

Wrangling Logs with Logstash and Elasticsearch

Nate Jones & David Castro

Media Temple

OSCON 2012

Why are we here?

Size

Quantity



Efficiency

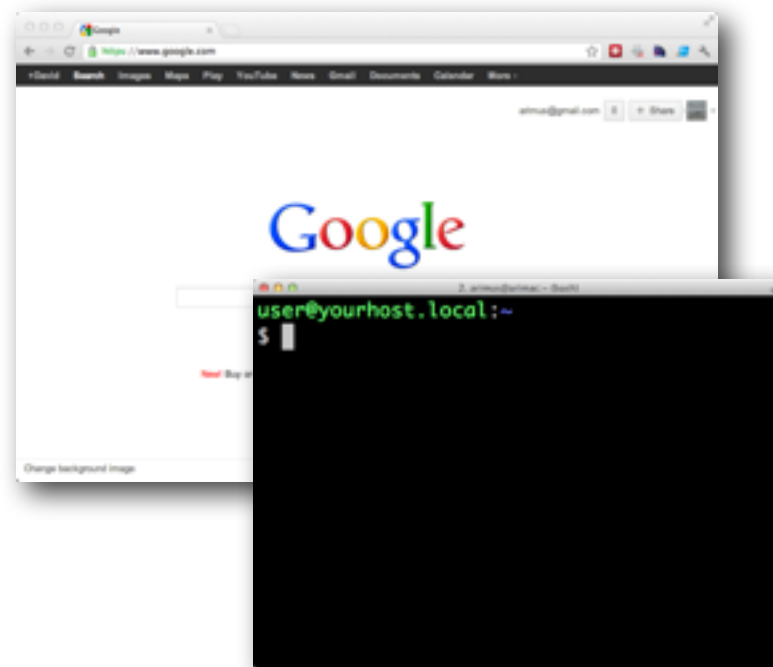
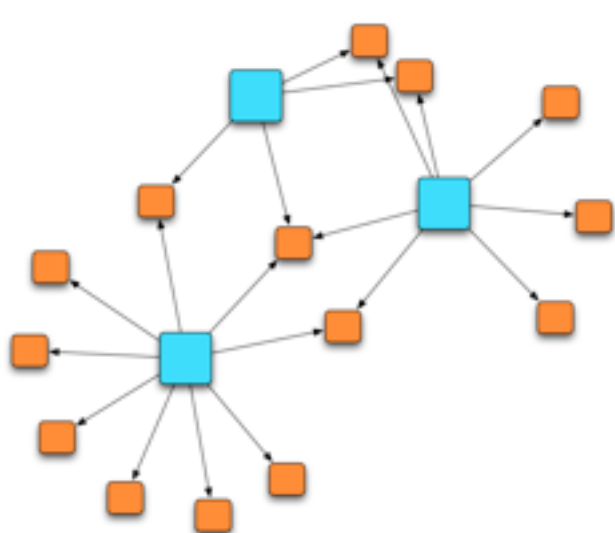


Access

Locality

Method

Filtering



Grokability

Noise

Structure

Metrics



Use Case: Mail Logs

Size

30 mail servers

2G logs / day / server

60GB / day total

1.8 TB / month

21 TB / year

1 billion log lines per week



Access

Front-line, easy access

No SSH

Shareable



Grokability

Operational

Did the email get delivered?

Why was the message marked as SPAM?

Are messages being rejected?



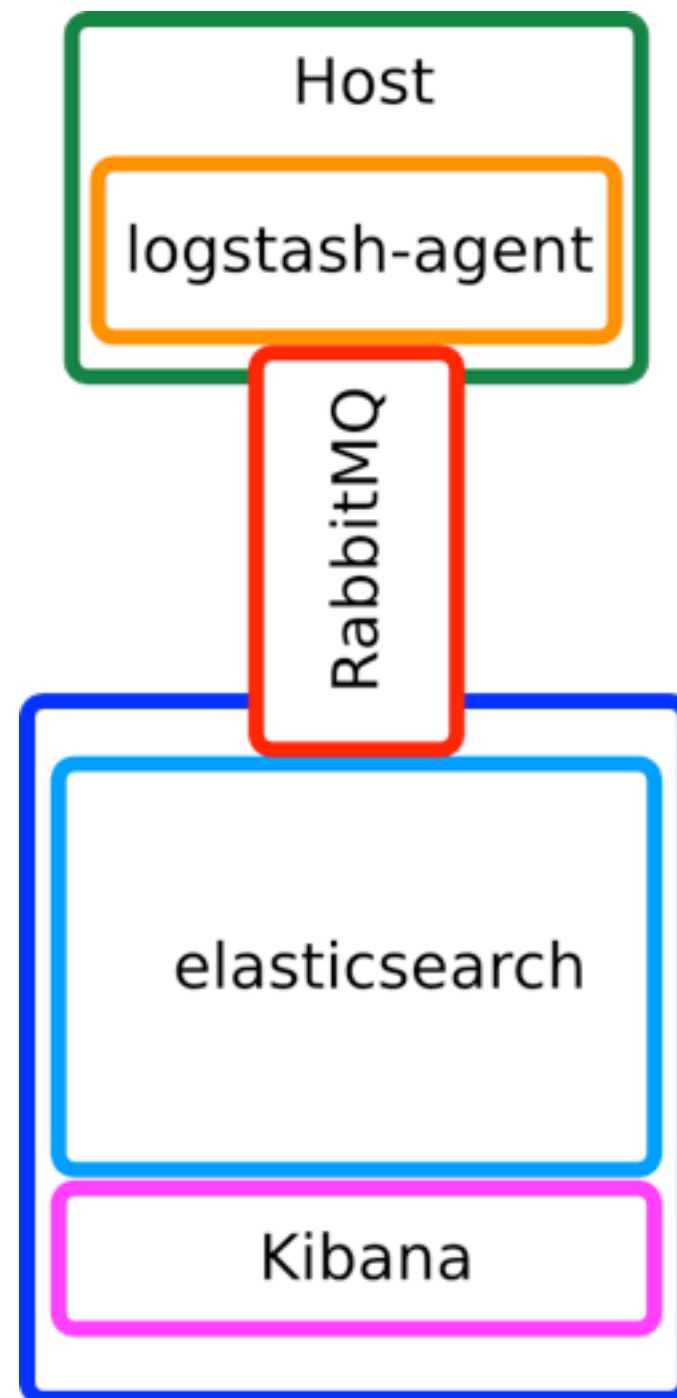
Metrics

What's the inbound/outbound message rate?

