

ElasticSearch

도서 대출 시스템에 ES 도입으로
검색 품질 높이기

목차

1. 기존 검색 시스템의 문제점
2. Elasticsearch 소개
3. Elasticsearch 도입
4. ES 도입할때 주의할 점
5. 마무리

1. 기존 검색 시스템의 문제점

1. 기존 검색 시스템의 문제점

```
SELECT [필드명] FROM [테이블명]  
WHERE [필드명] LIKE '%특정 문자열%';
```

1. 기존 검색 시스템의 문제점

100만건의 책 데이터 적재

도서 검색

제목 ▾ 자바

검색

[마이페이지로 이동](#)

검색 결과 (10건)

제목	분류	저자	옮긴이	출판일	ISBN	페이지	대출 여부	대출
자바의 정석 (기초편)	프로그래밍 및 소프트웨어	남궁성	null	2016-01-01	97911886112105	1024	대출 가능	<button>대출하기</button>
자바의 정석 (활용편)	프로그래밍 및 소프트웨어	남궁성	null	2017-03-15	97911886122143	880	대출 가능	<button>대출하기</button>
이펙티브 자바	프로그래밍 및 소프트웨어	Joshua Bloch	개앞맵시	2018-10-10	97889166262281	416	대출 가능	<button>대출하기</button>
모던 자바 인 액션	프로그래밍 및 소프트웨어	Raoul-Gabriel Urma	우정은	2019-07-01	921791162242025	592	대출 가능	<button>대출하기</button>
자바 퍼즐러	프로그래밍 및 소프트웨어	Joshua Bloch	박재호	2015-08-21	978897914438216	320	대출 가능	<button>대출하기</button>
혼자 공부하는 자바	프로그래밍 및 소프트웨어	신용권	null	2022-02-14	979116224466211	680	대출 가능	<button>대출하기</button>

```
SELECT b.*, bc.catName FROM dummy b JOIN book_category bc ON b.catCode = bc.catCode WHERE b.title  
timeElapsed = 5309ms
```


1. 기존 검색 시스템의 문제점

900만건의 책 데이터 적재

도서 검색

제목 ▾ 자바

검색

[마이페이지로 이동](#)

검색 결과 (10건)

