

고객 행동 데이터 보면서 개발하기 Elastic Stack

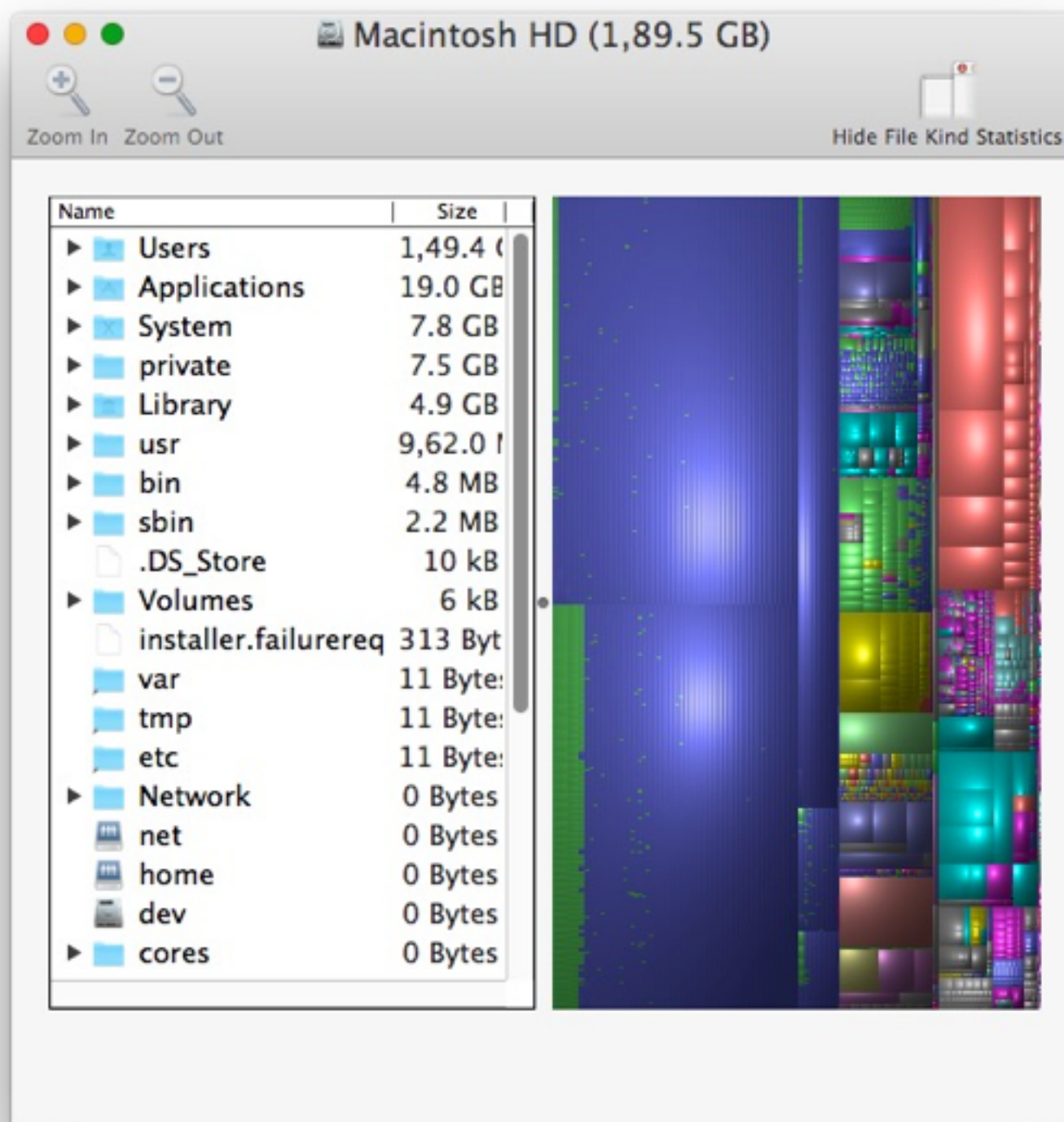
허광남

kenu@okky.kr

오늘 이야기

1. 데이터 시각화의 가치
2. 오픈소스 시각화 패키지 Elastic Stack
3. 유용한 플러그인 소개
4. ELK 적용 사례
 - AARRR 깔때기
 - A/B 테스트

