

Please evaluate  
my talk via the  
mobile app!

# How Elasticsearch powers the Guardian's newsroom

**elasticsearch.**

shay banon ■ @kimchy  
creator, co-founder and cto  
elasticsearch

**the guardian**

graham tackley ■ @tackers  
director of architecture  
guardian news and media

**FREE Great fairytales**

Today *Wisdom and folly*  
Tomorrow *Justice and punishment*

the guardian

How much TV should your children watch?  
G2 cover story

Dear leader  
Sven in line to manage North Korea

# Oil firm drops bid to gag Guardian over MP's question

- Web users publish details of legal issue raised in parliament
- Labour member asked about injunction over toxic waste case



How's he doin'?  
Matt LeBlanc's new TV comedy

Aditya Chakraborty is Facebook destroying friendship?

PLUS

the guardian

Ministers cave in to City over bank bonuses

No 10 rejects calls to tackle highest earners as it seeks face-saving deal

Obama lead moment of address

Undercover officer knew of second spy



Exclusive

# The whistleblower

I can't allow the US government to destroy privacy and basic liberties

the guardian

Edward Snowden, 29, emerges from hiding in Hong Kong

IT contractor says his concerns were ignored and he had to go public



**FREE Audiobook download**

for *Charlie Stinky Socks and the Really Big Adventure* by Kristina Stephenson

From our exclusive new series of audiobook downloads. See page 30 for details. Tomorrow Down and Out in Paris and London by George Orwell

the guardian

Revealed: Murder bill for hiding dirty

Tory PR chief under fire over tabloid hacking

Politicians and celebrities among victims

the guardian

PRISM

M I F



# Black inquiry was abandoned to avoid upsetting Met

- Leaked memo warned police would 'deeply resent' probe
- Senior ex-officer says new claims must be investigated



I think it is one of the most significant media stories of the year. It was systematic. News of the world out of control.

The whole thing is deplorable. Mr Cameron should be out of office.



amazon kindle  
theguardian The Observer  
Syrian refugees in Turkey:  
'People see the regime is lying.  
It is falling apart'  
Bloodied but defiant, those who have fled  
the besieged towns of Jar al-Shughour  
damn the rule of Bashar al-Assad  
Martin Chabou in Antakya, southern Turkey | 11:00  
words

the guardian  
9:41 AM  
Issues Sections Help  
Drought in east Africa  
the result of climate  
change and conflict  
Aid agencies say that weather in the  
region has become more erratic and  
leave populations especially vulnerable  
Wembley night to  
savour as NFL descends  
on London  
Funding triples for  
mental health  
research  
the guardian  
iPad edition  
Sport Wimbledon  
Financial

htc  
11:18  
football  
5 new updates  
Roy Hodgson welcomes his previous employers,  
Liverpool, to West Bromwich on Saturday, having  
masterminded a 2-1 victory in the equivalent fixture  
last season. Photograph: Ian Kington/AP/Getty  
Images  
Simon Burton  
Fri 28 October 2011  
Check for updates automatically  
11.46am:  
There's a statement up on Plymouth's

12:00  
theguardian  
Latest Trending  
Artists make HIV  
through  
World news • Aids and HIV  
Government gives go  
to high speed rail  
Support policy • Politics  
and lows at Milan  
week  
Style • Fashion  
Home Favourites Galleries Audio More

theguardian  
beta  
fashion

Osborne

quiz  
active houses

Sport

Ten things to watch

## The Scott Trust

**The Scott Trust is a unique form of media ownership in the UK**

The sole shareholder in Guardian Media Group, the Trust was created in 1936 to safeguard the journalistic freedom and liberal values of the Guardian. In 2008 it became a limited company, with the same protections for the Guardian enshrined in its constitution. Its core purpose is to secure the financial and editorial independence of the Guardian in perpetuity.

Rather than benefiting shareholders or a proprietor, GMG's profits are reinvested to sustain journalism that is free from commercial or political interference, and to uphold a set of values laid down by CP Scott, the great Manchester Guardian editor.

The remit of the Trust's charitable arm, the Guardian Foundation, reflects the Trust's aims of promoting press freedom and meeting its responsibilities to the community.



“created in 1936 ... to secure the financial and editorial independence of the Guardian in perpetuity”

elasticsearch.

# the guardian

# Ophan Dashboard

What's going on right now?

