한국산업인력공단

국가인적자원개발컨소시엄

Io

ELK를 이용한 실시간 데이터 수집, 분석, 시각화

2015. 6. 25 ~ 7. 2

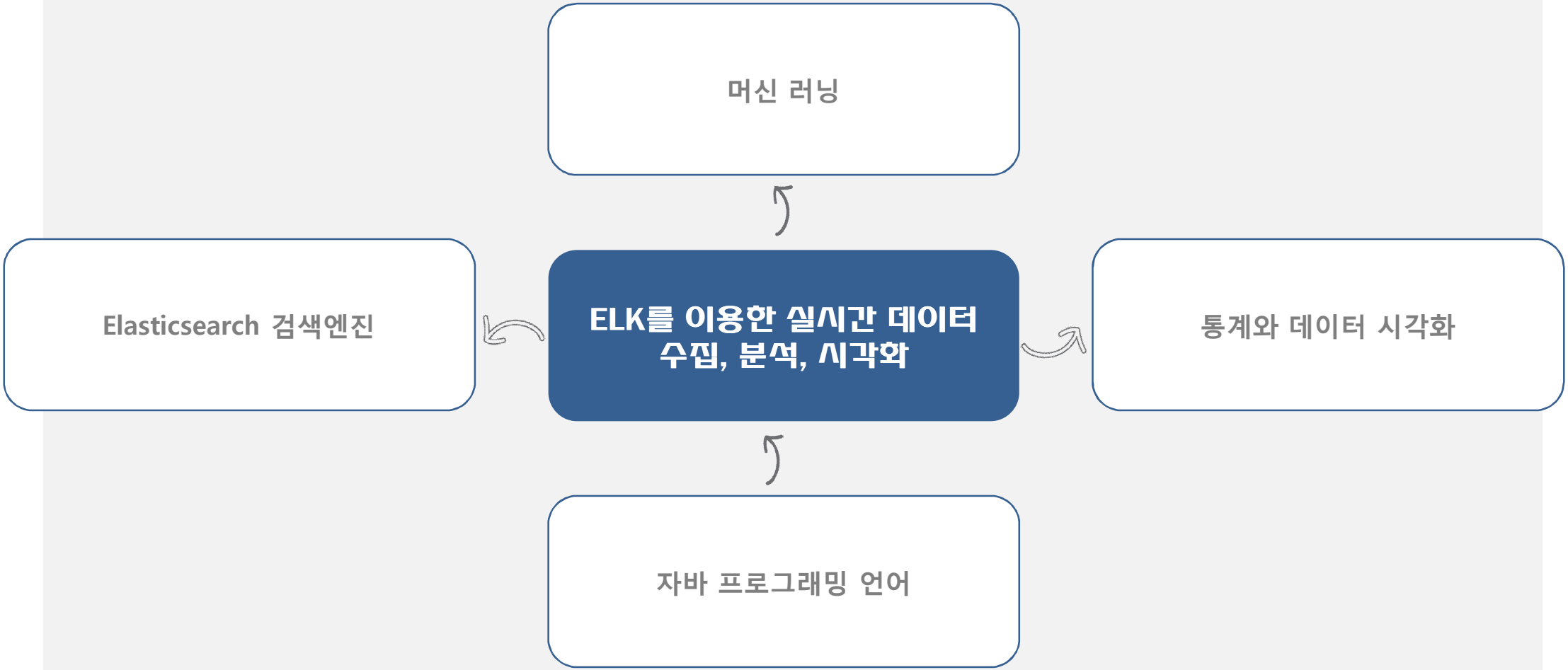
대한민국 소프트웨어 기술진흥 · 인력양성 대표기관

한국소프트웨어기술진흥협회

고용노동부 한국산업인력공단

한국소프트웨어기술진흥협회

과정 Learning Map



Learning Object(학습모듈) 및 커리큘럼

LO	커리큘럼
ELK를 이용한 실시간 데이터 수집 및 시각화 개요	ELK의 개요 실시간 데이터 수집, 검색, 분석, 시각화 절차의 이해 실시간 데이터 수집 및 시각화의 실제 예제
Logstash를 이용하여 데이터를 실시간으로 수집하기	Logstash 소개 Logstash 설치 및 환경 설정 플러그인 관리 및 사용하기 실제 예제를 통한 실습
Elasticsearch를 이용하여 실시간으로 데이터 검색 및 분석 하기	Elasticsearch 소개 Elasticsearch 설치 및 환경설정 인덱스 생성하기 질의 하기 실제 예제를 통한 실습
Kibana를 이용하여 실시간으로 수집한 데이터 시각화 하기	Kibana 소개 Kibana 설치 및 환경설정 대시보드 커스터마이징 하기 시각화 하기 실제 예제를 통한 실습

ELK를 이용한 실시간 데이터 수집 및 시각화 개요

-ELK의 개요
-실시간 데이터 수집, 검색, 분석, 시각화 절차의 이해
-실시간 데이터 수집 및 시각화의 실제 예제

◎ 학습지식 개요/요점

• ELK는 Elastic에서 개발한 ElasticSearch(검색엔진), LogStash(로그 수집기), Kibana(시각화도구)로 구성된 수집, 검색, 시각화를 실시간으로 처리할 수 있는 오픈소스 패키지이다.

