Untitled by Thomas Xu, used under CC BY-SA 2.0 / Converted to black and white

Taming logs with Elasticsearch, Logstash, and Kibana

Lansing DevOps Meetup
May 5, 2015
Dan Grabowski

- Introductions
- Problem
- Solution
- Demo
- Quick Start
- Experiences
- Questions

Vertafore®

Software developer, architect, manager

Work at Vertafore, formerly Sircon, since 2001

Primary experience developing and maintaining Java web applications

Currently evolving a suite of Java web apps towards a service-based architecture

The opinions expressed in this presentation are my own and do not represent the opinions of my employer.

My favorite programming language is Clojure.

My favorite technical talk is “Are We There Yet?” by Rich Hickey (http://www.infoq.com/presentations/Are-We-There-Yet-Rich-Hickey).

The most interesting thing I’ve read recently is Kyle Kingsbury’s Jepsen series of blog posts (http://aphyr.com/tags/Jepsen).

Non-technical topics I’ll talk forever about if you get me started include Formula 1 and The Wire.

Problem

Extracting information from large volumes of log data, making it accessible, and keeping it visible

Large

- On the order of
  - 10 million log events per day
  - A few dozen applications and services
  - Several dozen servers

Information

- Details of an occurrence of an issue
- Count and patterns of occurrences for an issue
- Future occurrences of an issue
- Performance trends and variations

Accessible

- Available to developers, business analysts, system administrators, customer support, etc.
- Can be used effectively by all these groups, perhaps with different levels of sophistication

Visible

- Keep monitoring in front of people
- … but don’t nag them

Solution

- Elasticsearch to store and index log events
- Logstash to monitor, parse, and load logs
- Kibana as a UI to search, visualize information, and build dashboards
- Elasticsearch HTTP API to extract data for further analysis

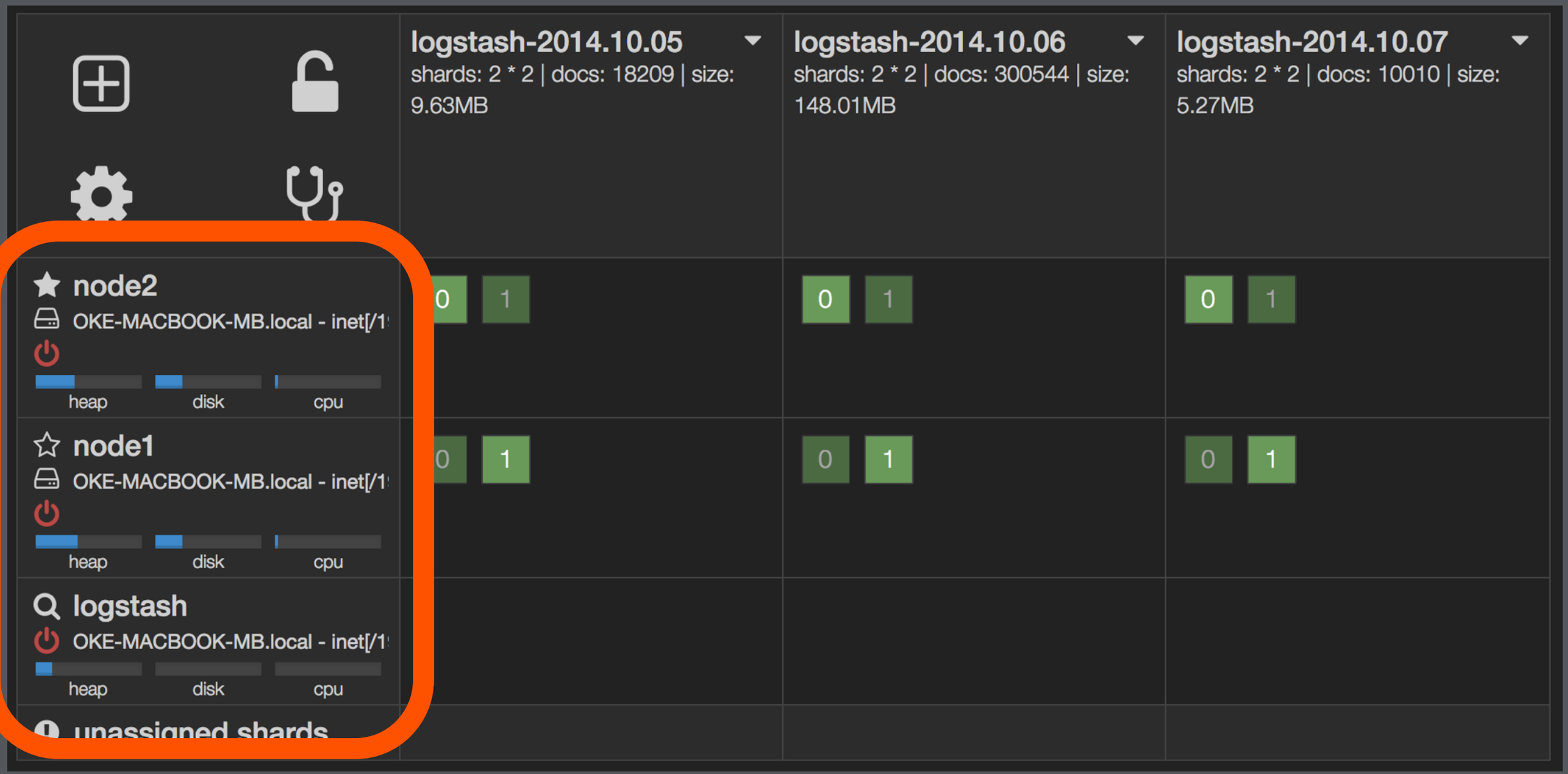
Elasticsearch (the product)

