Экспортер резервных копий Клаудвизор ElasticSearch версия v1.2

Инструмент экспорта резервных копий Клаудвизор ElasticSearch (ESBackupExporter.exe) предназначен для экспорта логов из вашего хранилища резервных копий ElasticSearch в сжатые текстовые файлы (*примечание: в настоящее время поддерживается только формат * .gz*) без восстановления в ES cluster. Инструмент поддерживает BLOB хранилища, бакет Хранилища S3 и общий файловый ресурс или локальную папку как для резервного копирования ES, так и для расположения экспорта, и позволяет выполнять как полный, так и инкрементный экспорт. Поскольку сжатые логи обычно занимают в 5-20 раз меньше места, чем Elasticsearch backup, инструмент можно запланировать на регулярной основе для хранения сжатых резервных копий в удобочитаемом формате, что позволит хранить больше исторических логов. Кроме того, фильтрация сообщений по индексу и логу позволяет использовать этот инструмент для создания исторических отчетов, экспортируя только интересующие логи.
После экспорта резервной копии вы можете использовать любой инструмент, считывающий файлы .gz, или использовать Клаудвизор (self-hosted) версию, чтобы получить наилучший опыт работы с логами.

Настройка местоположений резервного копирования и экспорта ES.

Перед использованием инструмента необходимо настроить папки резервного копирования и экспорта.
Откройте ESBackupExporter.exe.config и укажите свойства репозитория и выходных данных:
<добавить ключ=значение"репозитория"="elasticsearch-backup-location" />
<добавить ключ="выходное" значение="расположение экспортированных файлов" />
Эти свойства представлены в формате панели входа Клаудвизор и поддерживают следующие расположения:
бакет Хранилища S3:
тип=s3;ключ=*s3-ключхранилища***;секрет=***s3-хранилище-секрет***;регион=***s3-регион***;корзина=***s3-корзина*
Пример:
тип= s3; ключ= ABCDEF; секрет=1234567890; регион = запад США-1;сегмент=производство-логи-резервное копирование

Контейнер BLOB хранилища:
https:// хранилище**.blob.core.windows.net /**контейнер**?***параметры общего доступа*
Пример:
https://prodstorage.blob.core.windows.net/logs-backup?sv=2022-01-01&si=readonly&sr=c&sig=12345
Примечание: по соображениям безопасности для контейнеров BLOB хранилища поддерживаются только URL-адреса подписи общего доступа (SAS).

Локальная папка или общий ресурс:
Примеры:
e:/es-backup
\\резервное копирование \ логи\ prod\

Предупреждение: Когда этот инструмент запускается в хранилищах резервного копирования и / или экспорта в BLOB хранилище или Хранилище S3, очень важно запускать его на виртуальной машине в том же центре обработки данных BLOB хранилища или регионе Хранилища S3, чтобы избежать затрат на передачу данных! Если вы используете Клаудвизор для работы с экспортированными логами, также убедитесь, что вы запускаете его в том же центре обработки данных / регионе.

Запуск инструмента

Перед фактическим экспортом вы можете запустить инструмент в тестовом режиме, чтобы просмотреть и отфильтровать индексы, которые вам не нужно экспортировать:

ESBackupExporter.exe -тест

Примечание 1: обычно индексы, начинающиеся с точки ‘.’, представляют собой системные индексы elasticsearch / logstash / kibana, которые вы можете игнорировать при экспорте. В следующем разделе объясняется, как настроить фильтрацию индекса и другие параметры экспорта.
Примечание 2: вы можете экспортировать все индексы, оставив фильтр по умолчанию (‘’), но это может замедлить экспорт*

Чтобы запустить фактический экспорт, просто запустите инструмент без каких-либо параметров:

ESBackupExporter.exe

Инструмент прочитает все документы JSON из резервной копии Elasticsearch и запишет их в сжатые текстовые файлы, документ json в строке, один или несколько файлов в индексе ES. Инструмент выведет прогресс вместе с базовой статистикой по индексу на вывод консоли и запишет более подробные логи прогресса в файл exportlog.txt.
Примечание: при первом запуске инструмента для завершения работы может потребоваться много времени (в зависимости от возможностей вашего сервера)

Настройка параметров экспорта

Инструмент предоставляет разумные значения по умолчанию, с которыми вы можете начать работу, но позволяет настроить несколько параметров в зависимости от конфигурации вашего сервера. Вот список опций, которыми вы можете управлять:

<добавить ключ="потоки", значение="8" />

Указывает количество потоков для выполнения экспорта. Поскольку производительность растет линейно с увеличением количества используемых потоков, обычно рекомендуется иметь столько потоков, сколько процессоров на вашем компьютере, но вы можете уменьшить ее, если на сервере недостаточно доступной оперативной памяти или низкая пропускная способность сети.