Interested in a specific page?

Paste the url here and hit 'Go!'

Go!

or drag **View in Ophan** to your bookmark bar and click it when viewing any Guardian page

Welcome to Ophan, the Guardian's in-house, real-time traffic tool.

Get started with a [graph of total traffic](#) or the [top 20](#). Most pages in Ophan can be filtered or broken down in a number of ways, from length of time to editorial section, user location, referrers, desktop or mobile site and even browser. Use the "Filter By" bar at the top of the page to get started.

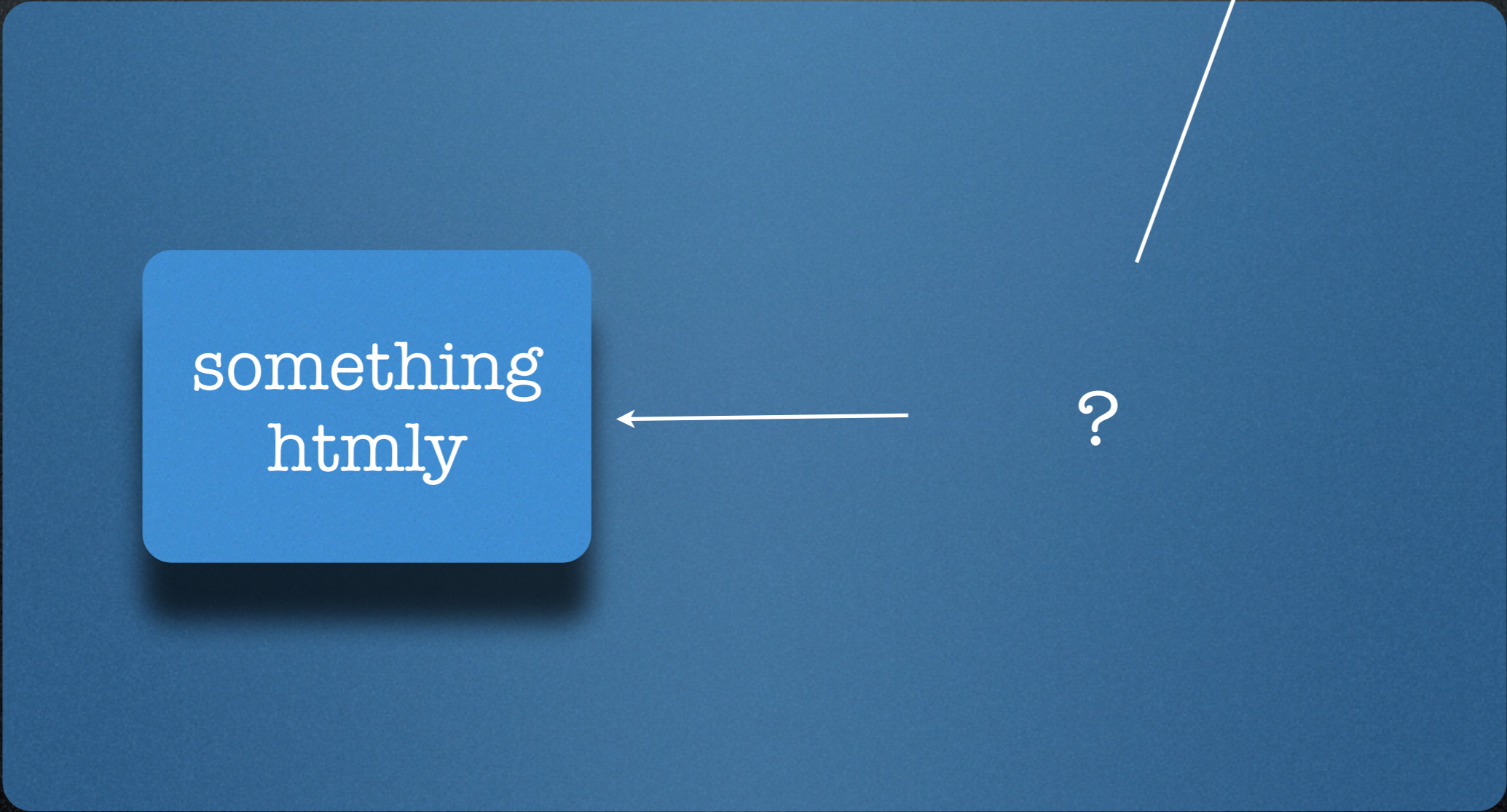
Other pages you might find interesting are the [recently published page](#), the [last 24 hours](#), [live Google search terms page](#). Just explore the options at the top of the screen.

