

Please evaluate
this talk via the
mobile app!



How Elasticsearch powers the Guardian's newsroom

elasticsearch.

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

theguardian

phil wills ■ @philwills
senior software architect
guardian news and media

FREE Great fairytales
 Today: *Wizards and fairy*
 Tomorrow: *Justice and punishment*

the guardian

Thursday 14 July 2011
 Published in London and Manchester
 guardian.co.uk

How much TV should your children watch?
 G2 cover story

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

David Leigh

The environmental campaigner's attempt to force the oil giant to disclose its activities in the Gulf of Mexico has been blocked by a court order. The Guardian has been told that it must not publish any information that could identify the company's operations in the region. The court order is a result of a legal challenge by the company, which claims that the publication of such information would be a breach of its confidentiality. The Guardian has argued that the information is in the public interest and should be made available to the public.

Dear leader:
 Even in line to manage North Korea



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



- Edward Snowden, 29, emerges from hiding in Hong Kong
- IT contractor says his concerns were ignored and he had to go public



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

Aditya Chakraborty
 Is Facebook destroying friendship?

the guardian

Ministers cave in to City over bank bonuses

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

Obama's second moment of silence

Undercover officer knew of second spy



FREE Audiobook download
 Mr Charlie Stinky Socks and the Really Big Adventure by Kristina Stephenson

From our exclusive new series of audiobook downloads. See page 20 for details. Tomorrow Down and Out in Paris and London by George Orwell

the guardian

Last sighting CCTV image of M16 man



Back inquiry was abandoned to avoid upsetting Met

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

A leaked memo from a senior police officer warned that a probe into the death of a man would 'deeply resent' the police. The memo was obtained by the Guardian and it is believed that it was written by a senior officer who was involved in the case. The memo stated that the police would be 'deeply resent' if a probe was launched into the death of the man. It also stated that the police would be 'deeply resent' if a probe was launched into the death of the man. The memo was written in 2009 and it is believed that it was written by a senior officer who was involved in the case.

Revealed: Murder bill for hiding dirty

- Tory PR chief under fire over tabloid hacking
- Politicians and celebrities among victims



I think it is one of the most significant modern stories of the system. News of the World out of control.

The whole thing is deplorable. Mr Cameron should be ashamed of himself.



theguardian
beta

life environment tech money travel fashion



ends

ussia's ac

Australia considers Russian

and of

Ukraine



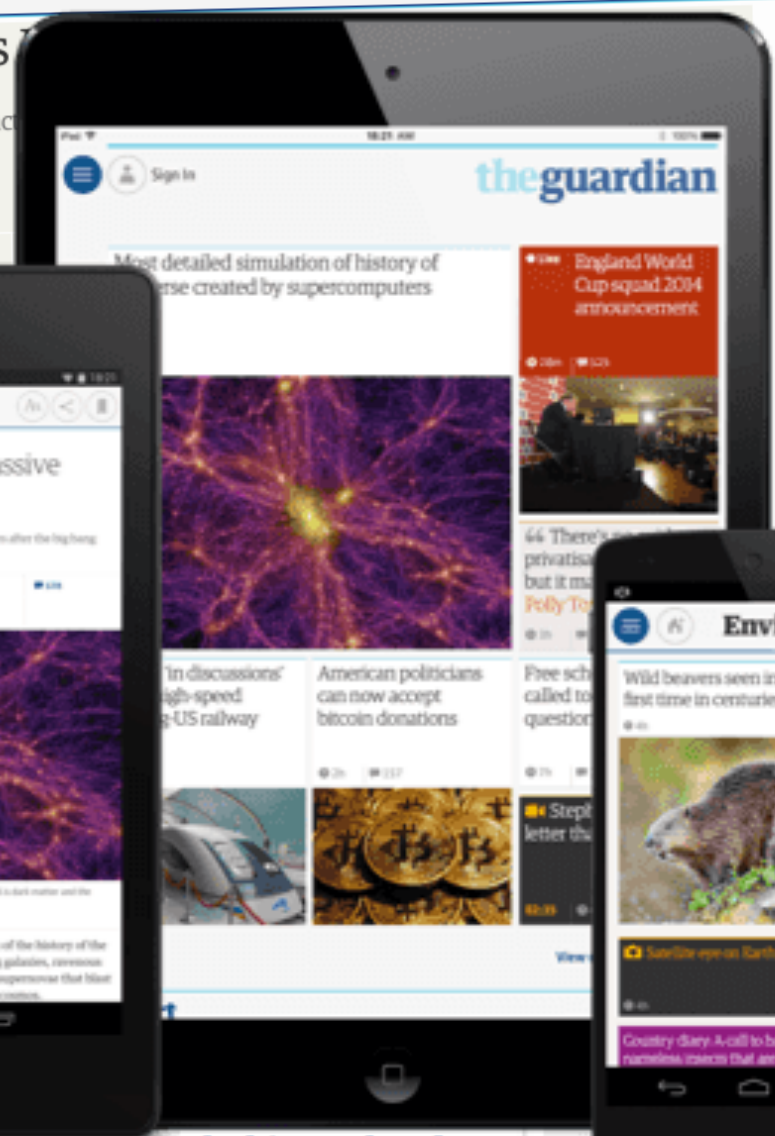
Lawyers to walk out for second time in protest against legal aid

Osborne faces £20bn black hole



Austerity may last until 2020 and chancellor may have to make further cuts to welfare, local

Travel
Disti



Ten things to watch this weekend

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.

Ophan

Top 20

Summary ▾

Recently Published

Graph

Last 24 Hours

More ▾

Labs ▾



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

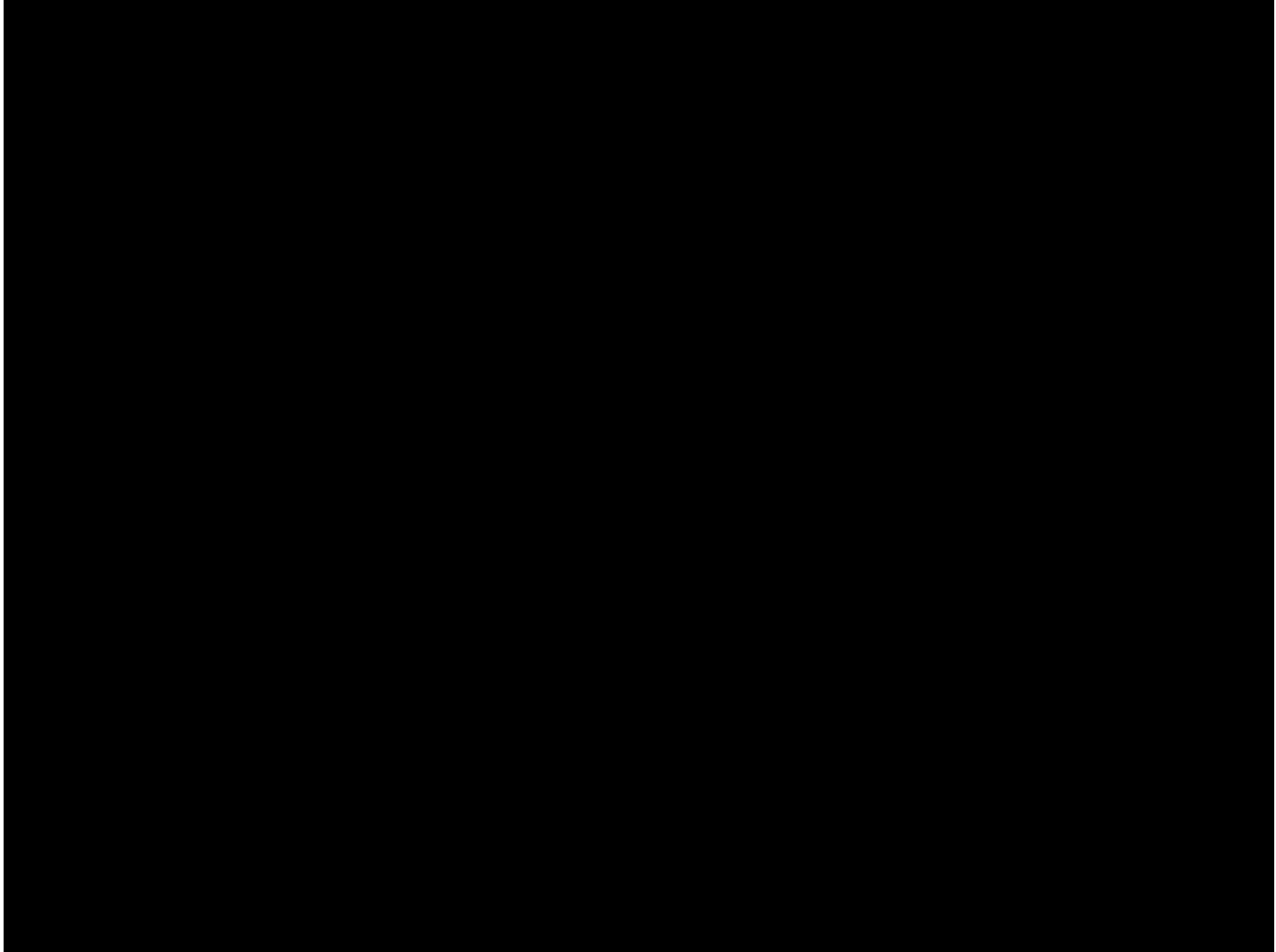
Welcome to Ophan, the Guardian's in-house, real-time traffic tool.

