아파치 루씬과 엘라스틱서치

오늘의 학습내용

- 아파치 루씬(Apache Lucene)
- ELK 개요
- ELK 사례

아파치 루씬(Apache Lucene)

아파치 루씬

- 루씬(Lucene)은 자바 언어로 이루어진 정보 검색 라이브러리 오픈 소스 소프트웨어
- 추가적인 개발을 통해 기존의 자바 언어 이외에 오브젝트 파스칼, 펄, C 샤프, C++, 파이썬, 루비, PHP 등 다른 프로그래밍 언어를 사용할 수 있도록 변경
- 전문 검색(Full text) 색인 및 검색 기능을 필요로 하는 모든 응용 프로그램에 적합
- 웹 검색 엔진 및 로컬 단일 사이트 검색 구현에서의 유용성
- 루씬 논리 아키텍쳐의 핵심은 텍스트를 가지고 있는 필드를 포함하는 문서의 개념
- 루씬은 색인 및 검색을 제공하는 라이브러리이며, 웹 크롤러나 HTML 구문 분석 등의 기능은 포함하지 않음.

아파치 루씬(Lucene)

아파치 루씬 기반 프로젝트

- 아파치 너치 - 웹 크롤러 및 HTML 구문 분석 제공
- 아파치 솔라 - 엔터프라이즈 검색 서버
- Compass - 엘라스틱서치의 전신
- CrateDB - 오픈소스, 루씬을 기반으로 하는 분산 SQL 데이터베이스
- DocFetcher - 크로스 플랫폼 데스크톱 환경 검색 애플리케이션
- 엘라스틱서치 - 2010년에 만들어진 엔터프라이즈 서버
- Kinosearch - 약간의 루씬 포팅과 함께 펄과 C로 작성한 검색 엔진. Social text의 위키, 모조모조 위키 엔진, Human Metabolome Database(HMDB)와 Toxin and Toxin-Target Database(T3DB)에서 사용.
- Swiftype - 루씬 기반의 엔터프라이즈 서버 스타트업

ELK 개요

ELK 개요

- ELK는 Elasticsearch, Logstash 및 Kibana 오픈 소스 프로젝트
- Elasticsearch는 검색 및 분석 엔진
- Logstash는 여러 소스에서 동시에 데이터를 수집하여 변환한 후 Elasticsearch 같은 "stash"로 전송하는 서버 사이드 데이터 처리 파이프라인
- Kibana는 사용자가 Elasticsearch에서 차트와 그래프를 이용해 데이터를 시각화

