logstash

a love story

@jordansissel
OSCON 2011

Who am I?

sysadmin.

coder.

Loggly, Inc.

rum and beer.

no tequila, please.

Data first.
Marketing second.

terminology?

the problem

the problem

unstructured unknown data

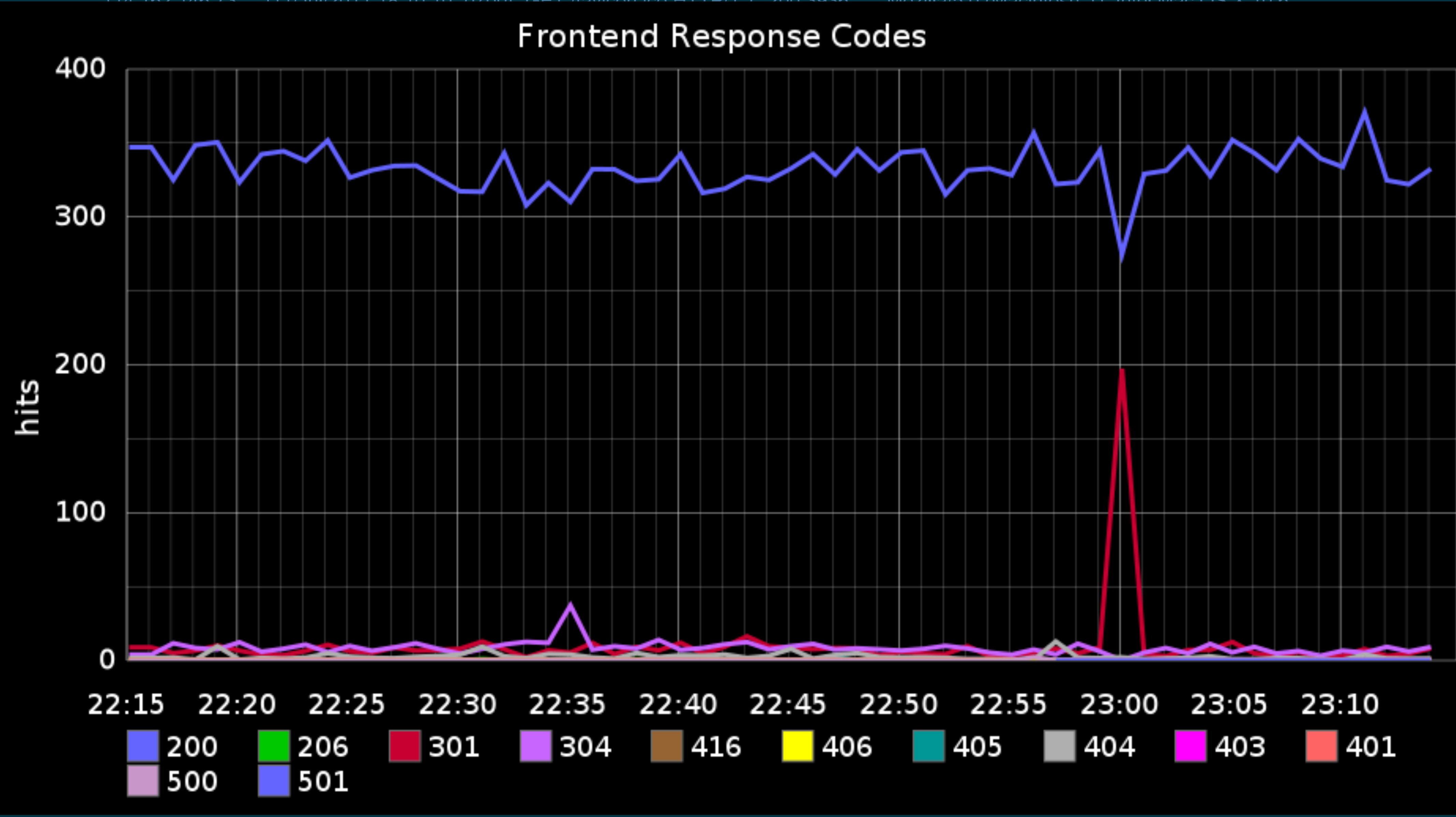
the problem

"keep them around just in case"

full disk? just add logrotate!

