Cloud Innovator MEGAZONE

elastic

MEGAZONE ELASTIC DAY

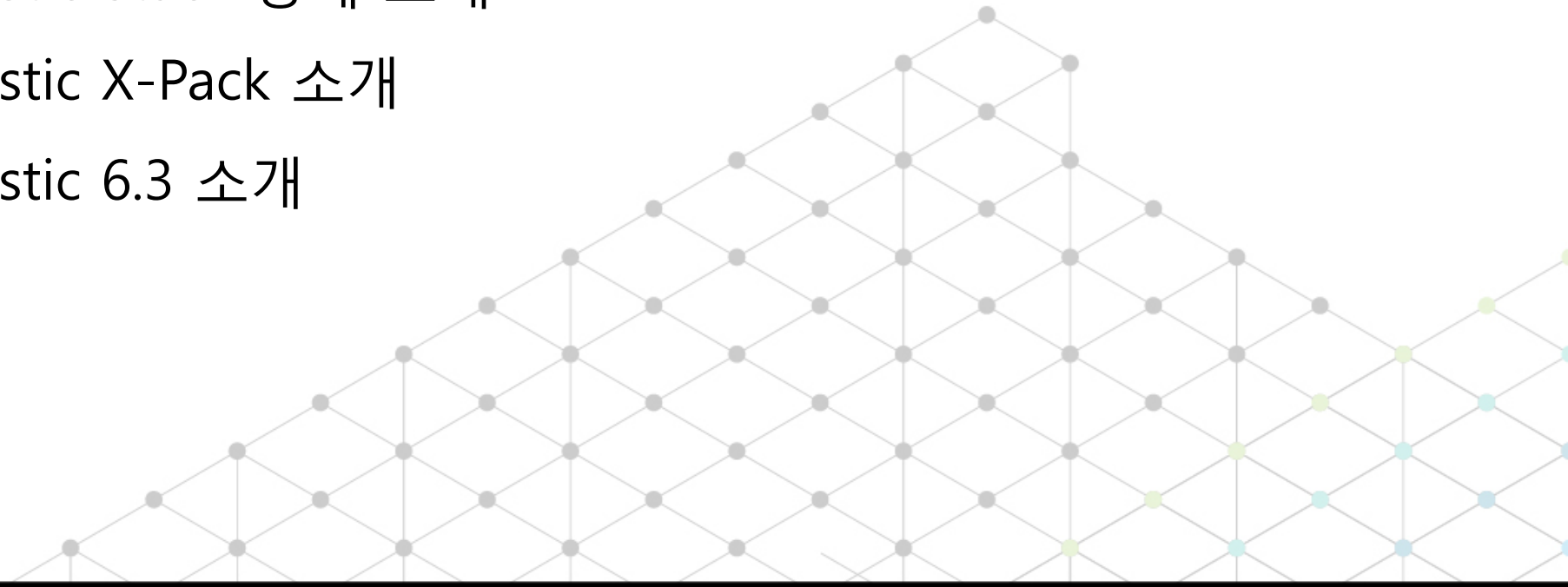
2018년 6월 27일(수)

Elastic Stack과 X-Pack의 기본구성과 활용사례

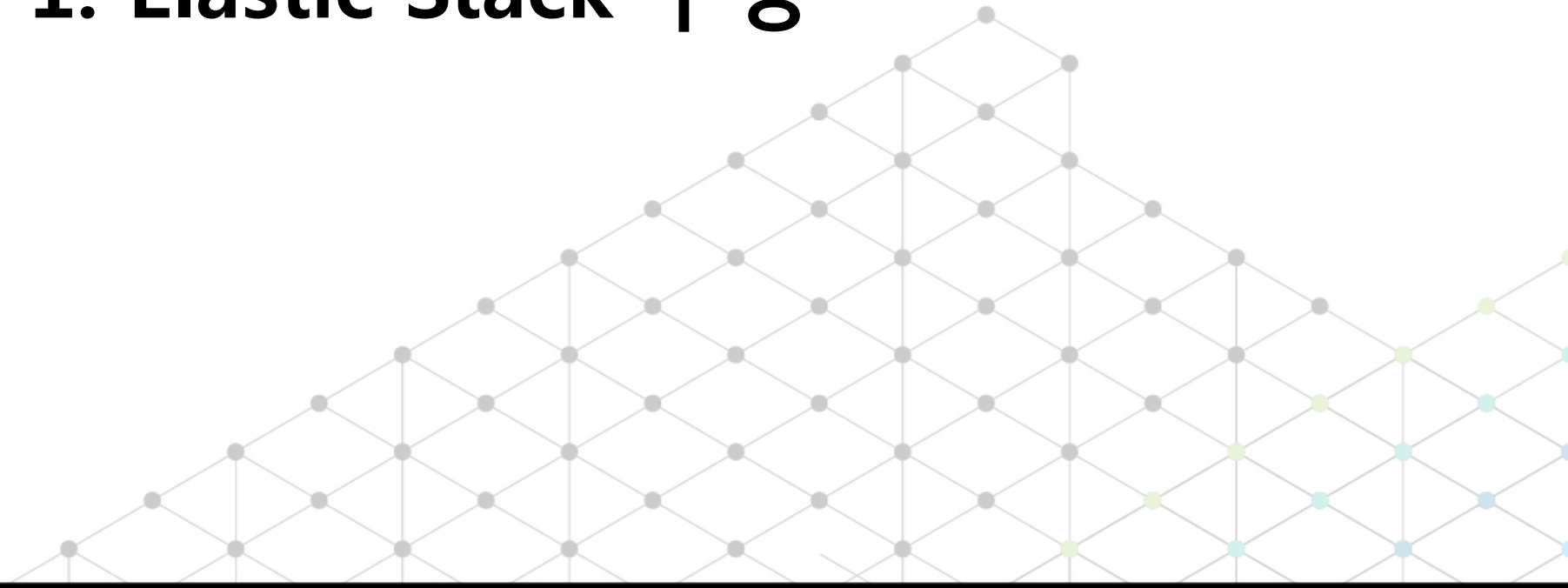
메가존 최재은 매니저

Contents

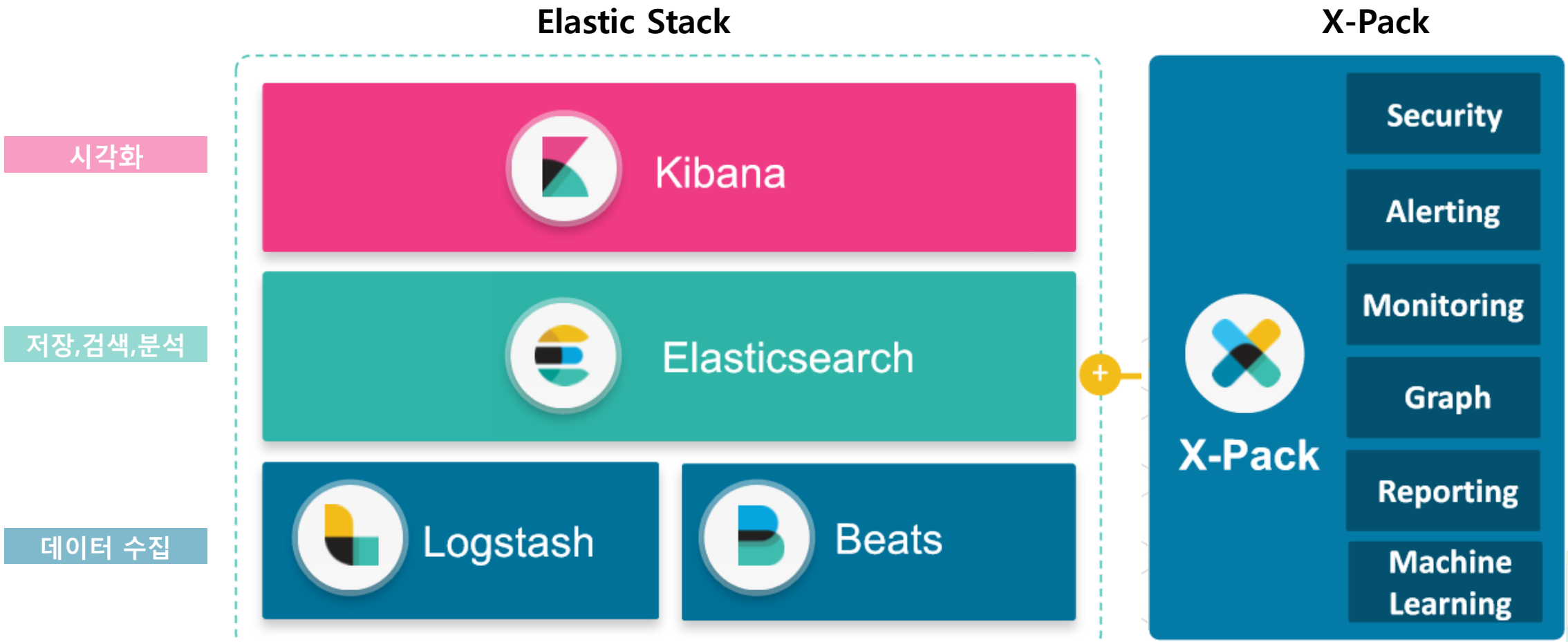
1. Elastic Stack 구성
2. Elastic Stack 상세 소개
3. Elastic X-Pack 소개
4. Elastic 6.3 소개



