Backups, snapshost y restore en Elasticsearch 8

Introducción

Uno de los temas más importantes, como siempre, es el de los backups. En el caso de la Elasticsearch, llamados snapshots. Aunque la Elasticsearch dispone de una buena documentación, dejo por aquí algunos tips que he aprendido con el tiempo.

> Ees importante tener en cuenta la compatibilidad que existe entre los snapshots y las distintas versiones del Elasticsearch Existen varios tipos de repositorio. Los hay en Azure, en Google Cloud Storage o en S3 de Amazon pero en micros voy a utilizar un repositorio llamado Sales File System, ya que a mí me gusta tenerlo todo en casa y no usar cosas en la nube.

Snapshot and Restore

Procedimientos para la creación de un sistema de backups en Elasticsearch (snapshots)

Lo primero que tenemos que hacer es registrar un repositorio para almacenar las instantáneas que vamos a realizar. Para ello debemos disponer de los siguientes permisos:

- privilegios de clase
- privilegios de índice

Repositorio del sistema de archivos compartido

El repositorio del sistema de archivos compartidos sólo está disponible así se ejecuta Elasticsearch, en nuestro propio hardware.

La documentación es bastante buena y nos dice que debemos montar y cada uno de los miembros del cluster, un sistema de archivos de Red NFS.

Repositorios

Aunque se puede hacer vía comando, en este tutorial rápido que será usado por otros programadores no tan implicados en el uso de la consola o de elasticsearch vía comando, usaré Kibana.

Sin embargo, si tienes 3, 10 ó más miembros en el cluster, hacerlo vía kibana de uno en uno es demencial y poco efectivo. :smile:

Punto de montaje

Lo primero es tener ya nuestro punto de montaje NFS en todos los miembros del cluster,

```
# df -h
nfs.server:/srv/storage/elasticsearch/  45T   39T  6.4T  86% /mnt/nfs/elasticsearch
```

Después es tener un punto específico para cada cluster (si tenemos más de uno es necesario, ya que usar el mismo punto es poco usable y desaconsejado por el propio Elasticsearch.

```
# ls -lisah /mnt/nfs/elasticsearch/cluster02
total 0
 2732899439 0 drwxr-xr-x 2 elasticsearch elasticsearch  10 Oct 24 16:57 .
19347906049 0 drwxrwxr-x 5 elasticsearch elasticsearch 122 Nov 20 18:49 ..
```

Bonus NFS

Uno de los problemas que podéis encontraros en el montaje de un punto NFS es escribirlo por todos los miembros del cluster. Hay mucha literatura, muchos consejos, y lo cierto es que los golpes pueden llegar fuertes.

En mi caso, yo tengo máquinas de backups y NFS, altamente secularizadas ya que son solamente accesible desde las máquinas que comparten NFS o usan SSH, y cada una estas autorizada en el firewall, con un modelo de **denegar todo primero-abrir puerto a quien lo necesita** Así que uso un NFS sin autenticación de usuario, y esto obliga a que todos usen el mismo usuario. Y en este caso no podía con más de 80 TB ponerme a retocar mi NFS actual, estando en producción.

Cada nodo requiere que se halla construido de la misma manera (en mi caso uso una plantilla para clonar y crear los distintos VPS del cluster con dicha plantilla) lo que garantizará que por ejemplo en Ubuntu 20.04 **elasticsearch** tenga como usuario el **uid 113**, y el **gid 117**

Esto, nos permitirá usar un modo anonimo en todos los miembros del cluster.

En el servidor nfs en `/etc/exports` añado el punto de compartición

```
/srv/storage/elasticsearch/IP_NODO_001(rw,nohide,insecure,no_subtree_check,sync,anonuid=113,anongid=117)
IP_NODO_002(rw,nohide,insecure,no_subtree_check,sync,anonuid=113,anongid=117)
...
IP_NODO_nnn(rw,nohide,insecure,no_subtree_check,sync,anonuid=113,anongid=117)
```

En cada nodo añado el montaje nfs a al `/etc/fstab`

```
[nfs_server]:/srv/storage/elasticsearch/ /mnt/nfs/elasticsearch nfs bg,hard,timeo=1200,rsize=1048576,wsize=1048576,sec=sys 0 0
```

Configuración del montaje en Elasticsearch

Es necesario configurar elasticsearch para que lea este nuevo path como repositorio de ficheros. Para ello necesitamos configurar la variable `path.repo` en el fichero `/etc/elasticsearch/elasticsearch.yml`

> Esta variable acepta valores separados por comas `,`

```
path.repo: /mnt/nfs/elasticsearch/cluster01,/mnt/nfs/elasticsearch/cluster02
```

Creación en Kibana del repositorio

Entraremos en Kibana a **registrar** nuestro repositorio.