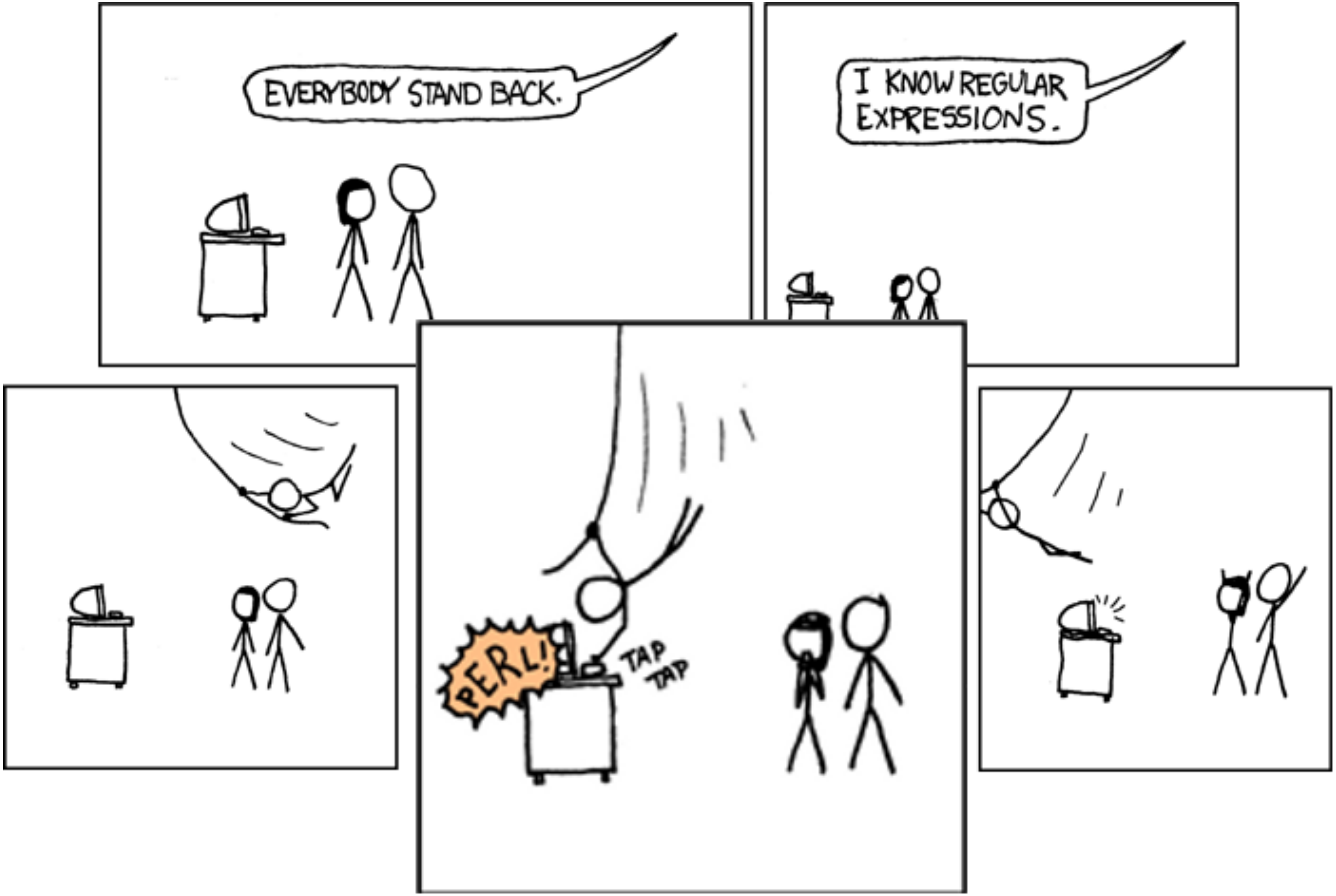
```
67.195.183.71 - - [11/Jul/2011:18:08:21 -0700] "POST /hackday08/randomtags.py HTTP/1.0" 200 142 "http://pipes.yahoo.com/pipes/pipe.
info?_id=oFqDT3KB3RG8tyjP073fcQ" "Yahoo Pipes 1.0"
173.203.57.156 - - [11/Jul/2011:18:08:27 -0700] "GET /files/logstash/logstash-1.0.14-monolithic.jar HTTP/1.1" 200 34408788 "-" "Chef
Client/0.9.16 (ruby-1.8.7-p249; ohai-0.5.8; x86_64-linux; +http://opscode.com)"
216.150.136.52 - - [11/Jul/2011:18:09:08 -0700] "GET /favicon.ico HTTP/1.0" 200 3898 "-" "Safari/6533.21.1 CFNetwork/454.12.4 Darwin/10.
8.0 (i386) (iMac11%2C1)"
114.162.146.73 - - [11/Jul/2011:18:10:10 -0700] "GET /favicon.ico HTTP/1.1" 200 3936 "-" "Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10.6;
en-US; rv:1.9.2.18) Gecko/20110614 Firefox/3.6.18"
95.65.77.99 - - [11/Jul/2011:18:10:22 -0700] "GET /blog/geekery/xvfb-firefox.html HTTP/1.0" 200 28389 "http://www.semicomplete.
com/blog/geekery/xvfb-firefox.html" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322; .NET CLR 2.0.50727)"
109.107.35.119 - - [11/Jul/2011:18:10:22 -0700] "GET / HTTP/1.1" 200 38544 "-" "check_http/v1.4.15 (nagios-plugins 1.4.15)"
95.65.77.99 - - [11/Jul/2011:18:10:23 -0700] "POST /blog/geekery/xvfb-firefox HTTP/1.0" 200 28389 "http://www.semicomplete.
com/blog/geekery/xvfb-firefox.html" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322; .NET CLR 2.0.50727)"
109.107.35.119 - - [11/Jul/2011:18:10:44 -0700] "GET / HTTP/1.1" 200 38544 "-" "check_http/v1.4.15 (nagios-plugins 1.4.15)"
123.126.50.75 - - [11/Jul/2011:18:11:18 -0700] "GET / HTTP/1.1" 200 12810 "-" "Sogou web spider/4.0(+http://www.sogou.
com/docs/help/webmasters.htm#07)"
67.195.111.152 - - [11/Jul/2011:18:11:26 -0700] "GET /misc/comment-form.html HTTP/1.0" 200 1555 "-" "Mozilla/5.0 (compatible; Yahoo!
Slurp; http://help.yahoo.com/help/us/ysearch/slurp)"
66.249.71.195 - - [11/Jul/2011:18:11:27 -0700] "GET /files/xdotool/docs/man/ HTTP/1.1" 200 710 "-" "Mozilla/5.0 (compatible; Googlebot/2.1;
+http://www.google.com/bot.html)"
87.248.122.128 - - [11/Jul/2011:18:12:05 -0700] "POST /hackday08/randomtags.py HTTP/1.0" 200 142 "http://pipes.yahoo.com/pipes/pipe.
info?_id=oFqDT3KB3RG8tyjP073fcQ" "Yahoo Pipes 1.0"
87.248.122.128 - - [11/Jul/2011:18:12:06 -0700] "POST /hackday08/randomtags.py HTTP/1.0" 200 142 "http://pipes.yahoo.com/pipes/pipe.
info?_id=oFqDT3KB3RG8tyjP073fcQ" "Yahoo Pipes 1.0"
87.248.122.128 - - [11/Jul/2011:18:12:06 -0700] "POST /hackday08/randomtags.py HTTP/1.0" 200 142 "http://pipes.yahoo.com/pipes/pipe.
info?_id=oFqDT3KB3RG8tyjP073fcQ" "Yahoo Pipes 1.0"
87.248.122.128 - - [11/Jul/2011:18:12:07 -0700] "POST /hackday08/randomtags.py HTTP/1.0" 200 142 "http://pipes.yahoo.com/pipes/pipe.
info?_id=oFqDT3KB3RG8tyjP073fcQ" "Yahoo Pipes 1.0"
209.85.224.88 - - [11/Jul/2011:18:13:06 -0700] "GET /?flav=rss20 HTTP/1.1" 200 27853 "-" "FeedBurner/1.0 (http://www.FeedBurner.com)"
190.245.143.36 - - [11/Jul/2011:18:13:43 -0700] "GET /favicon.ico HTTP/1.0" 200 3898 "-" "Safari/6533.21.1 CFNetwork/454.12.4 Darwin/10.
8.0 (i386) (MacBookPro5%2C1)"
64.183.2.50 - - [11/Jul/2011:18:14:26 -0700] "GET /blog/geekery/graphing-with-ruby.html HTTP/1.1" 200 5424 "http://www.google.com/url?
sa=t&source=web&cd=10&ved=0CFsQFjAJ&url=http%3A%2F%2Fwww.semicomplete.com%2Fblog%2Fgeekery%2Fgraphing-with-ruby.
html&rct=j&q=RVG%20ruby%20drawing%20lines&ei=zJ8bTon8D-bTiAL61ZD4CA&usg=AFQjCNE3YwOhs35ikC0o91NvcMI0_cBGPg"
"Mozilla/5.0 (Macintosh; Intel Mac OS X 10_6_7) AppleWebKit/534.30 (KHTML, like Gecko) Chrome/12.0.742.112 Safari/534.30"
64.183.2.50 - - [11/Jul/2011:18:14:26 -0700] "GET /reset.css HTTP/1.1" 200 910 "http://www.semicomplete.com/blog/geekery/graphing-with-
ruby.html" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_6_7) AppleWebKit/534.30 (KHTML, like Gecko) Chrome/12.0.742.112 Safari/534.30"
64.183.2.50 - - [11/Jul/2011:18:14:26 -0700] "GET /style2.css HTTP/1.1" 200 1819 "http://www.semicomplete.com/blog/geekery/graphing-
with-ruby.html" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_6_7) AppleWebKit/534.30 (KHTML, like Gecko) Chrome/12.0.742.112
Safari/534.30"
64.183.2.50 - - [11/Jul/2011:18:14:26 -0700] "GET /images/jordan-80.png HTTP/1.1" 200 6442 "http://www.semicomplete.
com/blog/geekery/graphing-with-ruby.html" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_6_7) AppleWebKit/534.30 (KHTML, like Gecko)
Chrome/12.0.742.112 Safari/534.30"
64.183.2.50 - - [11/Jul/2011:18:14:27 -0700] "GET /files/blogposts/20090519/testgraph.gif HTTP/1.1" 200 9164 "http://www.semicomplete.
com/blog/geekery/graphing-with-ruby.html" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_6_7) AppleWebKit/534.30 (KHTML, like Gecko)
Chrome/12.0.742.112 Safari/534.30"
```

```
67.195.183.71 - - [11/Jul/2011:18:08:21 -0700] "POST /hackday08/randomtags.py HTTP/1.0" 200 142 "http://pipes.yahoo.com/pipes/pipe.info?_id=oFqDT3KB3RG8tyjP073fcQ" "Yahoo Pipes 1.0"
173.203.57.156 - - [11/Jul/2011:18:08:27 -0700] "GET /files/logstash/logstash-1.0.14-monolithic.jar HTTP/1.1" 200 34408788 "-" "Chef Client/0.9.16 (ruby-1.8.7-p249; ohai-0.5.8; x86_64-linux; +http://opscode.com)"
216.150.136.52 - - [11/Jul/2011:18:09:08 -0700] "GET /favicon.ico HTTP/1.0" 200 3898 "-" "Safari/6533.21.1 CFNetwork/454.12.4 Darwin/10.8.0 (i386) (iMac11%2C1)"
114.162.146.73 - - [11/Jul/2011:18:10:10 -0700] "GET /favicon.ico HTTP/1.1" 200 3936 "-" "Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_6;
```



```
64.183.2.50 - - [11/Jul/2011:18:14:27 -0700] "GET /files/blogposts/20090519/testgraph.gif HTTP/1.1" 200 9164 "http://www.semicomplete.com/blog/geekery/graphing-with-ruby.html" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_6_7) AppleWebKit/534.30 (KHTML, like Gecko) Chrome/12.0.742.112 Safari/534.30"
```

```
67.195.183.71 - - [11/Jul/2011:18:08:21 -0700] "POST /hackday08/randomtags.py HTTP/1.0" 200 142 "http://pipes.yahoo.com/pipes/pipe.
info?_id=oFqDT3KB3RG8tyjP073fcQ" "Yahoo Pipes 1.0"
173.203.57.156 - - [11/Jul/2011:18:08:27 -0700] "GET /files/logstash/logstash-1.0.14-monolithic.jar HTTP/1.1" 200 34408788 "-" "Chef
Client/0.9.16 (ruby-1.8.7-p249; ohai-0.5.8; x86_64-linux; +http://opscode.com)"
216.150.136.52 - - [11/Jul/2011:18:09:08 -0700] "GET /favicon.ico HTTP/1.0" 200 3898 "-" "Safari/6533.21.1 CFNetwork/454.12.4 Darwin/10.
```



```
com/blog/geekery/graphing-with-ruby.html" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_6_7) AppleWebKit/534.30 (KHTML, like Gecko)
Chrome/12.0.742.112 Safari/534.30"
```