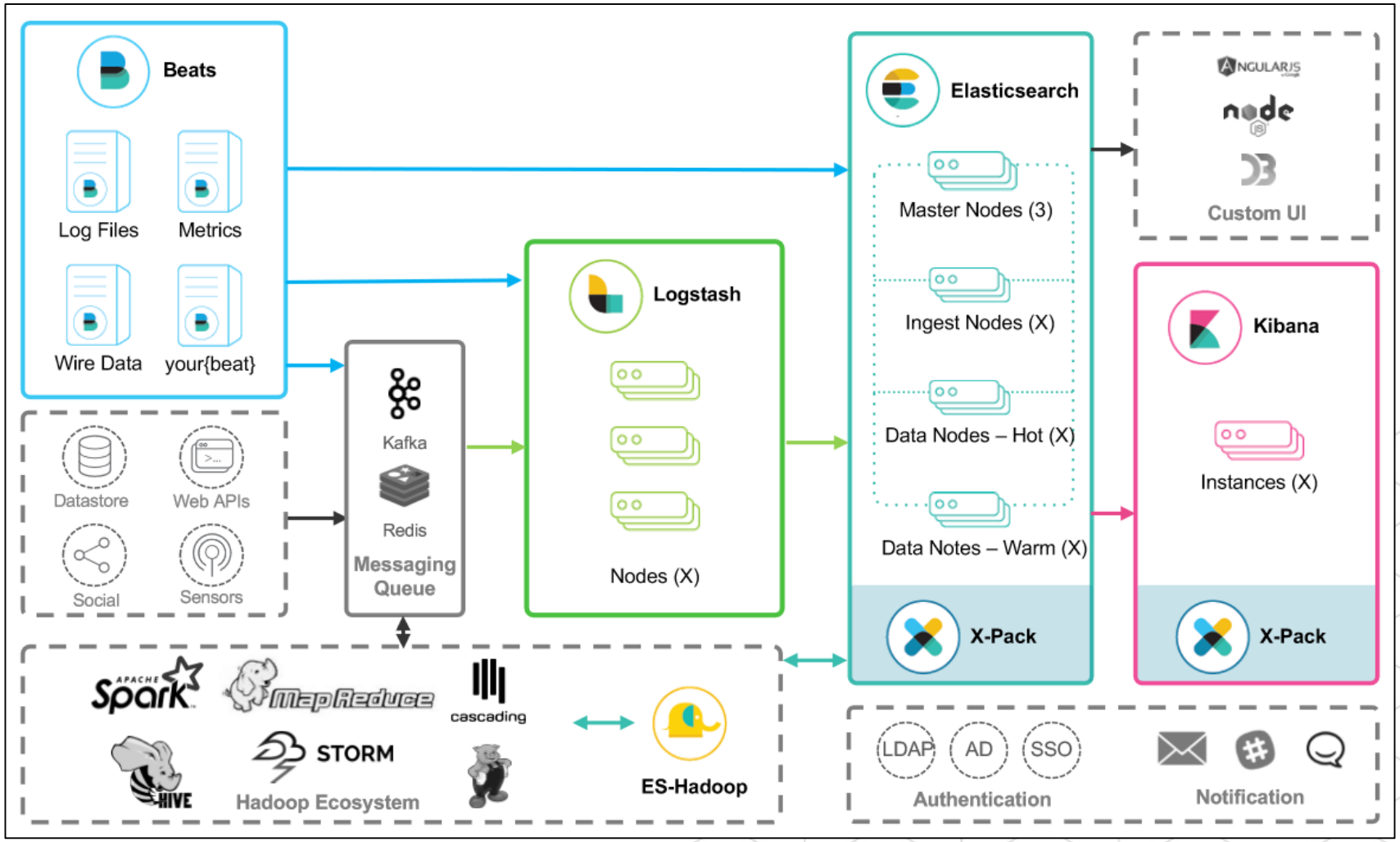
1. Elastic Stack 구성



Elastic Product Overview



Elastic 확장 구성



수집 영역 확장

- Beats를 활용한 다양한 데이터 수집
- 비정형 데이터 구조화(Logstash)
- Messaging Queue 구성

데이터 검색 및 분석 영역 세분화

- 노드의 세분화
- 성능과 확장성을 위해 elasticsearch의 노드를 용도에 맞게 분리 구성

빅데이터 클러스터 연계 확장

- ES-Hadoop을 통한 Hadoop 클러스터와의 연계

Elastic 권장 하드웨어 사양

기본 하드웨어 사양

- RAM : 64 GB (32 GB or 16 GB 가능)
- CPU : 2 ~ 8core
- SSD : 200GB
- RAID 0 (구성 디스크 속도 향상)

메모리 : 하드디스크 구성 비율

- 검색엔진 비율 <1:16>
- 로깅 시스템 <1:48, 1:96 등>

- CPU : 4 ~ 16core
- RAM : 8GB ~ 32GB
- HDD : Any

<master>
- CPU : 4core
- RAM : 8GB
- HDD : Any

<data node>
- CPU : 8core ~16core
- RAM : 16GB ~ 32GB
- SSD

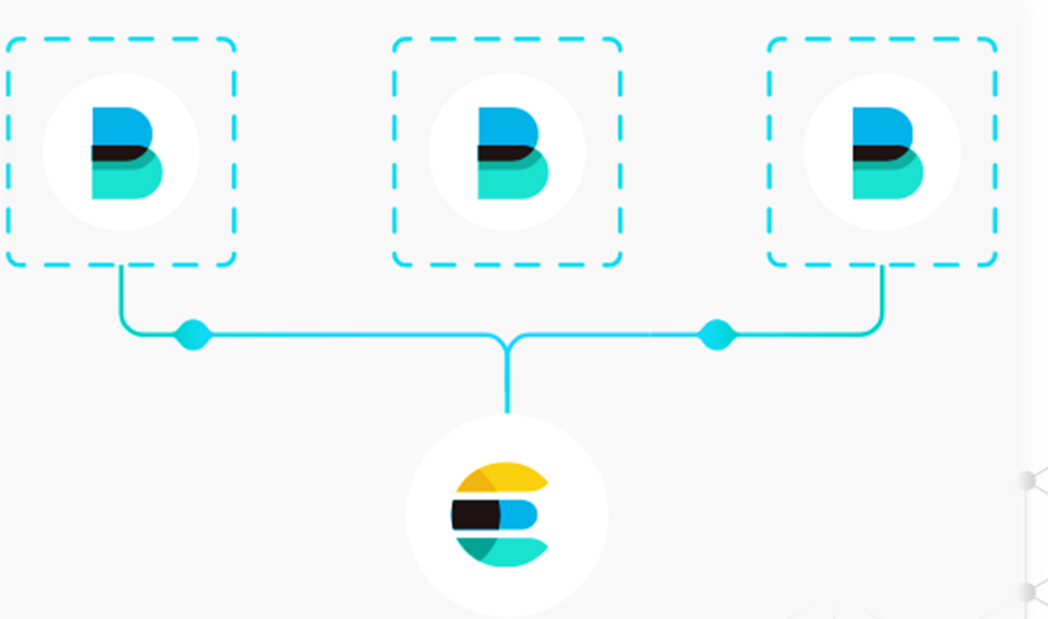
- CPU : 2core
- RAM : 4GB ~ 8GB
- HDD : Any

2. Elastic Stack 상세 소개

Beats

- 다양한 유형의 운영 데이터를 Elasticsearch에 전송하는 오픈 소스 데이터 shipper
- 데이터를 직접 Elasticsearch로 보내거나 Logstash를 통해 Elasticsearch로 전송

