Why we built the ELK stack

Continuous improvement of your data to achieve better business decisions

Alexander Reelsen
alexander.reelsen@elasticsearch.com

elasticsearch.

Agenda

Agenda

- Introduction

  The problem with data in your IT infrastructure

  Why your current approach is flawed

- The ELK stack

  Logstash

  Elasticsearch

  Kibana & Marvel

- Case Study

- Summary

Elasticsearch

- Founded in 2012 in Amsterdam
- Funded by Benchmark, Index Ventures and NEA Ventures
- Distributed company

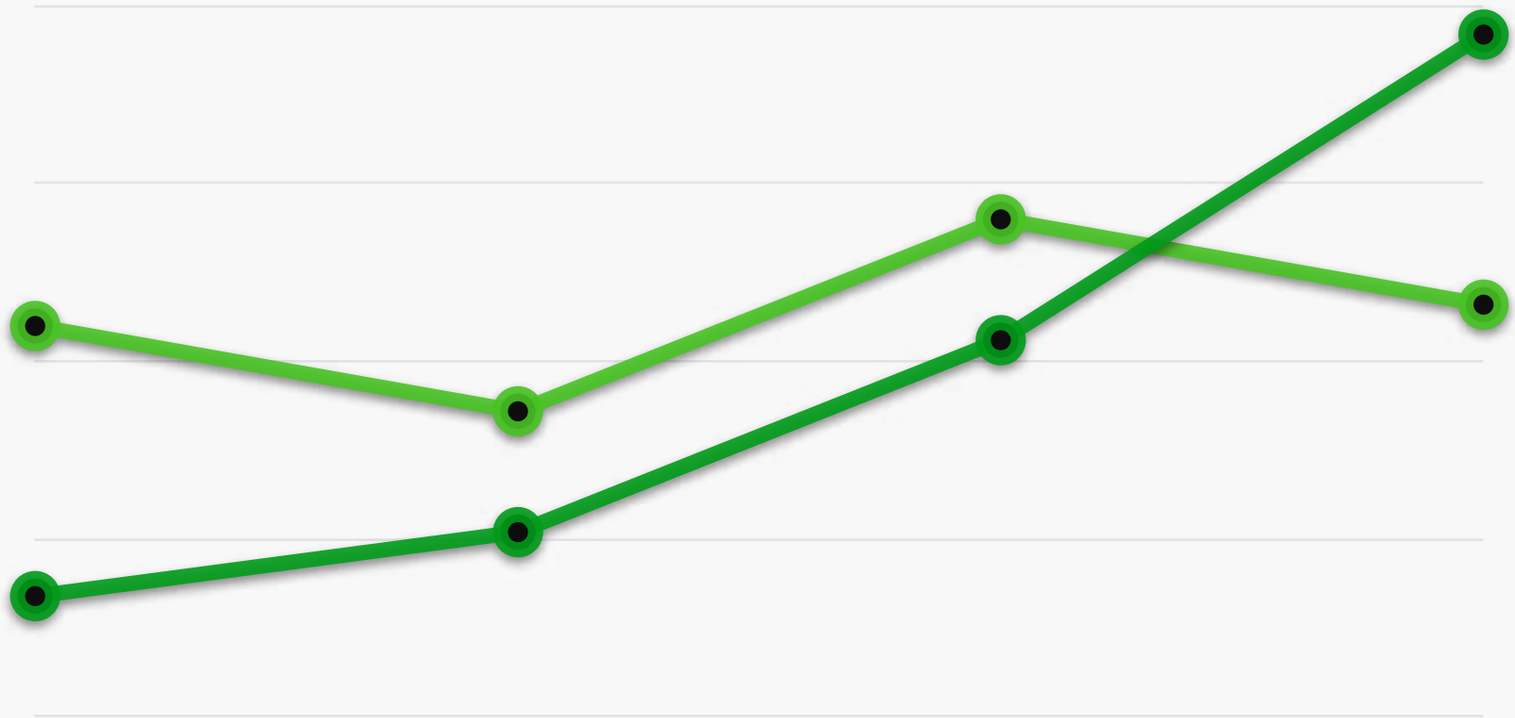
  HQs in Amsterdam & Los Altos

  Offices in Berlin, London & Phoenix
- Revenue from trainings, support subscriptions & monitoring product
- Employing experts in open source, search, logging & visualization

Introduction

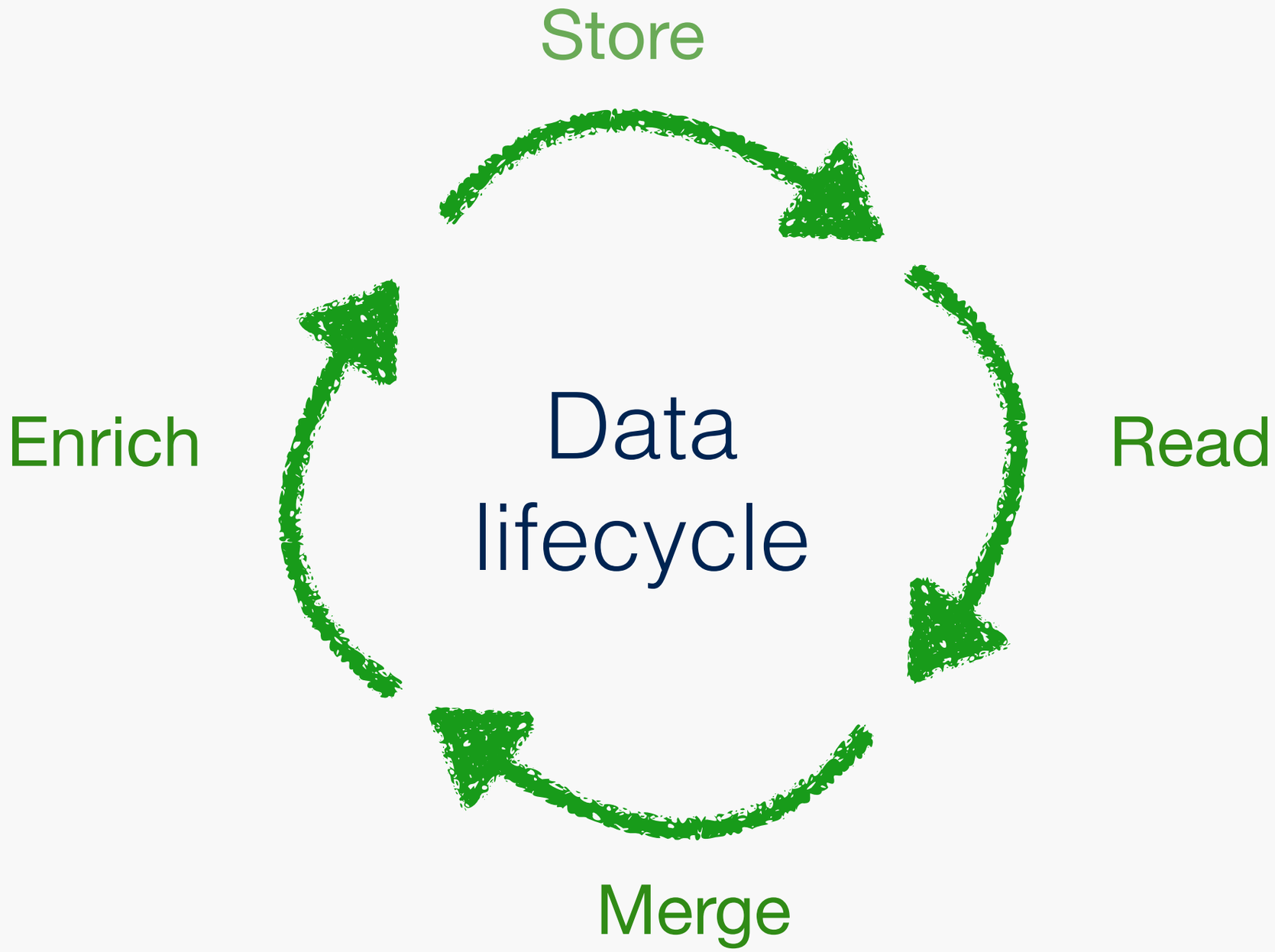
What is the core asset of your company?

- Ideas?
- Patents?
- Employees?
- Customers?
- Warehouse?
- Software?



- How do you decide where to invest?