- Distributed Lucene indexes
- HTTP API
- Runs on JVM
- Apache 2.0 license
- Maintained and supported by Elasticsearch (the company)

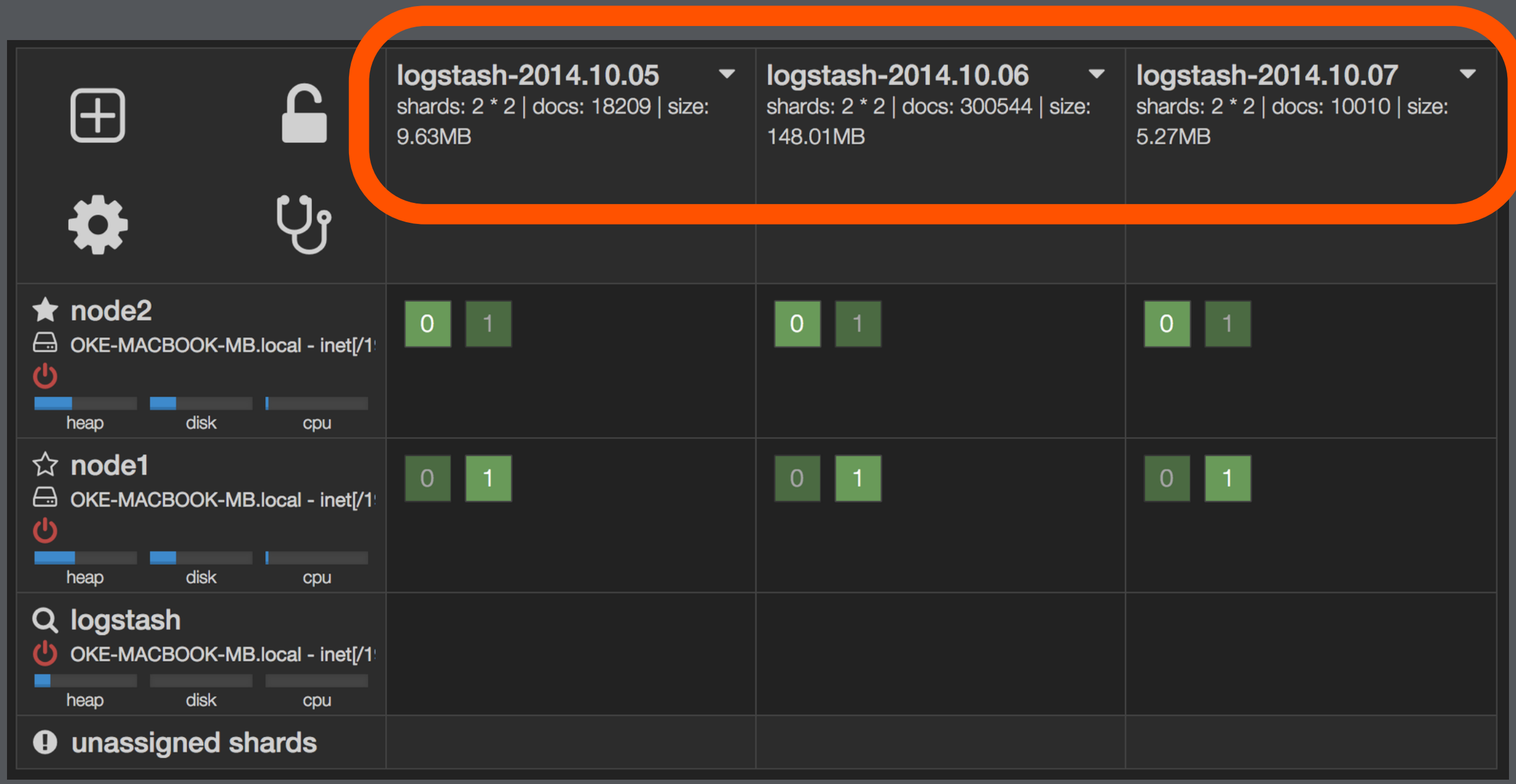
Cluster

	logstash-2014.10.05 shards: 2 * 2 | docs: 18209 | size: 9.63MB	logstash-2014.10.06 shards: 2 * 2 | docs: 300544 | size: 148.01MB	logstash-2014.10.07 shards: 2 * 2 | docs: 10010 | size: 5.27MB
★ node2 OKE-MACBOOK-MB.local - inet[/1 heap disk cpu	0 1	0 1	0 1
☆ node1 OKE-MACBOOK-MB.local - inet[/1 heap disk cpu	0 1	0 1	0 1
logstash OKE-MACBOOK-MB.local - inet[/1 heap disk cpu			
unassigned shards			