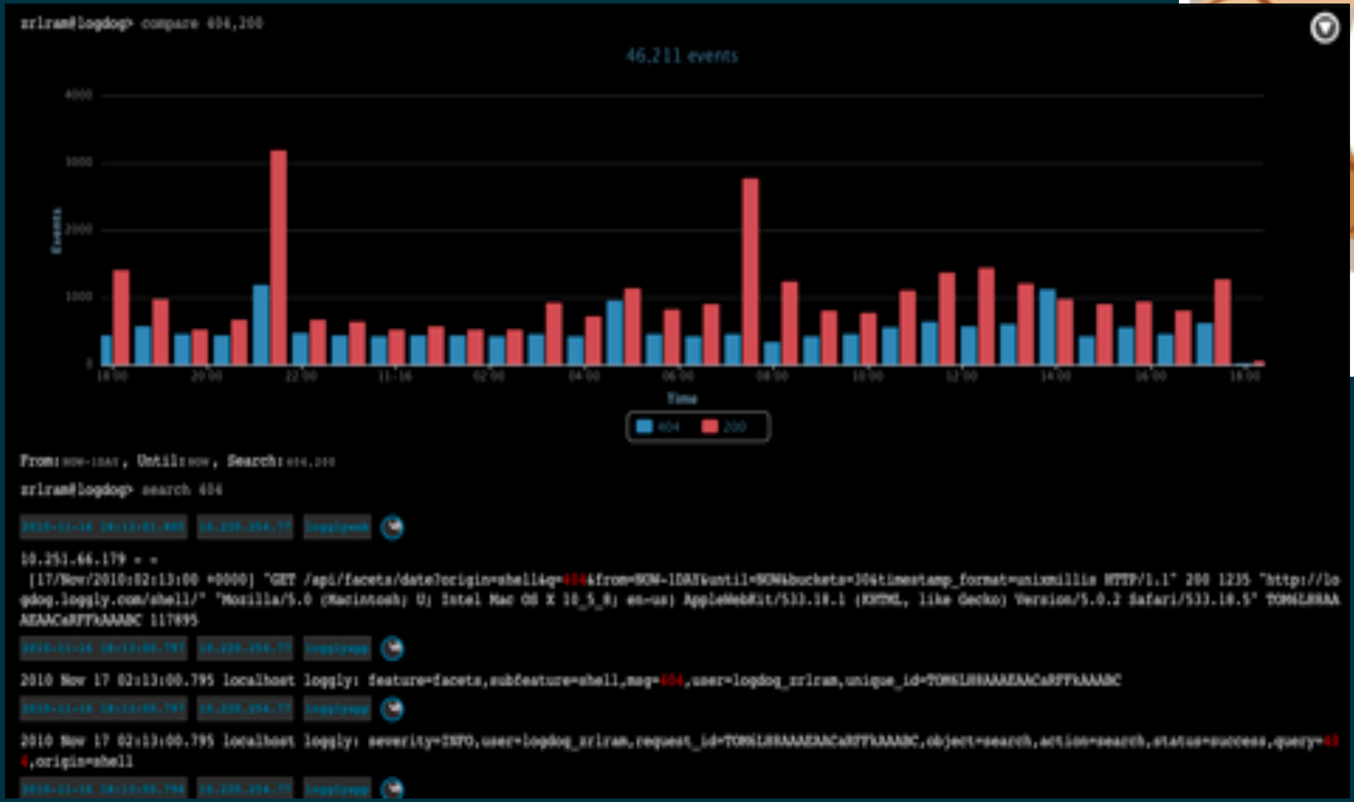
ssh, grep, and tail do not scale well

ssh, grep, and tail may not be available

Other stuff in this space

Loggly (I work here.)

Logging as a Service. Shell UI. APIs.



```
zrlram@logdog> compare 404,200
```

```
From: NOW-1DAY , Until: NOW , Search: 404,200
zrlram@logdog> search 404
2010-11-16 18:13:01.805  10.220.254.77  logglyweb
10.251.66.179 - -
 [17/Nov/2010:02:13:00 +0000] "GET /api/facets/date?origin=shell&q=404&from=NOW-1DAY&until=NOW&buckets=30&timestamp_format=unixmillis HTTP/1.1" 200 1235 "http://logdog.loggly.com/shell/" "Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_5_8; en-us) AppleWebKit/533.18.1 (KHTML, like Gecko) Version/5.0.2 Safari/533.18.5" TOM6LH8AAAEAACaRFFkAAABC 117895
2010-11-16 18:13:00.797  10.220.254.77  logglyapp
2010 Nov 17 02:13:00.795 localhost loggly: feature=facets,subfeature=shell,msg=404,user=logdog_zrlram,unique_id=TOM6LH8AAAEAACaRFFkAAABC
2010-11-16 18:13:00.797  10.220.254.77  logglyapp
2010 Nov 17 02:13:00.795 localhost loggly: severity=INFO,user=logdog_zrlram,request_id=TOM6LH8AAAEAACaRFFkAAABC,object=search,action=search,status=success,query=404,origin=shell
2010-11-16 18:13:00.796  10.220.254.77  logglyapp
```

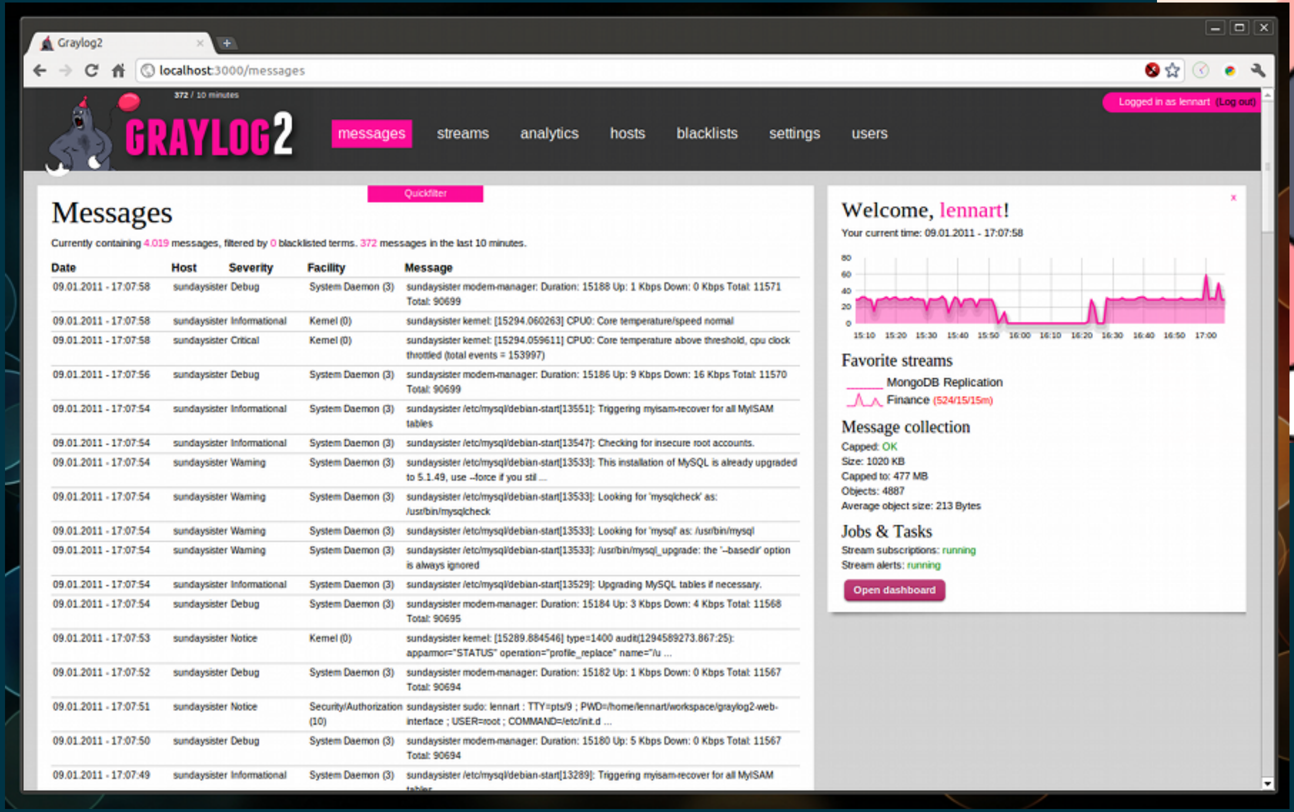
Nice Beaver

loggly

Other stuff in this space

Graylog2

Open Source. MongoDB. Ruby+Java



Messages

Currently containing 4.019 messages, filtered by 0 blacklisted terms. 372 messages in the last 10 minutes.

Date	Host	Severity	Facility	Message
09.01.2011 - 17:07:58	sundaysister	Debug	System Daemon (3)	sundaysister modem-manager: Duration: 15188 Up: 1 Kbps Down: 0 Kbps Total: 11571 Total: 90699
09.01.2011 - 17:07:58	sundaysister	Informational	Kernel (0)	sundaysister kernel: [15294.060263] CPU0: Core temperature/speed normal
09.01.2011 - 17:07:58	sundaysister	Critical	Kernel (0)	sundaysister kernel: [15294.059611] CPU0: Core temperature above threshold, cpu clock throttled (total events = 153997)
09.01.2011 - 17:07:56	sundaysister	Debug	System Daemon (3)	sundaysister modem-manager: Duration: 15186 Up: 9 Kbps Down: 16 Kbps Total: 11570 Total: 90699
09.01.2011 - 17:07:54	sundaysister	Informational	System Daemon (3)	sundaysister /etc/mysql/debian-start[13551]: Triggering myisam-recover for all MyISAM tables
09.01.2011 - 17:07:54	sundaysister	Informational	System Daemon (3)	sundaysister /etc/mysql/debian-start[13547]: Checking for insecure root accounts.
09.01.2011 - 17:07:54	sundaysister	Warning	System Daemon (3)	sundaysister /etc/mysql/debian-start[13533]: This installation of MySQL is already upgraded to 5.1.49, use --force if you stil ...
09.01.2011 - 17:07:54	sundaysister	Warning	System Daemon (3)	sundaysister /etc/mysql/debian-start[13533]: Looking for 'mysqlcheck' as: /usr/bin/mysqlcheck
09.01.2011 - 17:07:54	sundaysister	Warning	System Daemon (3)	sundaysister /etc/mysql/debian-start[13533]: Looking for 'mysql' as: /usr/bin/mysql
09.01.2011 - 17:07:54	sundaysister	Warning	System Daemon (3)	sundaysister /etc/mysql/debian-start[13533]: /usr/bin/mysql_upgrade: the '--basedir' option is always ignored
09.01.2011 - 17:07:54	sundaysister	Informational	System Daemon (3)	sundaysister /etc/mysql/debian-start[13529]: Upgrading MySQL tables if necessary.
09.01.2011 - 17:07:54	sundaysister	Debug	System Daemon (3)	sundaysister modem-manager: Duration: 15184 Up: 3 Kbps Down: 4 Kbps Total: 11568 Total: 90695
09.01.2011 - 17:07:53	sundaysister	Notice	Kernel (0)	sundaysister kernel: [15289.884546] type=1400 audit(1294589273.867:25): apparmor="STATUS" operation="profile_replace" name="/u ...
09.01.2011 - 17:07:52	sundaysister	Debug	System Daemon (3)	sundaysister modem-manager: Duration: 15182 Up: 1 Kbps Down: 0 Kbps Total: 11567 Total: 90694
09.01.2011 - 17:07:51	sundaysister	Notice	Security/Authorization (10)	sundaysister sudo: lennart : TTY=pts/9 ; PWD=/home/lennart/workspace/graylog2-web-interface ; USER=root ; COMMAND=/etc/init.d ...
09.01.2011 - 17:07:50	sundaysister	Debug	System Daemon (3)	sundaysister modem-manager: Duration: 15180 Up: 5 Kbps Down: 0 Kbps Total: 11567 Total: 90694
09.01.2011 - 17:07:49	sundaysister	Informational	System Daemon (3)	sundaysister /etc/mysql/debian-start[13289]: Triggering myisam-recover for all MyISAM tables

