

Why we built the ELK stack

Continuous improvement of your data to
achieve better business decisions

Alexander Reelsen
alexander.reelsen@elasticsearch.com

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

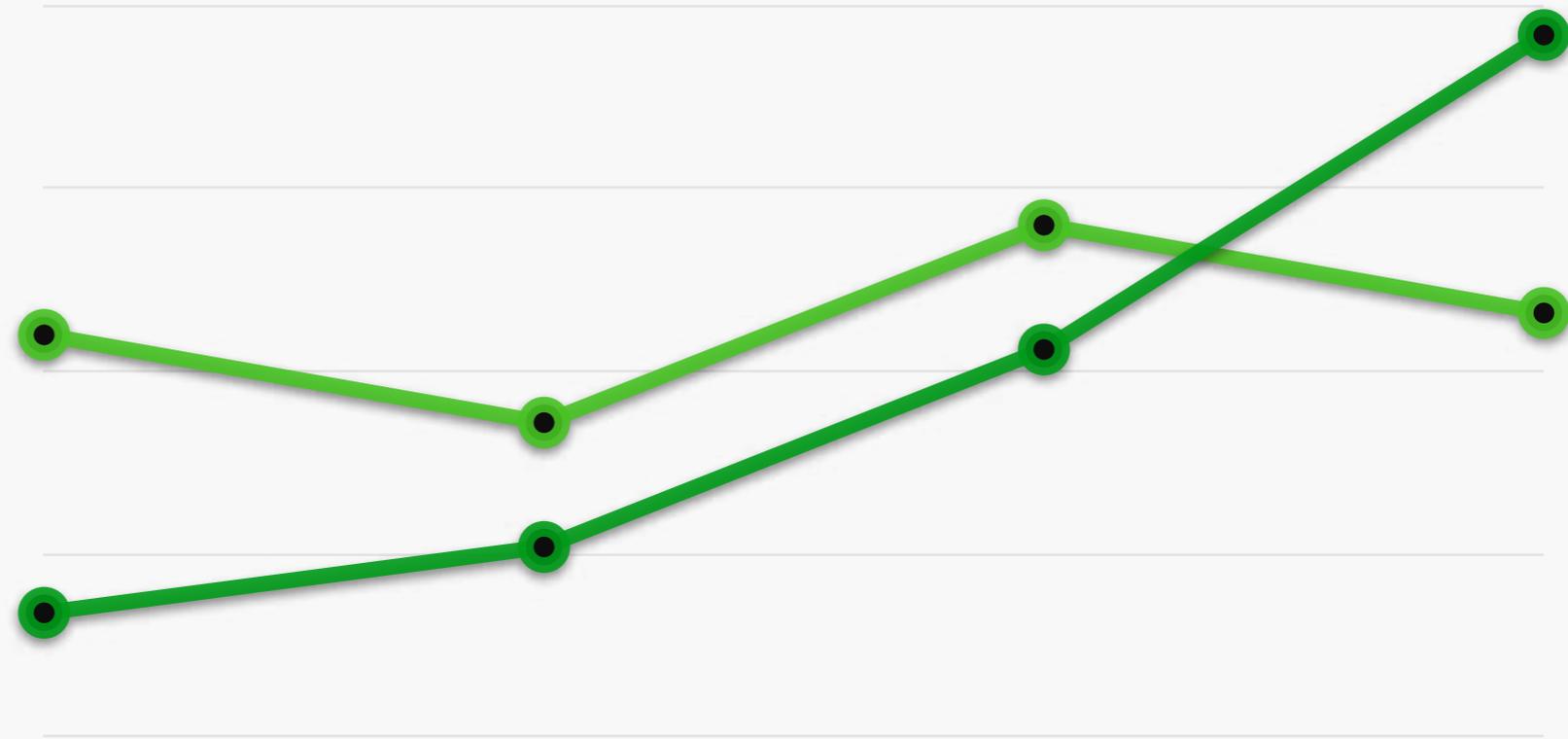
Elasticsearch

- 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

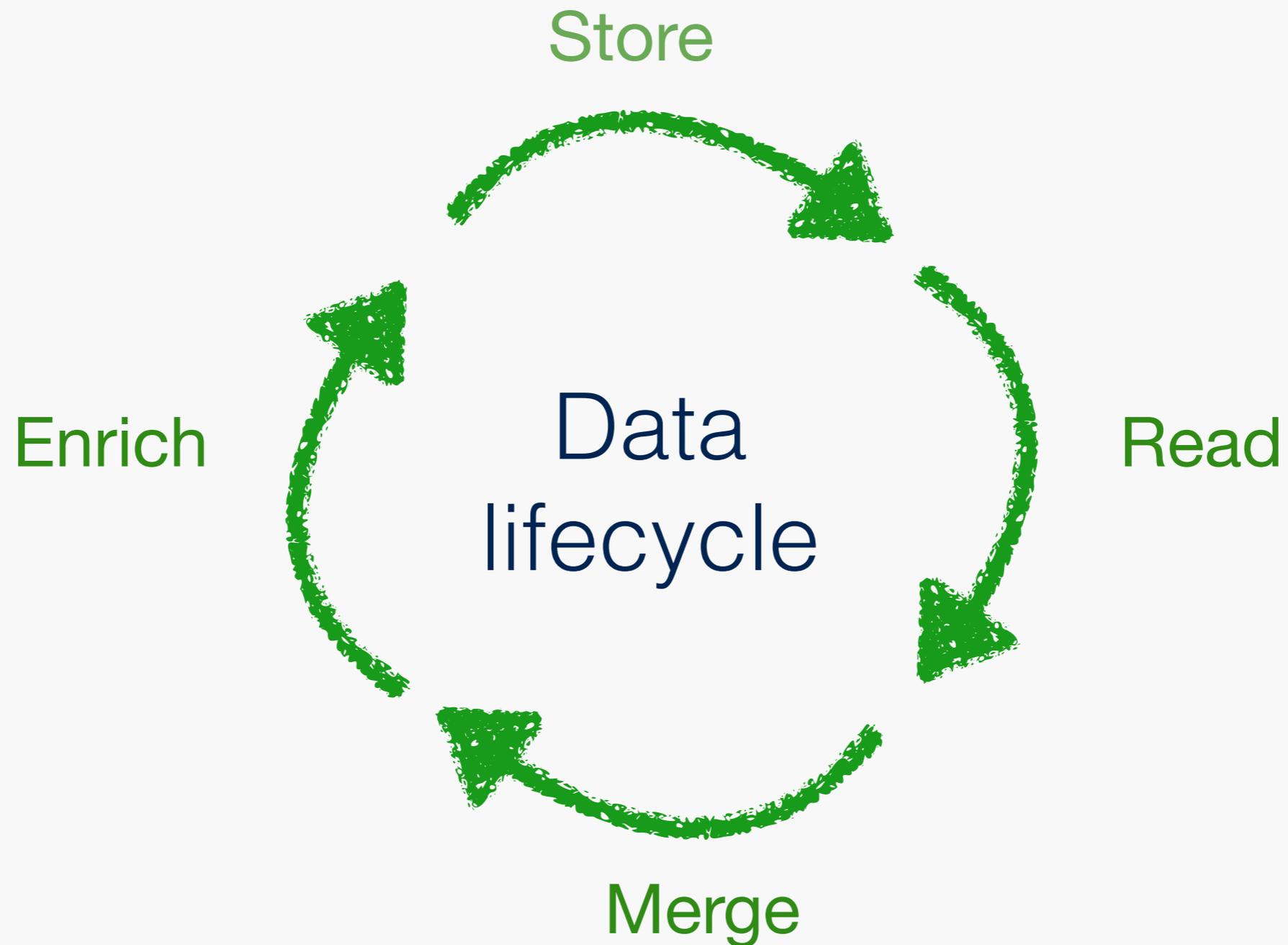
What is the core asset of your company?

