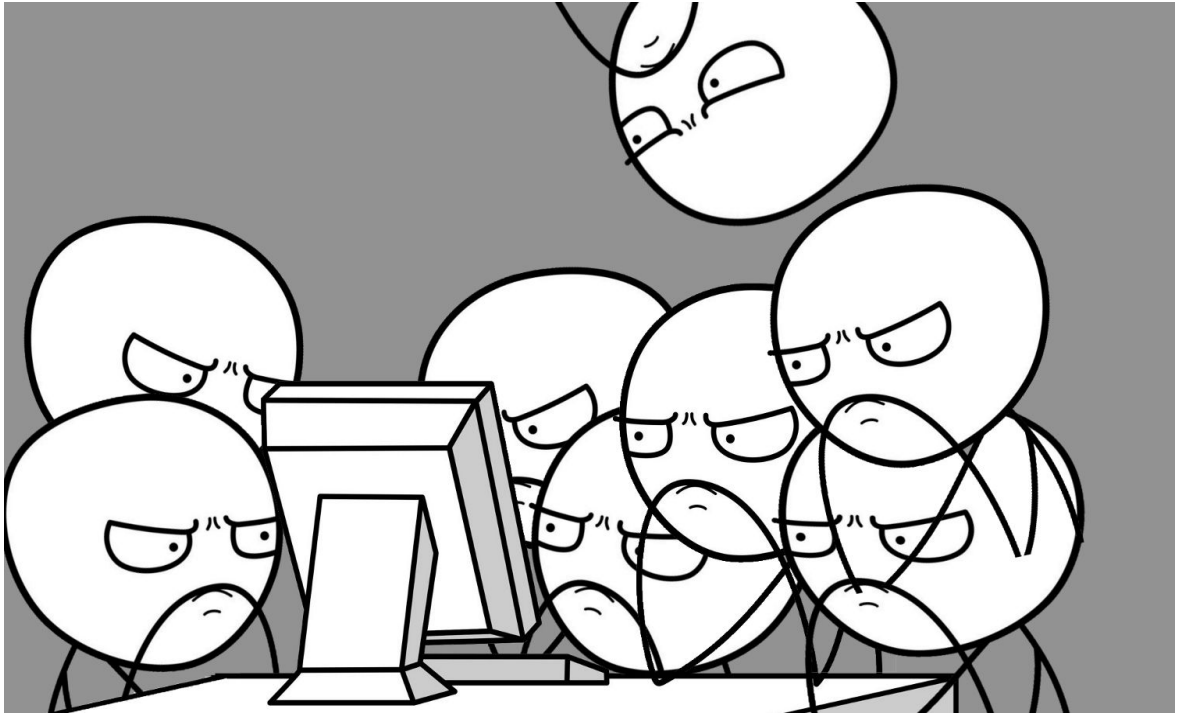
Мониторинг системы мониторинга



Кто мы и о чём расскажем в докладе

Часть 1. Мир Splunk
Рассказывает: Михаил Ефремов, системный администратор/devops
Часть 2. Make business monitoring great again
Рассказывает: Анна Манакова, системный администратор/devops
Часть 3. Данные о данных: Yet Another CMDB
Рассказывает: Михаил Ефремов

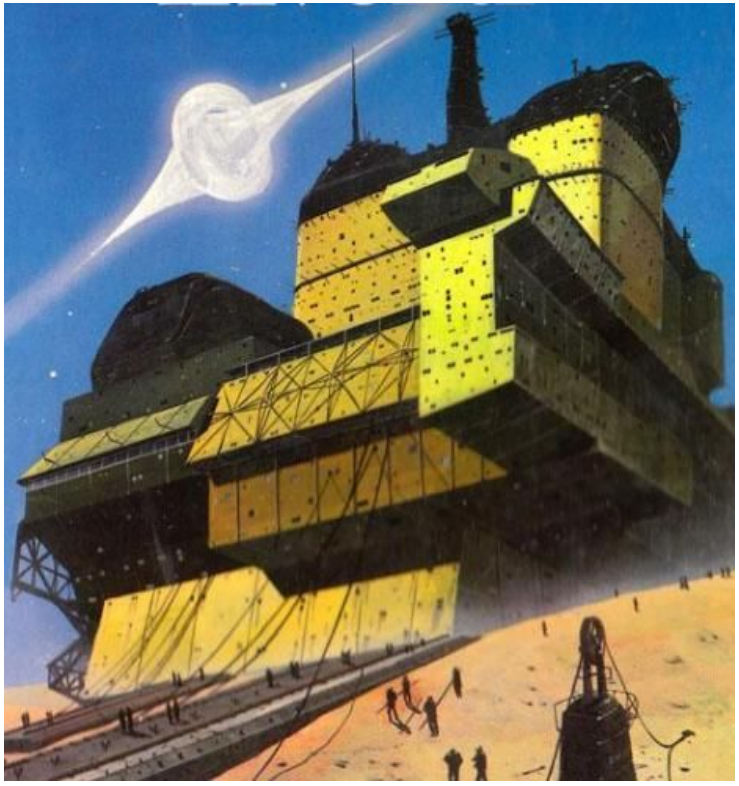
Часть 1. Мир Splunk

О чём расскажем в этой части:

- Мотивация доклада (чем хотим поделиться)
- «Категории» мониторинга
- Splunk как экосистема (в которой уживаются данные и код)
- Splunk как аппаратно-программный комплекс

Мотивация доклада (чем хотим поделиться)

Мотивация доклада (чем хотим поделиться)



Мотивация доклада (чем хотим поделиться)

Ощущение удовлетворения :)

Категории мониторинга

- Инфраструктурный мониторинг
- Бизнес-мониторинг

Экосистема Splunk

Splunk это

- Хранилище данных
- Средство разработки

Экосистема Splunk

Splunk как хранилище данных

- Логоподобные данные
- Справочники (lookups)
  - Файлы
  - Коллекции в MongoDB

Экосистема Splunk

Splunk как средство разработки

- Приложения, работающие в контексте самого Splunk
- REST API для интеграции с другими системами

Экосистема Splunk

Что ещё

- SPL (Search Processing Language)
  - Вычисляемые налету поля
  - Conditions
  - Regex
  - Статистические функции
  - Relational operators
  - Transactions processing
  - Predictions
  - и др.

Экосистема Splunk

Что ещё

- UI
  - Алерты
  - Отчёты
  - Реалтайм дашборы

Экосистема Splunk

Что хочет бизнес?

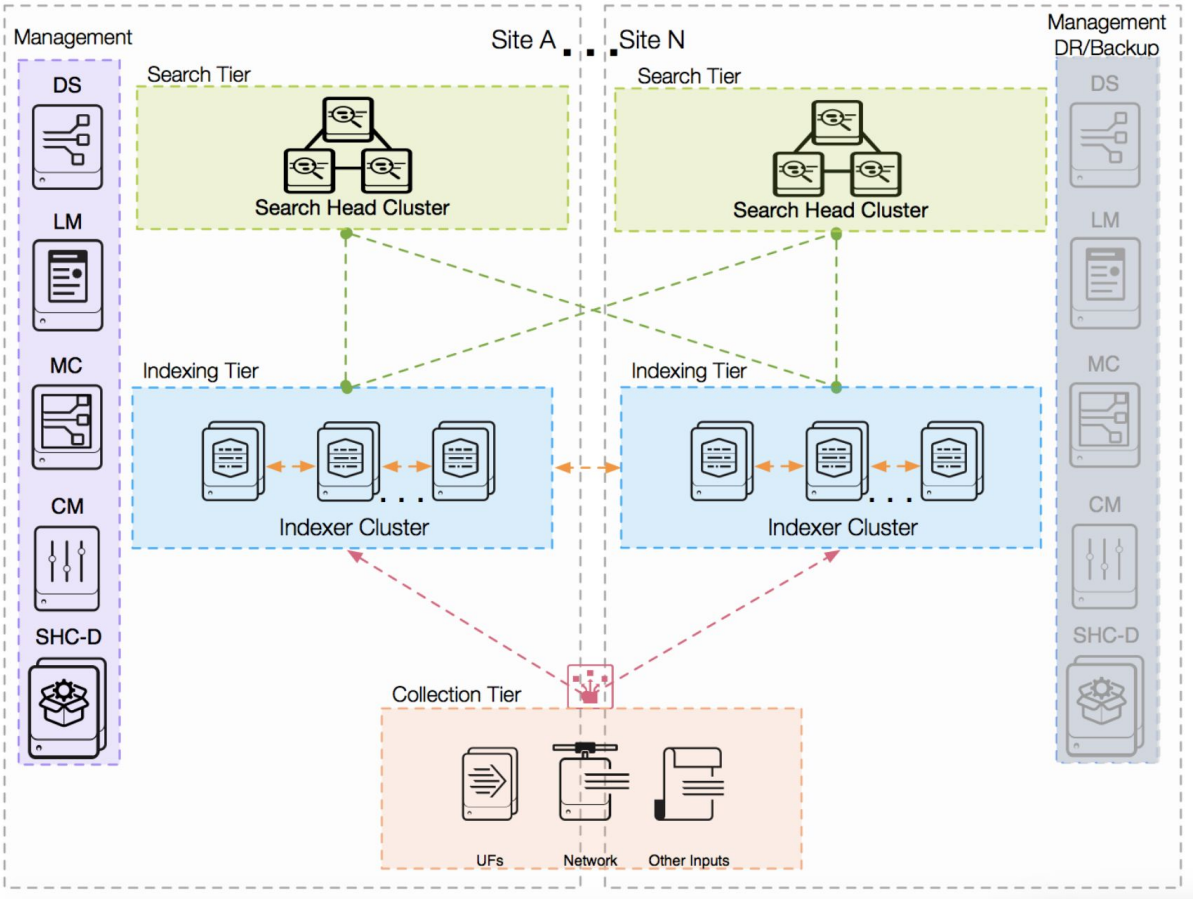
- Изощрённейшая аналитика с красивыми картинками
- Алерты по всему!
- Прямо сейчас!!