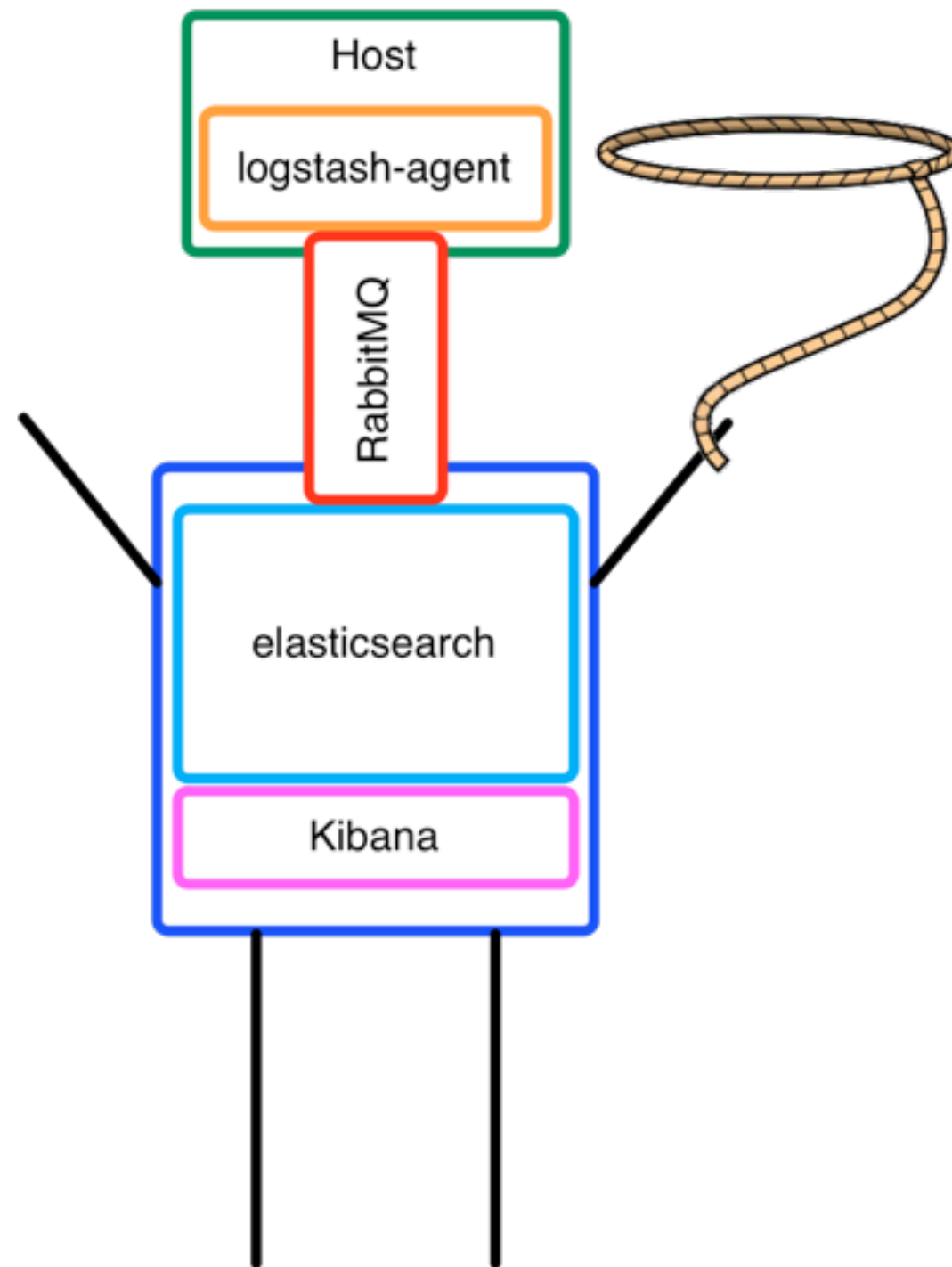
How often are we seeing particular errors?

The Solution

Overview



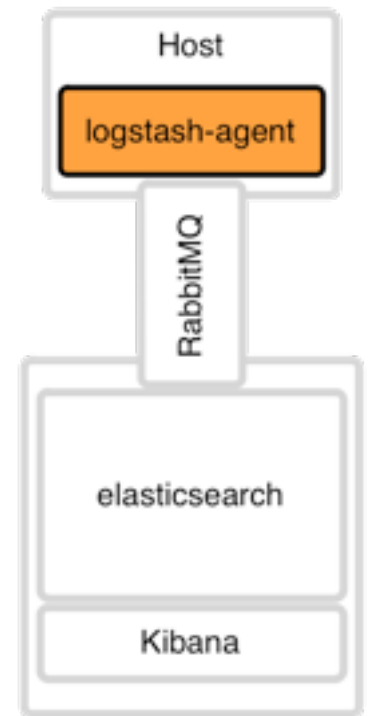
Overview



Logstash Overview

<http://logsta.sh/>

1. Parse log line
2. Transform/extract
3. Structure and send JSON



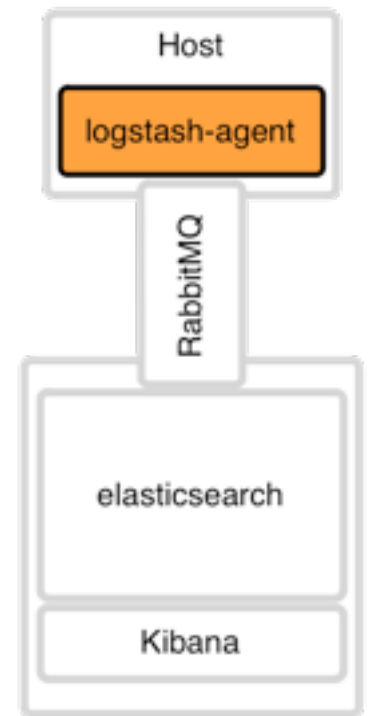
Logstash Parsing

Log line input

```
2012-07-10T20:00:02.446220-04:00 mail01 spamd[2478]: spamd: clean
message (-3.4/5.0) for nobody:93 in 0.0 seconds, 886 bytes.
```