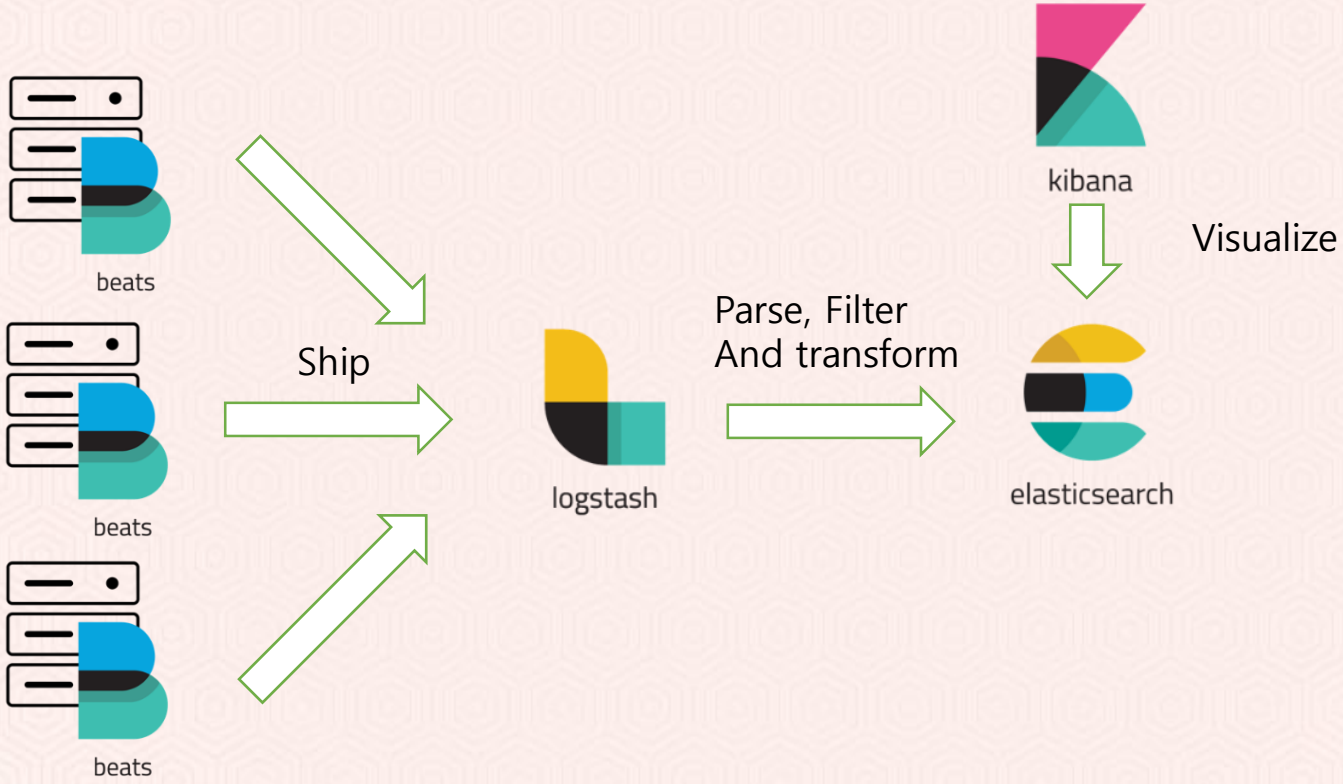
ELK 개요

ELK 장점

- 강력한 유연성과 호환성
- 자유 스키마 : JSON 방식의 Key-Value 형식의 데이터를 사용하므로 형식에 자유로움
- 인덱스 와일드카드 지원
- 확장(Scale-out) 가능 데이터베이스 : 클러스터링
- 사전에 준비된 시각화 도구와 부가기능
- 실시간 데이터 처리 : 메시지 큐(Message Queue, MQ)와 결합하면 강력한 실시간(Realtime) 데이터 수집 및 처리 시스템

ELK 개요

ELK 데이터 처리



ELK 개요

엘라스틱서치(Elasticsearch)

- JSON 기반의 분산형 오픈 소스 RESTful 검색 엔진
- Elasticsearch는 텍스트, 숫자, 위치 기반 정보, 정형 및 비정형 데이터 등 모든 유형의 데이터를 위한 분산형 오픈 소스 검색 및 분석 엔진
- Elasticsearch는 Lucene 기반으로 개발한 분산 검색엔진으로 Logstash를 통해 수신된 데이터를 저장소에 저장하는 역할을 담당
- 데이터를 중심부에 저장하여 예상되는 항목을 검색하고 예상치 못한 항목을 탐지
- 정형, 비정형, 위치정보, 메트릭 등 원하는 방법으로 다양한 유형의 검색을 수행하고 결합 가능