From Beats to Elasticsearch



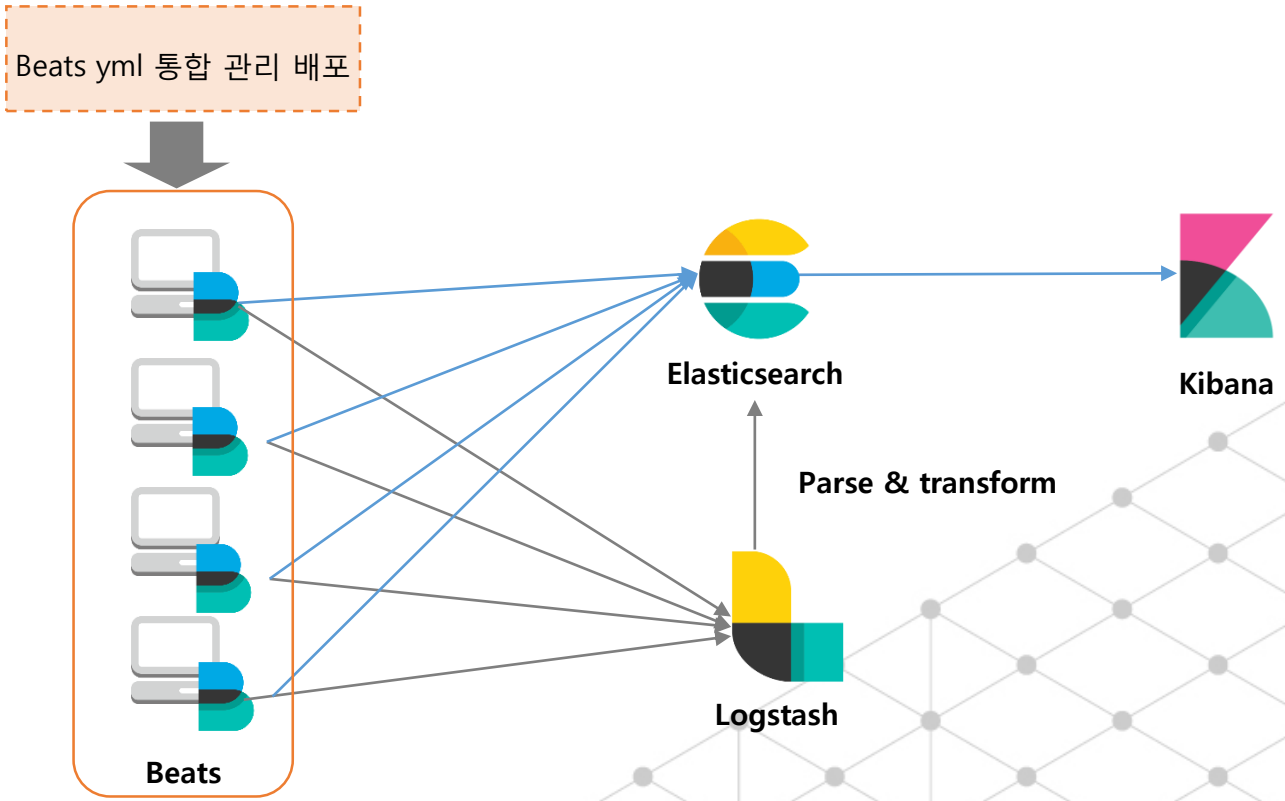
기능 및 강점

- Community Beats를 통해 필요한 Beat 사용 가능
- **로그와 데이터를 경량화 된 방식으로 전달**
- 다양한 시스템 서비스 통계 전송
- **서비스와 어플리케이션의 실시간 모니터링**
- 네트워크 트래픽 검색 및 분석
- **Windows 기반 인프라의 상태를 확인**
- Linux 시스템의 지속적인 감시
- 호스트들의 상태 모니터링

Beats Product Type

					
Filebeat	**Metricbeat**	**Packetbeat**	**Winlogbeat**	**Auditbeat**	**Heartbeat**
• 로그를 라인 별로 읽고 전달, 중단되는 경우 중단점 기억 재가동 • 일반적인 형식의 로그 데이터 간편한 수집, 파싱, 시각화 처리 (Apache, NGINX, System, MySQL)	• 다양한 시스템 통계 수집, 전송 (Linux, Widows, Mac 호스트) • CPU사용률, 메모리, 파일시스템, 디스크 IO 및 네트워크 IO 통계 제공 • 다양한 서비스의 메트릭을 수집하는 내장 모듈 제공	• 데이터에 실시간 접근, 내용 분석 • 네트웍 프로토콜로 접근하여 애플리케이션 지연시간 및 오류, 응답시간 SLA 성능, 사용자 엑세스 패턴 및 추이 확인	• Windows 기반 인프라의 상태 확인 • 보안로그, 장치연결, 신규 소프트웨어 설치 등 이벤트 로그를 정형화된 형식으로 수집	• Auditbeat는 Linux 감사 프레임워크와 직접 통신하여Audit 와 동일한 데이터 수집, 전송 • 파일 통합 모니터링 (메타데이터, 파일 내용의 암호화 해쉬 등 포함)	• 서비스 동작 모니터링 • 서비스의 가동 및 반응시간 데이터 생성 • 로드 밸런싱이 적용된 서버 호스트를 DNS분석 기법을 통해 모니터링 • 모니터링 대상의 추가 및 제거 프로세스 자동화

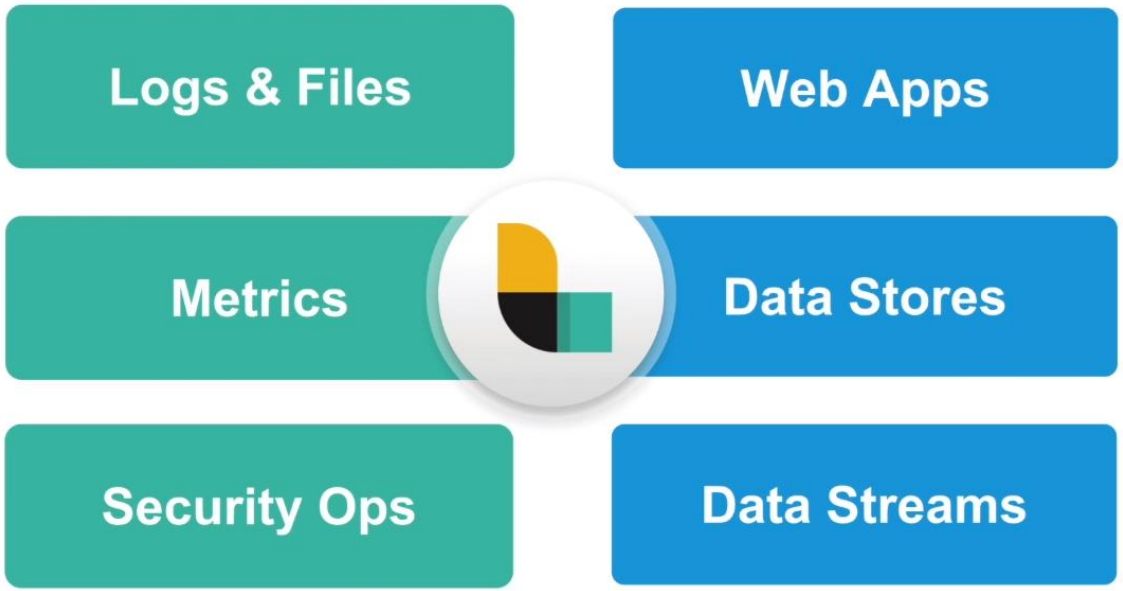
Beats를 이용한 Architecture



Logstash

- 실시간 파이프라인 기능을 가진 오픈소스 데이터 수집 엔진
- 서로 다른 소스의 데이터를 탄력적으로 통합, 사용자가 선택한 목적지로 데이터 정규화

Logstash IN/OUT



기능 및 강점

- 강력한 수집 기능– 수평확장 데이터 처리 파이프라인
- 플러그형 파이프라인 아키텍처
- **200여 개가 넘는 플러그인**
- 직접 플러그인을 만들어 제공할 수 있는 유연성
- **모든 유형의 로깅 데이터 처리**
- 메시징 큐가 제공하는 각종 데이터 스트림 통합
- 다양한 산업용 Application이 제공하는 데이터 수집
- **필터 기능으로 비정형 데이터 구조 도출**

Logstash In/Out

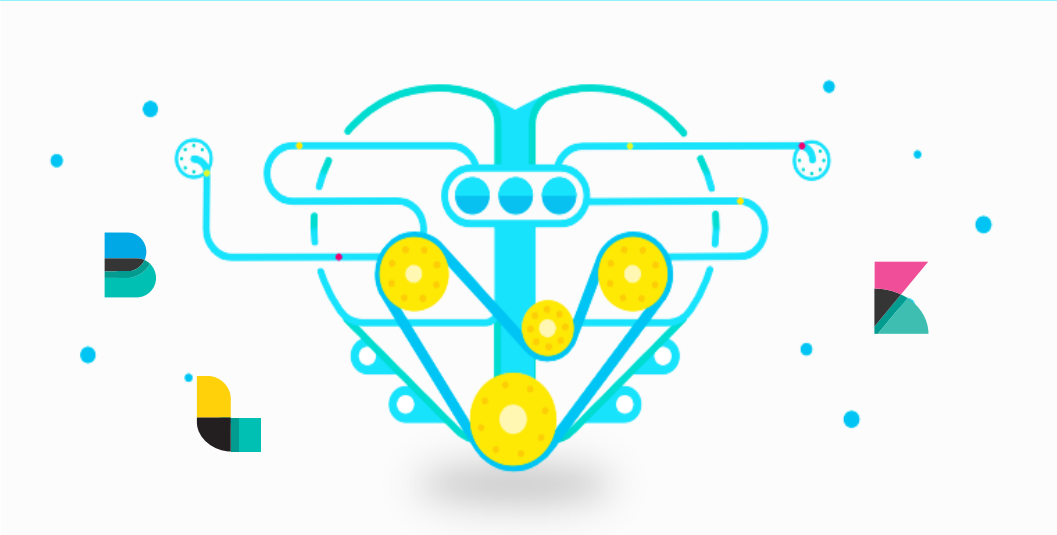
IN	
beats	Elastic Beats 프레임워크의 이벤트 입력
Cloudwatch	아마존 웹서비스 Cloudwatch API의 모든 이벤트 입력
File	Files의 스트림 이벤트 입력
graphite	Graphite 툴의 메트릭 데이터 입력
generator	테스트 목적의 랜덤 로그 이벤트 입력
github	Github webhook의 이벤트 입력
Http	Http 또는 Https상의 이벤트 입력
imap	IMAP 서버의 메일 입력
jdbc	JDBC 데이터의 이벤트 생성, 입력
Kafka	Kafka 토픽의 이벤트 입력
Kinesis	AWS 키네시스 스트림의 이벤트 입력
Redis	Redis 인스턴스의 이벤트 입력
S3	S3 버켓 Files의 스트림 이벤트 입력
syslog	이벤트들의 syslog 메시지 입력
그 외 다수	참고 사이트 : https://www.elastic.co/guide/en/logstash/current/input-plugins.html