By using data!



Lots of data!

- Product recommendations
- Page views
- Internet of Things
- Social media

- So, the more, the better? Sure, if...

The promise of "Big Data"

Create

01101001 11010011
11001001 10111001
00101010 00001101
00100110 11000101
11001010 00010001
00110011 10101101
00111101 00110010
11000110 11011110
01011110 00010111
01010010 10110101
01101001 11100010
01101011 10000000
11111010 00001111

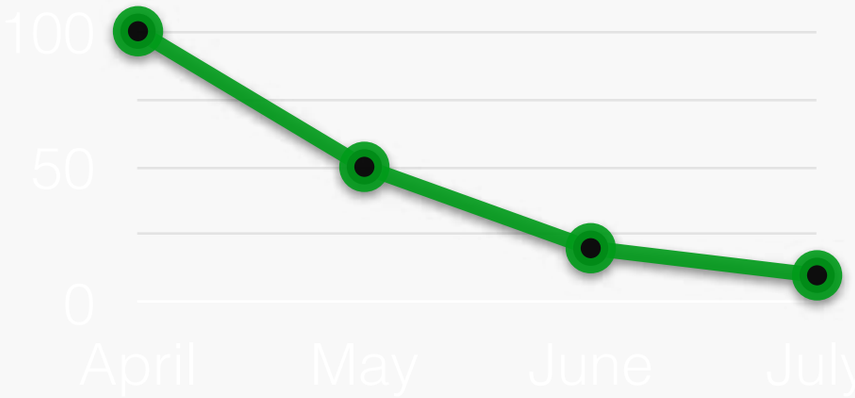
Store

Insight

- Problem 1: Missing key factor: TIME
- Problem 2: Merging different data sources
- Problem 3: Storing the data does not lead to insights

Correlation between time and event

- The value of an event changes based on how quickly you can store and analyze it



- Examples
  - Outage notification
  - Stock ticker value
  - Social media posts
  - Page views on frontpage (used for further ranking)

Merging data sources

- An event may increase its value if it is merged with different data

You just got your biggest order ever - what do you know about this customer?

Sudden traffic spike

Geo information from a mobile device when searching for a restaurant

Social media generated page views

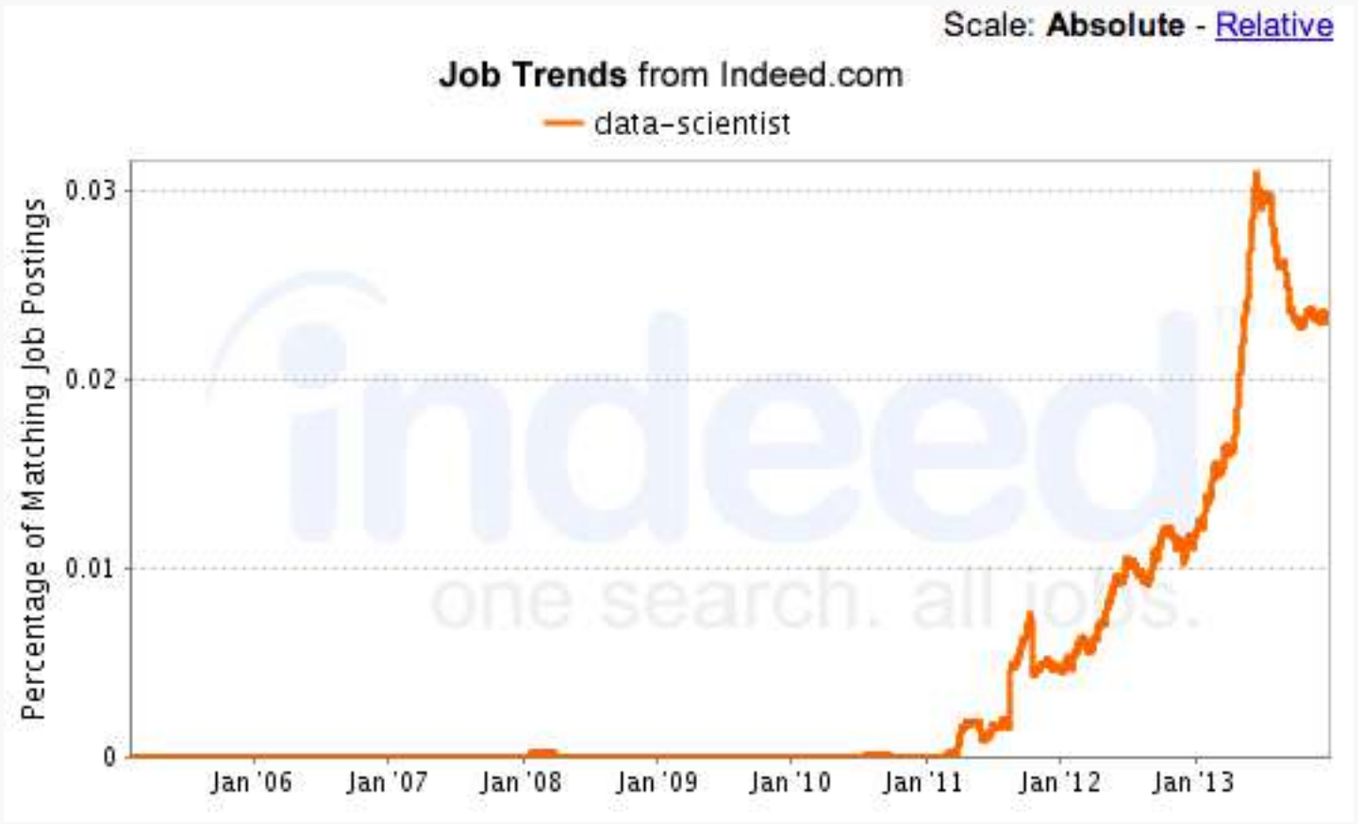
Fraud detection for payments

Storing data != insight

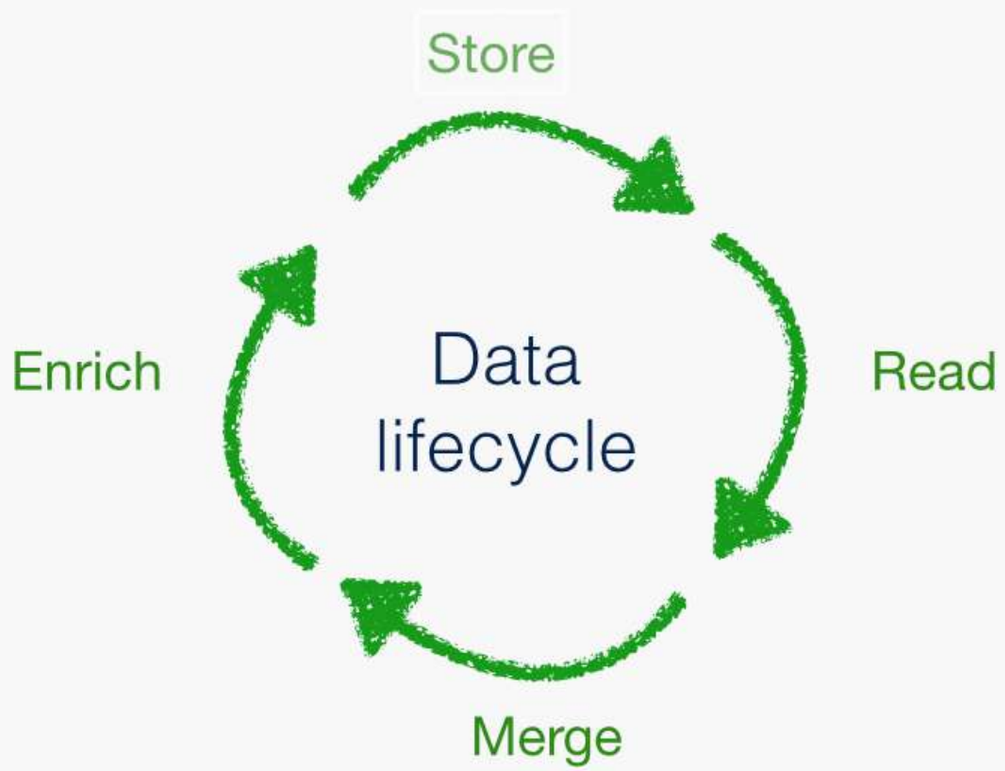
- Just because you are writing terabytes of data does not give you any value
- SQL example: We are trained to normalize our data as well as possible, until we denormalize it again to counter performance issues
- Data should never be optimised for writing, but always optimised for reading and information extraction.

elasticsearch.