Welcome, lennart!

Your current time: 09.01.2011 - 17:07:58

Favorite streams

MongoDB Replication

Finance (524/15/15m)

Message collection

Capped: OK
Size: 1020 KB
Capped to: 477 MB
Objects: 4887
Average object size: 213 Bytes

Jobs & Tasks

Stream subscriptions: running
Stream alerts: running

What is logstash?

take events,
ship them somewhere,
make them useful,

but don't be annoying.

logstash is open source.

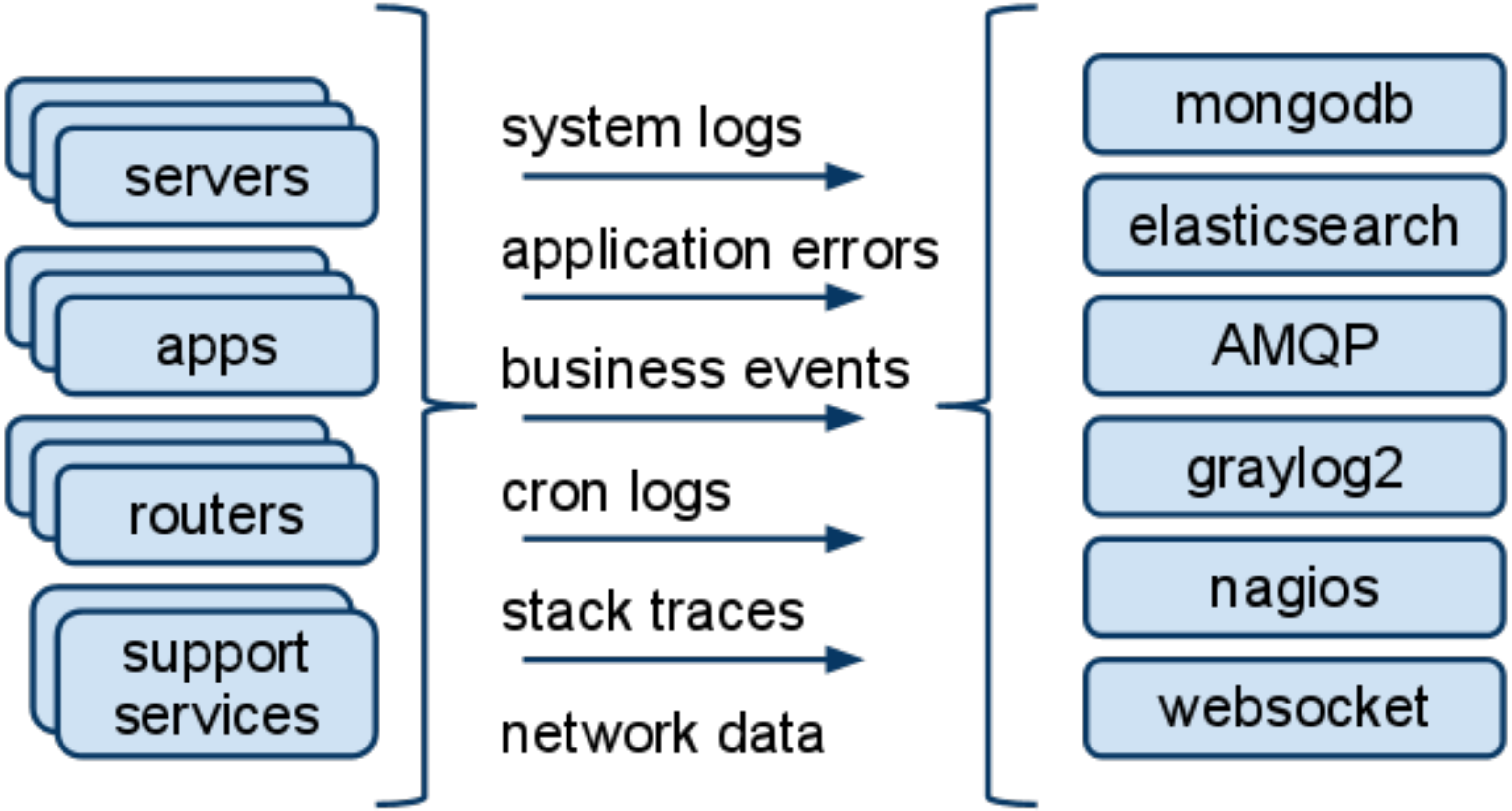
logstash should fit your environment.

Not the other way around.

logstash should be usable in parts as well as whole.

logstash should be extendable.

logstash - sane log management



servers

apps

routers

support services

system logs

application errors

business events

cron logs

stack traces

network data

mongodb

elasticsearch

AMQP

graylog2

nagios

websocket

ship any event to anywhere over any protocol input | filter | output

[input] | filter | output

email

files

syslog

AMQP

Flume

tcp socket

HTTP

Beanstalk

STOMP

twitter

input | [filter] | output

timestamp parsing

drop events

anonymize

parse fields

multiline joins

input | filter | **[output]**

AMQP

TCP

STOMP

Graylog2/GELF

ElasticSearch

syslog

Redis

MongoDB

beanstalk

Nagios

WebSockets

input: file

forever tail a file. tracks log rotation.

one line is one event.

input: syslog

take syslog messages over the network

input: stdin

one event per line read.

input: messaging tools

AMQP, STOMP, beanstalk, redis

input: twitter's stream api

tracking Biebers per second

filter: date parsing

because everyone invents their own crappy time format.

filter: grep

drop stuff you don't want, etc...

filter: grok

pattern parsing ninja

filter: grok discovery

inverted grok, more on this later.

filter: multiline

join related log lines in a single event.

output: ElasticSearch

scalable indexing for your logs

elasticsearch.

output: messaging tools

AMQP, STOMP, Beanstalk, Redis

RabbitMQ™

Messaging that just works

output: graylog2 (GELF)

ship events to a graylog2 server

output: mongodb

write your events to whatever the cool kids are using today.

mongoDB

output: nagios

trigger nagios alerts from events

Nagios®

output: websocket

stream events to your browser

grok

regular expressions are pretty awesome

grok

regular expressions are pretty awful

grok

so if regular expressions are awful...