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 a length of time or 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.

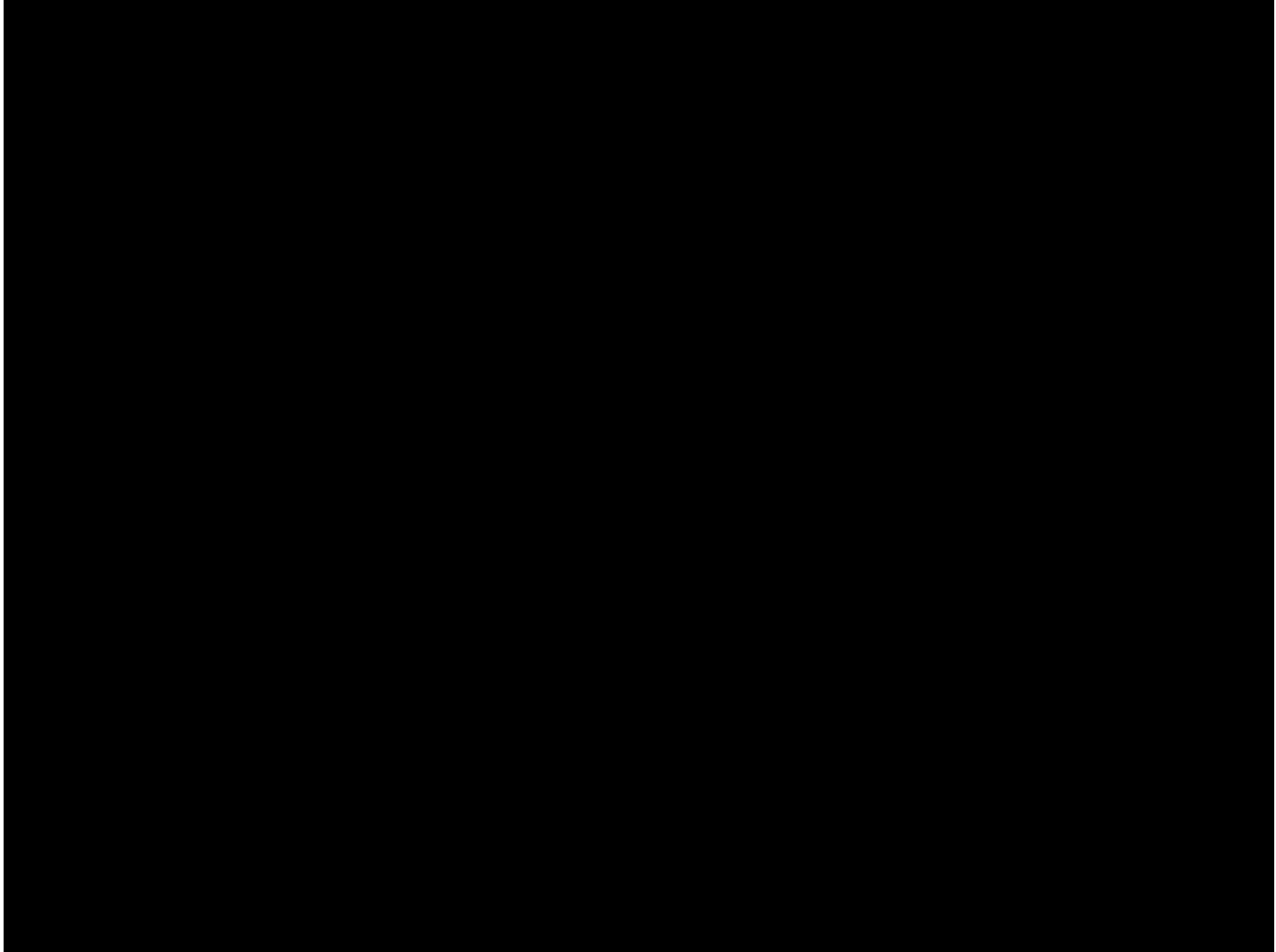
Other pages you might find interesting are the [recently published page](#), the [top 20](#) and the [live Google search terms page](#). Just explore the options at the top of the screen.

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.](#)

Version 982 created September 20, 2013 10:43:47 AM BST. Ophan is currently 3 seconds behind.







desktop
workstation



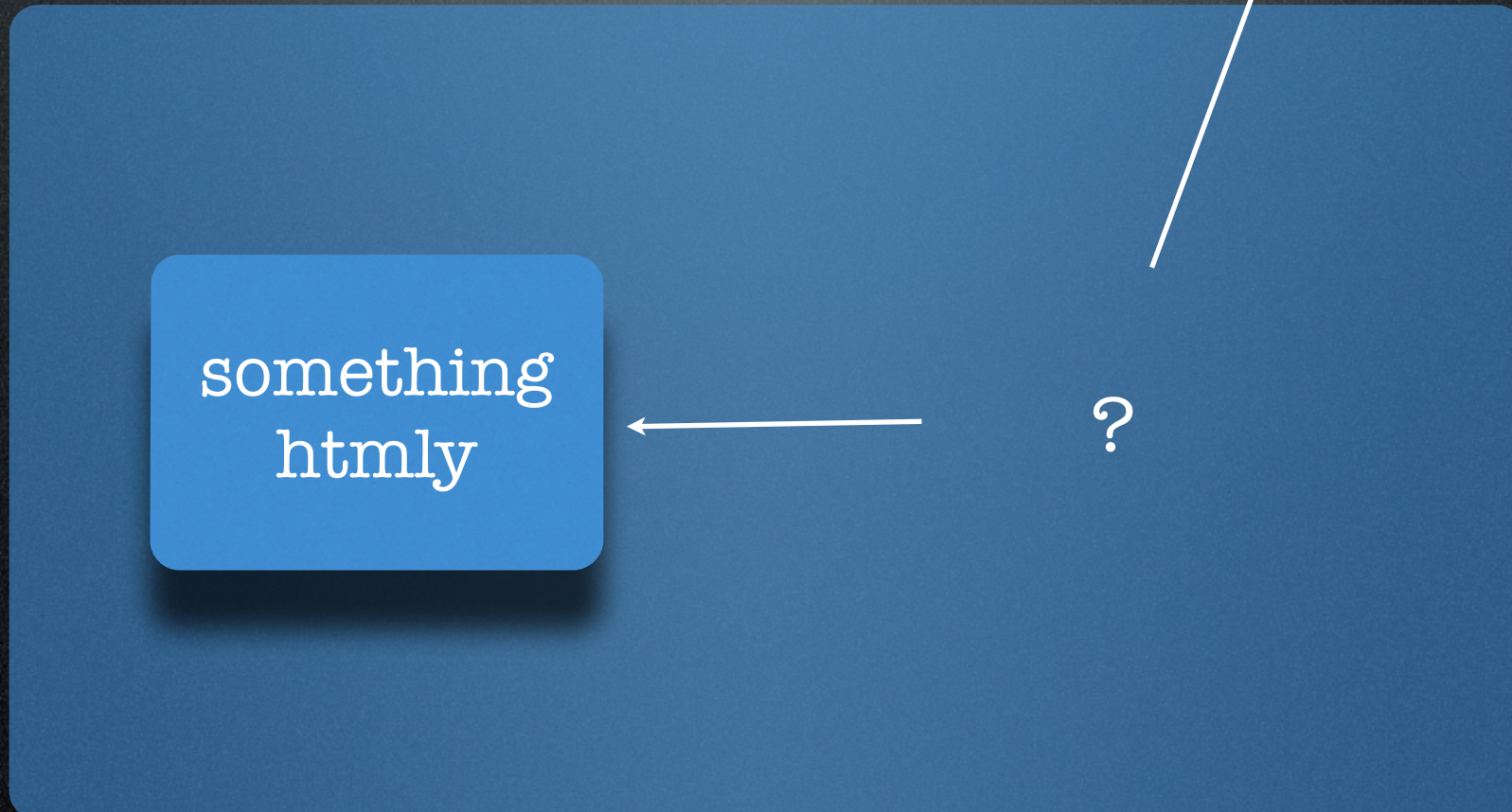
production
apaches



something
htmlly



?