Прямо сейчас!!1

Splunk как **аппаратно**-программный комплекс

- Кластеры indexer-ов
- Кластеры search head-ов
- Служебные машины
- Источники данных

Splunk как **аппаратно**-программный комплекс



Splunk как **аппаратно**-программный комплекс

Индексеры

- Полтора десятка железных серверов
- Сотни гиг RAM на каждом
- Десятки ядер (Xeon) на каждом
- NVME/SSD для горячих и тёплых данных
- 10+ терабайтные локальные RAID
- Резервирование: податацентровое + реплики внутри ДЦ

Splunk как **аппаратно**-программный комплекс

Search head-ы

- Примерно то же самое, но без забористых хранилок

Splunk как **аппаратно**-программный комплекс

Источники

- DBconnect
-
-
-
-

Splunk как **аппаратно**-программный комплекс

Источники

- DBconnect
- Syslog
-
-
-

Splunk как **аппаратно**-программный комплекс

Источники

- DBconnect
- Syslog
- HTTP Event Collector
-
-

Splunk как **аппаратно**-программный комплекс

Источники

- DBconnect
- Syslog
- HTTP Event Collector
- **Universal Forwarders**
-

Splunk как **аппаратно**-программный комплекс

Источники

- DBconnect
- Syslog
- HTTP Event Collector
- **Universal Forwarders**
- Кое-что ещё...