grok

why do developers keep making crappy log formats

```
(((?:(\b(?:[0-9A-Za-z][0-9A-Za-z-]{0,62})(?:\.(?:[0-9A-Za-z][0-9A-
Za-z-]{0,62}))*(\.?|\b))|(((?<![0-9])(?:(?:25[0-5]|2[0-4][0-9]|[0-1]?
[0-9]{1,2})[.](?:25[0-5]|2[0-4][0-9]|[0-1]?[0-9]{1,2})[.](?:25[0-5]|2
[0-4][0-9]|[0-1]?[0-9]{1,2})[.](?:25[0-5]|2[0-4][0-9]|[0-1]?[0-9]
{1,2}))(?![0-9])))) (([a-zA-Z0-9_-]+)) (([a-zA-Z0-9_-]+)) \[(((?:3
[01]|[1-2]?[0-9]|0?[1-9]))/(\b(?:Jan(?:uary)?|Feb(?:ruary)?|Mar(?:ch)?|Apr
(?:il)?|May|Jun(?:e)?|Jul(?:y)?|Aug(?:ust)?|Sep(?:tember)?|Oct(?:ober)?|Nov(?:
ember)?|Dec(?:ember)?)\b)/([0-9]+):((?!<[0-9])((?:2[0123]|[01][0-9])):((?:[0-5][0-
9]))(?::((?:(?:[0-5][0-9]|60)(?:[.,][0-9]+)?)))(?![0-9])) ((?:[+-]?(?:[0-9]+))))\]
"(\b\w+\b) (((?:/[A-Za-z0-9$.+!*'(),~:#%_-]*)+)(?:(\?[A-Za-z0-9$.+!*'(),~#%&/=:;_-
]*))?) HTTP/((?:((?<![0-9.+-])(?>[+-]?(?:(?:[0-9]+(?:\.[0-9]+)?)|(?:\.[0-9]+))))))" ((?:((?<![0-9.+-])(?>[+-]?
(?:(?:[0-9]+(?:\.[0-9]+)?)|(?:\.[0-9]+)))))) (?:((?:((?<![0-9.+-])(?>[+-]?(?:(?:[0-9]+(?:\.[0-9]+)?)|(?:\.[0-9]+))))))|-
) "(?:(([A-Za-z]+(\+[A-Za-z+]+)?)://(?:(([a-zA-Z0-9_-]+))(?::[^@]*)?@)?(?:(((?:(\b(?:[0-9A-Za-z][0-9A-Za-z-]
{0,62})(?:\.(?:[0-9A-Za-z][0-9A-Za-z-]{0,62}))*(\.?|\b))|(((?<![0-9])(?:(?:25[0-5]|2[0-4][0-9]|[0-1]?[0-9]{1,2})[.](?:25[0-5]|2[0-4][0-9]|
[0-1]?[0-9]{1,2})[.](?:25[0-5]|2[0-4][0-9]|[0-1]?[0-9]{1,2})[.](?:25[0-5]|2[0-4][0-9]|[0-1]?[0-9]{1,2}))(?![0-9]))))(?::(\b(?:[0-9]+)\b))?))?
(?:(((?:/[A-Za-z0-9$.+!*'(),~:#%_-]*)+)(?:(\?[A-Za-z0-9$.+!*'(),~#%&/=:;_-]*))?))?)|-)" (((?:(?<!\\)(?:"(?:\\.|[^\\"])*"|(?:'(?:\\.|[^\\'])*')|(?:
`(?:\\.|[^\\`])*`))))))
(((?:(\b(?:[0-9A-Za-z][0-9A-Za-z-]{0,62})(?:\.(?:[0-9A-Za-z][0-9A-
Za-z-]{0,62}))*(\.?|\b))|(((?<![0-9])(?:(?:25[0-5]|2[0-4][0-9]|[0-1]?
[0-9]{1,2})[.](?:25[0-5]|2[0-4][0-9]|[0-1]?[0-9]{1,2})[.](?:25[0-5]|2
[0-4][0-9]|[0-1]?[0-9]{1,2})[.](?:25[0-5]|2[0-4][0-9]|[0-1]?[0-9]
{1,2}))(?![0-9])))) (([a-zA-Z0-9_-]+)) (([a-zA-Z0-9_-]+)) \[(((?:3
[01]]|[1-2]?[0-9]|0?[1-9]))/(\b(?:Jan(?:uary)?|Feb(?:ruary)?|Mar(?:ch)?|Apr
(?:il)?|May|Jun(?:e)?|Jul(?:y)?|Aug(?:ust)?|Sep(?:tember)?|Oct(?:ober)?|Nov(?:
ember)?|Dec(?:ember)?)\b)/([0-9]+):((?!<[0-9])((?:2[0123]|[01][0-9])):((?:[0-5][0-
9]))(?::((?:(?:[0-5][0-9]|60)(?:[.,][0-9]+)?)))(?![0-9])) ((?:[+-]?(?:[0-9]+)))))\]
"(\b\w+\b) (((?:/[A-Za-z0-9$.+!*'(),~:#%_-]*)+)(?:(\?[A-Za-z0-9$.+!*'(),~#%&/=:;_-
]*))?) HTTP/((?:((?<![0-9.+-])(?>[+-]?(?:(?:[0-9]+(?:\.[0-9]+)?)|(?:\.[0-9]+))))))" ((?:((?<![0-9.+-])(?>[+-]?
(?:(?:[0-9]+(?:\.[0-9]+)?)|(?:\.[0-9]+))))))) (?:((?:((?<![0-9.+-])(?>[+-]?(?:(?:[0-9]+(?:\.[0-9]+)?)|(?:\.[0-9]+)))))))|-
) "(?:(([A-Za-z]+(\+[A-Za-z+]+)?)://(?:(([a-zA-Z0-9_-]+))(?::[^@]*)?@)?(?:(((?:(\b(?:[0-9A-Za-z][0-9A-Za-z-]
{0,62})(?:\.(?:[0-9A-Za-z][0-9A-Za-z-]{0,62}))*(\.?|\b))|(((?<![0-9])(?:(?:25[0-5]|2[0-4][0-9]|[0-1]?[0-9]{1,2})[.](?:25[0-5]|2[0-4][0-9]|
[0-1]?[0-9]{1,2})[.](?:25[0-5]|2[0-4][0-9]|[0-1]?[0-9]{1,2})[.](?:25[0-5]|2[0-4][0-9]|[0-1]?[0-9]{1,2}))(?![0-9]))))(?::(\b(?:[0-9]+)\b))?))?
(?:(((?:/[A-Za-z0-9$.+!*'(),~:#%_-]*)+)(?:(\?[A-Za-z0-9$.+!*'(),~#%&/=:;_-]*))?))?|-)" (((?:(?<!\\)(?:"(?:\\.|[^\\"])*"|(?:'(?:\\.|[^\\'])*')|(?:
`(?:\\.|[^\\`])*`))))))
```

This pattern is for apache access logs.

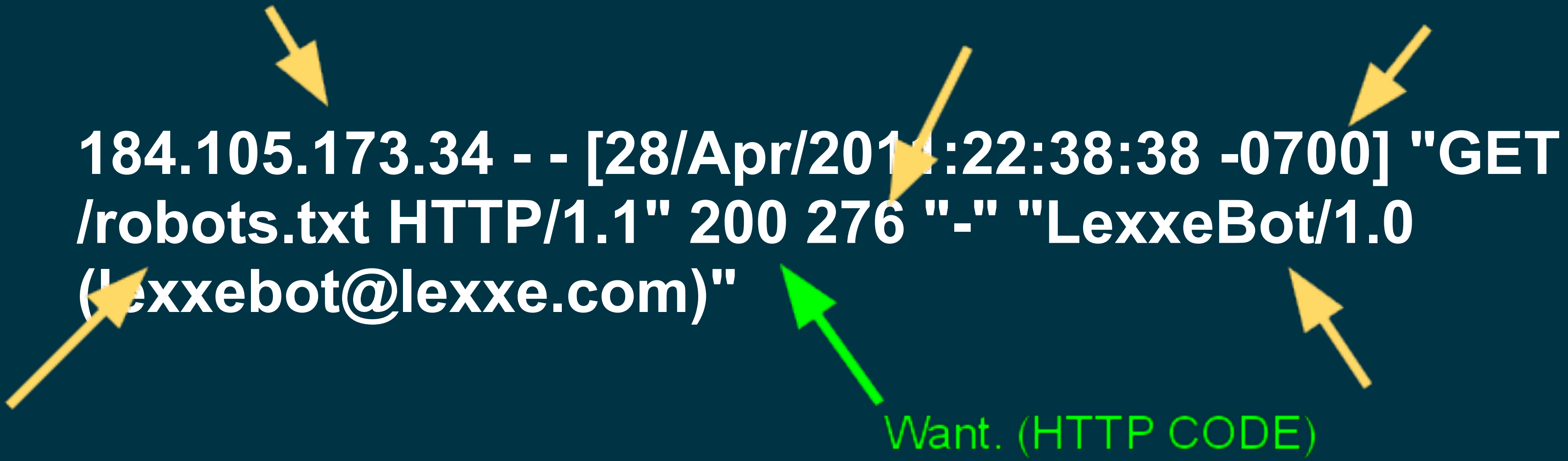
Good luck writing, debugging, and
maintaining that crap.

Humans are no good at logs.

Computers can't either if there's no format.

Thanks to crappy, random-format log data.

% grep 200 access.log



```
184.105.173.34 - - [28/Apr/2011:22:38:38 -0700] "GET
/robots.txt HTTP/1.1" 200 276 "-" "LexxeBot/1.0
(lexxebot@lexxe.com)"
```

Want. (HTTP CODE)

Any of these fields can include '200': ip, bytes, request path, user agent, time zone

Better.

```
{
  "client address": "184.105.173.34",
  "timestamp": "2011-04-28T22:38:38-0700",
  "verb": "GET",
  "path": "/robots.txt",
  "http version": 1.1,
  "response code": 200,
  "bytes": 276,
  "referrer": null
  "user agent": "LexxeBot/1.0"
}
```

Structured.

Easily parsed.

Easily queried.

Easily aggregated.

No nasty regexp.

```
{
  "client address": "184.105.173.34",
  "user": null,
  "timestamp": "2011-04-28T22:38:38-0700",
  "verb": "GET",
  "path": "/robots.txt",
  "query": null,
  "http version": 1.1,
  "response code": 200,
  "bytes": 276,
  "referrer": null
  "user agent": "LexxeBot/1.0"
}
```

It's fine if you don't use JSON.

thrift

avro

msgpack

xml

protobuf

csv

STOP INVENTING TIME FORMATS

STOP INVENTING TIME FORMATS

Nagios uses "1304060505"

Some syslogs use "Oct 11 20:21:47"

Apache uses "[29/Apr/2011:07:05:26 +0000]"

MySQL uses "020805 13:51:24"

Loggly's java apps use "110429.071055,118"

2011-05-01T10:15:33.144523-0500

2011-05-01T15:15:33.144Z

2011-05-01T15:15:33Z

2011-05-01

ISO8601

RFC3339

DEAR DEVELOPERS,

Please stop inventing terrible log formats.

but since I know you'll never change.

grok

Write regular expressions once. Reuse them later.