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


desktop
workstation



2 x production
apaches

ssh "tail"

SEO
dashboard

zeromq

publisher



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)

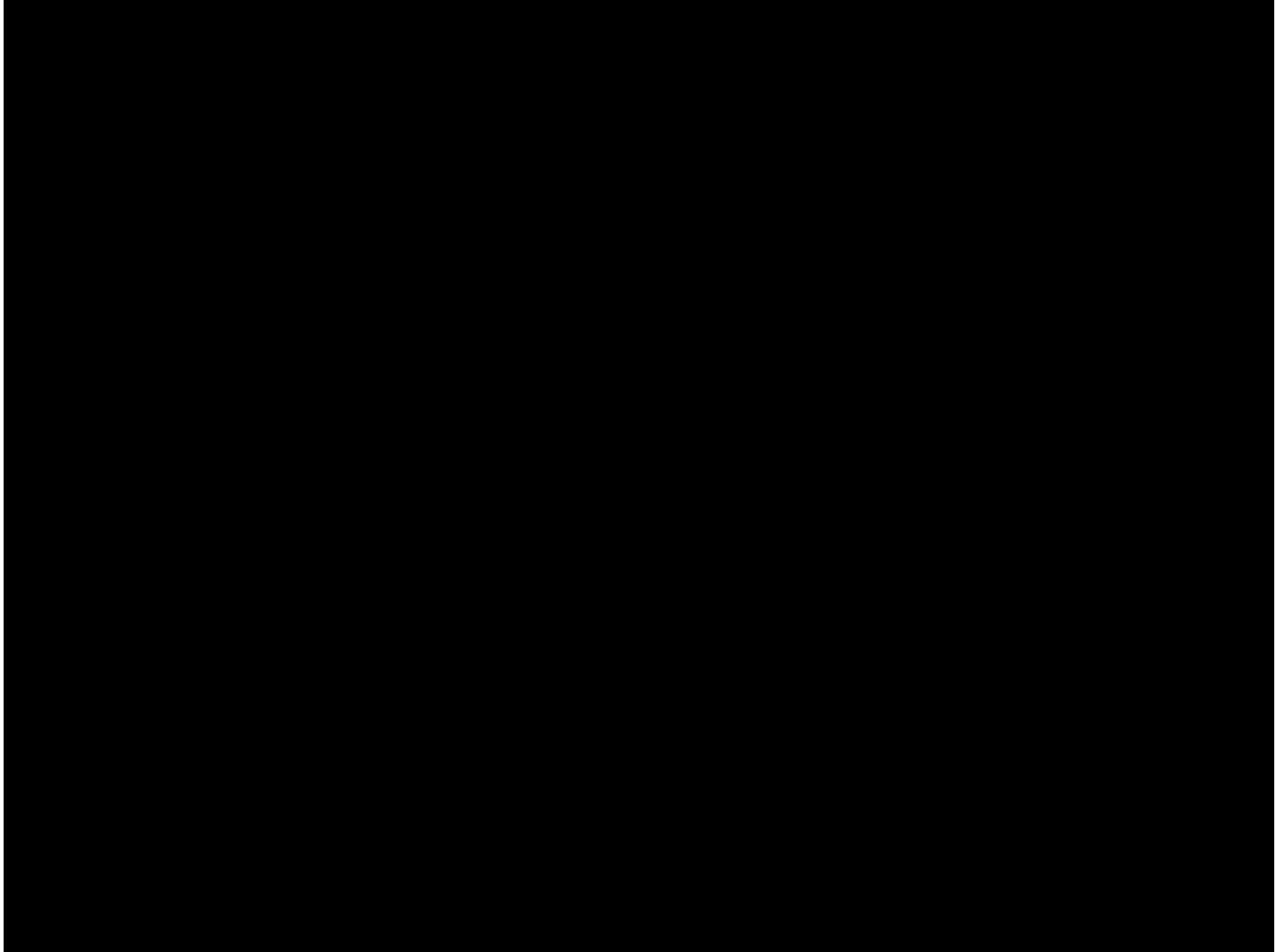
Revealed: Qatar's World Cup 'slaves'