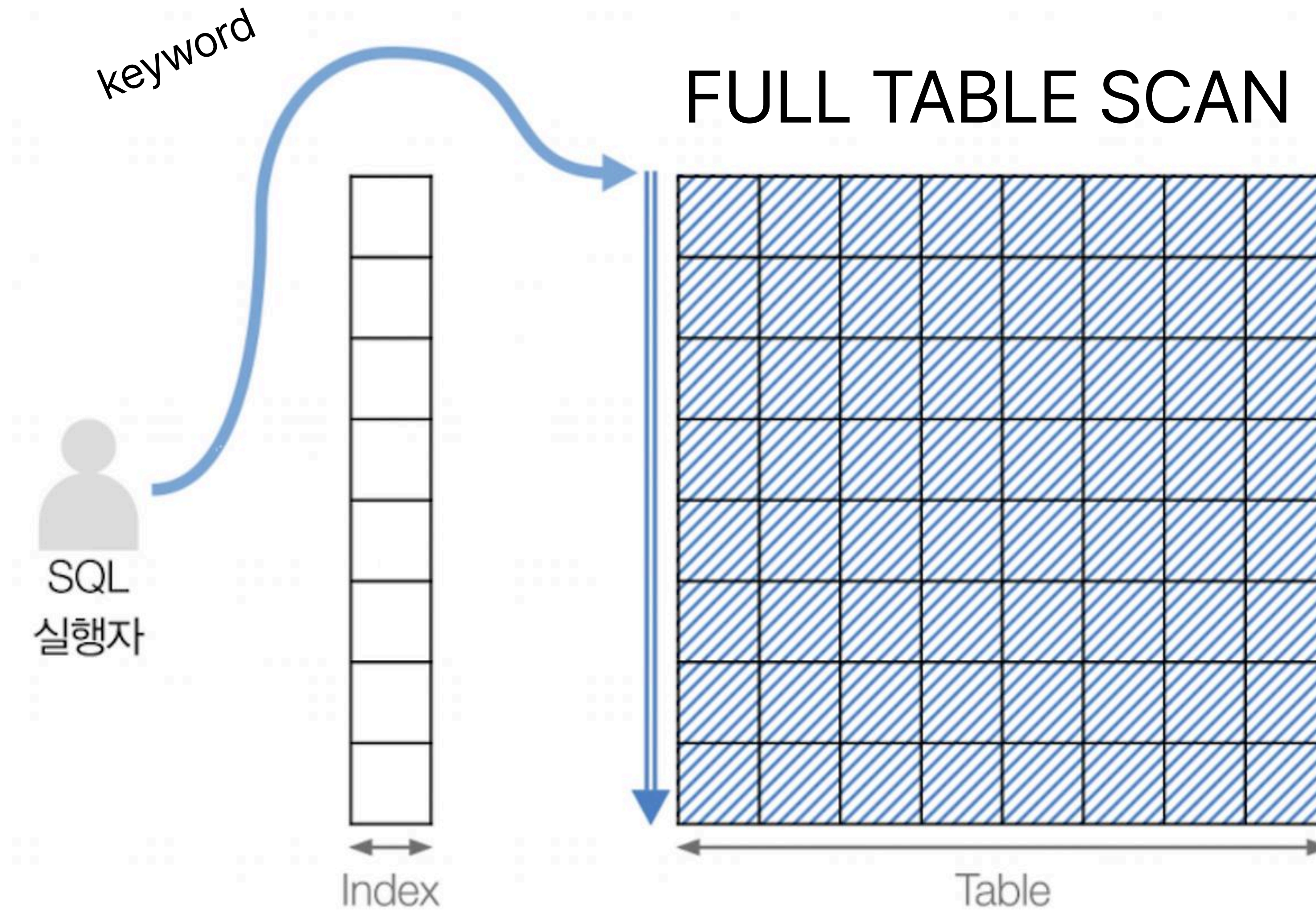
제목	분류	저자	옮긴이	출판일	ISBN	페이지	대출 여부	대출
자바의 정석 (기초편)	프로그래밍 및 소프트웨어	남궁성	null	2016-01-01	97911886112105	1024	대출 가능	<button>대출하기</button>
자바의 정석 (활용편)	프로그래밍 및 소프트웨어	남궁성	null	2017-03-15	97911886122143	880	대출 가능	<button>대출하기</button>
이펙티브 자바	프로그래밍 및 소프트웨어	Joshua Bloch	개앞맵시	2018-10-10	97889166262281	416	대출 가능	<button>대출하기</button>
모던 자바 인 액션	프로그래밍 및 소프트웨어	Raoul-Gabriel Urma	우정은	2019-07-01	921791162242025	592	대출 가능	<button>대출하기</button>
자바 퍼즐러	프로그래밍 및 소프트웨어	Joshua Bloch	박재호	2015-08-21	978897914438216	320	대출 가능	<button>대출하기</button>
혼자 공부하는 자바	프로그래밍 및 소프트웨어	신용권	null	2022-02-14	979116224466211	680	대출 가능	<button>대출하기</button>

```
SELECT b.*, bc.catName FROM dummy b JOIN book_category bc ON b.catCode = bc.catCode WHERE b.title  
timeElapsed = 10078ms
```