데이터 시각화의 가치

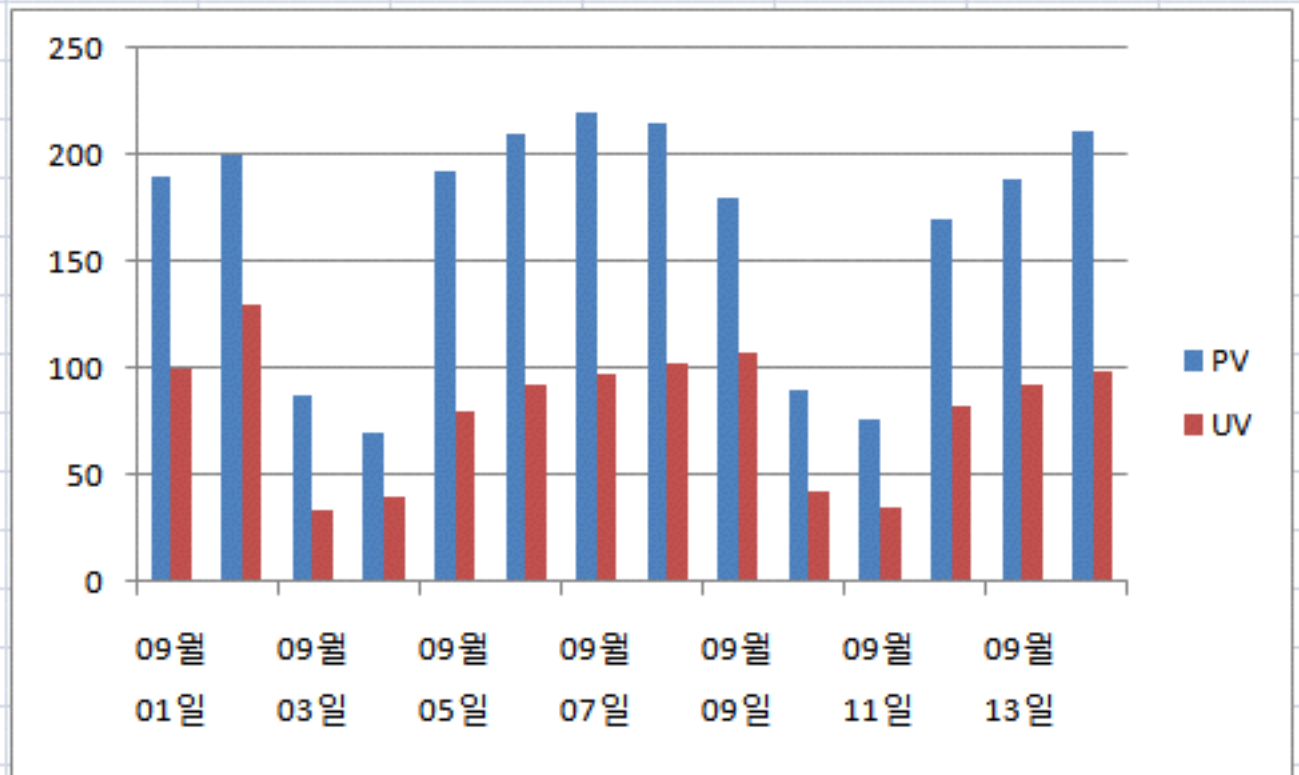


Color	Kind	Size	Files
	Preview.app Document	1,09.9 GB	67414
	Application	20.2 GB	397
	QuickTime Player.app D	13.6 GB	1373
	TextEdit.app Document	9.8 GB	42635
	Document	6.3 GB	116655
	iTunes.app Document	5.3 GB	1174
	MacBinary archive	4.1 GB	87
	iMovie Library	3.2 GB	5
	Disk Image	2.3 GB	56
	Terminal.app Document	1.9 GB	1294
	gzip compressed archiv	1.8 GB	1211
	Zip Archive	1.7 GB	120
	7-Zip Archive	1.3 GB	5
	TrueType® font collecti	9,62.5 MB	158
	CAF Audio file	8,12.4 MB	2072
	TextWrangler text docur	7,84.4 MB	24777
	GarageBandMagicMento	5,06.0 MB	2
	OS X Preference Pane	3,67.0 MB	35
	NDIF Disk Image	3,47.2 MB	1
	Mail Message Part	3,46.9 MB	424
	TrueType® font	3,27.5 MB	446
	Bill of Materials	2,90.2 MB	149
	Plain Text Document	2,52.1 MB	3287
	DAT file	2,43.3 MB	274
	Chat transcript	2,36.0 MB	1560

from: <http://osxdaily.com/2016/04/29/best-disk-storage-analyzers-mac/>

	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	
16				

	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								
16											



용어집

- ELK : Elasticsearch(검색엔진) + Logstash(로그수집기) + Kibana(시각화도구)
- GA : Google Analytics
- DataViz : 넷플릭스에서 유행하는 단어 Data Visualization

GA와 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-6707625-5', 'auto');
ga('send', 'pageview');