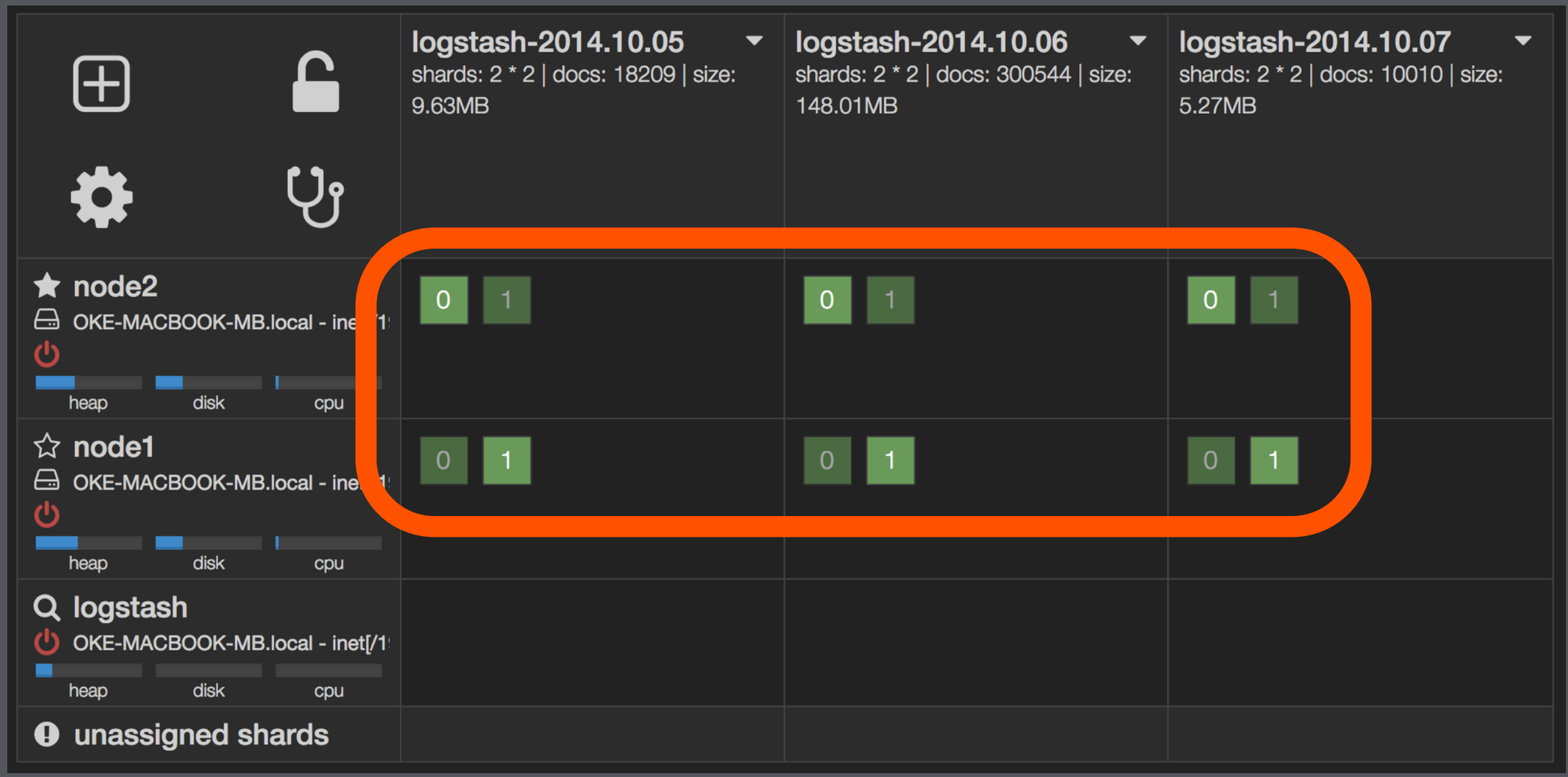
Nodes


logstash-2014.10.05
shards: 2 * 2 | docs: 18209 | size: 9.63MB
logstash-2014.10.06
shards: 2 * 2 | docs: 300544 | size: 148.01MB
logstash-2014.10.07
shards: 2 * 2 | docs: 10010 | size: 5.27MB
node2
OKE-MACBOOK-MB.local - inet[/1
heap
disk
cpu
node1
OKE-MACBOOK-MB.local - inet[/1
heap
disk
cpu
logstash
OKE-MACBOOK-MB.local - inet[/1
heap
disk
cpu
unassigned shards
0
1

Indexes


logstash-2014.10.05
shards: 2 * 2 | docs: 18209 | size: 9.63MB
logstash-2014.10.06
shards: 2 * 2 | docs: 300544 | size: 148.01MB
logstash-2014.10.07
shards: 2 * 2 | docs: 10010 | size: 5.27MB
node2
OKE-MACBOOK-MB.local - inet[/1
heap
disk
cpu
0
1
node1
logstash
unassigned shards

