

## 1. 데이터 색인

- 색인 (indexing) : 원본 문서 → 검색이 용이하도록 변환하여 저장
- 인덱스 (index) : 색인 과정을 거친 결과를 또는 색인된 데이터를 저장하는 저장소

## 2. 클러스터

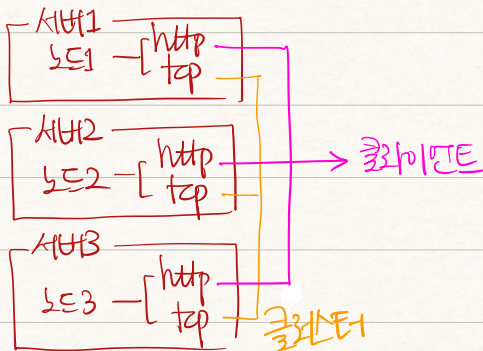
- 1개의 노드만 있어도 클러스터 구성
- 노드의 네트워크 통신

http 포트 : 클라이언트와의 통신  
tcp 포트 : 노드 간의 데이터 교환

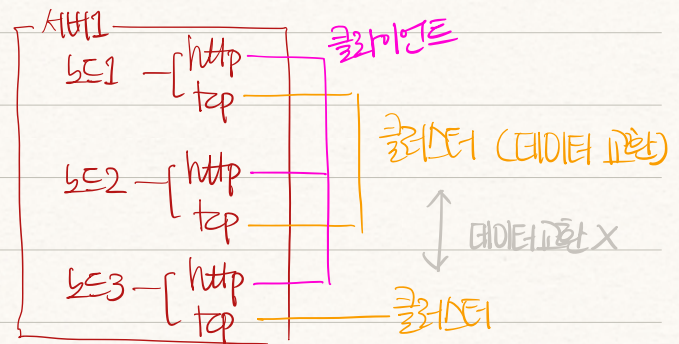
- 클러스터명 설정 (cluster.name) 동일한 노드끼리 하나의 클러스터로 묶임
- 종류

1마리 서버 1노드

일반적



1마리 서버 여러 노드 ~ 클러스터



- 디스커버리 and 미리 설정된

노드가 처음 실행될 때 같은 서버의 노드 찾아 하나의 클러스터로 바인딩

## 3. 인덱스 & 샤드

- 문서 : 단일 데이터 단위
- 인덱스 / 인디스 : 문서의 집합
- 색인 : 데이터를 Elasticsearch에 저장하는 행위

## • 인덱스

샤드 단위로 분리되고 각 노드에 분산되어 저장

↳ 루션의 단일 검색 인스턴스

## • 샤드

[ 프라이머리 샤드 : 처음 생성된 샤드

레플리카 : 복제본. 원본과 동일 데이터. 서로 다른 노드

⇒ 노드 추가하면 프라이머리 샤드 분산 & 1개 레플리카 생성

최소 3개의 노드 추천. 노드 유실되어도 분산 복제 데이터 살아남음

인덱스 처음 생성할 때 개수 지정

→ 재배인 아니면 바꿀 수 X

레플리카 개수는 바꿀 수 있음!

## • 노드 종류

마스터 노드: 인덱스 메타 데이터, 샤드 위치 등 클러스터 상태 관리

없으면 클러스터 작동 중지

— 마스터 후보 노드는 마스터 노드 정보 갖고있어 전환가능

`node.master` 설정 true/false 관리 가능

데이터 노드: 실제 색인된 데이터 저장

`node.data` 설정 true/false 관리 가능

여하튼 따라  
저장 데이터 설정

## • Split brain

마스터 후보 노드는 3개 이상의 활성

↳ 그렇지 않으면 네트워크 단절 → 각자 동작 → 복구 → 데이터 정합성 문제



minimum\_master\_node : ( 전체 노드 / 2 ) + 1

↳ 7.0부터는 자동 관리