- 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



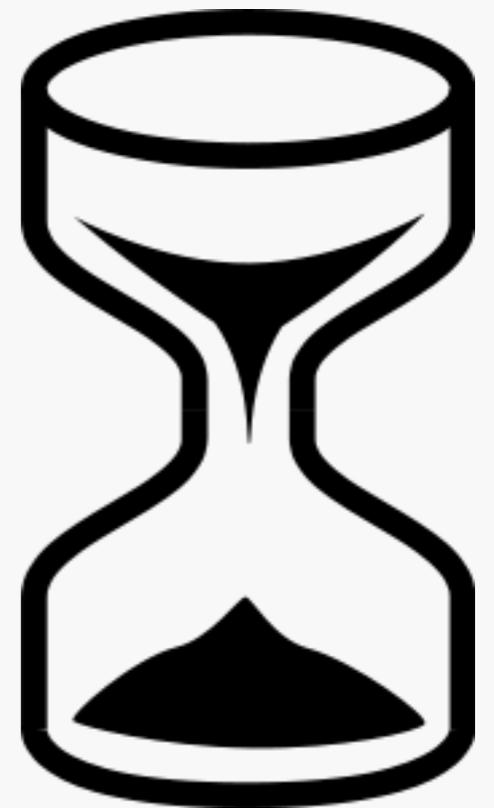
- **Examples**

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

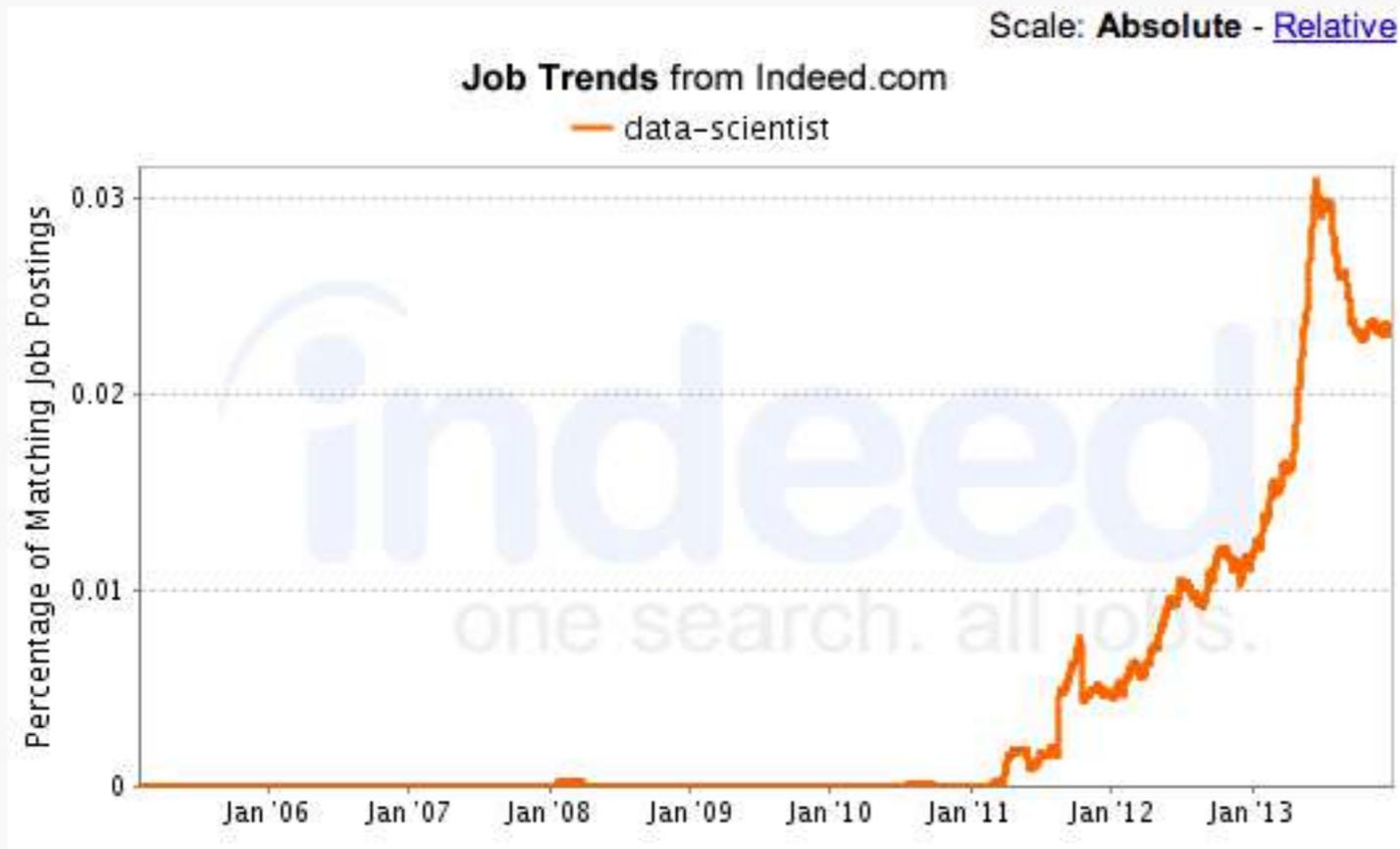
Fraud detection for payments

Storing data != insight

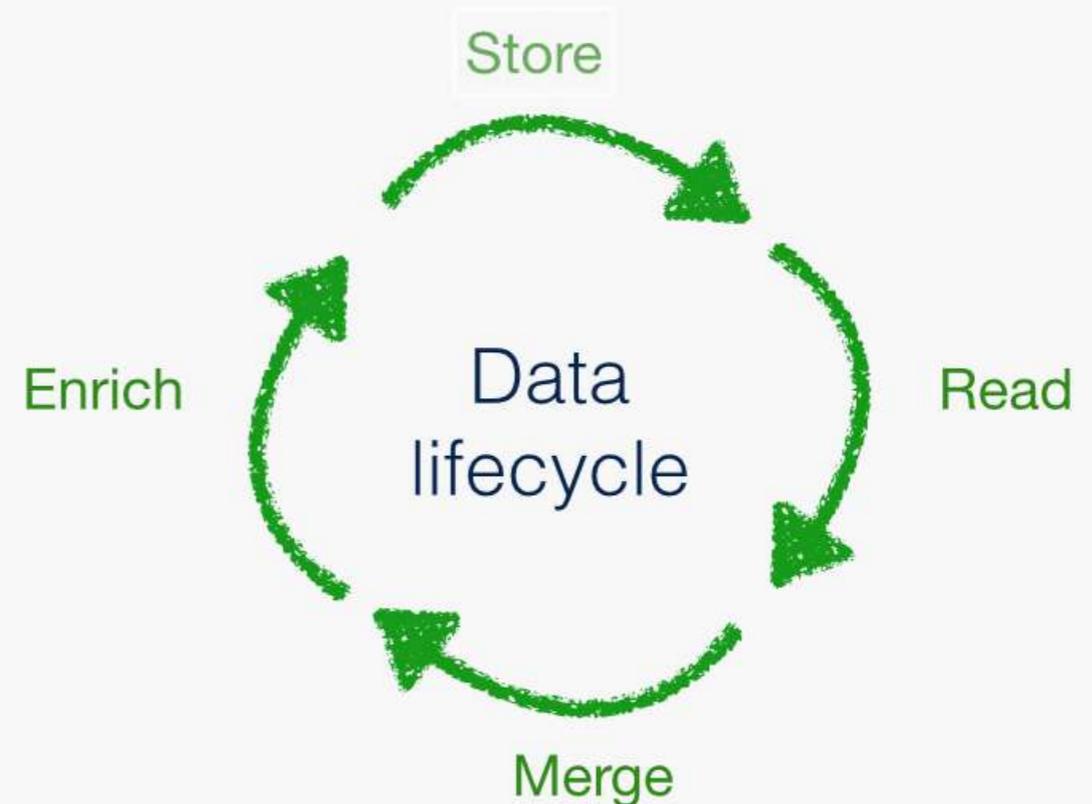
- Just because you are writing terabytes of data does not give you any value