Splunk как **аппаратно**-программный комплекс

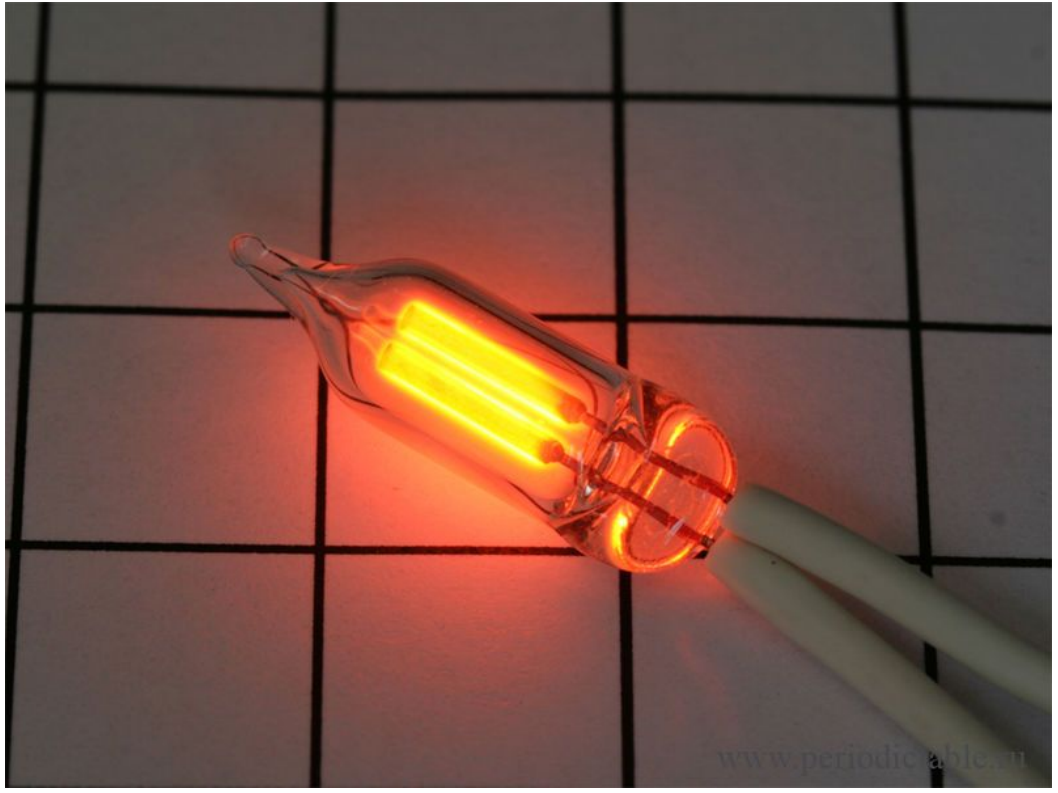
Ещё немного цифр

- Терабайты данных суммарно в сутки
- Несколько тысяч источников
- Сотни активных пользователей
- Сотни RPS, десятки SPS
- Хранение в среднем пару недель

Splunk как **аппаратно**-программный комплекс

А что же внутри?

Splunk как **аппаратно**-программный комплекс



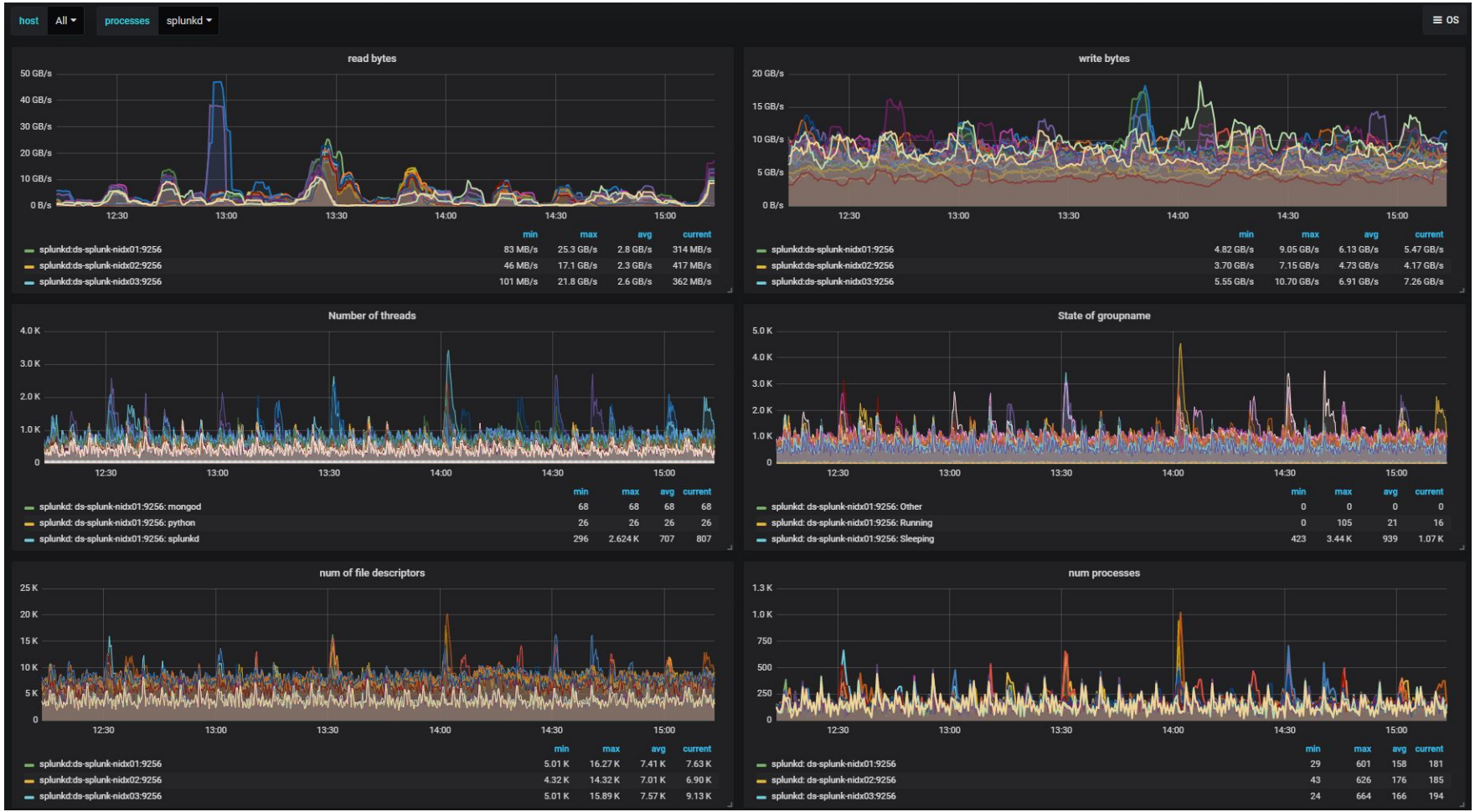
И в какой-то момент что-то пошло не так...