</script>
```


오픈소스 시각화 패키지

Elastic Stack

- ELK Stack
- Elasticsearch + Logstash + Kibana
- ELK는 Elastic에서 개발한 ElasticSearch(검색엔진), LogStash(로그 수집기), Kibana(시각화도구)로 구성된 수집, 검색, 시각화를 실시간으로 처리할 수 있는 오픈소스 패키지이다.

🔍 보고서 및 도움말 검색

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



세션

598,898

사용자

264,761

페이지뷰 수

2,931,830

세션당 페이지수

4.90

평균 세션 시간

00:04:09

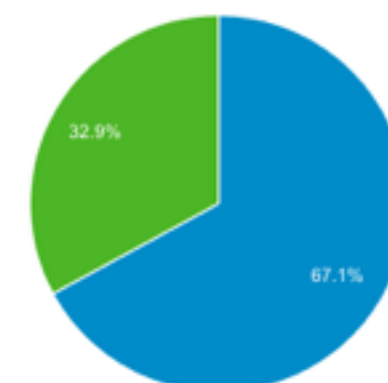
이탈률

60.73%

새로운 세션 %

32.92%

■ Returning Visitor ■ New Visitor



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

Quick

Relative

Absolute

From:

YYYY-MM-DD HH:mm:ss.SSS

< April 2016 >						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
27	28	29	30	31	01	02
03	04	05	06	07	08	09
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
01	02	03	04	05	06	07

To: [Set To Now](#)

YYYY-MM-DD HH:mm:ss.SSS

< April 2016 >						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
27	28	29	30	31	01	02
03	04	05	06	07	08	09
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
01	02	03	04	05	06	07

Go



*

logstash-*

Selected Fields

? _source

20,000

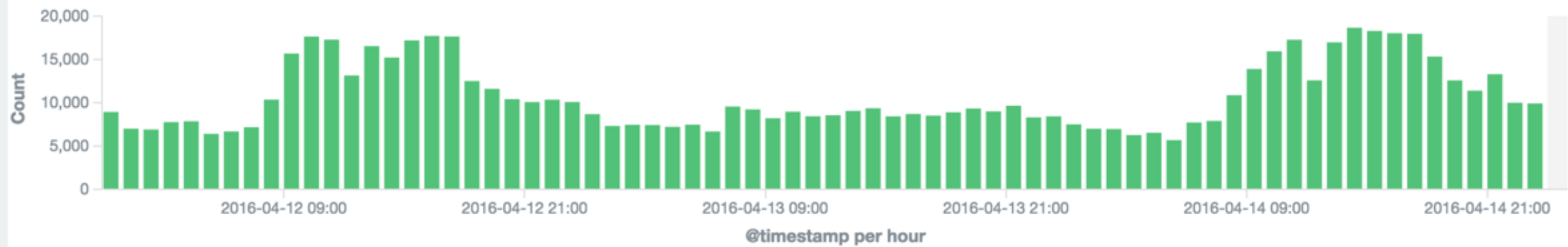
April 12th 2016, 00:00:00.000 - April 15th 2016, 00:

10	11	12	13	14	15	16	10	11	12	13	14	15	16
17	18	19	20	21	22	23	17	18	19	20	21	22	23
24	25	26	27	28	29	30	24	25	26	27	28	29	30
01	02	03	04	05	06	07	01	02	03	04	05	06	07



773,626

April 12th 2016, 00:00:00.000 - April 15th 2016, 00:00:00.000 — [by 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)

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

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"

"_"

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}

ELK

- 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

ELK





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](#)



elastic

Products

Cloud

Subscrip

A Search
Powering
Imagine T

Read the B

elasticsearch

elasticsearch as a service

logstash

kibana

beats

watcher

shield

marvel

graph

hadoop

downloads

ati

Now

Get S

Elastic Stack 특징

- Google Analytics(GA)의 데이터로 사이트 접속 통계를 구할 경우 원하는 대로 데이터를 획득하기 어렵다.
- 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.

ELK 설치, 참 쉽죠

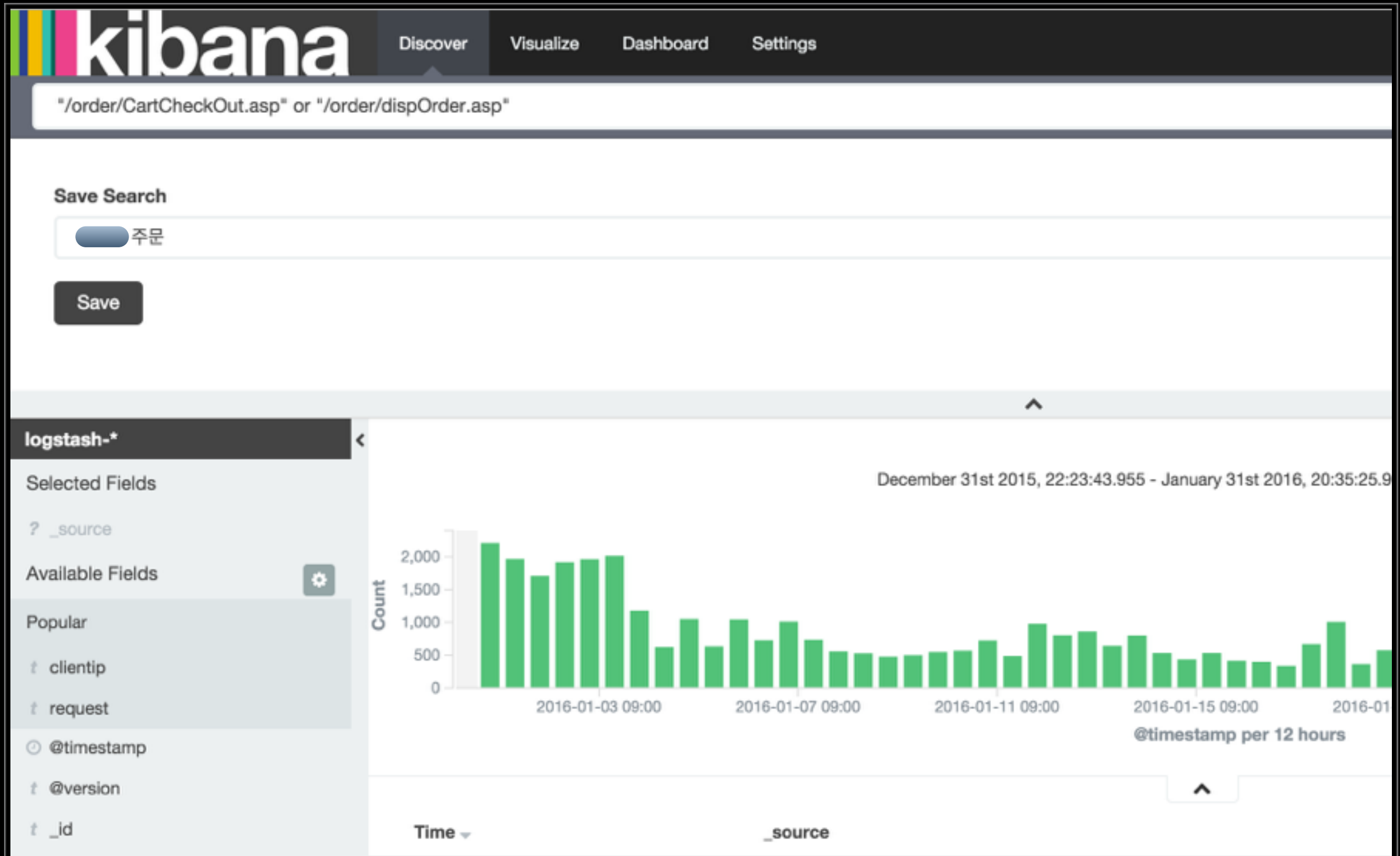
- <https://okdevtv.com/mib/elk>
- 제가 한 번 해보겠습니다.
- 사전 준비 nginx + AWS 포트 설정
- Elasticsearch 설치 + Kibana 설치 + Logstash 설치
- Logstash 설정

logstash conf


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 {}  
}
```