OUT	
stdout	Standard output으로 events 프린트
cloudwatch	AWS CloudWatch에 요약된 메트릭 데이터 전송
CSV	일정한 포맷으로 디스크에 이벤트 쓰기
datadog	로그스태시 이벤트 기반으로 DataDogHQ에 이벤트 전송
elasticsearch	Elasticsearch에 로그 저장
email	Output 확인 시 특정 어드레스로 이메일 전송
file	디스크 파일에 이벤트 쓰기
Google_bigquery	구글 BigQuery에 이벤트 쓰기
graphite	Graphite에 메트릭 데이터 쓰기
http	Http 또는 Https 엔드포인트로 이벤트 전송
Kafka	카프카 topic에 이벤트 쓰기
MongoDB	MongoDB에 이벤트 쓰기
Redis	RPUSH 커맨드로 Redis큐에 이벤트 전송
S3	아마존 S3로 로그스태시 이벤트 전송
그 외 다수	참고 사이트 : https://www.elastic.co/guide/en/logstash/current/output-plugins.html

Elasticsearch

- 확장성이 뛰어난 오픈소스 풀텍스트 검색 및 분석 엔진
- 방대한 양의 데이터를 신속하게, 준 실시간으로 저장, 검색, 분석 지원

The Heart of the Elastic Stack

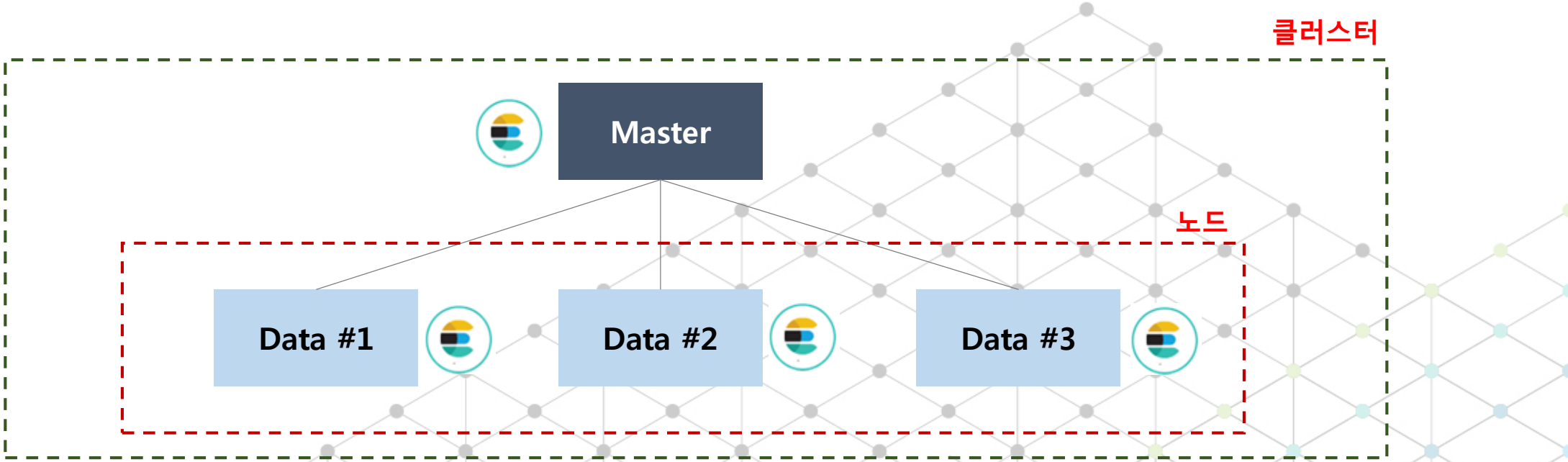


기능 및 강점

- **방대한 데이터 조사, 분석, 시각화, 임시 질의 수행**
- Aggregation 기능 지원
- 복잡한 비즈니스 인텔리전스 쿼리를 수행
- 데이터 분석, 추이, 통계, 요약 정보 확인
- 문서를 색인화하는 시점부터 검색 가능 대기 시간(약 1초)
- 콘텐츠 볼륨의 수평 분할/확장
- **샤드 분산 배치로 성능/처리량 증가**
- **리블리카 생성으로 고가용성 제공**

Elasticsearch 구성

클러스터	**클러스터는 하나 이상의 노드(서버)가 모인 것이며, 이를 통해 전체 데이터를 저장**, 모든 노드를 포괄하는 통합 색인화 및 검색 기능 제공
노드	노드는 클러스터에 포함된 단일 서버로서 **데이터를 저장하고 클러스터의 색인화 및 검색 기능에 할당** 하나의 클러스터에서 원하는 개수의 노드로 구성

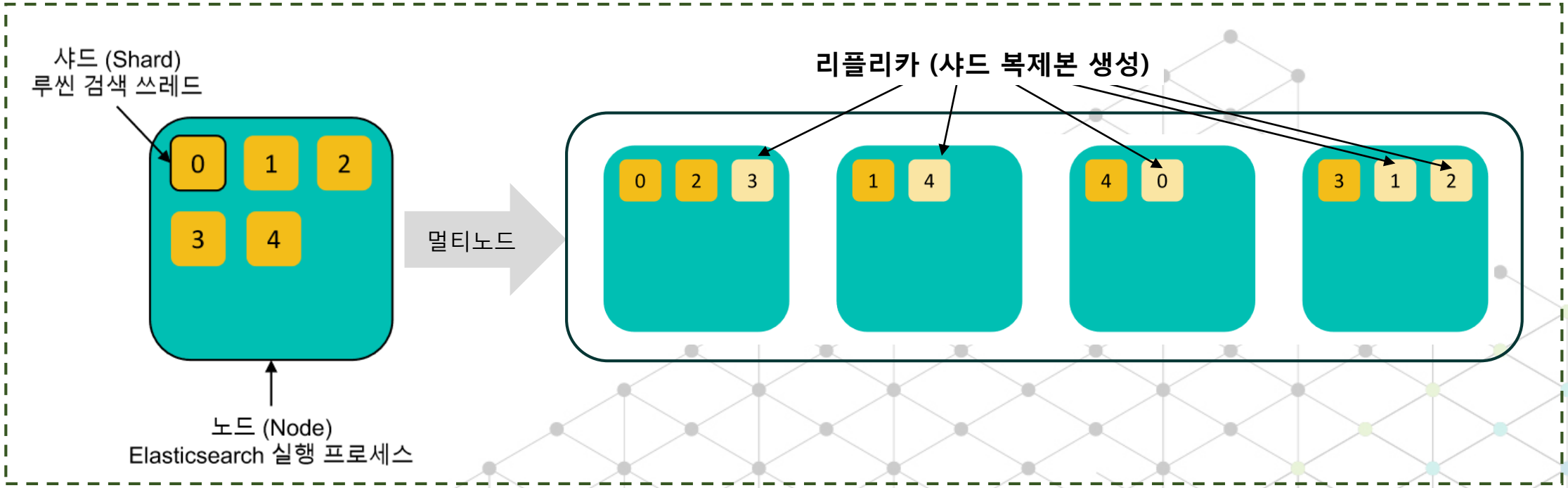


Elasticsearch 구성

Index(색인)	**색인은 다소 비슷한 특성을 가진 문서의 모음.** 예) 고객 데이터에 대한 색인, 제품 카탈로그에 대한 색인, 주문 데이터에 대한 색인.
Type	**하나의 색인에서 하나의 Type을 정의.** Type이란 색인을 논리적으로 분류/구분한 것이며 그 의미 체계는 전적으로 사용자가 결정 https://www.elastic.co/guide/en/elasticsearch/reference/master/removal-of-types.html
Document	문서를 **색인화할 수 있는 기본 정보 단위**. JSON(JavaScript Object Notation) 형식. DB의 Row와 비교. 예) 단일 고객, 단일 제품, 단일 주문에 대한 문서가 각각 존재
Field	Document를 구성하는 가장 최소 단위, Elasticsearch는 필드에 여러 종류의 데이터 타입 지원 예) String, Numeric, Date, Boolean, Binary, Range, Array, Object, Geo-point, IP, Token 외 다수 지원

Elasticsearch 기본 구조

샤드	방대한 양의 데이터를 저장할 때 **색인을 이른바 샤드(shard)라는 조각으로 분할** 단일 노드의 디스크에서 수용하지 못하거나 검색 요청 처리 시 속도가 느려지는 것을 방지
리플리카	오류가 일어날 가능성이 있는 네트워크/클라우드 환경에서 샤드에 대해 **하나 이상의 복사본 생성** 하는 개념 **기본적으로 Elasticsearch의 각 색인은 샤드 5개 , 리플리카 1개를 갖습니다.**



