

# Elasticsearch Snapshot 備份實作

## 實作對象

- Elasticsearch 7.15.2
  - elasticsearch-infra-01
  - elasticsearch-infra-02
  - elasticsearch-infra-03
- Elasticsearch 6.3.0
  - elasticsearch-backend-01
  - elasticsearch-backend-02
  - elasticsearch-backend-03

## 實作細節

- 以下每個動作，每台機器都要進行！
- 網路環境是走 Amazon *Virtual Private Cloud* (Amazon VPC) 到 S3，外網無法使用

## 查看服務是否正在運行

```
## 建議全部都以 root 身分進行
sudo su

## 查看
# 或
sudo systemctl status elasticsearch
```

## 若服務尚未啟動，則用以下方式啟動

```
## 若指令有寫入 systemctl 就可以這樣做
systemctl start elasticsearch

#####

## 若指令未寫入 systemctl 就必須這樣做

# 確認是否有建立 elasticsearch 專用 user，如 els/elasticsearch/esuser 等
# elasticsearch 不允許用 root 身分來啟動服務
ls -al /home/

# 找尋資料夾位置
find / -name "elasticsearch" -type f

# 進入所找到的資料夾位置，以 elasticsearch 專用 user 啟動服務
cd /opt/elasticsearch-6.3.0/
su els
./bin/elasticsearch &

#####

## 確認服務是否順利啟動：9200 & 9300
netstat -ntulp | grep :9
# 或
systemctl status elasticsearch
```

## 安裝必要套件 (在允許外網連線的時候)

```
## 安裝 repository-s3 這外掛套件
cd /opt/elasticsearch-6.3.0/bin
./elasticsearch-plugin install repository-s3

## 順利安裝成功，會印出 repository-s3
cd /opt/elasticsearch-6.3.0/bin
./elasticsearch-plugin list
# 或
curl -XGET `hostname -i | cut -d ' ' -f 2`:9200/_nodes?filter_path=nodes.*.plugins

#####

## 下載 AWS CLI
# 參考網頁：https://docs.aws.amazon.com/cli/latest/userguide/getting-started-install.html
# 以下指令可能不同，以網頁內如為主
# 建議先回到 cd ~ 的位置

curl "https://awscli.amazonaws.com/awscli-exe-linux-x86_64.zip" -o "awscliv2.zip"

# 為了解壓縮而需要先安裝
yum install -y unzip zip

unzip awscliv2.zip
sudo ./aws/install

# 確認安裝成功 by 查看版本
/usr/local/bin/aws --version
```

## 基本設定之前

- 必須要向帳號管理者取得專案的 AccessKey & SecretKey，才能繼續以下動作

## 基本設定

```
## 針對 repository-s3 設定
cd /opt/elasticsearch-6.3.0/bin
echo "${access_key}" | ./elasticsearch-keystore add s3.client.default.access_key
echo "${secret_key}" | ./elasticsearch-keystore add s3.client.default.secret_key

#####

## 針對 aws cli 設定
/usr/local/bin/aws configure

# AWS Access Key ID [None]: 填入 Access Key
# AWS Secret Access Key [None]: 填入 Secret Access Key
# Default region name [None]: 填入專案 region，例如 VS/VST 都是 ap-northeast-1
# Default output format [None]: 可以不填寫
```

## elasticsearch 6.3.0 要額外做的事情

- 在 elasticsearch.yml 裡面新增一行：`s3.client.default.endpoint:s3.[專案所在的_region].amazonaws.com`
- 系統預設：`s3.client.default.endpoint: s3.amazonaws.com`
- 若未設定此項而繼續下面的動作，elasticsearch 的 snapshot 會永遠只連線到 `s3.amazonaws.com` 而找不到 bucket