kibana

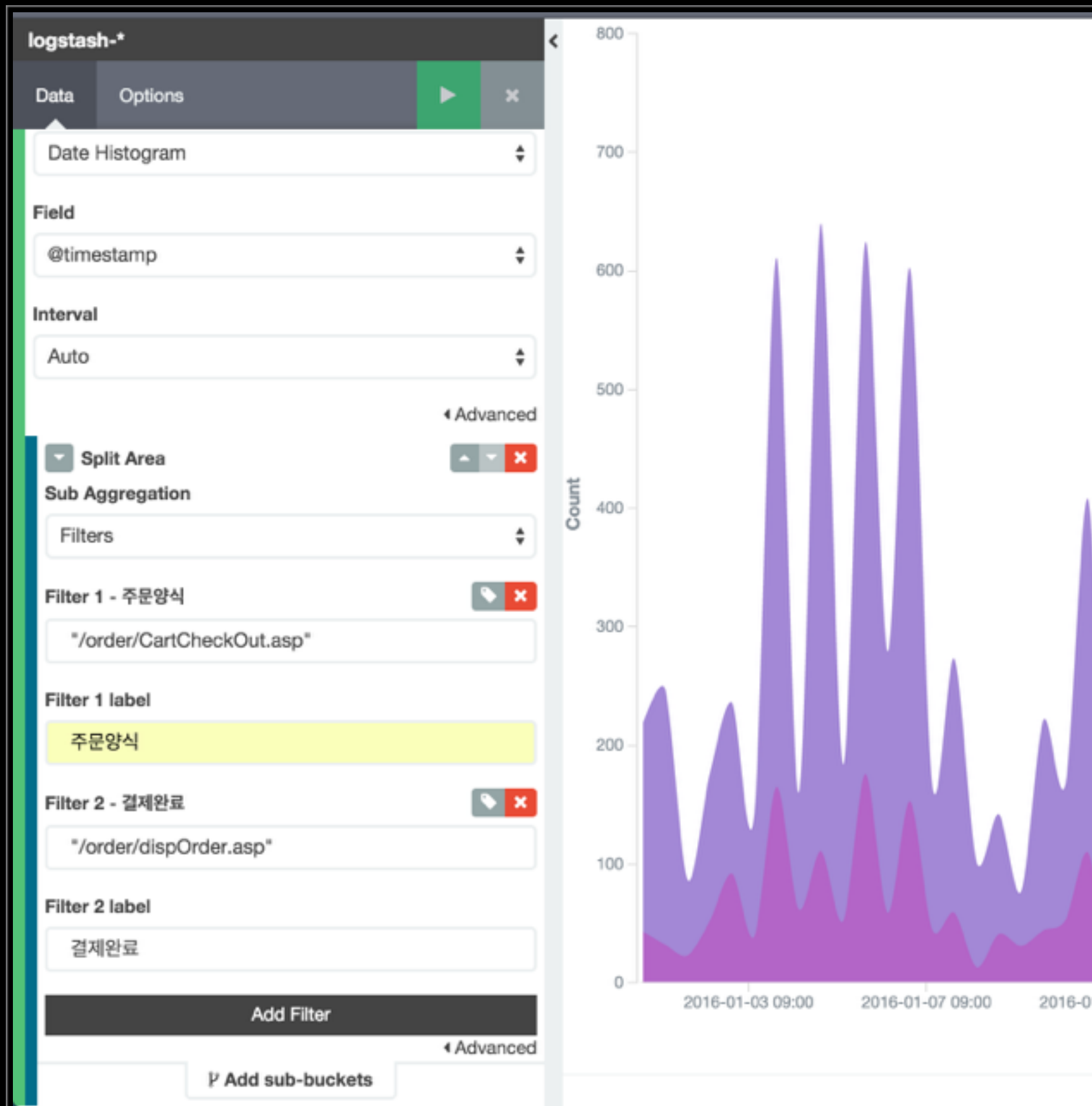


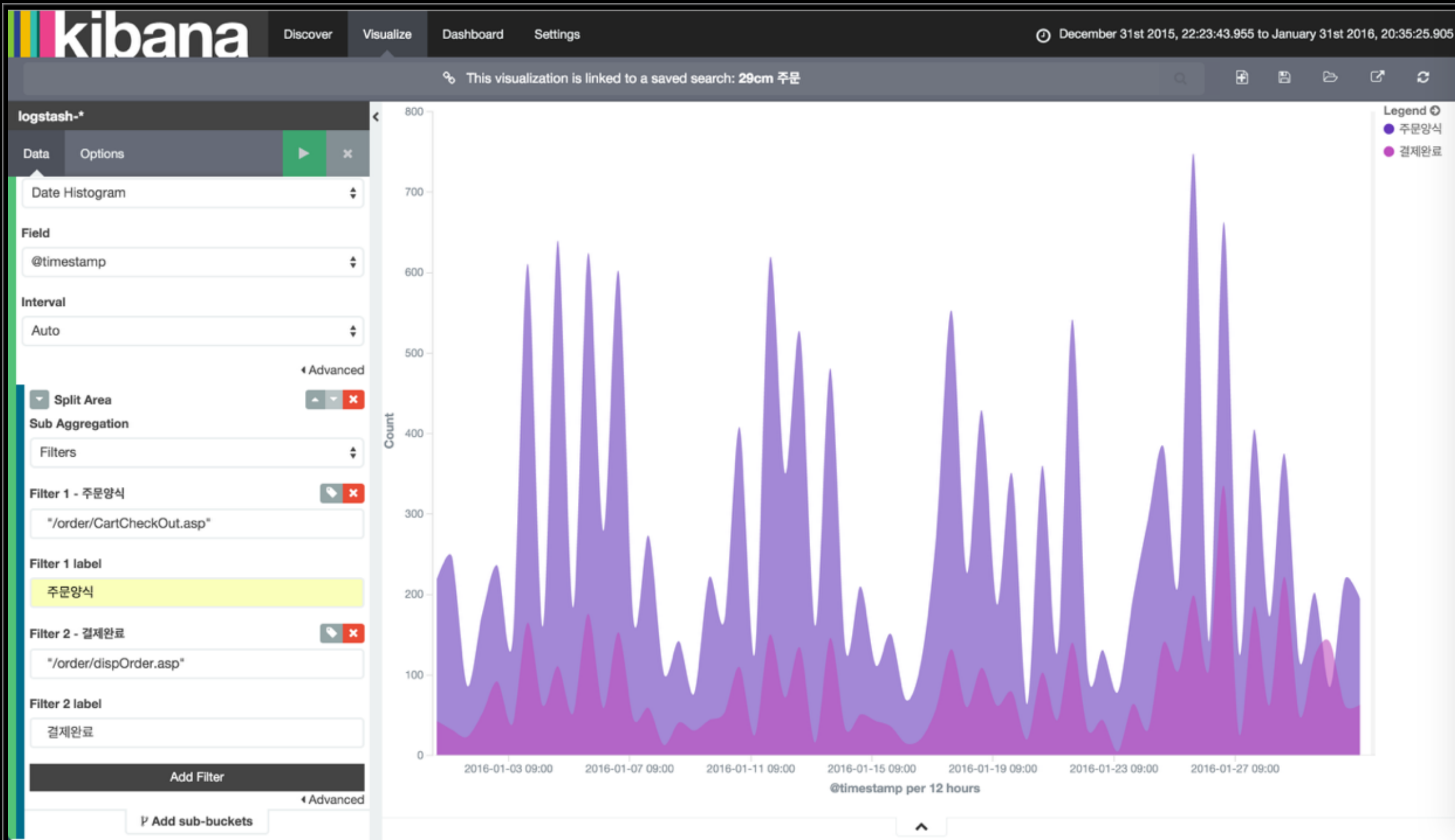
visualization

DiscoverVisualizeDashboardSettings

Create a new visualization

	Area chart	Great for stacked timelines in which the total of all series is more important than comparing change of unrelated data points as changes in a series lower down the stack will have a
	Data table	The data table provides a detailed breakdown, in tabular format, of the results of a comp charts by clicking grey bar at the bottom of the chart.
	Line chart	Often the best chart for high density time series. Great for comparing one series to another can be misleading.
	Markdown widget	Useful for displaying explanations or instructions for dashboards.
	Metric	One big number for all of your one big number needs. Perfect for showing a count of hits