<добавить ключ="memoryCachePerThreadMB" значение="256" />

Память, доступная каждому потоку для обработки и сортировки событий ES перед сохранением их в файл. Ее увеличение приведет к более быстрому экспорту, но потребует больше памяти. При настройке по умолчанию использование памяти может отличаться, но инструменту обычно требуется 0,5 – 1 ГБ оперативной памяти **на поток**.

<добавить ключ="репозиторий" значение="es-backup-storage" />

Исходное хранилище, в котором хранится резервная копия ES

<добавить ключ="выходное" значение="экспортировать-местоположение" />

Целевое хранилище, где будут храниться сжатые файлы .gz с экспортированными логами

<добавить ключ="запрос" значение="*" />

Фильтрует отдельные события. Фильтр выполнен в формате панели регистрации Клаудвизор (в xml-кодировке):
Простое слово: ошибка
Слово с подстановочными знаками: 10.*.*.27

Фраза: "ошибка сервера"
Фраза с подстановочными знаками: "Ошибка: *исключение"
Фильтр исключения: -отладка
Предложение AND (через пробел): исключение ошибки

Примечание: двойные кавычки должны быть закодированы как ' в формате xml.

Пример:
<добавить ключ="запрос" значение="ошибка -'ошибка клиента'" -debug />

Примечание 1: вы можете использовать опцию фильтрации для создания исторических отчетов через ES backup вместо полного экспорта.
Примечание 2: Оставьте значение по умолчанию '' для экспорта всех логов*.
Примечание 3: Предложение OR появится в следующем выпуске

<добавить ключ="indexNameFilter" значение="*" />

Фильтрует индексы ES по названию. Фильтр также доступен в формате панели регистрации Клаудвизор.

Пример:
<добавить ключ="indexNameFilter" значение="-.kibana* -logstash* -*2018*" />

Этот запрос фильтрует индексы, которые не начинаются с ".kibana" или "logstash" и не содержат в себе "2018" год.

Примечание 1: вы можете использовать этот параметр для исключения системных индексов, которые обычно начинаются с '.', указывая
<добавить ключ="indexNameFilter" значение="-." />*
Примечание 2: Оставьте значение по умолчанию '' для экспорта всех индексов ES*.

<добавить ключ="подробное" значение="false" />

В случае ошибок или других проблем (например, связанных с памятью или производительностью) установите для этого параметра значение true, чтобы инструмент записывал более подробные логи в файл exportlog.txt.
Примечание: включение этой опции приведет к тому, что размер exportlog.txt увеличится очень сильно

<добавить ключ="incrementalExport" значение="true" />

Если включено, инструмент пропускает уже экспортированные индексы при повторном запуске, при условии, что индексы не были изменены. Это полезно при планировании регулярного экспорта или для повторного экспорта только подмножества индексов, например, когда предыдущий экспорт не удался из-за проблем с сетью. Отключайте его только в особых случаях, например, при получении более новой версии инструмента, которая исправила некорректный экспорт в предыдущей версии.

<добавить ключ="sortEvents" значение="true" />

Поскольку сообщения лога не сортируются по индексам Elasticsearch, инструмент сортирует их по мере сбора и записывает частями фиксированного размера (в зависимости от параметра memoryCachePerThreadMB). После обработки всех сообщений, если для этого параметра установлено значение true, части объединяются в один или несколько файлов, где сообщения сортируются по временной метке, а дубликаты удаляются. Если эта опция отключена, части остаются при экспорте без объединения их вместе. Обратите внимание, что сообщения внутри этих частей по-прежнему сортируются, но у частей могут

пересекаться диапазоны временных меток. Кроме того, дедупликация не будет работать, поэтому две части могут содержать одно и то же событие (в зависимости от других настроек). Кроме того, при отключении этой опции инструмент работает в 2-5 раз быстрее и использует до 2 раз меньше памяти.
Примечание: если вас устраивают экспортированные индексные файлы, имеющие пересекающиеся диапазоны временных меток, и ваши индексы никогда не изменялись (т. Е. Добавлялись только новые сообщения), то следующая комбинация параметров обеспечит вам гораздо лучшую производительность: sortEvents=false; deduplicate=false; parshShardIndex=true; lastSnapshotOnly=true.

<добавить ключ="дедупликировать" значение="true" />

Поскольку резервная копия Elasticsearch может содержать несколько моментальных снимков, в резервной копии может быть несколько копий сообщений лога. Эта опция позволяет инструменту оставлять только одну копию таких сообщений, если экспортируется несколько снимков. Обычно эту опцию не следует отключать, поскольку она повышает производительность лишь на несколько процентов.
Примечание: для корректной работы дедупликации необходимо включить опцию сортировки.