Часть 2. Make Business Monitoring Great Again

О чём расскажем в этой части:

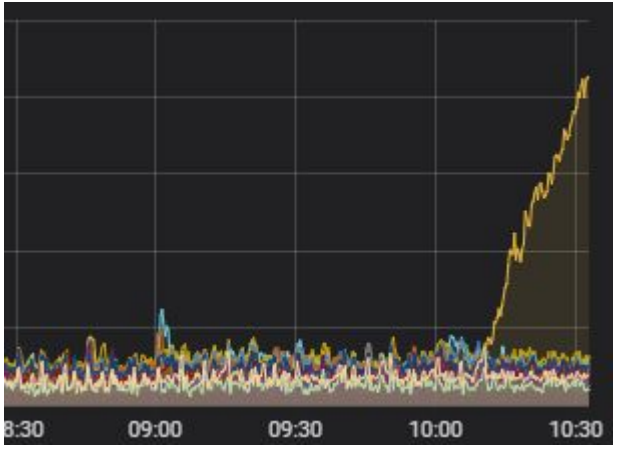
- Формулировка проблемы
- Мониторинг процессов, что нас удивило и где нашли зацепку
- Жизнь внутри индекса и как за этим можно смотреть

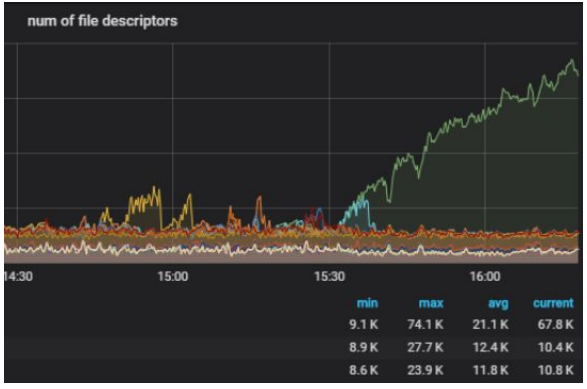
Мониторинг процессов

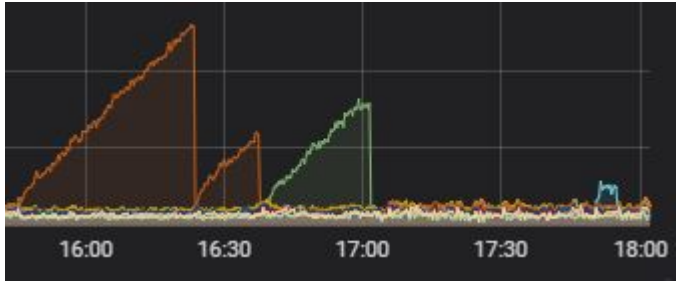


https://github.com/ncabatoff/process-exporter

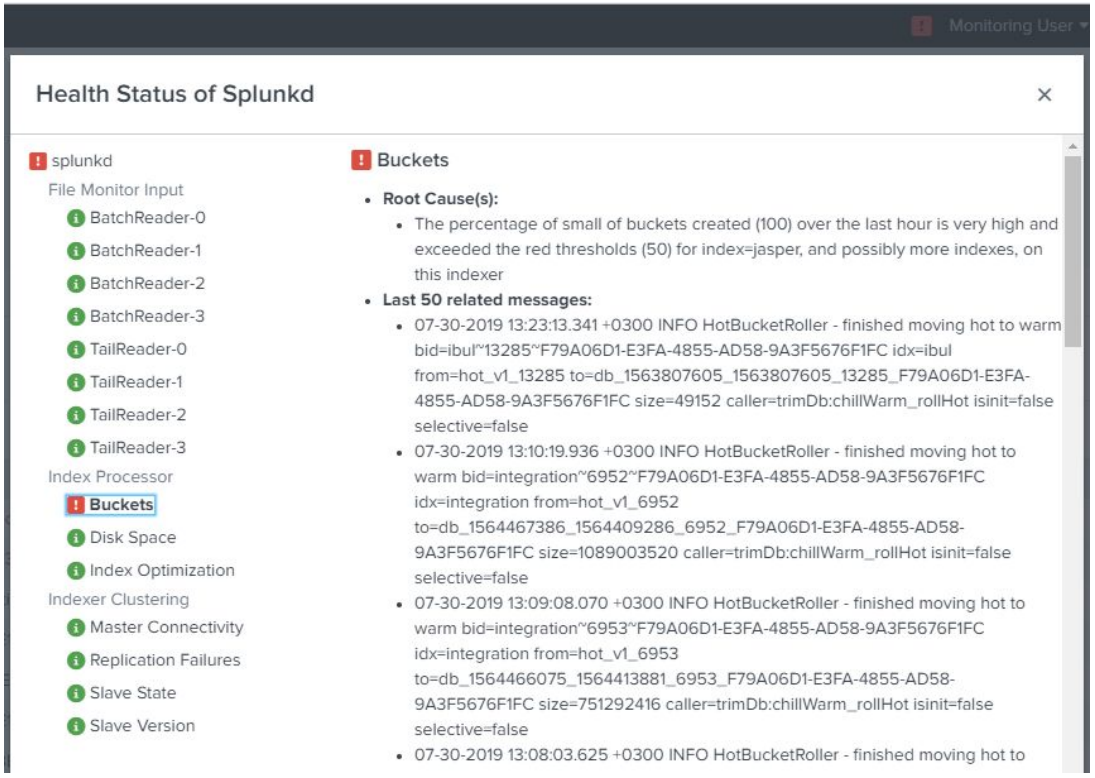
В начале было...