Shards

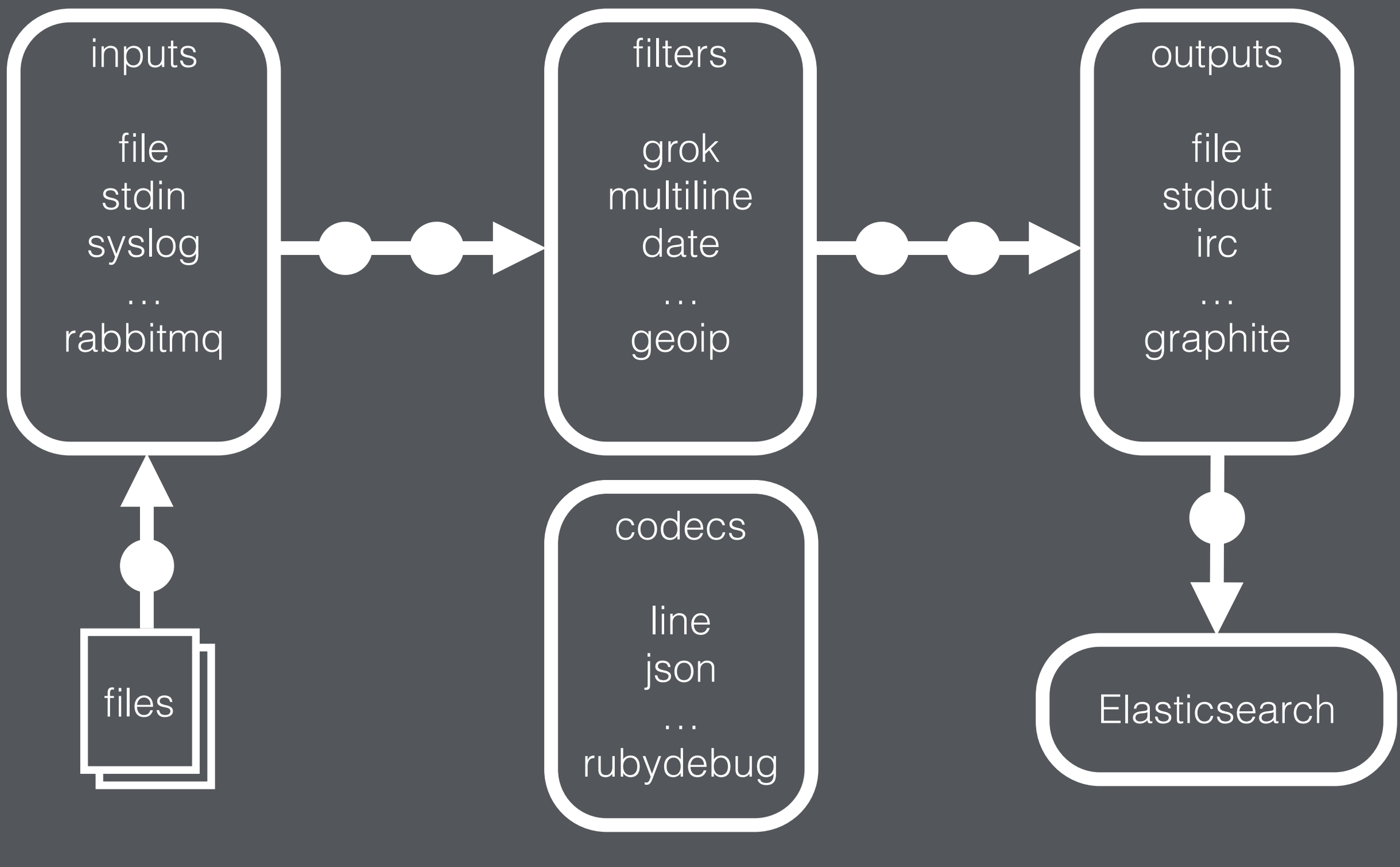

logstash-2014.10.05
shards: 2 * 2 | docs: 18209 | size: 9.63MB
logstash-2014.10.06
shards: 2 * 2 | docs: 300544 | size: 148.01MB
logstash-2014.10.07
shards: 2 * 2 | docs: 10010 | size: 5.27MB
node2
OKE-MACBOOK-MB.local - ine
heap
disk
cpu
0
1
node1
OKE-MACBOOK-MB.local - ine
heap
disk
cpu
0
1
logstash
OKE-MACBOOK-MB.local - inet[/1
heap
disk
cpu
unassigned shards

Logstash

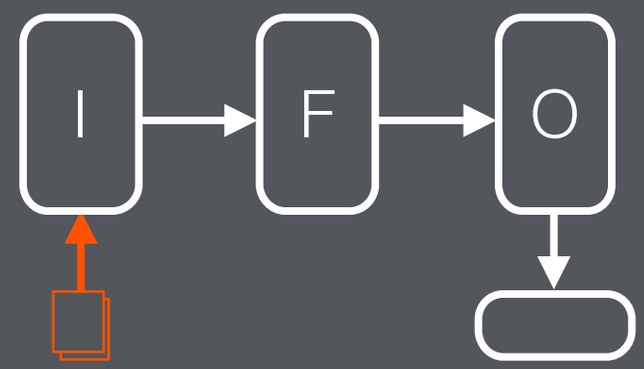
- Log stream processing
- Implemented with JRuby, runs on JVM
- Apache 2.0 license
- Created by Jordan Sissel
- Maintained and supported by Elasticsearch (the company)