We want Ophan to help you make useful decisions about your content so let us know what you find useful and what you'd like to see added. Bear in mind that we have [Site Catalyst](#) for longer-term data and ophan isn't meant to replicate that. But we'd love to hear your ideas. [You can find our contact details and read more about how Ophan was created and how it got its name here.](#)

our in-house real-time traffic tool



my desktop  
workstation



production  
apaches

something  
htmlly

?

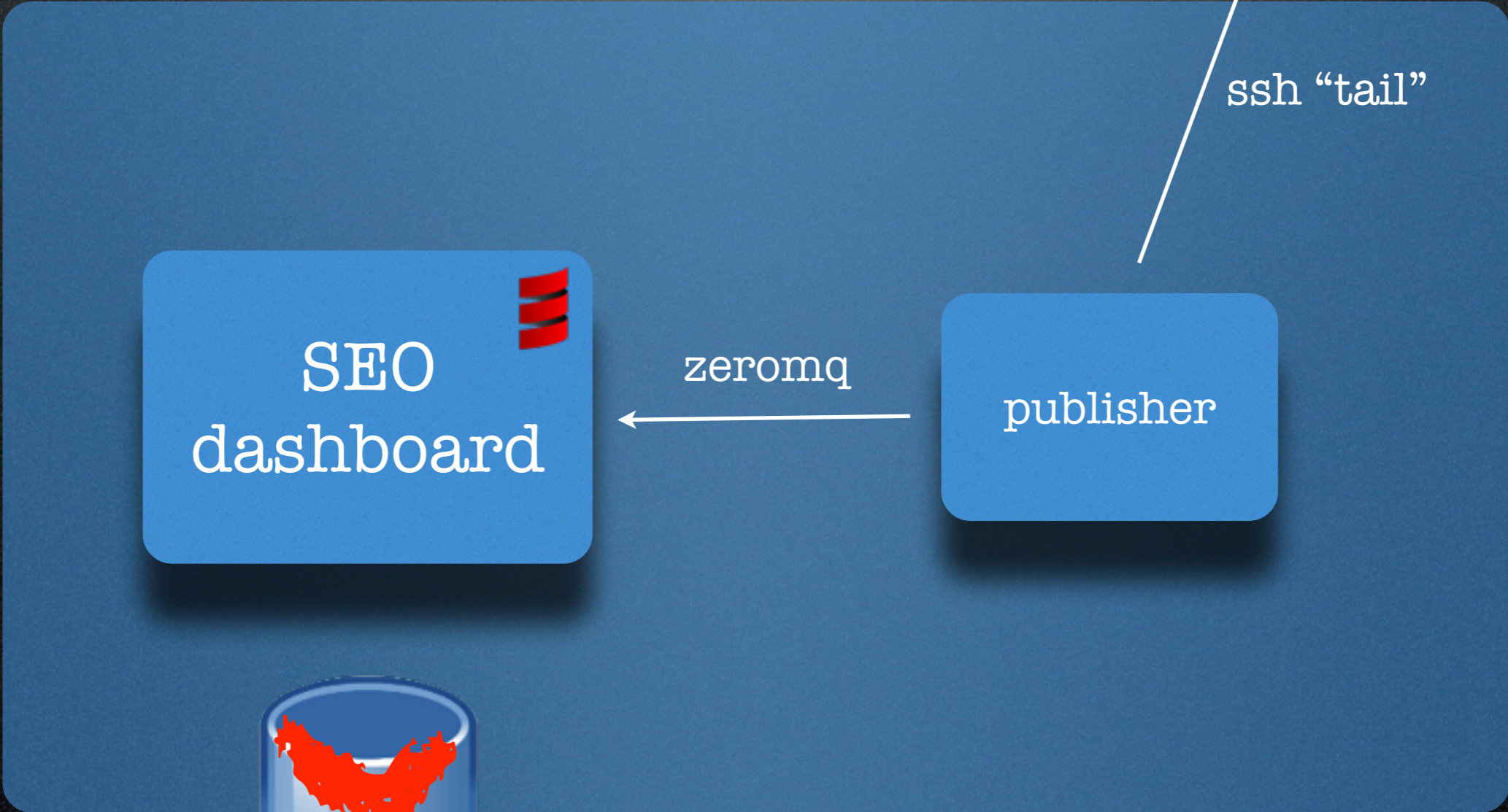


```
ssh $SERVER "nice tail -f /apache2/logs/guardian-access_log"
```

my desktop  
workstation



2 x production  
apaches



SEO  
dashboard



zeromq

publisher

ssh "tail"



