

Elasticsearch 튜토리얼

안녕하세요 정보통신공학부 김준성입니다.

이번 튜토리얼에서는 Elasticsearch에 대해 알아보고 빌드 및 실행 해보도록 하겠습니다.

1. Elasticsearch란?

Elasticsearch는 Apache Lucene(아파치 루씬) 기반의 Java 오픈소스 분산 검색 엔진입니다. Elasticsearch를 통해 루씬 라이브러리를 단독으로 사용할 수 있게 되었으며, 방대한 양의 데이터를 신속하게, 거의 실시간으로 저장, 검색, 분석할 수 있습니다.



Elasticsearch는 검색을 위해 단독으로 사용되기도 하며, ELK(Elasticsearch / Logstatsh / Kibana)스택으로 사용되기도 합니다.

Logstash

다양한 소스(DB, csv 파일 등)의 로그 또는 트랜잭션 데이터를 수집, 집계, 파싱하여 Elasticsearch로 전달

Elasticsearch

Logstash로부터 받은 데이터를 검색 및 집계를 하여 필요한 관심 있는 정보를 획득

Kibana

Elasticsearch의 빠른 검색을 통해 데이터를 시각화 및 모니터링

2. 예제

Elasticsearch를 사용하는 간단한 예제를 해보겠습니다. 이 예제는 Restful API로 document를 추가하고 조회하는 예제입니다.

Elasticsearch CRUD 작업은 API를 호출해서 이루어지며, curl로 데이터를 넘겨줄 수도 있지만, json 포맷으로 queryDSL을 작성해서 API를 호출할 수도 있습니다.

```
# curl -XPOST 'localhost:9200/victolee/blog/1?pretty' -d '{"postName": "elasticsearch", "category": "IT"}' -H 'Content-Type: application/json'
```

-d 옵션: 추가할 데이터를 json 포맷으로 전달합니다.

-H 옵션: 헤더를 명시합니다.

이렇게 curl 요청을 하면 blog 타입으로 id 값이 1인 document가 저장됩니다.

```
[root@victolee ~]# curl -XPOST 'localhost:9200/victolee/blog/1?pretty' -d '{"postName": "elasticsearch", "category": "IT"}' -H 'Content-Type: application/json'
{
  "_index": "victolee",
  "_type": "blog",
  "_id": "1",
  "_version": 1,
  "result": "created",
  "shards": {
    "total": 2,
    "successful": 1,
    "failed": 0
  },
  "_seq_no": 0,
  "_primary_term": 1
}
```

document 조회하는 방법을 알아보겠습니다.

```
# curl -XGET 'localhost:9200/victolee/blog/1?pretty'
```

이렇게 curl 요청을 해주면 document의 값이 조회됩니다.

```
[root@victolee ~]# curl -XGET 'localhost:9200/victolee/blog/1?pretty'
{
  "_index" : "victolee",
  "_type" : "blog",
  "_id" : "1",
  "_version" : 1,
  "_seq_no" : 0,
  "_primary_term" : 1,
  "found" : true,
  "_source" : {
    "postName" : "elasticsearch",
    "category" : "IT"
  }
}
```

Elasticsearch를 잘 다루기 위해서는 Elasticsearch에서 제공하는 API를 알아둘 필요가 있으며 queryDSL을 통해 작성해 원하는 방식으로 다양하게 조회할 수 있어야 합니다.