Logstash logo from http://logstash.net/images/logstash.png

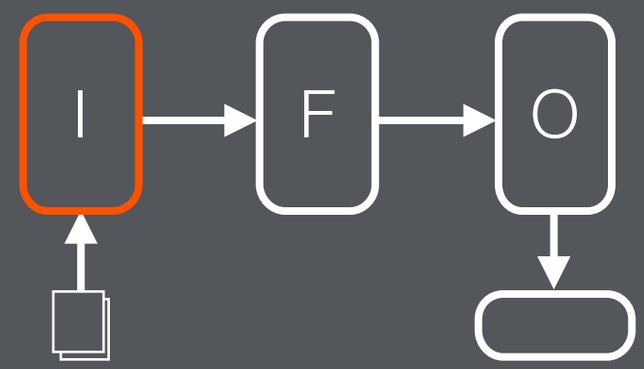
Logstash concepts



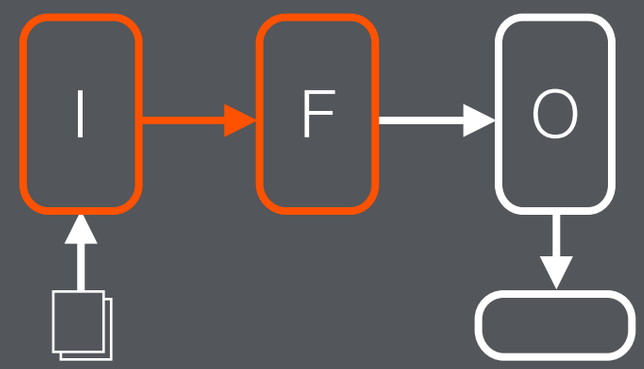
```
2014-10-11 12:52:22 GET /cgi-bin/ 404 10049 192.168.0.1 - - - 0.0
2014-10-11 12:52:22 GET /default.asp 404 10049 192.168.0.1 - - - 0.0
2014-10-11 12:52:22 GET /index.jsp 301 0 192.168.0.1 - - - 0.0010
2014-10-11 12:52:22 GET /scripts/formmail.html 404 10049 192.168.0.1 - - - 0.0
2014-10-11 12:52:22 GET /demo/../%3f.jsp 404 10049 192.168.0.1 - - - 0.0
2014-10-11 12:52:22 GET /q79w_38jg__.shtml 404 10049 192.168.0.1 - - - 0.0010
2014-10-11 12:52:22 GET /displaytable.php 404 10049 192.168.0.1 - - - 0.0
2014-10-11 12:52:22 GET /default.jsp 404 10049 192.168.0.1 - - - 0.0010
2014-10-11 12:52:22 GET /scripts/mailform.html 404 10049 192.168.0.1 - - - 0.0
2014-10-11 12:52:22 GET /forum/default.asp 404 10049 192.168.0.1 - - - 0.0010
```



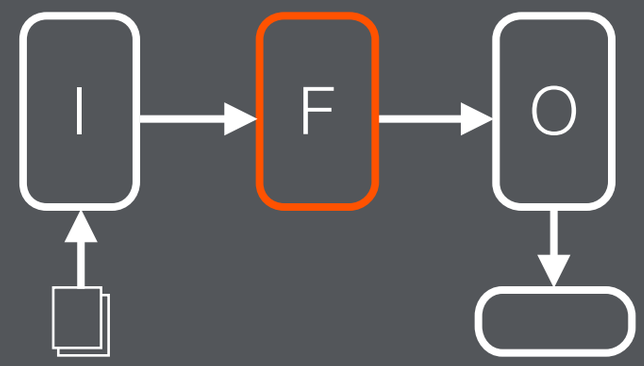
```
2014-10-11 12:52:22 GET /cgi-bin/ 404 10049 192.168.0.1 - - - 0.0
2014-10-11 12:52:22 GET /index.jsp 301 0 192.168.0.1 - - - 0.0010
2014-10-11 12:52:22 GET /demo/../%3f.jsp 404 10049 192.168.0.1 - - - 0.0
2014-10-11 12:52:22 GET /q79w_38jg__.shtml 404 10049 192.168.0.1 - - - 0.0010
2014-10-11 12:52:22 GET /displaytable.php 404 10049 192.168.0.1 - - - 0.0
2014-10-11 12:52:22 GET /default.jsp 404 10049 192.168.0.1 - - - 0.0010
2014-10-11 12:52:22 GET /scripts/mailform.html 404 10049 192.168.0.1 - - - 0.0
2014-10-11 12:52:22 GET /forum/default.asp 404 10049 192.168.0.1 - - - 0.0010
```



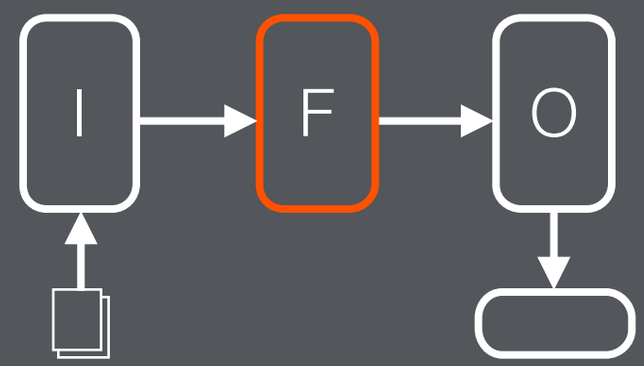
```
{
     "message" => "2014-10-11 12:52:22 GET /index.jsp 301 0
192.168.0.1 - - - 0.0010",
    "@version" => "1",
  "@timestamp" => "2014-11-13T20:39:41.753Z",
        "host" => "logstash.example.com",
        "path" => "/logs/access.log"
}
```



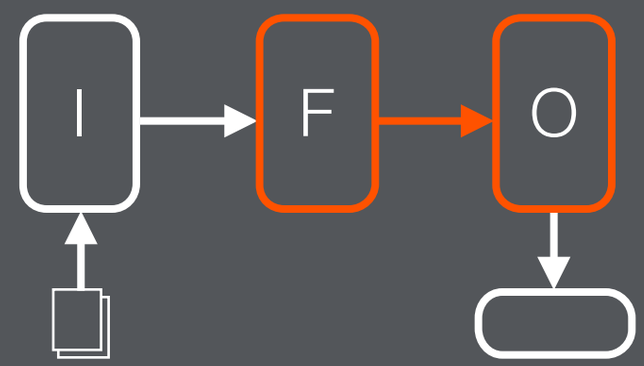
```
{
      "message" => "2014-10-11 12:52:22 GET /index.jsp 301 0
192.168.0.1 - - - 0.0010",
     "@version" => "1",
   "@timestamp" => "2014-11-13T20:39:41.753Z",
         "host" => "logstash.example.com",
         "path" => "/logs/access.log",
         "date" => "2014-10-11",
         "time" => "12:52:22",
       "method" => "GET",
     "uri-path" => "/index.jsp",
       "status" => 301,
        "bytes" => 0,
   "ip-address" => "192.168.0.1",
        "query" => "-",
     "referer" => "-",
   "user-agent" => "-",
    "elapsed_s" => 0.001
}
```