## 逐一重啟服務

```
## 若指令有寫入 systemctl 就可以這樣做
systemctl restart elasticsearch

## 若指令未寫入 systemctl 就可以這樣做
# 確認是否有建立 elasticsearch 專用 user，如 els/elasticsearch/esuser 等
ls -al /home
# 若有建立 elasticsearch 專用 user，則要讓它有以下兩檔案的權限
find / -name "elasticsearch" -type f
chown (els)(elasticsearch)(esuser):root elasticsearch.keystore
chown (els)(elasticsearch)(esuser):root elasticsearch.yml

# 以此找尋 elasticsearch 相關資料夾位置
```

```
find / -name "elasticsearch" -type f

#####

## 重啟服務：為了 reload 以得到新設定
# 在 prod 要依序重開，然後才能做其他動作

# 服務停止
netstat -ntlp | grep -E ':9200|:9300' | awk '{print $7}' | awk -F '/' '{print $1}' | xargs kill

# 確認服務是否順利停止: 9200 & 9300
netstat -ntulp | grep :9
# 或
systemctl status elasticsearch

## 服務啟動
cd /opt/elasticsearch-6.3.0/bin
su els
./elasticsearch &
```

## 在 elasticsearch 裡面建立 S3 的資訊

- 這指令只要在其中一台執行就好
- bucket：要先在 s3 建立好 (指令：`/usr/local/bin/aws s3 mb s3://[_bucket_的名字_] --region [專案所在的_region]`)
- region：先查好專案的 region，例如 VS/VST 都是 ap-northeast-1

```
## _snapshot/backup -> _snapshot/[任何 repository_name]
curl -XPUT `hostname -i | cut -d ' ' -f 2`:9200/_snapshot/backup -H 'Content-Type: application/json' -d '
{
  "type": "s3",
  "settings": {
    "bucket": "[_bucket_的名字_]",
    "region": "[專案所在的_region]",
    "endpoint": "s3.[專案所在的_region].amazonaws.com",
    "compress": "true",
    "max_snapshot_bytes_per_sec": "8mb"
  }
}'

## 確認上述指令有確實寫入 elasticsearch
curl -X GET `hostname -i | cut -d ' ' -f 2`:9200/_snapshot/backup?pretty"

## 確認現有的 repositories
curl -XGET `hostname -i | cut -d ' ' -f 2`:9200/_cat/repositories?v=true
```

## 設定 elasticsearch snapshot 週期

- 參考網頁

### Tutorial: Elasticsearch Snapshot Lifecycle Management (SLM) - Coralogix

Let's face it, nothing is perfect. The better we architect our systems, though, the more near-perfect they become. But even so, someday, something is likely to go wrong, despite our best effort. Part of preparing for the unexpected is regularly backing up our data to help us recover from eventual failures and this tutorial

🟢 <https://coralogix.com/blog/tutorial-elasticsearch-snapshot-lifecycle-management-slm/>



Snapshot Lifecycle Management

Scaling with large numbers of subfields

```
## 設定內容說明
# schedule: a cron expression: <second> <minute> <hour> <day_of_month> <month> <day_of_week> [year], with the year parameter being optional
# indices: indices: By using the special asterisk wildcard character "*" which include all indices in the cluster.
# retention and expire_after: periodically remove all snapshots that are older than 60 days.

## Elasticsearch 7.15.2 的指令
curl -XPUT `hostname -i | cut -d ' ' -f 2`:9200/_slm/policy/backup_policy_hourly_s3 -H 'Content-Type: application/json' -d '
{
  "schedule": "0 0 * * * ?",
  "name": "<backup-{now{YYYYMMdd.hhmm}}>",
  "repository": "backup",
  "config": {
```

```

        "indices":["*"]
    },
    "retention": {
        "expire_after": "60d"
    }
}
}'

## Elasticsearch 6.3.0 的指令
curl -XPUT `hostname -i | cut -d ' ' -f 2`:9200/slm/policy/backup_policy_hourly_s3 -H 'Content-Type: application/json' -d '
{
    "schedule": "0 0 * * * ?",
    "name": "<backup-{now{YYYYMMdd.hhmm}}>",
    "repository": "backup",
    "config": {
        "indices":["*"]
    },
    "retention": {
        "expire_after": "60d"
    }
}
}'