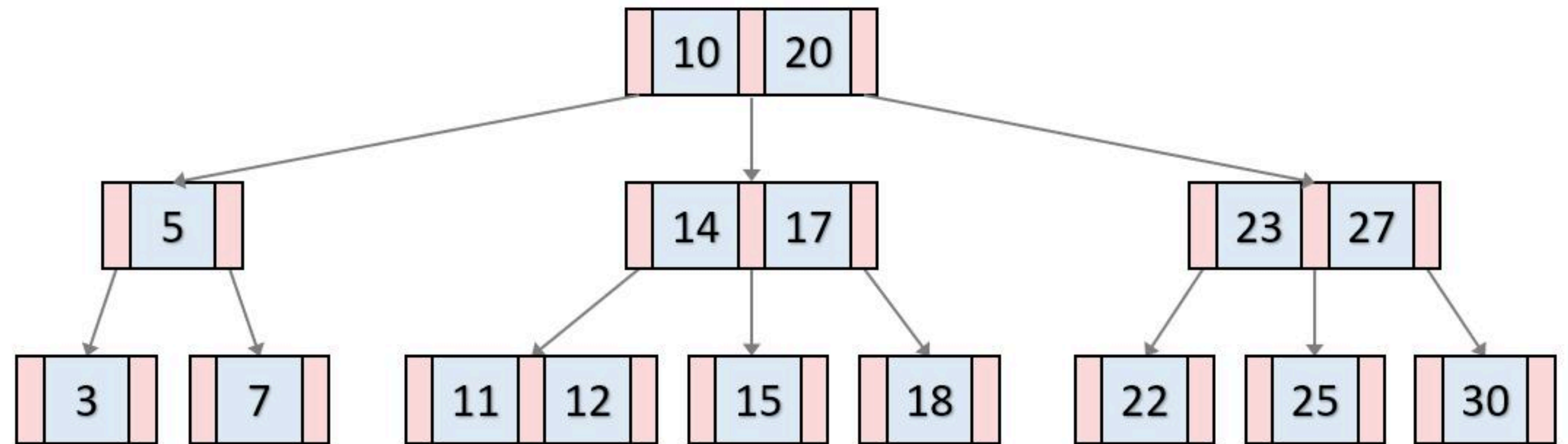
1. 기존 검색 시스템의 문제점

찾아야하는 테이블의
ROW가 1억개라면?

1. 기존 검색 시스템의 문제점



인덱스로 개선하기 V1



LIKE %keyword%

문자열의 시작이 무엇인지 알 수 없음

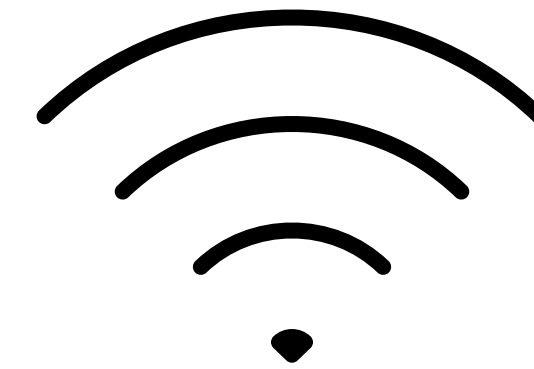
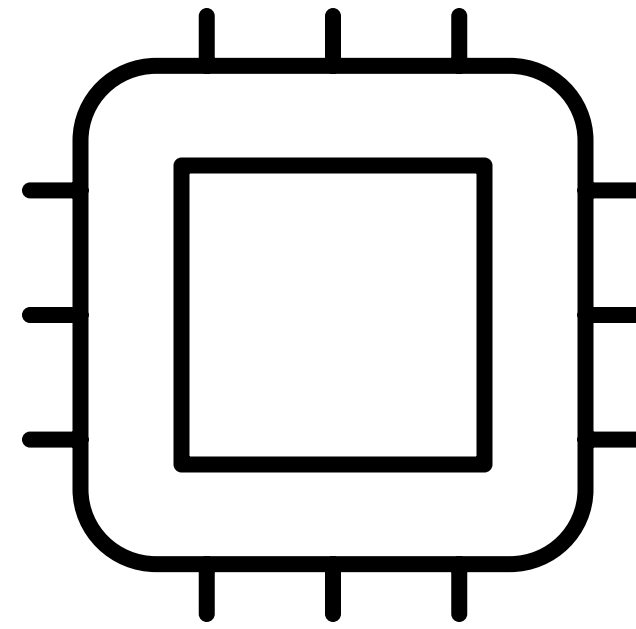
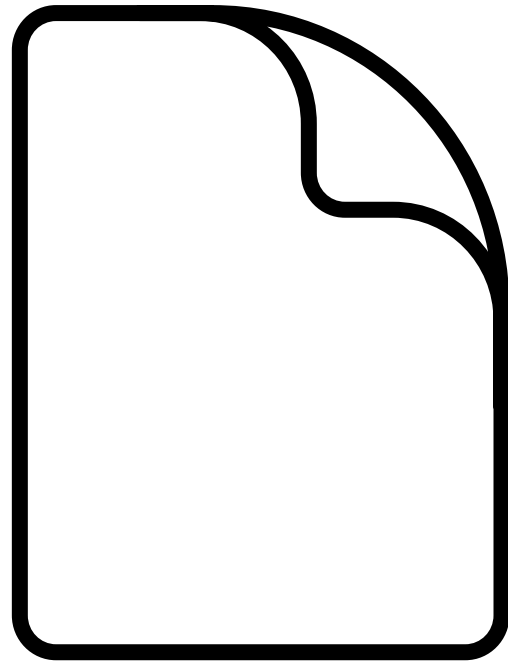
FULL TABLE SCAN