ELK 개요

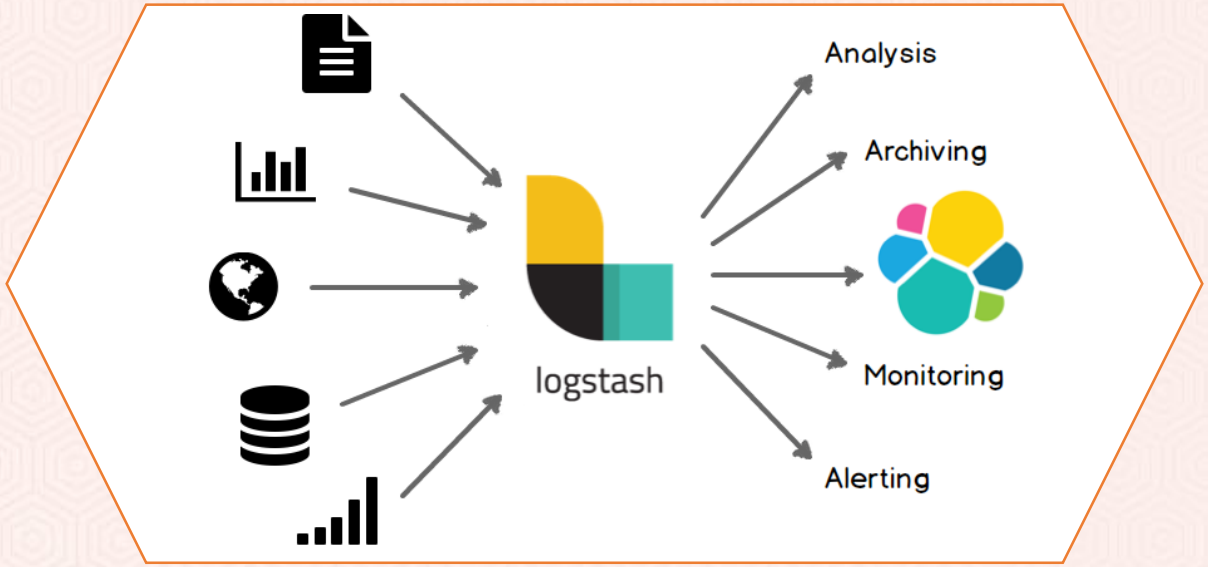
Logstash

- 서버 사이드 오픈 소스 데이터 처리 파이프라인
- 다양한 소스에서 동시에 데이터를 수집하고 변환하여 stash 보관소 저장
- Logstash는 데이터를 집계하고 처리하여 Elasticsearch로 전송하는 데 사용
- 다양한 고급 다운스트림 분석 및 시각화 활용 사례를 위해 모든 데이터를 정리하고 대중화(democratization)
- 사용자는 다양한 소스에서 동시에 데이터를 수집하고 강화하고 변환한 다음 Elasticsearch에서 색인 가능하도록 전송하는 역할을 담당하는 소프트웨어
- 어떤 유형의 이벤트도 다양한 입력, 필터, 출력 플러그인을 통해 강화하고 전환할 수 있으며 기본 제공되는 여러 코덱으로 수집(ingestion) 프로세스를 한층 더 간소화

ELK 개요

Logstash

- ❖ 수평 확장이 가능한 데이터 처리 파이프라인
- ❖ 플러그형 파이프라인 아키텍처 : 다양한 입력, 필터, 출력을 조정하여 운용
- ❖ 확장 가능하고 개발자에게 편리한 플러그인 에코시스템 : 200여개 플러그인



ELK 개요

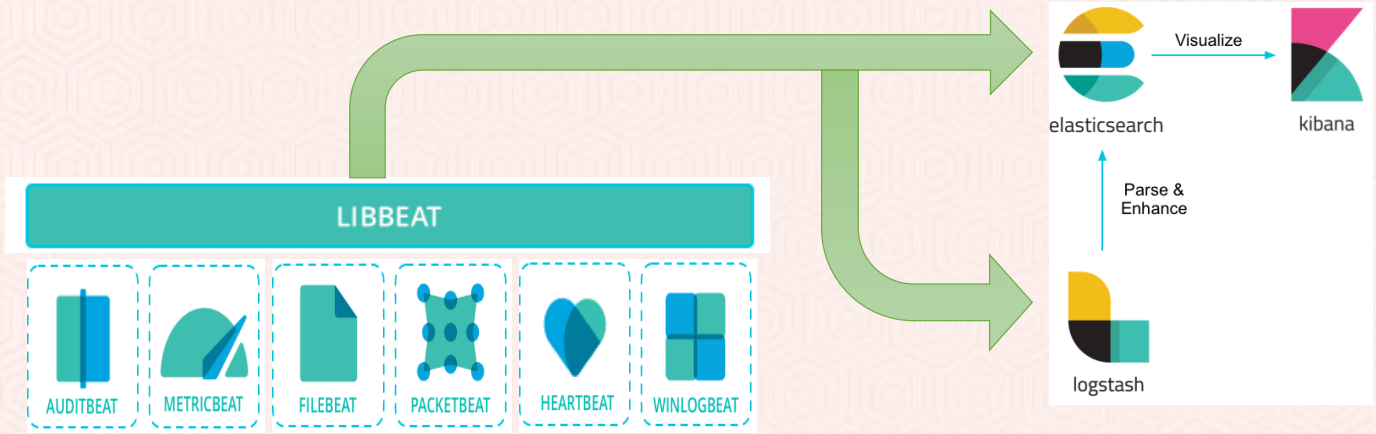
Kibana

- Elasticsearch와 함께 사용하도록 설계된 오픈소스 분석 및 시각화 플랫폼
- Kibana를 사용하여 Elasticsearch 색인에 저장된 데이터를 검색하고 보고 상호 작용을 수행 가능
- 손쉽게 고급 데이터 분석을 수행하고 다양한 차트, 테이블, 지도의 형태로 데이터를 시각화 가능
- 간단한 브라우저 기반 인터페이스에서 Elasticsearch 쿼리의 변경 사항을 실시간으로 표시하는 동적 대시보드를 신속하게 생성하고 공유 가능