The data scientist fallacy



- Result of a flawed IT infrastructure
- Often doing someone else's job
- Human process of that graph
- Gathering data != insights



elasticsearch.

Do it yourself!

- Why not let everyone create their own reports? Customized, straight to the point, near real-time
- Requirements

  Clean data to work on

  Fast analysis chain

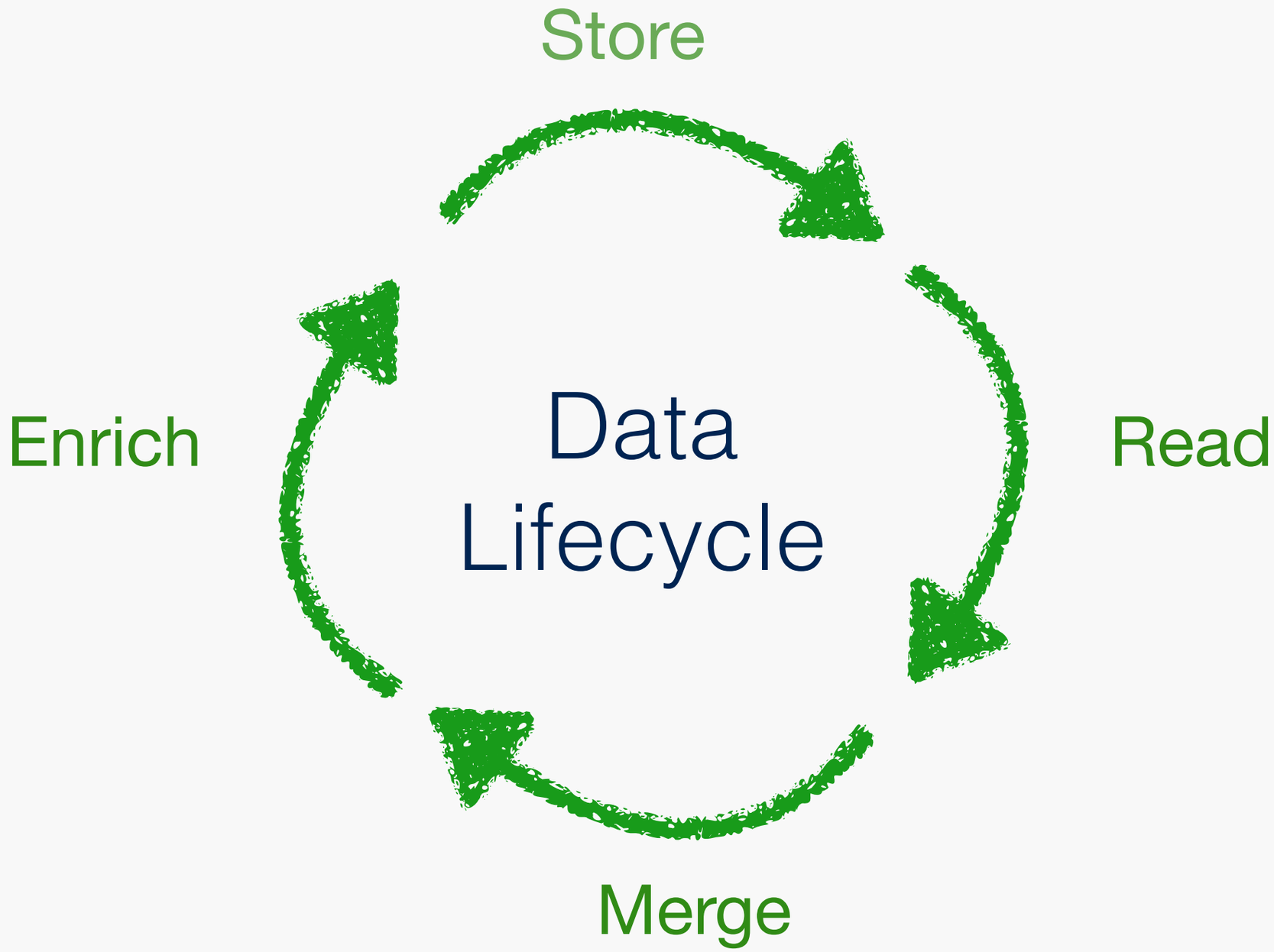
  Easy to use front-end

Meet the ELK stack

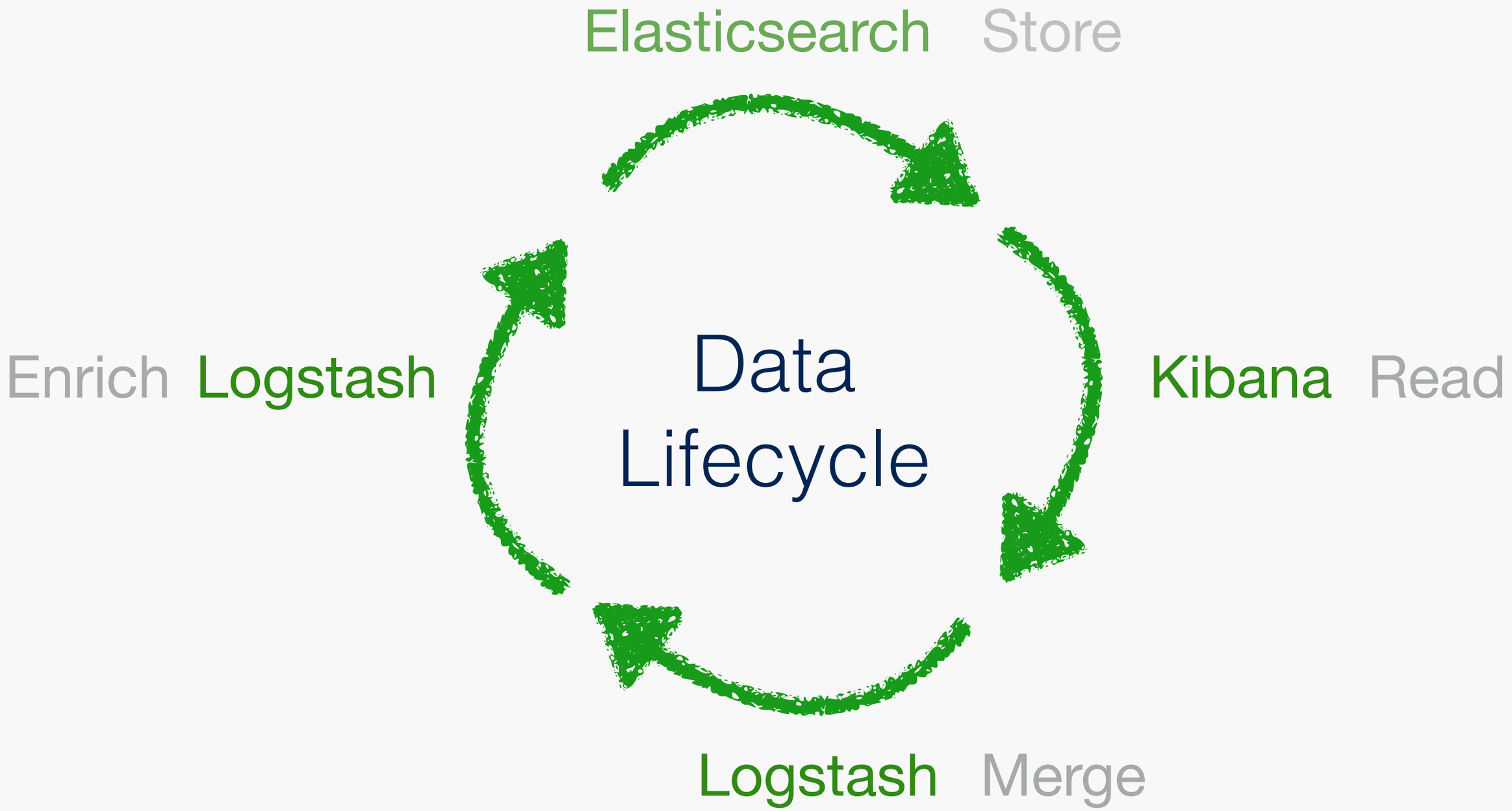
The ELK stack

elasticsearch.

