

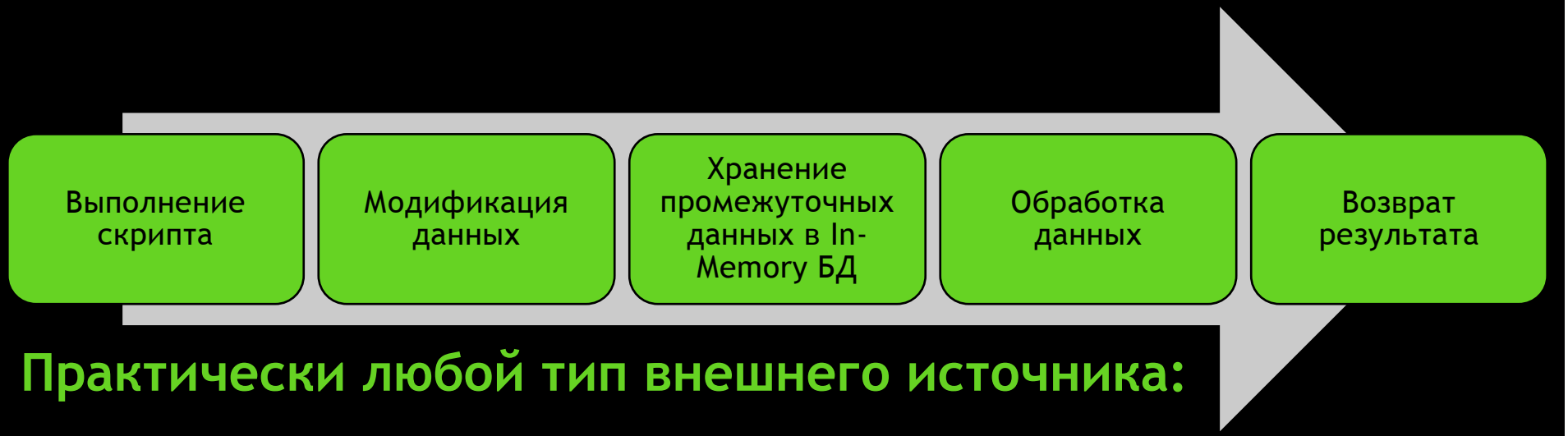
Не надо ждать! Первые результаты практически сразу!

Поисковый конвейер (Elasticsearch)



Мы не теряем данные поиска!

Поисковый конвейер (Scripts)



Практически любой тип внешнего источника:

- внешний web сервис
- база данных
- файл
- физический сенсор
- и т.д. ….

Field Alias

VB-Trend 2019 plan>b VolgaBlob

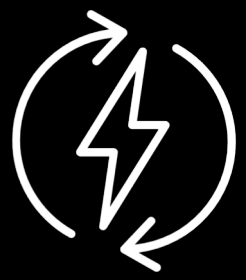
Проблема

- Добавление новых данных в Elasticsearch всегда должно соответствовать определенной структуре
- Индексы не должны быть слишком большие (20 Гб)
- Новые данные – новые index mappings

Решение – SME «Field Alias»

- Поддержка множественности индексов
- Mapping данных и приведение к общему результату
- Минимум настроек

Акселерация запросов



Запрос перед выполнением анализируется и оптимизируется

Данные обрабатываются порциями и в памяти, это быстро

Промежуточные результаты сохраняются и выгружаются in-memory БД

Статистика

VB-Trend 2019
plan>b
VolgaBlob

10K событий:

Dense:
- diff: +75 ms / + 2%

Sparse:
- diff: +71 ms / + 1%

Super-sparse:
- diff: +54 ms / + 0.05%

1M событий:

Dense:
- diff: +12 sec / + 1.5%

Sparse:
- diff: +2 sec / + 0.5%

Super-sparse:
- diff: +5 sec / + 5%

Примеры вызовов SME

1: source ‘access_logs' | limit 1,5 | fields - username

2: source ‘access_logs' qsize=500 | eval name= substr (tostring (fullname,"commas") , 1,8) | fields name

