Elasticsearch, Logstash & Kibana

Bessere Entscheidungen durch bessere Daten

Alexander Reelsen
@spinscale
alexander.reelsen@elasticsearch.com

Agenda

- Einführung
- Der ELK Stack
- Beispiele
- Zusammenfassung

Agenda

- Einführung
- Der ELK Stack
- Beispiele
- Zusammenfassung

- Demo

Über...

- ... Elasticsearch

  2012 gegründet

  Büros in Mountain View, Amsterdam, London, Berlin, Phoenix

  VC von Benchmark, Index Ventures & NEA

  Trainings, Development/Production Support

  Produkte: Elasticsearch, Logstash, Kibana, Marvel, Shield

Über...

- ## ... Elasticsearch

  2012 gegründet

  Büros in Mountain View, Amsterdam, London, Berlin, Phoenix

  VC von Benchmark, Index Ventures & NEA

  Trainings, Development/Production Support

  Produkte: Elasticsearch, Logstash, Kibana, Marvel, Shield

- ## ... mich

  seit Anfang 2013 bei Elasticsearch

  interessiert an Skalierung & Concurrency

  Core/Shield Entwickler, Blogger, Trainer, Supporter, Speaker

Einführung

Entscheiden, aber wie?

- Limitierte Ressourcen

  Budget, Zeit, Mitarbeiter, Kunden…

- Entscheidungsfindung

  Logdateien

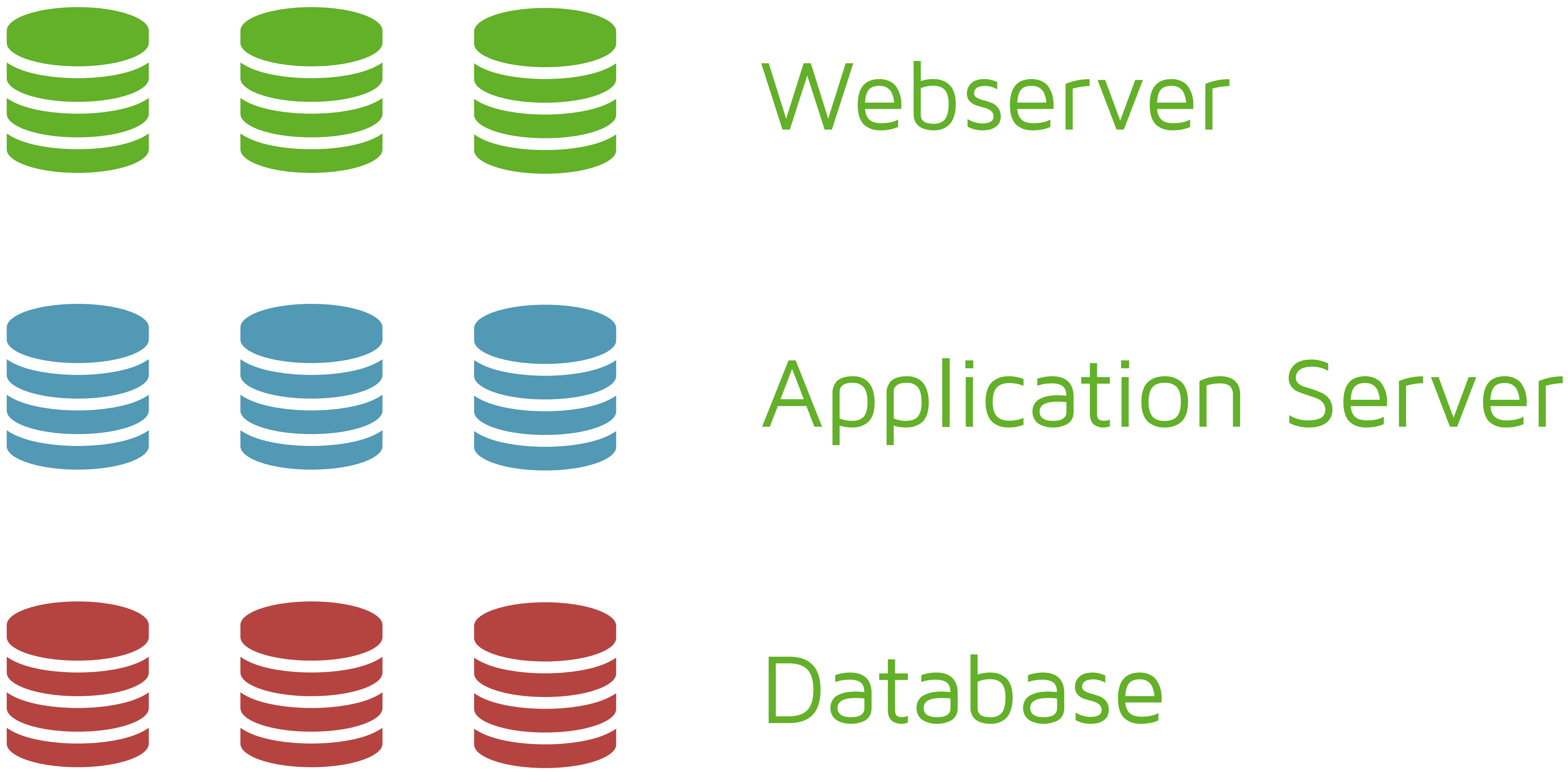
  Hochregallager

  Kreditkartenbetrug

  Fließband

Daten sind verteilt

- über Systeme



Webserver

Application Server

Database

Daten sind verteilt

- über Abteilungen

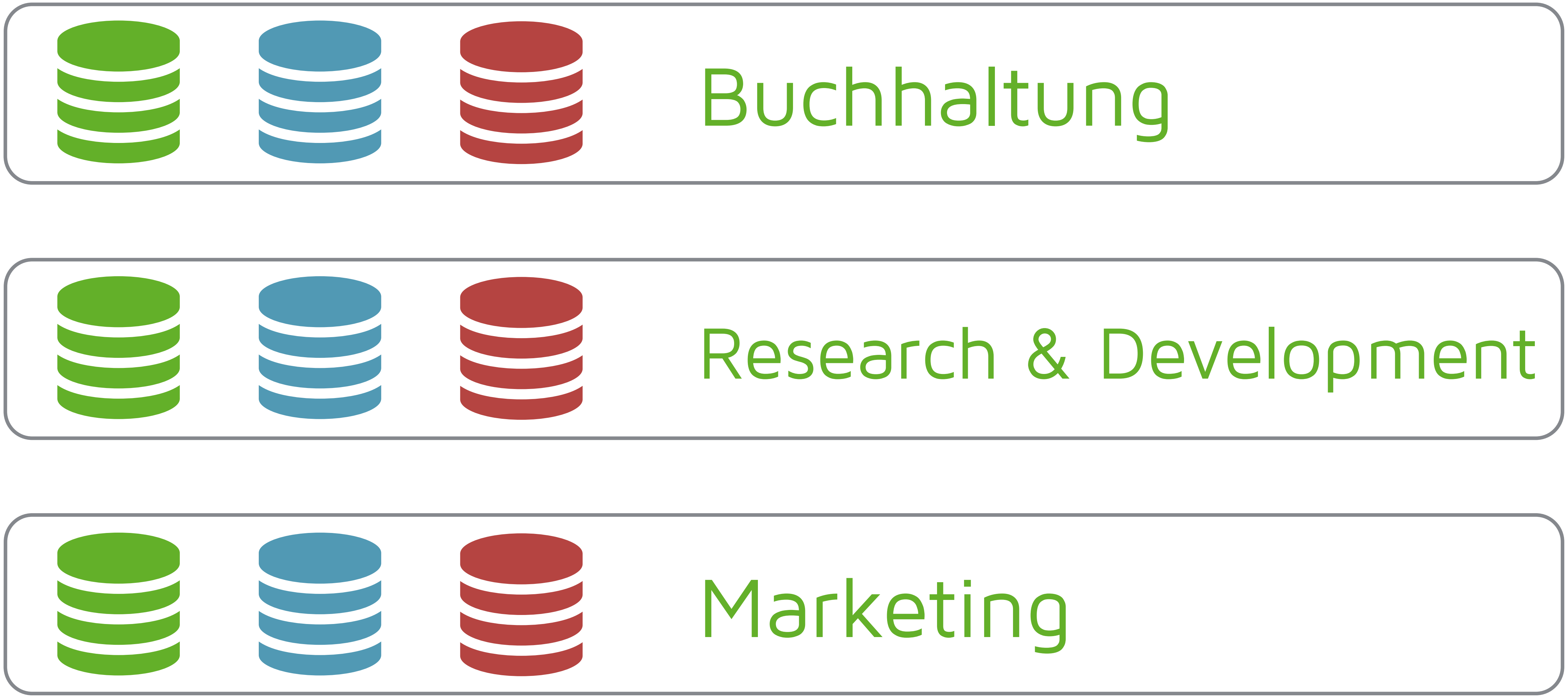
Buchhaltung

Research & Development

Marketing

Daten sind verteilt

- Silos



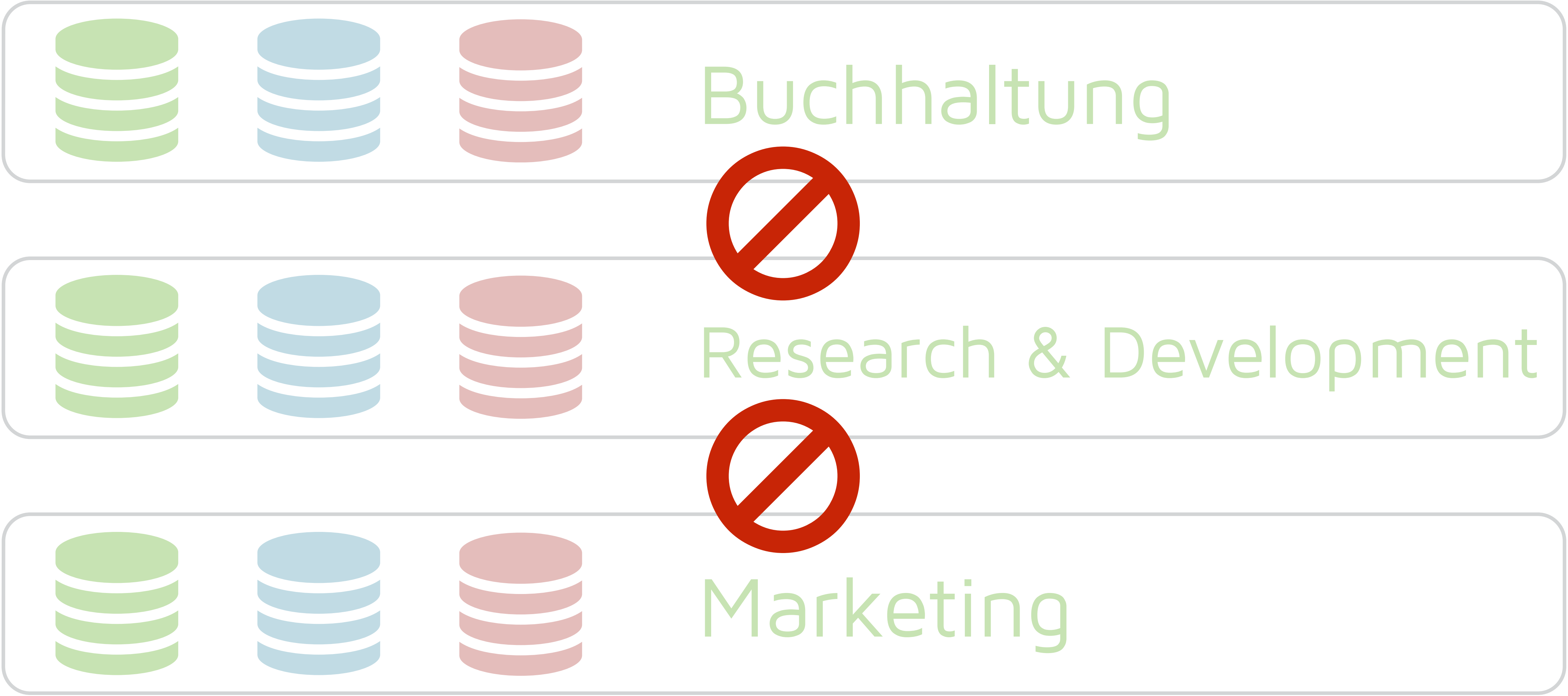
Buchhaltung

Research & Development

Marketing

Daten sind verteilt

- Keine Kommunikation



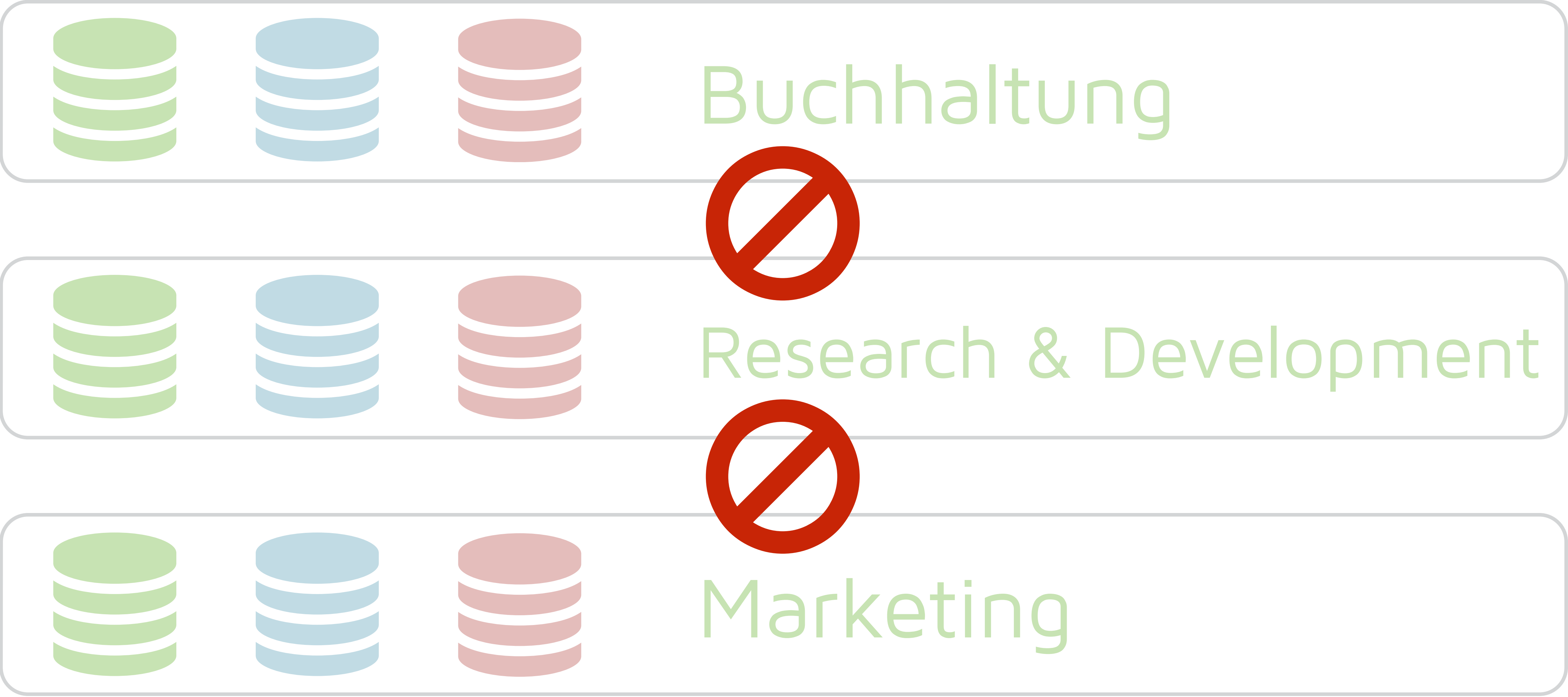
Buchhaltung