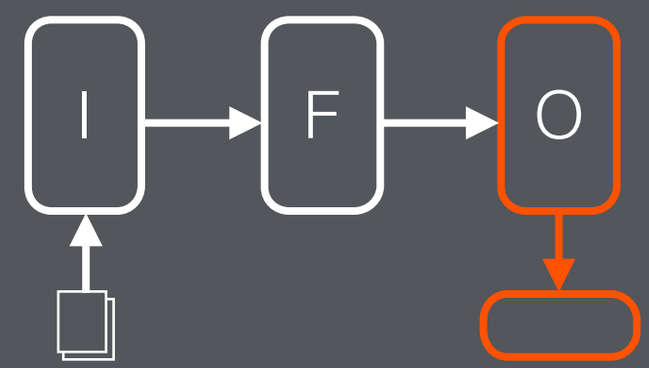
```
{
     "message" => "2014-10-11 12:52:22 GET /index.jsp 301 0
192.168.0.1 - - - 0.0010",
    "@version" => "1",
  "@timestamp" => "2014-10-11T17:52:22.000Z",
        "host" => "logstash.example.com",
        "path" => "/logs/access.log",
      "method" => "GET",
    "uri-path" => "/index.jsp",
      "status" => 301,
       "bytes" => 0,
  "ip-address" => "192.168.0.1",
       "query" => "-",
     "referer" => "-",
  "user-agent" => "-",
   "elapsed_s" => 0.001
}
```



```
{
     "message" => "2014-10-11 12:52:22 GET /index.jsp 301 0
192.168.0.1 - - - 0.0010",
    "@version" => "1",
  "@timestamp" => "2014-10-11T17:52:22.000Z",
        "host" => "server1.example.com",
        "path" => "/logs/access.log",
      "method" => "GET",
    "uri-path" => "/index.jsp",
      "status" => 301,
       "bytes" => 0,
  "ip-address" => "192.168.0.1",
       "query" => "-",
     "referer" => "-",
  "user-agent" => "-",
   "elapsed_s" => 0.001
}
```



```
{
     "message": "2014-10-11 12:52:22 GET /index.jsp 301 0
192.168.0.1 - - - 0.0010",
    "@version": "1",
  "@timestamp": "2014-10-11T17:52:22.000Z",
        "host": "server1.example.com",
        "path": "/logs/access.log",
      "method": "GET",
    "uri-path": "/index.jsp",
      "status": 301,
       "bytes": 0,
  "ip-address": "192.168.0.1",
       "query": "-",
    "referer": "-",
  "user-agent": "-",
   "elapsed_s": 0.001
}
```

Kibana

- Javascript application that interacts with Elasticsearch HTTP API
- Provides search, visualization, and dashboard capabilities
- Apache 2.0 license
- Maintained and supported by Elasticsearch (the company)

Kibana logo from https://github.com/elasticsearch/kibana/blob/3.0/src/img/kibana.png

Logstash Search

query

time window, refresh

load, save, configure

QUERY

*

7 days ago to a few seconds ago

FILTERING

time must
field : @timestamp
from : now-7d
to : now

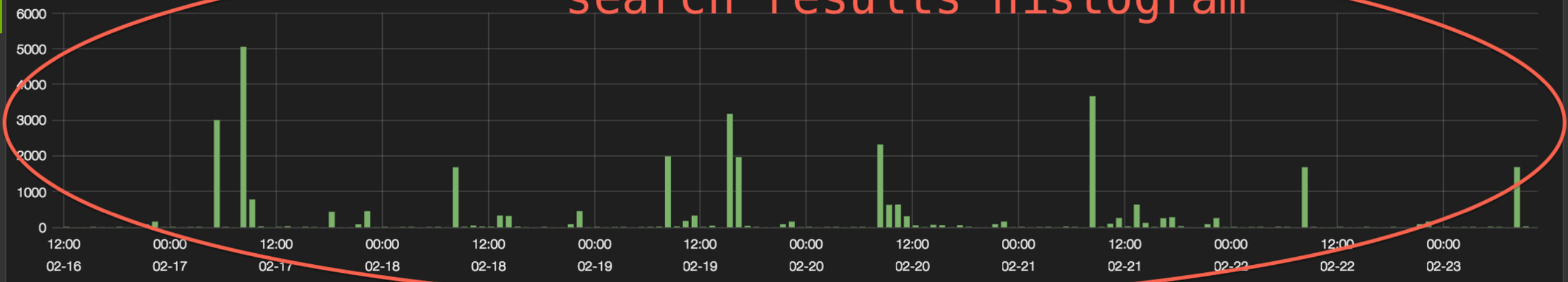
querystring must
query : type:application

filters

EVENTS OVER TIME

View | Zoom Out | * (35637) count per **1h** | (**35637** hits)

search results histogram



ALL EVENTS

field selection

0 to **100** of 500 available for paging

search results table

Fields

All (933) / **Current (18)**

Type to filter...

- [x] **@timestamp**
- [ ] @version
- [ ] _id
- [ ] _index
- [ ] _type
- [x] **application**
- [ ] correlationId
- [x] **host**
- [ ] logger
- [x] **logMessage**
- [x] **priority**
- [ ] sourceApplication
- [ ] threadName
- [ ] throwable.className