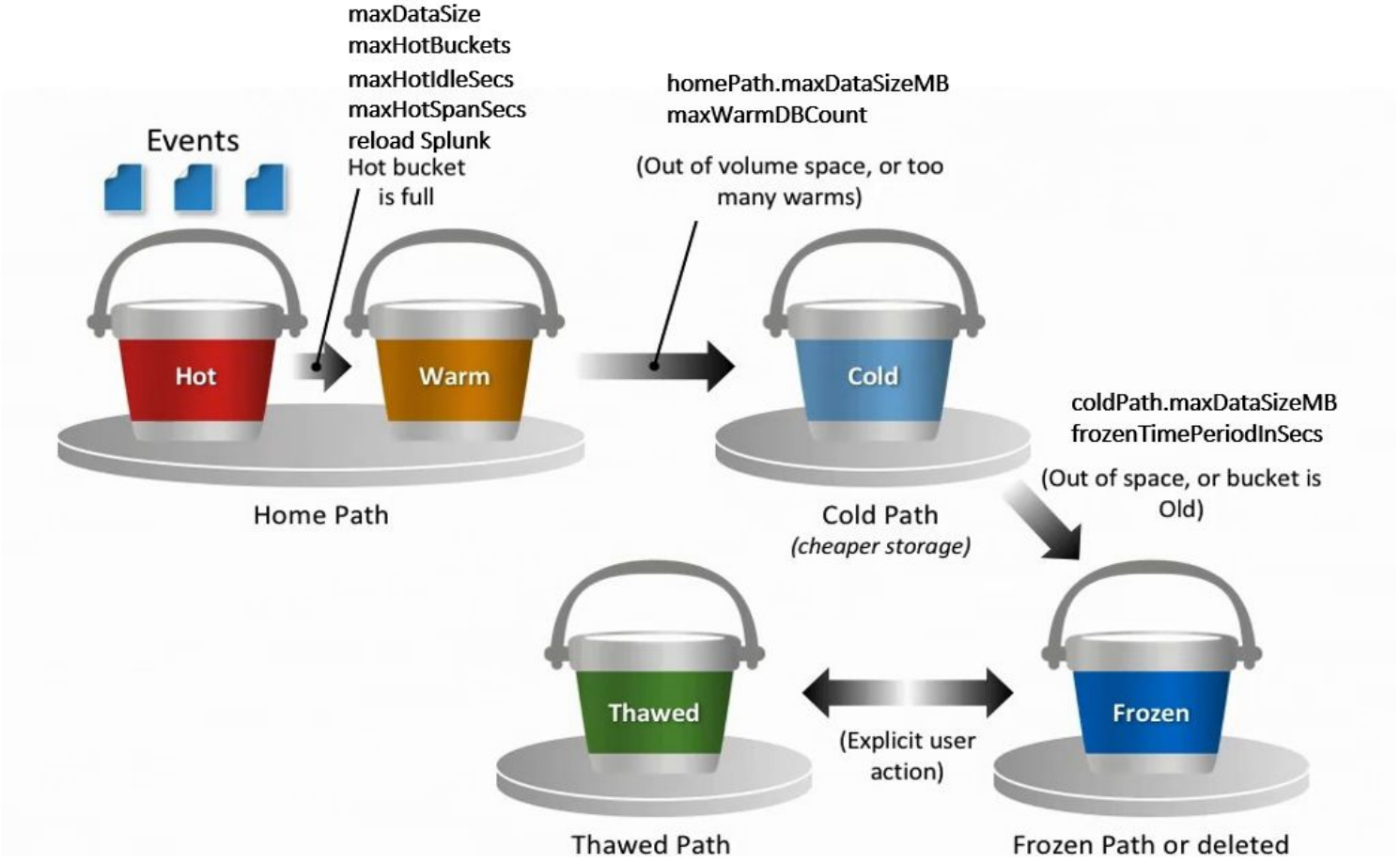




Помоги себе сам



Жизненный цикл данных в индексе

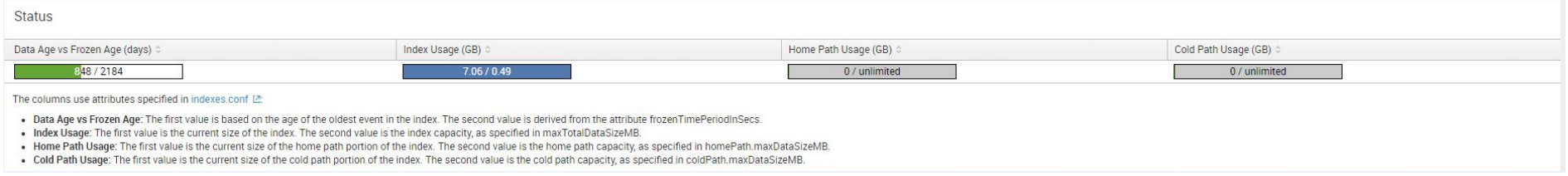


Откуда горки растут

```
[filenet]
homePath = $SPLUNK_DB/cl_filenet/db
coldPath = $SPLUNK_DB/cl_filenet/colddb
thawedPath = $SPLUNK_DB/cl_filenet/thaweddb
maxHotBuckets = 10
maxHotIdleSecs = 86400
repFactor = auto
maxDataSize = auto_high_volume