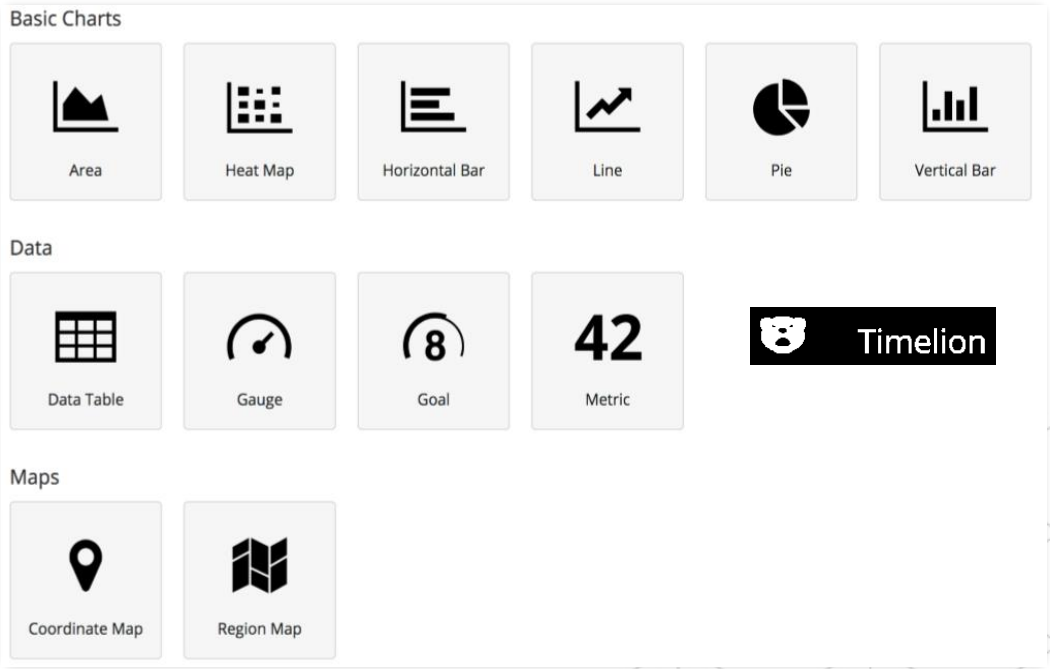
Elasticsearch 노드 타입

Node Type	Description
Master nodes	클러스터 관리 노드, 클러스터 상태 정보 관리
Data nodes	데이터 입/출력, 검색 수행
Ingest nodes	데이터 가공 기능 제공
Coordinating Nodes	검색 및 대량 색인 요청 시 분산 저장된 데이터를 단일 결과 집합화
Alerting Nodes	X-Pack Alerting 실행 노드
Machine Learning Nodes	X-Pack machine learning job을 수행하기 위한 전용 노드

Kibana

- Kibana는 Elasticsearch와 함께 사용하도록 설계된 오픈소스 분석 및 시각화 플랫폼
- Elasticsearch 쿼리의 변경 사항을 실시간으로 표시하는 동적 대시보드 생성, 공유

Kibana 지원 분석도구

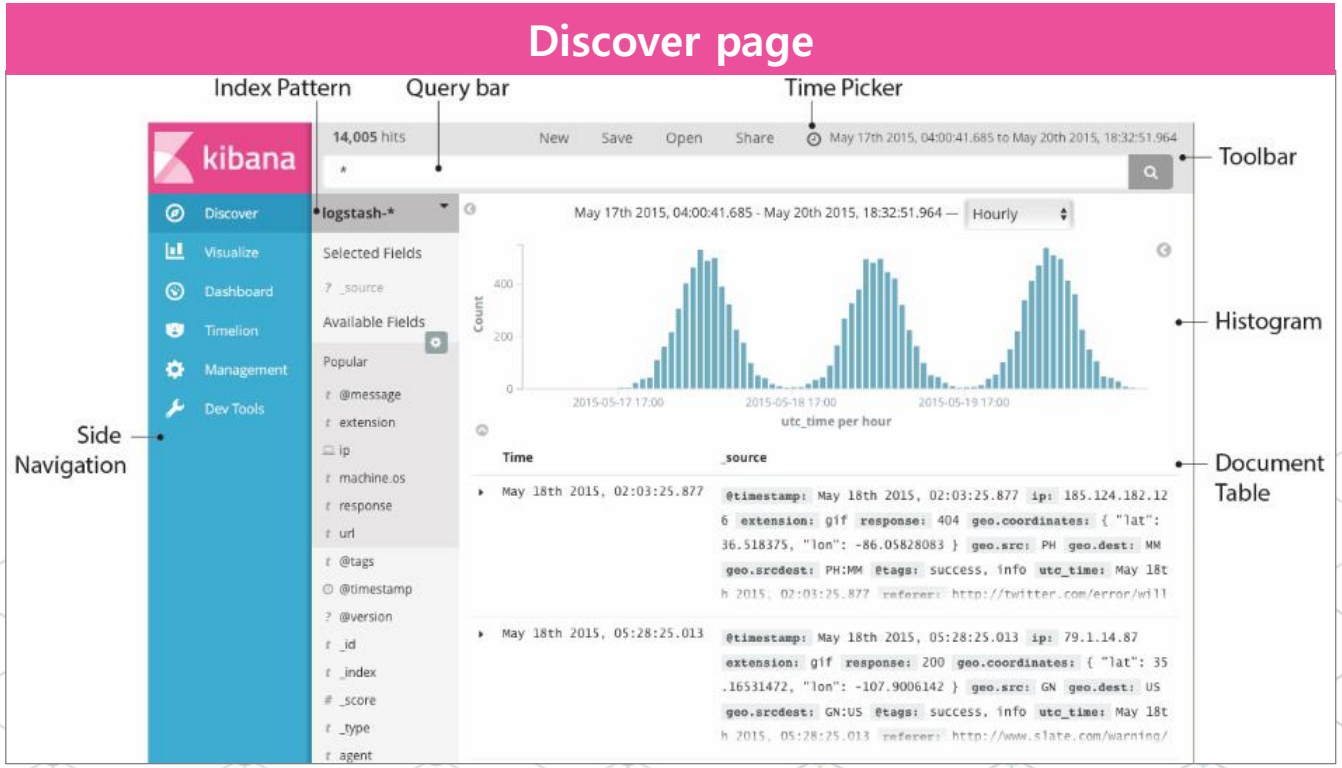


기능 및 강점

- **다양한 형태의 데이터 분석도구 제공**
- 수집된 데이터의 관계성 확인
- 시계열 분석을 통해 다양한 그래프 생성
- Kibana에서 제공하는 시각화 도구
  - Basic Charts
  - Data (Metic 등)
  - Maps (위치정보)
- Timelion

Kibana(Discover)

- 대화식으로 데이터 탐색 지원
- 선택한 색인 패턴과 일치하는 모든 색인 문서에 액세스 및 확인 가능
- 문서의 시간별 분포 히스토그램으로 표기
- 검색 쿼리 제공



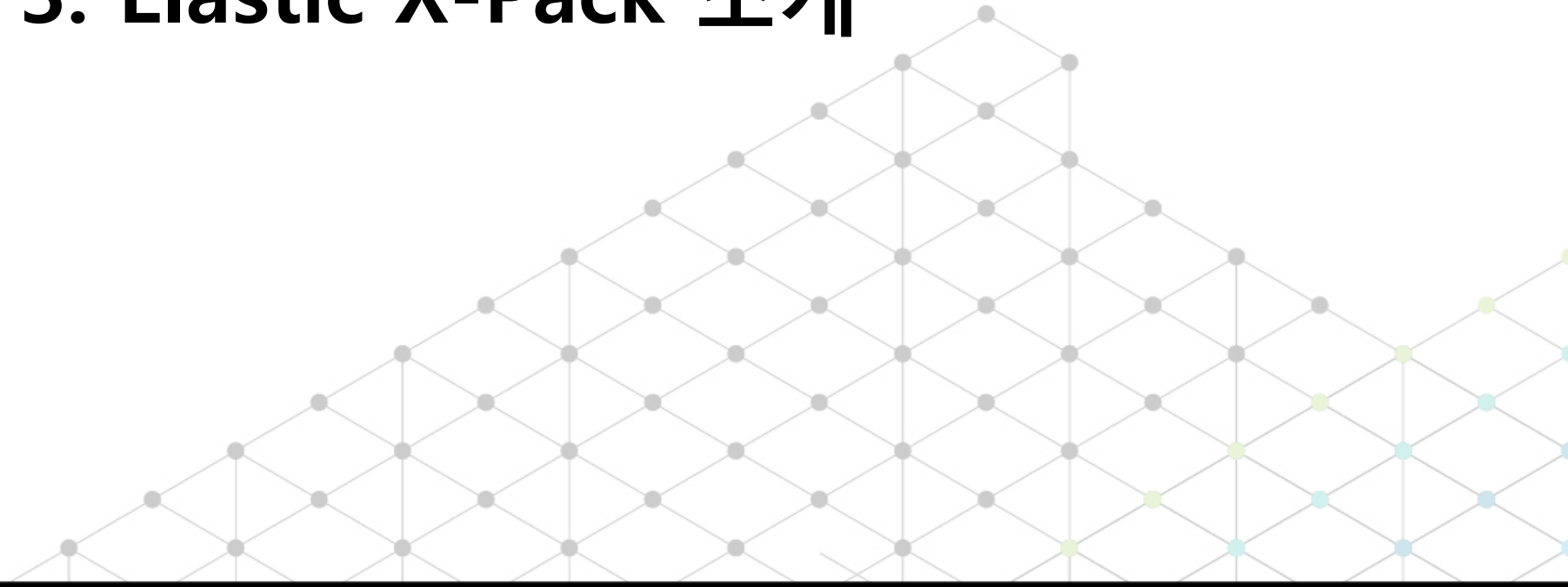
Kibana(aggregations)

