**Elastic Search 자료조사**

1. **기본 개념**

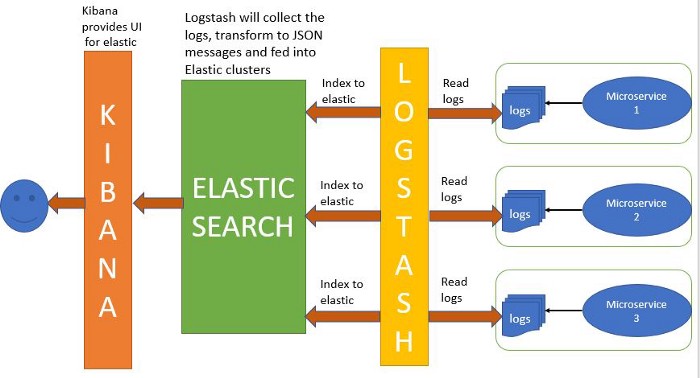
Elasticsearch는 텍스트, 숫자, 위치 기반 정보, 정형 및 비정형 데이터 등 모든 유형의 데이터를 위한 무료 검색 및 분석 엔진으로 분산형 및 개방형을 특징으로 합니다. Elasticsearch는 Apache Lucene을 기반으로 구축되었으며, Elasticsearch N.V.(현재 명칭 Elastic)가 2010년에 최초로 출시했습니다. 간단한 REST API, 분산형 특징, 속도, 확장성으로 유명한 Elasticsearch는 데이터 수집, 보강, 저장, 분석, 시각화를 위한 무료 개방형 도구 모음인 Elastic Stack의 핵심 구성 요소입니다. 보통 **ELK Stack**(Elasticsearch, Logstash, Kibana의 머리글자)이라고 하는 Elastic Stack에는 이제 데이터를 Elasticsearch로 전송하기 위한 경량의 다양한 데이터 수집 에이전트인 Beats가 포함되어 있습니다.

<출처 : Elastic Search 공식홈페이지>

* **ELK Stack(엘라스틱 스택)** : 엘라스틱 서치(데이터검색, 분석) + 키바나(시각화) + 비츠 & 로그스태시(데이터수집 & 정규화)

## 엘라스틱 스택 과정

1. Beats(Raw Data수집) 🡪 Logstach(Raw Data정제 및 저장)
2. ElasticSearch : Logstach에서 받은 데이터를 클러스터(데이터베이스)에 저장한 후 분석
3. Kabana : Elastic Search에서 분석한 자료 시각화



1. **Elasticsearch 활용분야**

애플리케이션 검색/웹사이트 검색/엔터프라이즈 검색/로깅과 로그 분석/인프라 메트릭과 컨테이너 모니터링/애플리케이션 성능 모니터링/위치 기반 정보 데이터 분석 및 시각화/보안 분석/비즈니스 분석

1. **Elastic Search의 기본용어(🡪 RDBMS용어와 비교)**

* 인덱스/타입/문서(document) 🡪 데이터베이스/테이블/로우
* 필드 🡪 컬럼
* 매핑(mapping) 🡪 인덱스/타입/문서의 규칙을 정의한 것
* 클러스터(cluster) 🡪 ES의 가장 큰 시스템 단위, 하나의 클러스터는 여러 개의 노드로 구성되어 있음
* 노드(node) 🡪 각 클러스터의 메타정보관리 및 실제 데이터 저장
* 셔드(shard) 🡪 데이터 검색을 위해 구분되는 최소단위

1. **실행방법**

다운로드 받은 엘라스틱서치폴더 > bin > elasticsearch.bat 실행

🡪 크롬에 localhost:9200 입력 🡪 주피터 노트북 실행 🡪 사용

1. **코드 과정**

총 3개의 인덱스를 생성하였으며,

* 첫번째 인덱스는 인덱스 생성/확인/삭제, 데이터 확인, 개별키 접근, 데이터 검색과 관련된 내용을 다뤘습니다.
* 두번째 인덱스는 중복된 단어(Hack)를 가진 2개의 document를 만들고 match\_phrase, must, must\_not, should쿼리를 작성하는 내용을 다뤘습니다.
* 세번째 인덱스는 중복된 단어(Hack)을 가진 3개의 document를 만들고 정규표현식(regexp)쿼리와 관련된 내용을 다뤘습니다(진행중)