Research & Development

Marketing

Daten sind verteilt

- Keine Korrelation



Buchhaltung

Research & Development

Marketing

Daten haben einen Wert

Reaktion

Anreicherung

Information

Daten ändern ihren Wert

Reaktion

Anreicherung

Information

- Dauer von Ereignis bis Auswertung

Daten ändern ihren Wert

Reaktion

Anreicherung

Information

- Mehrwert durch Anreicherung

Daten ändern ihren Wert

Reaktion

Anreicherung

Information

- Optimieren für Extraktion

  Speicheroptimierung zweirangig!

Anforderungen

Struktur

Echtzeit

UI

Anforderungen

Struktur

Echtzeit

UI

- Datenaufbereitung

Anforderungen

Struktur

Echtzeit

UI

- Zeitnahe Bereitstellung der Daten

Anforderungen

Struktur

Echtzeit

UI

- Visualisierung

ELK Stack

ELK Stack

Logstash

Elasticsearch

Kibana

ELK Stack

Logstash

Elasticsearch

Kibana

- Annahme & Konvertierung
- Anreicherung
- Weitergabe

ELK Stack

Logstash

Elasticsearch

Kibana

- Speicherung

ELK Stack

Logstash

Elasticsearch

Kibana

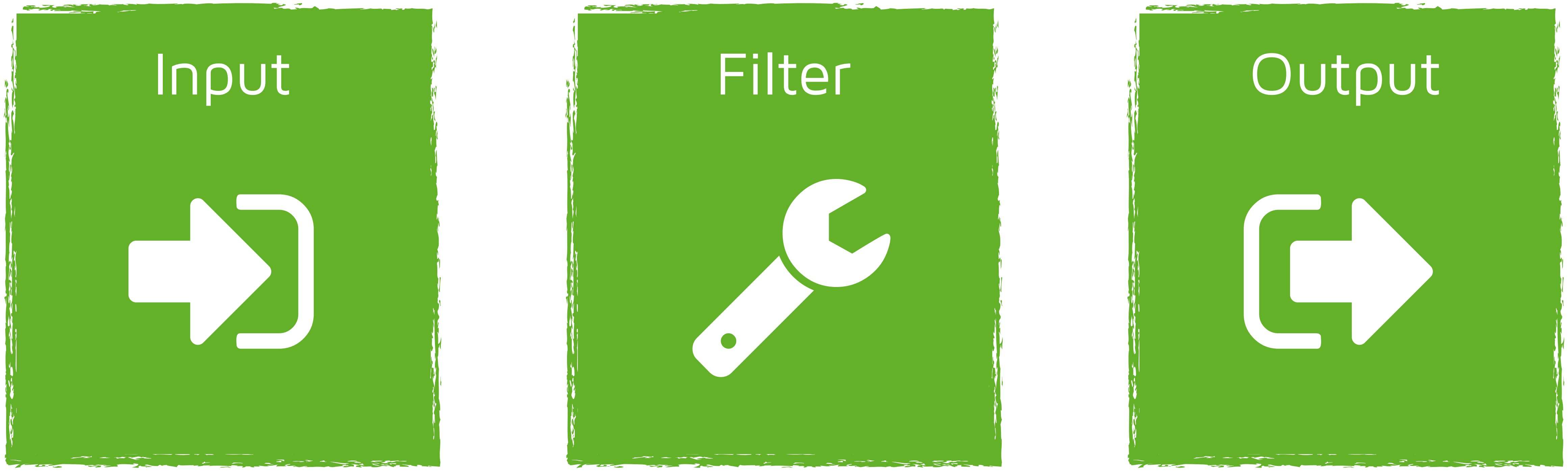
- Anfrage

Logstash

Logstash

- Events and Logs verwalten
- Daten sammeln
- Daten auswerten
- Daten anreichern
- Daten speichern
- Open Source: Apache License 2.0

Architektur

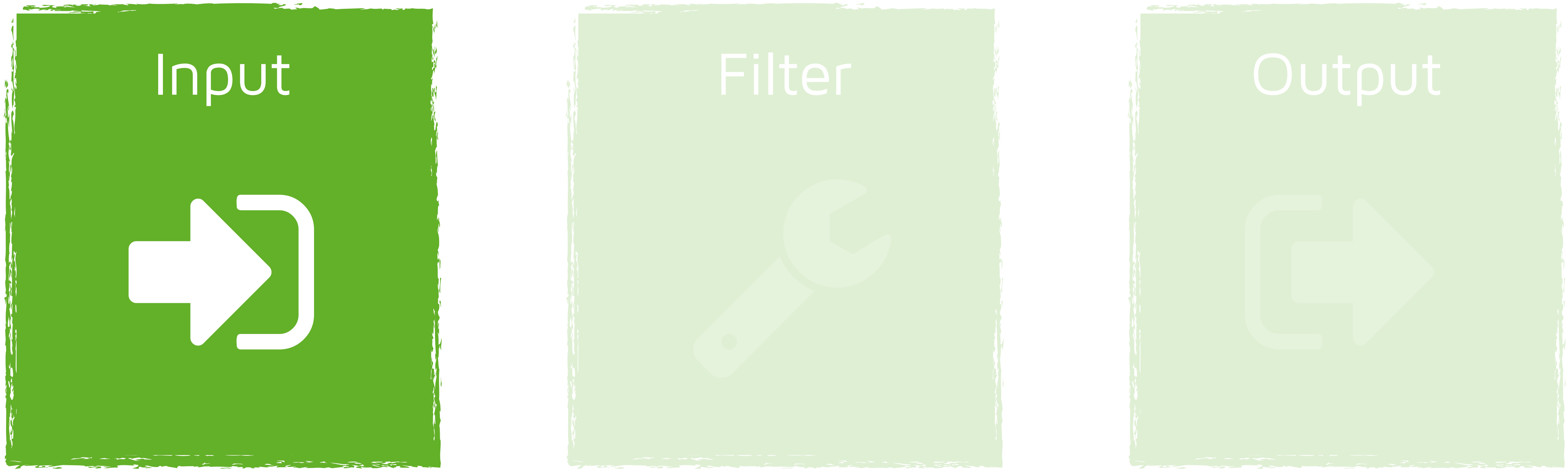


Input

Filter

Output

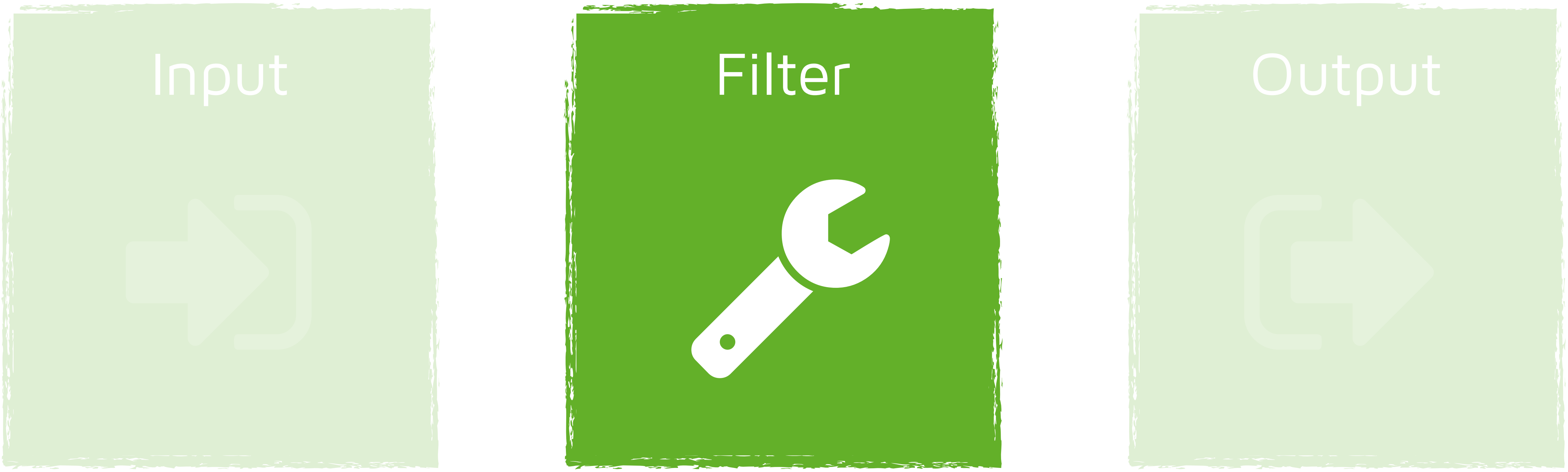
Inputs



- Lesen des Ereignisses

  datastore, stream, log files, files, monitoring, queues, network

Filter


Input

Filter

Output

- Ändern des Ereignisses

  parse, enrich, tag, drop

Output

Input
Filter
Output

- Schreiben des Ereignisses

  datastore, files, email, pager, monitoring, chat, API, queues

It's a pipe!