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





~~desktop
workstation~~



Javascript in browser

hidden pixel



Tracker



SNS

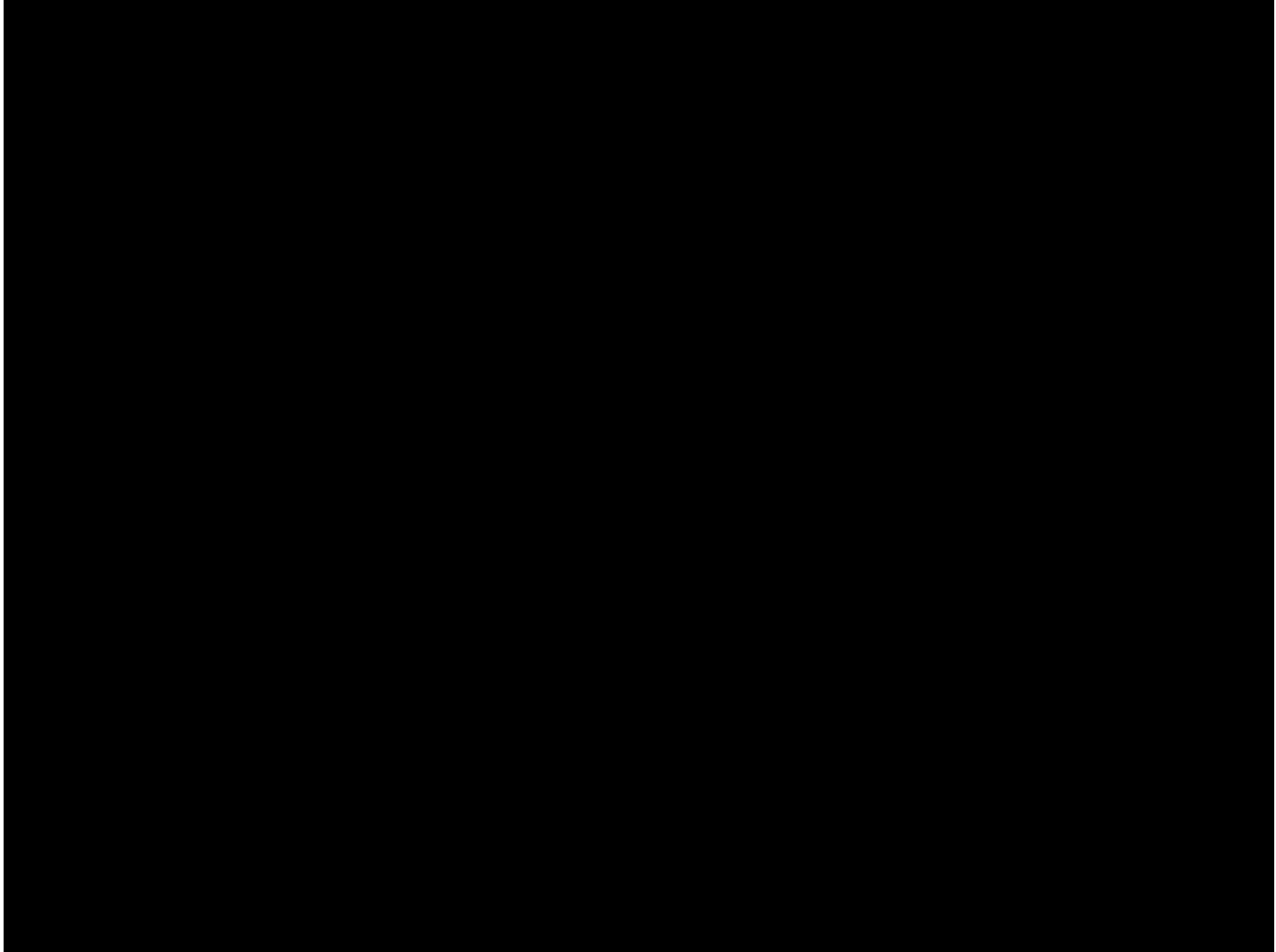


SQS



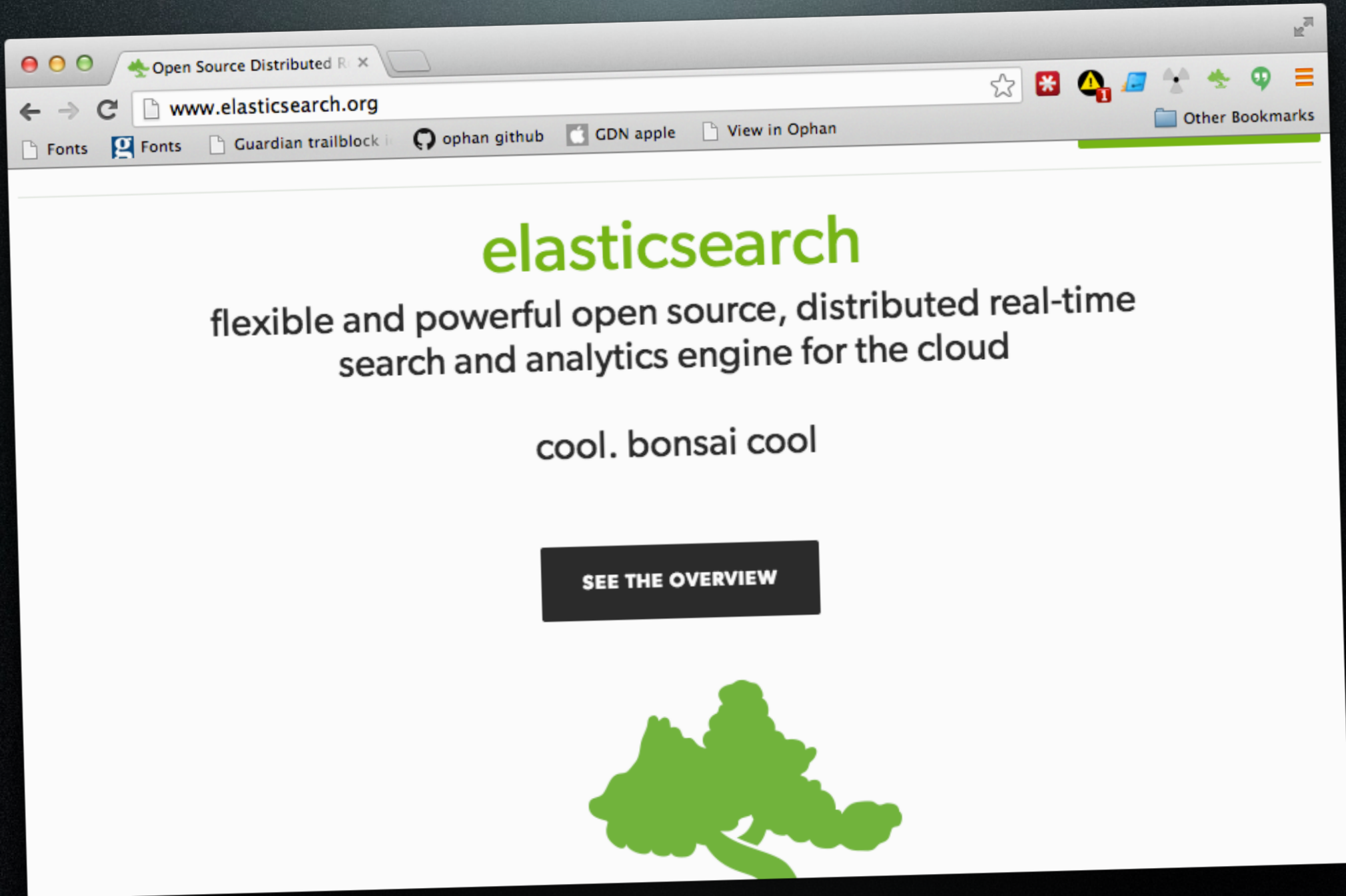
Dashboard





Elasticsearch

“you know, for search”