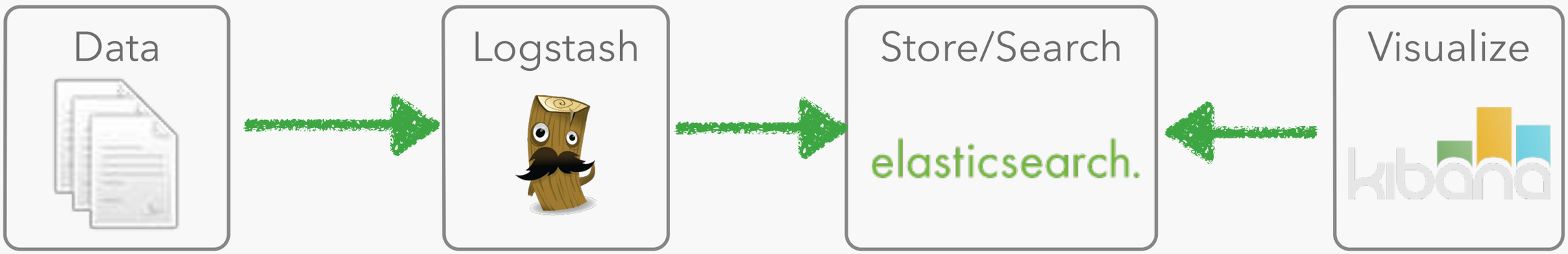
The ELK stack



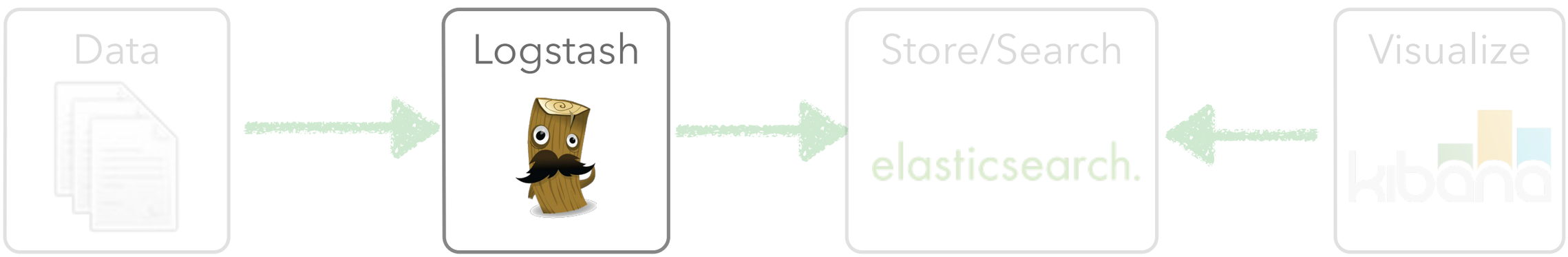
The ELK stack



The ELK stack



Logstash



elasticsearch.

Logstash

- Managing events and logs
- Collect data
- Parse data
- Enrich data
- Store data
- Open Source: Apache License 2.0

Logstash architecture

Input

datastore

stream

log files

files

monitoring

queues

network

parse, enrich, tag, drop

Output

datastore

files

email

pager

monitoring

chat

API

queues

Logstash architecture

Input

datastore

stream

log files

files

monitoring

queues

network

ip: 141.1.1.1

Filter

Logstash

parse, enrich, tag, drop

ip: 141.1.1.1

city: Zurich

country: CH

Output

datastore

files

email

pager

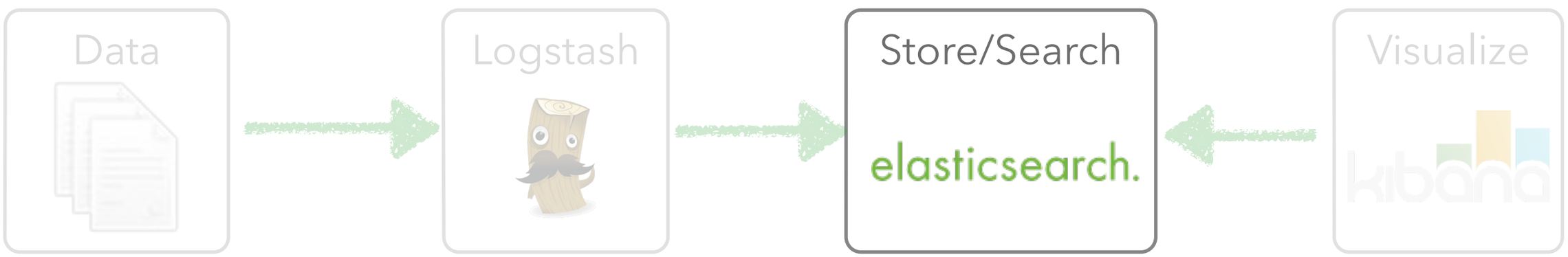
monitoring

chat

API

queues

Elasticsearch



elasticsearch.

Elasticsearch

- Schema-free, REST & JSON based distributed search engine
- Open Source: Apache License 2.0

- Easy to understand, yet very powerful query language

  Full text search (phrase, fuzzy)

  Numeric search (support ranges, dates, ipv4 addresses)

  Highlighting

  Aggregations

  Suggestions

XING

stackoverflow

SOUNDCLOUD

Fog Creek SOFTWARE

foursquare

LiveChat!com

loggly

NETWORKED INSIGHTS

elasticsearch.

Create knowledge from data

- Orders

  How many orders were created every day in the last month?

  How many orders were created per country in the last month?

- Money

  What is the average revenue per shopping cart?

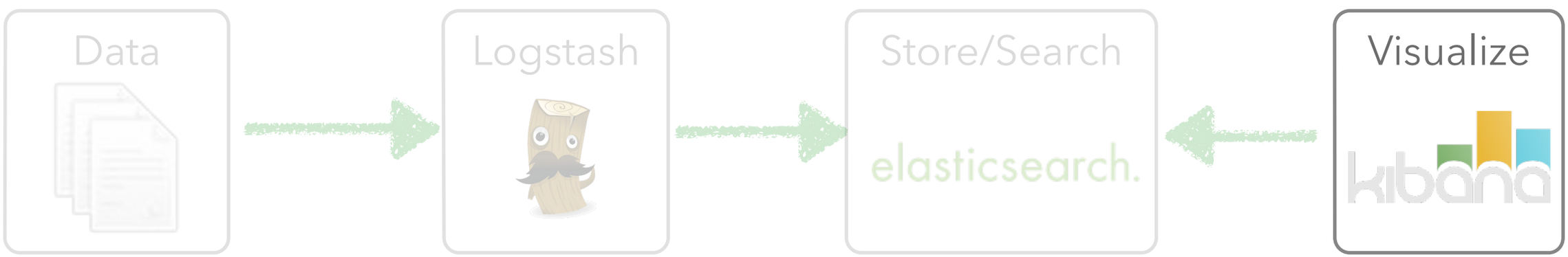
  What is the average shopping cart size per order per hour?

- Product portfolio

  Take the location of people into account for special offers?

  Analyze page views - premium or low-budget ecommerce site?