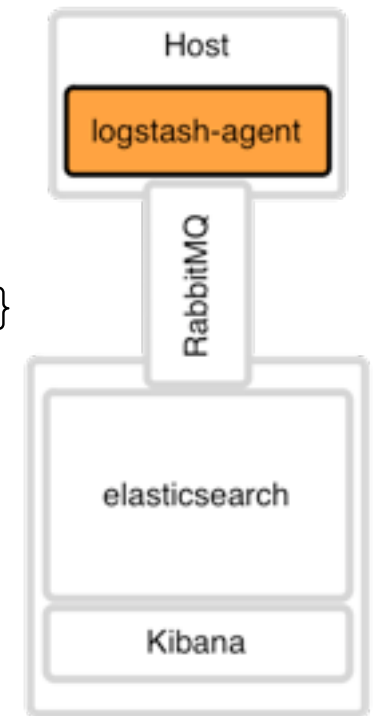
JSON output

```
{
  "@timestamp" : "2012-07-16T06:44:00.548000Z",
  "@tags" : [],
  "@fields" : {},
  "@source_path" : "/client/127.0.0.1:40010",
  "@source" : "tcp://0.0.0.0:6999/client/127.0.0.1:40010",
  "@source_host" : "0.0.0.0",
  "@message" : "2012-07-10T20:00:02.446220-04:00 mail01 spamd[2478]:
spamd: clean message (-3.4/5.0) for nobody:93 in 0.0 seconds, 886
bytes.",
  "@type" : "maillog"
}
```



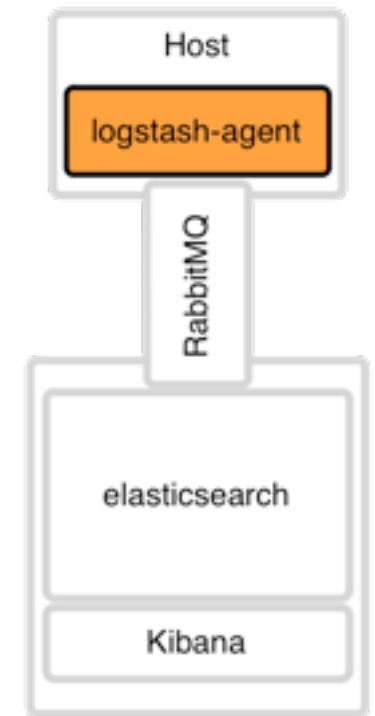
Logstash Parsing

```
grok {
  type => "maillog"
  pattern => "%{TIMESTAMP_ISO8601:timestamp} %{WORD:host}
%{SYSLOGPROG:service}: %{GREEDYDATA:message}"
}
mutate {
  type => "maillog"
  # replace the timestamp, correcting import timestamp
  replace => ["@timestamp", "%{timestamp}"]
  # replace the message sans-timestamp/host/service
  replace => ["@message", "%{message}"]
}
```



Logstash Parsing

```
{
  "@timestamp" : "2012-07-10T20:00:02.446220-04:00",
  "@tags" : [],
  "@fields" : {
    "pid" : [
      "2478"
    ],
    "service" : [
      "spamd[2478]"
    ],
    "program" : [
      "spamd"
    ],
    "host" : [
      "mail01"
    ]
  },
  "@source_path" : "/client/127.0.0.1:39998",
  "@source" : "tcp://0.0.0.0:6999/client/127.0.0.1:39998",
  "@source_host" : "0.0.0.0",
  "@message" : "spamd: clean message (-3.4/5.0) for nobody:93 in 0.0
seconds, 886 bytes.",
  "@type" : "maillog"
}
```