**Top 10** from 26 Sep 2013 20:55:20 to 26 Sep 2013 20:58:20 (BST)

## The US promised a visa to my Afghan interpreter. Now it's been revoked | Matt Zeller

576

• 66% from www.reddit.com • 24% from (unknown) • 5% from www.theguardian.com • 2% from native-app.guardian.co.uk

## José Mourinho 'cried' after being overlooked for Manchester United job

287

• 56% from www.theguardian.com • 19% from native-app.guardian.co.uk • 10% from (unknown) • 6% from t.co  
• 2% from m.bleacherreport.com

## Interpol issues international arrest warrant for Samantha Lewthwaite

161

• 62% from www.theguardian.com • 31% from native-app.guardian.co.uk • 4% from (unknown)

## Revealed: Qatar's World Cup 'slaves'

157

• 21% from (unknown) • 17% from www.facebook.com • 16% from m.facebook.com • 10% from www.theguardian.com  
• 3% from native-app.guardian.co.uk

## Pasta firm Barilla boycotted over 'classic family' remarks





my ~~desktop~~  
workstation



Javascript in browser

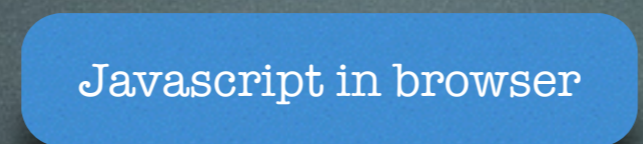
hidden pixel

Tracker

SNS

SQS

Dashboard







# elasticsearch

flexible and powerful open source, distributed real-time search and analytics engine for the cloud

cool. bonsai cool

[SEE THE OVERVIEW](#)



Javascript in browser

hidden pixel

Tracker

SNS

SQS

SQS

Serf

Dashboard

elasticsearch

Dashboard

Overview [Browser](#) [Structured Query \[+\]](#) [Any Request \[+\]](#)

	ads	content	facebook	kibana-int	ophan-2014-02-25	ophan-2014-02-26	ophan-2014-02-27	ophan-2014-02-28	ophan-2014-03-01	ophan-2014-03-02	ophan-2014-03-03	ophan-2014-03-04	ophan-2014-03-05	ophan-2014-03-06	ophan-2014-03-07
size:	4.09M (12.2M)	98.9M (297M)	7.43M (18.9M)	174k (347k)	17.2G (34.5G)	17.1G (34.3G)	17.6G (35.2G)	17.8G (35.5G)	15.5G (30.9G)	18.1G (36.1G)	21.5G (43.0G)	20.0G (40.0G)	18.9G (37.7G)	13.6G (27.1G)	891B (1.56k)
docs:	17,661 (18,027)	93,709 (93,938)	4,925 (5,057)	16 (16)	74,537,915 (74,543,873)	74,747,073 (74,752,706)	74,088,774 (74,094,771)	69,641,335 (69,645,792)	58,720,858 (58,724,227)	69,112,564 (69,117,449)	84,470,896 (84,478,138)	78,253,806 (78,257,987)	72,844,310 (72,864,157)	52,461,492 (52,470,934)	0 (0)

● <b>Beyonder</b>	0	1	2	2	1	4	2	5	5	4	3	1	1	1	0	4	7
● <b>Blindside</b>	2	3	1				1		4	1	4	6	5	2	3	1	6
● <b>Cardiac</b>	2			2	0	0	0	3	0	8	1	7	8	4	2	0	3
● <b>Electron</b>	1	2		1	0	3	0	3	2	2	2	0	3	2	0	3	0
● <b>Foolkiller</b>	0	1	2	1		1		5	4	3						7	4
● <b>Human Robot</b>		3	0	0		2		1	2		6	0	6	0	3	2	5
● <b>Jacob "Jake" Fury</b>			2	1	2	5	2	0	0	3	1	4	2	5	2	2	5
● <b>Johnny Storm</b>	4	5	0	1		5	4	5	3	4	3			7	8	6	8
● <b>Katherine Reynolds</b>	5	7	2				7	0	7	7	5	1	7	3	2	7	4
● <b>Nebulos</b>	4	6		2	2	3		6	3	1	1	7	4	5	4	6	8
● <b>Paul Bailey</b>	0	1	1	0	1	4	4	3	5	2	8	0	3	1	4	1	4
★ <b>Professor Power</b>		3	0	0		4	0	3	1	8	0	2		7	6	0	7

