ПОЧЕМУ И КАК ВЫПОЛНЯТЬ ПОИСК 2-ЛЕТНЕЙ ДАВНОСТИ В ВАШИХ ЛОГАХ ELASTIC SEARCH

последняя важная часть управления вашим жизненным циклом индексов (ILM)

cloudvyzor.ru

В идеальном мире

- Все мои журналы помещаютсяв мой Elastic Search кластер
- Мне не нужны старые логи, я решаю все проблемы немедленно
- Все необходимые показатели уже предопределены и предварительно рассчитаны

В реальном мире

- Размер журналов составляет терабайты
- Только журналы за 4-8 недель помещаются в мой Elastic Search кластер
- Возможно, мне придется ВЕРНУТЬСЯ к старым журналам …

3 возможные причины вернуться

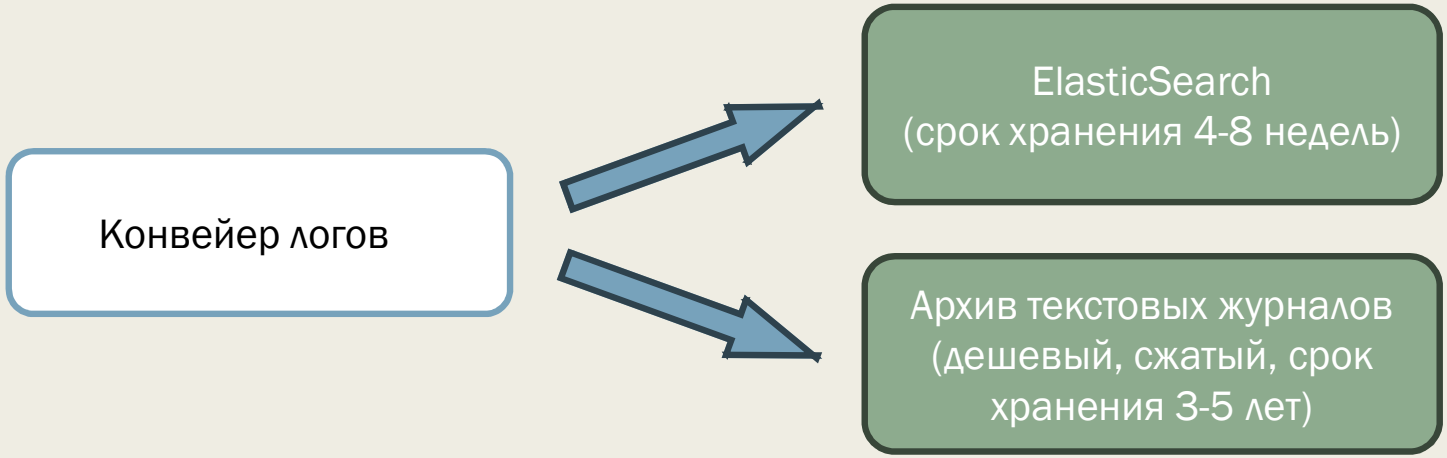
- Поддержка
- Интеллектуальный анализ данных
- Соответствие

3 возможные причины вернуться

- Поддержка:
  – *“Давайте выясним, когда появился этот недостаток в бизнес-логике и какие клиенты пострадали”.*
- Интеллектуальный анализ данных
  – *“Давайте рассчитаем этот новый показатель на основе данных за последние 2 года”*
- Соответствие
  – *“Давайте докажем, что только авторизованные инженеры имели доступ к производству в течение последнего года”*