RabbitMQ Overview

<http://www.rabbitmq.com/>

Message Queue

AMQP

Clustered



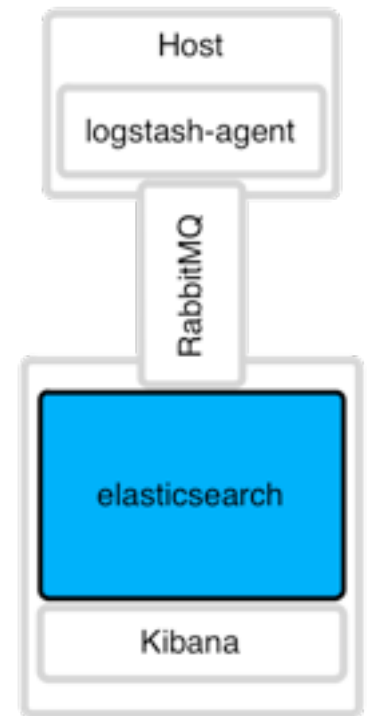
Elasticsearch Intro

<http://www.elasticsearch.org/>

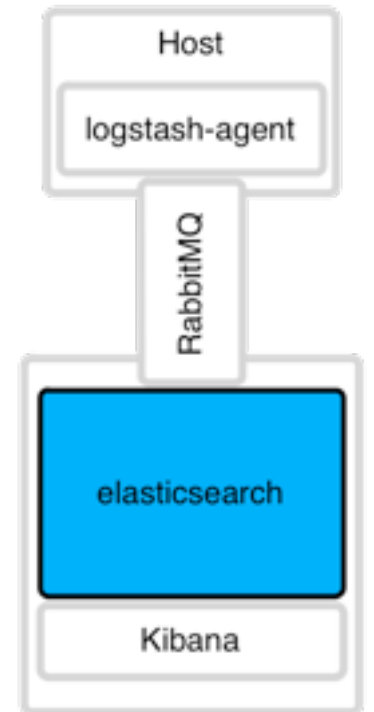
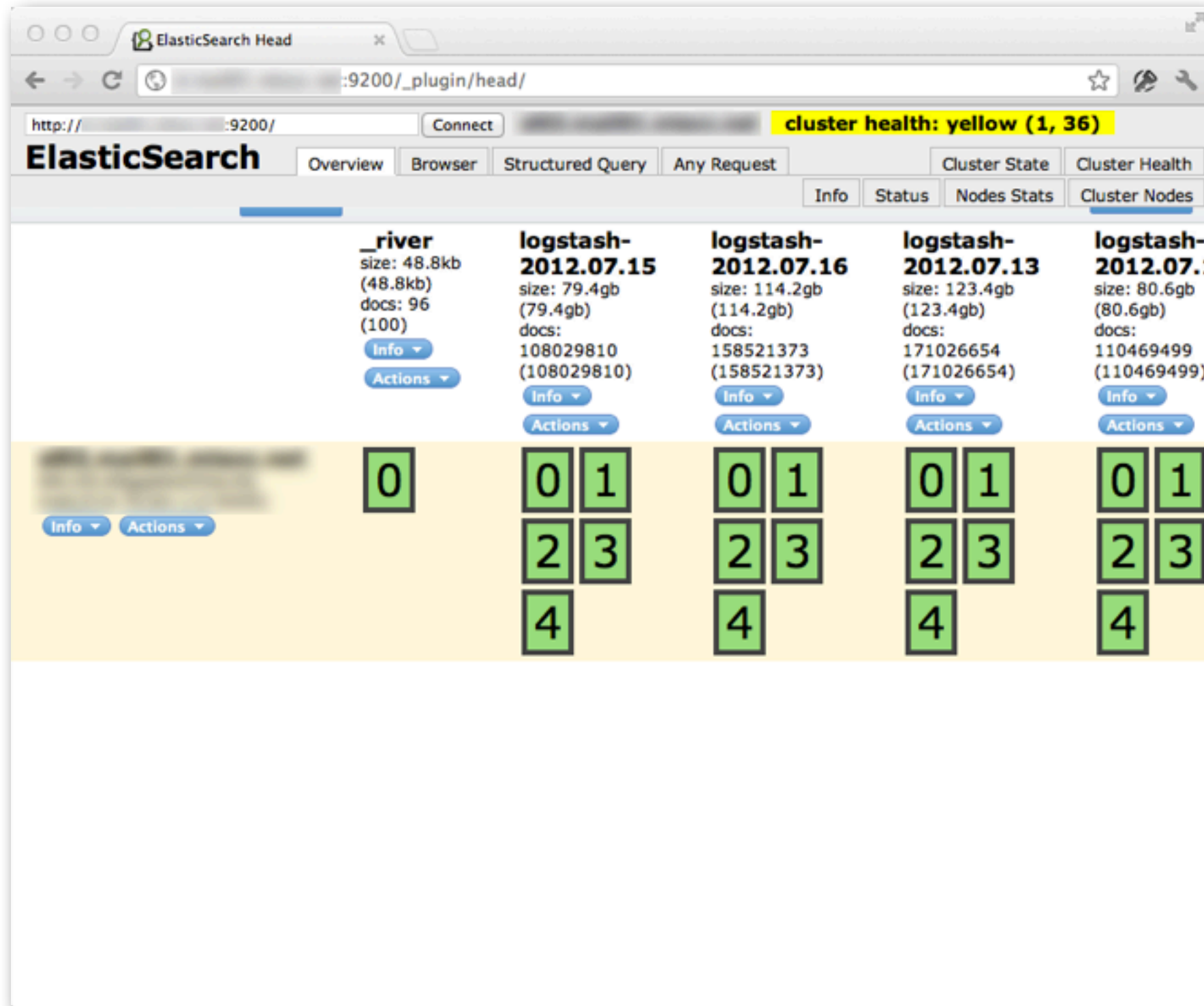
Index in Lucene shards