Aggregations 프레임 워크는 검색 쿼리를 기반으로 Aggregations 된 데이터 제공

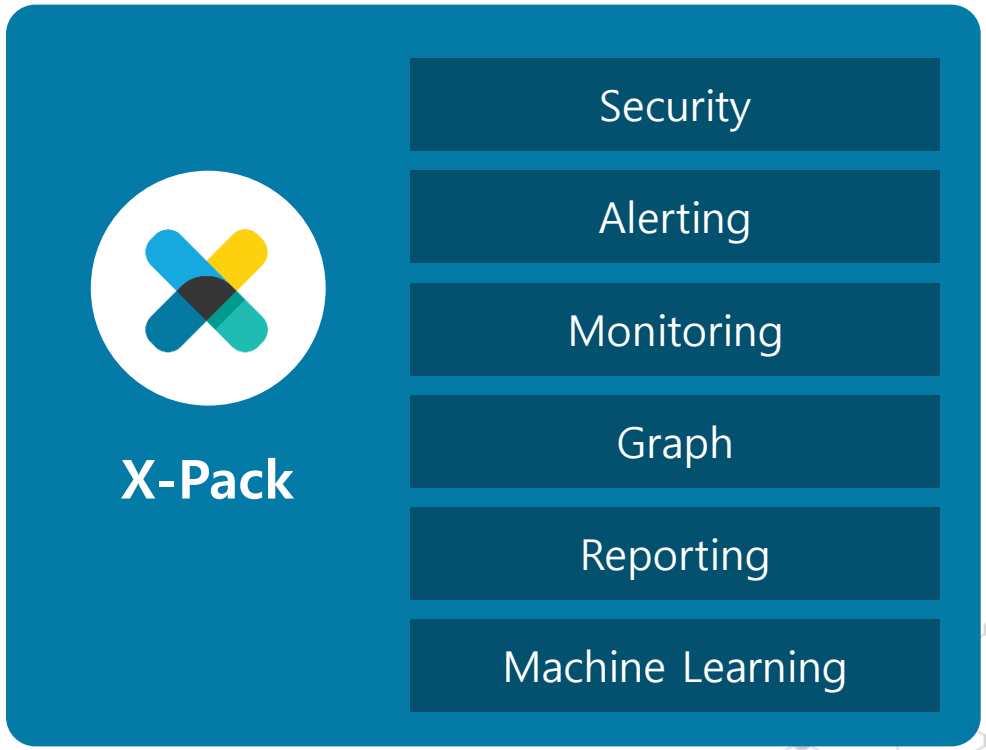
Aggregations에는 여러 가지 유형이 있으며, 각각 고유 한 목적과 결과 도출

Bucketing	• 키와 문서의 기준과 연관되어있는 버킷을 빌드하는 Aggregation 모음 • Aggregation이 실행되면 모든 버킷 기준이 컨텍스트의 모든 문서에서 평가되고 기준에 일치하면 해당 버킷에 "속하는"것으로 간주됨
Metric	• 집합 문서를 통해 메트릭을 추적하고 계산하는 Aggregation
Matrix	• 여러 필드에서 실행하고 요청 된 문서 필드에서 추출한 값을 기반으로 행렬 결과 생성
Pipeline	• 다른 Aggregation 및 관련 메트릭의 출력을 집계하는 Aggregations • 파이프 라인 Aggregation은 문서 집합이 아닌 다른 집계에서 생성 된 출력에서 동작

3. Elastic X-Pack 소개



Elastic X-Pack 소개



Elastic Stack의 활용 가치를
“UP”

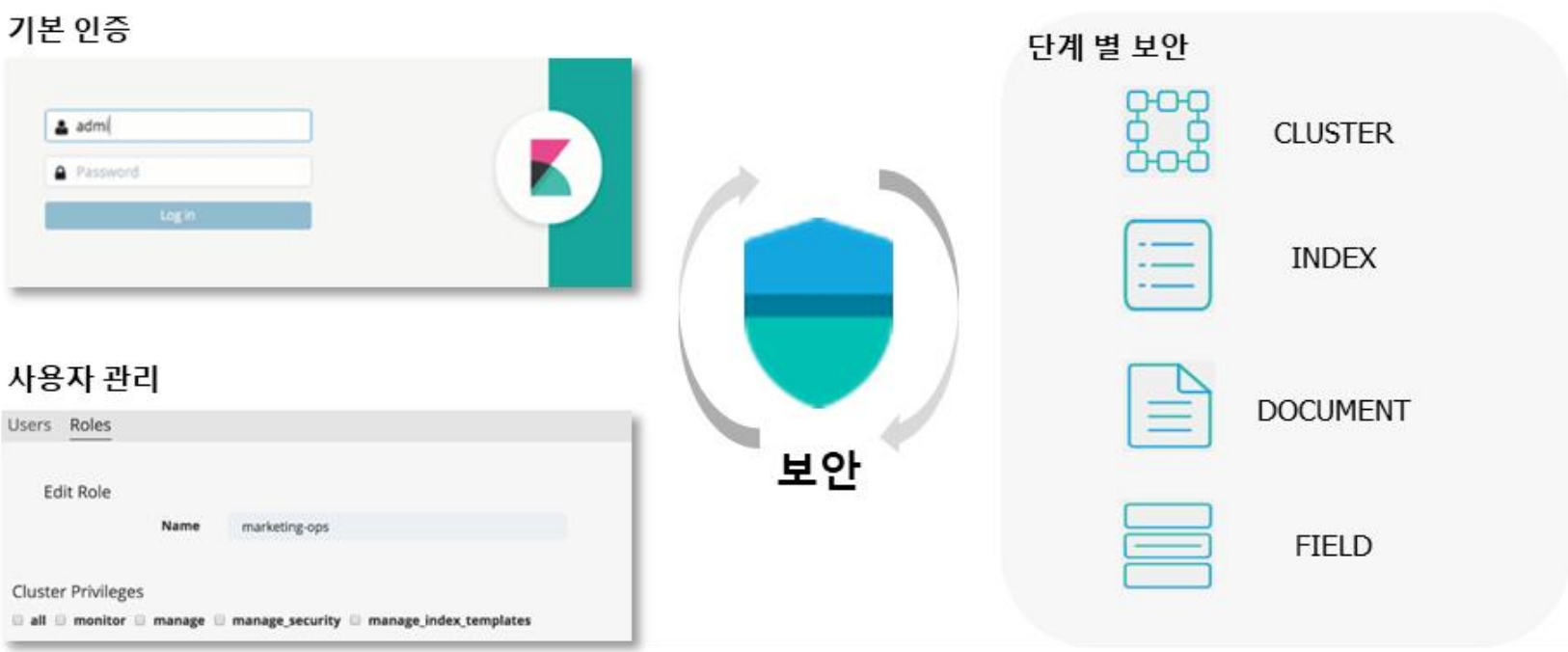
X-Pack은 보안, 알림, 모니터링, 보고, 그래프 머신러닝 기능을 단일 패키지로 번들 구성한 Elastic Stack 확장 프로그램

Security

특징

- Elastic Stack의 사용자 역할과 권한 관리
- 필드 및 도큐먼트 수준의 보안 제공
- SSL/TLS 암호화로 통신 트래픽 전송 보호
- ABAC(Attribute-based access control) 제공

Security Overview

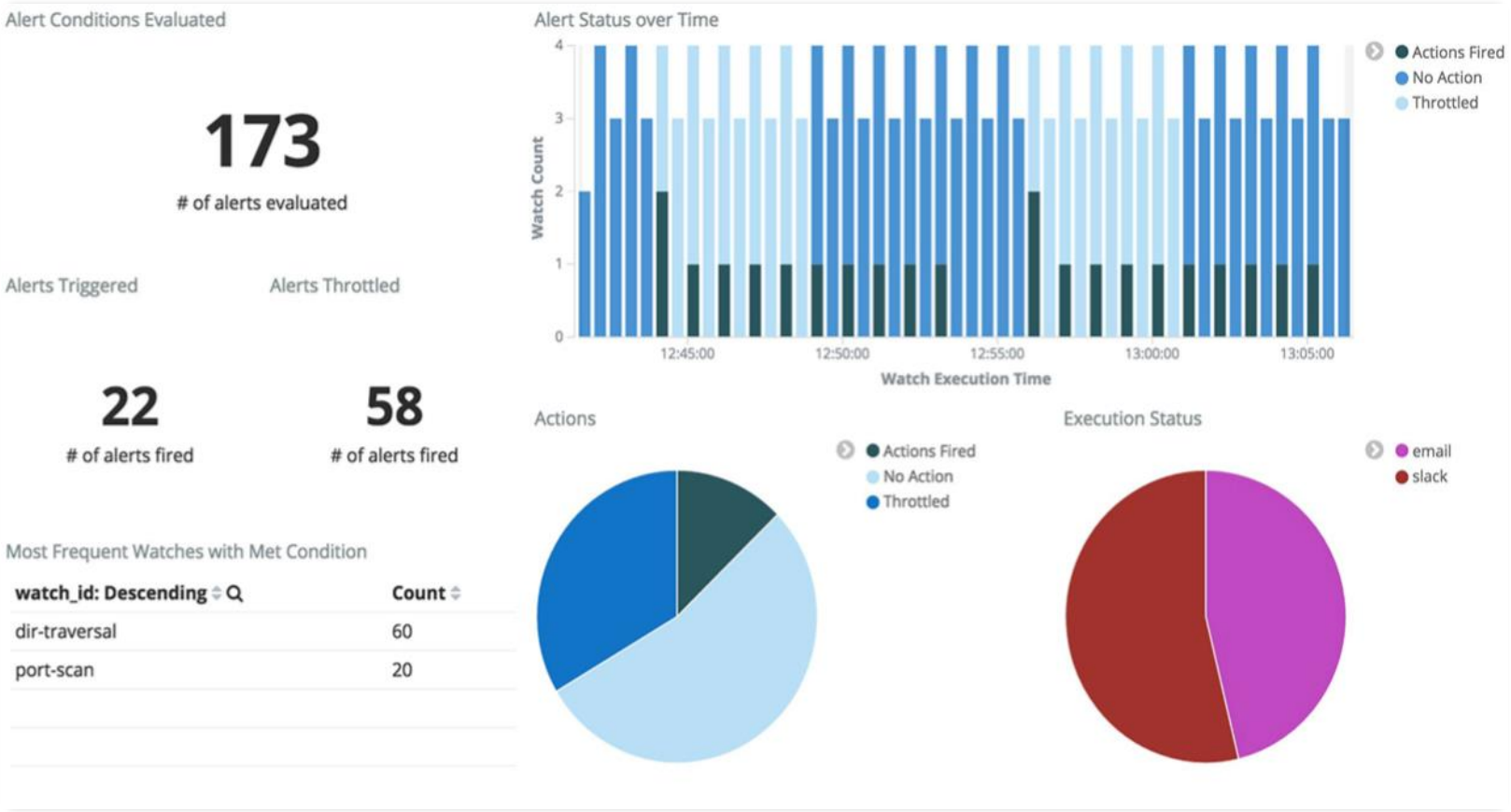


Alerting

특징

- Elastic Stack의 데이터 변경 사항 식별
- 머신러닝 기능과 조합하여 이상징후 알림
- 다양한 알림 방식 지원

Alerting Dashboard



Alerting 활용 예

- 데이터의 변경 또는 이상 요소를 감시하고 그에 대응하여 필요한 작업 수행
- X-Pack은 워치를 생성, 관리, 테스트할 수 있는 API 제공