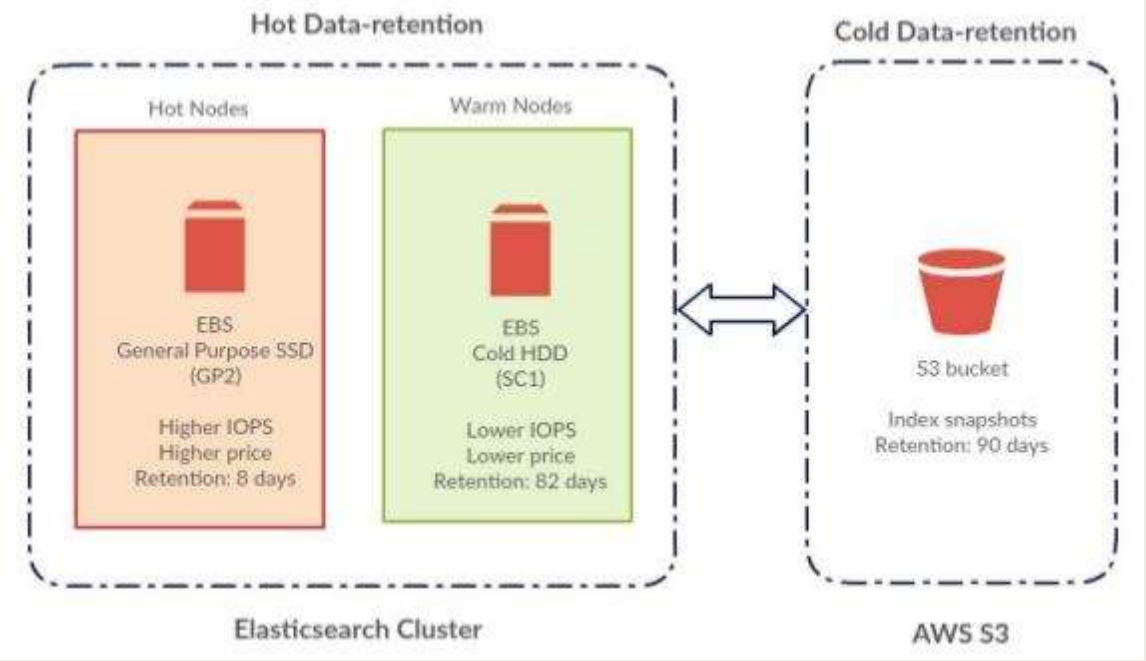
Как я могу вернуться?

Некоторые компании используют архив текстовых журналов с возможностью поиска



Lazada: https://youtu.be/NAeedJv-S3I?t=1335
Яндекс: https://youtu.be/ydwuccVwYBM?t=354

Подождите! Но я уже делаю ежедневные резервные копии индекса!



Но можете ли вы найти?

Вероятно:
в простом случае
“Покажите мне все журналы для этого клиента за 23 октября”

- Мне нужно выполнить поиск в пределах небольшого известного временного интервала
- Я знаю, какие индексы монтировать
- Их всего несколько, я могу быстро смонтировать
- Чтобы я мог быстро искать в своем кластере ELK

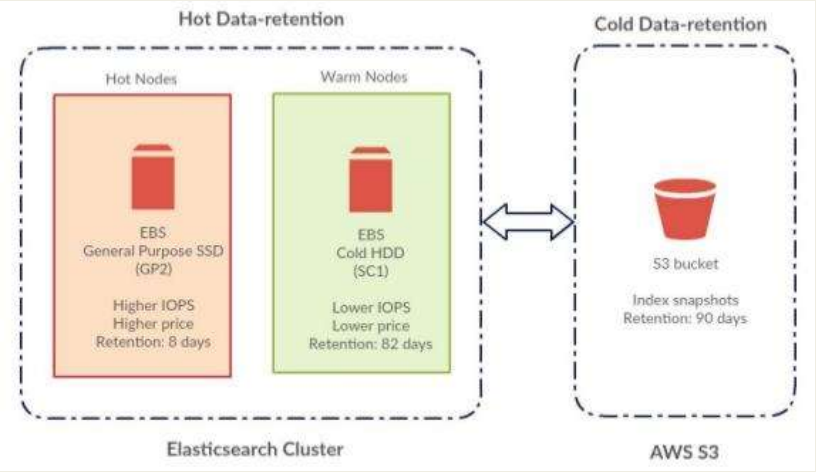
Но можете ли вы найти?

Не совсем:
в трудном случае
“Давайте поищем эту ошибку 6 месяцев назад, чтобы узнать, когда она впервые возникла”
“Давайте вычислим новую метрику, которую мы раньше не анализировали”

- Мне нужно выполнить поиск некоторых событий по всем или нескольким индексам многолетней давности
- Мне может понадобиться полнотекстовый поиск
- Некоторые поля еще не были проанализированы, нужно проанализировать сейчас
- Будет медленно монтировать сотни ежедневных снимков один за другим и выполнять поиск