학습내용(Table of Contents)

- ELK의 개요
- 실시간 데이터 수집, 검색, 분석, 시각화 절차의 이해
- 실시간 데이터 수집 및 시각화의 실제 예제

ELK의 개요

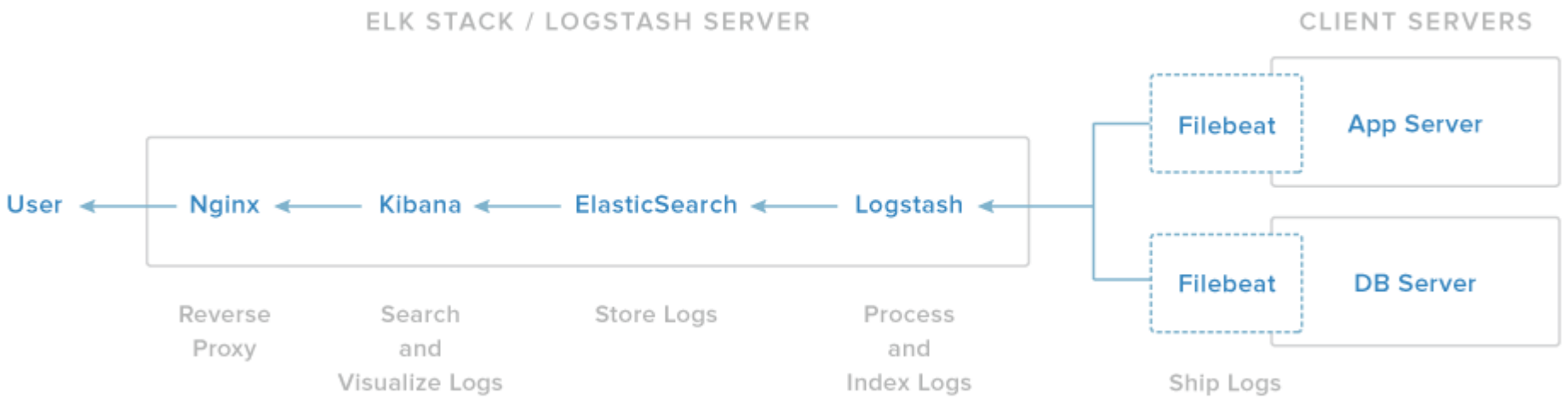
- ELK : Elasticsearch(검색엔진) + Logstash(로그수집기) + Kibana(시각화도구)
- Elasticsearch는 Apache Lucene 기반의 실시간 분산 검색 엔진
- Logstash는 각종 로그를 가져와 JSON형태로 만들어 Elasticsearch로 전송
- Kibana는 Elasticsearch에 저장된 Data를 사용자에게 Dashboard 형태로 보여주는 솔루션
- Beats를 추가해 Elastic Stack이라 부름