Javascript in browser

image pixel

Tracker

SNS

SQS

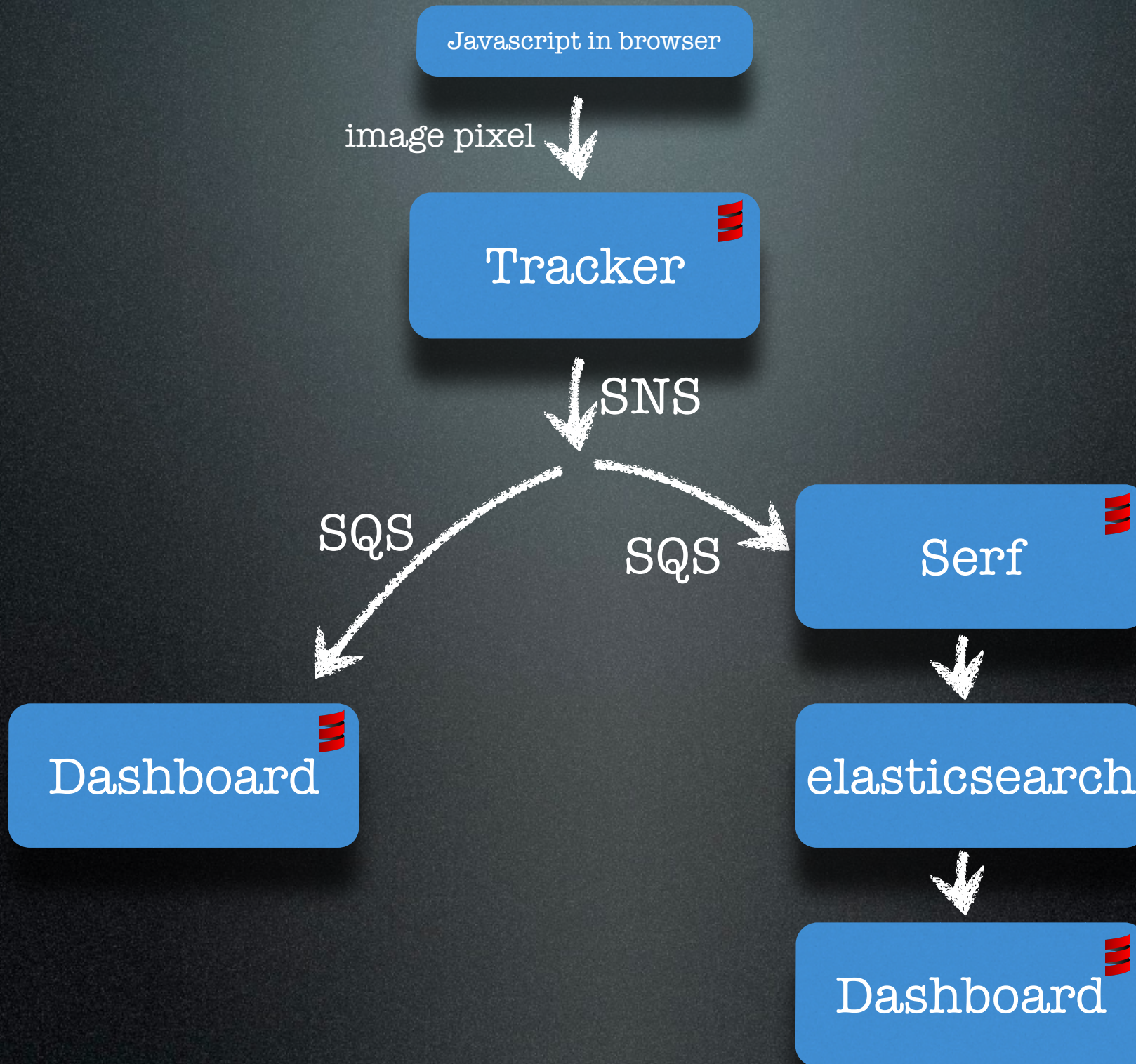
SQS

Serf

Dashboard

elasticsearch

Dashboard



	ads size: 1.27M (3.82M) docs: 5,828 (5,838) Info Actions	content size: 163M (490M) docs: 146,371 (148,566) Info Actions	facebook size: 12.1M (36.2M) docs: 10,379 (10,546) Info Actions	kibana-int size: 234k (468k) docs: 22 (22) Info Actions	ophan-2014-06-09 size: 30.1G (60.1G) docs: 94,656,801 (94,662,673) Info Actions	ophan-2014-06-10 size: 28.9G (57.8G) docs: 91,509,805 (91,515,263) Info Actions	ophan-2014-06-11 size: 27.8G (55.7G) docs: 88,273,234 (88,278,777) Info Actions	ophan-2014-06-12 size: 27.6G (55.1G) docs: 85,831,178 (85,836,307) Info Actions	ophan-2014-06-13 size: 27.3G (54.5G) docs: 85,865,848 (85,869,754) Info Actions	ophann-2014-06-14 size: 24.8.0G (48.0G)) docs: 74,982,51582 (74,986,3,316) Info Actions	ophan-2014-06-15 size: 26.4G (52.9G) docs: 82,822,325 (82,827,958) Info Actions	ophan-2014-06-16 size: 32.9G (58.0G) docs: 104,594,614 (104,600,391) Info Actions	ophan-2014-06-17 size: 30.4G (58.7G) docs: 95,663,775 (95,699,956) Info Actions	ophan-2014-06-18 size: 9.31G (18.6G) docs: 28,939,809 (28,944,601) Info Actions	ophan-2014-06-19 size: 494B (968B) docs: 0 (0) Info Actions	tag size: 12.1M (36.2M) docs: 42,751 (42,751) Info Actions
● Alysande Stuart Info Actions	0 2	1	1 2		3 4	4	0 4	0 4	0 4	2	3	3	0 2		0 2	0
● Cancer Info Actions			2				2		0							
★ Flatman Info Actions	0 1			1		2			1	1	0	0 1	1	0 1	1	
		2			4	4	4	4 5	4 5	5	5	4	5	5	4 5	2
● Hellcat Info Actions	1	0 2	0	0 2	3 5	3 5	3 5	2 3	2		1 5	1 4			3 2	1
● Jessica Drew Info Actions								3		0	2	2		4		
● Katie Power Info Actions	1					3		4		4					1	
● Kid Nova Info Actions	2	0 1	0 1		0 1	0 1	1		0		2 3	2	2	2 3	0	0 1
				2			2			4			4		3	
● Master Man Info Actions		1	1	0	0		0	1					1		1	1
	2				3	5		3	2 3		4 5	5	4 5	4 5	5	
● Spike Freeman Info Actions	0	0	0	1	1	1	1	0	1	1	0 1	0	0	0		0
		2	2		2	2	2	3	2			3		3		2
								4 5		5			3		4	

PROD

Last Updated: 3 secs ago Philip Wills

ophan-events

ophan-raw-event-log

6 * c3.4xlarge

0

dashboard

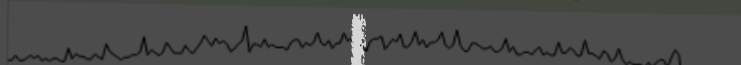
instance store (SSD)



Instance	AutoScaling	ELB	Uptime	Version
i-7fff343f	InService	InService	20 hours	1740
i-dd04d49e	InService	InService	20 hours	1740
i-023b8c40	InService	InService	20 hours	1740

in an autoscaling group
(with manual scaling)

Instance	AutoScaling	ELB	Uptime	Version
i-996a71d8	InService	InService	56 days	1591
i-6b015	InService	InService	56 days	1591
i-0c2fd64e	InService	InService	50 days	1591



elasticsearch

Suspended activities: Terminate

Instance	AutoScaling	Uptime	Version
i-4ffe090f	InService	24 days	1.1.2 Aleta Ogord
i-9bdb24db	InService	18 days	1.1.2 Mist Mistress
i-d7916a94	InService	24 days	1.1.2 Asteroth
i-4b976c08	InService	24 days	1.1.2 Doop
i-1e19c45c	InService	24 days	1.1.2 Dragon Man
i-1f19c45d	InService	24 days	1.1.2 Achelous



trigr

Instance	AutoScaling	ELB	Uptime	Version
i-7923df39	InService	InService	18 days	1642



logger

Instance	AutoScaling	ELB	Uptime	Version
i-bb6328fa	InService	InService	50 days	1621



<https://github.com/guardian/status-app>

Ophan

Top 20

Summary ▾

Recently Published

Graph

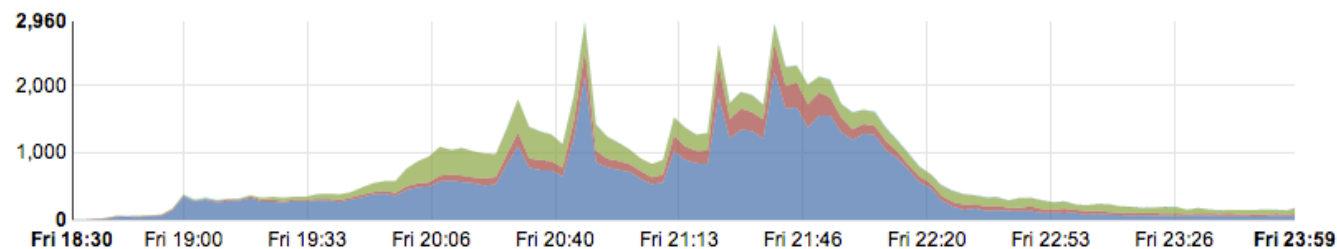
Last 24 Hours

More ▾

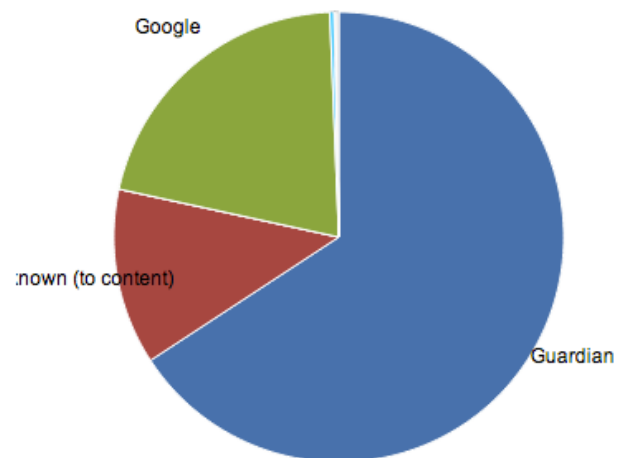


Spain v Holland: World Cup 2014 - as it happened

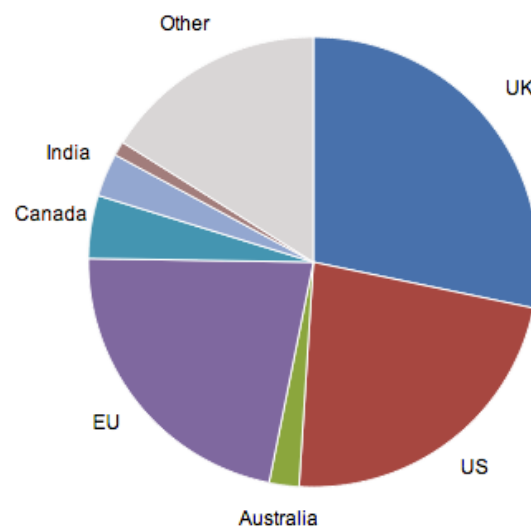
261,064 page views



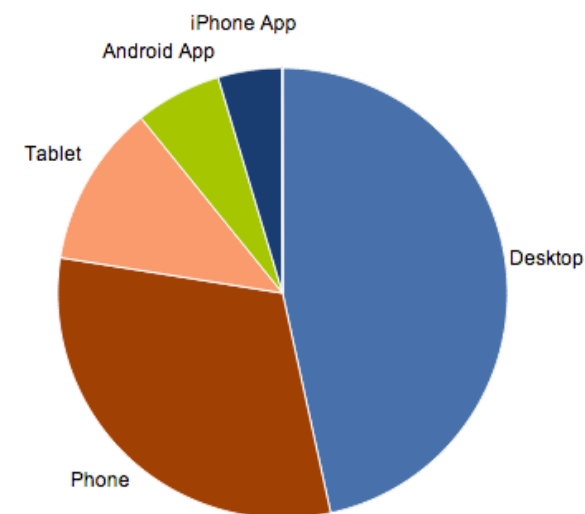
Referrers [see all »](#)