LIKE keyword% 는 가능

모든 자바 인 액션" → "자바"만 검색하면 안 나옴

사용성 ↓

1. 기존 검색 시스템의 문제점



인덱스 활용 불가 → 디스크 I/O 증가 → 응답 지연

⇒ DB에 과도한 부하를 줘서 전체 시스템 성능 저하

1. 기존 검색 시스템의 문제점

인덱스로 개선하기 V2 - Full-Text Index

도서 검색

제목 ▾ 자바

검색

[마이페이지로 이동](#)

검색 결과 (4건) <= 6건 누락

제목	분류	저자	옮긴이	출판일	ISBN	페이지	대출 여부	대출
자바의 정석 (기초편)	프로그래밍 및 소프트웨어	남궁성	null	2016-01-01	97911886112105	1024	대출 가능	<button>대출하기</button>
자바의 정석 (활용편)	프로그래밍 및 소프트웨어	남궁성	null	2017-03-15	97911886122143	880	대출 가능	<button>대출하기</button>
이것이 자바다	프로그래밍 및 소프트웨어	신용권	null	2021-09-15	979116224512215	900	대출 가능	<button>대출하기</button>
모던 자바스크립트 Deep Dive	프로그래밍 및 소프트웨어	이웅모	null	2020-09-25	979115128392239	956	대출 가능	<button>대출하기</button>

```
Executing SQL: SELECT b.*, bc.catName, MATCH(b.title, b.author, b.isbn) AGAINST (? IN BOOLEAN MODE
timeElapsed = 31ms|
```