- GA : Google Analytics
- DataViz : 넷플릭스에서 유행하는 단어 Data Visualization

ELK의 개요

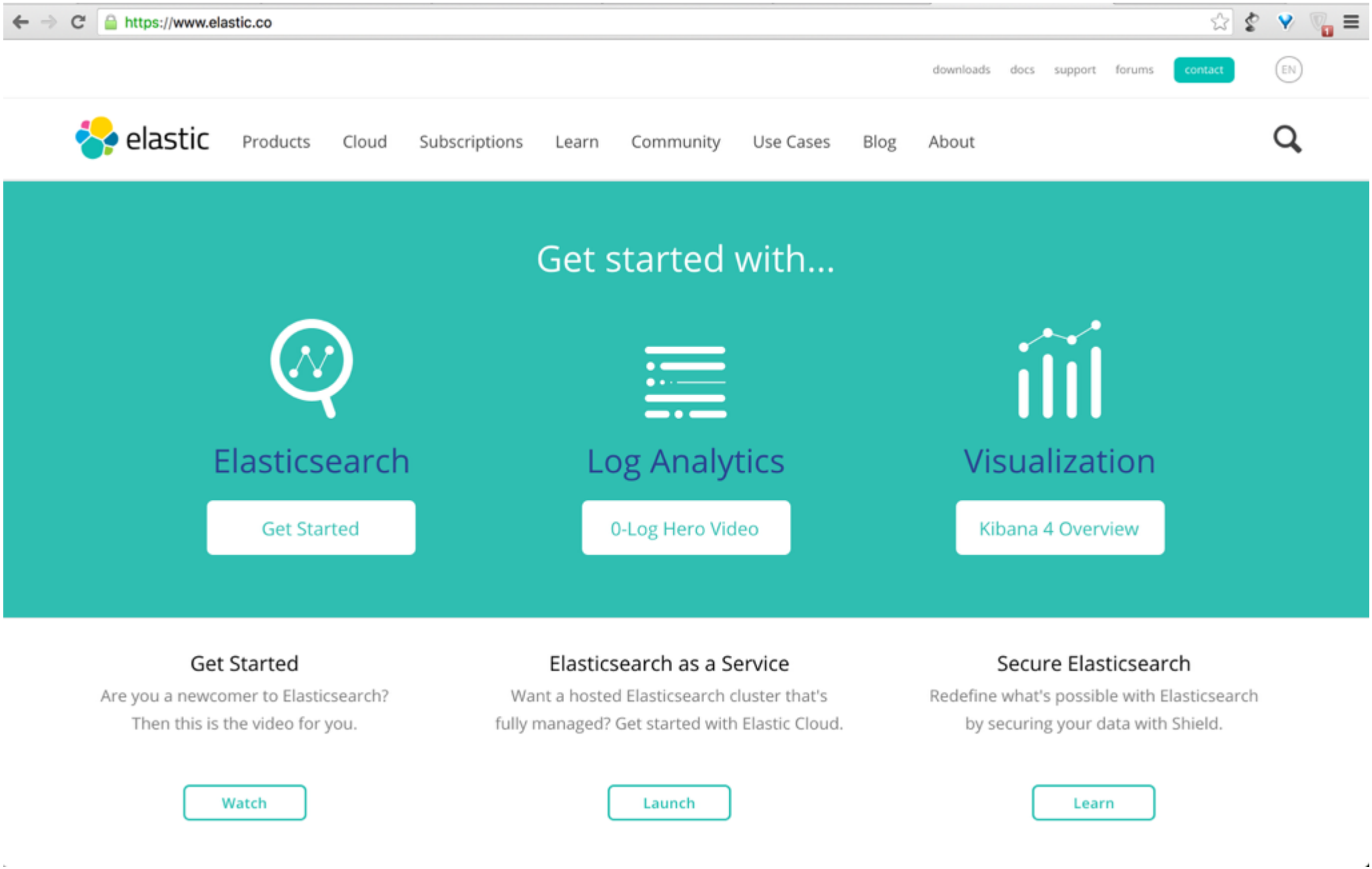
http://elastic.co 사이트 오픈소스 제품

ELK의 개요

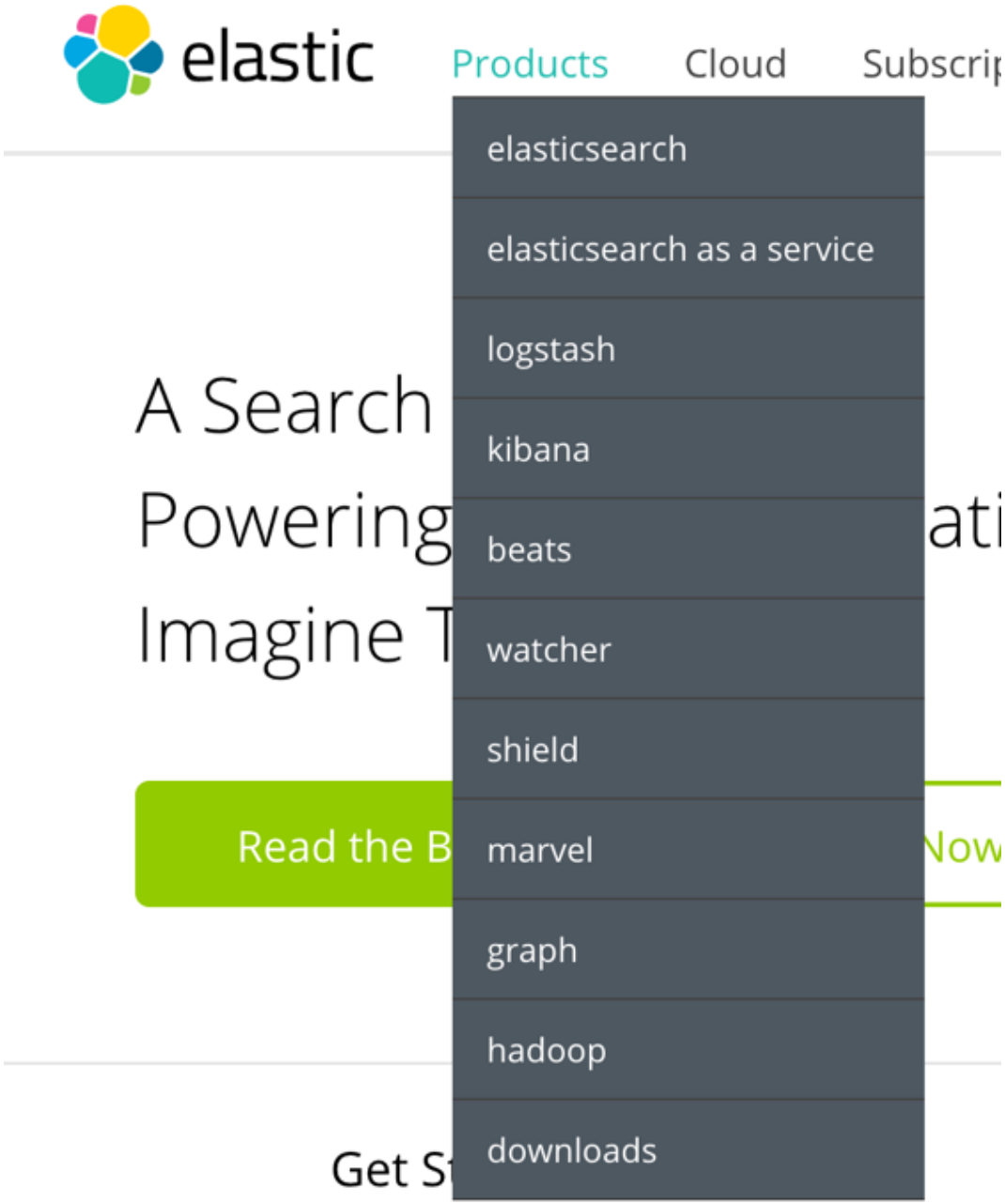


https://www.digitalocean.com/community/tutorials/how-to-install-elasticsearch-logstash-and-kibana-elk-stack-on-ubuntu-14-04

ELK의 개요


https://www.elastic.co
downloads
docs
support
forums
contact
EN
elastic
Products
Cloud
Subscriptions
Learn
Community
Use Cases
Blog
About
Get started with...
Elasticsearch
Get Started
Log Analytics
0-Log Hero Video
Visualization
Kibana 4 Overview
Get Started
Are you a newcomer to Elasticsearch?
Then this is the video for you.
Watch
Elasticsearch as a Service
Want a hosted Elasticsearch cluster that's
fully managed? Get started with Elastic Cloud.
Launch
Secure Elasticsearch
Redefine what's possible with Elasticsearch
by securing your data with Shield.
Learn

ELK의 개요


elastic
Products
Cloud
Subscrip
elasticsearch
elasticsearch as a service
logstash
kibana
beats
watcher
shield
marvel
graph
hadoop
downloads
A Search
Powering
ati
Imagine T
Read the B
Now
Get St