Registrar el repositorio

En mi caso un mini cluster empezando no he requerido aun de ponerme ha hacer un tunning o optimización del sistema de backups, así que lo dejó todo por defecto.

Sólo pongo la situación del repositorio.

Configuración del repositorio

Con eso ya está creado nuestro repositorio.

Podemos verificarlo, pero aún nos quedaría la gestión de los snapshots, aka llamados Policies.

Políticas

Básicamente el sistema nos pide que creemos una política de snapshots. Las variantes son muchas, y no es el alcance de este tutorial. En mi caso y por las características de mi sistema (Cluster de VPS formato KVM con snapshots diarios), solo requiero de backups cada hora de cada

uno de los índices del sistema, de forma separada, por si ocurriera algún desastre procedente de una manipulación o edición indebida.

Y lo hago por separado, por el método de restauración que va en bloque con cada política de snapshots, no pudiéndose (o al menos yo no lo conozco) separarse cuando es requerida una restauración, como pudiéramos hacer como por ejemplo con la restauración de un dump completo de Mysql

Logística

- Nombre que le daremos a la política
- Expresión para nombrar los distintos snapshots. Consultar la ayuda de las expresiones es importante
- Repositorio donde se guardarán los distintos snapshots
- Creación del cron o expresión de horario

> Atención que despista un poco el constructor de cron, porque la sintaxis despista con el **?** que en realidad es la variable del comando que ejecutará. Nada más.

CReate Policy : Logistics

Configuración

Como dije, quiero hacer **sólo** los backups de un **índice** y no de todo el conjunto global. Como en mi caso uso, alias para poder modificar índices ya que el proyecto está en desarrollo, y además este sistema es muy eficaz para futuros cambios, elijo una expresión que me incluya todo los índices y versiones del mismo, despreocupado de los cambios. Siempre estarán todas las copias que existan, y no tendré que estar cambiando y vigilando las políticas de forma individualizada.

Snapshot settings

Política de retención

Esto es personal y adaptable a las necesidades de cada uno, como todo.

Snaopshot retention (optional)

Revisión

Ya sólo queda revisar la configuración y darle a crear.

Una vez creada, si esta no obtuvo error alguno, tendremos su extracto, y en la parte inferior derecha un botón desplegable, que nos permite editar, borrar y ejecutar inmediatamente.

Restauración

Para la restauración ya lo dejo en tus manos. Es una cuestión compleja esta, y no está a mi alcance explicaría. Lo que sí puedo decir, es, que como todo sistema de backups, **no sirve para nada si no existe una política de comprobación de las copias de seguridad**

Es decir, que tienes que probar en un cluster de pruebas, que esto te funciona, que puedes restaurarlo, que lo documentas, y que regularmente lo compruebas. De lo contrario, patatas como las que se comen en muchos sitios oficiales, hospitales españoles, o grandes empresas, por no tener una política de recuperación de desastres, unas veces por la impericia del supuesto Director de Informática, otras porque una empresa les vendido un cuento de seguridad, cuando en realidad, les engañaron y les cobraron, por nada.

Aviso

Esta documentación y su contenido, no implica que funcione en tu caso o determinados casos. También implica que tienes conocimientos sobre lo que trata, y que en cualquier caso tienes copias de seguridad. El contenido el contenido se entrega, tal y como está, sin que ello implique ningún obligación ni responsabilidad por parte de Castris

Si necesitas soporte profesional puedes contratar con Castris soporte profesional.

Revision #4
Created 2023-03-10 13:41:04 UTC by Abkrim
Updated 2023-03-17 19:54:44 UTC by Abkrim