- SQL example: We are trained to normalize our data as well as possible, until we denormalize it again to counter performance issues
- Data should never be optimised for writing, but always optimised for reading and information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



The data scientist fallacy



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



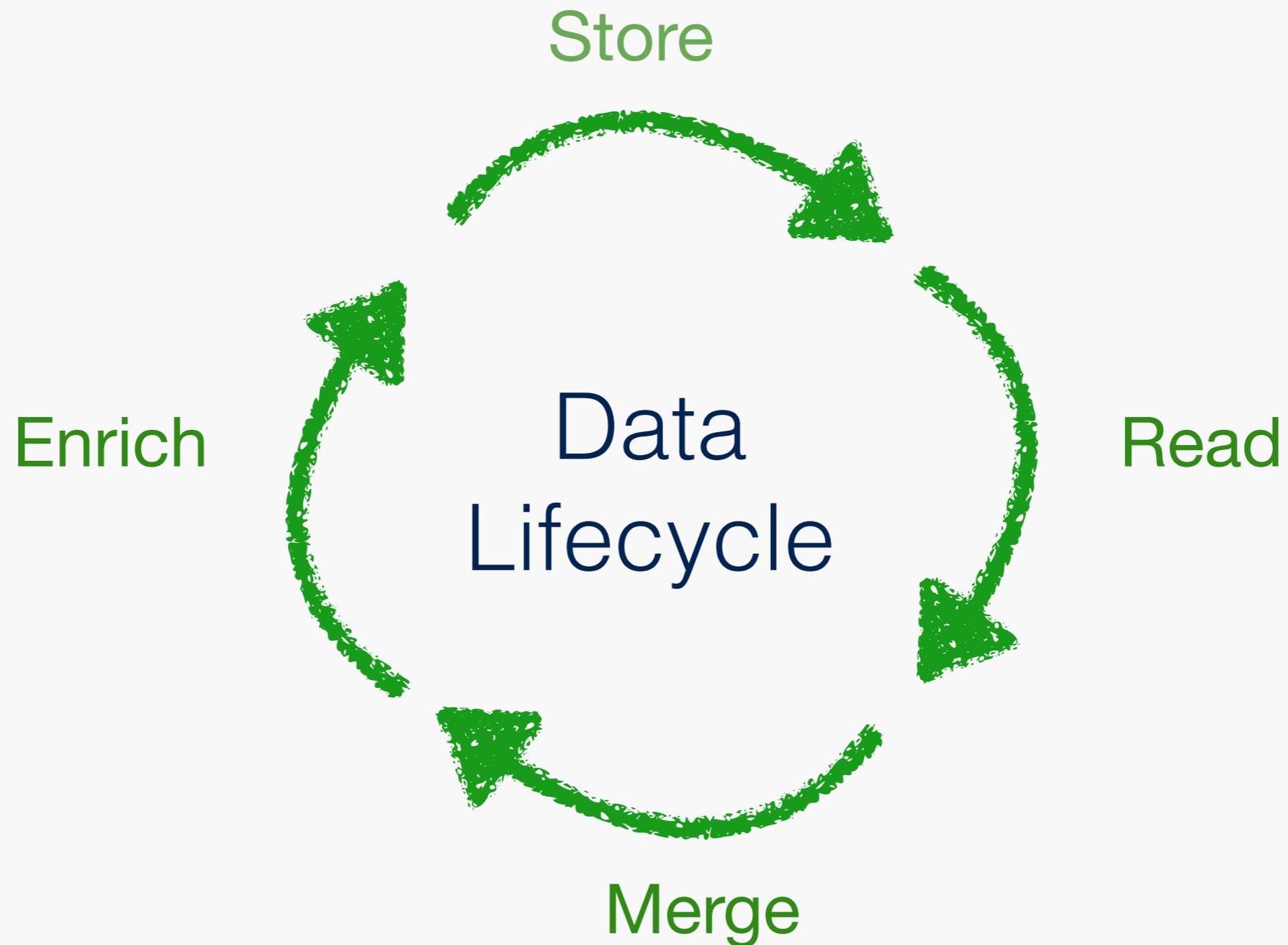
Do it yourself!