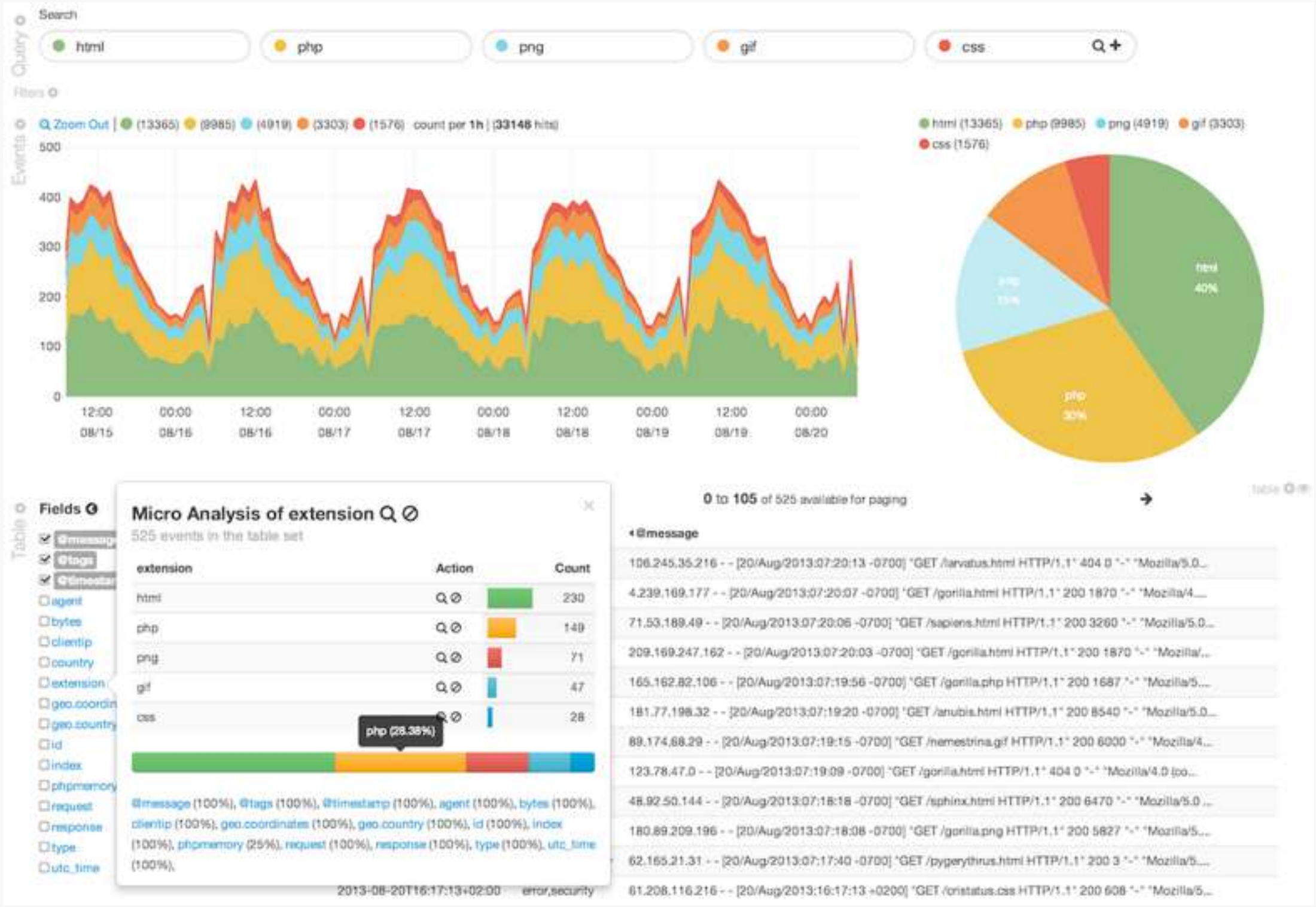
Kibana



Kibana

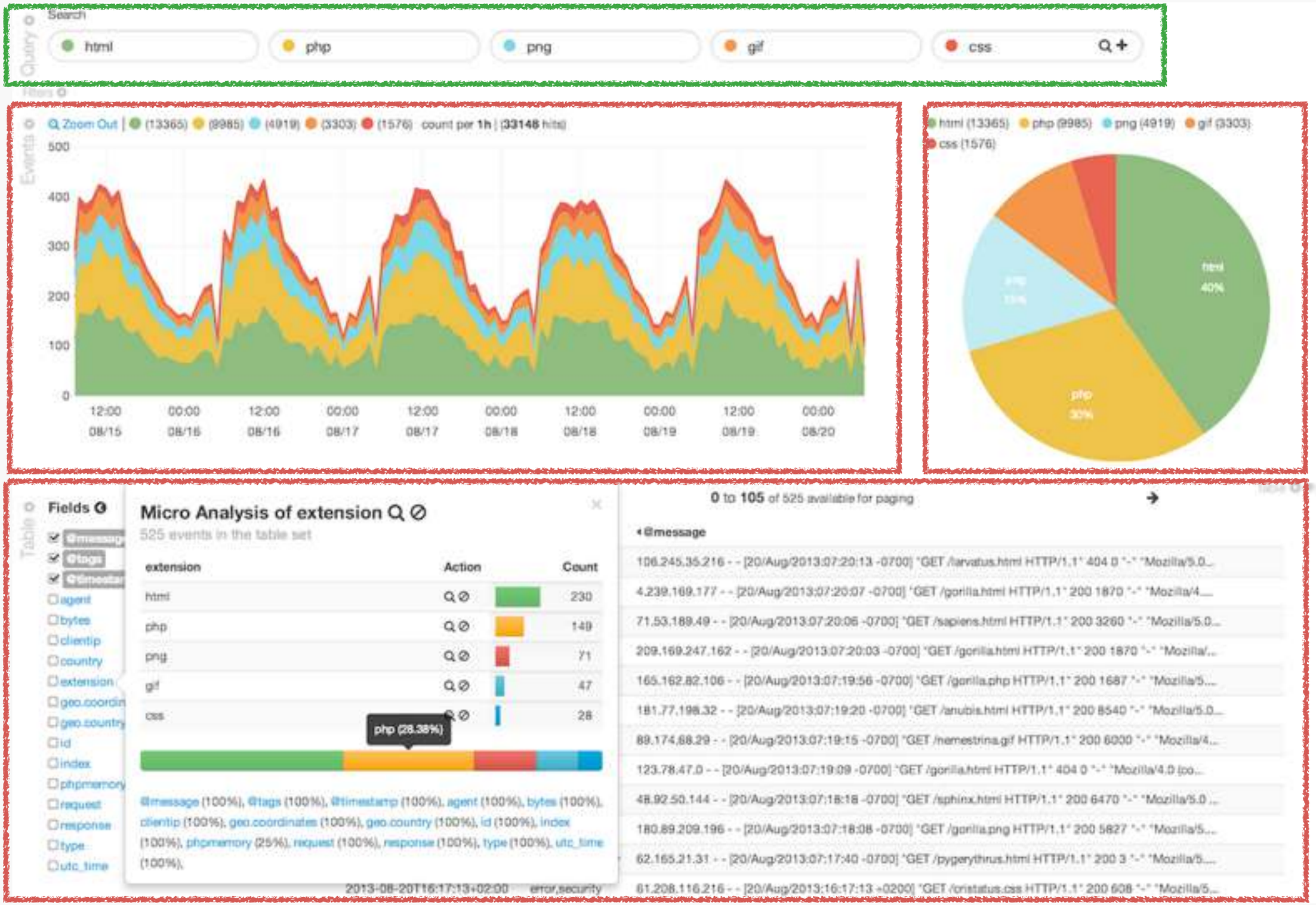
- Execute queries on your data & visualize results
- Add/remove widgets
- Share/Save/Load dashboards
- Open Source: Apache License 2.0

Kibana



elasticsearch.