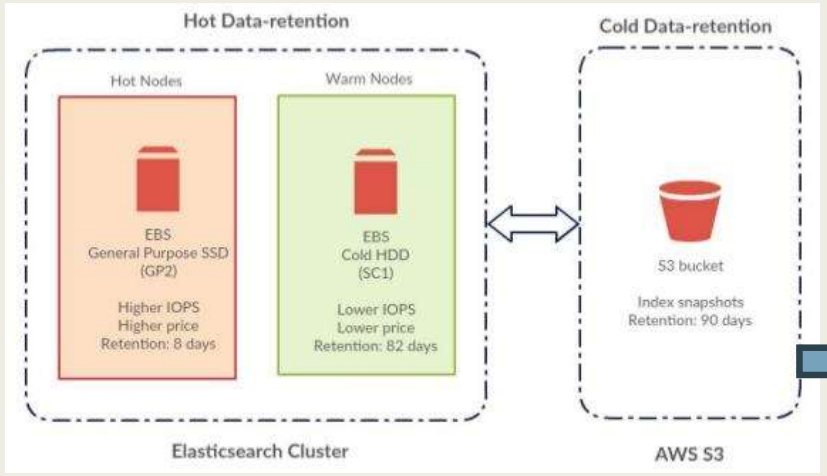
Как быть готовым?

- Сохраняйте то, что у вас есть
- Добавьте еще несколько шагов

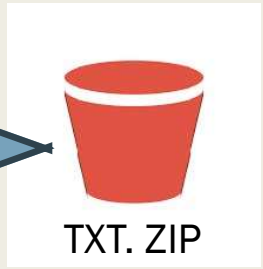


Как быть готовым?

- Сохраняйте то, что у вас есть
- Добавьте еще несколько шагов



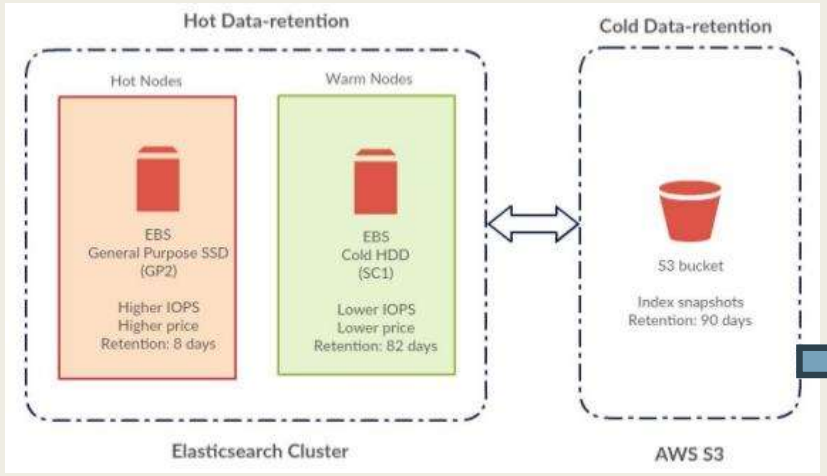
Преобразуйте ваши снимки в текстовые архивы



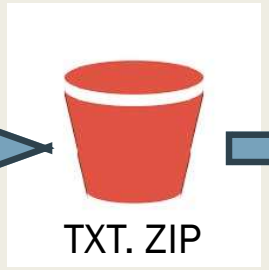
TXT. ZIP

Как быть готовым?

- Сохраняйте то, что у вас есть
- Добавьте еще несколько шагов



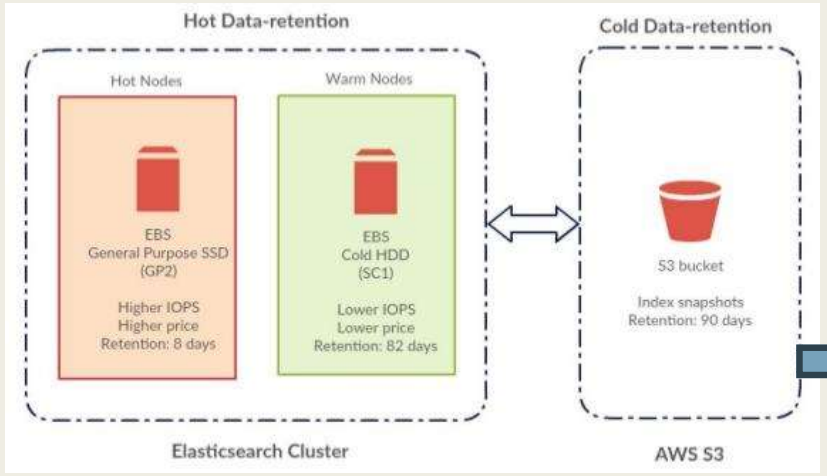
Преобразуйте ваши снимки в текстовые архивы



Выполните поиск в текстовых архивах и получите отфильтрованный CSV-файл

Как быть готовым?