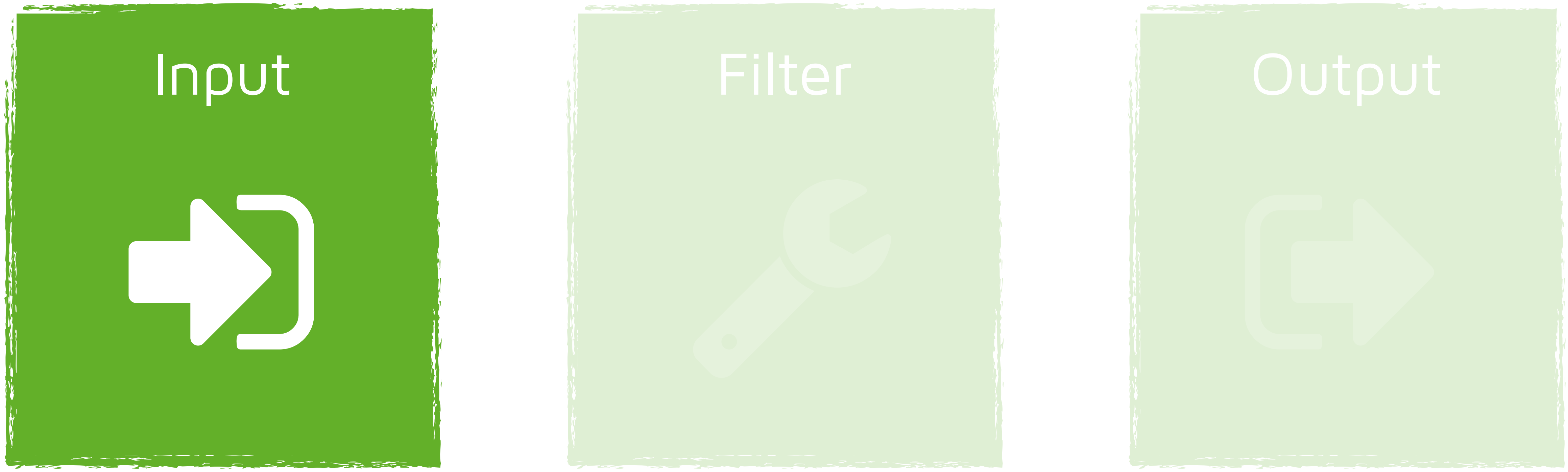
Input

Filter

Output

- Unix pipe?

Beispiel: GeoIP Filter



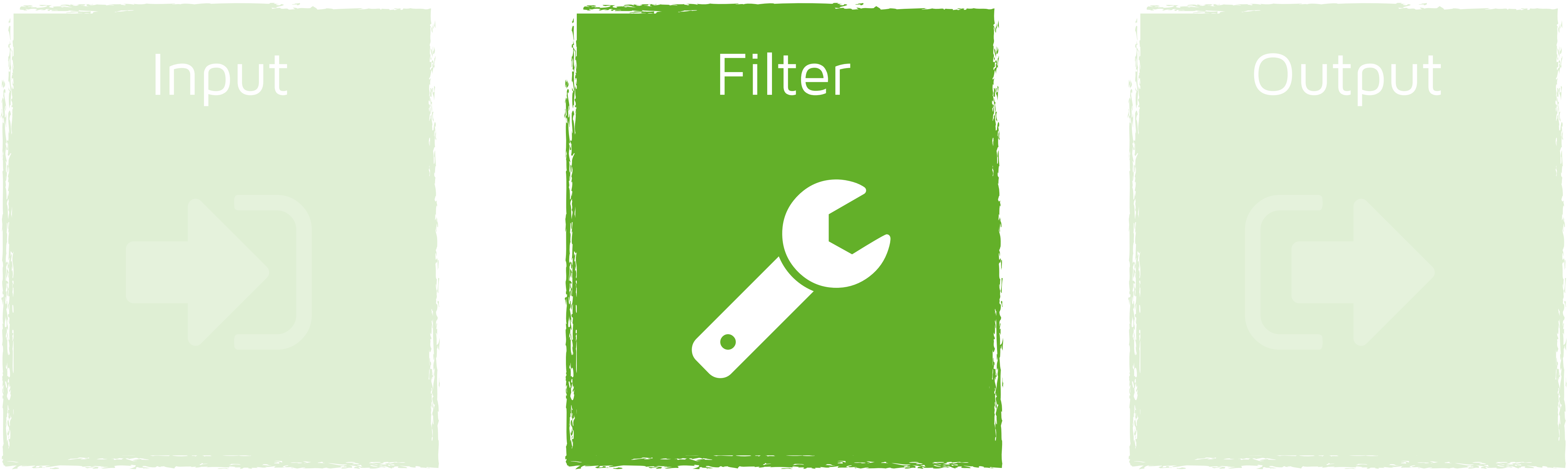
Input

Filter

Output

- Input stdin

Beispiel: GeoIP Filter



Input

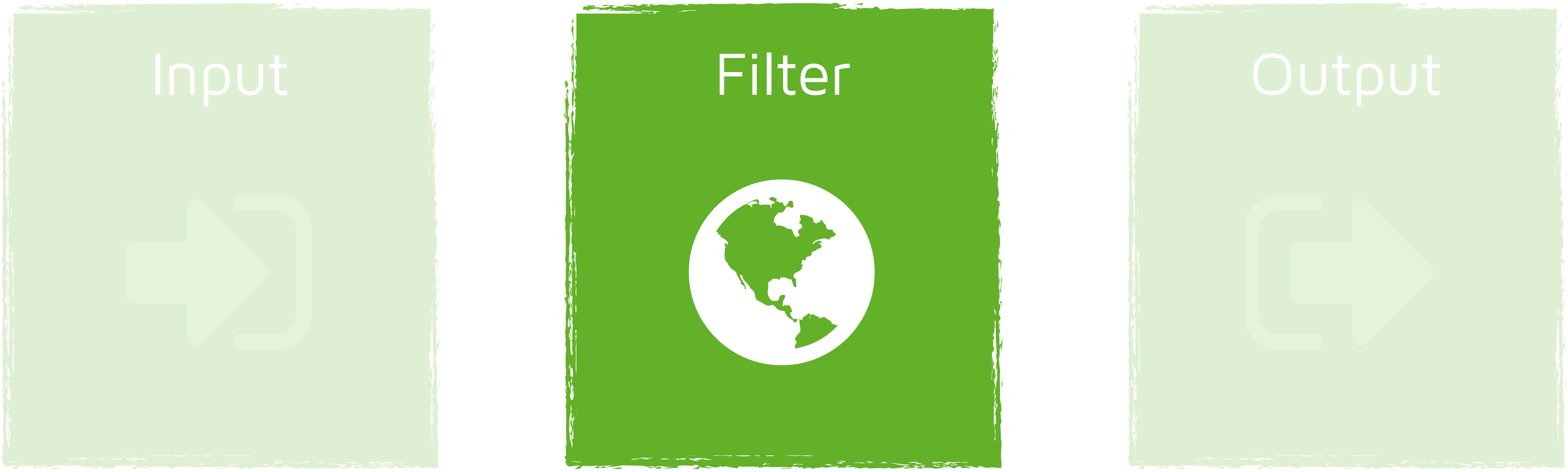
Filter

Output

- Bearbeiten des Ereignisses

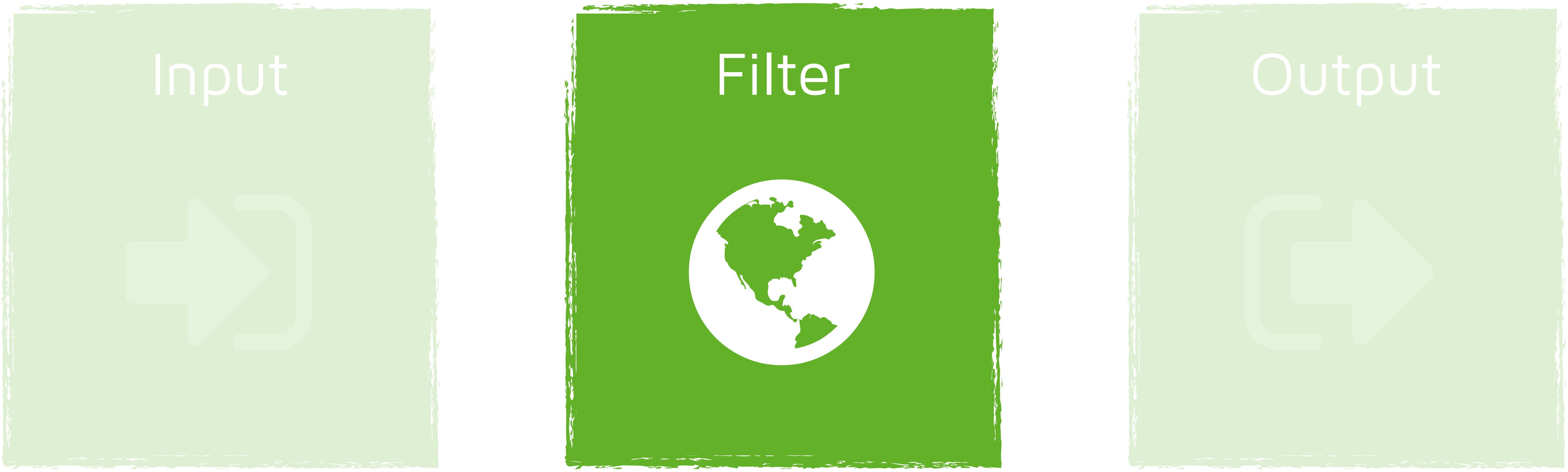
  grok, date, useragent, geoip

Beispiel: GeoIP Filter



- ip: 141.1.1.1

Beispiel: GeoIP Filter



- ip: 141.1.1.1
- city: Munich
- country: GER

Beispiel: GeoIP Filter

Input
Filter
Output

- Speichern

  elasticsearch

Elasticsearch

Was ist Elasticsearch?

- HTTP & JSON

Was ist Elasticsearch?

- HTTP & JSON
- Schema-less

Was ist Elasticsearch?

- HTTP & JSON
- Schema-less
- distributed

Was ist Elasticsearch?

- HTTP & JSON
- Schema-less
- distributed
- document-oriented

Was ist Elasticsearch?

- HTTP & JSON
- Schema-less
- distributed
- document-oriented
- near-realtime

Was ist Elasticsearch?

- HTTP & JSON
- Schema-less
- distributed
- document-oriented
- near-realtime
- search

Was ist Elasticsearch?

- HTTP & JSON
- Schema-less
- distributed
- document-oriented
- near-realtime
- search
- analytics

