데이터 시각화 환경 구축하기

허광남 차장
GS SHOP 벤처투자팀

튜토리얼 목차

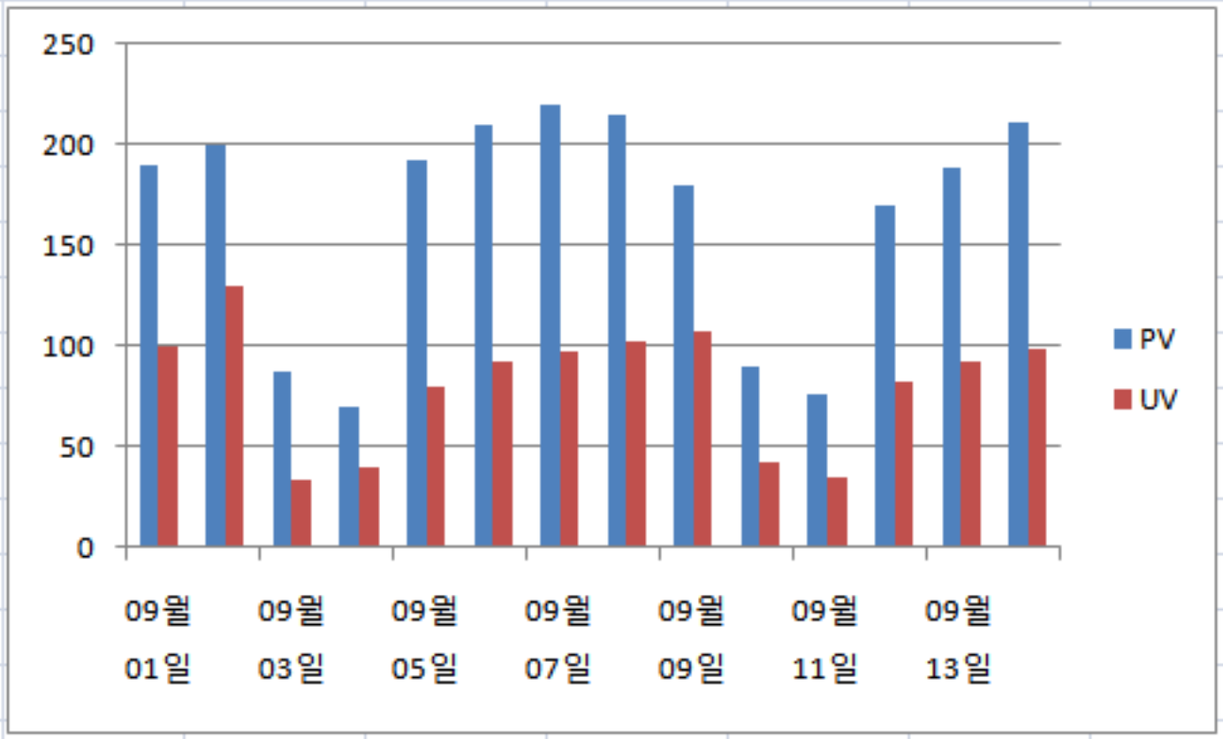
- 데이터 시각화의 중요성
- Elastic Stack
- 설치와 구성
- 로그 적재(Logstash)
- 검색 활용(Elasticsearch)
- 차트 그리기(Kibana)