12 \* m3.xlarge

instance store (SSD)

in an autoscaling group  
(with manual scaling)

ophan-elasticsearch

**Warning!** The following autoscaling activities are currently suspended: Terminate

Instance	AutoScaling	Uptime	Version
i-b13ed8ff	InService	37 days	1.0.0 Cardiac
i-bf3ed8f1	InService	37 days	1.0.0 Human Robot
i-b33ed8fd	InService	37 days	1.0.0 Blindside
i-ab3ed8e5	InService	37 days	1.0.0 Paul Bailey
i-3a3bc27b	InService	35 days	1.0.0 Jacob "Jake" Fury
i-4ee12e0f	InService	35 days	1.0.0 Katherine Reynolds
i-48e12e09	InService	35 days	1.0.0 Nebulos
i-e3d601a2	InService	35 days	1.0.0 Johnny Storm
i-17e5d259	InService	35 days	1.0.0 Foolkiller

ophan-trigr

Instance	AutoScaling	ELB	Uptime	Version
i-dcecbe9d	InService	InService	8 days	1456

ophan-raw-event-logger

Instance	AutoScaling	ELB	Uptime	Version
i-581c011b	InService	InService	4 hours	1422
i-b0aa80f0	InService	InService	3 hours	1422

Filter by

Time ▾

Section ▾

Device ▾

Browser ▾

Country ▾

Referrer ▾

Breakdown by ▾

Summary

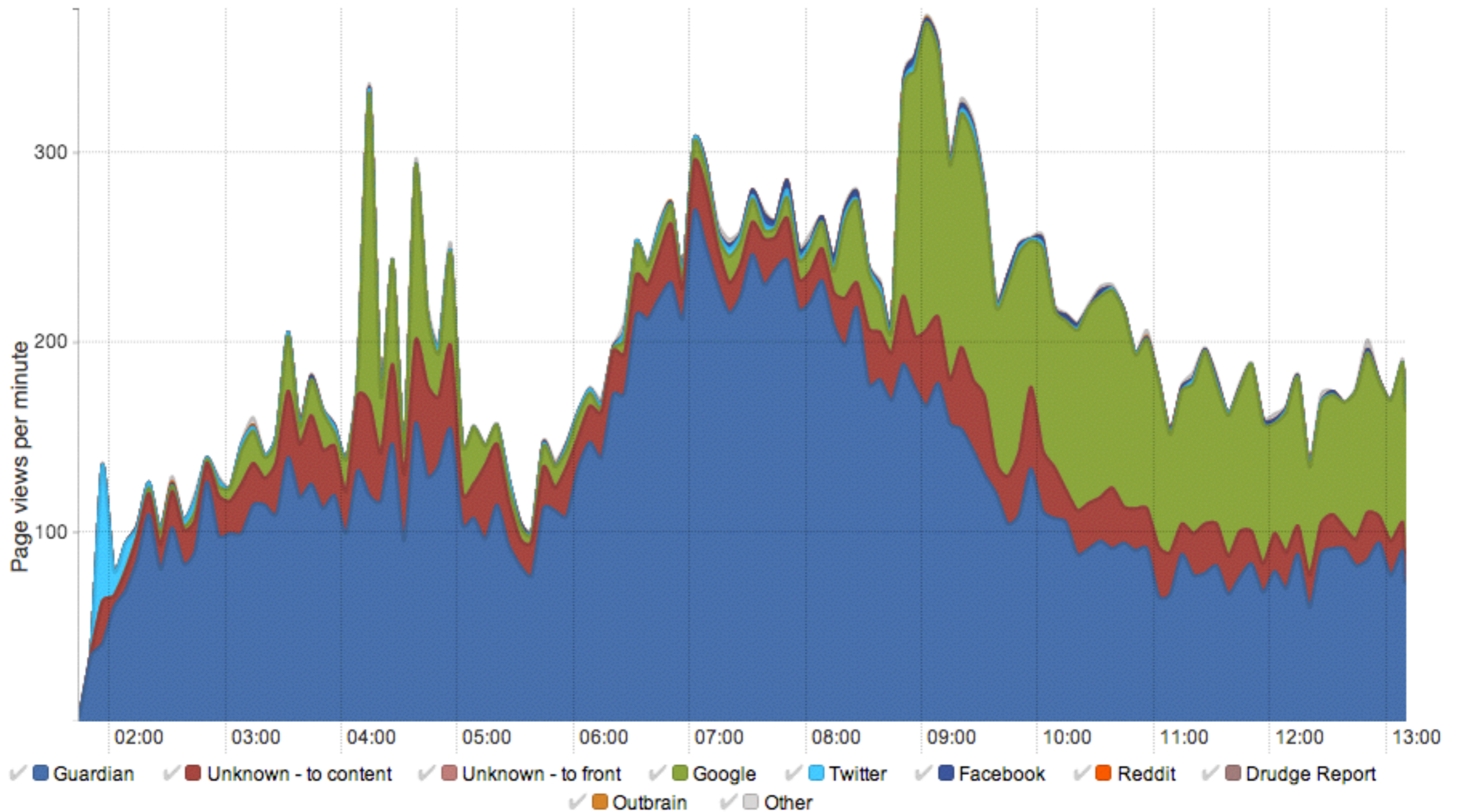
Graph

Where from?

Where next?

Heatmap

path: /film/2014/mar/03/oscars-2014-winners-list, Period: today



```
{
  "dt": "2014-03-03T02:01:48.026Z", ←... count per minute
  "url": "http://www.theguardian.com/film/2014/mar/03/oscars-2014-winners-list",
  "queryString": "",
  "host": "www.theguardian.com",
  "path": "/film/2014/mar/03/oscars-2014-winners-list", ←... filter