Use-Cases

Search

Logging

Analytics

Use-Cases

Search

Logging

Analytics

- Volltextsuche

  Dokumente, Produkte, Quellcode

Use-Cases

Search

Logging

Analytics

- Zentralisierung & Suchbarkeit

  System, Webserver, Audit, Firewall

Use-Cases

Search

Logging

Analytics

- **Aggregation**

  Reporting, BI, Facetted Navigation

Use-Cases: Kombination

Search & Logging & Analytics

Kibana

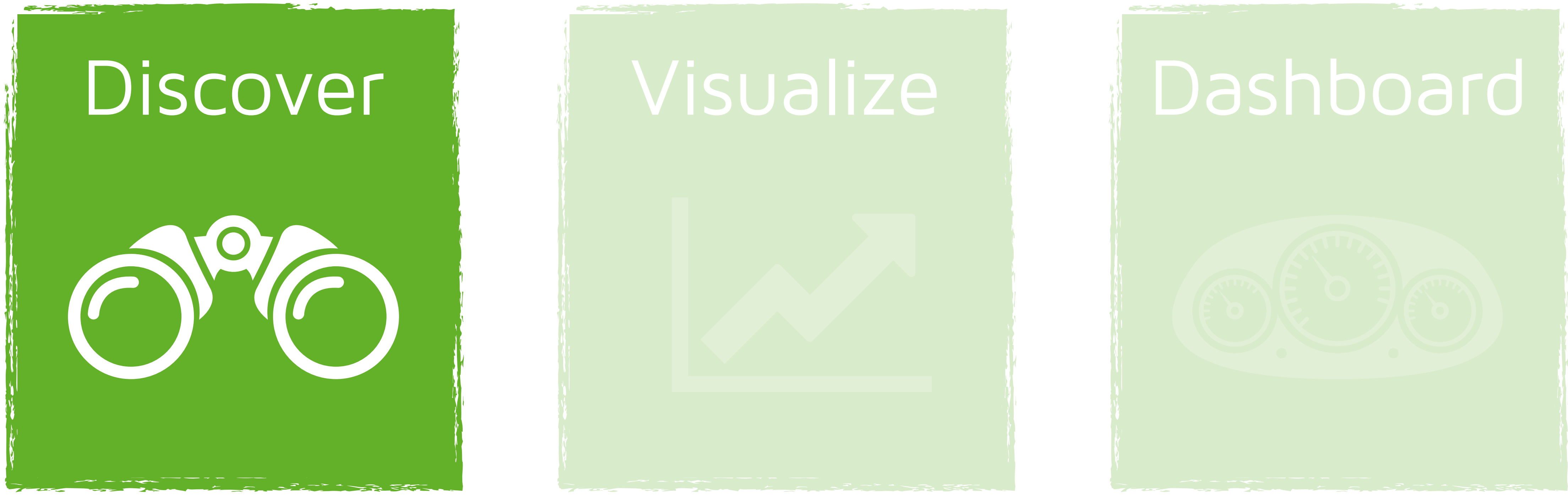
Kibana

Discover

Visualize

Dashboard

Kibana



Discover

Visualize

Dashboard

- Interaktive Datenexploration
- Ad-Hoc Suchanfragen

Kibana

Discover

Visualize

Dashboard

- Visualisierungen und Charts

Kibana

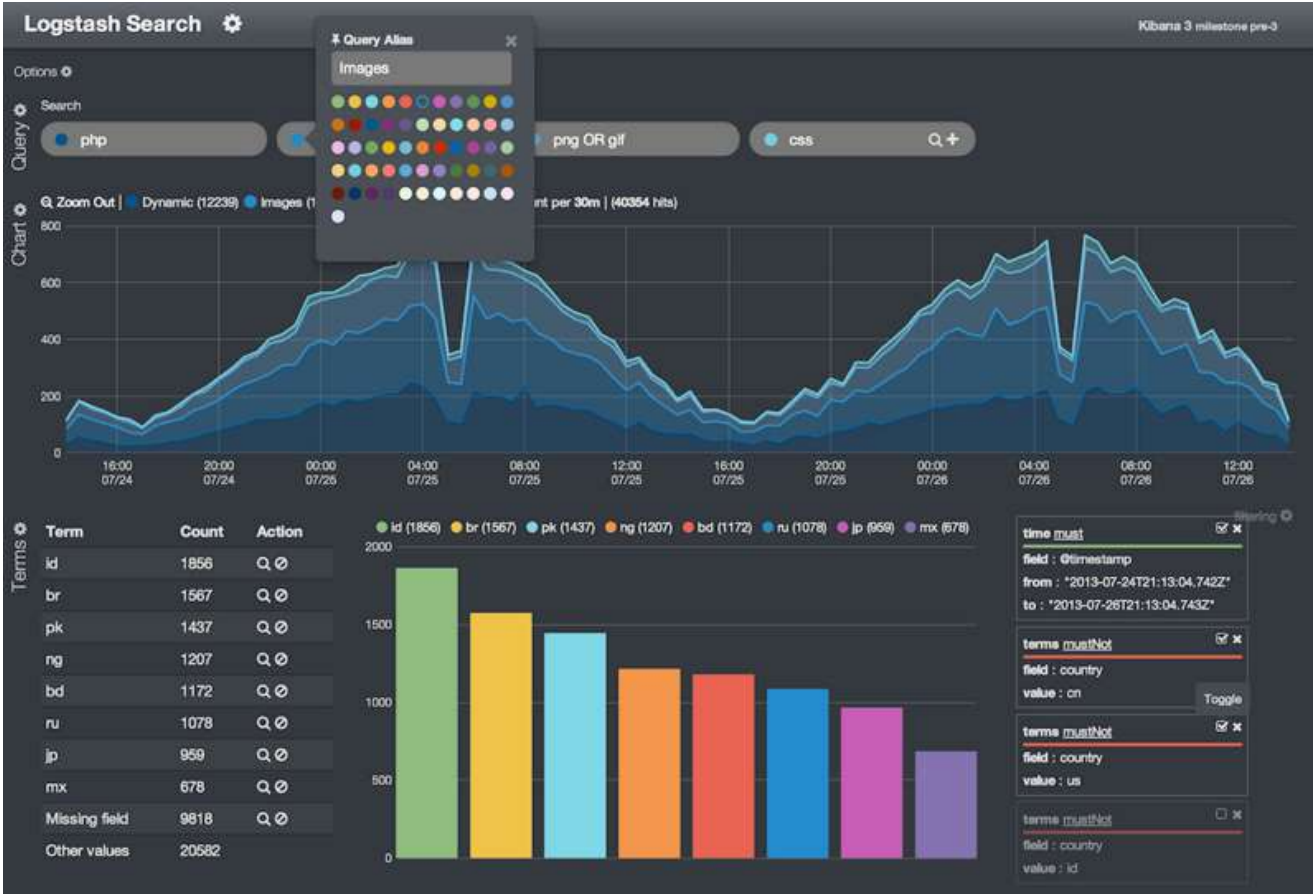
Discover

Visualize

Dashboard

- Flexible, interaktive Dashboards

Kibana



Term	Count	Action
id	1856	🔍 ⊘
br	1567	🔍 ⊘
pk	1437	🔍 ⊘
ng	1207	🔍 ⊘
bd	1172	🔍 ⊘
ru	1078	🔍 ⊘
jp	959	🔍 ⊘
mx	678	🔍 ⊘
Missing field	9818	🔍 ⊘
Other values	20582	

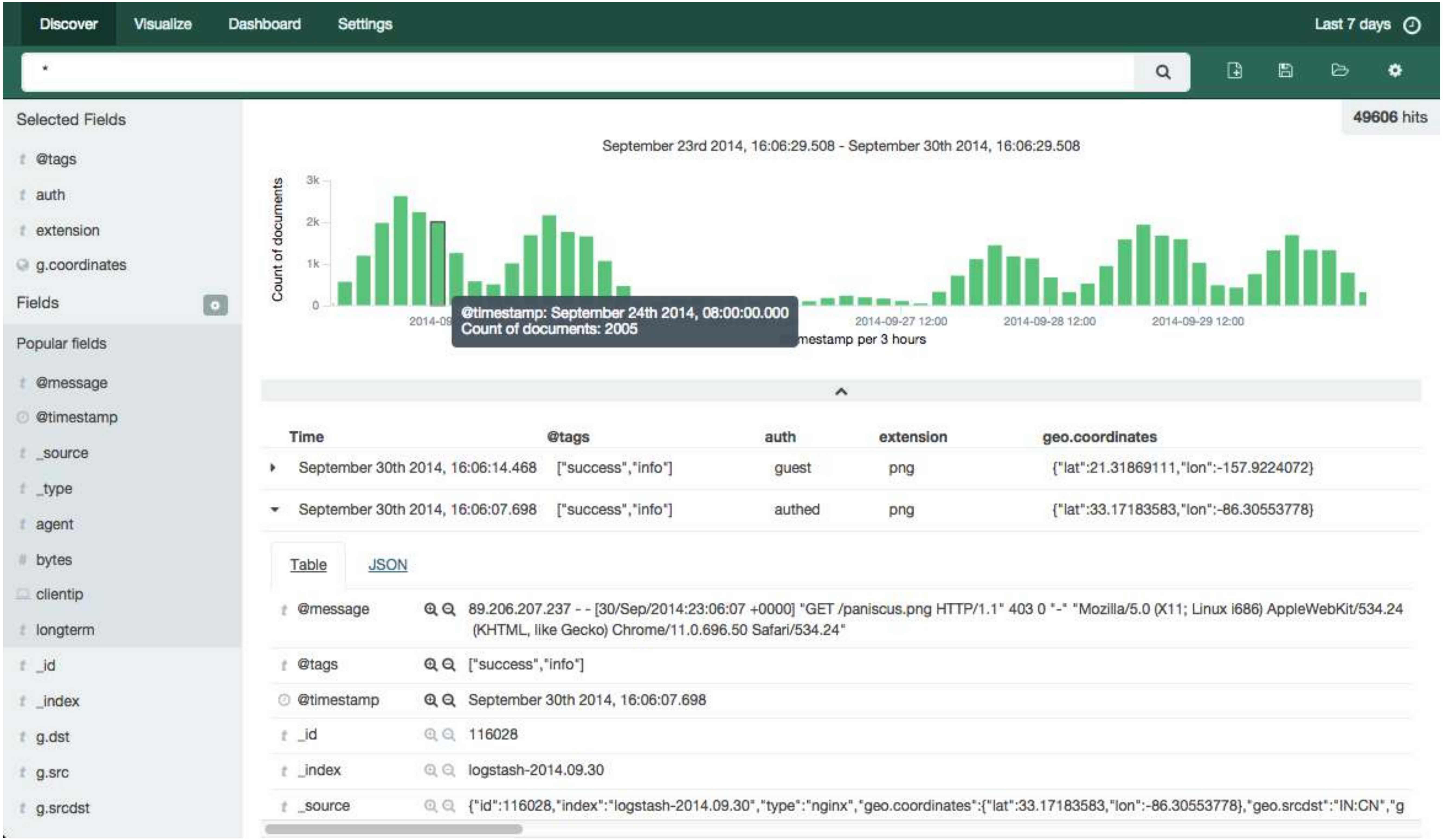
time must ☑ ✖
field : @timestamp
from : "2013-07-24T21:13:04.742Z"
to : "2013-07-26T21:13:04.743Z"

terms mustNot ☑ ✖
field : country
value : cn

terms mustNot ☑ ✖
field : country
value : us

terms mustNot □ ✖
field : country
value : id

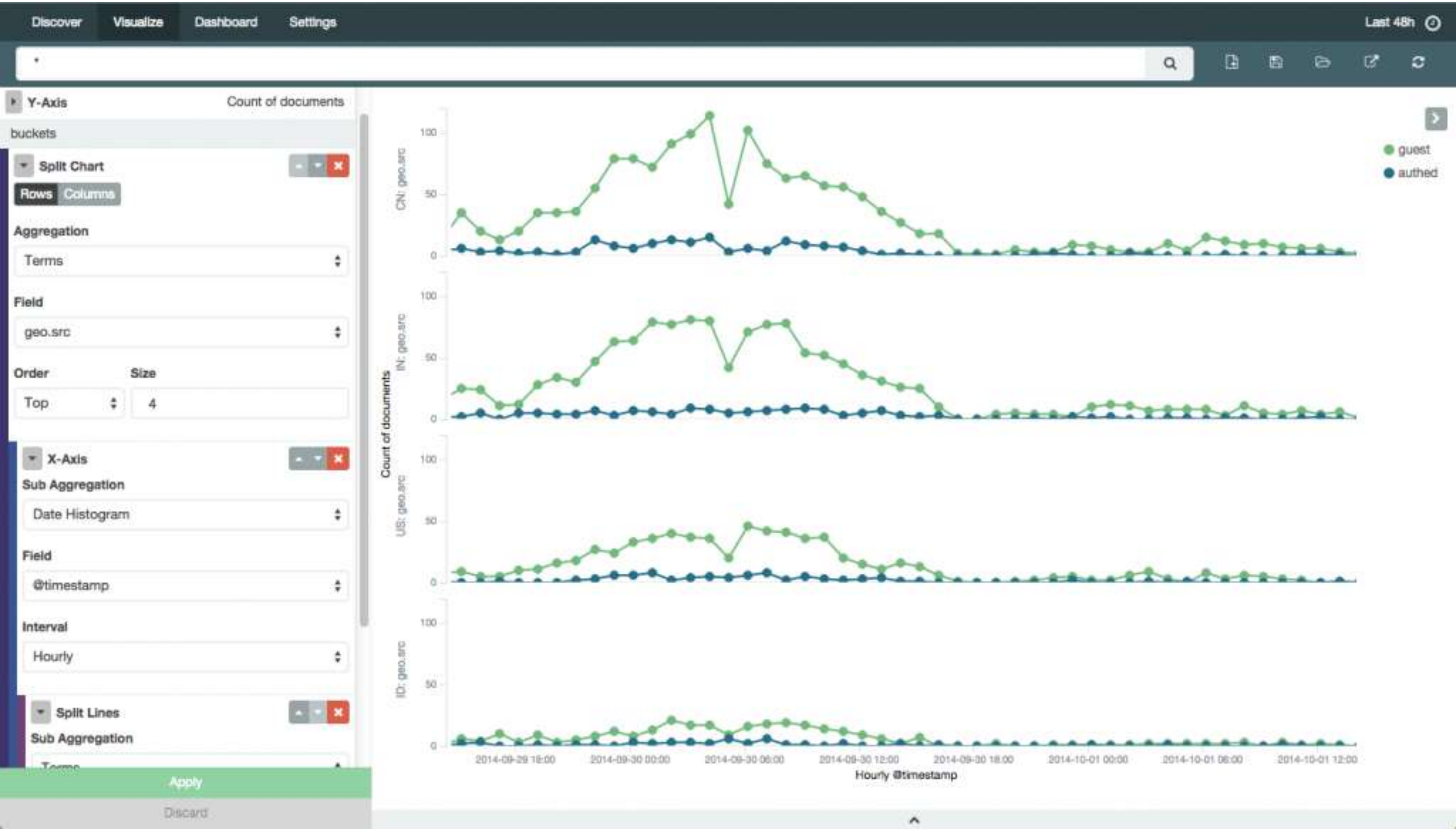
Kibana 4



Time	@tags	auth	extension	geo.coordinates
September 30th 2014, 16:06:14.468	["success","info"]	guest	png	{"lat":21.31869111,"lon":-157.9224072}
September 30th 2014, 16:06:07.698	["success","info"]	authed	png	{"lat":33.17183583,"lon":-86.30553778}