<добавить ключ="parseShardIndex" значение="true" />

Эта опция позволяет инструменту анализировать файлы конфигурации резервной копии Elasticsearch, чтобы избежать чтения ненужных частей файлов резервной копии. Если он выключен, инструмент будет считывать все данные резервной копии ES, увеличивая использование сети в два раза. Если файлы конфигурации не могут быть проанализированы (например, из-за неподдерживаемого формата), инструмент автоматически вернется к чтению всех файлов данных, поэтому отключайте эту опцию, только если вы подозреваете, что инструмент неправильно считывает эти файлы конфигурации.

<добавить ключ="lastSnapshotOnly" значение="false" />

Резервная копия ES может содержать несколько моментальных снимков для каждого индекса. Логи в ES удаляются по индексу, а не по сообщению, поэтому последний снимок обычно содержит все данные, которые есть в предыдущих снимках, поэтому включите эту опцию, чтобы экспортировать только последний снимок и пропустить все предыдущие снимки. Однако, если индекс был удален между снимками или некоторые сообщения лога были удалены вручную (с помощью ES API), то предыдущие снимки могут содержать данные, которых нет в последнем снимке. Это также может произойти, если последние снимки находятся в неполном или сбойном состоянии. По умолчанию эта опция отключена, поскольку, хотя ее включение обычно увеличивает производительность на 5-40%, это может привести к отсутствию сообщений лога при экспорте. Включайте его только в том случае, если вы уверены, что сообщения не были удалены между снимками, или если вас интересует только последний снимок, и ваши снимки всегда получаются успешными.

Ограничения

Инструмент разработан для обеспечения максимально возможной производительности, поэтому он не использует API Elasticsearch; вместо этого он анализирует файлы резервных копий и извлекает документы json, используя свой собственный код. Инструмент был протестирован на резервных копиях Elasticsearch 7.4, но если более новые версии Elasticsearch изменят формат данных, инструмент может начать работать некорректно. Пожалуйста, свяжитесь со службой поддержки Клаудвизор, если вы хотите поддерживать другие версии Elasticsearch.

Инструмент делает все возможное, чтобы извлечь временную метку из логов, но иногда бывает сложно определить, какая временная метка является правильной, потому что логи обычно имеют несколько разных временных меток. В настоящее время он поддерживает базовые варианты формата даты и времени ISO 8601, например 2020-01-01T00: 00: 00.000. Пожалуйста, свяжитесь с Клаудвизор, если вы хотите, чтобы он поддерживал больше форматов.

Инструмент пытается найти временную метку в следующем порядке:

- Выполняет поиск поля “@timestamp”, которое обычно используется Elasticsearch /Logstash
- Выполняет поиск первого поля, содержащего временную метку в ожидаемом формате
- Выполняет поиск временной метки в ожидаемом формате в любом месте сообщения лога.

Пожалуйста, обратитесь в службу поддержки Клаудвизор, если в ваших логах используется другой шаблон временных меток.

Производительность

Примечание: цифры в этом разделе могут значительно отличаться в зависимости от вашего случая

С настройками по умолчанию и папкой резервного копирования и экспорта ES в BLOB хранилище или Хранилище S3 инструмент обычно обрабатывает 100-200 ГБ файлов резервных копий ES в час и создает 5-20 ГБ экспортированных логов индекса. Лучший способ повысить производительность - использовать сервер с большим объемом процессора (обычно пропускная способность сети не является узким местом в BLOB хранилище / Хранилище S3).
Пример 1:
ES backup содержит 100 индексов и 200 ГБ общих данных в корзине Хранилища S3
Инструмент запускается на сервере с 4 ядрами процессора и 8 ГБ оперативной памяти и настройками по умолчанию
Полный экспорт занимает 2 часа и дает 25 ГБ сжатых текстовых файлов.
Пример 2:
ES backup содержит 10000 индексов и 25 ТБ общих данных в корзине Хранилища S3
Инструмент запускается на сервере с 72 ядрами процессора и 180 ГБ оперативной памяти
Для параметра "потоки" установлено значение 100, а все остальные настройки установлены по умолчанию
Полный экспорт занимает 20 часов и позволяет получить сжатые текстовые файлы объемом 3 ТБ.
Пример 3:
Резервная копия ES содержит 10000 индексов и 25 ТБ общих данных в корзине Хранилища S3, а экспорт был запущен два дня назад.
Инструмент запускается на сервере с 4 ядрами процессора и 8 ГБ оперативной памяти и indexNameFilter=”-*beat-* -elastalert* -logstash -.*”
Частичный экспорт занимает 2 часа и позволяет получить дополнительные 13 ГБ сжатых текстовых файлов.