GA와의 차이점

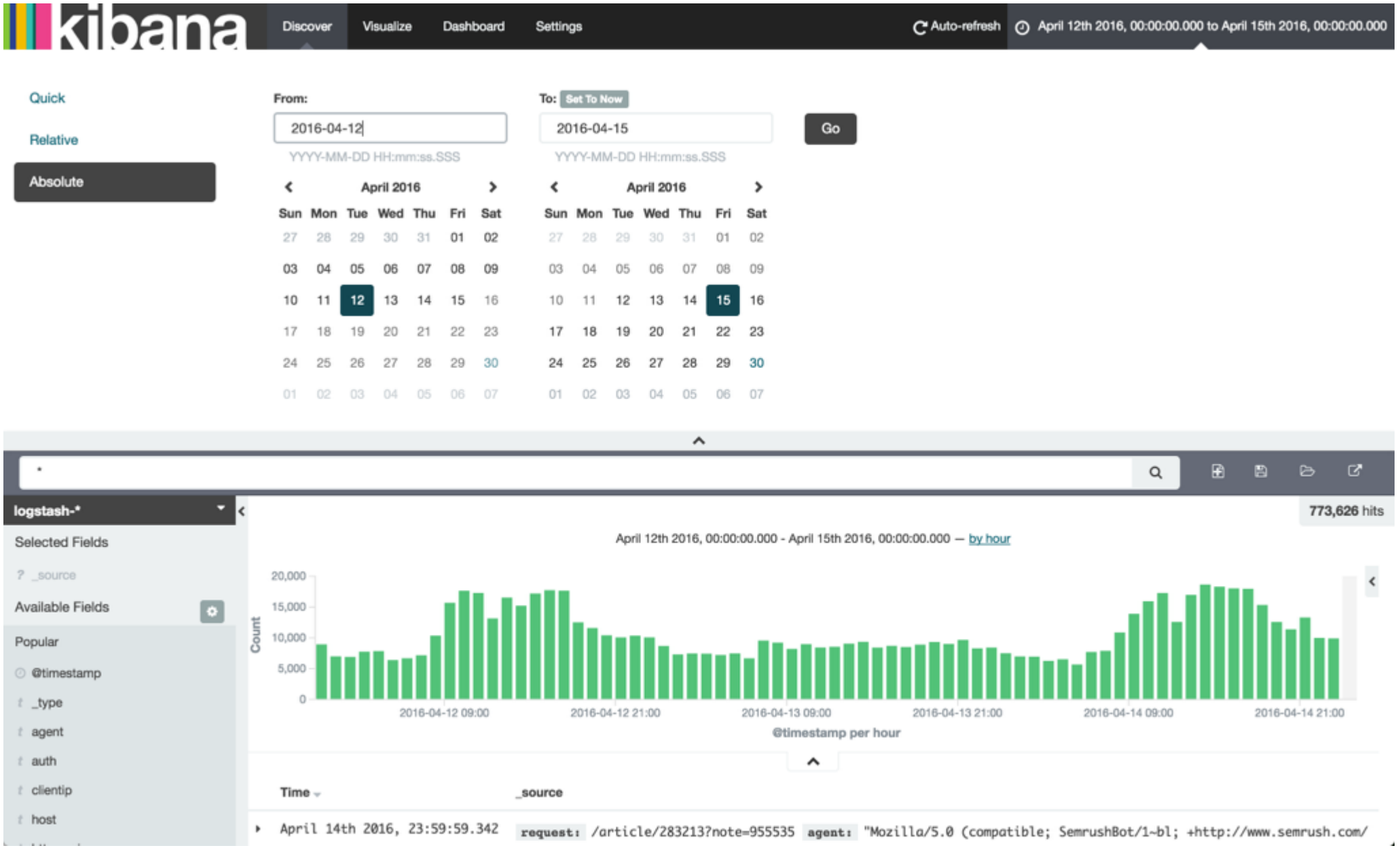
- GA는 페이지마다 스크립트 삽입
- ELK는 서버의 AccessLog 기반이라 누락 없음

```
<script>
    (function(i,s,o,g,r,a,m){i['GoogleAnalyticsObject']=r;i[r]=i[r]||function(){
        (i[r].q=i[r].q||[]).push(arguments)},i[r].l=1*new Date();a=s.createElement(o),
            m=s.getElementsByTagName(o)[0];a.async=1;a.src=g;m.parentNode.insertBefore(a,m)
    })(window,document,'script','//www.google-analytics.com/analytics.js','ga');

    ga('create', 'UA-6[redacted]25-5', 'auto');
    ga('send', 'pageview');

</script>
```

ELK의 개요


Google Analytics
홈
보고서
맞춤설정
관리
kenu.heo@gmail.com
OKKY - http://okky.kr
전체 웹사이트 데이터
보고서 및 도움말 검색
대시보드
바로가기
지능형 이벤트
실시간
잠재고객
개요
활성 사용자
동질 집단 분석 베타
사용자 탐색기 베타
인구통계
관심분야
지역
방문 형태
기술환경
모바일
맞춤
벤치마킹
사용자 흐름
잠재고객 개요
2016. 3. 30. - 2016. 4. 29.
이메일
내보내기
대시보드에 추가
바로가기
이 보고서는 598,898회의 세션(전체 세션의 100%)을 기반으로 작성되었습니다. 자세히 알아보기
느리지만 정확도가 높음
모든 사용자
100.00% 세션
+ 세그먼트
개요
세션
vs.
측정항목 선택
시간
일
주
월
세션
40,000
20,000
4월 5일
4월 12일
4월 19일
4월 26일
세션
598,898
사용자
264,761
페이지뷰 수
2,931,830
세션당 페이지수
4.90
평균 세션 시간
00:04:09
이탈률
60.73%
새로운 세션 %
32.92%
Returning Visitor
New Visitor
32.9%
67.1%

ELK의 개요


kibana
Discover
Visualize
Dashboard
Settings
Auto-refresh
April 12th 2016, 00:00:00.000 to April 15th 2016, 00:00:00.000
Quick
Relative
Absolute
From:
2016-04-12
YYYY-MM-DD HH:mm:ss.SSS
To:
Set To Now
2016-04-15
YYYY-MM-DD HH:mm:ss.SSS
Go
April 2016
Sun Mon Tue Wed Thu Fri Sat
logstash-*
Selected Fields
_source
Available Fields
Popular
@timestamp
_type
agent
auth
clientip
host
773,626 hits
April 12th 2016, 00:00:00.000 - April 15th 2016, 00:00:00.000 — by hour
Count
20,000
15,000
10,000
5,000
0
2016-04-12 09:00
2016-04-12 21:00
2016-04-13 09:00
2016-04-13 21:00
2016-04-14 09:00
2016-04-14 21:00
@timestamp per hour
Time
_source
April 14th 2016, 23:59:59.342
request: /article/283213?note=955535 agent: "Mozilla/5.0 (compatible; SemrushBot/1~bl; +http://www.semrush.com/