@message	89.206.207.237 - - [30/Sep/2014:23:06:07 +0000] "GET /paniscus.png HTTP/1.1" 403 0 "-" "Mozilla/5.0 (X11; Linux i686) AppleWebKit/534.24 (KHTML, like Gecko) Chrome/11.0.696.50 Safari/534.24"
@tags	["success","info"]
@timestamp	September 30th 2014, 16:06:07.698
_id	116028
_index	logstash-2014.09.30
_source	{"id":116028,"index":"logstash-2014.09.30","type":"nginx","geo.coordinates":{"lat":33.17183583,"lon":-86.30553778},"geo.srcdst":"IN:CN","g

Kibana 4



Y-Axis Count of documents

buckets

Split Chart

Rows Columns

Aggregation

Terms

Field

geo.src

Order Top

Size 4

X-Axis

Sub Aggregation

Date Histogram

Field

@timestamp

Interval

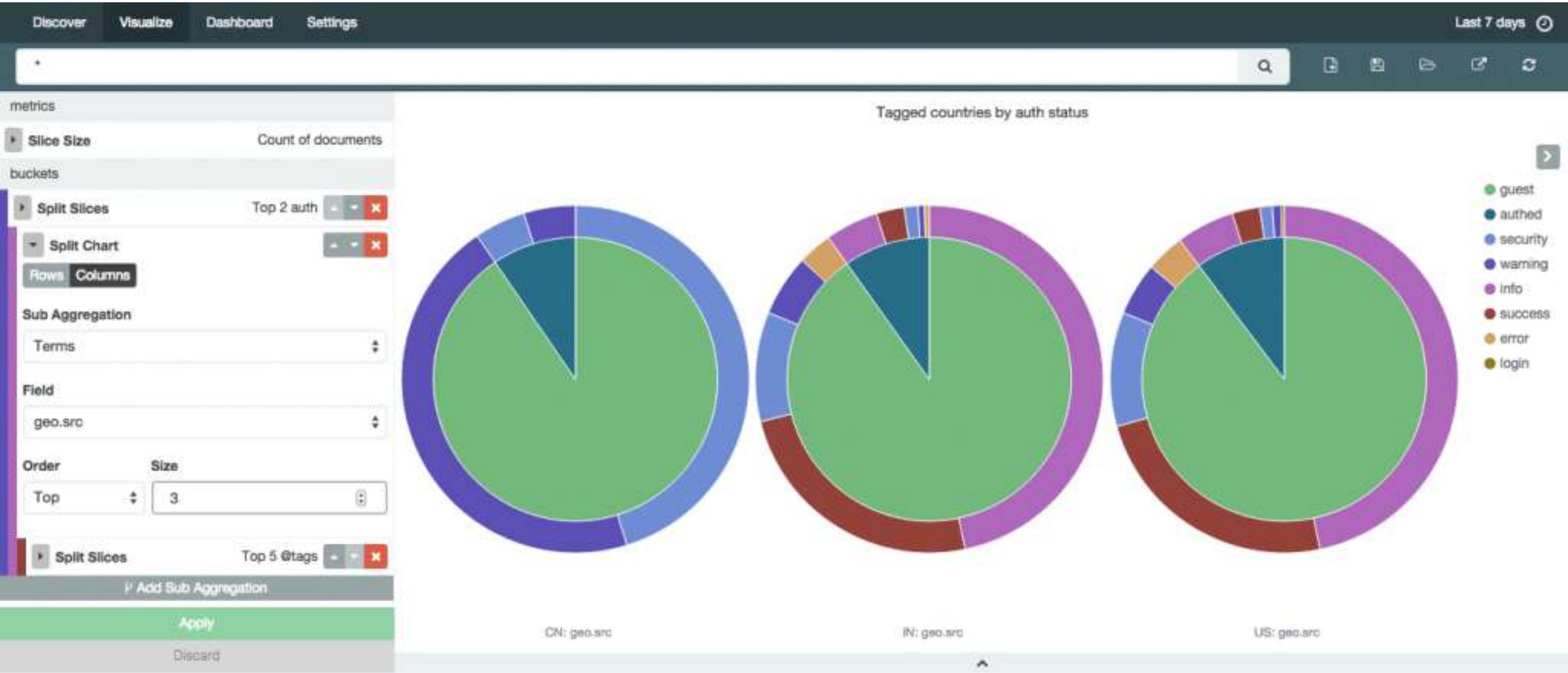
Hourly

Split Lines

Sub Aggregation

Terms

Kibana 4



metrics

Slice Size: Count of documents

buckets

Split Slices: Top 2 auth

Split Chart: Rows Columns

Sub Aggregation: Terms

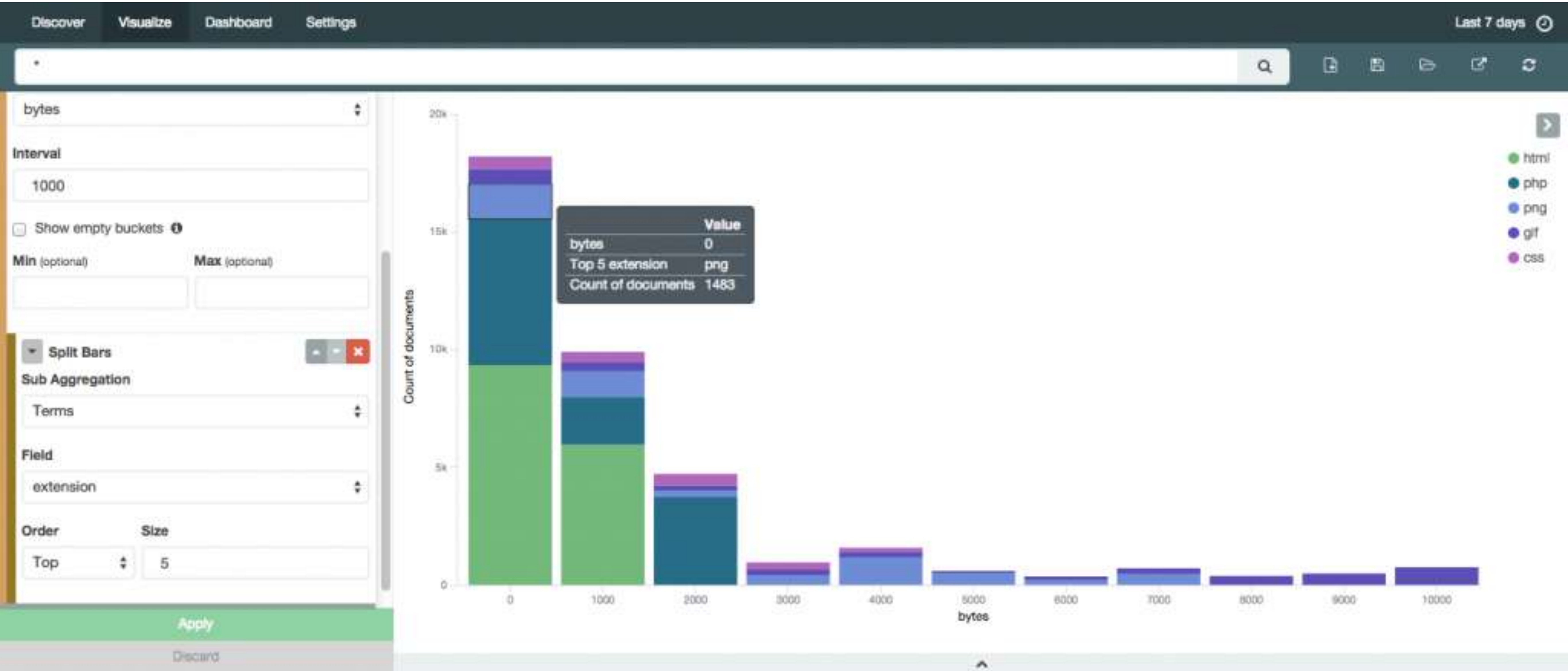
Field: geo.src

Order: Top

Size: 3

Split Slices: Top 5 @tags

Kibana 4



bytes

Interval: 1000

□ Show empty buckets

Min (optional)

Max (optional)

Split Bars

Sub Aggregation: Terms

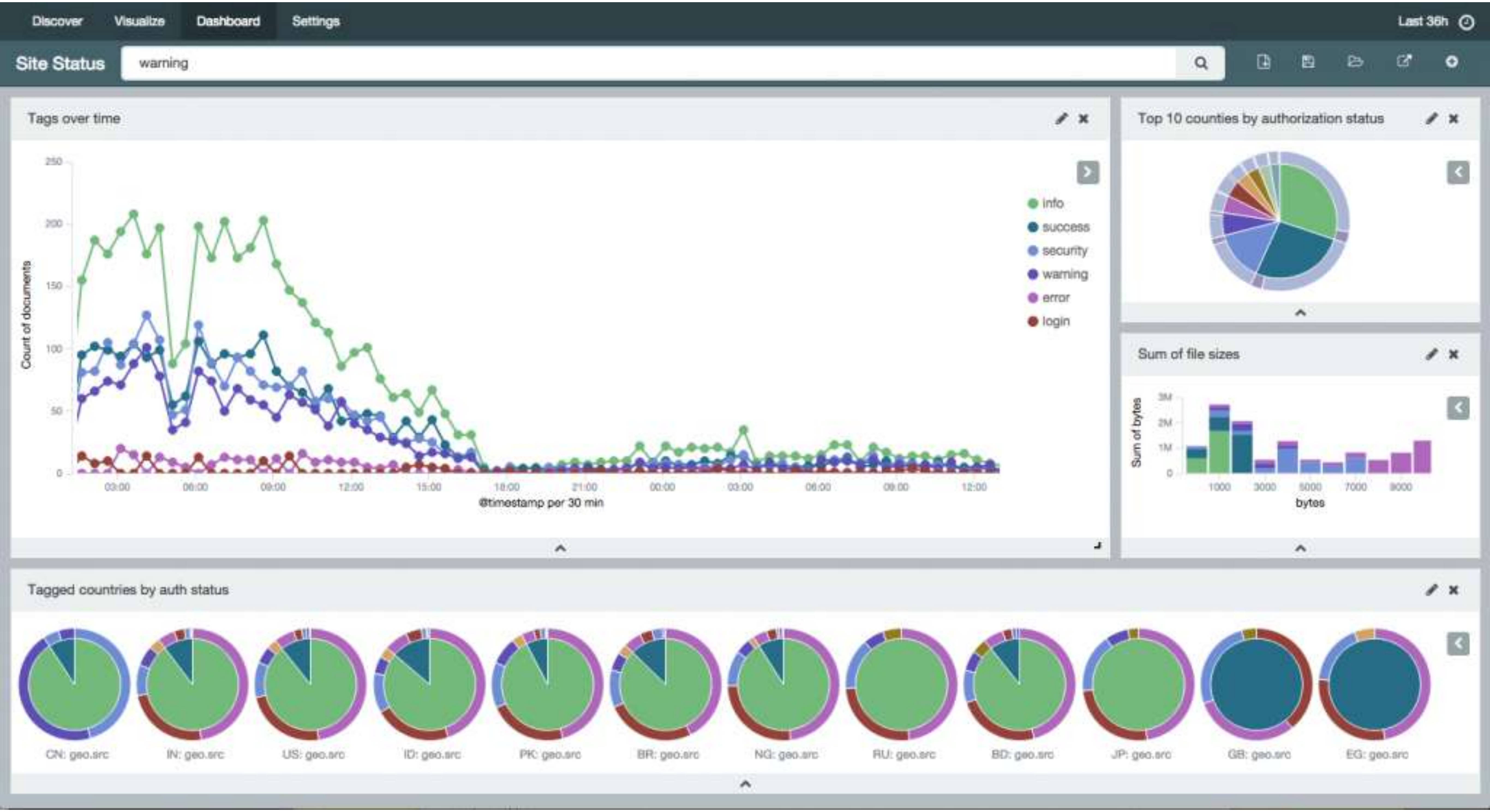
Field: extension

Order: Top

Size: 5

	Value
bytes	0
Top 5 extension	png
Count of documents	1483

Kibana 4



Discover
Visualize
Dashboard
Settings
Last 36h
Site Status
warning
Tags over time
250
200
150
100
50
0
Count of documents
03:00
06:00
09:00
12:00
15:00
18:00
21:00
00:00
03:00
06:00
09:00
12:00
@timestamp per 30 min
info
success
security
warning
error
login
Top 10 counties by authorization status
Sum of file sizes
3M
2M
1M
0
Sum of bytes
1000
3000
5000
7000
9000
bytes
Tagged countries by auth status
CN: geo.src
IN: geo.src
US: geo.src
ID: geo.src
PK: geo.src
BR: geo.src
NG: geo.src
RU: geo.src
BD: geo.src
JP: geo.src
GB: geo.src
EG: geo.src