- 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

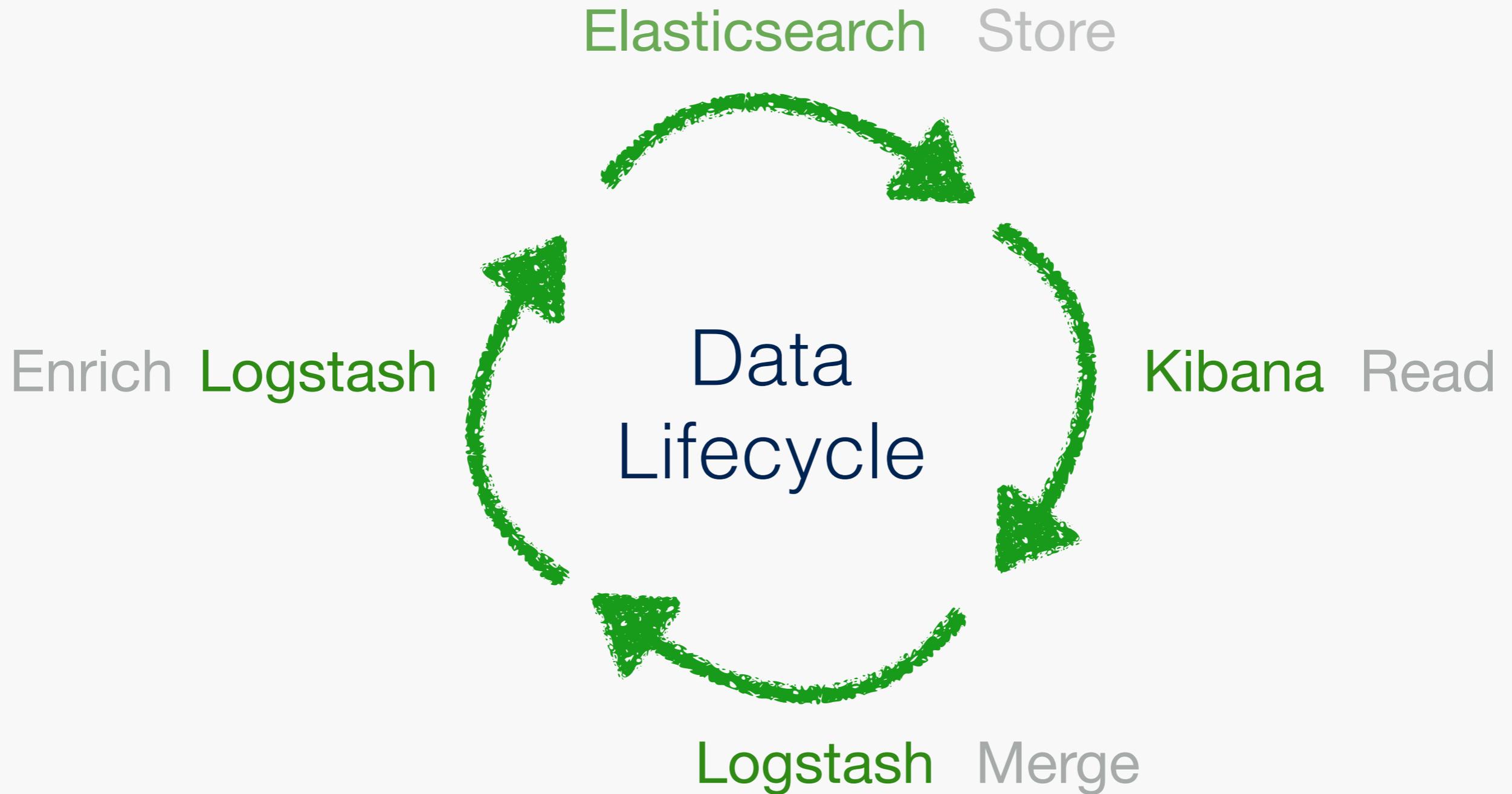
Meet the ELK stack

The ELK stack

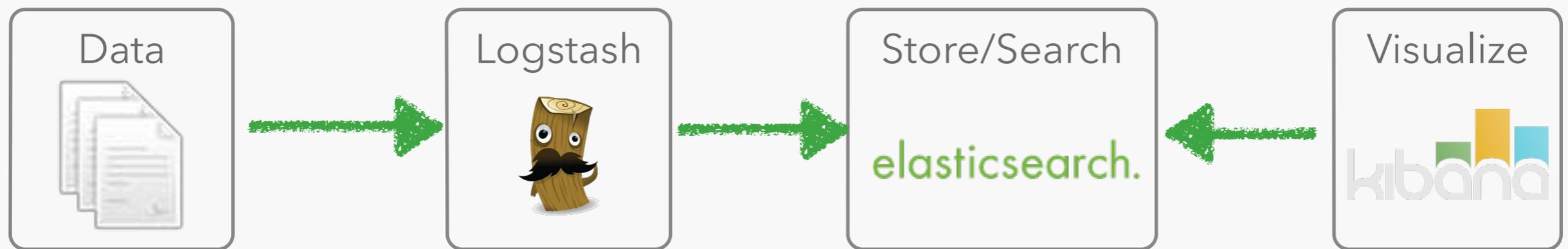
The ELK stack



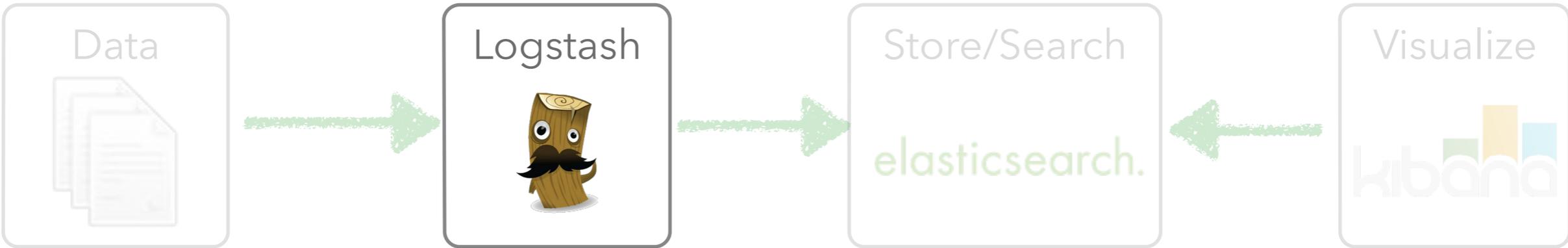
The ELK stack



The ELK stack



Logstash



Logstash

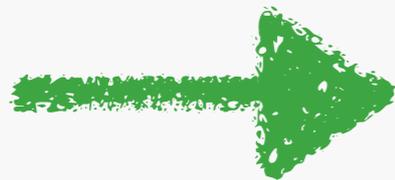
- Managing events and logs
- Collect data
- Parse data
- Enrich data
- Store data
- Open Source: Apache License 2.0



Logstash architecture

Input

datastore
stream
log files
files
monitoring
queues
network



Filter



parse, enrich, tag, drop

Output

datastore
files
email
pager
monitoring
chat
API
queues



Logstash architecture

Input

datastore
stream
log files
files
monitoring
queues
network



Filter



parse, enrich, tag, drop

Output

ip: 141.1.1.1
city: Zurich
country: CH



datastore
files
email
pager
monitoring
chat
API
queues

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



Mc
Graw
Hill

Fog Creek
SOFTWARE

foursquare



LiveChat.com

loggly



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?

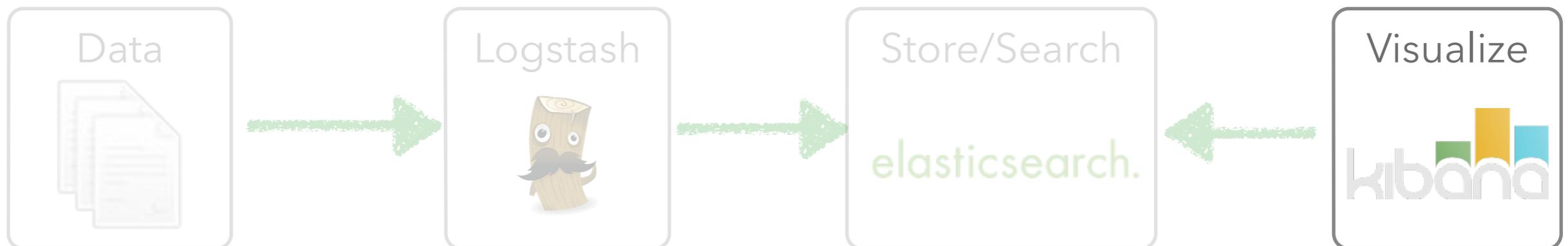
What is the average shopping cart size per order per hour?

