

Elasticsearch Modeling

정호욱

- **미미박스 검색개발팀 리더**



- [실무 예제로 배우는 Elasticsearch 검색엔진 – 활용편] / 한빛미디어
- [실무 예제로 배우는 Elasticsearch 검색엔진 – 기본편] / 한빛 미디어
- Elasticsearch 성능 최적화 / DeView 2014

- [Meetup] 이커머스 스타트업에서 검색 인프라 구축하기 – 2016.03.31
- [Meetup] Elasticsearch on Apache Tajo – 2015.06.04

오늘 다룰 주제

Modeling

Modeling

- Elasticsearch cluster.
- Elasticsearch node.
- Elasticsearch index.
- Elasticsearch shard.
- Elasticsearch document.
- Elasticsearch store.

Elasticsearch cluster

- Single cluster 구성을 추천 합니다.
 - Cluster = Classification = Category
- Cluster HA(High Availability) 구성이 필요한가요?
- Multiple clusters 구성이 필요한가요?
- Tribe 노드를 활용하세요.

Elasticsearch node

- Master node는 꼭 HA 구성을 해주세요.
- 서비스 특성을 먼저 파악해 주세요.
 - 검색 질의가 많나요?
 - 집합 질의가 많나요?
 - 벌크 색인이 많나요?
 - 데이터에 대한 변경이 많나요?
- Data node는 검색과 색인 용도에 따라 검토해 주세요.
- Preference를 활용하세요.
- Box type(hot/warm)을 활용하세요.

Elasticsearch index

- Time series data ? non-time series data ?
- Index split ? or type split ?
- Index template 을 활용하세요.

Elasticsearch shard

- 색인이 중요? 검색질의가 중요? 또는 둘 다 중요?
 - 장비 스펙이 중요합니다.
- Core 크기는 얼마나 되나요?
- Data node는 몇 개 인가요?
- Primary shard는 기본 색인 성능과 관계가 있습니다.
- Replica shard는 기본 검색 질의 성능과 관계가 있습니다.
- Shard 하나에 배치 시킬 문서 수와 shard 하나의 크기에 대한 평가를 합니다.
- Routing을 활용하세요.
- Replica를 활용하세요.

Elasticsearch document

- 문서 하나의 크기에 대한 평가를 합니다.
- Denormalization 작업이 중요합니다.
- Field에 대한 정의가 중요 합니다.
 - 검색 대상 필드인가요?
 - Exact or full text search
 - 정렬 대상 필드인가요?
 - 강조 대상 필드인가요?
 - 집합 대상 필드인가요?
 - 저장 대상 필드인가요?

Elasticsearch store

- _all
- _source
- store
- fields (multi-field)
- index_options
- term_vector