IPORHOST	(?:%{HOSTNAME}|%{IP})
HOSTNAME	\b(?:[0-9A-Za-z].......
IP	(?<![0-9])(?:(?:25[0-5]|2[0-4][0-9]....

grok

Ships with about 100 patterns.

Patterns you don't have to write yourself.

It is easy to add new patterns.

grok

```
Loaded suite test/alltests
Started
...................................
Finished in 4.39923 seconds.

35 tests, 72461 assertions, 0 failures, 0 errors
```

Mostly pattern-correctness tests

grok discovery

Have big library of patterns?

Take plain text, try to find patterns that match.

grok discovery

Original:

```
Apr 20 00:53:46 rickastley roll: Never gonna give you up.
```

Run through grok discover:

```
%{SYSLOGBASE}\Q Never gonna give you up.\E
```

```
%{SYSLOGTIMESTAMP} %{SYSLOGHOST} %{SYSLOGPROG}:
```

\Q\E%{SYSLOGBASE}\Q Never gonna give you up.\E

```
\Q\E(?<0000>(?<0001>(?<0002>\b(?:Jan(?:uary)?|Feb(?:ruary)?|Mar(?:ch)?|Apr(?:il)?|May|Jun(?:e)?|Jul(?:y)?|Aug(?:ust)?|Sep(?:tember)?|Oct(?:ober)?|Nov(?:ember)?|Dec(?:ember)?)\b) +(?<0003>(?:3[01]|[1-2]?[0-9]|0?[1-9])) (?<0004>(?!<[0-9])(?<0005>(?:2[0123]|[01][0-9])):(?<0006>(?:[0-5][0-9]))(?::(?<0007>(?:(?:[0-5][0-9]|60)(?:[.,][0-9]+)?)))(?![0-9]))) (?:(?<0008><(?<0009>\b(?:[0-9]+)\b).(?<000a>\b(?:[0-9]+)\b)>) )?(?<000b>(?<000c>(?:(?<000d>\b(?:[0-9A-Za-z][0-9A-Za-z-]{0,62})(?:\.(?:[0-9A-Za-z][0-9A-Za-z-]{0,62}))*(\.?|\b))|(?<000e>(?<![0-9])(?:(?:25[0-5]|2[0-4][0-9]|[0-1]?[0-9]{1,2})[.](?:25[0-5]|2[0-4][0-9]|[0-1]?[0-9]{1,2})[.](?:25[0-5]|2[0-4][0-9]|[0-1]?[0-9]{1,2})[.](?:25[0-5]|2[0-4][0-9]|[0-1]?[0-9]{1,2}))(?![0-9])))))) (?<000f>(?<0010>(?:[\w._/-]+))(?:\[(?<0011>\b(?:[0-9]+)\b)\])?):)\Q Never gonna give you up.\E
```

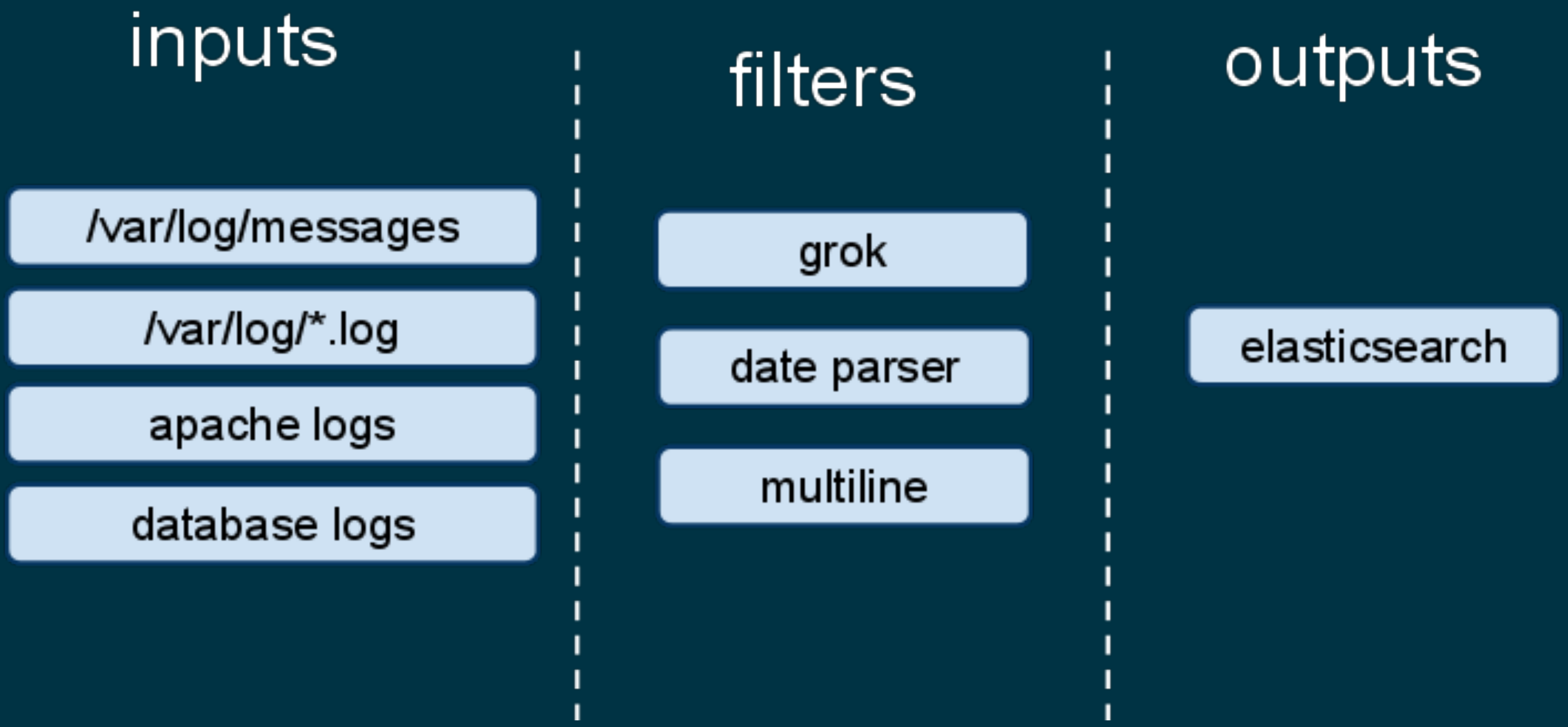
Patterns from known-good parts.

Tested. **Reusable.**

You didn't have to write it by hand.

```
 \Q\E(?<0000>(?<0001>(?<0002>\b(?:Jan(?:uary)?|Feb(?:
ruary)?|Mar(?:ch)?|Apr(?:il)?|May|Jun(?:e)?|Jul(?:y)?|Aug(?:
ust)?|Sep(?:tember)?|Oct(?:ober)?|Nov(?:ember)?|Dec(?:
ember)?)\b) +(?<0003>(?:3[01]|[1-2]?[0-9]|0?[1-9])) (?<0004>
(?!<[0-9])(?<0005>(?:2[0123]|[01][0-9])):(?<0006>(?:[0-5][0-9]))
(?:(?<0007>(?:(?:[0-5][0-9]|60)(?:[.,][0-9]+)?))(?![0-9]))) (?:(?
<0008><(?<0009>\b(?:[0-9]+)\b).(?<000a>\b(?:[0-9]+)\b)>) )?(?
<000b>(?<000c>(?:(?<000d>\b(?:[0-9A-Za-z][0-9A-Za-z-]
{0,62})(?:\.(?:[0-9A-Za-z][0-9A-Za-z-]{0,62}))*(\.?|\b))|(?<000e>
(?<![0-9])(?:(?:25[0-5]|2[0-4][0-9]|[0-1]?[0-9]{1,2})[.](?:25[0-5]|2
[0-4][0-9]|[0-1]?[0-9]{1,2})[.](?:25[0-5]|2[0-4][0-9]|[0-1]?[0-9]
{1,2})[.](?:25[0-5]|2[0-4][0-9]|[0-1]?[0-9]{1,2}))(?![0-9]))))) (?
<000f>(?<0010>(?:[\w._/-]+))(?:\[(?<0011>\b(?:[0-9]+)\b)\])?):)
\Q Never gonna give you up\E
```