Beispiele

Beispiele

- Guardian case study
- Web server logs
- meetup.com RSVP stream
- kippo SSH honeypot

Case study: The Guardian

theguardian

- Ophan: In-house analytics software

- Empower the organization

  Give the entire organization real-time insight into audience engagement

  Democratize analytics access for more than 500 users

  Encourage a culture of exploration and innovation for all employees

- Leverage real-time analytics

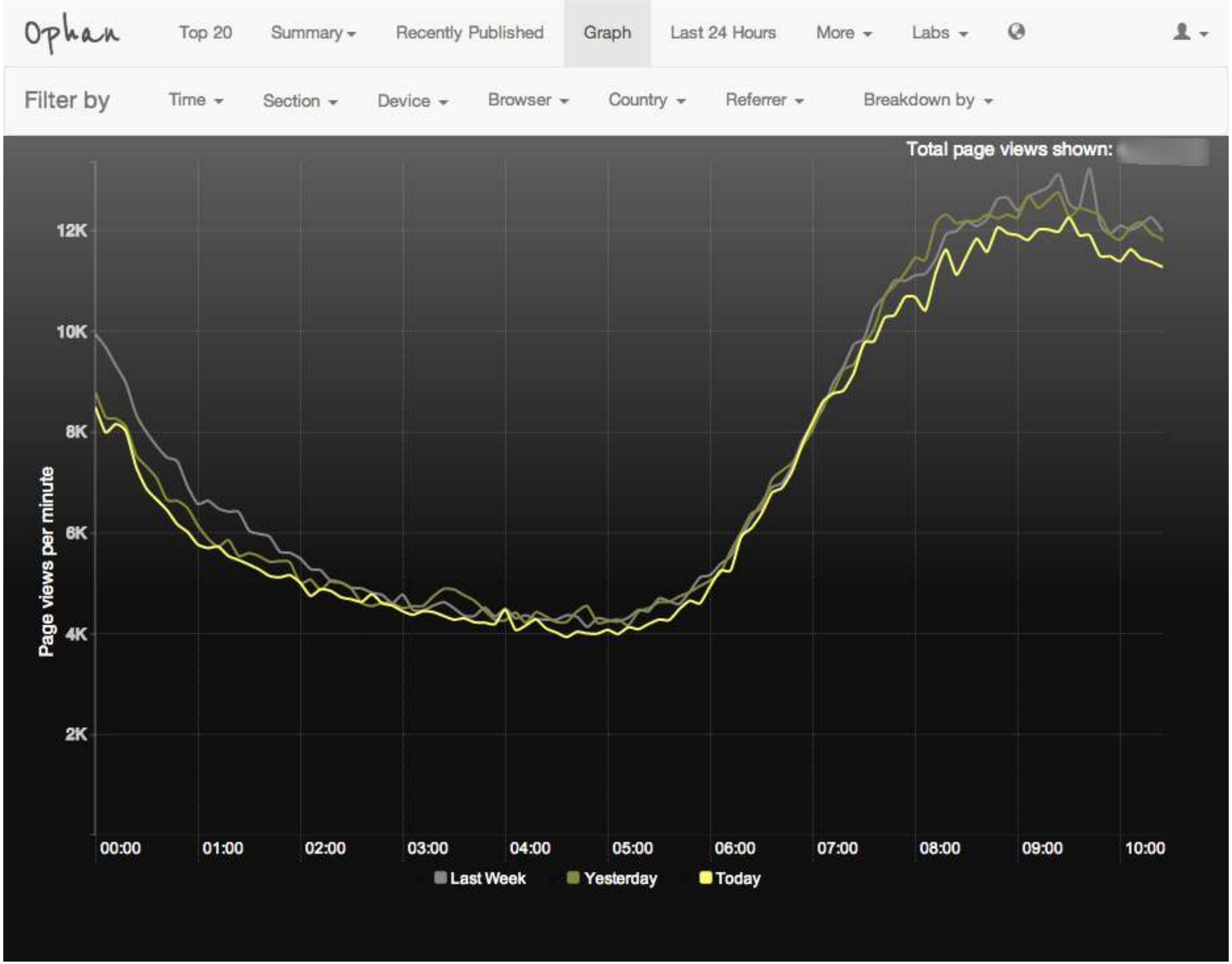
  Easily query 360 million documents

  See traffic for all content as it happens

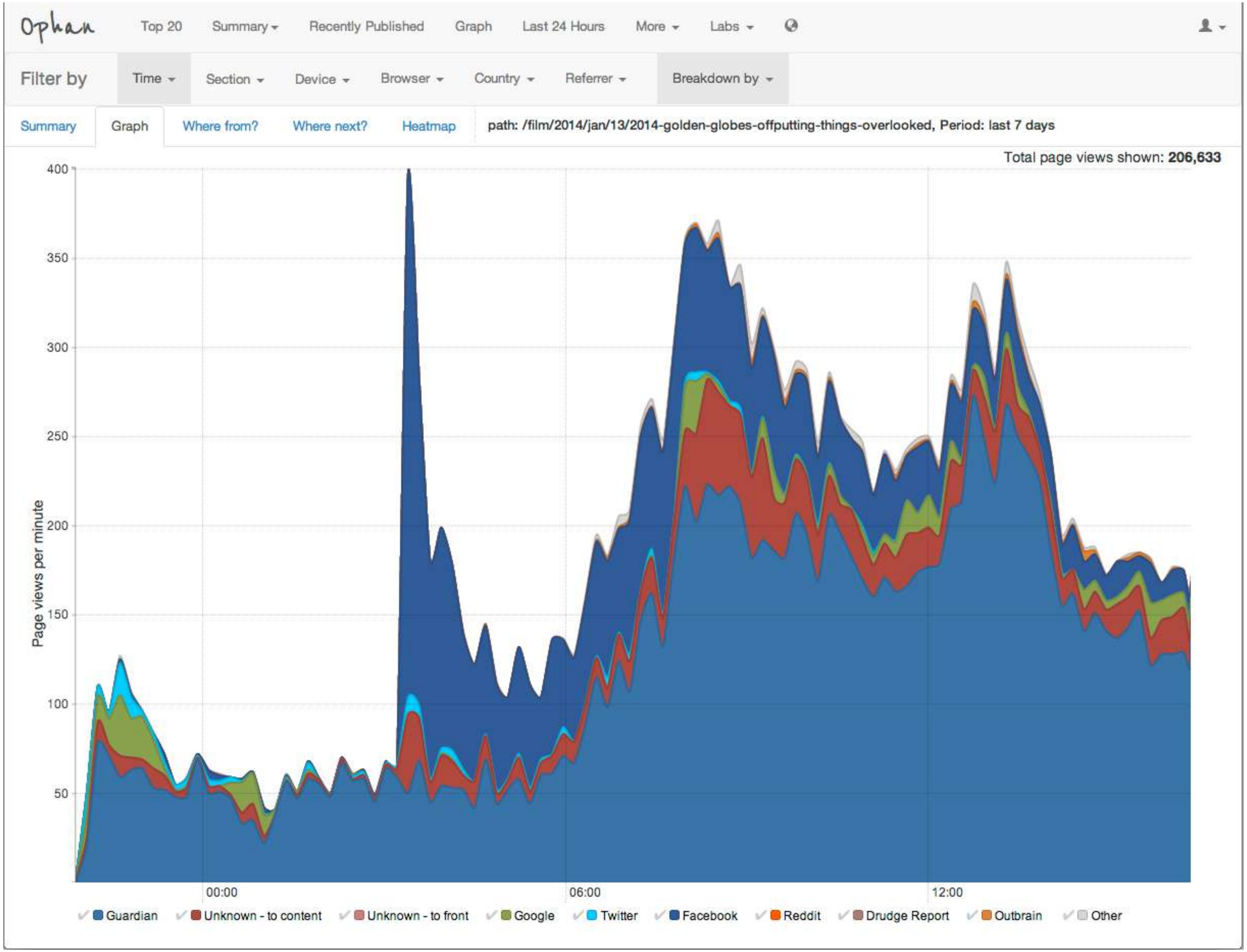
  Gain insight into how updates impact site traffic

http://www.elasticsearch.com/case-study/guardian/

Case study: The Guardian



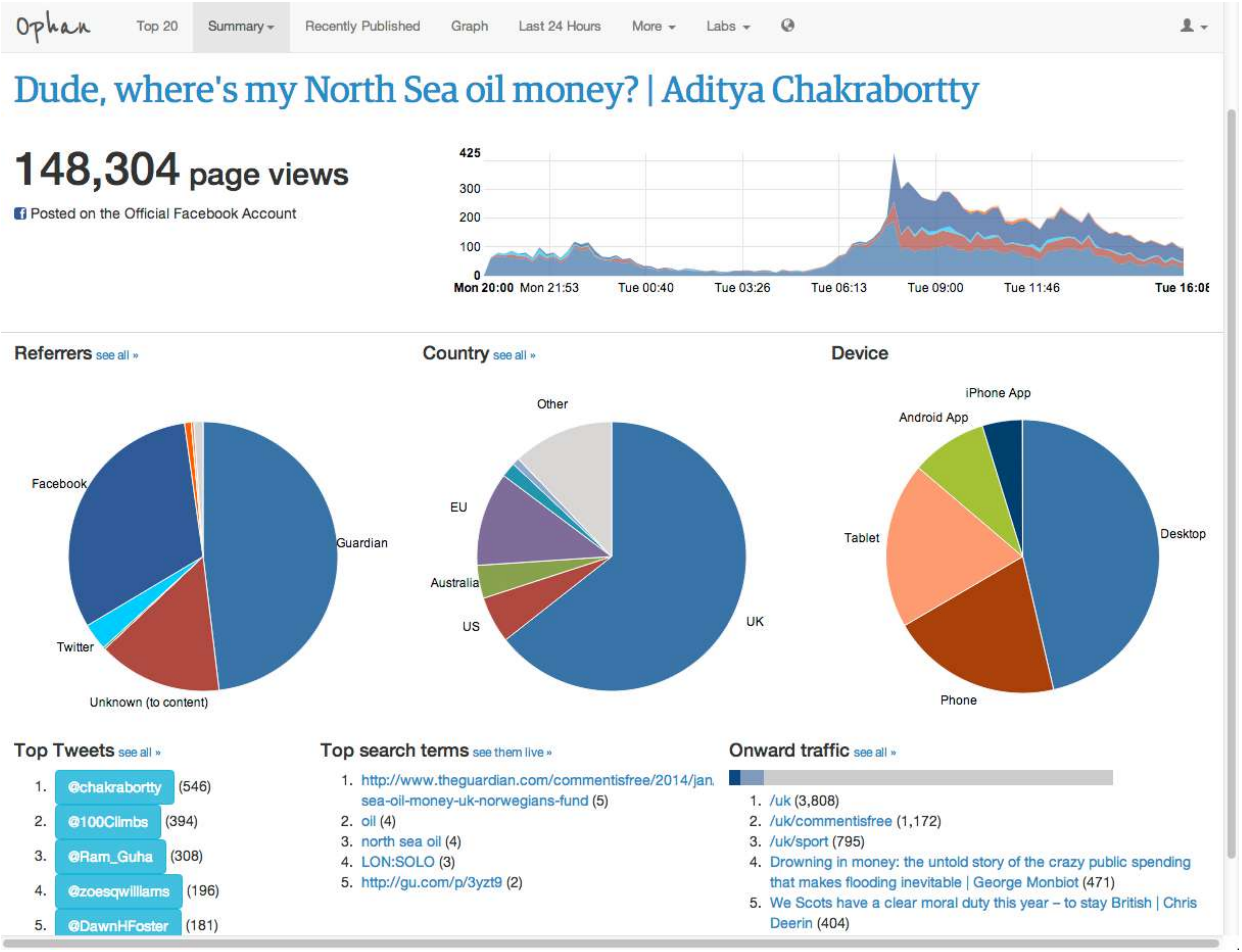
Case study: The Guardian



path: /film/2014/jan/13/2014-golden-globes-offputting-things-overlooked, Period: last 7 days

Total page views shown: **206,633**

Case study: The Guardian



Dude, where's my North Sea oil money? | Aditya Chakrabortty

148,304 page views

Posted on the Official Facebook Account

Referrers see all »