Kibana

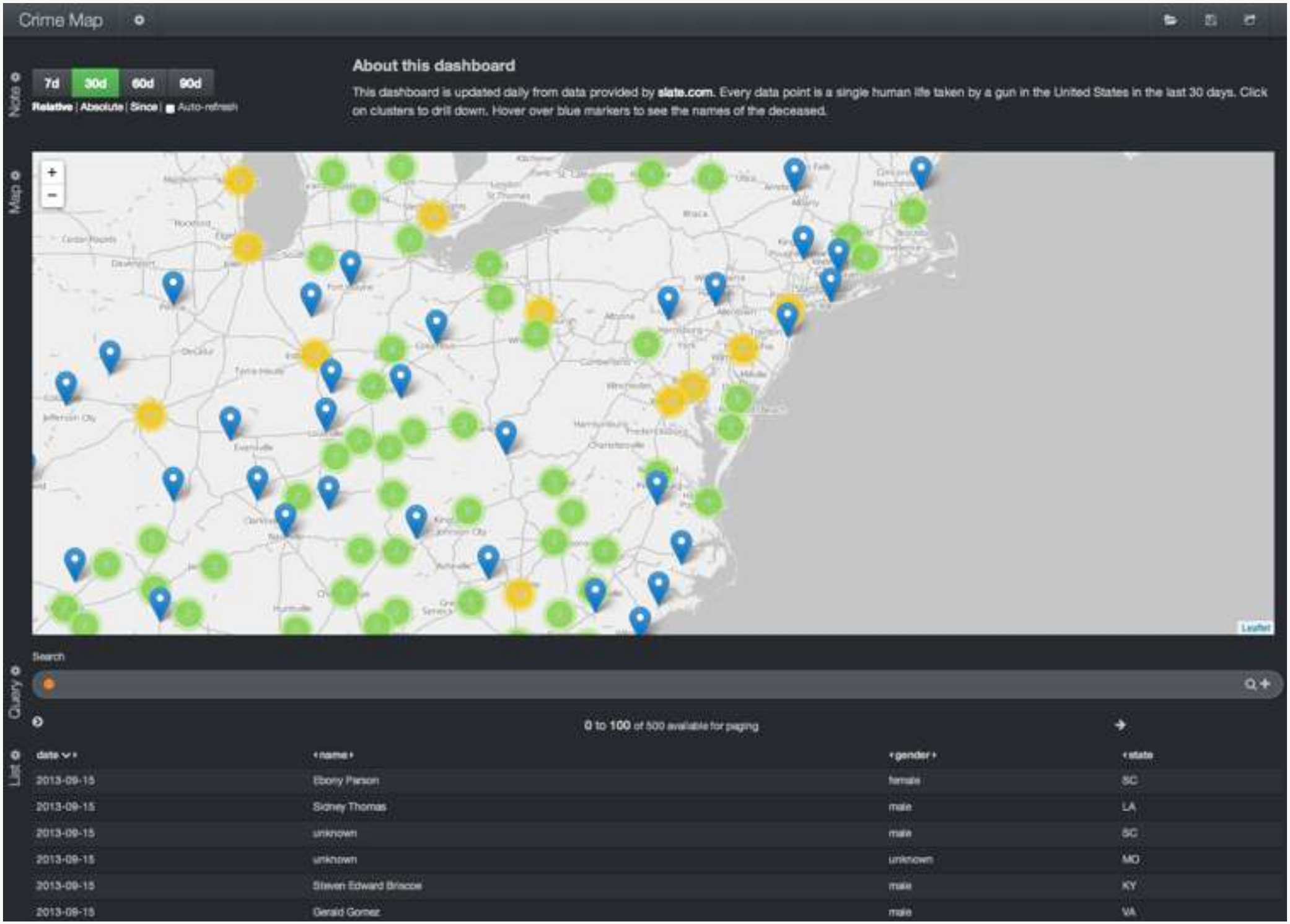


elasticsearch.

Kibana

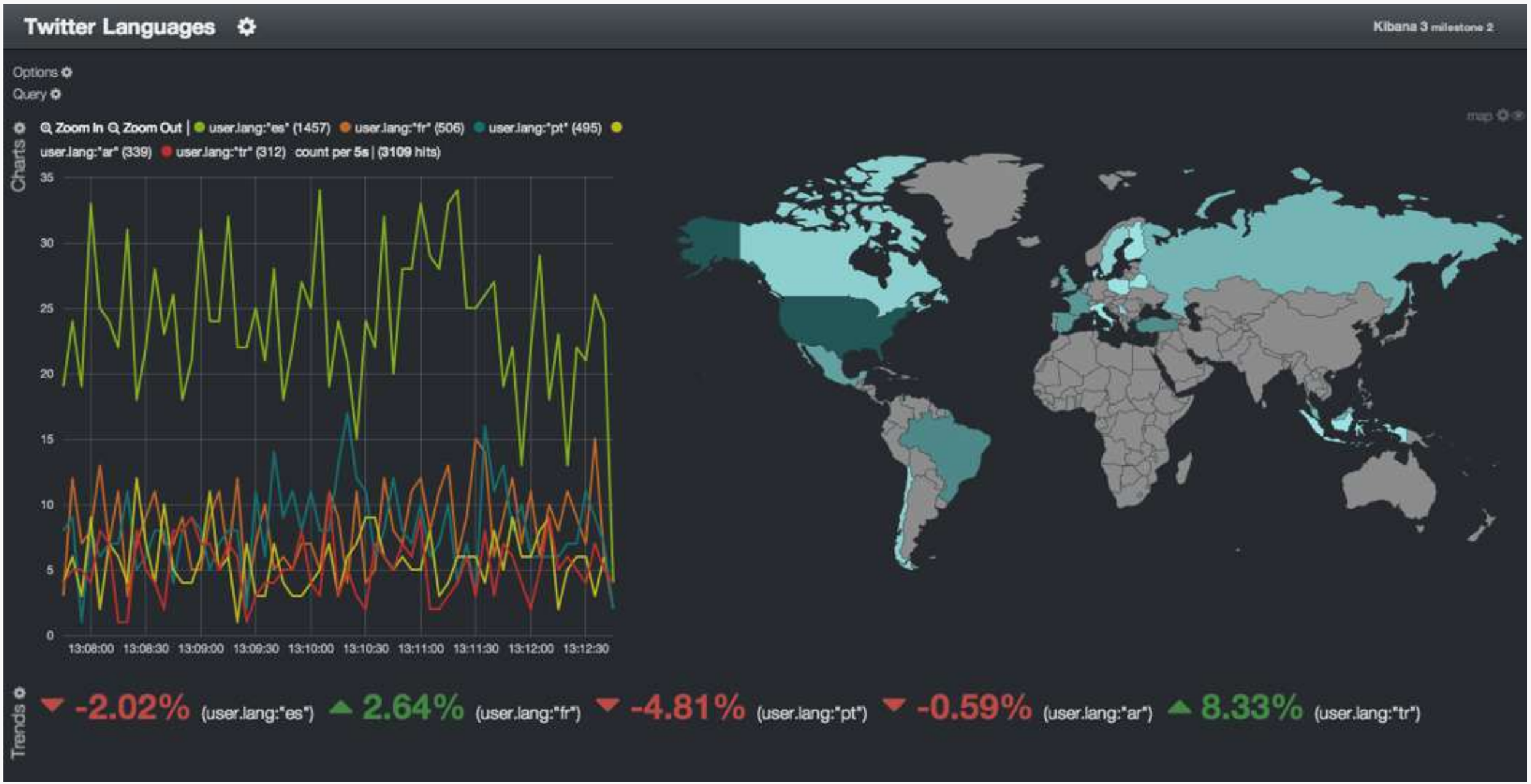
elasticsearch.

Kibana



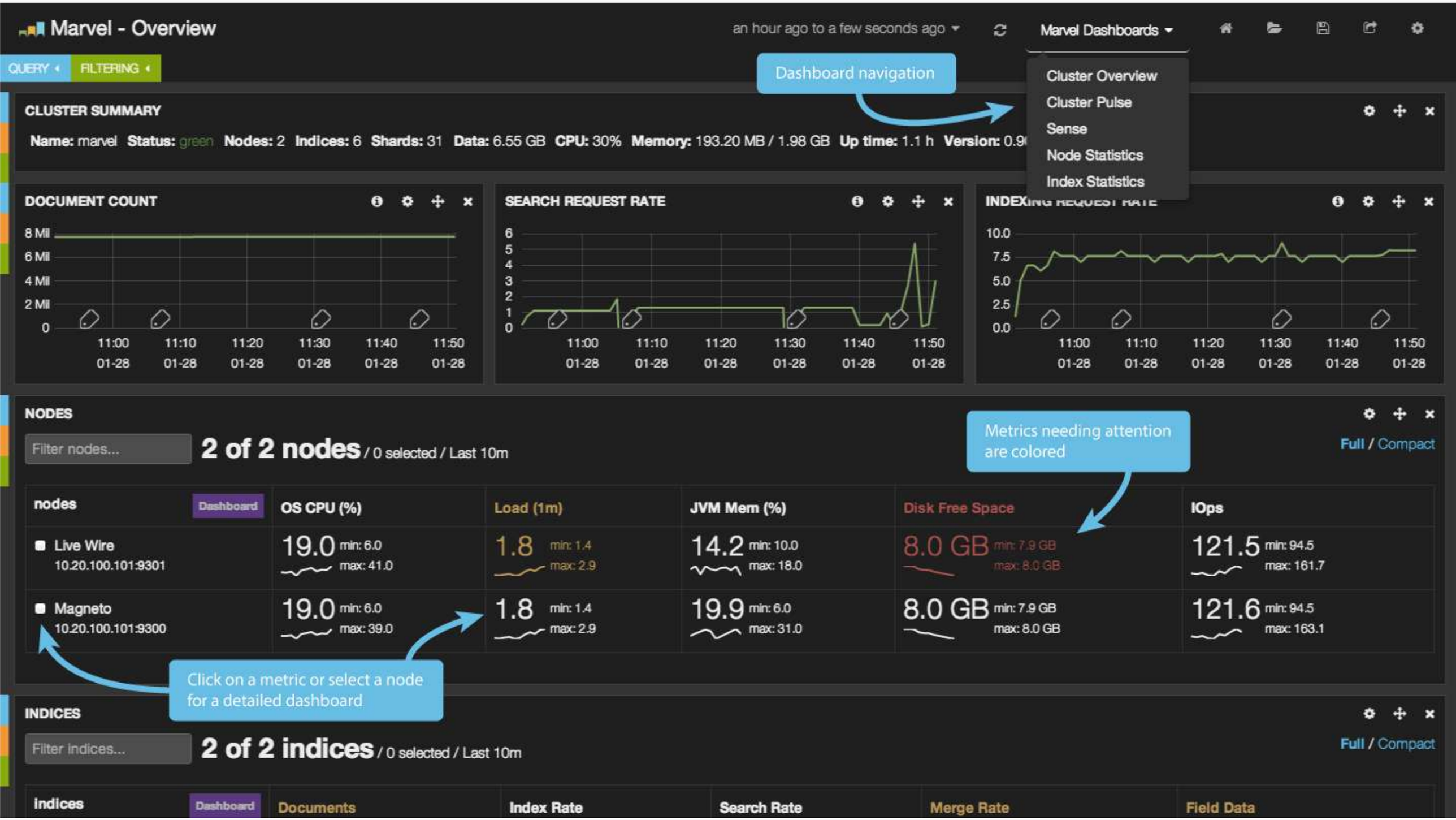
elasticsearch.