  "section": "film",
  "platform": "r2",
  "userAgent": {
    "type": "Browser",
    "family": "Safari 5.1.9",
    "os": "OS X 10.6.8",
    "device": "Personal computer"
  },
  "documentReferrer": "http://www.theguardian.com/world",
  "browser": {
    "id": "gA6RUFLhWNQvWdt0rW4r78Fg",
    "isNew": false
  },
  "referringHost": "theguardian.com", ←... filter
  "referringPath": "/world",
  "isContent": true,
  "contentPublicationDate": "2014-03-03",
  "countryCode": "US",
  "countryName": "United States",
  "location": {
    "lonlat": [-73.4409, 41.2094]
  }
}
```

```
{
  "query" : {
    "filtered" : {
      "query" : { "match_all" : { } },
      "filter" : {
        "term" : { "path" : "/film/2014/mar/03/oscars-2014-winners-list" }
      }
    }
  },
  ...
}
```

```
...
"facets": {
  "Reddit": {
    "date_histogram": { "field": "dt", "interval": "1m" },
    "facet_filter": { "term": { "referringHost": "reddit.com" } }
  },
  "Facebook": {
    "date_histogram": { "field": "dt", "interval": "1m" },
    "facet_filter": { "term": { "referringHost": "facebook.com" } }
  },
  "Google": {
    "date_histogram": { "field": "dt", "interval": "1m" },
    "facet_filter": {
      "or": {
        "filters": [
          { "prefix": { "referringHost": "www.google." } },
          { "prefix": { "referringHost": "news.google." } }
        ]
      }
    }
  }
}
}
```