Cluster-able

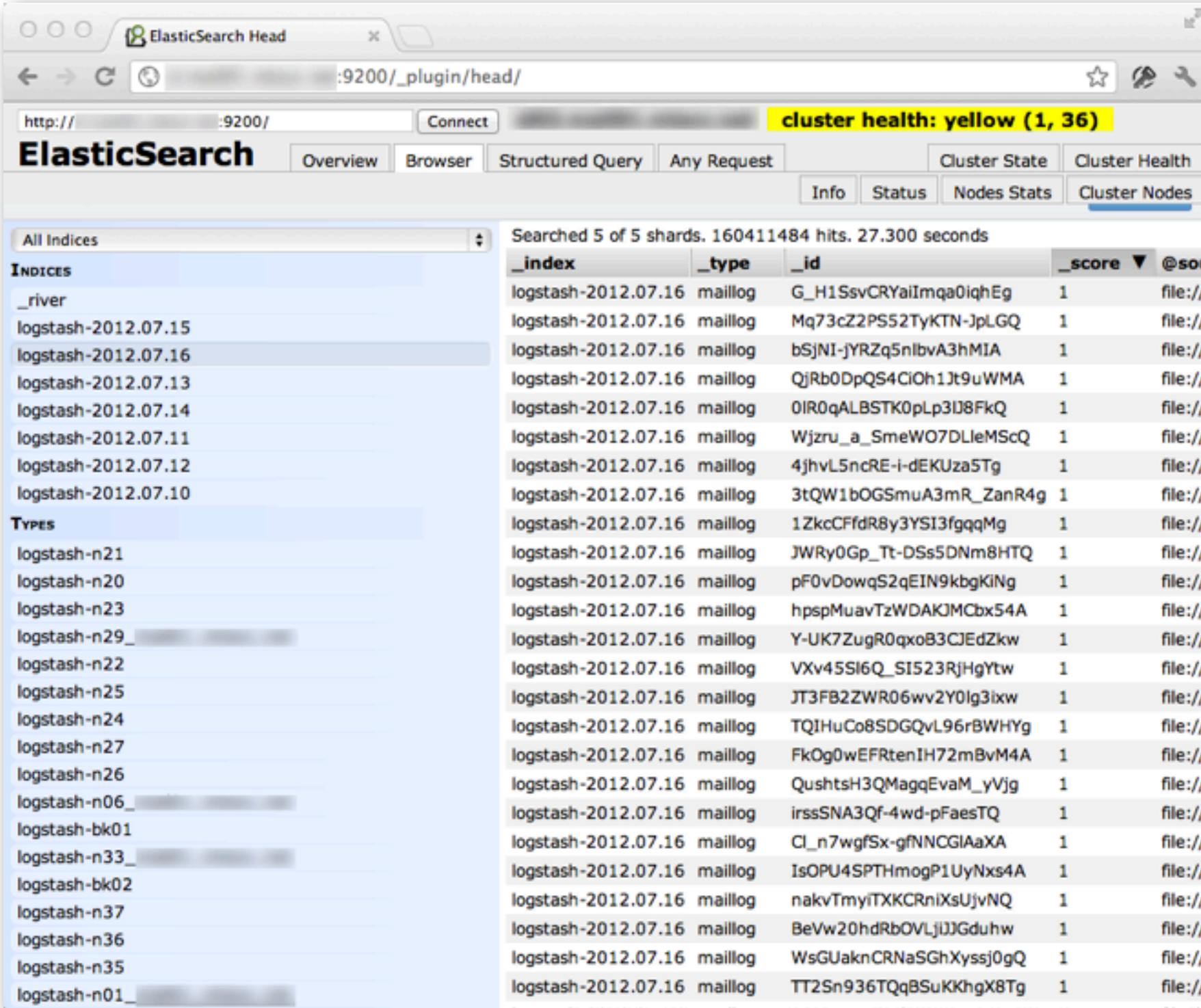
Fault tolerant



Elasticsearch Head

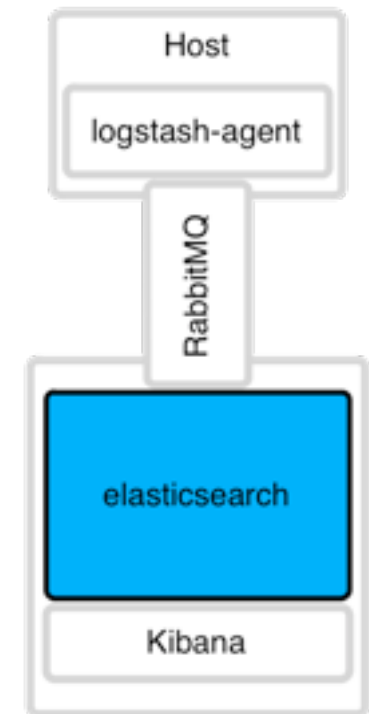


Elasticsearch Browser



The screenshot shows the Elasticsearch Head browser interface. The top navigation bar includes tabs for Overview, Browser, Structured Query, Any Request, Cluster State, and Cluster Health. The Cluster Health tab is active, displaying a yellow status for 1 node out of 36. The left sidebar shows a list of indices, with 'logstash-2012.07.16' selected. The main content area displays a search results table with columns: _index, _type, _id, _score, and @source. The table contains 36 rows of logstash data from July 2012.

_index	_type	_id	_score	@source
logstash-2012.07.16	maillog	G_H1SsvCRYaiImqa0iqhEg	1	file://...
logstash-2012.07.16	maillog	Mq73cZ2PS52TyKTN-JpLGQ	1	file://...
logstash-2012.07.16	maillog	bSjNI-jYRZq5nlbvA3hMIA	1	file://...
logstash-2012.07.16	maillog	QjRb0DpQS4CioH1Jt9uWMA	1	file://...
logstash-2012.07.16	maillog	0IR0qALBSTK0pLp3IJ8FkQ	1	file://...
logstash-2012.07.16	maillog	Wjzru_a_SmeWO7DLleMScQ	1	file://...
logstash-2012.07.16	maillog	4jHvL5ncRE-i-dEKUza5Tg	1	file://...
logstash-2012.07.16	maillog	3tQW1bOGSmuA3mR_ZanR4g	1	file://...
logstash-2012.07.16	maillog	1ZkcCFdR8y3YSI3fgqqMg	1	file://...
logstash-2012.07.16	maillog	JWRy0Gp_Tt-DSs5DNm8HTQ	1	file://...
logstash-2012.07.16	maillog	pF0vDowqS2qEIN9kbGKINg	1	file://...
logstash-2012.07.16	maillog	hpspMuavTzWDAKJMCbx54A	1	file://...
logstash-2012.07.16	maillog	Y-UK7ZugR0qxoB3CJEdZkw	1	file://...
logstash-2012.07.16	maillog	VXv45SI6Q_SI523RjHgYtw	1	file://...
logstash-2012.07.16	maillog	JT3FB2ZWR06wv2Y0lg3ixw	1	file://...
logstash-2012.07.16	maillog	TQIHuCo8SDGQvL96rBWHYg	1	file://...
logstash-2012.07.16	maillog	FkOg0wEFRtenIH72mBvM4A	1	file://...
logstash-2012.07.16	maillog	QushtsH3QMagqEvaM_yVjg	1	file://...
logstash-2012.07.16	maillog	irssSNA3Qf-4wd-pFaesTQ	1	file://...
logstash-2012.07.16	maillog	Cl_n7wgfSx-gfNNCGIAaXA	1	file://...
logstash-2012.07.16	maillog	IsOPU4SPTHmogP1UyNxs4A	1	file://...
logstash-2012.07.16	maillog	nakvTmyiTXXCRniXsUjvNQ	1	file://...
logstash-2012.07.16	maillog	BeVw20hdRbOVLjJJGduhw	1	file://...
logstash-2012.07.16	maillog	WsGUaknCRNaSGhXyssj0gQ	1	file://...
logstash-2012.07.16	maillog	TT2Sn936TQqBSuKKhgX8Tg	1	file://...