Country [see all »](#) [map »](#)



Device



Filter by

Time ▾

Section ▾

Device ▾

Browser ▾

Country ▾

Referrer ▾

Breakdown by ▾

Summary

Graph

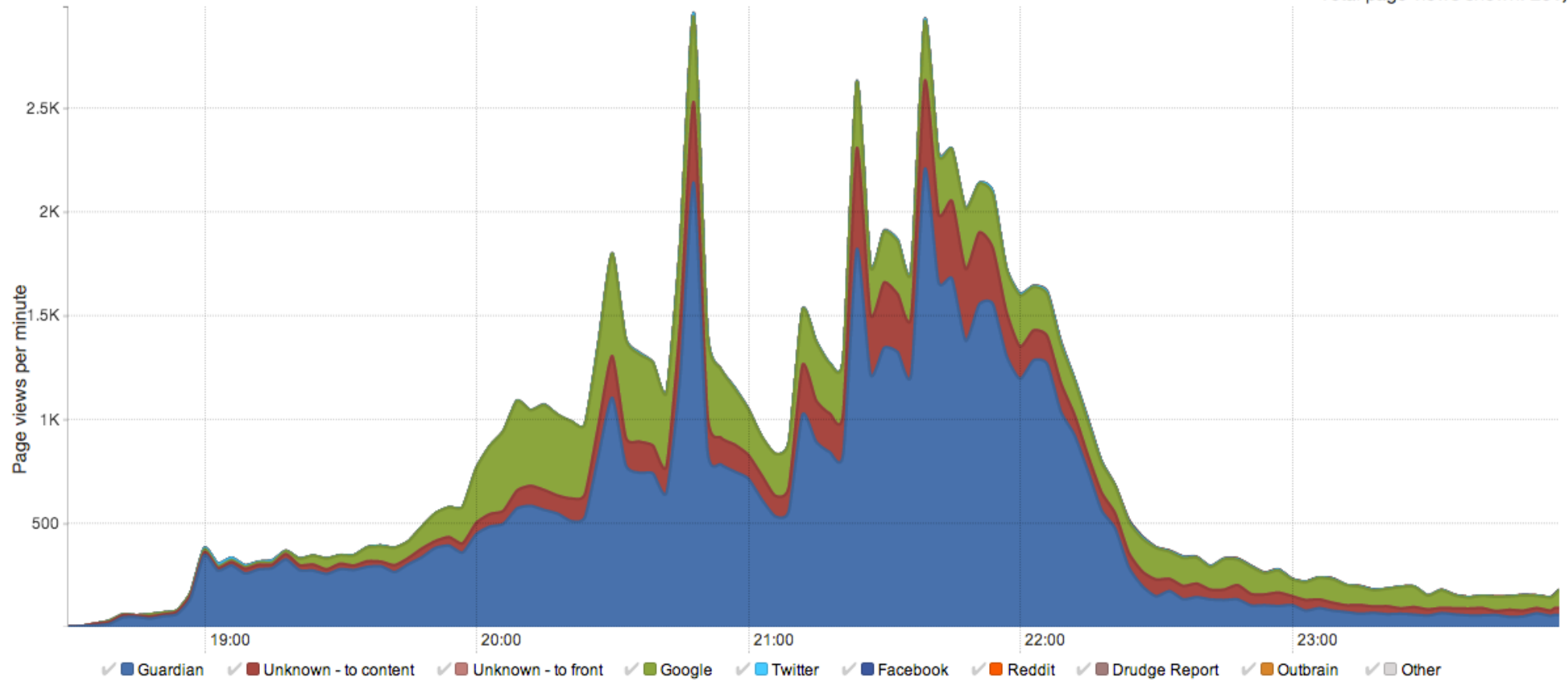
Where from?

Where next?

Heatmap

path: /football/2014/jun/13/spain-v-holland-world-cup-2014-live-report, Period: 2014-06-13