- **Product portfolio**

Take the location of people into account for special offers?

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

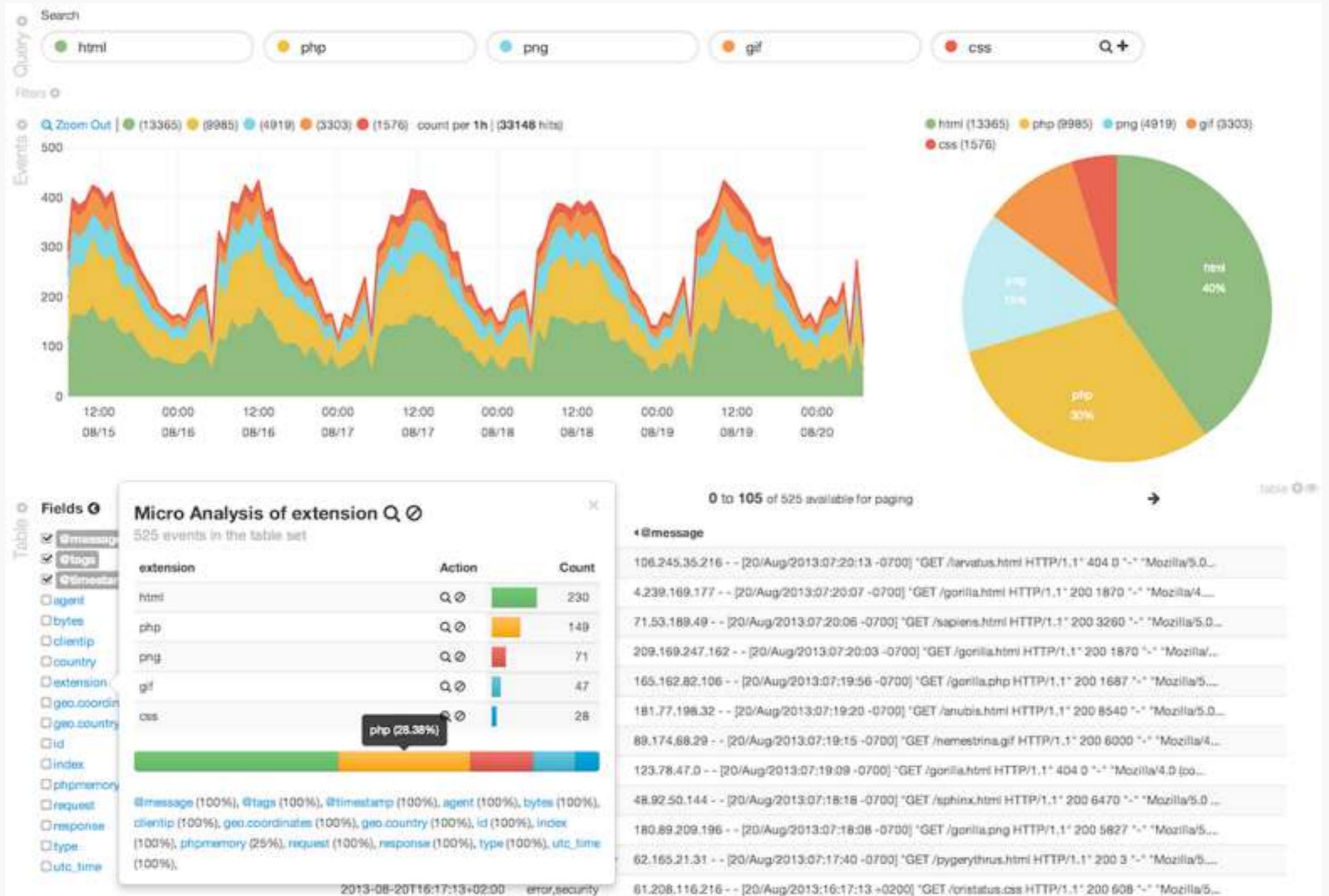
Kibana



Kibana

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

Kibana

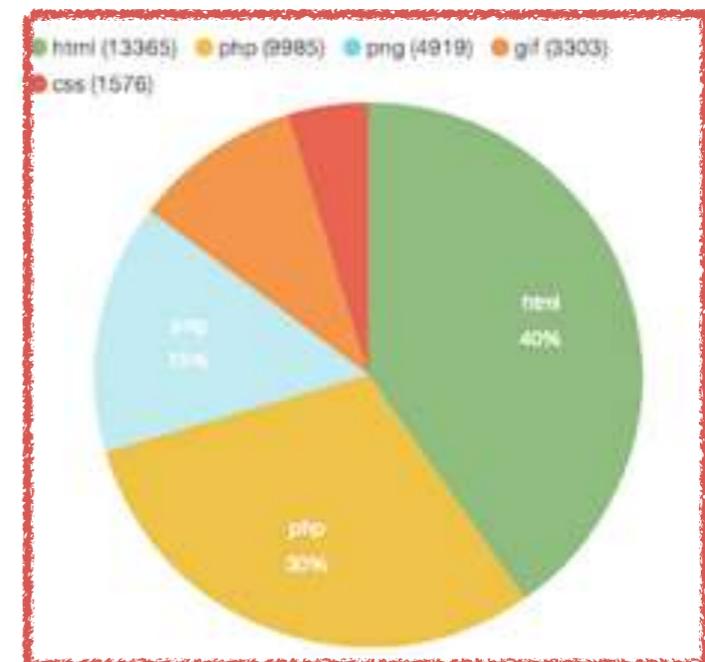
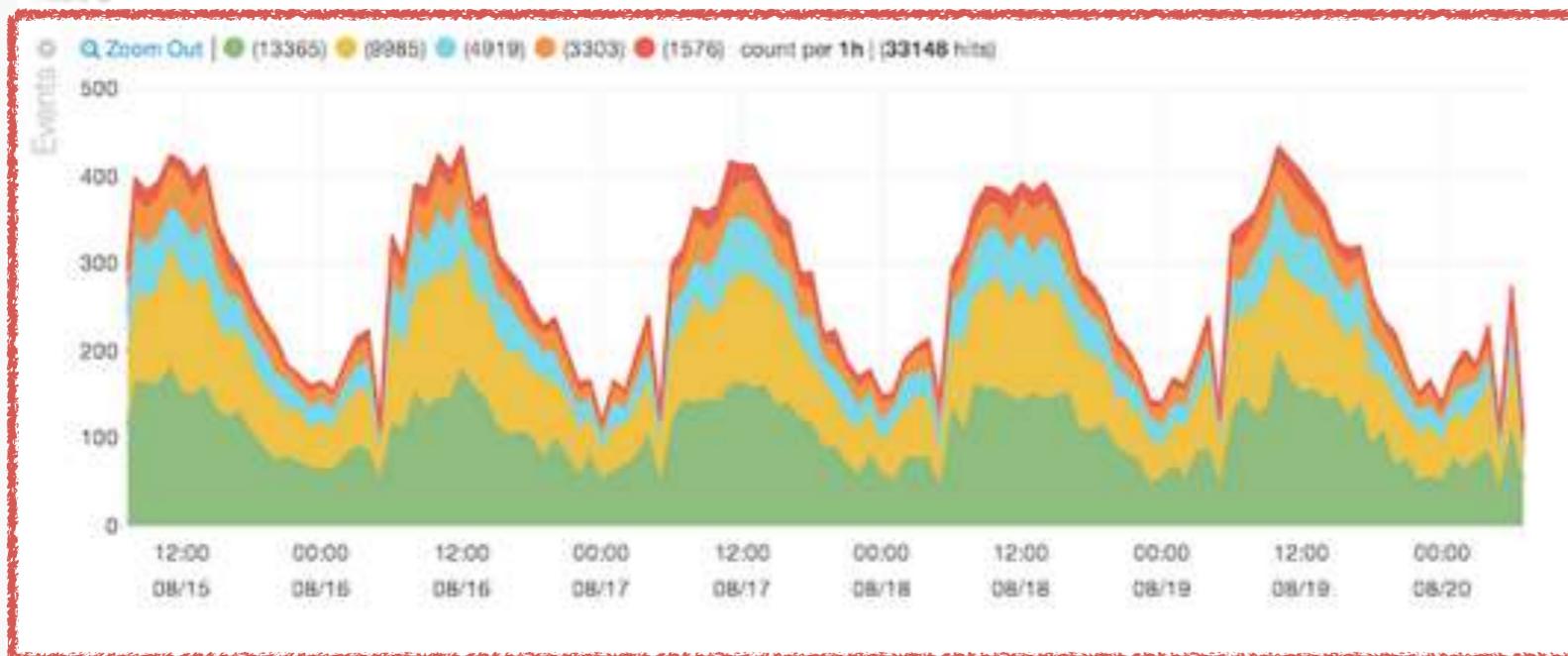