Kibana Intro

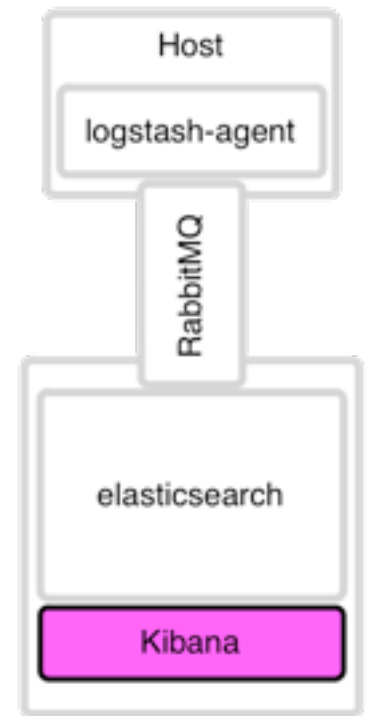
<http://rashidkpc.github.com/Kibana/>

User friendly front-end to elasticsearch

Search log lines

Graph, score, trend

Streaming dashboard



Kibana Queries

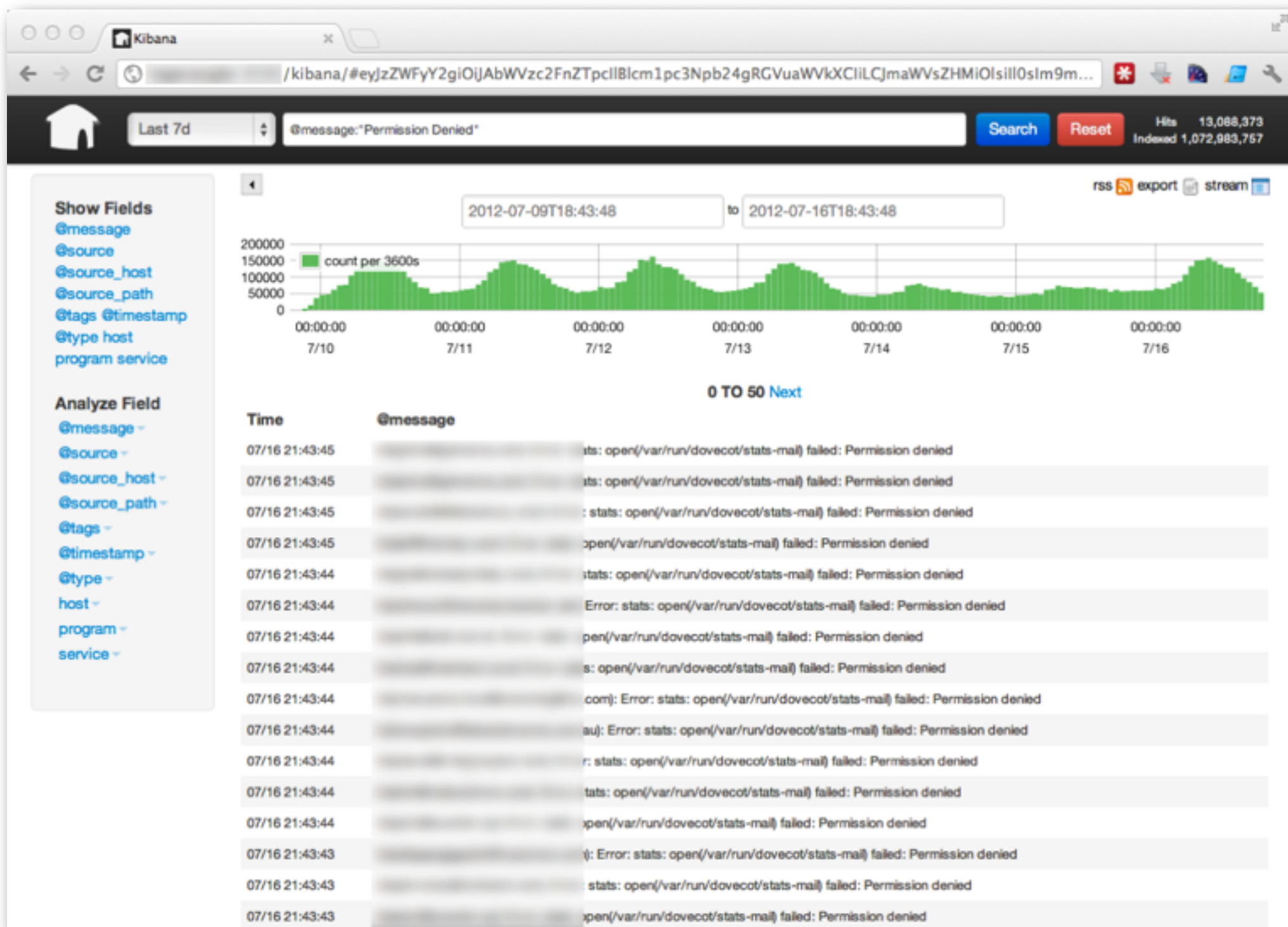
Question

How many errors of a particular type are we seeing in the logs?

