

Elasticsearch

6장 클러스터 운영

6.1 클러스터 설정 API

```
_cluster/settings
{
  "persistent" : {},
  "transient" : {}
}
```

Persistent 밑에 지정하는 설정은 클러스터를 풀 리스타트로 전체 재시작해도 유지된다.

Transient 밑에 지정하는 설정을 클러스터를 전체 재시작하면 내용이 사라진다. 그러므로 임시 설정을 지정하기에 좋다

Persistent 와 Transient에 동일한 설정이 지정되면 Transient 설정이 우선 적용된다.

엘라스틱 서치의 설정 적용 우선 순위는 Transient > Persistent > config/elasticsearch.yml 파일 순이다. 설정은 모든 마스터 후보 노드의 path.data 경로내 파일로 저장된다.

Cat API

엘라스틱의 여러 현재 상태를 조회할 수 있는 API

대부분의 엘라스틱서치 REST API는 JSON 형식이 기본이다. Cat API는 터미널에서 사람이 조회했을때 보기 편한 형태로 응답하는 것이 목적이다.

cat API의 기본 엔드 포인트는 `/_cat`이다. 매개 변수가 없이 호출하면 cat API에 사용할 수 있는 모든 엔드 포인트를 표시한다.

```
curl -XGET 'localhost:9200/_cat'
=^,^=
/_cat/allocation
/_cat/shards
/_cat/shards/{index}
/_cat/master
/_cat/nodes
/_cat/tasks
/_cat/indices
/_cat/indices/{index}
/_cat/segments
/_cat/segments/{index}
/_cat/count
/_cat/count/{index}
/_cat/recovery
/_cat/recovery/{index}
/_cat/health
/_cat/pending_tasks
/_cat/aliases
/_cat/aliases/{alias}
/_cat/thread_pool
/_cat/thread_pool/{thread_pools}
/_cat/plugins
/_cat/fielddata
/_cat/fielddata/{fields}
/_cat/nodeattrs
/_cat/repositories
/_cat/snapshots/{repository}
/_cat/templates
```

각각의 cap api 요약을 한다면 다음과 같다.

- 클러스터에서 현재 실행 중인 작업
- 세그먼트 통계
- 세그먼트 통계(특정 인덱스로 제한)
- 샤드 할당 - 관련 정보
- 필드 데이터 캐시 크기
- 개별 필드의 필드 데이터 캐시 크기
- 복구 정보
- 복구 정보(특정 인덱스로 제한)
- 클러스터에 등록된 스냅샷 저장소 정보
- 사용자 정의 노드의 속성 정보
- 색인 통계
- 색인 통계(특정 색인으로 제한)
- 특정 저장소에 속한 모든 스냅샷에 대한 정보
- 각 노드에 설치된 플러그인
- 특정 별명의 인덱스 앨리어스와 인덱스
- 마스터 정보(선출된 마스터 표시 포함)
- 마스터 노드 정보
- 클러스터 상태
- 실행 대기 중인 작업
- 클러스터당 노드별 스레드 풀 정보
- 클러스터당 노드별 단일 또는 다중 스레드 풀에 대한 스레드 풀 정보
- 전체 클러스터 또는 개별 인덱스의 다큐먼트 개수
- 모든 샤드 관련 정보(특정 인덱스로 제한)

6.1 클러스터 설정 API

대량 색인

서비스 출시를 위한 초기 데이터 세팅이나 마이그레이션, reindex를 통한 인덱스 재생성 등 대량 색인이 필요할때 색인 도중 데이터를 조회할 필요가 없으므로 해당 설정으로 색인 속도를 높일수 있다. 작업 끝나면 원래설정으로 복원한다.

PUT my_index/_settings

```
{
  "refresh_interval" : "-1",
  "number_of_replicas" : 0
}
```

샤드의 크기와 개수 조정

클러스터에 샤드가 너무 많아지면 클러스터 성능이 눈에 띄게 떨어진다.

샤드 하나당 20~40GB 크기가 적절하다고 한다. 하지만 실제로 하면 20GB도 다양한 상황에서 꽤 느리고 무겁다. 수 GB 내외 수준에서 조정하자.

합 1GB 당 20 개 이하의 샤드를 들고 있는 것이 적절하다고 한다.

스냅샷과 복구

엘라스틱서치는 데이터를 파일 기반으로 저장한다. 그러므로 데이터 디렉토리를 그대로 복제해 두면 백업이 된다고 생각하지만 지원하지 않는다. 스냅샷기능을 사용하자.

Put _snapshot/[저장소이름]

```
{
  "type" : "fs",
  "settings" : {
    "location" : "~~"
  }
}
```

Fs 타입은 공유 파일 시스템을 저장소 타입으로 사용할 때 이용한다.

Location 설정에는 공유파일시스템이 마운트된 경로를 입력한다. 이 마운트 경로는 elasticsearch.yml에 path.repo 설정으로 등록해야 한다.

6.1 클러스터 설정 API

스냅샷을 생성하고 조회하기

PUT _snapshot/[저장소이름]/[스냅샷이름]?wait_for_complete=false

```
{
  "indices": "my-index-202108*"
}
```

스냅샷에서 인덱스 복구하기

Post _snapshot/[저장소이름]/[스냅샷이름]/_restore

```
{
  "indices": "my-index-202108*",
  "include_global_state": false,
  "feature_states": ["kibana"]
}
```

서킷 브레이커

엘라스틱 서치는 과도한 요청이 들어오면 거부하는 정책을 채택했다.

슬로우 로그 설정

– 검색이나 색인 작업시 너무 오랜시간이 소요되면 별도로 로그를 남기도록 설정할 수 있다.