maxTotalDataSizeMB = 500
```

homePath.maxDataSizeMB = (maxWarmDBCount + maxHotBuckets) * maxDataSize

coldPath.maxDataSizeMB = maxTotalDataSizeMB - homePath



«Что делать?»

1. Для горячих бакетов ключевой параметр - *maxDataSize*:

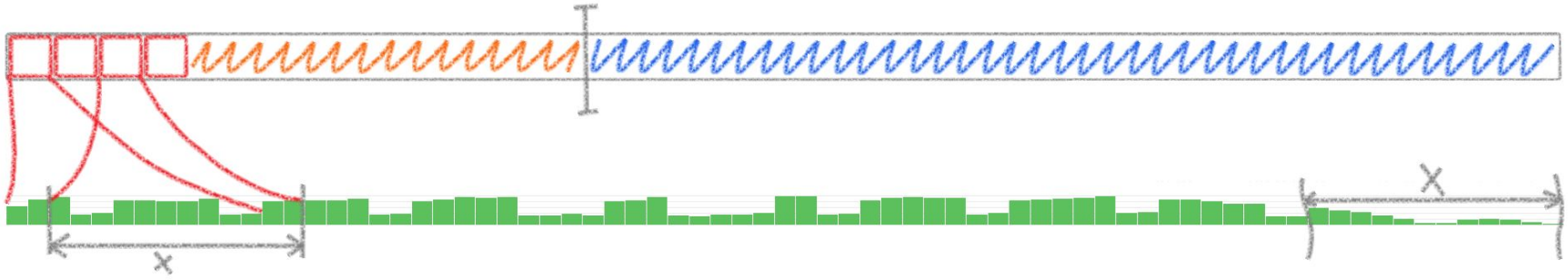
$$\limsup_{maxDataSize \to \infty} maxHotBuckets$$

2. Для теплых бакетов ключевой параметр - *homePath*:

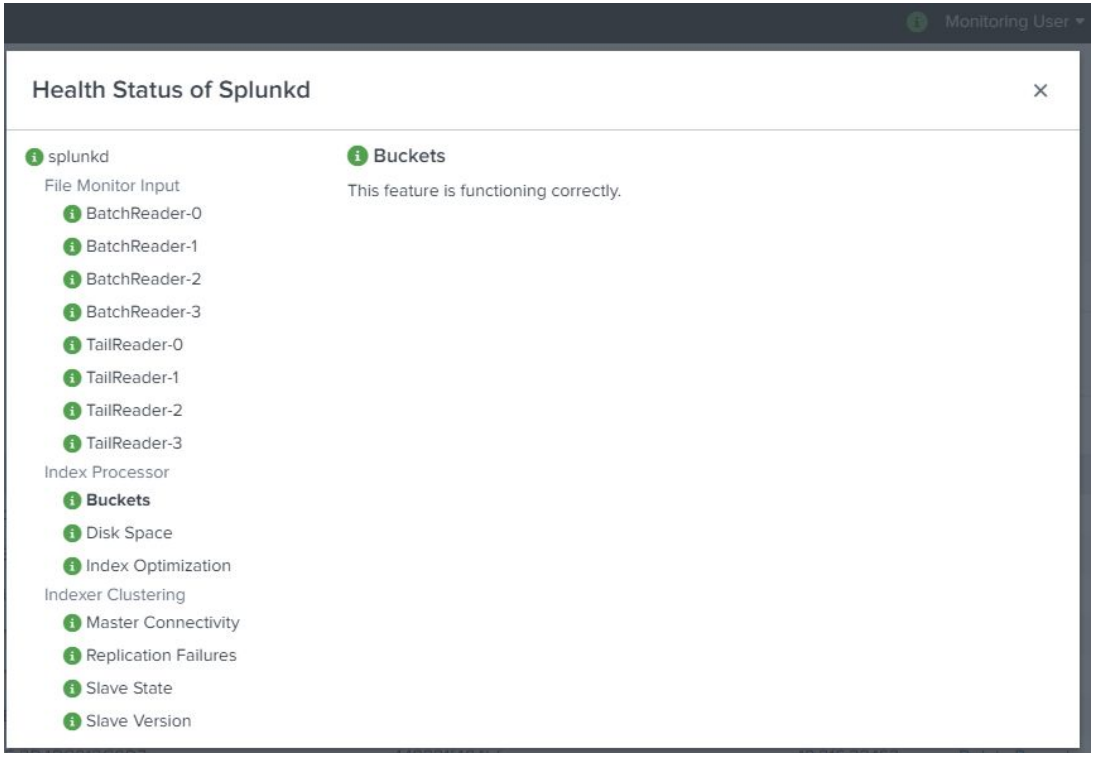
$$\limsup_{maxDataSize \to 0} maxWarmDBCount$$

3. Исключается maxHotIdleSecs = 0:

$$\limsup_{maxHotSpanSecs \to x} maxDataSize$$



Результат внутреннего мониторинга сейчас



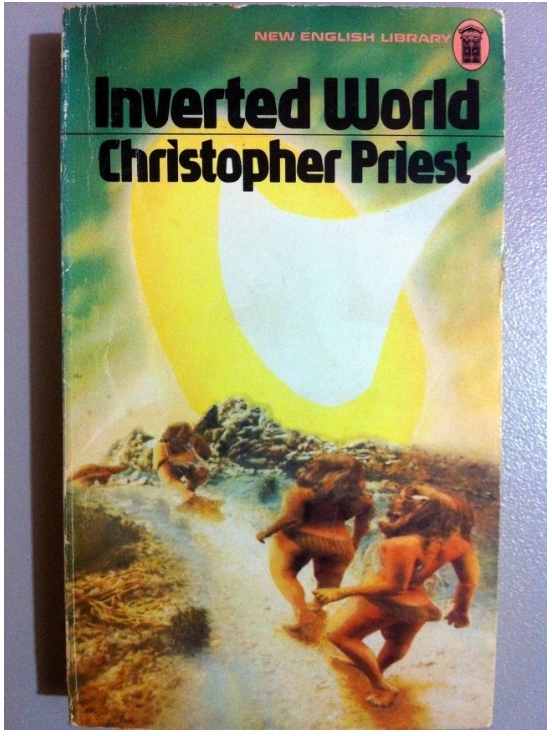
Часть 3. Данные о данных: Yet Another CMDB

Часть 3. Данные о данных: Yet Another CMDB

О чём (не)расскажем в этой части:

- Summary по предыдущей части
- Затупившаяся бритва Оккама как мотивация создания Splunk CMDB
- Splunk REST API → Django-based CMDB
- CMDB → Splunk и связанная с этим «девопсня»

Ответ на загадку



Спасибо за внимание!

Докладывали обстановку
- Ефремов Михаил - e-mail: meechanic@gmail.com, telegram: @mike_meechanic
- Манакова Анна - e-mail: manakova.anna@gmail.com, telegram: @mannabanana