Kibana

Search

Query

html php png gif css Q+



Fields

Micro Analysis of extension Q

525 events in the table set

extension	Action	Count
html	Q	230
php	Q	149
png	Q	71
gif	Q	47
css	Q	28

php (28.38%)

4 @message

106.245.35.216 - - [20/Aug/2013:07:20:13 -0700] "GET /arvatus.html HTTP/1.1" 404 0 "-" Mozilla/5.0...

4.239.169.177 - - [20/Aug/2013:07:20:07 -0700] "GET /gorilla.html HTTP/1.1" 200 1870 "-" Mozilla/4...

71.53.189.49 - - [20/Aug/2013:07:20:06 -0700] "GET /sapiens.html HTTP/1.1" 200 3260 "-" Mozilla/5.0...

209.169.247.162 - - [20/Aug/2013:07:20:03 -0700] "GET /gorilla.html HTTP/1.1" 200 1870 "-" Mozilla/...

165.162.82.106 - - [20/Aug/2013:07:19:56 -0700] "GET /gorilla.php HTTP/1.1" 200 1687 "-" Mozilla/5...

181.77.198.32 - - [20/Aug/2013:07:19:20 -0700] "GET /anubis.html HTTP/1.1" 200 8540 "-" Mozilla/5.0...

89.174.68.29 - - [20/Aug/2013:07:19:15 -0700] "GET /nemestrina.gif HTTP/1.1" 200 6000 "-" Mozilla/4...

123.78.47.0 - - [20/Aug/2013:07:19:09 -0700] "GET /gorilla.html HTTP/1.1" 404 0 "-" Mozilla/4.0 (co...

48.92.50.144 - - [20/Aug/2013:07:18:18 -0700] "GET /sphinx.html HTTP/1.1" 200 6470 "-" Mozilla/5.0...

180.89.209.196 - - [20/Aug/2013:07:18:08 -0700] "GET /gorilla.png HTTP/1.1" 200 5827 "-" Mozilla/5...

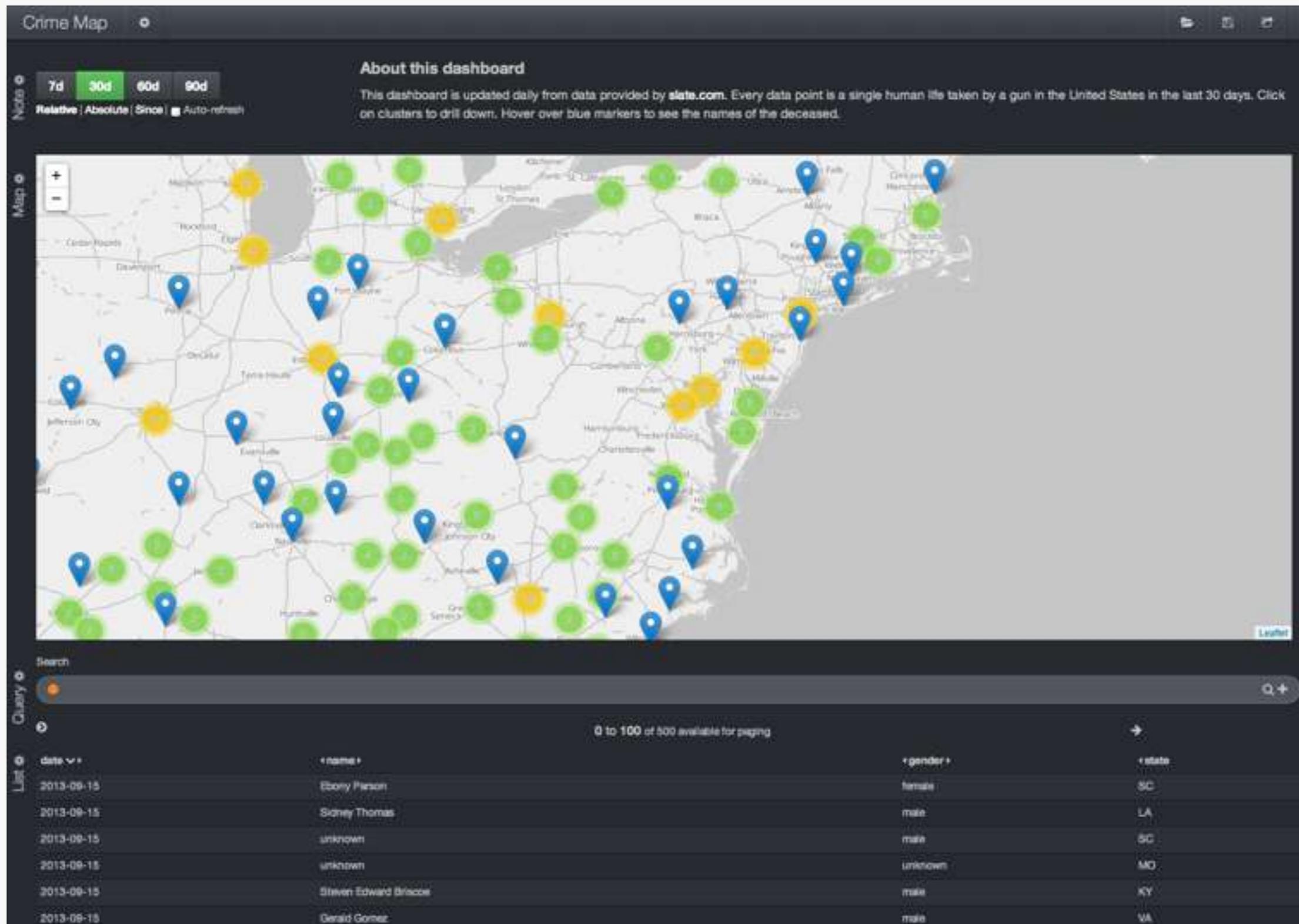
62.165.21.31 - - [20/Aug/2013:07:17:40 -0700] "GET /pyperythrus.html HTTP/1.1" 200 3 "-" Mozilla/5...