데이터 시각화의 중요성

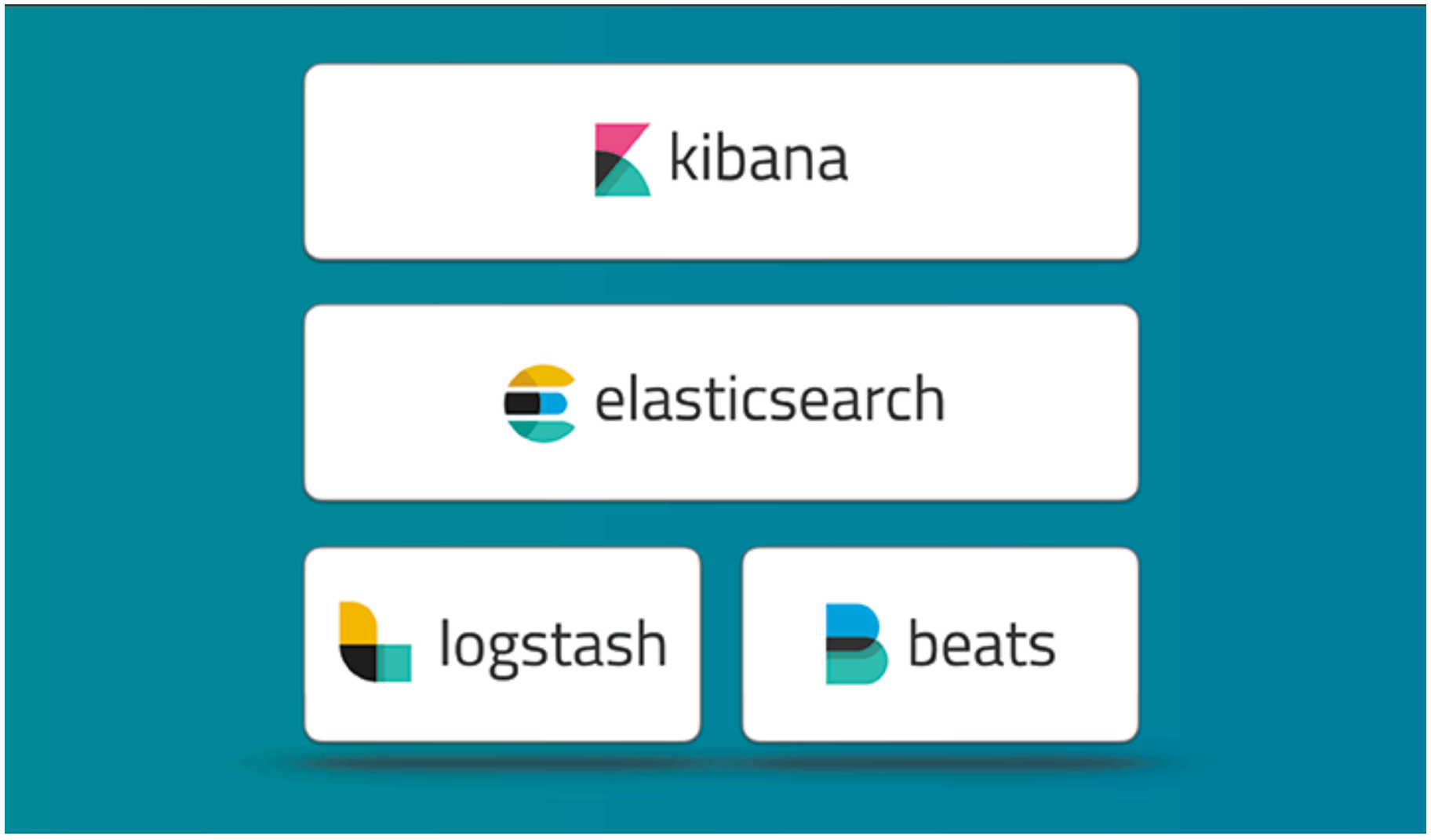
- DataViz : Data Visualization

	A	B	C
1	**날짜**	**PV**	**UV**
2	09월 01일	190	100
3	09월 02일	200	130
4	09월 03일	87	34
5	09월 04일	70	40
6	09월 05일	193	80
7	09월 06일	210	93
8	09월 07일	220	98
9	09월 08일	215	102
10	09월 09일	180	107
11	09월 10일	90	43
12	09월 11일	76	35
13	09월 12일	170	82
14	09월 13일	189	92
15	09월 14일	211	99

	A	B	C	D	E	F	G	H	I	J	K
1	날짜	PV	UV								
2	09월 01일	190	100								
3	09월 02일	200	130								
4	09월 03일	87	34								
5	09월 04일	70	40								
6	09월 05일	193	80								
7	09월 06일	210	93								
8	09월 07일	220	98								
9	09월 08일	215	102								
10	09월 09일	180	107								
11	09월 10일	90	43								
12	09월 11일	76	35								
13	09월 12일	170	82								
14	09월 13일	189	92								
15	09월 14일	211	99								



Elastic Stack



https://www.elastic.co/kr/blog/heya-elastic-stack-and-x-pack

ELK

http://elastic.co 사이트 오픈소스 제품

설치와 구성

- Elasticsearch 1+ (저장소)
- Kibana (시각화 도구)
- Logstash (로그 파서)
- Beats 1+ (로그 전송)

검색 활용(Elasticsearch)

- Shay Banon
- Lucene(http://lucene.apache.org) 라이브러리 사용
- Lucene 검색 엔진을 잘 이용한 제품
- 경쟁제품 Solr, Tica
- JSON 입출력

로그 적재(Logstash)

input {}
filter {}
output {}

```
input {
    file {
        path => "/var/log/nginx/access.log"
        start_position => beginning
    }
}
filter {
    grok {
        match => { "message" => "%{COMBINEDAPACHELOG}"}
    }
    geoip {
        source => "clientip"
    }
}
output {
    elasticsearch {}
    stdout {}
}
```

AccessLog

- 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"

AccessLog

IP	112.72.239.19
인증아이디	- -
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 log 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

차트 그리기(Kibana)

- 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

실습

- https://okdevtv.com/mib/elk/elk5

참고

- http://elastic.co
- http://okdevtv.com/mib/elk/elk5
- http://okdevtv.com/mib/elk/kibana
- http://okdevtv.com/mib/elasticsearch

감사합니다!

kenu.heo@gmail.com