실시간 데이터 수집, 검색, 분석, 시각화 절차의 이해

- 목표 데이터 선정
    - 액세스 로그, OpenAPI로 끌어온 로그, 시스템 로그 등
- 파싱 패턴
    - 로그의 패턴 설정
- 검색엔진에 인덱싱
    - 로그 필드별 매핑 설정 후 인덱싱
- 시각화 도구로 항목 선정
    - 필드별 통계 및 질의어에 대한 통계
- 차트로 시각화
    - 각 데이터의 특성에 맞게 차트 생성

ELK 특징

- Google Analytics(GA)는 사이트 접속 통계를 구할 경우 원하는 대로 데이터를 획득하기 어렵다. (시간별, IP별, 등)
- 자체 서버의 모든 로그를 100% 수집할 수 있기 때문에 데이터에 대한 신뢰성이 높다.
- 파라미터 값별로 통계를 볼 수 있기 때문에 정확한 데이터 분석이 가능하다.
- 검색엔진(lucene)이 포함되어 있어, 빠르게 데이터를 검색할 수 있다.
- 모두 오픈소스이며 자유롭게 사용이 가능하다.

AccessLog 예

```
112.72.239.19 - - [14/Apr/2016:23:59:54 +0900] "GET /article/321382
HTTP/1.1" 200 4460 "http://okky.kr/articles/evalcom" "Mozilla/5.0
(Windows NT 6.3;WOW64; Trident/7.0; MASMJS; rv:11.0) like Gecko" "-"
```

IP	112.72.239.19
인증정보	- -
시간	[14/Apr/2016:23:59:54 +0900]
Method URI	"GET /article/321382 HTTP/1.1"
상태코드	200
용량	4460
referer	"http://okky.kr/articles/evalcom"
user-agent	"Mozilla/5.0 (Windows NT 6.3;WOW64; Trident/7.0; MASMJS; rv:11.0) like Gecko" "-"

grok pattern

- **COMMONAPACHELOG**

```
%{IPORHOST:clientip} %{HTTPDUSER:ident} %{USER:auth}
\[%{HTTPDATE:timestamp}\] "(?:%{WORD:verb} %{NOTSPACE:request}(?:
HTTP/%{NUMBER:httpversion})?|%{DATA:rawrequest})" %{NUMBER:response}
(?:%{NUMBER:bytes}|-)
```

- **COMBINEDAPACHELOG**

```
%{COMMONAPACHELOG} %{QS:referrer} %{QS:agent}
```

https://github.com/logstash-plugins/logstash-patterns-core/blob/master/patterns/grok-patterns

실습 예제 및 수행가이드

- AWS 또는 클라우드 사용
- 메모리 4G 이상
- JDK1.8 설치
- git-scm.com 의 git bash와 cmd 사용(Windows)