Query

@message:"Permission Denied"

Kibana Queries



Kibana Queries

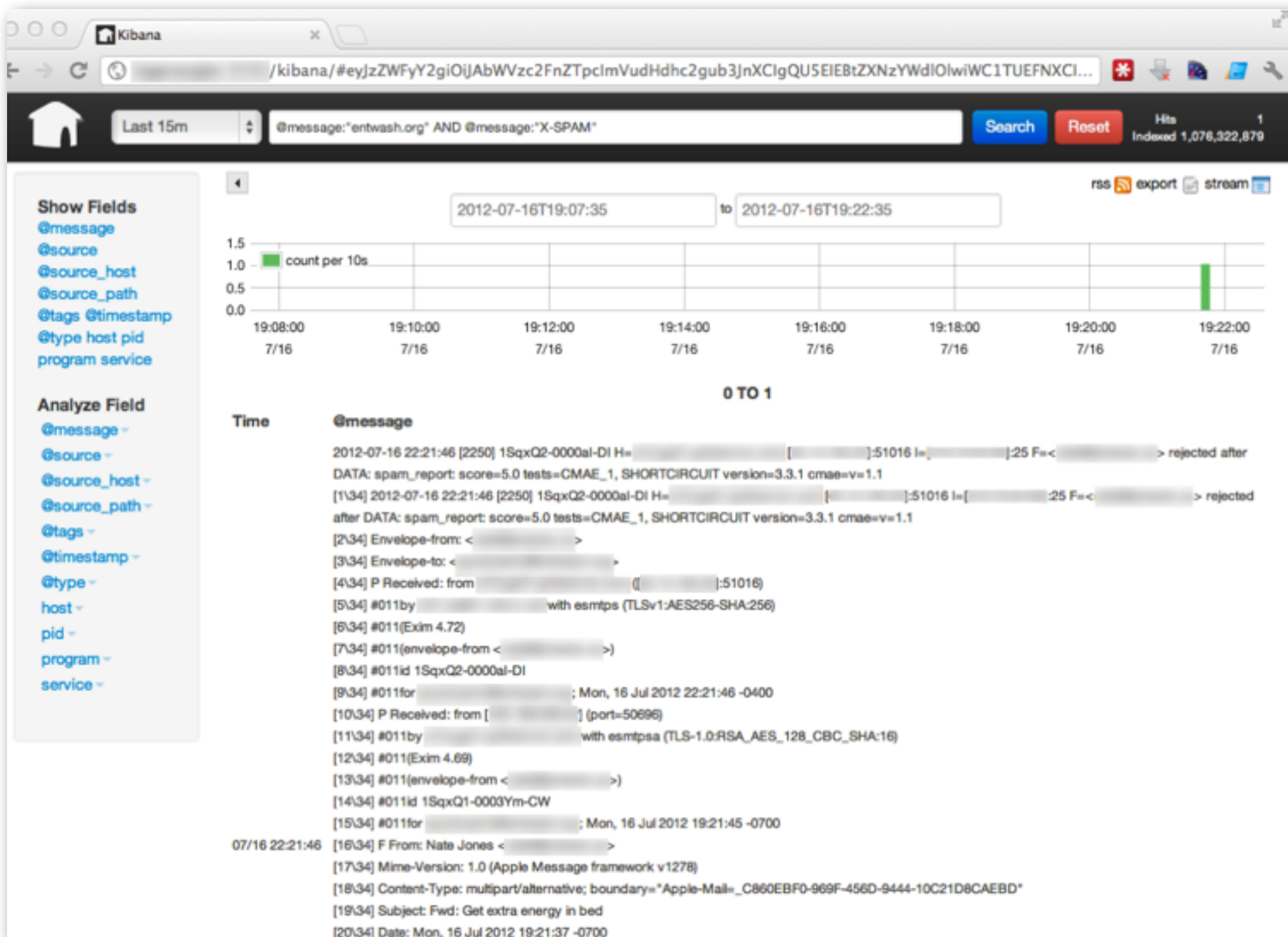
Question

Why did the mail for user X get marked as SPAM?

Query

@message:"domain.com" AND @message:"X-SPAM"