81.208.116.216 - - [20/Aug/2013:16:17:13 +0200] "GET /crisstatus.css HTTP/1.1" 200 608 "-" Mozilla/5...

Kibana



Kibana



Kibana



Marvel Overview

The screenshot shows the Elasticsearch Marvel Overview dashboard. At the top, there's a header with 'Marvel - Overview', a time range selector set to 'an hour ago to a few seconds ago', and a 'Marvel Dashboards' dropdown menu. The dropdown menu is open, showing options: 'Cluster Overview', 'Cluster Pulse', 'Sense', 'Node Statistics', and 'Index Statistics'. A blue arrow points from the 'Dashboard navigation' label to this menu.

Below the header is the 'CLUSTER SUMMARY' section, displaying: Name: marvel, Status: green, Nodes: 2, Indices: 6, Shards: 31, Data: 6.55 GB, CPU: 30%, Memory: 193.20 MB / 1.98 GB, Up time: 1.1 h, Version: 0.9.0.

The middle section contains three line graphs: 'DOCUMENT COUNT' (showing a steady increase from 0 to 8 Mil), 'SEARCH REQUEST RATE' (showing a fluctuating rate between 1 and 6), and 'INDEXING REQUEST RATE' (showing a fluctuating rate between 0.0 and 10.0). Each graph has a time axis from 11:00 to 11:50 on 01-28.

The bottom section is titled 'NODES' and shows '2 of 2 nodes / 0 selected / Last 10m'. It contains a table with columns: nodes, OS CPU (%), Load (1m), JVM Mem (%), Disk Free Space, and IOps. The 'Disk Free Space' column has a red header and red values, with a blue arrow pointing to it from the 'Metrics needing attention are colored' label. The 'IOps' column has blue values. A blue arrow points from the 'Click on a metric or select a node for a detailed dashboard' label to the 'Magneo' node.

Below the nodes section is the 'INDICES' section, showing '2 of 2 indices / 0 selected / Last 10m'. It contains a table with columns: Indices, Documents, Index Rate, Search Rate, Merge Rate, and Field Data.