Country see all »

Device

Top Tweets see all »

1. @chakrabortty (546)
2. @100Climbs (394)
3. @Ram_Guha (308)
4. @zoesqwilliams (196)
5. @DawnHFoster (181)

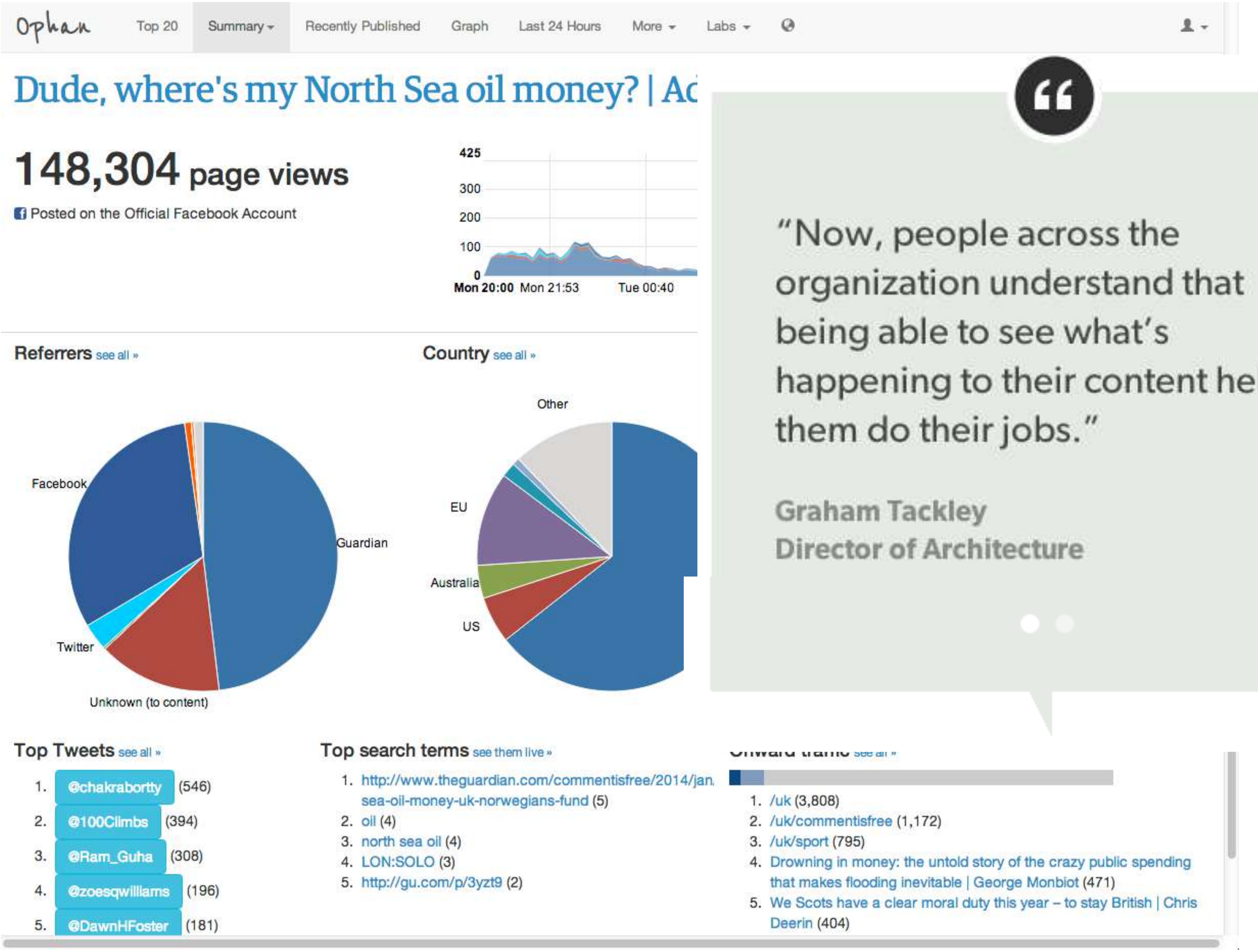
Top search terms see them live »

1. http://www.theguardian.com/commentisfree/2014/jan sea-oil-money-uk-norwegians-fund (5)
2. oil (4)
3. north sea oil (4)
4. LON:SOLO (3)
5. http://gu.com/p/3yzt9 (2)

Onward traffic see all »

1. /uk (3,808)
2. /uk/commentisfree (1,172)
3. /uk/sport (795)
4. Drowning in money: the untold story of the crazy public spending that makes flooding inevitable | George Monbiot (471)
5. We Scots have a clear moral duty this year – to stay British | Chris Deerin (404)

Case study: The Guardian



Dude, where's my North Sea oil money? | Ad

148,304 page views

Posted on the Official Facebook Account

Referrers see all »

Country see all »

Top Tweets see all »

1. @chakrabortty (546)
2. @100Climbs (394)
3. @Ram_Guha (308)
4. @zoesqwilliams (196)
5. @DawnHFoster (181)

Top search terms see them live »

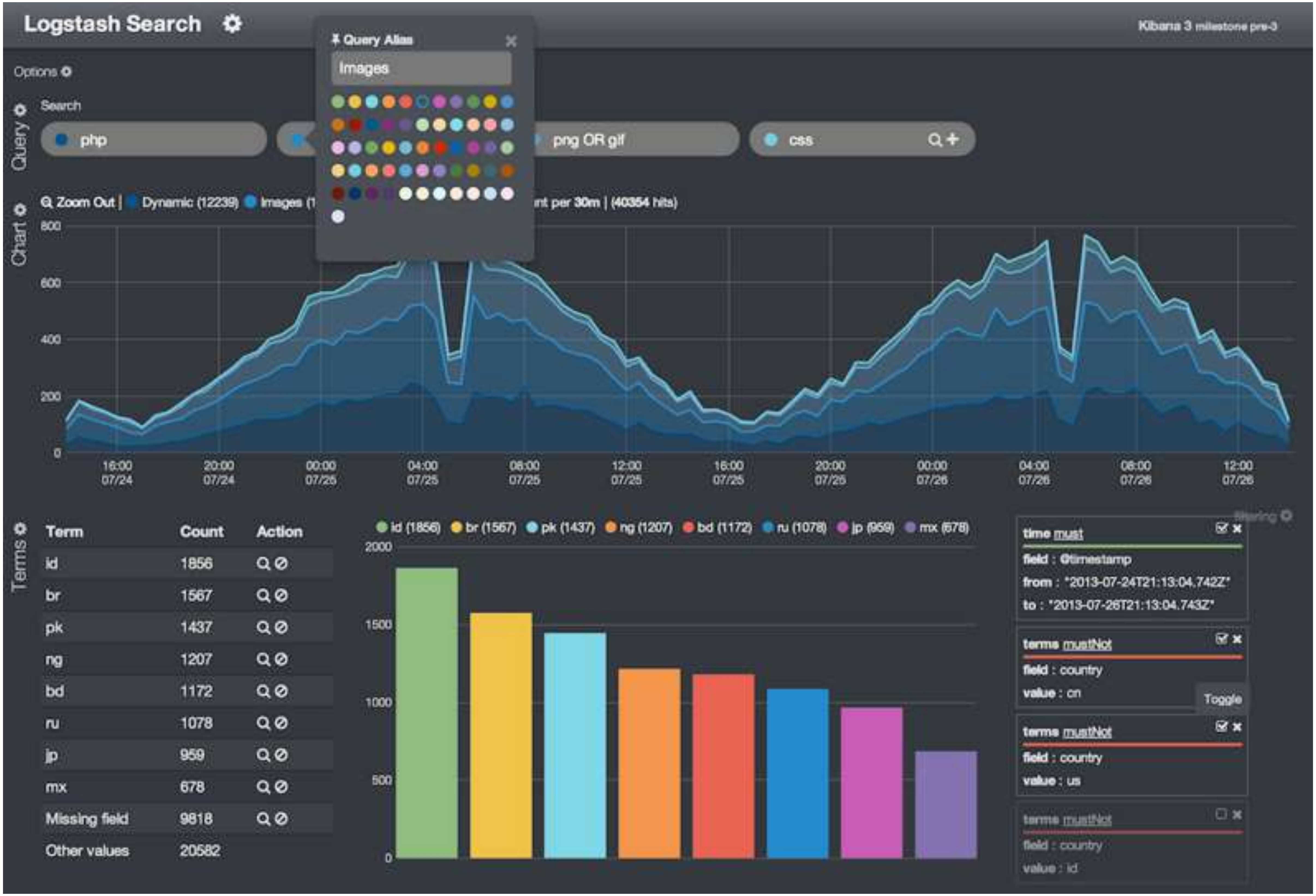
1. http://www.theguardian.com/commentisfree/2014/jan sea-oil-money-uk-norwegians-fund (5)
2. oil (4)
3. north sea oil (4)
4. LON:SOLO (3)
5. http://gu.com/p/3yzt9 (2)

Onward traffic see all »

1. /uk (3,808)
2. /uk/commentisfree (1,172)
3. /uk/sport (795)
4. Drowning in money: the untold story of the crazy public spending that makes flooding inevitable | George Monbiot (471)
5. We Scots have a clear moral duty this year – to stay British | Chris Deerin (404)

> "Now, people across the organization understand that being able to see what's happening to their content helps them do their jobs."
>
> **Graham Tackley**
> **Director of Architecture**

Beispiel: Web server logs



Term	Count	Action
id	1856	
br	1567	
pk	1437	
ng	1207	
bd	1172	
ru	1078	
jp	959	
mx	678	
Missing field	9818	
Other values	20582	

time must ☑ ✕
field : @timestamp
from : "2013-07-24T21:13:04.742Z"
to : "2013-07-26T21:13:04.743Z"

terms mustNot ☑ ✕
field : country
value : cn

terms mustNot ☑ ✕
field : country
value : us

terms mustNot □ ✕
field : country
value : id

Beispiel: Logstash Konfiguration

```
input { stdin {} }

filter {
  grok { match => { "message" => "%{COMBINEDAPACHELOG}" } }

  date { match => [ "timestamp", "dd/MMM/YYYY:HH:mm:ss Z" ] }

  geoip { source => “clientip" }

  useragent {
    source => "agent"
    target => "useragent"
  }
}

output {
  elasticsearch {
    protocol => "http"
    host => "localhost"
  }
}
```

Beispiel: Logstash Konfiguration

```
input { stdin {} }

filter {
  grok { match => { "message" => "%{COMBINEDAPACHELOG}" } }

  date { match => [ "timestamp", "dd/MMM/YYYY:HH:mm:ss Z" ] }

  geoip { source => “clientip" }

  useragent {
    source => "agent"
    target => "useragent"
  }
}

output {
  elasticsearch {
    protocol => "http"
    host => "localhost"
  }
}
```