```

```

##### 針對 Elasticsearch 6.3.0 的特別說明 #####
## 前面都已經把 slm 都設定完成，但是經試驗後它並沒有照排程時間執行
## 故在此利用主機的 crontab 將它設定為類手動執行，非由 slm 來主導
$ vim snapshot.sh
#!/bin/bash
#####
# Project: Run the elasticsearch snapshot
# Branch:
# Author: Gok, the DBA
# Created: 2023-02-07
# Updated: 2023-02-07
# Note: Snapshot lifecycle management (SLM) doesn't work as expected
#####

date=`date +%Y%M%d.%H%M`

## 執行指令
curl -XPUT `hostname`:9200/_snapshot/backup/backup-$date

```

```

## 確認 snapshot 狀態
$ curl -XGET `hostname`:9200/_snapshot/_status?pretty

## 直接查看特定 snapshot 的語法
## 時間戳為系統時間，故 20230207.14:00 所做備份為 20230207.0600
## 現行的 snapshot 名稱後面會帶一組亂數

## 建議可以先用以下方式逐一尋找切確的 snapshot 名稱
$ curl -XGET `hostname`:9200/_snapshot/backup/_all?pretty | grep 'snapshot'
% Total % Received % Xferd Average Speed Time Time Time Current
Dload Upload Total Spent Left Speed
100 7419 100 7419 0 0 1290k 0 --:--:-- --:--:-- "snapshots" : [
-- "snapshot" : "backup-2023.02.07-leqaxcbqrb2kg6z-a-txyg",
--:- "snapshot" : "backup-20230207.0600-oecnvoozt181abtu1kgja",
-:- 1449k

## 深入查找
$ curl -XGET `hostname`:9200/_snapshot/backup/[snapshot 名稱]
$ curl -XGET `hostname`:9200/_snapshot/backup/[snapshot 名稱]/_status

## 實際例子
$ curl -XGET `hostname`:9200/_snapshot/backup/backup-20230207.0600-oecnvoozt181abtu1kgja
$ curl -XGET `hostname`:9200/_snapshot/backup/backup-20230207.0600-oecnvoozt181abtu1kgja/_status

```

## 參考網頁

- [elasticsearch 官方網頁](#)

### S3 Repository Plugin | Elasticsearch Plugins and Integrations [6.3] | Elastic


Elastic Docs » Elasticsearch Plugins and Integrations [6.3] » Snapshot/Restore Repository Plugins

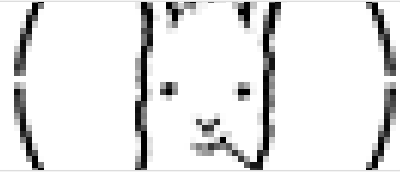
 <https://www.elastic.co/guide/en/elasticsearch/plugins/6.3/repository-s3.html>

- 其他

#### elasticsearch 6.3.0 快照


```
install plugin bin/elasticsearch-plugin install repository-s3 set up client bin/elasticsearch-keystore add  
s3.client.default.access_key bin/elasticsearch-keystore add s3.client.default.secret_key vim  
/etc/elasticsearch/elasticsearch.yml add configure s3.client.default.endpoint: s3.cn-northwest-
```

 <https://www.twblogs.net/a/5f0484cde94612e3fceabbb4>



#### Take and Restore Snapshots

The Open Distro project is archived. Open Distro development has moved to OpenSearch. The Open Distro plugins will continue to work with legacy versions of Elasticsearch OSS, but we recommend upgrading to OpenSearch to take advantage of the latest features and improvements. Snapshots are backups of a cluster's indices and state.

 <https://opendistro.github.io/for-elasticsearch-docs/docs/elasticsearch/snapshot-restore/>

#### AWS CLI S3 Configuration - AWS CLI 1.27.29 Command Reference

These are the configuration values you can set specifically for the command set: max\_concurrent\_requests - The maximum number of concurrent requests.  
max\_queue\_size - The maximum number of tasks in the task queue. multipart\_threshold - The size threshold the CLI uses for multipart transfers of individual files.

 <https://docs.aws.amazon.com/cli/latest/topic/s3-config.html>