Kibana



elasticsearch.

Marvel Overview



elasticsearch.

Case Study

Case Study: The Guardian

- 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

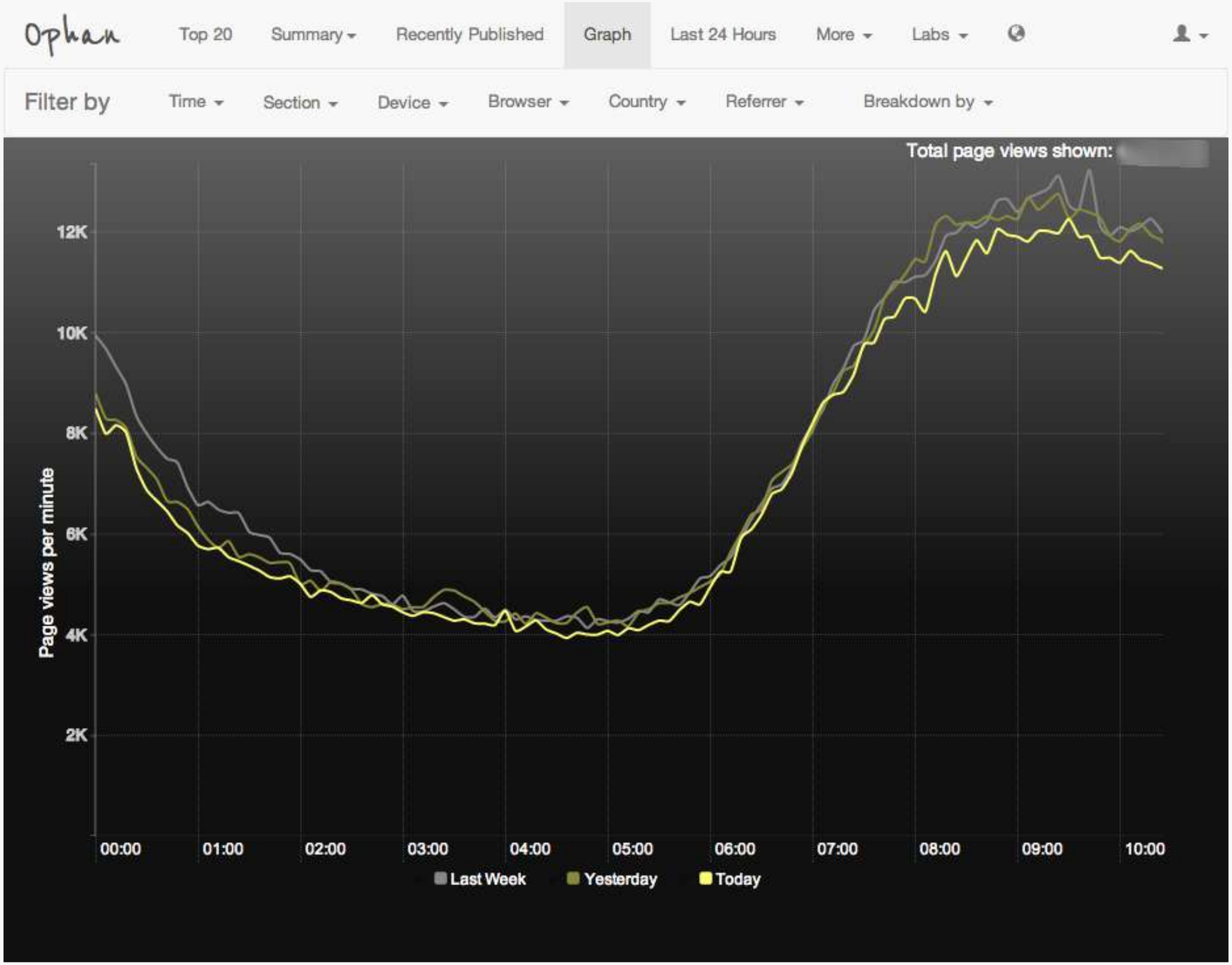
  Easily query 360 million documents

  See traffic for all content as it happens

  Gain insight into how updates impact site traffic

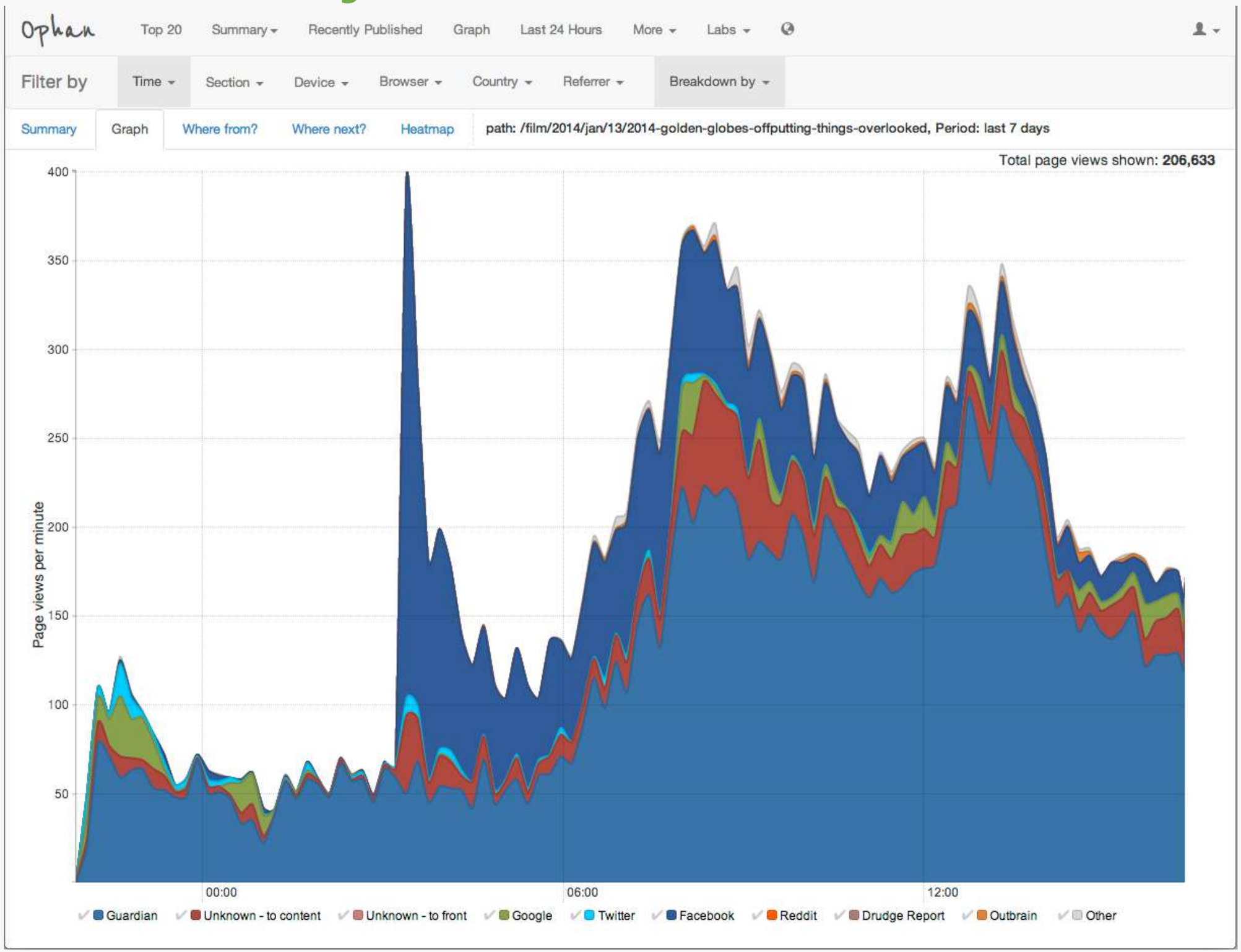
- http://www.elasticsearch.com/case-study/guardian/