3: source ‘access_logs' qsize=1000 | stats count by username, ip

4: | script_mc process.py | fields id totals

Пример запроса в Query DSL

Многоуровневый JSON-запрос (60+ строк)

```
GET radius_logs/_search
{
  "size": 10000,
  "query": {
    "bool": {
      "must_not": [
        {
          "match": {
            "user": "0"
          }
        }
      ]
    }
  },
  "script_fields": {
    "user": {
      "script": {
        "lang": "painless",
        "source": "doc['user'].value.substring(8).toUpperCase()“,
    }
   }…
},
    "cert": {
      "script": {
        "lang": "painless",
        "source": "doc['cert'].value.toUpperCase()"
      }
    },
    "date": {
      "script": {
        "lang": "painless",
        "source": "doc['@timestamp'].value.toString(\"HH:mm:ss
dd.MM.yyyy\")"
      }
    },
    "type": {
      "script": {
        "lang": "painless",
        "source": "doc['type'].value + \":\" + doc['event'].value"
      }
    },
    "start_ip": {
```

Пример запроса в SME

22 строки читаемого кода

```
source 'radius_logs' qsize=10000
| search user!='0'
| fields - raw_event
| eval user = upper(substr(user, 8, -1))
| eval cert = upper(cert)
| eval date = strptime(@timestamp, "y-M-d'T'H:m:s.SSSZ")
| eval date = strftime(date, "HH:mm:ss dd.MM.yyyy")
| eval type = type + ":" + event
| eval coordinate_x_start = tonumber(substr(src_ip, 1, 2), 16)
| eval coordinate_y_start = tonumber(substr(src_ip, 3, 4), 16)
| eval coordinate_x_end = tonumber(substr(assigned_ip, 1, 2), 16)
| eval coordinate_y_end = tonumber(substr(assigned_ip, 3, 4), 16)
| eval lon = coordinate_x_end * 3.141592 / 180
| eval lat = coordinate_y_end * 3.141592 / 180
| eval prev_lon = coordinate_x_start * 3.141592 / 180
| eval prev_lat = coordinate_y_start * 3.141592 / 180
| eval lon_different = lon - prev_lon
| eval y = sqrt(pow(cos(prev_lat) * sin(lon_different), 2) + pow(cos(lat) *
sin(prev_lat) - sin(lat) * cos(prev_lat) * cos(lon_different), 2))
| eval x = sin(lat) * sin(prev_lat) + cos(lat) * cos(prev_lat) *
cos(lon_different)
| eval distance = atan2(y, x) * 6371
| eval calc_time = (distance / 900) * 60
| eval start_ip = src_ip, end_ip = assigned_ip
```

SME vs Discover

VB-Trend 2019 plan>b VolgaBlob

Выбор данных из разных индексов		
	10 Прямой поиск	10 Прямой поиск
Поиск по ключевому слову		
	9 Поиск по любому полю, нужны настройки	10 Поиск по любому полю
Статистика		
	7 Агрегации Elasticsearch	9 Полная статистика по запросу + данные
Строка поиска		
	3 Команды DSL	10 Любые команды SME

Скриншоты интерфейса

New Search

```
1 source 'kibana_sample_data_ecommerce'
2 | fields order_date customer_gender
3 | stats count by customer_gender
```

2,500 documents matched in 4,333 ms (11-10-2019 14:30:36 to 10-11-2019 14:30:36)

Last 30 days | Show dates | Refresh

Documents (2500) | Statistics (2) | Visualization

count	customer_gender
1236	MALE
1349	FEMALE

Rows per page: 10

Скриншоты интерфейса

VB-Trend 2019 plan>b VolgaBlob

New Search

```
1 source 'kibana_sample_data_ecommerce'
2 | fields order_date customer_gender
3 | stats count by customer_gender
```

2,500 documents matched in 4,333 ms (11-10-2019 14:30:36 to 10-11-2019 14:30:36)