속도는 데이터가 크지 않다는 선에서 어느정도 개선이 가능

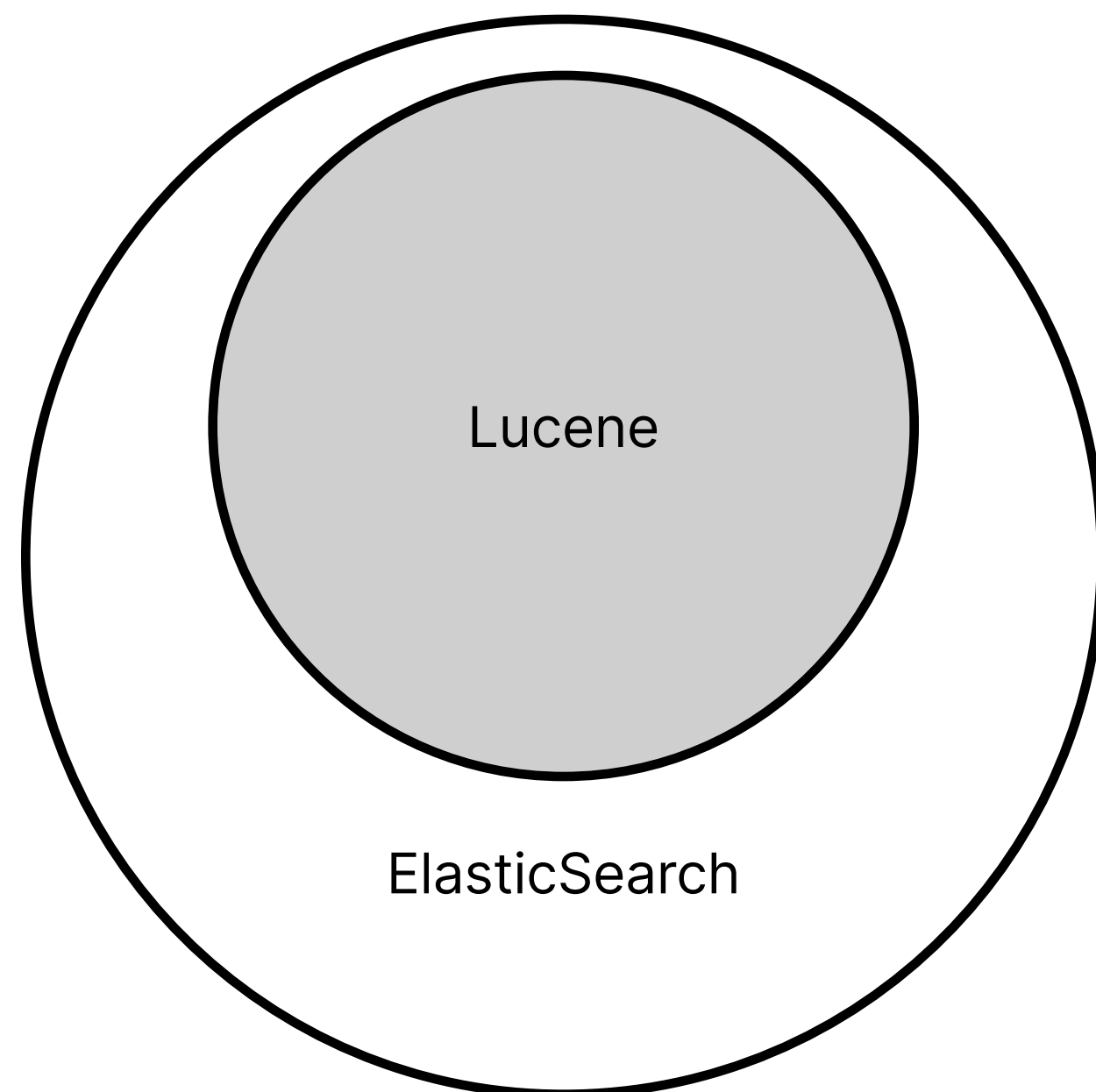
2. Elasticsearch 소개

2. Elasticsearch 소개

ELASTICSEARCH란?

루씬(Lucene)을 기반으로 개발된 오픈
소스 검색엔진

문서를 JSON 형태로 색인하여 저장하고,
이를 대상으로 빠르게 검색과 분석 수행



특징

I. 준 실시간 검색

II. 분산형 아키텍처

III. 다양한 검색 기능

IV. 유연한 데이터 스키마

V. REST API 지원

2. Elasticsearch 소개

I. 준 실시간 검색

역 인덱스(invert index)?
키워드를 통해 문서를 찾아내는 방식을 의미한다.

inverted index

```
{  
  "type": "book",  
  "title": "linux kernel",  
  "author": "alden",  
  "rating": 4.0  
}
```

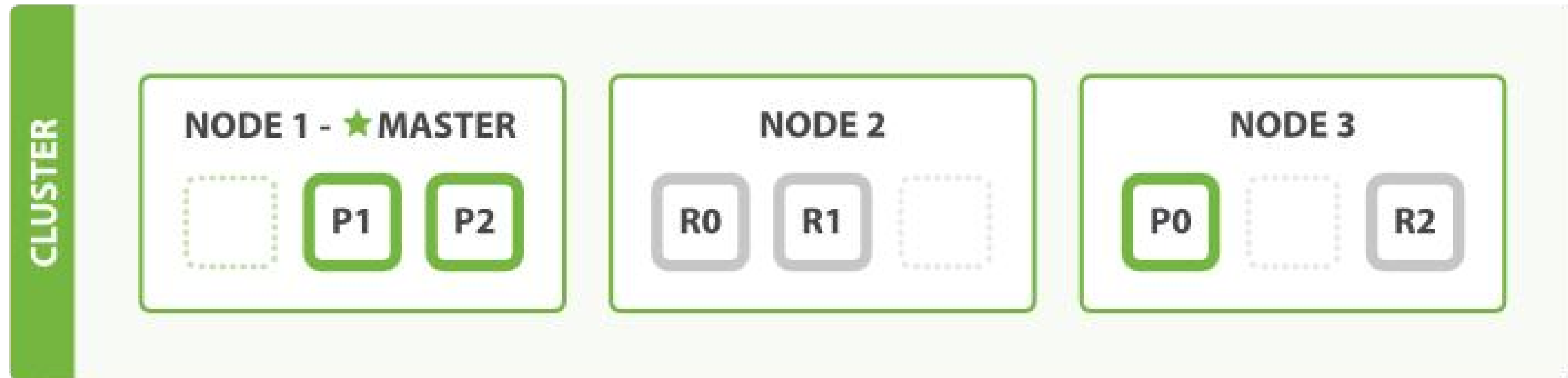
```
{  
  "type": "book",  
  "title": "elasticsearch essential",  
  "author": "alden",  
  "rating": 5.0  
}
```

Tokens	Documents
book	1, 2
linux	1
kernel	1
alden	1, 2
elasticsearch	2
essential	2

II. 분산형 아키텍처

ElasticSearch 클러스터

여러 대의 노드들이 각자의 역할을 바탕으로 연결되어 하나의 시스템처럼 동작하게 되어 있음



⇒ 수평 확장 용이, 처리량을 높여 퍼포먼스를 향상시킬수있음,고가용성

III. REST API 지원

GET,POST,PUT,DELTE 등 지원으로 프로그래밍 언어와 무관하게 사용자가 쉽게 사용할 수있음.