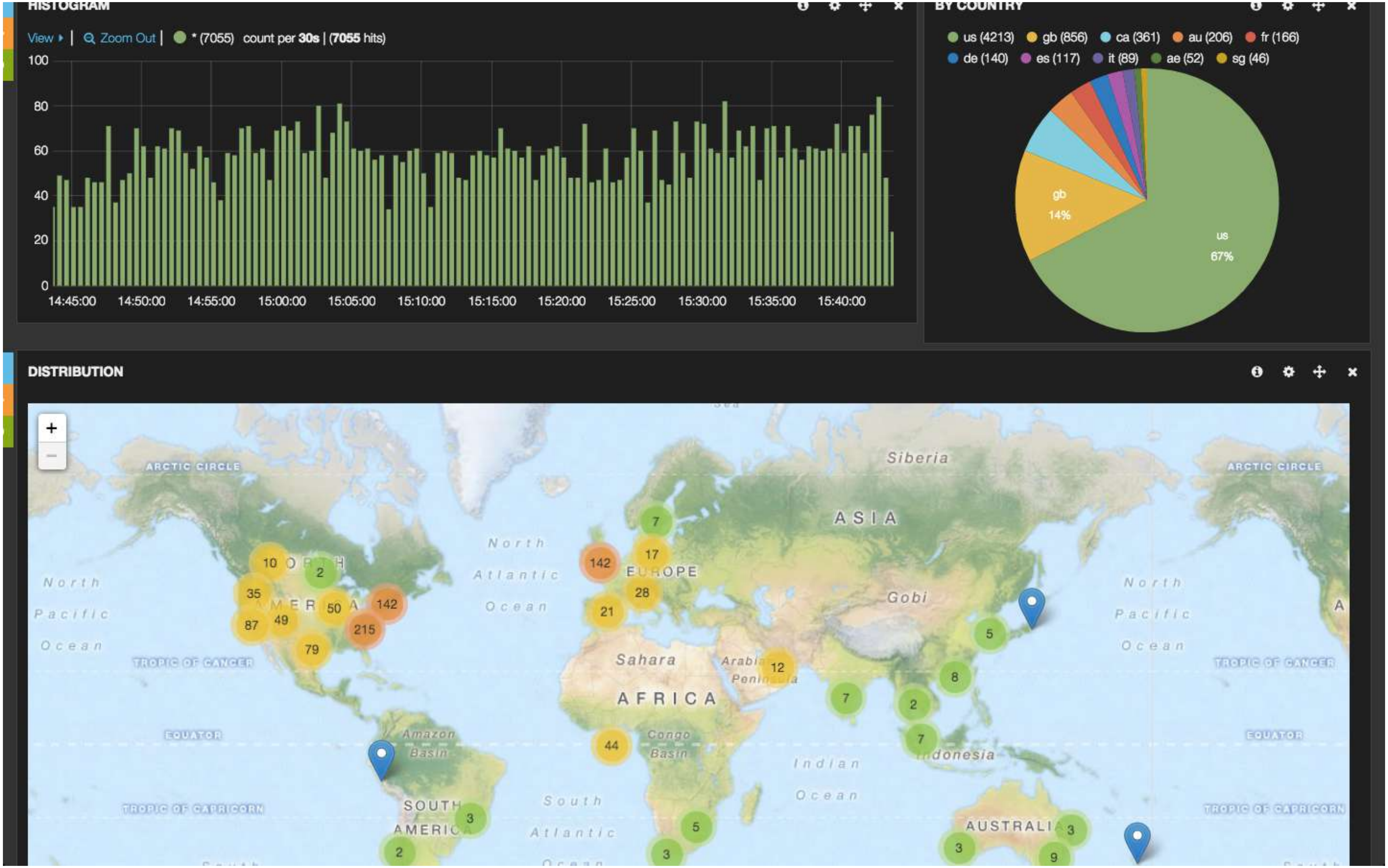
```
cat access.log | logstash agent -f logstash.conf
```

meetup.com RSVP stream

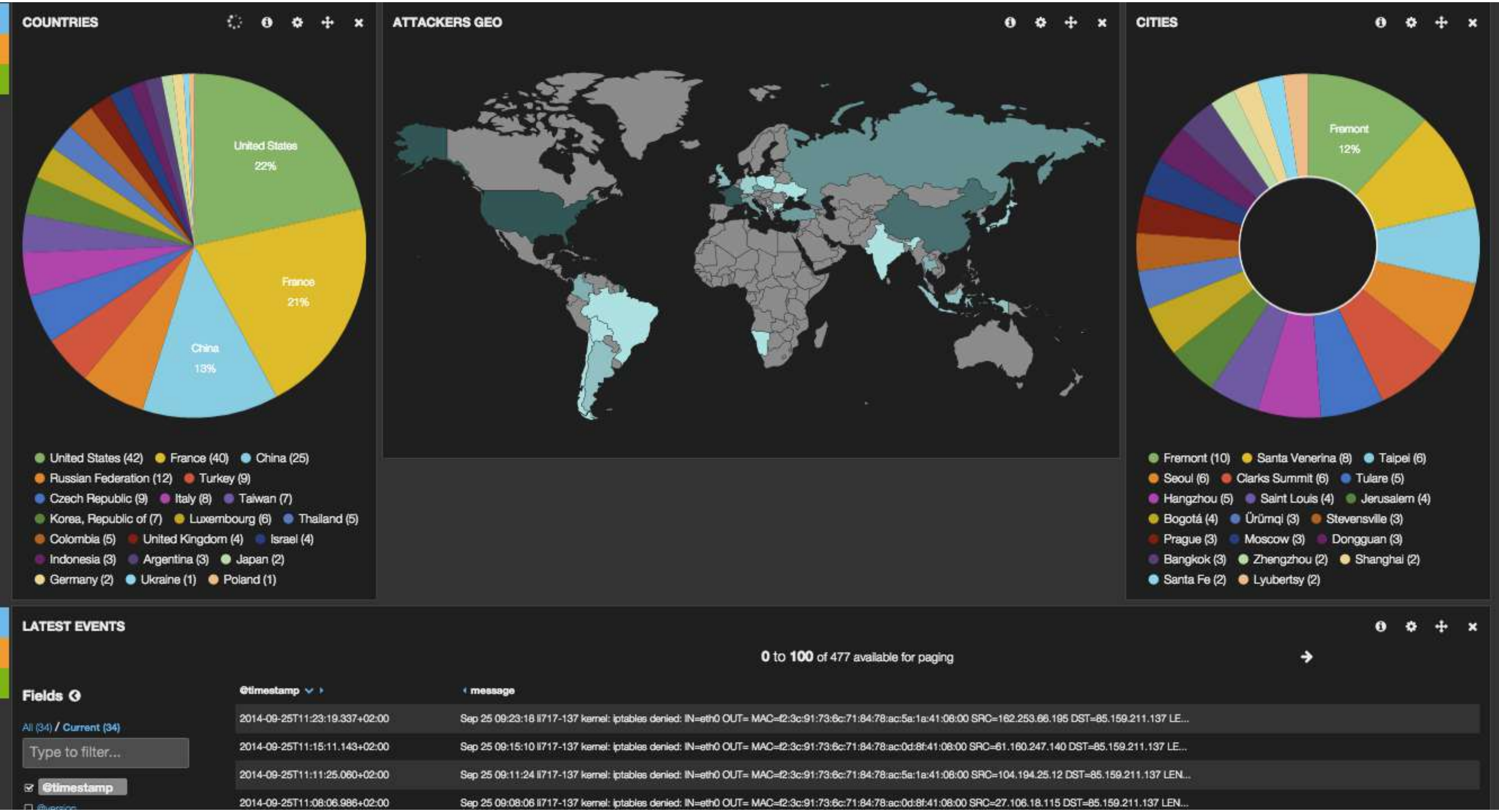
- Alle Reservierungen werden an einen HTTP Stream gesendet
- Jede Zeile ist ein eigenes JSON Dokument (= Event)

http://stream.meetup.com/2/rsvps

meetup.com RSVP stream



kippo SSH honeypot



0 to 100 of 477 available for paging

```
@timestamp                      message
2014-09-25T11:23:19.337+02:00   Sep 25 09:23:18 li717-137 kernel: iptables denied: IN=eth0 OUT= MAC=f2:3c:91:73:6c:71:84:78:ac:5a:1a:41:08:00 SRC=162.253.66.195 DST=85.159.211.137 LE...
2014-09-25T11:15:11.143+02:00   Sep 25 09:15:10 li717-137 kernel: iptables denied: IN=eth0 OUT= MAC=f2:3c:91:73:6c:71:84:78:ac:0d:8f:41:08:00 SRC=61.160.247.140 DST=85.159.211.137 LE...
2014-09-25T11:11:25.060+02:00   Sep 25 09:11:24 li717-137 kernel: iptables denied: IN=eth0 OUT= MAC=f2:3c:91:73:6c:71:84:78:ac:5a:1a:41:08:00 SRC=104.194.25.12 DST=85.159.211.137 LEN...
2014-09-25T11:08:06.986+02:00   Sep 25 09:08:06 li717-137 kernel: iptables denied: IN=eth0 OUT= MAC=f2:3c:91:73:6c:71:84:78:ac:0d:8f:41:08:00 SRC=27.106.18.115 DST=85.159.211.137 LEN...
```

Zusammenfassung

Zusammenfassung

Daten

Fokus

Iteration

Zusammenfassung

Daten

Fokus

Iteration

- Extrahieren & zentralisieren
- Bereitstellen & demokratisieren

Zusammenfassung

Daten

Fokus

Iteration

- Informationsgewinnung
- Visualisierung

Zusammenfassung

Daten

Fokus

Iteration

- Kontinuierliche Verbesserung

Einfach loslegen!

- Elasticsearch, Logstash & Kibana herunterladen & anfangen

```
# elasticsearch-1.4.2/bin/elasticsearch

# kibana-4.0.0/bin/kibana

# logstash-1.5.0/bin/logstash agent -f logstash.conf

# open localhost:5601
```

Resourcen

Links

- Elasticsearch, Logstash & Kibana

  http://www.elasticsearch.org

  http://www.elasticsearch.org/blog/

  http://www.elasticsearch.org/guide/

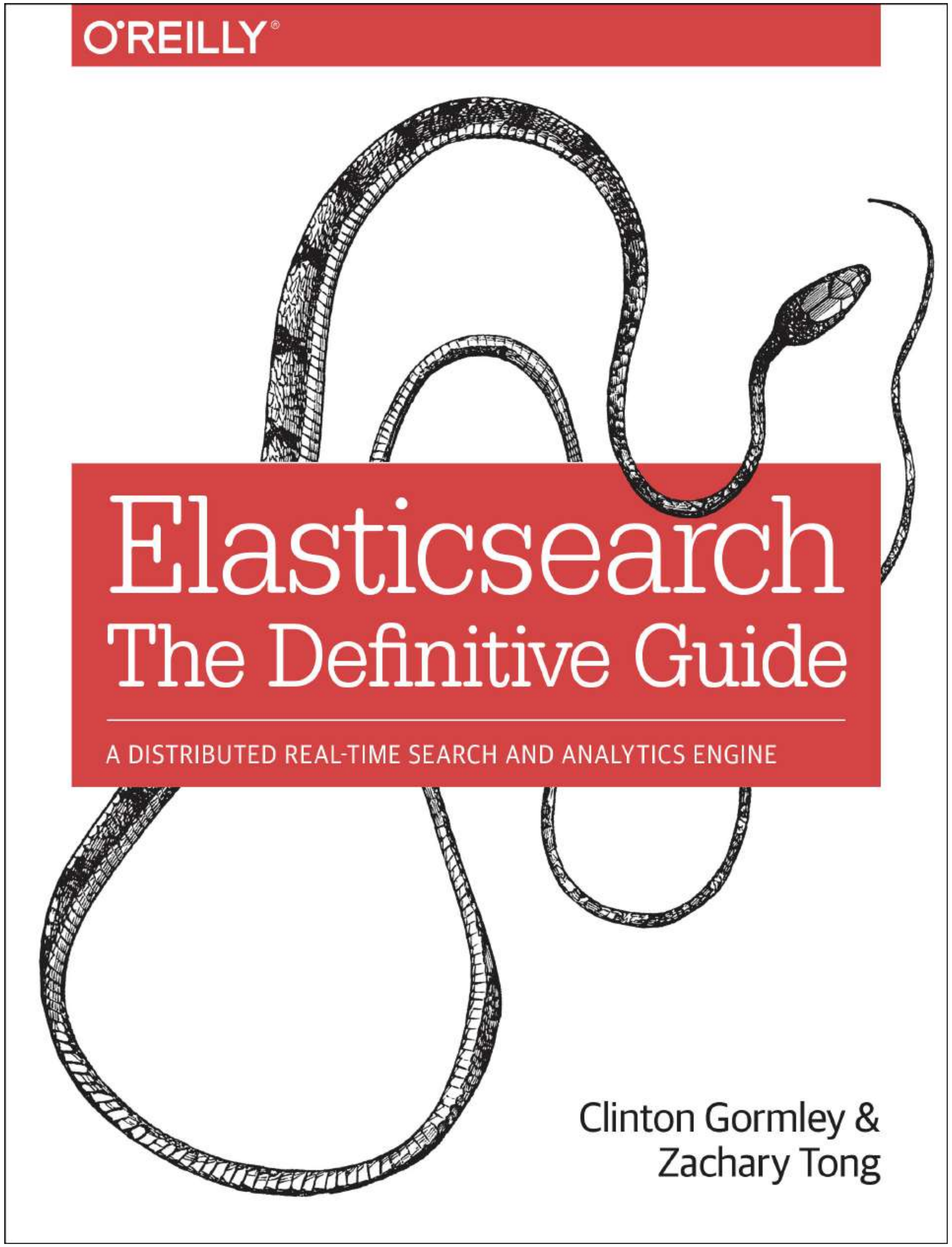
  http://www.elasticsearch.com

  http://www.elasticsearch.com/customers/

  http://www.elasticsearch.com/products/shield

  http://www.elasticsearch.com/products/marvel

Resources



O'REILLY

Elasticsearch
The Definitive Guide

A DISTRIBUTED REAL-TIME SEARCH AND ANALYTICS ENGINE

Clinton Gormley &
Zachary Tong

Vielen Dank!

We're hiring!
http://elasticsearch.com/jobs

We're helping!
http://elasticsearch.com/support
http://elasticsearch.com/training

Alexander Reelsen
@spinscale
alexander.reelsen@elasticsearch.com