Total page views shown: 261,064




```
{
  "dt": "2014-06-13T20:01:48.026Z", ←... count per minute
  "url": "http://www.theguardian.com/football/2014/jun/13/spain-v-holland-world-cup-2014-
live-report",
  "queryString": "",
  "host": "www.theguardian.com",
  "path": "/football/2014/jun/13/spain-v-holland-world-cup-2014-live-report", ←... filter
  "section": "football",
  "platform": "r2",
  "userAgent": {
    "type": "Browser",
    "family": "Safari 5.1.9",
    "os": "OS X 10.6.8",
    "device": "Personal computer"
  },
  "documentReferrer": "http://www.theguardian.com/football",
  "browser": {
    "id": "gA6RUFLhWNQvWdt0rW4r78Fg",
    "isNew": false
  },
  "referringHost": "theguardian.com", ←... filter
  "referringPath": "/football",
  "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": "/football/2014/jun/13/spain-v-holland-world-cup-2014-live-report" }
      }
    }
  },
  ...
}
```



```

...
"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

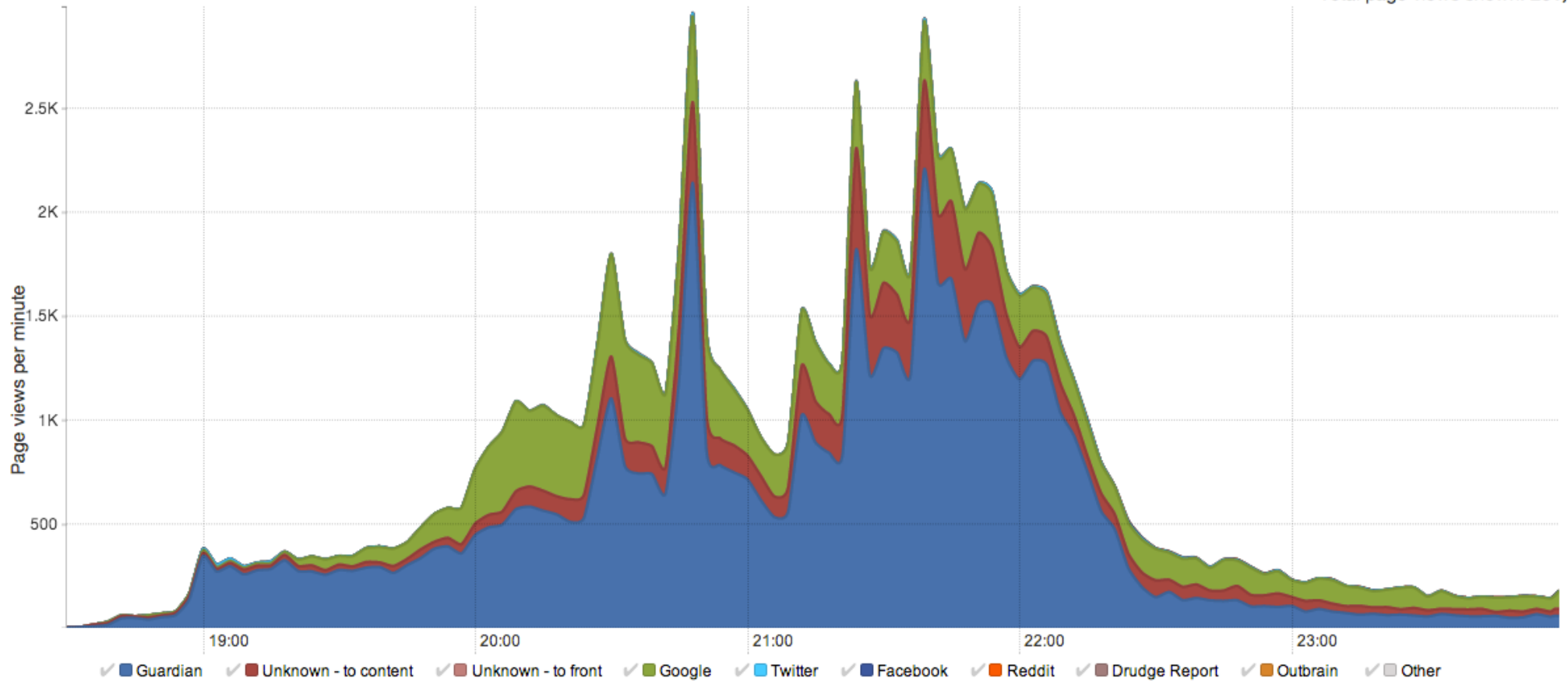
Where from?

Where next?

Heatmap

path: /football/2014/jun/13/spain-v-holland-world-cup-2014-live-report, Period: 2014-06-13

Total page views shown: 261,064



Ophan

Top 20

Summary ▾

Recently Published

Graph

Last 24 Hours

More ▾



Filter by

Time ▾

Section ▾

Device ▾

Browser ▾

Country ▾

Referrer ▾

Breakdown by ▾

Stacked

DNS

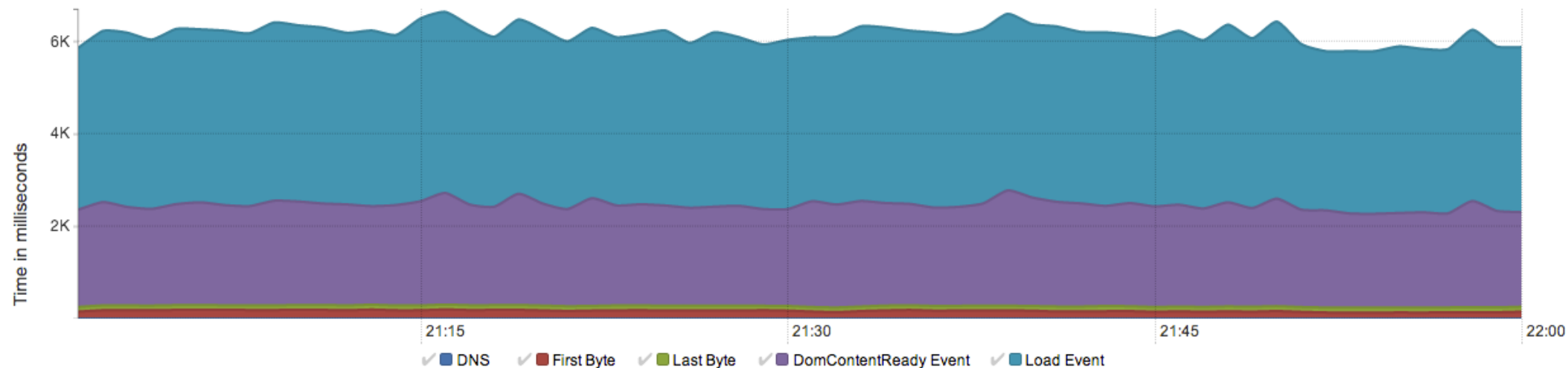
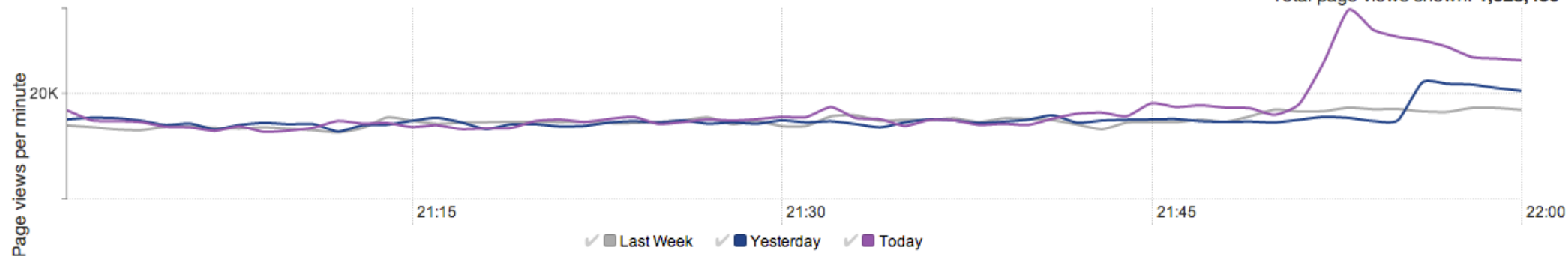
First byte

Last byte

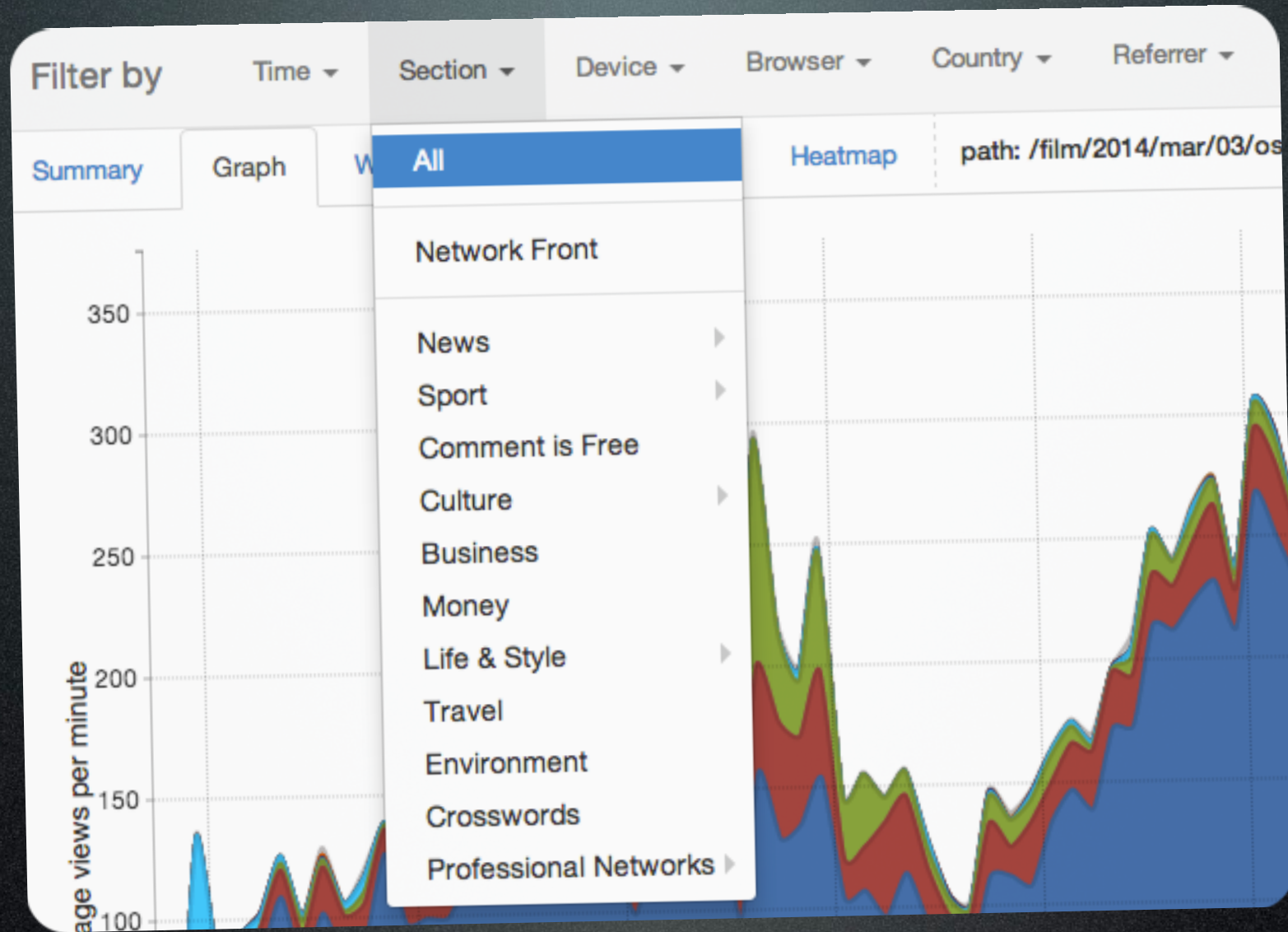
DOM content loaded

Loaded

Total page views shown: 1,028,450