@timestamp	application	host	priority	logMessage
2015-02-23T10:38:34.086-05:00	[redacted]	[redacted]	WARN	Failed executing POST [redacted]
2015-02-23T10:25:00.000-05:00	[redacted]	[redacted]	WARN	Running job Perform Pending Reimbursements
2015-02-23T10:20:00.002-05:00	[redacted]	[redacted]	WARN	Running job Abandon Transactions
2015-02-23T10:15:00.002-05:00	[redacted]	[redacted]	WARN	Running job Payments Pending Settlement
2015-02-23T10:05:00.166-05:00	[redacted]	[redacted]	WARN	Running job Cancel Old Authorized Payments
2015-02-23T09:54:20.275-05:00	[redacted]	[redacted]	WARN	Failed executing POST [redacted]
2015-02-23T09:51:15.971-05:00	[redacted]	[redacted]	WARN	Failed executing POST [redacted]
2015-02-23T09:40:56.567-05:00	[redacted]	[redacted]	ERROR	An unexpected error occurred that prevented education transc...
2015-02-23T09:31:14.957-05:00	[redacted]	[redacted]	INFO	Finished UnsponsorProducersJob
2015-02-23T09:31:14.918-05:00	[redacted]	[redacted]	ERROR	[redacted]
2015-02-23T09:31:14.852-05:00	[redacted]	[redacted]	WARN	Resource not found
2015-02-23T09:31:14.796-05:00	[redacted]	[redacted]	WARN	An object was not found

Kibana Demo

7 days ago to a few seconds ago

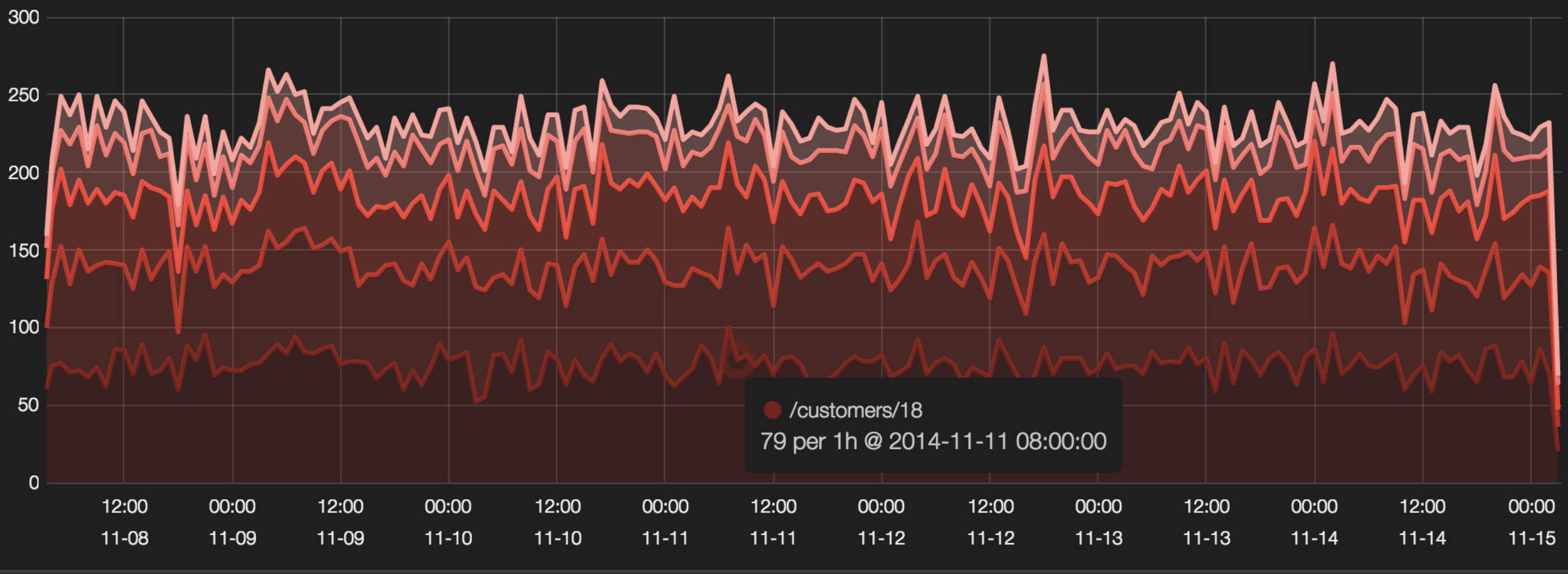
QUERY

host:"server1" AND method:"GET" AND uri-path.raw:/\/customers\/.*/

FILTERING

EVENTS OVER TIME

View | Zoom Out | /customers/18 (12747) /customers/18/orders (10484) /customers/5 (7724) /customers/5/orders (5173) /customers/32 (2602) count per **1h** | (**38730** hits)