Kibana Queries



Kibana Queries

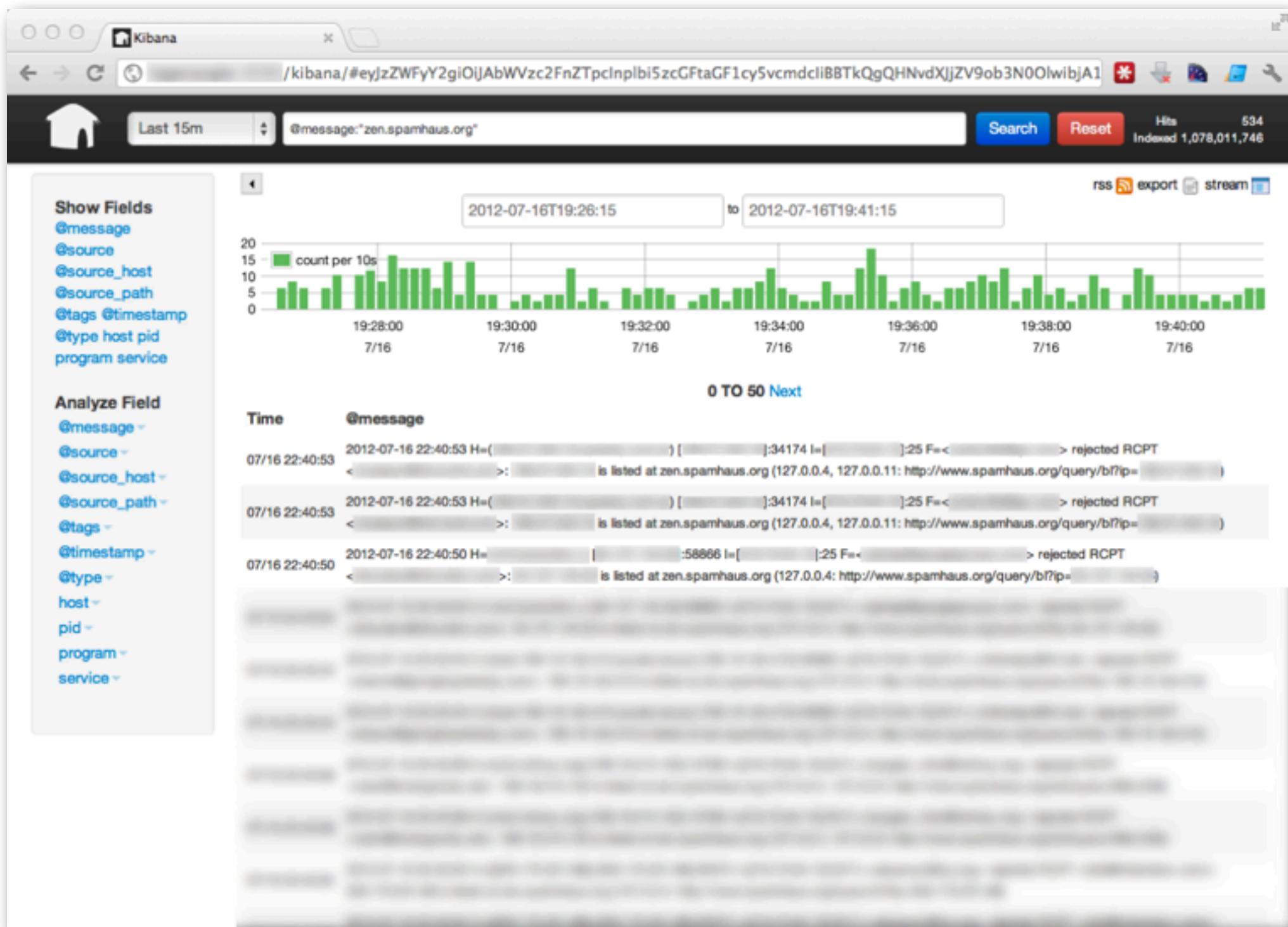
Question

How many messages are being rejected due to the sending host being listed in an RBL?

Query

@message:"zen.spamhaus.org"

Kibana Queries



Kibana Queries

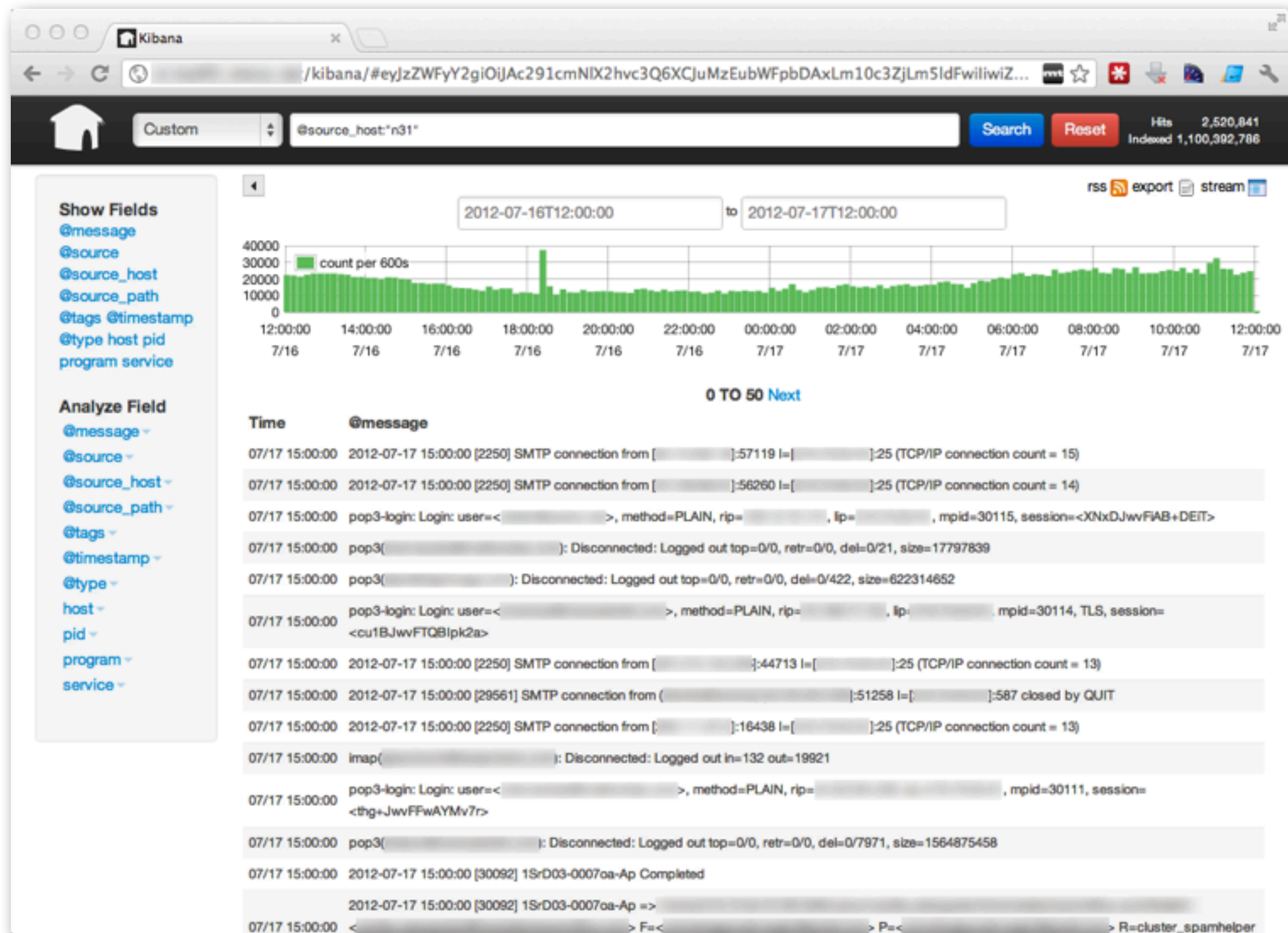
Question

How many log messages do we have for a specific mail host?

Query

```
@source_host:"n31"
```

Kibana Queries



Report Card

Size

Quantity



Efficiency



Access

Locality

Method

Filtering



Grokability

Noise

Structure

Metrics



Next Steps

Push more stats into graphite

Further breaking down log messages

More stuff

Everything you need

Instructions and software

<http://logwrangler.mtcode.com/>

Puppet code and slides

<http://github.com/mediatemple/logwrangler>

Local wifi share: logwrangler (guest/guest)

Demo

Netcat port for Logstash

RabbitMQ

Elasticsearch

Kibana

Contact Info

Nate Jones

@ndj

nate@mediateemple.net

David Castro

@arimus

dcastro@mediateemple.net

Questions?