ELK 개요

Beats

- ❖ 서버에 에이전트 형식으로 설치하는 오픈소스 경량 데이터 수집기
- ❖ 단일 목적의 데이터 수집기 플랫폼인 Beats는 수백 수천 개의 장비와 시스템으로부터 Logstash나 Elasticsearch에 데이터를 전송



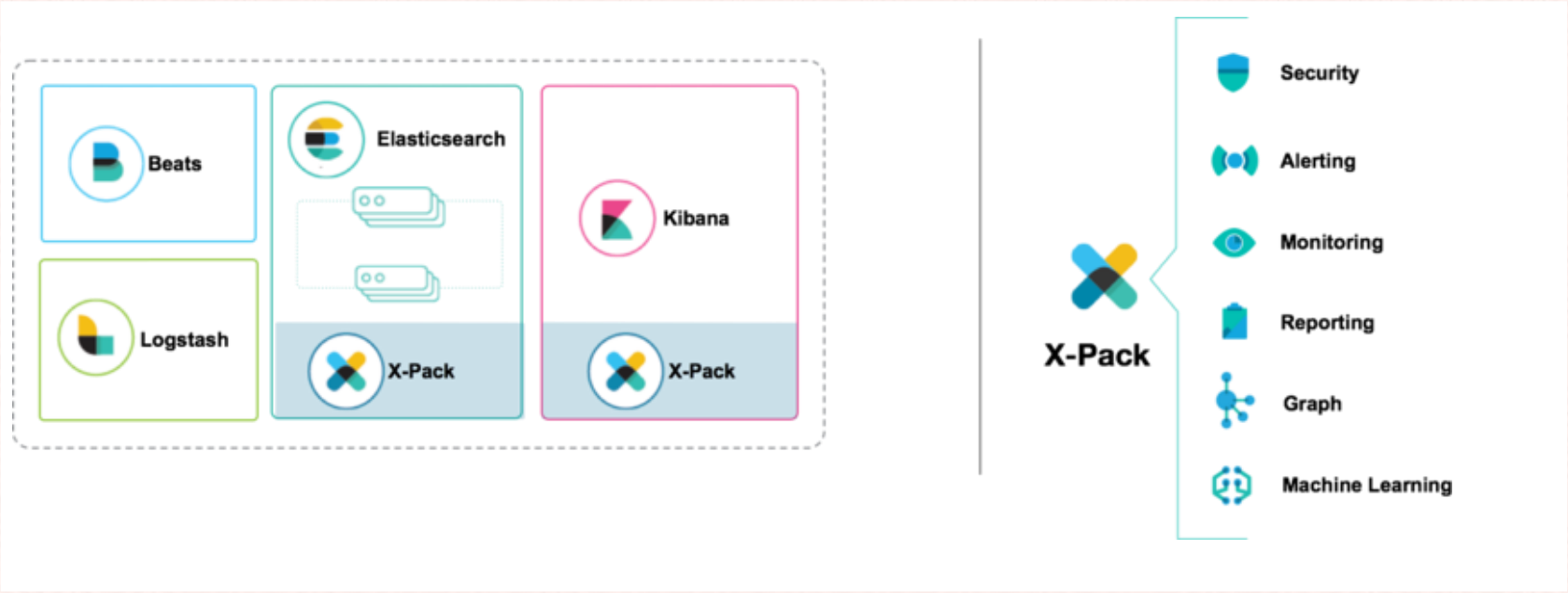
ELK 개요

X-Pack

- X-Pack은 보안, 알림, 모니터링, 보고, 그래프 기능을 설치하기 편리한 단일 패키지로 번들 구성한 Elastic Stack 확장 프로그램
- X-Pack 구성 요소는 서로 원활하게 연동할 수 있도록 설계되었지만 사용할 기능을 손쉽게 활성화하거나 비활성화 가능
- X-Pack 머신러닝 기능은 비지도(unsupervised) 학습 방식을 통해 "시계열 데이터로부터 이상징후 탐지" 기능을 제공하는 데 중점
- Elasticsearch에서 로그 파일, 응용 프로그램 및 성능 수치, 네트워크 통신 그리고 금융 거래 데이터와 같은 시계열 데이터를 다루는 사용자에게 부가 가치를 제공하는 데 초점

ELK 개요

X-Pack


Beats
Logstash
Elasticsearch
X-Pack
Kibana
X-Pack
X-Pack
Security
Alerting
Monitoring
Reporting
Graph
Machine Learning

ELK 사례

ELK 사용 사례

- 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