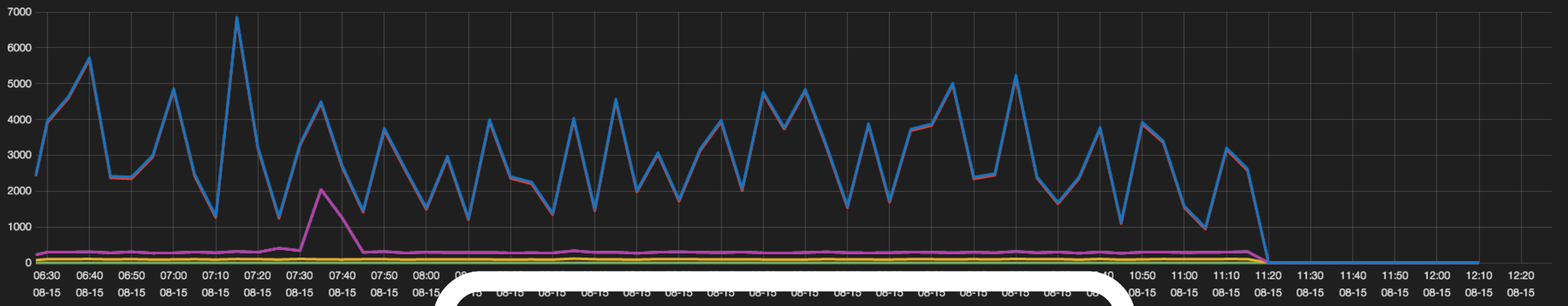
Access Dashboard

6 hours ago to 2 minutes ago

QUERY | FILTERING ★

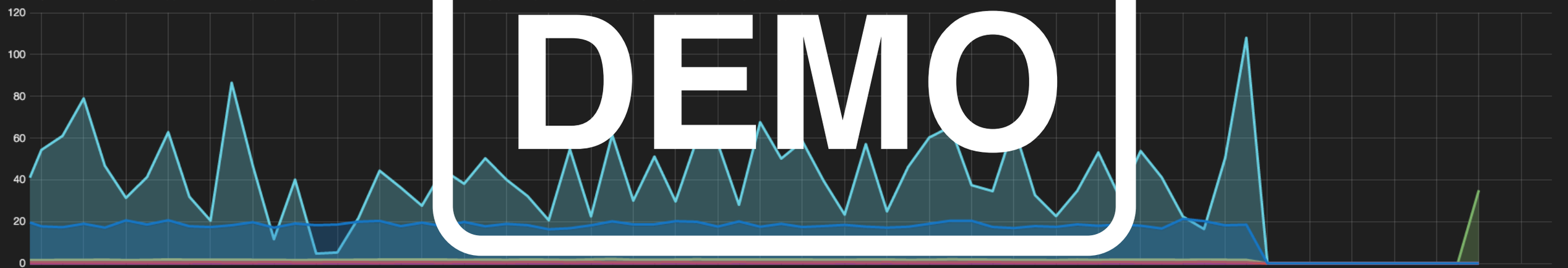
REQUEST RATE

View | Zoom Out | echo (1) | connection (5708) | edge (14201) | notification (0) | identity (154755) | license (2825) count per **5m** | (**177490** hits)

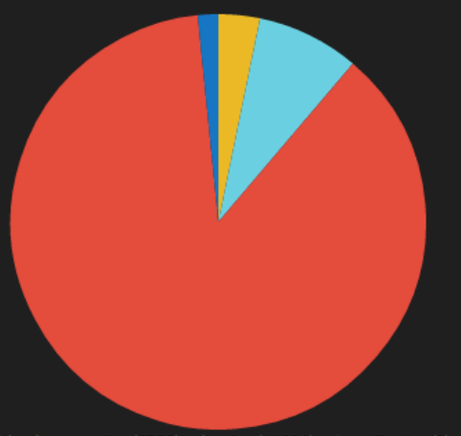


AVERAGE RESPONSE TIMES

View | Zoom Out | echo (1) | connection (5708) | edge (14201) | notification (| identity (154755) | license (2825) **elapsedTimeMillis** mean per **5m** | (**177490** hits)



REQUESTS



echo (1) | connection (5708) | edge (14201) | notification (0) | identity (154755) | license (2825)

RESPONSE SIZE (BYTES)

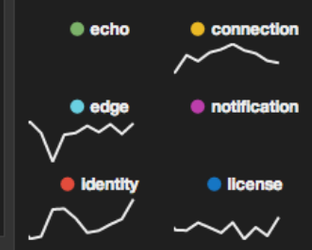
453 (mean)

Query	count	min	max	mean	total	variance	std_deviation
echo	1	6	6	6	6	0	0
connection	5,708	106	263,551	2,814	16,062,867	219,314,862	14,809
edge	14,201	0	642,733	3,003	42,648,398	536,104,447	23,154
notification	0	Infinity	-Infinity	0	0	0	0
identity	154,755	102	221	124	19,163,573	167	13
license	2,825	307	15,807	904	2,554,249	3,986,432	1,997

REQUEST RATE

- ▲0% (echo)
- ▼-1.77% (connection)
- ▼-1.42% (edge)
- ▲?% (notification)
- ▼-1.73% (identity)
- ▼-1.7% (license)

RESPONSE TIME



Demo/Quick Start Project

- Downloads, installs, and configures Elasticsearch, Logstash, Kibana, and kopf plugin
- Provides scripts for starting, stopping, and resetting data
- Generates log data to index
- Prerequisites
  - JDK (1.7 or 1.8 should work)
  - node.js
- Caveats
  - Shell scripts require OS X or Linux

Setup Steps

1. https://github.com/dgrabows/elk-demo
2. Clone or download zip
3. Run ./install.sh
4. Run ./start-all.sh
5. Kibana: http://localhost/9200
6. Admin UI (kopf): http://localhost:9200/_plugin/kopf

Experiences

- Don’t underestimate power of rsync, find, grep, awk, sed, wc, cron, etc.
- Elasticsearch and Kibana very effective for exploring log data
- Elasticsearch has additional capabilities not exposed through Kibana (e.g. percentile aggregates)

Experiences (cont.)

- One effective pattern
  - Analyze aggregate metrics with Elasticsearch, node.js, and spreadsheets
  - Drilled down into problem areas with Kibana and Elasticsearch
- You could get a lot done with a 4-8 core/32 GB RAM/1 TB disk server

Questions?

References

- https://github.com/dgrabows/presentations (this presentation)
- https://github.com/dgrabows/log-gen (access log generation code)
- http://www.elasticsearch.org
- http://logstash.net/
- https://github.com/elasticsearch (Elasticsearch, Logstash, Kibana repos)
- https://github.com/lmenezes/elasticsearch-kopf (kopf plugin)

Attributions

Full attributions provided inline, where practical

Title Slide

Untitled (https://www.flickr.com/photos/81disasters/7094336635/) by Thomas Xu (https://www.flickr.com/photos/81disasters/), used under CC BY-SA 2.0 (https://creativecommons.org/licenses/by-sa/2.0/) / Converted to black and white

Dan Grabowski

dan.grabowski@gmail.com