인프라 모니터링, 디스크 사용 경향 추적

악성 활동을 탐지하도록 네트워크 활동 추적, 악성 사용자 거부

모니터링 중 노드가 클러스터를 떠나거나 쿼리 처리량이 예상 범위를 초과할 경우 알림

관련 데이터 또는 데이터 변경 사항은 정기적인 Elasticsearch 쿼리로 식별

Monitoring

특징

- Elastic Stack의 지속적인 성능 확인
- Elastic 클러스터의 헬스체크를 통해 높은 가용성 유지

Monitoring Dashboard

모니터링 지표

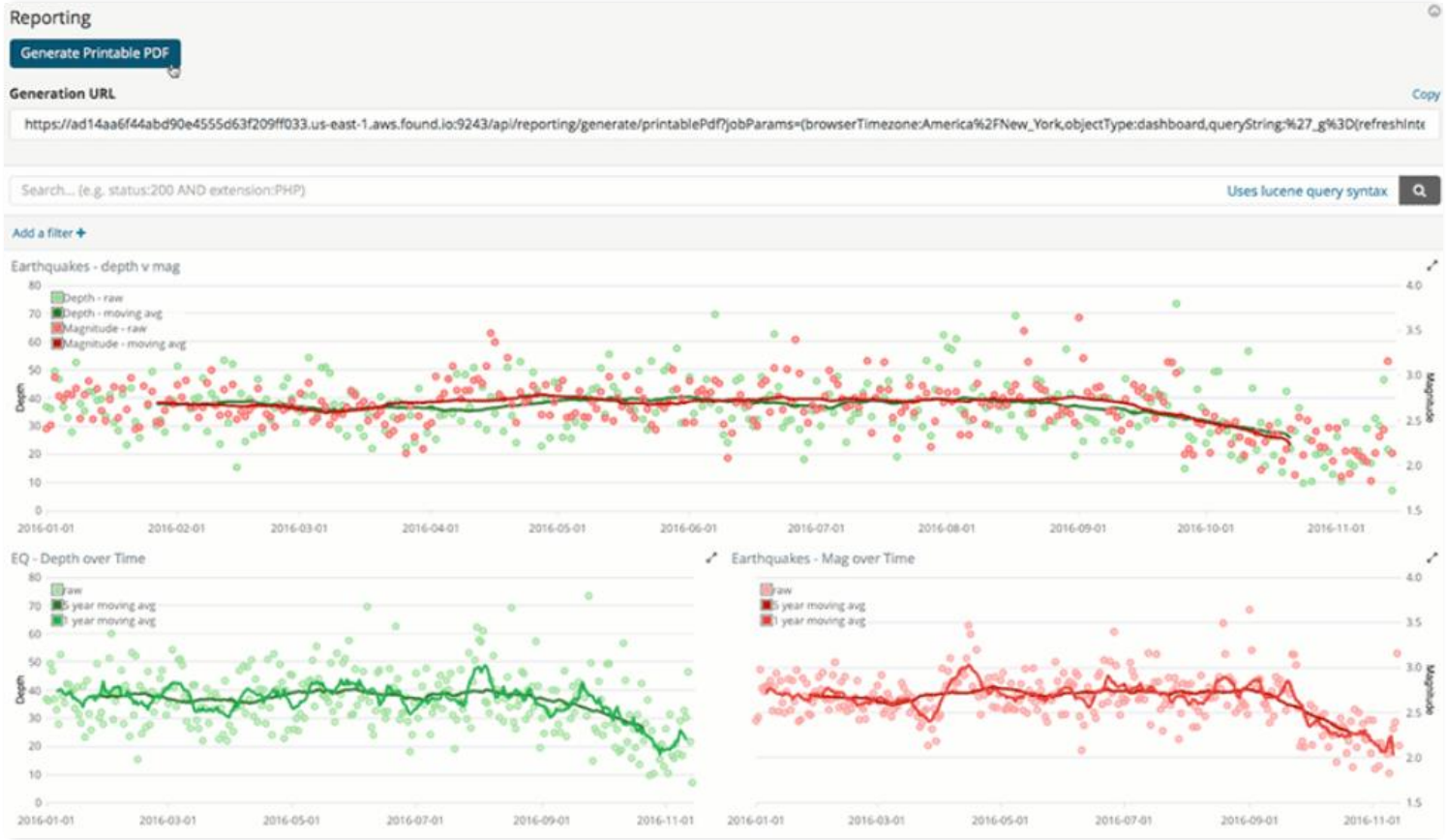
- Cluster
  - Cluster Alerts
- Node
  - Jvm Heap
  - Free disk space
  - Documents
  - Data
  - Indices
  - Shards
- Indices
  - Status
  - Index Rate
- Kibana
  - Memory size
  - System load
  - HTTP Connection
- Logstash
  - Events Received Rate
  - Event Latency
  - CPU Utilization

Reporting

특징

- 편리한 보고서 생성
  CSV, PDF 형식 다운로드
- 일정 또는 이벤트 기준
  보고서 생성
- 리포팅 기능으로
  원하는 시각화,대시보드를
  손쉽게 공유

Reporting Dashboard

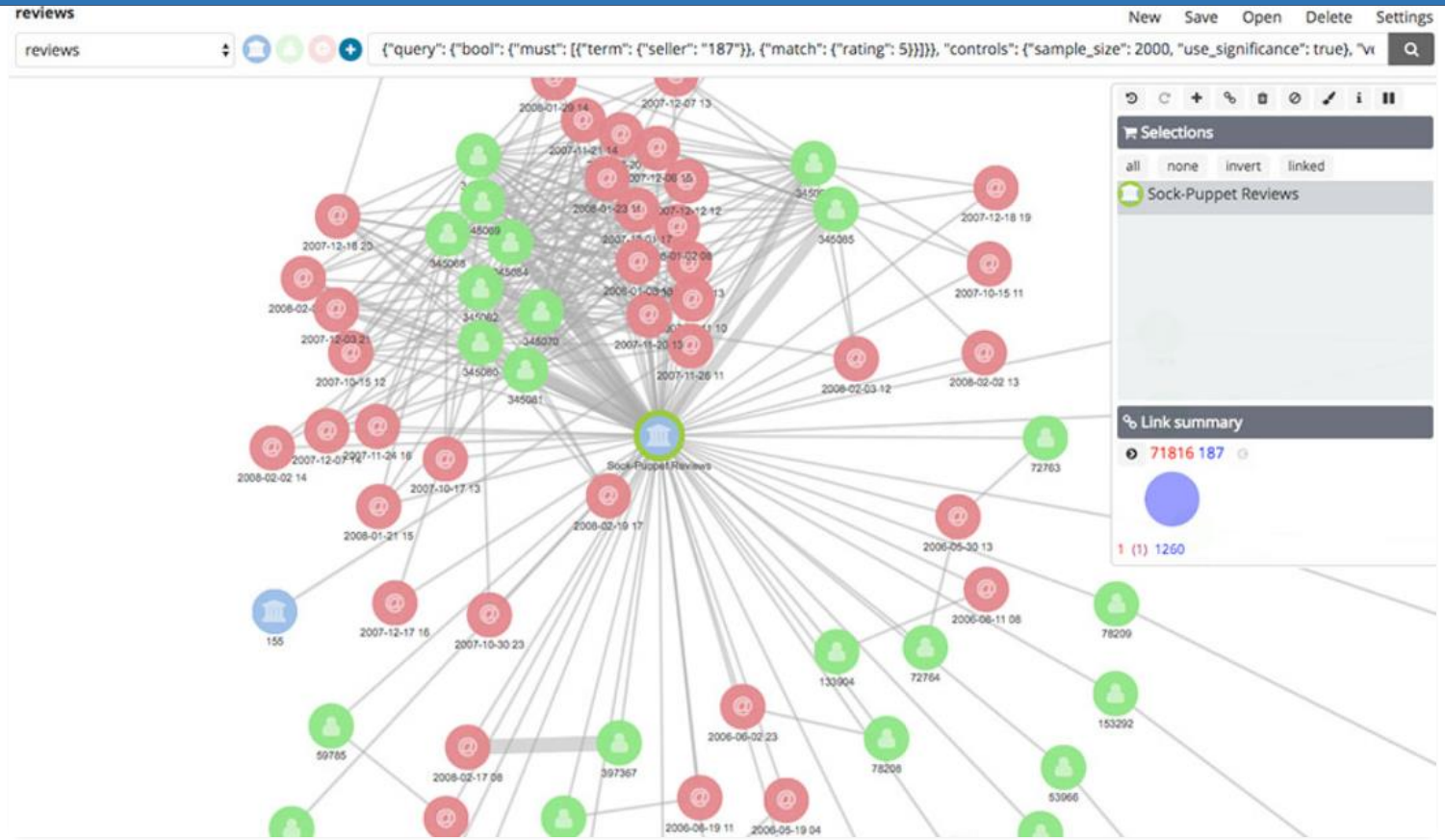


Graph

특징

- 데이터 관계 분석
- 분산 쿼리 실행, 실시간 데이터 활용 및 대용량 인덱싱 기능
- 데이터 유사 관계 파악 API 및 UI 도구

Graph Dashboard



Graph 활용 예

- X-Pack Graph 기능을 사용하여 Elasticsearch 색인에 포함된 항목의 연관성 파악
- 색인화된 용어의 상호 연결 관계를 탐색하고 가장 중요한 연결 확인