Simple single-host agent example



inputs

/var/log/messages

/var/log/*.log

apache logs

database logs

filters

grok

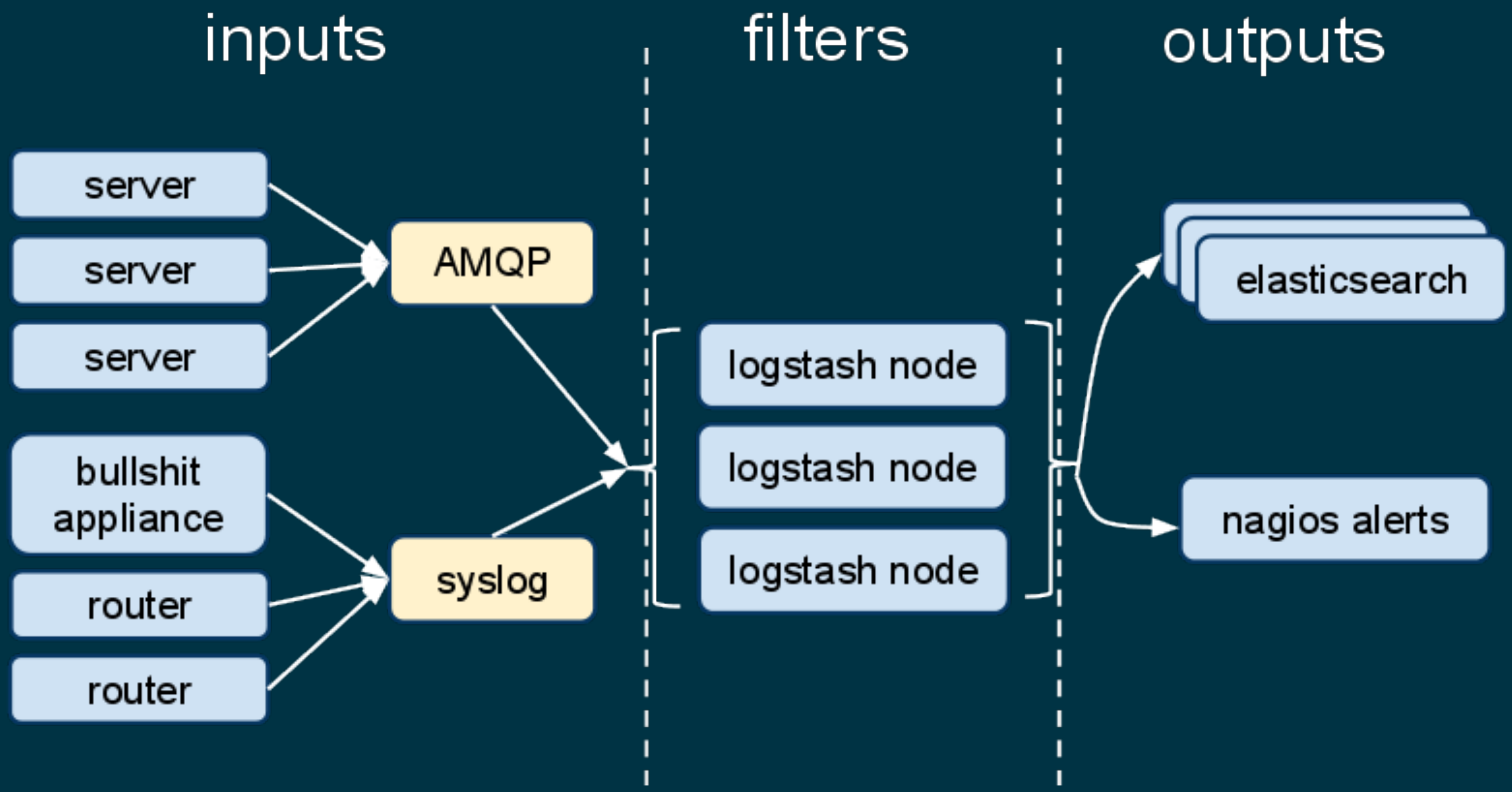
date parser

multiline

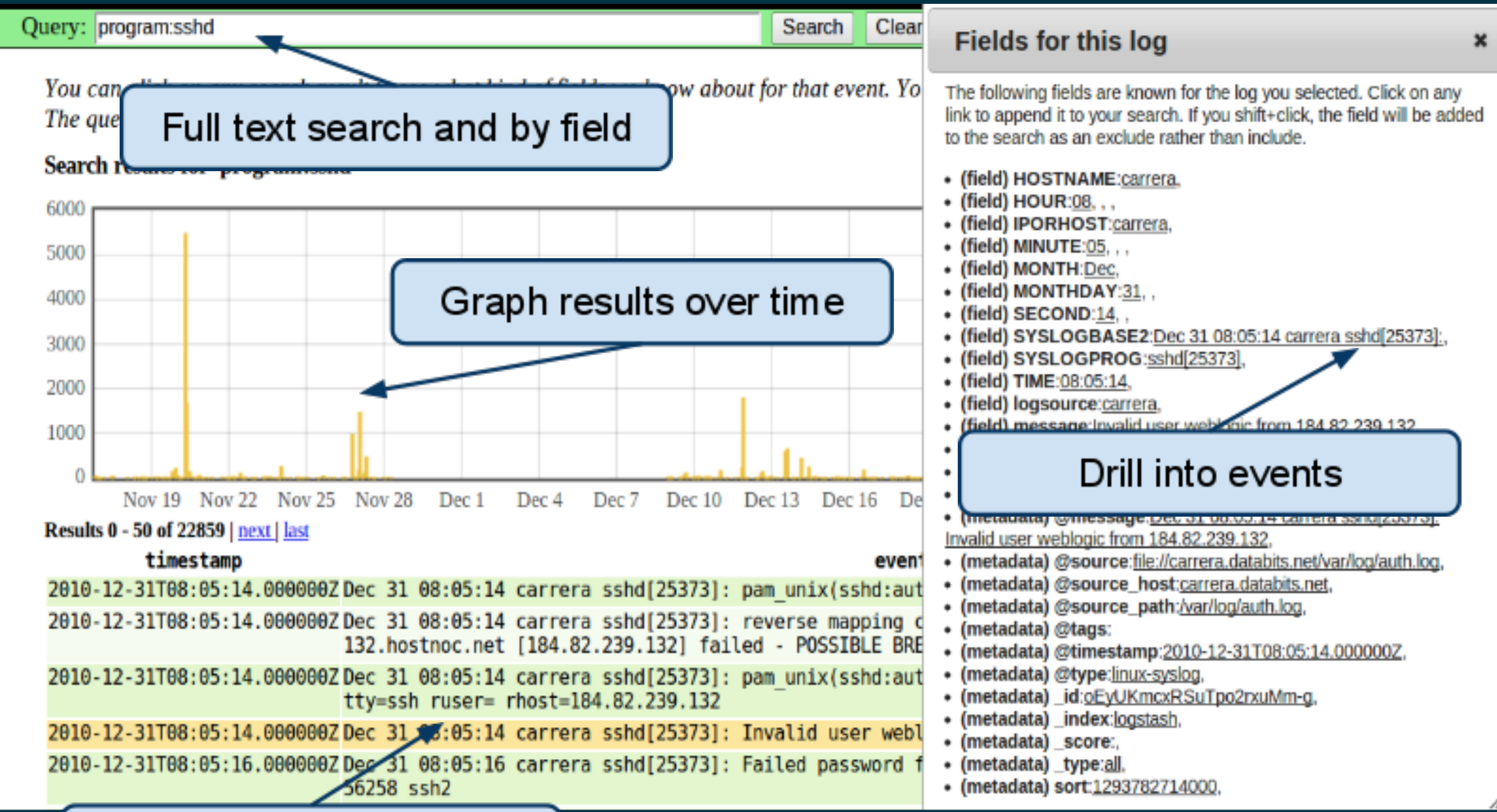
outputs

elasticsearch

Complex full-infrastructure setup.



Search.



Query: program:sshd

You can … ow about for that event. Yo
The que

Search r

Results 0 - 50 of 22859 | next | last

```
timestamp                 event
2010-12-31T08:05:14.000000Z Dec 31 08:05:14 carrera sshd[25373]: pam_unix(sshd:aut
2010-12-31T08:05:14.000000Z Dec 31 08:05:14 carrera sshd[25373]: reverse mapping c
                            132.hostnoc.net [184.82.239.132] failed - POSSIBLE BRE
2010-12-31T08:05:14.000000Z Dec 31 08:05:14 carrera sshd[25373]: pam_unix(sshd:aut
                            tty=ssh ruser= rhost=184.82.239.132
2010-12-31T08:05:14.000000Z Dec 31 08:05:14 carrera sshd[25373]: Invalid user webl
2010-12-31T08:05:16.000000Z Dec 31 08:05:16 carrera sshd[25373]: Failed password f
                            56258 ssh2
```

Fields for this log ×

The following fields are known for the log you selected. Click on any link to append it to your search. If you shift+click, the field will be added to the search as an exclude rather than include.

- **(field) HOSTNAME**:carrera,
- **(field) HOUR**:08, , ,
- **(field) IPORHOST**:carrera,
- **(field) MINUTE**:05, , ,
- **(field) MONTH**:Dec,
- **(field) MONTHDAY**:31, ,
- **(field) SECOND**:14, ,
- **(field) SYSLOGBASE2**:Dec 31 08:05:14 carrera sshd[25373]:,
- **(field) SYSLOGPROG**:sshd[25373],
- **(field) TIME**:08:05:14,
- **(field) logsource**:carrera,
- **(field) message**:Invalid user weblogic from 184.82.239.132
- **(metadata) @message**:Dec 31 08:05:14 carrera sshd[25373]: Invalid user weblogic from 184.82.239.132,
- **(metadata) @source**:file://carrera.databits.net/var/log/auth.log,
- **(metadata) @source_host**:carrera.databits.net,
- **(metadata) @source_path**:/var/log/auth.log,
- **(metadata) @tags**:
- **(metadata) @timestamp**:2010-12-31T08:05:14.000000Z,
- **(metadata) @type**:linux-syslog,
- **(metadata) _id**:oEyUKmcxRSuTpo2rxuMm-g,
- **(metadata) _index**:logstash,
- **(metadata) _score**:,
- **(metadata) _type**:all,
- **(metadata) sort**:1293782714000,