Case Study: The Guardian



elasticsearch.

Case Study: The Guardian



elasticsearch.

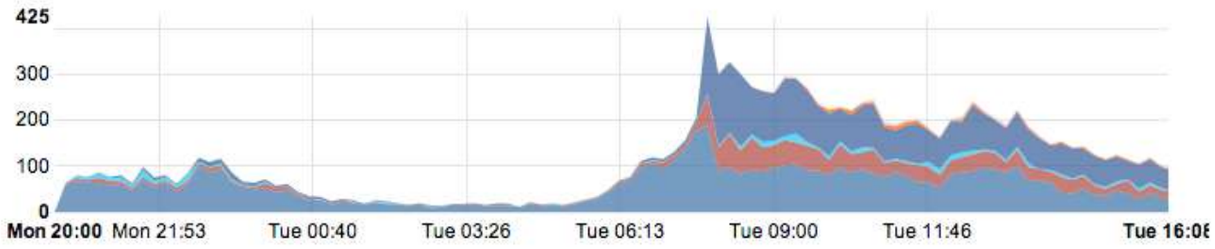
Case Study: The Guardian

Ophan | Top 20 | Summary | Recently Published | Graph | Last 24 Hours | More | Labs

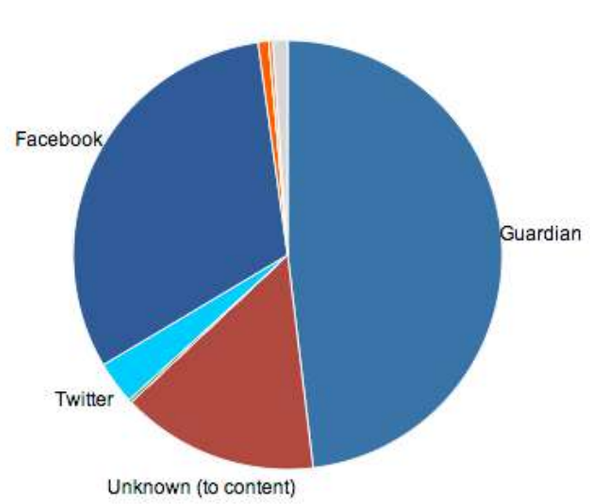
Dude, where's my North Sea oil money? | Aditya Chakrabortty

148,304 page views

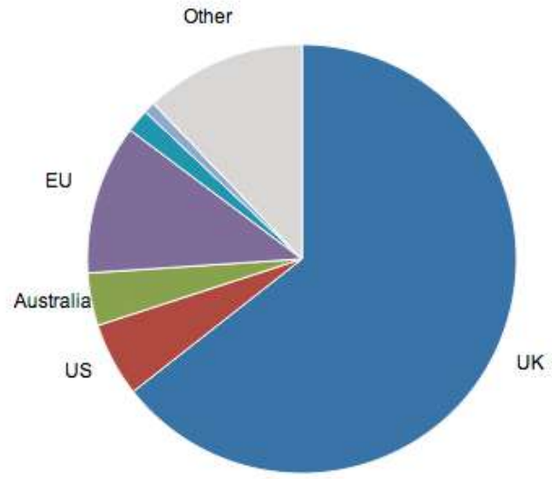
Posted on the Official Facebook Account



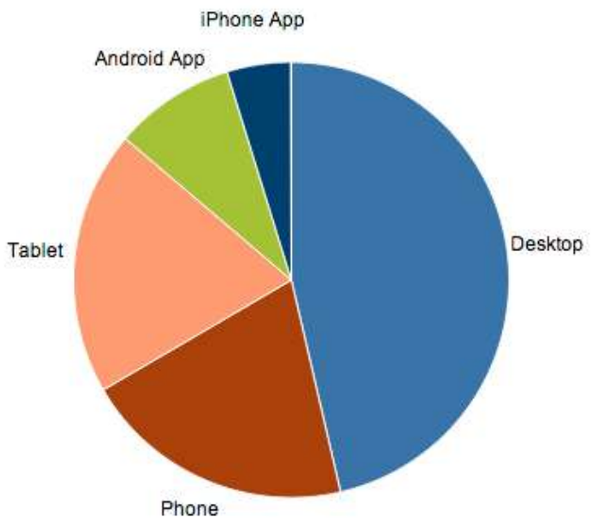
Referrers see all »



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)

elasticsearch.

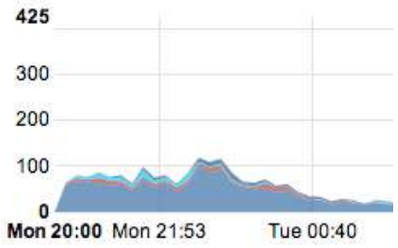
Case Study: The Guardian

Ophan | Top 20 | Summary | Recently Published | Graph | Last 24 Hours | More | Labs

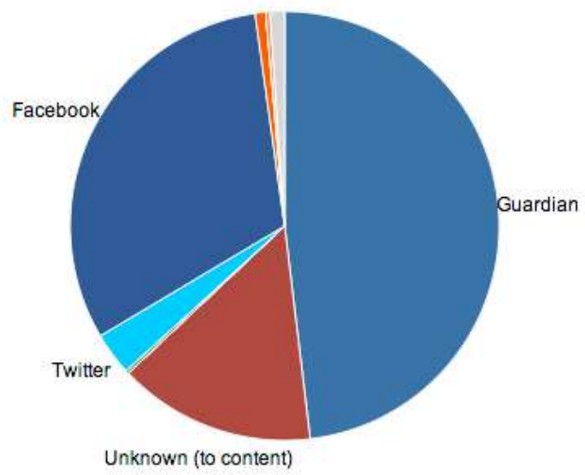
Dude, where's my North Sea oil money? | Ac

148,304 page views

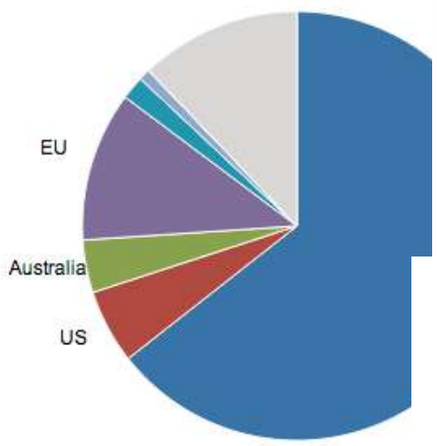
Posted on the Official Facebook Account



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**

elasticsearch.

Summary

Data driven decisions!

elasticsearch.

kibana

- Do not create data silos. Free your data!
- Make sure data is easy to query, not to store
- Visualize
- Give everyone the opportunity to query
- Reiterate
- Let the ELK stack help you to enable data driven decisions all across your company

Thanks for listening!

Q & A

P.S. We're hiring
http://elasticsearch.com/about/jobs
http://elasticsearch.com/support

Alexander Reelsen
@spinscale
alexander.reelsen@elasticsearch.com

elasticsearch.