Filter by

Time ▾

Section ▾

Device ▾

Browser ▾

Country ▾

Referrer ▾

Breakdown by ▾

Summary

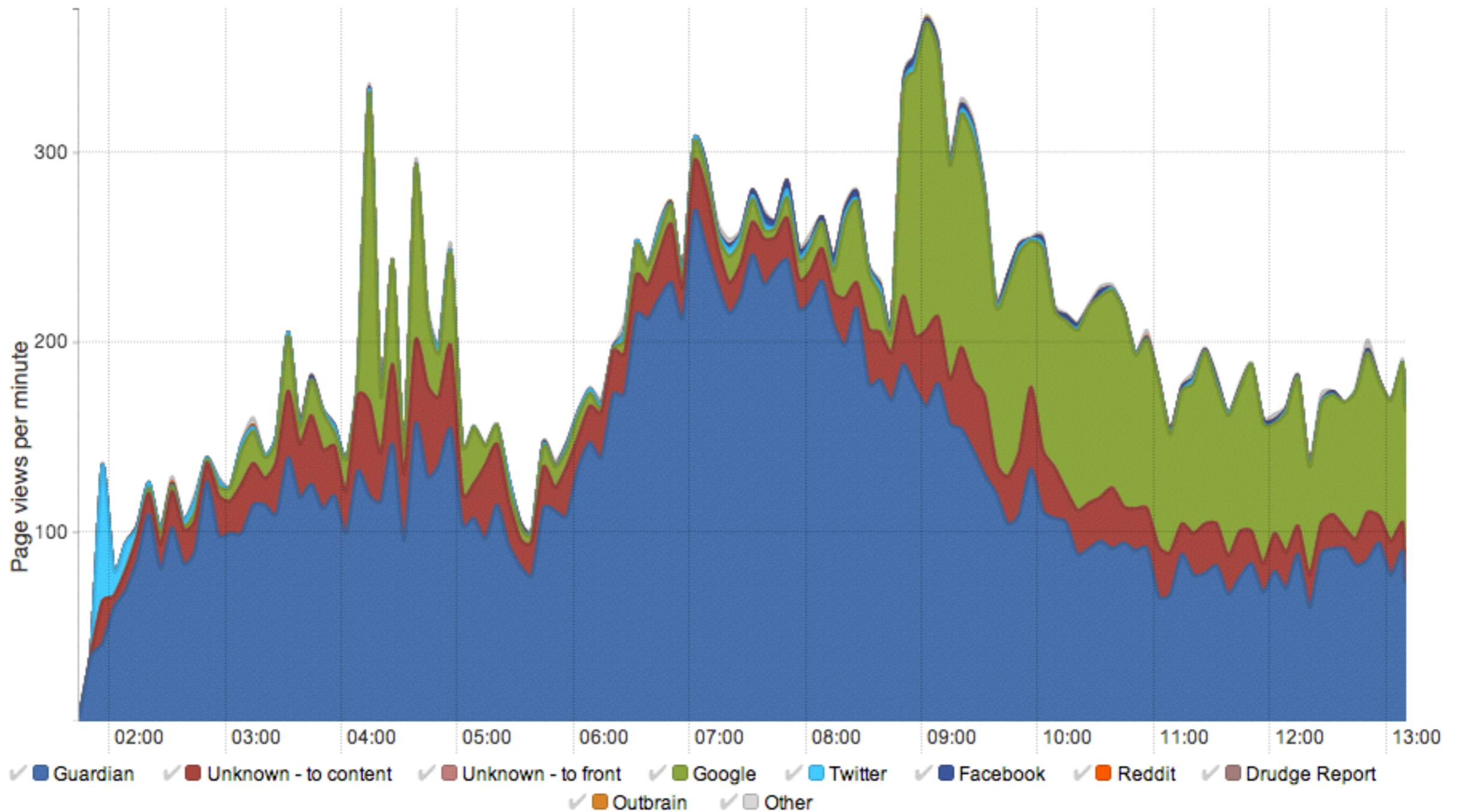
Graph

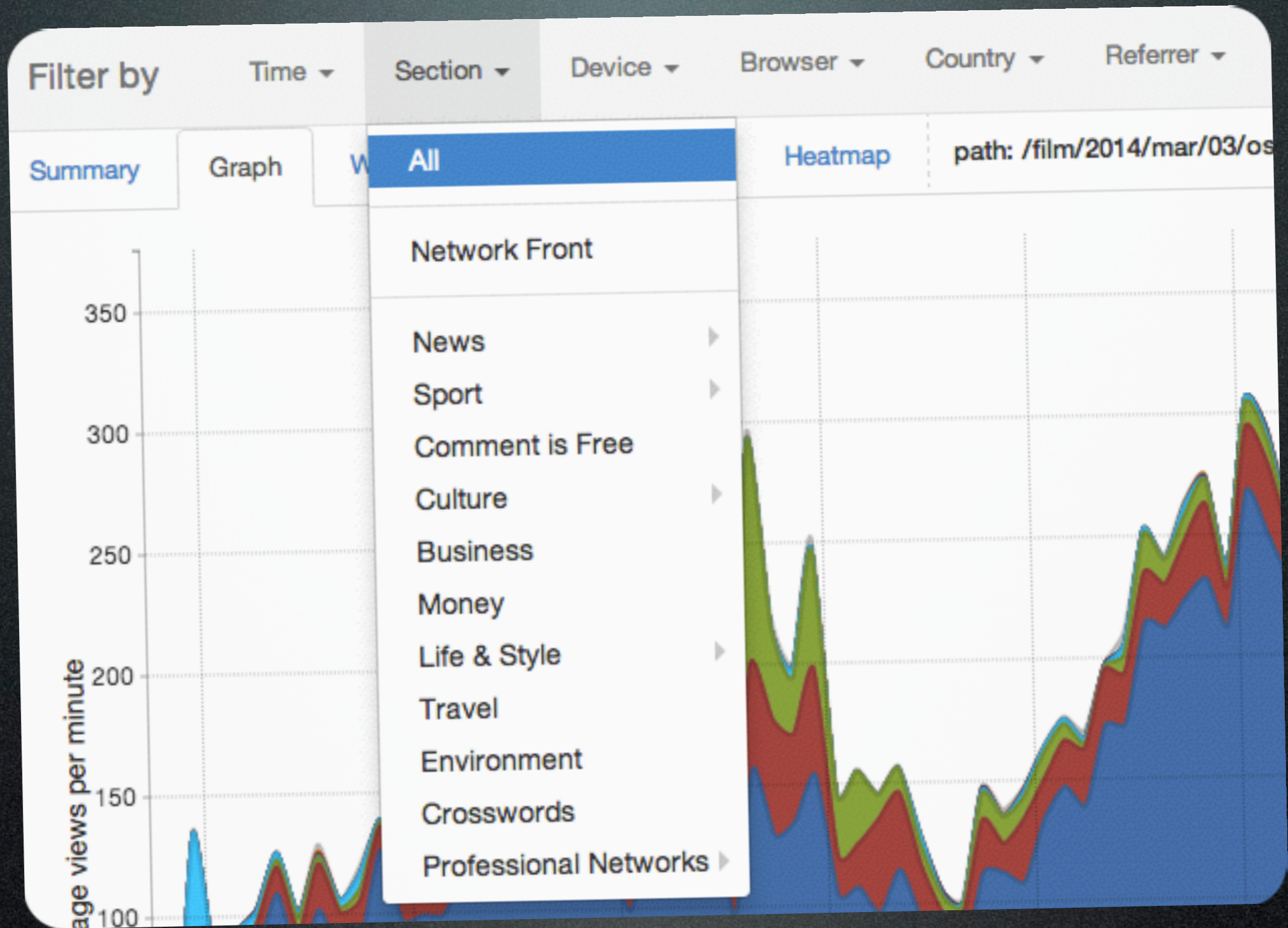
Where from?

Where next?

Heatmap

path: /film/2014/mar/03/oscars-2014-winners-list, Period: today





</graph/breakdown?section=commentisfree>

?section=commentisfree



```
ophan.StandardFilters
```

```
ophan.StandardFiltersToElasticsearch
```

```
org.elasticsearch.index.  
query.FilterBuilder
```



```
{
  "query" : {
    "filtered" : {
      "query" : { "match_all" : { } },
      "filter" : {
        "term" : { "path" : "/film/2014/mar/03/oscars-2014-winners-list" }
      }
    }
  },
  ...
}
```

```
"filter": {
  "and": {
    "filters": [
      {
        "range": {
          "dt": {
            "from": "2014-03-03T00:00:00.000Z",
            "to": "2014-03-03T22:30:59.999Z",
            "include_lower": true,
            "include_upper": false
          }
        }
      },
      {
        "not": {
          "filter": { "term": { "countryCode": "GNM" } }
        }
      },
      {
        "not": {
          "filter": { "term": { "userAgent.type": "Robot" } }
        }
      },
      {
        "filter": { "terms": { "section": [ "commentisfree" ] }}
      }
    ]
  }
}
```





QCon

Please evaluate  
my talk via the  
mobile app!



elasticsearch.

shay banon ■ @kimchy  
creator, co-founder and cto  
elasticsearch

the guardian

graham tackley ■ @tackers  
director of architecture  
guardian news and media