Last 30 days | Show dates | Refresh

Documents (2500) | Statistics (2) | Visualization

Time	_source
2019-11-24T03:15:22+04:00	category: ["Women's Clothing"] currency: EUR customer_first_name: Abigail customer_full_name: Abigail Garner customer_gender: FEMALE customer_id: 46 customer_last_name: Garner customer_phone: day_of_week: Saturday day_of_week_i: 5 email: abigail@garner-family.zzz geoip.city_name: Birmingham geoip.continent_name: Europe geoip.country_iso_code: GB geoip.location: { "lon": -1.9, "lat": 52.5 } geoip.region_name: Birmingham manufacturer: ["Tigress Enterprises", "Champion Arts"] order_date: 2019-11-23T23:15:22+00:00 order_id: 592082

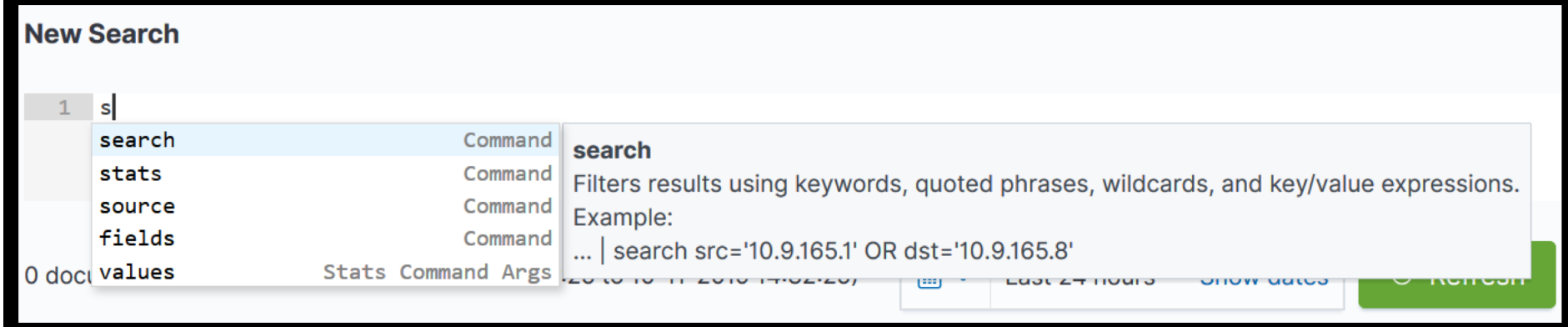
Expanded Document

Fields | JSON

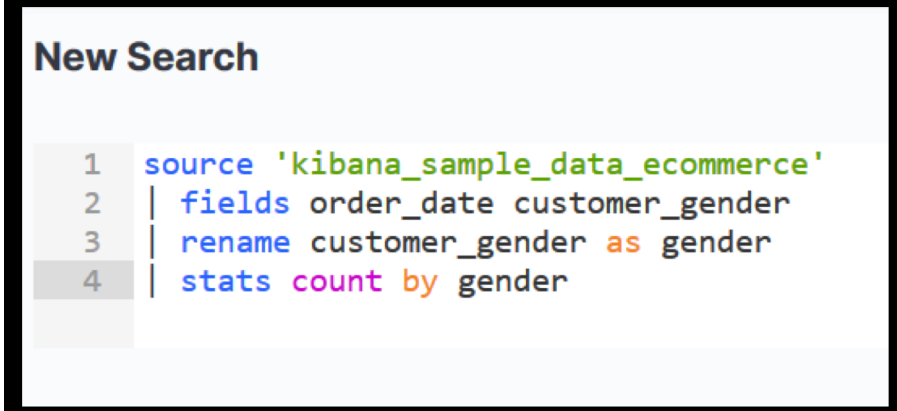
t	category	["Women's Clothing"]
t	currency	EUR

Скриншоты интерфейса

Подсказки в командах и описание к ним



Понятная цветовая схема



Roadmap

Команды – subsearch, lookup, join, map

Оптимизация производительности

Больше визуализации:

Single Value с поддержкой текстовых значений

Data Tables – подсветка в зависимости от значений

Расширение типов визуализации Controls

И все это… очень скоро!

Тестовые данные

1: Sample eCommerce Orders

данные о совершенных транзакциях (покупки в магазинах одежды)

2: Sample Flight Data

данные работы авиакомпании, статусы вылета и прилета

Демонстрация

Спасибо за внимание!