Full text search and by field

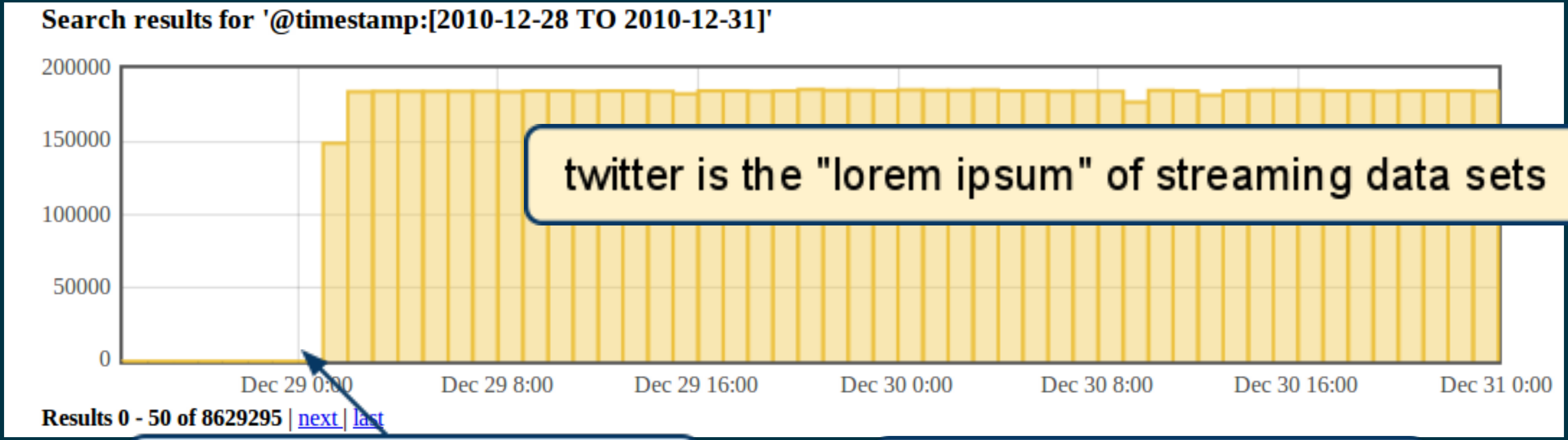
Graph results over time

Drill into events

Search results

Searching for failed ssh attempts

Ship data around. Use it awesomely.



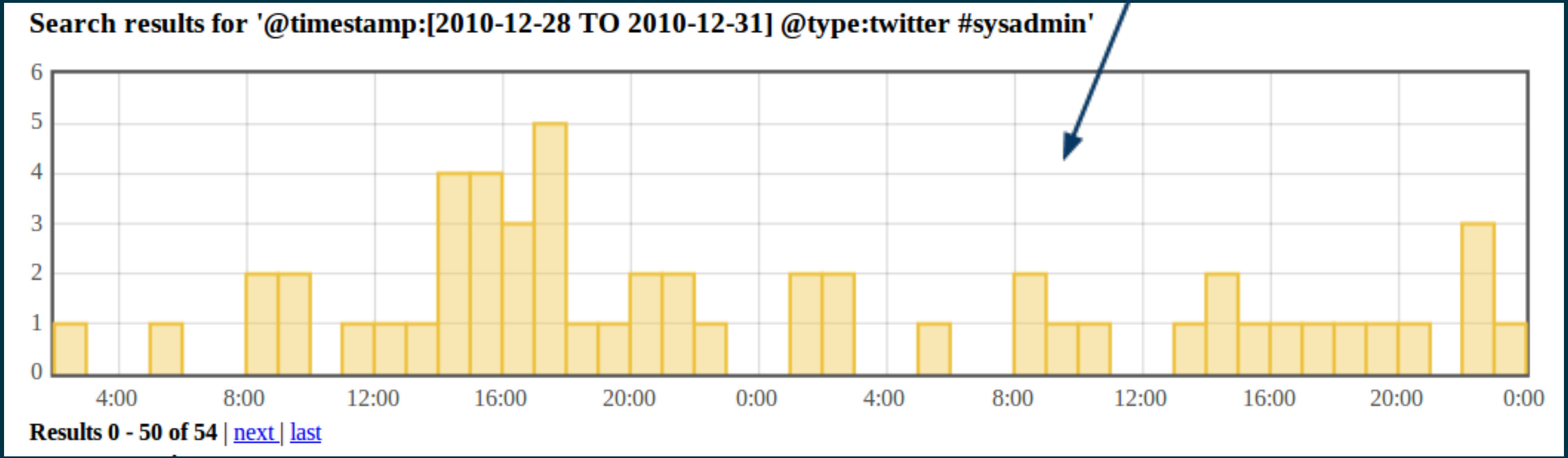
Search results for '@timestamp:[2010-12-28 TO 2010-12-31]'

Results 0 - 50 of 8629295 | next | last

twitter is the "lorem ipsum" of streaming data sets

Started indexing twitter

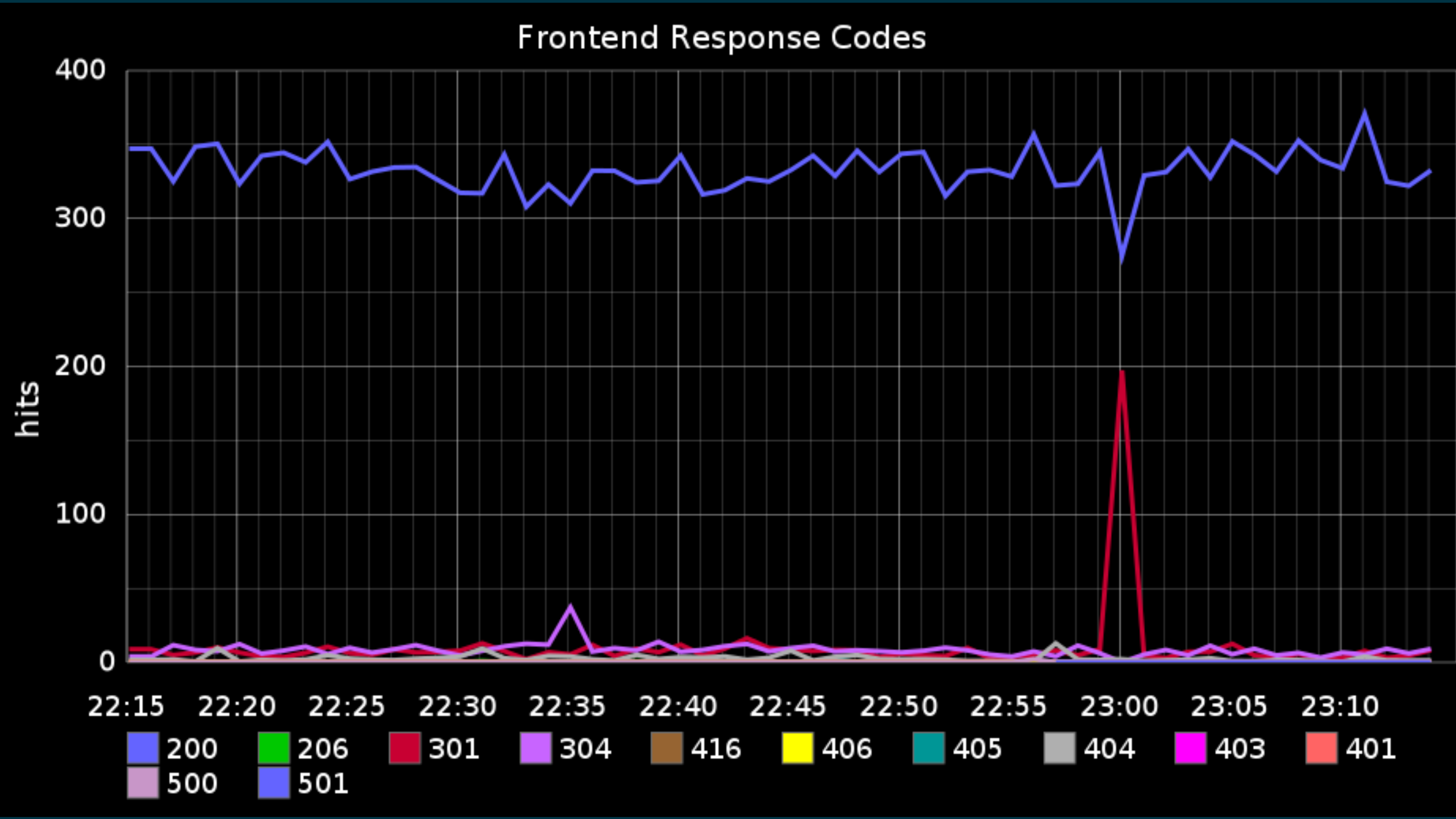
Tweets with '#sysadmin'



Search results for '@timestamp:[2010-12-28 TO 2010-12-31] @type:twitter #sysadmin'

Results 0 - 50 of 54 | next | last

Integration.



apache logs ⟶ logstash ⟶ graphite

past and future.

- grok started in 2005
- Logstash first released in November 2010.

And then?

Project site: http://logstash.net

Code: http://github.com/logstash/logstash

Issues: http://logstash.jira.com

IRC: #logstash on freenode.

Mailing list: logstash-users@googlegroups.com

References

- logstash mascot by Andre Jolicoeur http://andrejolicoeur.com/
- graylog2 screenshots from http://www.graylog2.org/about/screenshots
- graylog2 mascot (party gorilla) from http://theoatmeal.com/

ssh, grep, and tail

FAIL

humans + epic oneliners

FAIL

too many tools required

FAIL

FAIL