https://okdevtv.com/mib/elk/elk

학습진단 평가문제

1. ELK의 구성 패키지별 특징
2. Elasticsearch 검색엔진의 특징
3. 실시간 데이터 수집 및 시각화의 실제 예제

Logstash를 이용하여 데이터를 실시간으로 수집하기

-Logstash 소개
-Logstash 설치 및 환경 설정
-플러그인 관리 및 사용하기
-실제 예제를 통한 실습

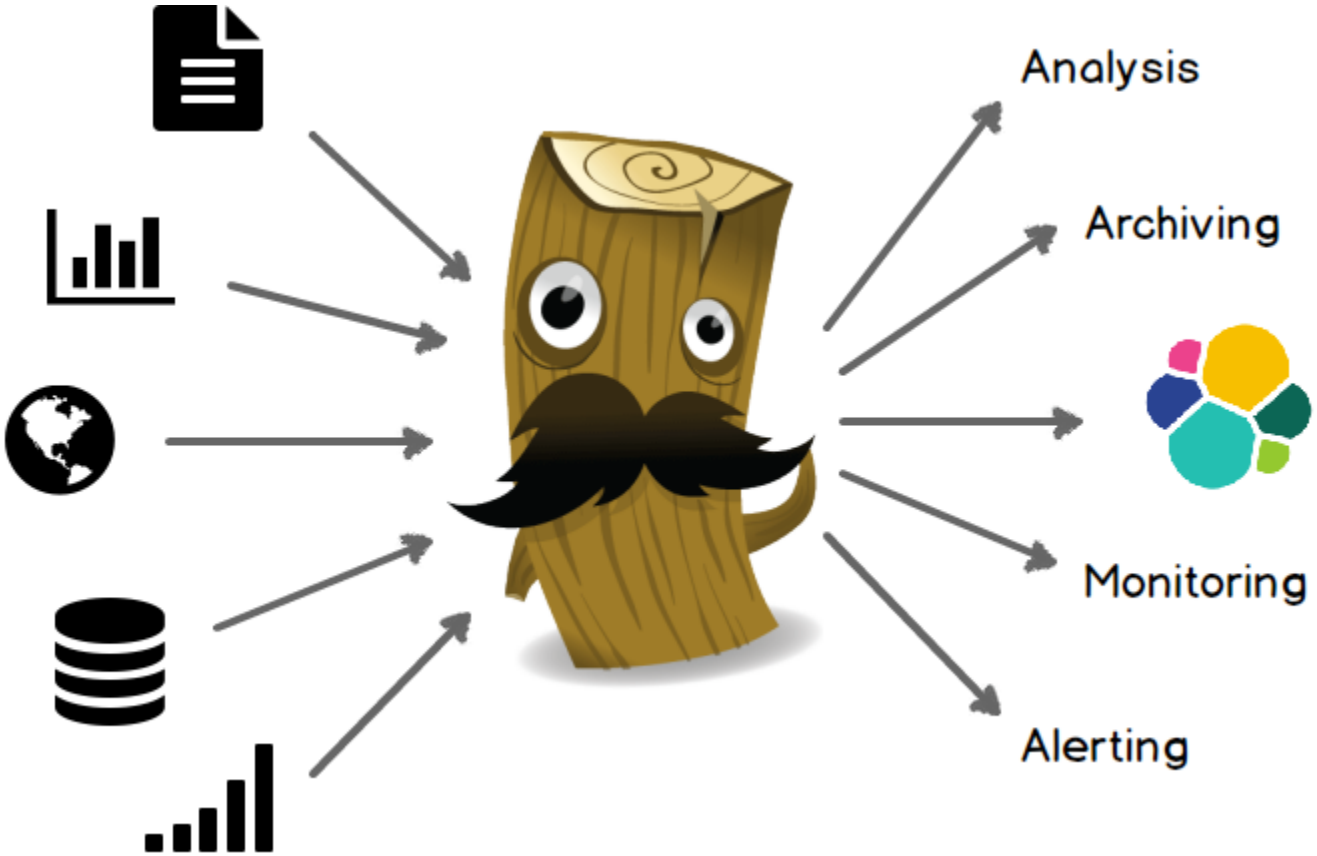
◎ 학습지식 개요/요점

• Logstash는 다양한 시스템에서 생성되는 데이터를 실시간으로 수집하고 가공하는 로그 수집기이다.

학습내용(Table of Contents)

- Logstash 소개
- Logstash 설치 및 환경 설정
- 플러그인 관리 및 사용하기
- Input Plugin
- Output Plugin
- Filter Plugin
- Codec Plugin
- 실제 예제를 통한 실습

Logstash 소개



https://www.elastic.co/guide/en/logstash/current/introduction.html

Logstash 설치 및 환경 설정

- JDK8 이상 필요
- 메모리 많이 차지함

```
cd ~/local
wget https://download.elastic.co/logstash/logstash/logstash-2.3.2.tar.gz
tar xvfz logstash-2.3.2.tar.gz
ln -s logstash-2.3.2 logstash
cd logstash
```

플러그인 관리 및 사용하기

- 기본 플러그인
- input {}
- filter {}
- output {}
- 설정 파일을 만들어서 관리
- https://www.elastic.co/guide/en/logstash/current/first-event.html

```
cd logstash-2.3.0

bin/logstash -e 'input { stdin { } } output { stdout {} }'
```

Input Plugin

- 입력 소스 설정
- **beats**, couchdb_changes, drupal_dblog, elasticsearch, exec, eventlog, **file**, ganglia, gelf, generator, graphite, github, heartbeat, heroku, http, http_poller, irc, imap, jdbc, jmx, kafka, log4j, lumberjack, meetup, pipe, puppet_facter, relp, rss, rackspace, rabbitmq, redis, salesforce, snmptrap, **stdin**, sqlite, s3, sqs, stomp, syslog, tcp, twitter, unix, udp, varnishlog, wmi, websocket, xmpp, zenoss, zeromq
- https://www.elastic.co/guide/en/logstash/current/input-plugins.html

Output Plugin

- 출력 형태 설정
- https://www.elastic.co/guide/en/logstash/current/output-plugins.html

Filter Plugin

- 필터 설정
- https://www.elastic.co/guide/en/logstash/current/filter-plugins.html

Codec Plugin

- 이벤트 표시 설정
- avro, cef, cloudtrail, cloudfront, collectd, compress_spooler, dots, edn, edn_lines, es_bulk, fluent, graphite, gzip_lines, json, json_lines, line, msgpack, multiline, netflow, nmap, oldlogstashjson, plain, **rubydebug**, s3_plain, spool
- https://www.elastic.co/guide/en/logstash/current/codec-plugins.html