GET books/_count

```
{
  "count": 9000021,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  }
}
```

2. Elasticsearch 소개

다양한 검색 기능

한국어 형태소 분석, 자동완성, 오타 교정 등

형태소란? 일정한 의미가 있는 가장 작은 말의 단위"

```
1 GET _analyze
2 {
3   "tokenizer": "nori_tokenizer",
4   "text": [
5     "혼자 공부하는 운영체제와 컴퓨터 구조"
6   ]
7 }
```

미 사용시

혼자 / 공부하는 / 운영체제와
/ 컴퓨터 / 구조

형태소 분석(nori tokenizer) 적용 시

혼자 / 공부 / 하 / 는 / 운영
/ 체제 / 와 / 컴퓨터 / 구조

⇒ 좀 더 나은 검색 품질을 얻을 수 있음

2. Elasticsearch 소개

유연한 데이터 스키마

동적 매핑

새로운 필드가 들어오면 자동으로 매핑에 추가됨

PUT dynamic-test



```
POST dynamic-test/_doc/1
{
  "title": "혼자 공부하는 운영체제와 컴퓨터구조",
  "price": 25000,
  "isAvailable": true,
  "publishedDate": "2023-09-01"
}
```

→ GET dynamic-test/_mapping

2. Elasticsearch 소개

동적 매핑

GET dynamic-test/_mapping →

필드의 매핑이 자동으로 생성 되었음.
타입은 Elasticsearch가 추론

```
{
  "dynamic-test": {
    "mappings": {
      "properties": {
        "isAvailable": {
          "type": "boolean"
        },
        "price": {
          "type": "long"
        },
        "publishedDate": {
          "type": "date"
        },
        "title": {
          "type": "text",
          "fields": {
            "keyword": {
              "type": "keyword",
              "ignore_above":

```

256

}

3. Elasticsearch 도입

3. Elasticsearch 도입

100만건의 책 데이터 적재

도서 검색

제목 ▼ 자바

검색 [마이페이지로 이동](#)

검색 결과 (10건)

제목	분류	저자	옮긴이	출판일	ISBN	페이지	대출 여부	대출
자바 퍼즐러	null	Joshua Bloch	박재호	2015-08-21	978897914438216	320	대출 가능	<button>대출하기</button>
이것이 자바다	null	신용권	null	2021-09-15	979116224512215	900	대출 가능	<button>대출하기</button>
이펙티브 자바	null	Joshua Bloch	개앞맵시	2018-10-10	97889166262281	416	대출 가능	<button>대출하기</button>
모던 자바 인 액션	null	Raoul-Gabriel Urma	우정은	2019-07-01	921791162242025	592	대출 가능	<button>대출하기</button>
혼자 공부하는 자바	null	신용권	null	2022-02-14	979116224466211	680	대출 가능	<button>대출하기</button>
자바의 정석 (기초편)	null	남궁성	null	2016-01-01	97911886112105	1024	대출 가능	<button>대출하기</button>
자바의 정석 (활용편)	null	남궁성	null	2017-03-15	97911886122143	880	대출 가능	<button>대출하기</button>

```
SELECT b.*, bc.catName FROM dummy b JOIN book_category bc ON b.catCode = bc.catCode WHERE b.title
timeElapsed = 5309ms
```

↓ 개선

```
timeElapsed = 22ms
```


3. Elasticsearch 도입

900만건의 책 데이터 적재

도서 검색

제목 ☒ 자바

검색 [마이페이지로 이동](#)

검색 결과 (10건)