filebeat

- logstash forwarder(deprecated) 의 lightweight 버전

유용한 플러그인 소개

- elasticsearch-head
- elastic-HQ

ELK 적용사례

- 스타트업
- 신규 개발
- okky.kr

Count

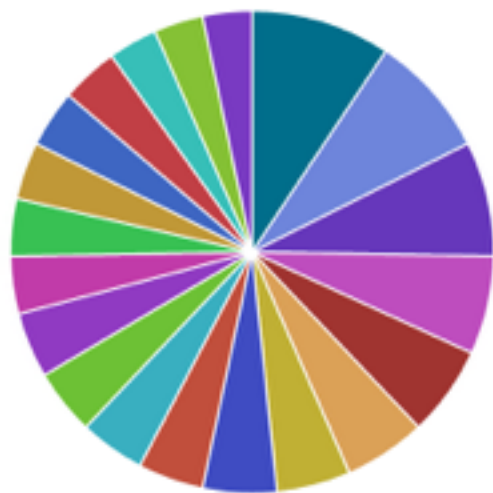
158,332

Count

16,629

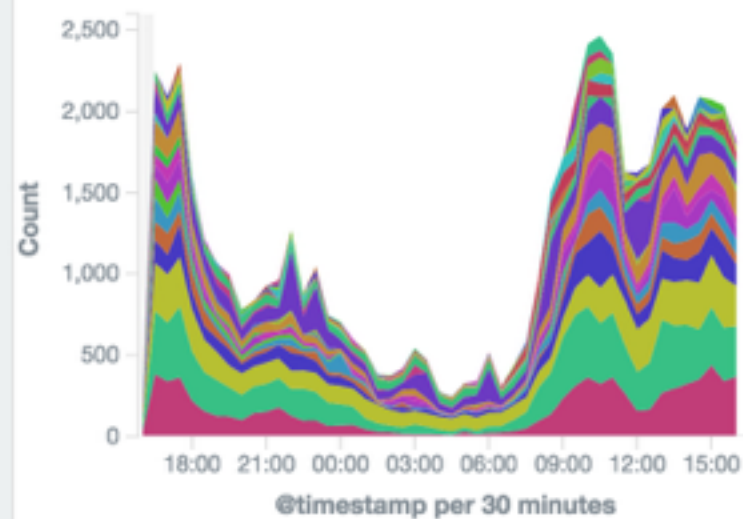
Unique count of clientip.raw

searchUniquelp



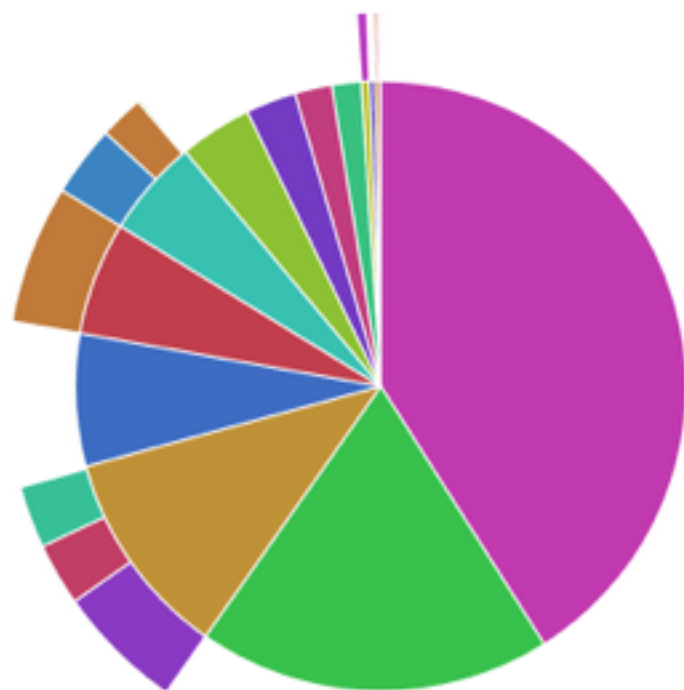
- 대전
- 대구
- 신입
- 단가
- 연봉
- 고졸+신입
- 나주
- 이직
- 노트북
- 안드로이드
- 초급
- 쿠팡
- KT
- encular

req



- /articles/community
- /
- /notification/count.json
- /articles/jobs
- /articles/recruit
- /articles/life
- /article/350098
- /css/PIE.htc
- /articles/tech
- /article/350082
- /articles/questions
- /article/350028
- /article/349636
- /article/350072

OS



- Windows 7
- Windows 10
- Android
- Other
- Mac OS X
- iOS
- Windows 8.1
- Windows
- Windows XP
- Windows 8
- Linux
- Windows Vista
- Ubuntu
- Windows 2000
- Windows ME
- BlackBerry OS
- Chrome OS
- Fedora
- Windows 98

best_article

p_no.raw: Descending Q

Count

350213	1,390
350098	1,070
350221	713
350242	687
350082	648
349966	572
350246	546
350028	487
350198	484
350193	469
350285	461