```
"aggregations" : {  
  "dns" : {  
    "date_histogram" : {  
      "field" : "dt",  
      "interval" : "1m"  
    },  
    "aggregations" : {  
      "dns" : {  
        "percentiles" : {  
          "field" : "dns",  
          "percents" : [ 50.0 ],  
          "estimator" : "tdigest",  
          "compression" : 10.0  
        }  
      }  
    }  
  }  
}
```

`/graph/breakdown?section=commentisfree`

?section=commentisfree



ophan.StandardFilters

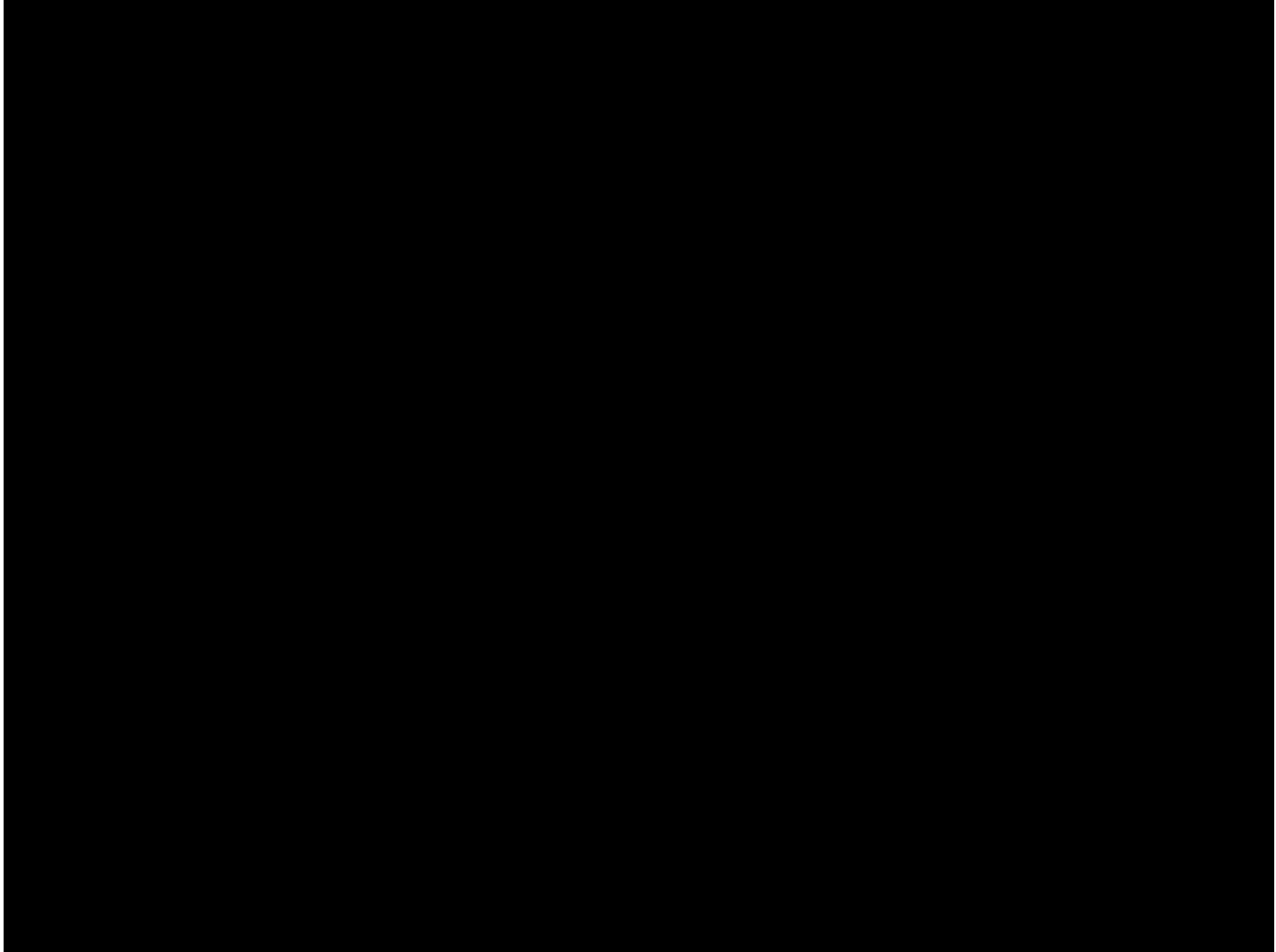
ophan.StandardFiltersToElasticsearch



org.elasticsearch.index.
query.FilterBuilder


```
{  
  "query" : {  
    "filtered" : {  
      "query" : { "match_all" : { } },  
      "filter" : {  
        "term" : { "path": "/football/2014/jun/13/spain-v-holland-world-cup-2014-live-report" }  
      }  
    }  
  },  
  ...  
}
```

```
"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" ] }}
      }
    ]
  }
}
```





This repository ▾

Search or type a command



Explore Gist Blog Help



philwills



PRIVATE



guardian / ophan

Unwatch ▾

9

★ Star

1

🍴 Fork

13

Contributors

Traffic

Commits

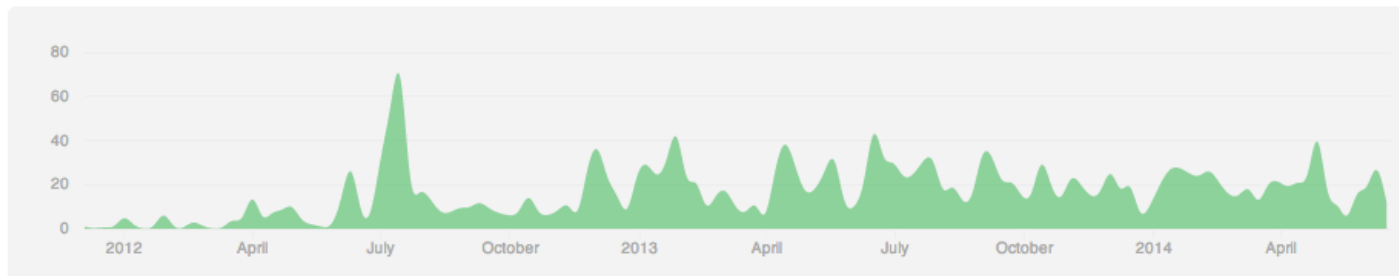
Code frequency

Punch card

Dec 4, 2011 - Jun 18, 2014

Contributions to master, excluding merge commits

Contribution type: **Commits** ▾



tackley

1,037 commits / 217,300 ++ / 79,432 --

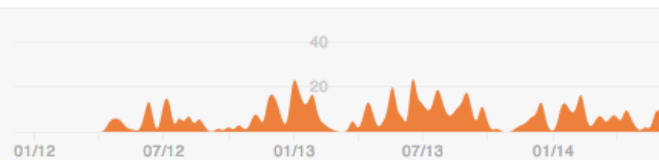
#1



philwills

806 commits / 104,891 ++ / 51,305 --

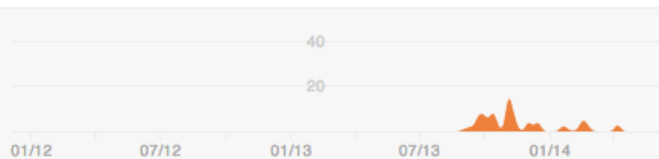
#2



lindseydew

97 commits / 14,388 ++ / 2,147 --

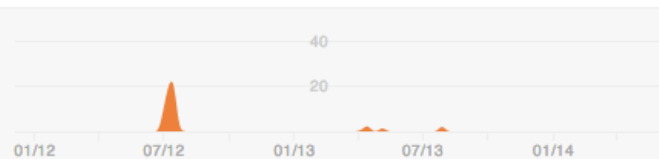
#3



daithiocruailaoich

67 commits / 17,179 ++ / 86,473 --

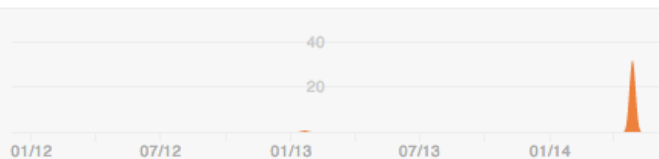
#4



mattandrews

49 commits / 1,318 ++ / 576 --

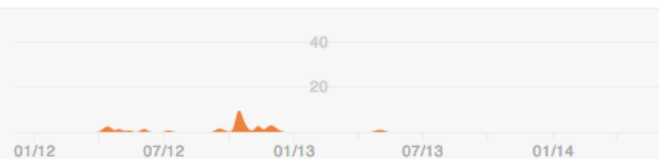
#5



ahume

46 commits / 2,120 ++ / 1,632 --

#6



Please evaluate
this talk via the
mobile app!



Follow us @gotoamst

elasticsearch.

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

theguardian

phil wills ■ @philwills
senior software architect
guardian news and media