제목	분류	저자	옮긴이	출판일	ISBN	페이지	대출 여부	대출
자바 퍼즐러	null	Joshua Bloch	박재호	2015-08-21	978897914438216	320	대출 가능	<button>대출하기</button>
이것이 자바다	null	신용권	null	2021-09-15	979116224512215	900	대출 가능	<button>대출하기</button>
이펙티브 자바	null	Joshua Bloch	개앞맵시	2018-10-10	97889166262281	416	대출 가능	<button>대출하기</button>
모던 자바 인 액션	null	Raoul-Gabriel Urma	우정은	2019-07-01	921791162242025	592	대출 가능	<button>대출하기</button>
혼자 공부하는 자바	null	신용권	null	2022-02-14	979116224466211	680	대출 가능	<button>대출하기</button>
자바의 정석 (기초편)	null	남궁성	null	2016-01-01	97911886112105	1024	대출 가능	<button>대출하기</button>
자바의 정석 (활용편)	null	남궁성	null	2017-03-15	97911886122143	880	대출 가능	<button>대출하기</button>

```
SELECT b.*, bc.catName FROM dummy b JOIN book_category bc ON b.catCode = bc.catCode WHERE b.title  
timeElapsed = 10078ms
```

↓ 개선

100만건 900만건 차이 X

```
timeElapsed = 23ms ↔ timeElapsed = 22ms
```


3. Elasticsearch 도입

900만건의 책 데이터 적재

도서 검색

제목 ▾ 자바

검색 [마이페이지로 이동](#)

유사도 정렬

검색 결과 (10건)

제목	분류	저자	옮긴이	출판일	ISBN	페이지	대출 여부	대출
자바 퍼즐러	null	Joshua Bloch	박재호	2015-08-21	978897914438216	320	대출 가능	<button>대출하기</button>
이것이 자바다	null	신용권	null	2021-09-15	979116224512215	900	대출 가능	<button>대출하기</button>
이펙티브 자바	null	Joshua Bloch	개앞맵시	2018-10-10	97889166262281	416	대출 가능	<button>대출하기</button>
모던 자바 인 액션	null	Raoul-Gabriel Urma	우정은	2019-07-01	921791162242025	592	대출 가능	<button>대출하기</button>
혼자 공부하는 자바	null	신용권	null	2022-02-14	979116224466211	680	대출 가능	<button>대출하기</button>
자바의 정석 (기초편)	null	남궁성	null	2016-01-01	97911886112105	1024	대출 가능	<button>대출하기</button>
자바의 정석 (활용편)	null	남궁성	null	2017-03-15	97911886122143	880	대출 가능	<button>대출하기</button>

```
SELECT b.*, bc.catName FROM dummy b JOIN book_category bc ON b.catCode = bc.catCode WHERE b.title
timeElapsed = 10078ms
```

```
timeElapsed = 23ms
```

오타가 있어도 검색가능

제목 ▾ 혼자 공부하는

검색 [마이페이지로 이동](#)

검색 결과 (2건)

제목	분류	저자	옮긴이	출판일	ISBN	페이지	대출 여부	대출
혼자 공부하는 자바	null	신용권	null	2022-02-14	979116224466211	680	대출 가능	<div>대출하기</div>

자동 완성

도서 검색

제목 ▾

자 ▾

검색

[마이페이지로 이동](#)

자바 퍼즐러

이것이 자바다

이펙티브 자바

모던 자바 인 액션

혼자 공부하는 자바

자바의 정석 (기초편)

자바의 정석 (활용편)

Do it! 자바 프로그래밍 입문

자바 ORM 표준 JPA 프로그래밍

모던 자바스크립트 Deep Dive

제목 ▾

자바의 ▾

검색

[마이페이지로 이동](#)

자바의 정석 (기초편)

자바의 정석 (활용편)

4. ES 도입할 때 주의 할 점

4. ES 도입할때 주의할 점

완전 실시간은 아니다.

Transaction Rollback을 지원하지 않는다.

데이터의 업데이트를 제공하지 않는다.

기존 매핑을 변경할수 없다.

4. ES 도입할때 주의할 점

동적 매핑(Dynamic Mapping)만 믿으면 생기는 문제

정적 매핑은 문서의 필드들이 가지는 값에 따라 타입을 정의해 둘 필요가 있거나 불필요한 색인이 발생하지 않게 하기 위해 필요함.

적절한 샤드의 갯수 고려(기본은 1개)

샤드의 갯수에 따라 검색 성능, 메모리 사용량, 노드 확장성, 장애 복구 속도 모두 영향

실제 운영 단계라면 좀 더 알아보고 설계를 할것

대용량 데이터에 빠른 검색이 필요할 경우 우선적으로 고려해봐야하는 데이터베이스

결론 : 검색에는 Elasticsearch를 쓰자

감사합니다.

Q&A