웹 사이트 해킹 시도에 대한 공통된 행동 패턴 분석

특정 상품 구매자들의 선호도 분석

데이터 구조 및 연관성 순위를 활용, 연결된 데이터에서 유용한 시그널과 노이즈 구분

계정 A와 B 간의 모든 거래를 검색하지 않고 그 관계를 나타내는 하나의 관계를 유추

Machine Learning

특징

- 시계열 데이터 활용, 모든 종류의 이상징후 탐지
- 데이터의 트렌드와 주기성 실시간 모델링
- 데이터 행동 유형 학습 및 근본 원인 분석

Machine Dashboard

머신러닝 Start Option

- Choose index or Input index
- Select the Indices
- Select Type
- Choose Time-field name
- Create a new job
  - Job details
  - Analysis Configuration
    (bucket_span, Detectors)
  - Datafeed
    (Query, Frequency, Scroll size)
  - Edit Json
- Loading Kibana

Machine Learning 활용 예

- X-Pack 머신 러닝은 Elasticsearch 데이터의 트렌드, 주기성 등을 자동으로 실시간 모델링
- 빠르게 문제를 식별하고 근본 원인 분석, 간소화

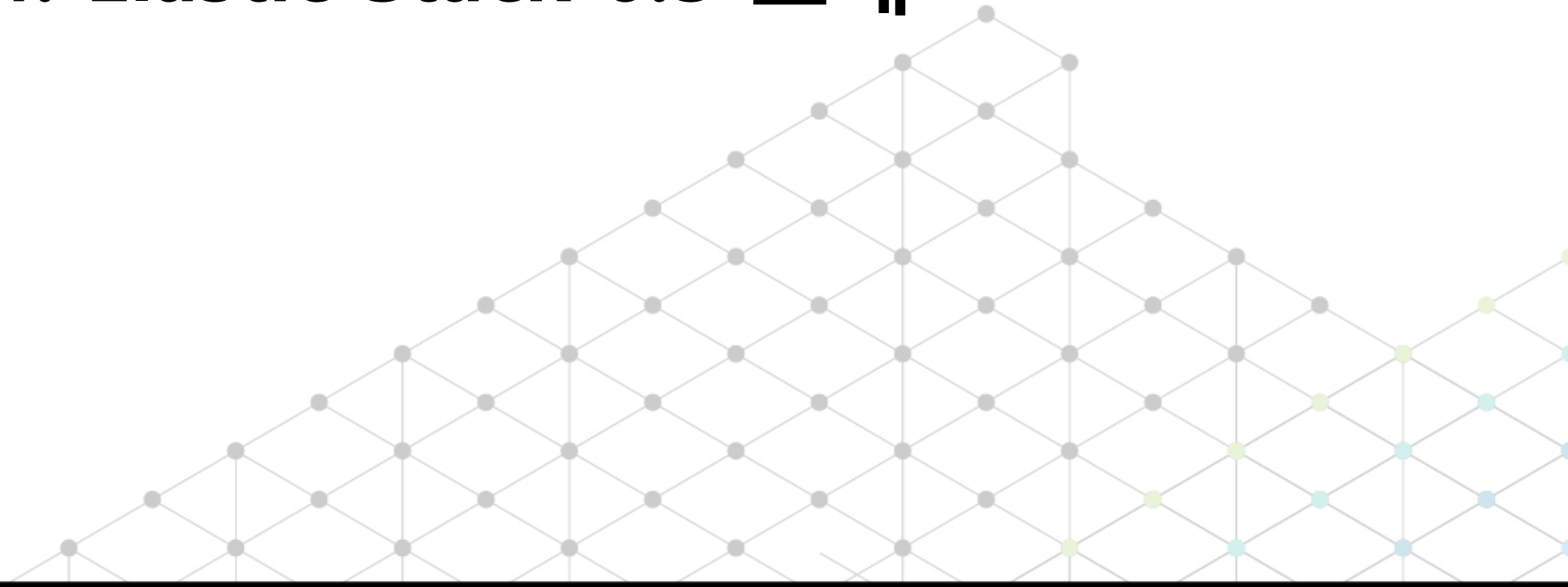
애플리케이션 요청이 비정상적으로 증가하거나, 감소한 원인 파악

비정상적인 네트워크 활동이나 사용자 행동 식별

데이터의 정상적인 행동 유형을 스스로 학습하여 비정상적인 유형 식별

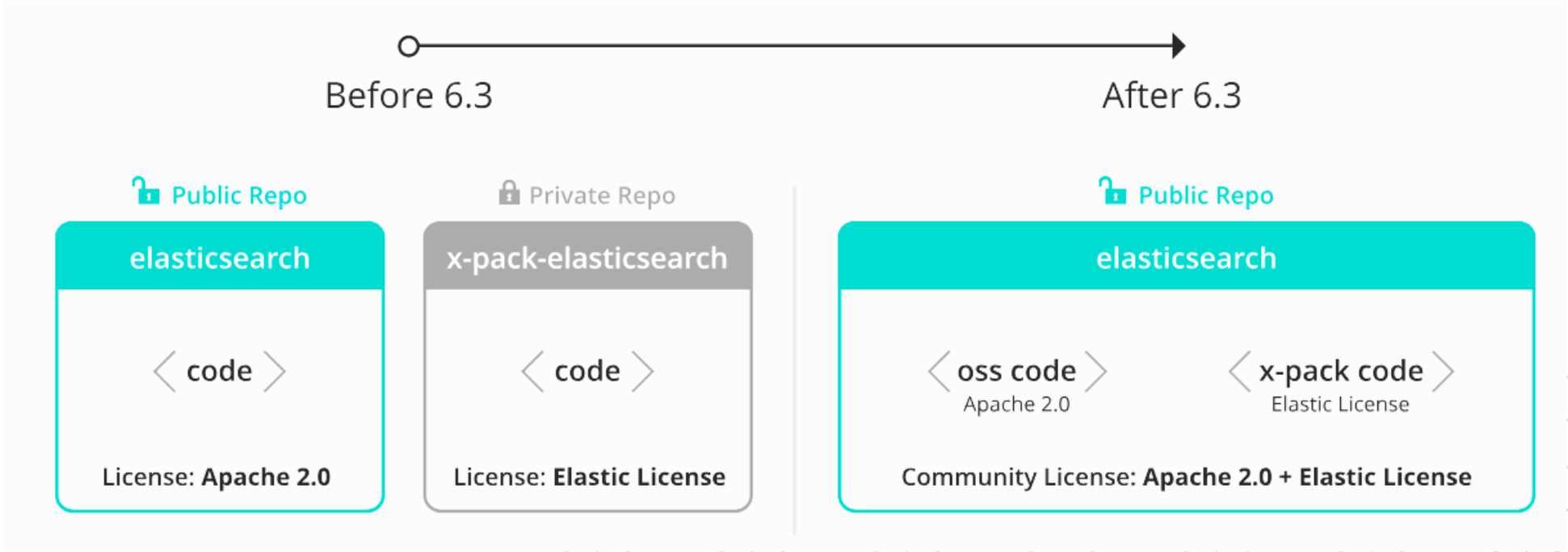
로그 메시지 분류, 특이한 이벤트 또는 비정상적인 유형의 메시지 파악

4. Elastic Stack 6.3 소개



Elastic Stack 6.3버전 소개

- 최근 릴리즈 된 6.3버전은 아래와 같이 X-Pack code가 Elastic License로 배포
- X-Pack의 일부 기능은 기간 제한 없이 사용 가능



- 최근 릴리즈 된 6.3버전은 아래와 같이 free basic 라이선스로 X-Pack의 일부 기능 사용 가능
- X-Pack 기능 중 Security, Alerting, Machine learning, Graph 기능은 해당 사항(X)
- 그 외 기능 일부 또는 전부 "Free basic" 라이선스로 이용 가능.

	FREE	
	OPEN SOURCE	BASIC
	Download	
X-PACK		
Security (formerly Shield)		
Monitoring (formerly Marvel)		✓
Management		✓
Alerting (via Watcher)		
Machine Learning		
APM		✓
Graph Analytics & Visualization		
Reporting		✓
SQL Access		✓
Modules		✓
Dev Tools		✓

* Monitoring

Full stack monitoring		✓
Multi stack monitoring support		
Configurable retention policy		
Automatic alerts on stack issues		

* Management

Upgrade Assistant APIs & UI		✓
Logstash pipeline management		

* Reporting

CSV exports		✓
PDF reporting		

* SQL Access

REST and Translate APIs		✓
Command Line Interface (CLI) tools		✓
JDBC client		

참고 URL : https://www.elastic.co/subscriptions