플러그인 관리 및 사용하기

- filebeat 플러그인

```
cd ~/local/logstash

./bin/logstash-plugin install logstash-input-beats
```

플러그인 관리 및 사용하기

- logstash 설정

```
input {
  beats {
    port => 5044
  }
}
```

플러그인 관리 및 사용하기

• filebeat 설정

```
cd ~/local
wget https://download.elastic.co/beats/filebeat/filebeat-1.2.3-
x86_64.tar.gz
tar xvfz filebeat-1.2.3-x86_64.tar.gz
ln -s filebeat-1.2.3-x86_64 filebeat
cd filebeat
# elasticsearch 부분 #으로 주석 처리
  # elasticsearch:
    #hosts: ["localhost:9200"]
# logstash 부분 # 주석 해제
  logstash:
    hosts: ["localhost:5044"]

# filebeat.yml 내용 중 로그 위치 변경 `/var/log/nginx/*.log`
```

•실행

```
./filebeat -e -c filebeat.yml
```

실제 예제를 통한 실습

- 로컬에 설치해서 실습

https://okdevtv.com/mib/elk/elk

실습 예제 및 수행가이드

https://okdevtv.com/mib/elk/elk

학습진단 평가문제

Elasticsearch를 이용하여 실시간으로 데이터 검색 및 분석하기

-Elasticsearch 소개
-Elasticsearch 설치 및 환경설정
-인덱스 생성하기
-질의 하기
-실제 예제를 통한 실습

◎ 학습지식 개요/요점

• Elasticsearch는 Apache Lucene 검색엔진을 기반으로 하는 고성능 분산 검색 엔진으로써 데이터를 수집한 후 수집한 데이터의 인덱스를 생성하고 검색할 수 있도록 한다.

학습내용(Table of Contents)

- Elasticsearch 소개
- Elasticsearch 설치 및 환경설정
- 인덱스 생성하기
- 질의 하기
- 분석 하기
- 실제 예제를 통한 실습

실습 예제 및 수행가이드

https://okdevtv.com/mib/elasticsearch/elasticsearch

Kibana를 이용하여 실시간으로 수집한 데이터 시각화 하기

-Kibana 소개
-Kibana 설치 및 환경설정
-대시보드 커스터마이징 하기
-시각화 하기
-실제 예제를 통한 실습

◎ 학습지식 개요/요점

• Kibana는 대시보드를 통해 데이터를 시각화 하기 위한 강력한 실시간 시각화 오픈소스이다.

학습내용(Table of Contents)

- Kibana 소개
- Kibana 설치 및 환경설정
- 대시보드 커스터마이징 하기
- 이벤트 질의 하기
- 연산자 사용하기
- 시각화 하기
- 실제 예제를 통한 실습

Kibana

데이터 시각화 도구 Data Visualization Tool

검색엔진(elasticsearch) 데이터를 이용해서 시간에 따른 차트를 자동으로 그려줌

기능

키워드 검색

라인차트, 파이차트, 영역차트, 지도차트 가능

시간 선택 가능

Discover

좌측 Field목록에서 보기 원하는 항목 add 또는 remove

IP 검색

검색어는 따옴표("")로 감싼다.

"123.123.123.123"

clientip : "123.123.123.123"

URL 검색

"/order/form.html"

request : "/order/form.html"

device 검색 크롤러 제외

데이터 수집시 useragent 플러그인이 된 경우 가능

-device: "Spider"

여러 URL 검색

"/order/form.html" OR "/order/end.html"

OR 또는 AND는 대문자

실습 예제 및 수행가이드

https://dl.dropboxusercontent.com/u/2385737/Kibana-basic.pdf

https://okdevtv.com/mib/elk/kibana

학습진단 평가문제

- kibana의 데이터 저장소 역할을 하는 것은?
- 데이터를 질의해서 이벤트별 상세 정보를 볼 수 있는 메뉴는?
- 데이터를 차트로 표현할 수 있는 메뉴는?
- Discover 메뉴에서 가능한 기능은?
- 저장된 차트를 구성해서 한 눈에 현황을 볼 수 있는 메뉴는?

참고 자료

- http://elastic.co
- http://okdevtv.com/mib/elk/elk
- http://okdevtv.com/mib/elk/kibana
- http://okdevtv.com/mib/elasticsearch/elasticsearch

한국소프트웨어기술진흥협회
KOSTA Korea Software Technology Association

판교 | 경기도 성남시 분당구 삼평동 대왕판교로 670길 유스페이스2 B동 8층 T. 070-5039-5805
가산 | 서울시 금천구 가산동 371-47 이노플렉스 1차 2층 T. 070-5039-5815
웹사이트 | http://edu.kosta.or.kr **팩스 |** 070-7614-3450

고용노동부 한국산업인력공단

한국소프트웨어기술진흥협회