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

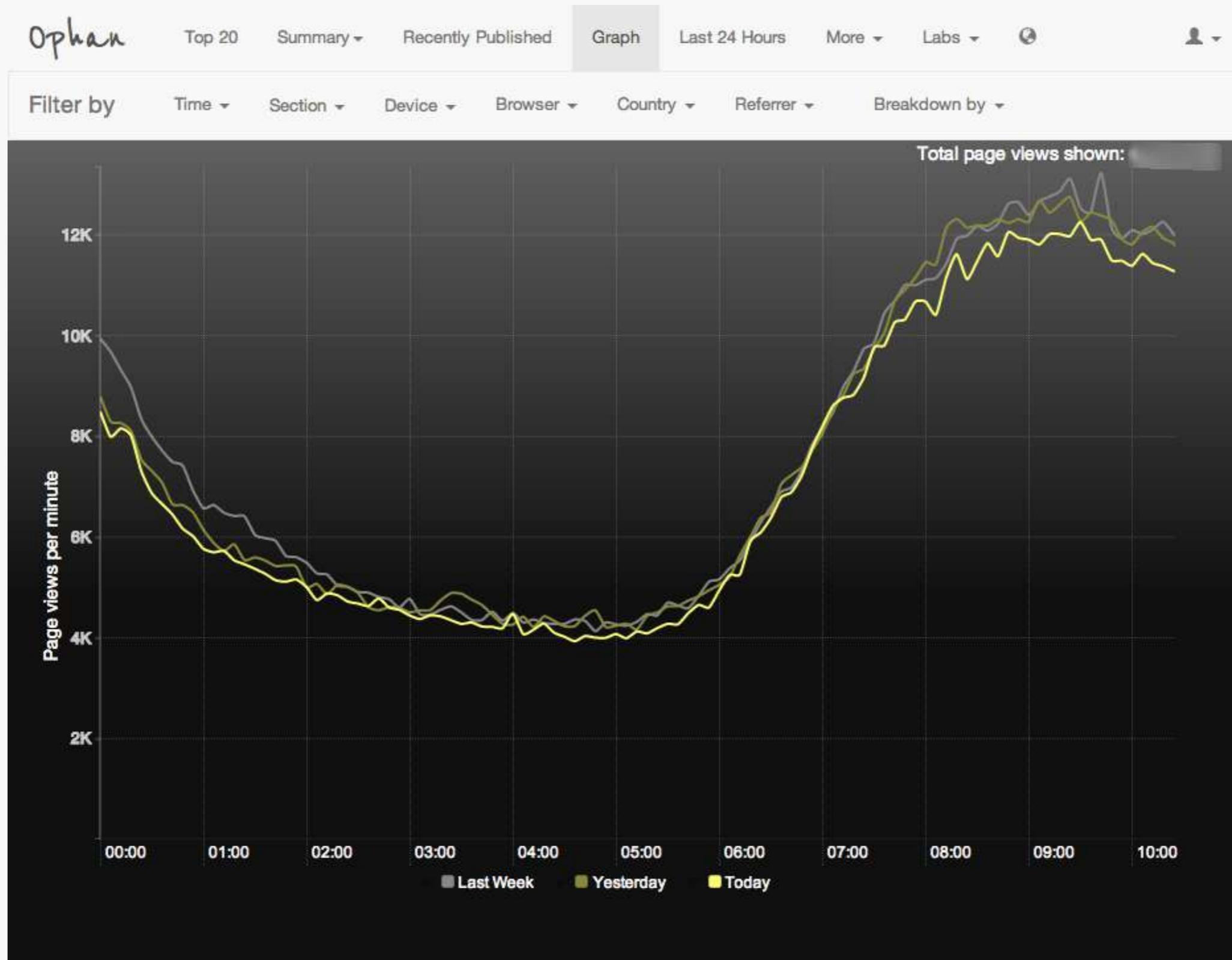
Easily query 360 million documents

See traffic for all content as it happens

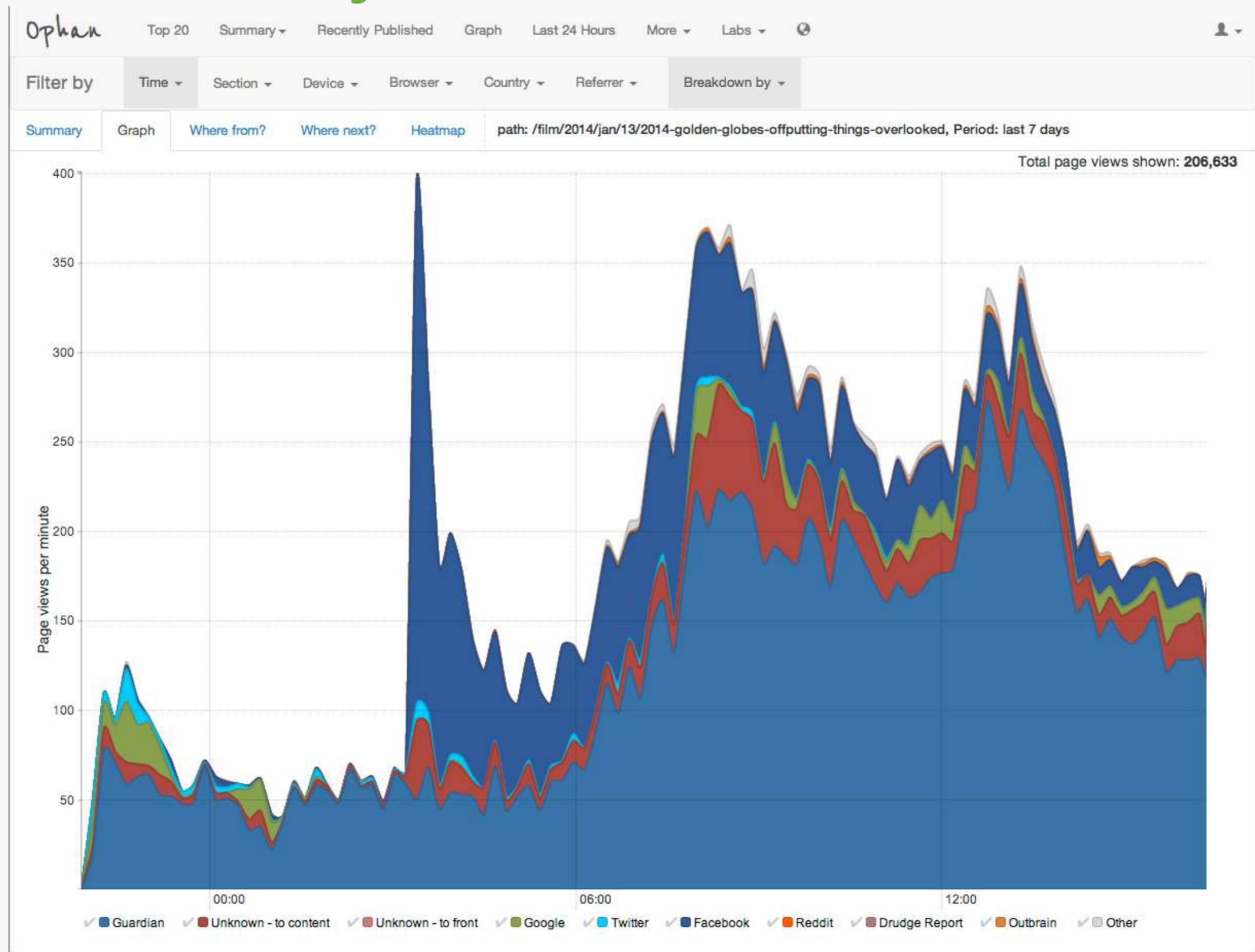
Gain insight into how updates impact site traffic

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

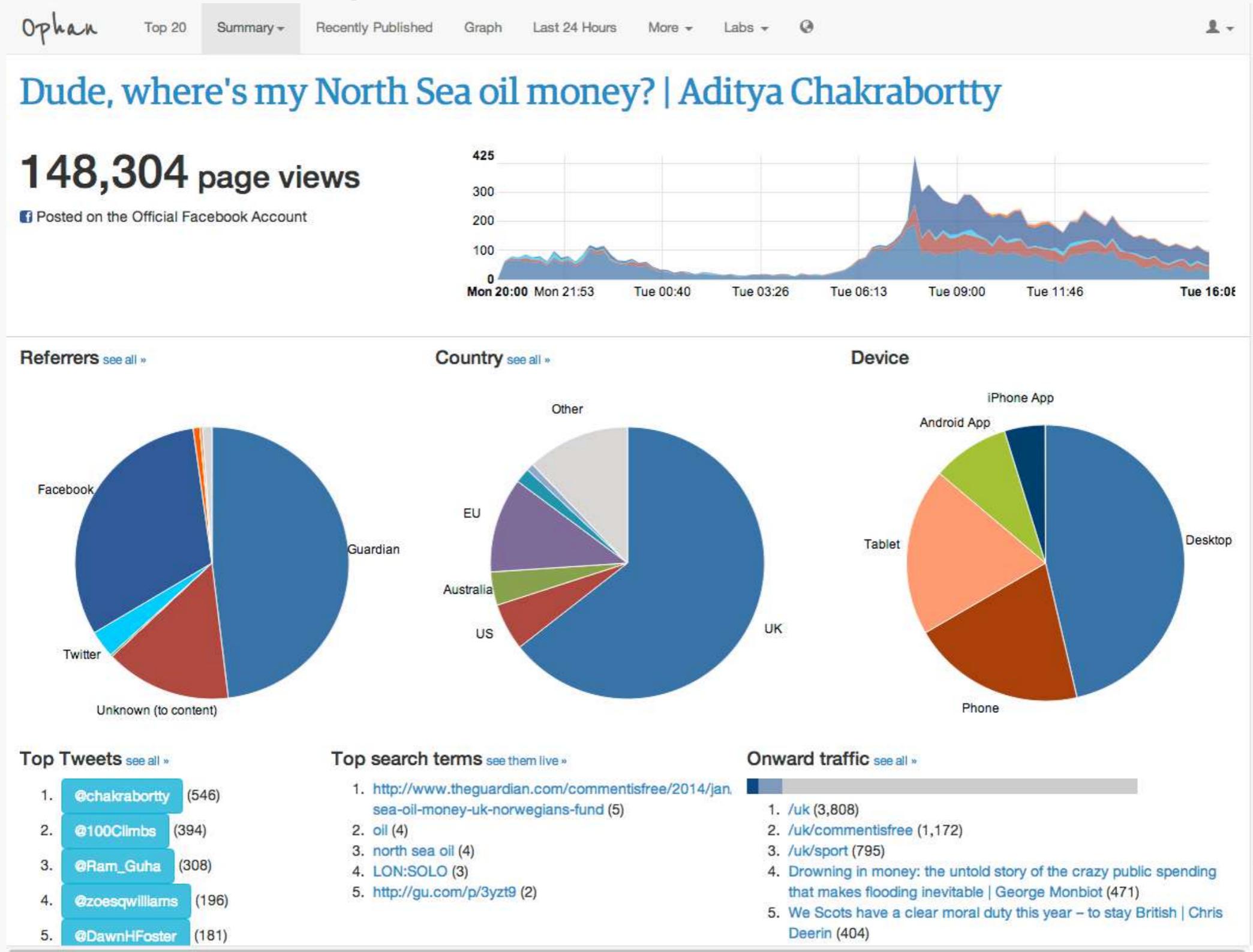
Case Study: The Guardian



Case Study: The Guardian



Case Study: The Guardian



Case Study: The Guardian



“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



Summary

Data driven decisions!

- 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



elasticsearch.



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.