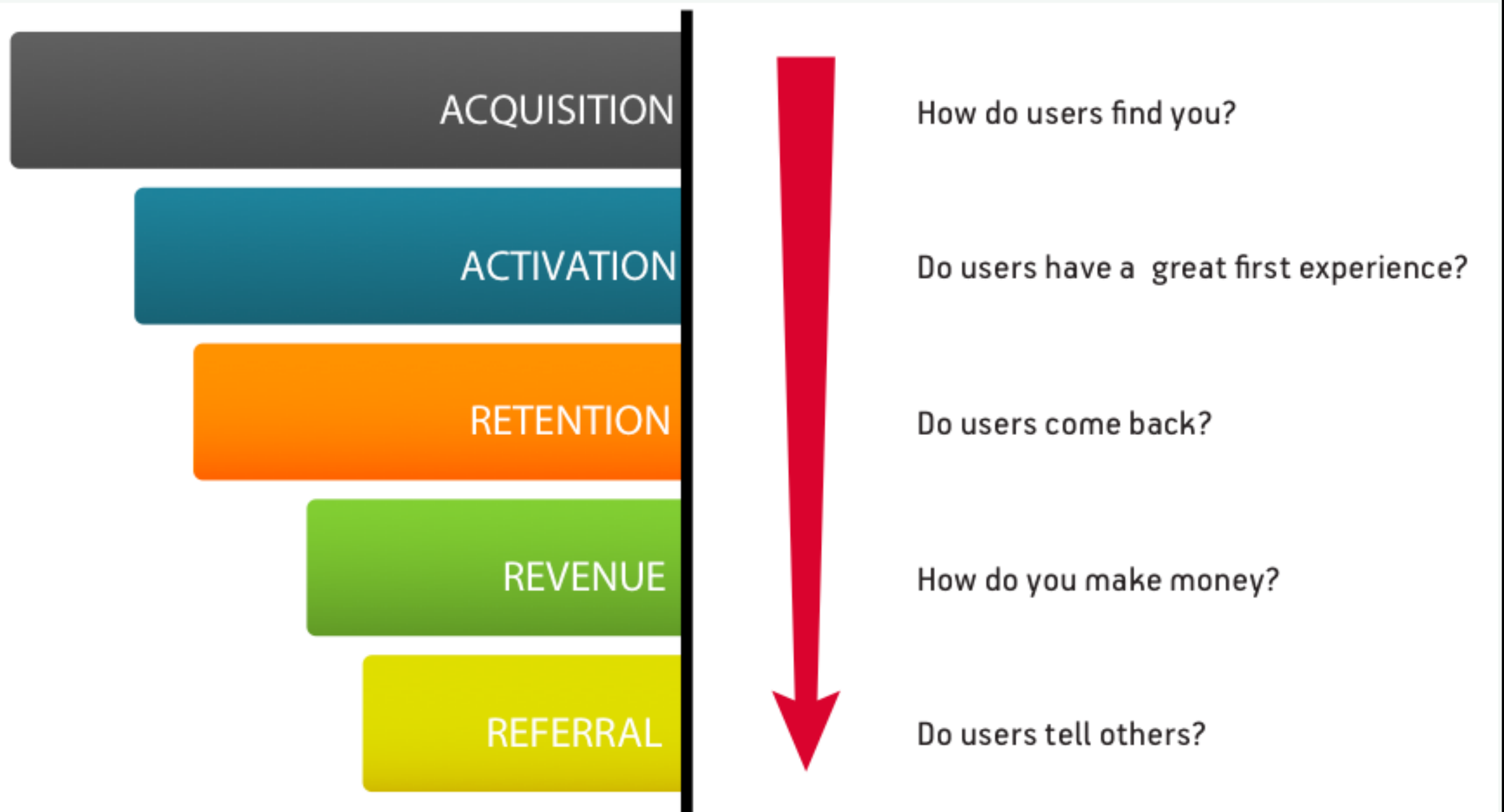
ELK를 통해서

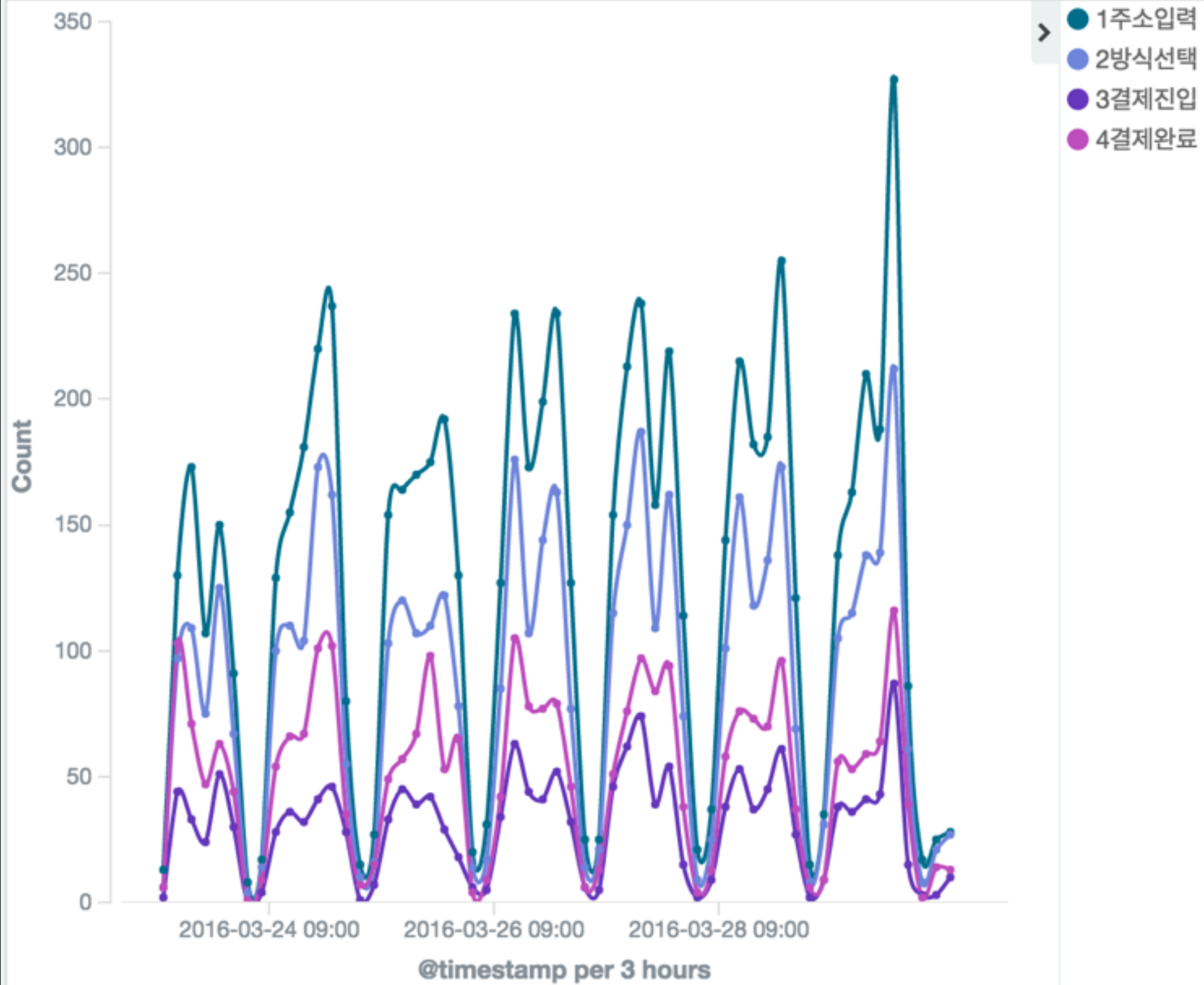
고객 액세스 트래킹이 가능

기존 막연한 마케팅 성과를
명확한 숫자로 분석할 수 있고,

A/B테스트도 자체적으로 가능

AARRR funnel





A/B Test

- 비교군을 정해서 Segmentation
- 해당 그룹에 다른 콘텐츠를 보여주고 A/B
- 반응률이 높은 쪽을 전체에 적용하는 테스트 방법

ELK를 통해서 고객 액세스 분석이 가능해졌기 때문에
기존 막연한 마케팅 성과를 명확한 숫자로 분석할 수 있고,
AB테스트도 자체적으로 실시하고 있습니다.

외부 추천 서비스, AB테스트로 검증 후 제거, 월 600
만원 절감

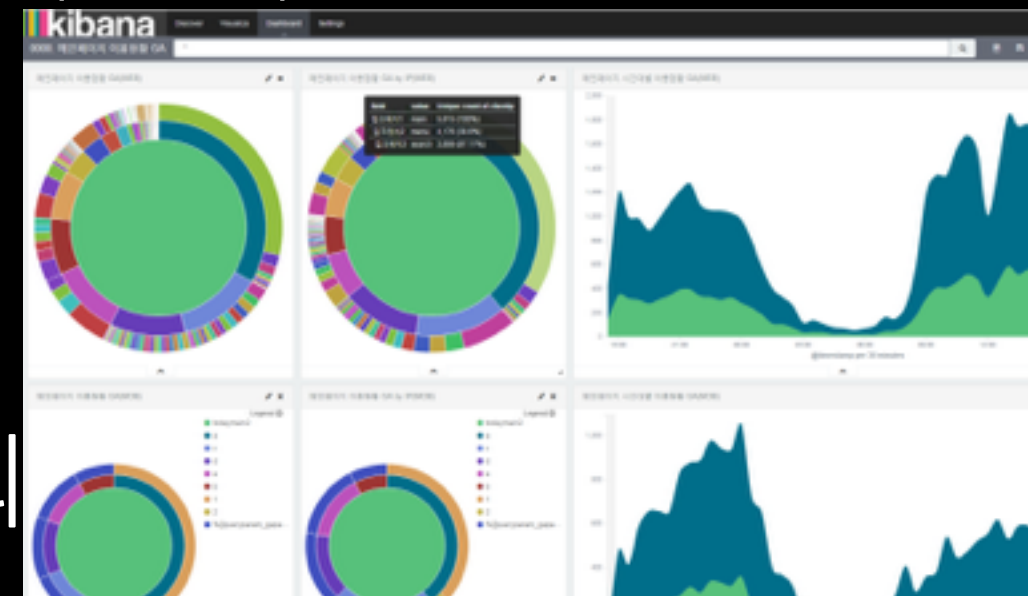
명확한 고객 활동 데이터 조회 가능(>GA)

메인 영역별 클릭율 조회 가능

웹, 모바일 모두 측정 가능

검색 키워드 랭킹 등을 기간별 조회

유입 채널 및 검색어 통계를 쉽게



참고

- EMOCON 2015 F/W ELK 스택을 사용한 서울시 지하철 대시보드 만들기
https://youtu.be/ec-XzM6_CgU
- ELKR (ElasticSearch + Logstash + Kibana + Redis) 를 이용한 로그분석 환경 구축하기
<http://brantiffy.axisj.com/archives/418>
- Splunk 대체 Solution으로서의 ELK Stack
<http://blog.embian.com/18>

참고

- Elasticsearch study
<https://okdevtv.com/mib/elasticsearch>
- ELK 초간단 설치
<https://okdevtv.com/mib/elk>
- ELK 프로그래밍 방송 영상
<http://bit.ly/okdevtv-elk>

참고

- Elasticsearch(Lucene) Query Syntax
https://lucene.apache.org/core/2_9_4/queryparsersyntax.html
- Logstash Configuration
<https://www.elastic.co/guide/en/logstash/current/event-dependent-configuration.html>
- Logstash grok patterns
<https://github.com/logstash-plugins/logstash-patterns-core/blob/master/patterns/grok-patterns>

감사합니다