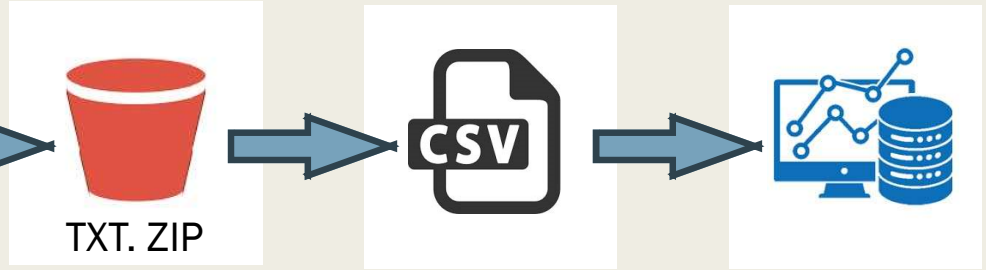
- Сохраняйте то, что у вас есть
- Добавьте еще несколько шагов



Преобразуйте ваши снимки в текстовые архивы

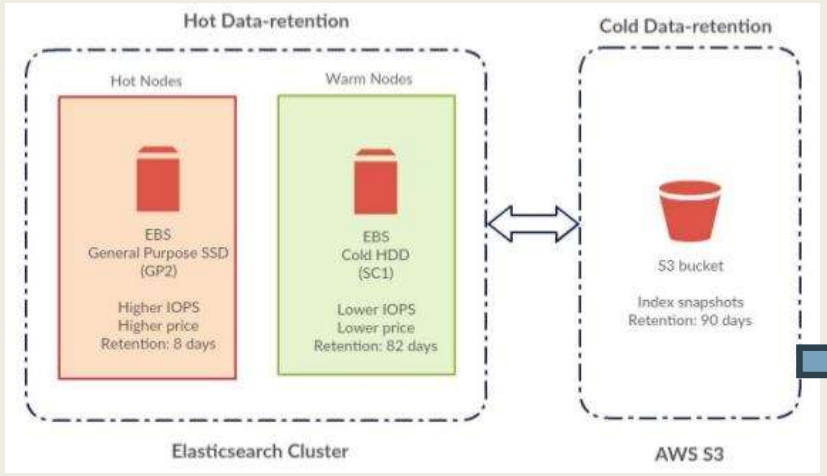
Выполните поиск в текстовых архивах и получите отфильтрованный CSV-файл

Проанализируйте с помощью вашего любимого инструмента BI (или вернитесь к ES)



Как быть готовым?

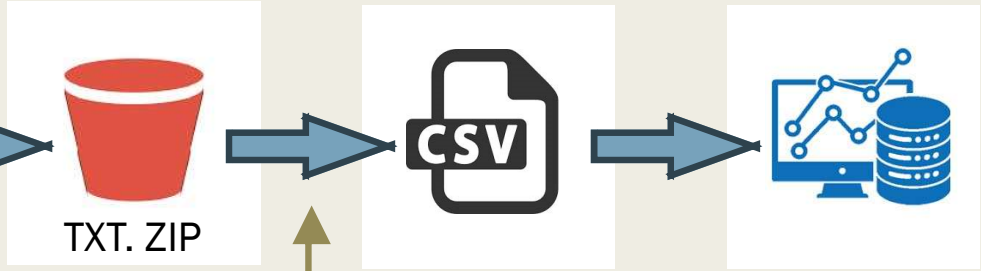
- Сохраняйте то, что у вас есть
- Добавьте еще несколько шагов



Преобразуйте ваши снимки в текстовые архивы

Выполните поиск в текстовых архивах и получите отфильтрованный CSV-файл

Проанализируйте с помощью вашего любимого инструмента BI (или вернитесь к ES)



Клаудвизор
Экспортер
резервных копий ES

Клаудвизор

Тематическое исследование

- 2 года ежедневных снимков индекса ES на S3
  - *60 TB снимков*
  - *13 B логов*

Задача 1:

Обеспечить доступ всех инженеров к prod за последний год

Задача 2:

Получить сводную таблицу выставленных счетов за прошедший год

Шаг 1: Преобразуйте снимки в текстовые файловые архивы

- Инструмент: Экспортер резервных копий Клаудвизор ES
- Входные данные: корзина AWS S3,60 ТБ, снимки за 2 года
- Машина: AWS EC2 c5n.18xlarge
  - *72 vCPU, 196 RAM, сеть 100 ГБ*
- Прошедшее время: 30 часов
- Результат: корзина AWS S3, 3 ТБзаархивированных текстовых файлов
- Потрачено: 250 долларов США

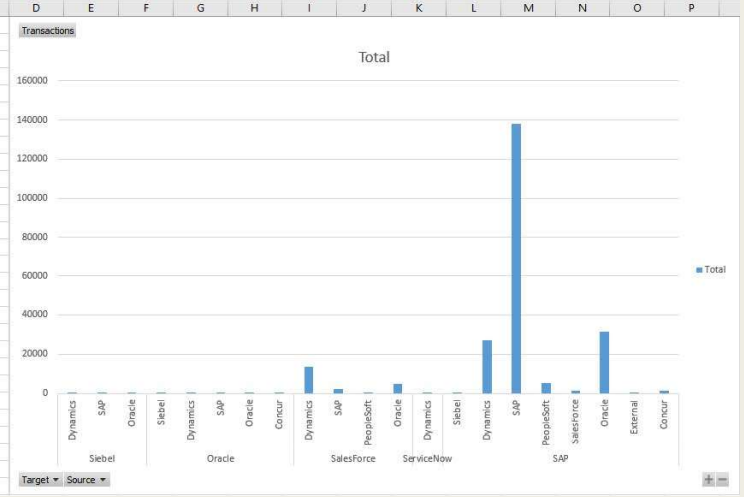
Шаг 2: Поиск и экспорт

- Инструмент: Панель входа в Клаудвизор (On-premise версия)
- Входные данные: корзина AWS S3,3 ТБ сжатых текстовых файлов
- Компьютер: AWS EC2 c4.xl большой: 4 VCPU, 8 ГБ оперативной памяти
- Стоимость 1 часа интерактивного поиска и экспорта: 0,5 доллара США
- Поиск1: Все входы в prod в 2019 году
  - Время поиска: 2 минуты, отсканировано 36 ГБ журналов
  - Время экспорта: 15 минут, 360 тысяч событий
  - Выходной файл: CSV, 36 МБ
- Search2: Все выставленные счета за транзакции в 2019 году
  - Время поиска: 5 минут, отсканировано 245 ГБ
  - Время экспорта: 11 минут, 260 тыс. Событий
  - Выходной файл: CSV, 40 МБ

Шаг 3: Анализ

- Задача 1: Расширить доступ
  - *вы получили: CSV - это ваш окончательный отчет*
- Задача 2: Сводная информация по выставленным счетам за транзакции
  - *Инструмент: MS Excel*
  - *Ввод: CSV, 260 тыс. Строк*
  - *Машина: Настольная*
  - *Время сборки стержня: 10 мин*

Row Labels	Transactions
Siebel	**117**
Dynamics	8
SAP	95
Oracle	14
Oracle	**811**
Siebel	3
Dynamics	532
SAP	6
Oracle	239
Concur	31
SalesForce	**20490**
Dynamics	13517
SAP	2159
PeopleSoft	2
Oracle	4812
ServiceNow	**7**
Dynamics	7
SAP	**205155**
Siebel	36
Dynamics	27278
SAP	138032
PeopleSoft	5401
SalesForce	1338
Oracle	31539
External	23
Concur	1508
Grand Total	**226580**



Общее затраченное время

- 1,5 дня и 250 долларов США в первый раз
  - если у вас никогда не было файлового архива
  - 30 часов на преобразование снимков за 2 года (60 ТБ)
  - 15 минут на поиск и анализ
- 0,5 часа и 1 доллар США за следующий раз
  - если вы поддерживаете свой файловый архив в актуальном состоянии
    - путем планирования инкрементного преобразования последних снимков за ночь
    - на компьютере с 4 процессорами преобразование снимков за последний день занимает 15 минут
  - 15 минут на поиск и анализ

Вы можете попробовать

- Клаудвизор
- Клаудвизор ES обозреватель резервных копий

И спасибо вам за ваше внимание!