Roll Your Own SIEM

ELK, Python, and Pattern Recognition For Fun and Profit

TASK June 2015

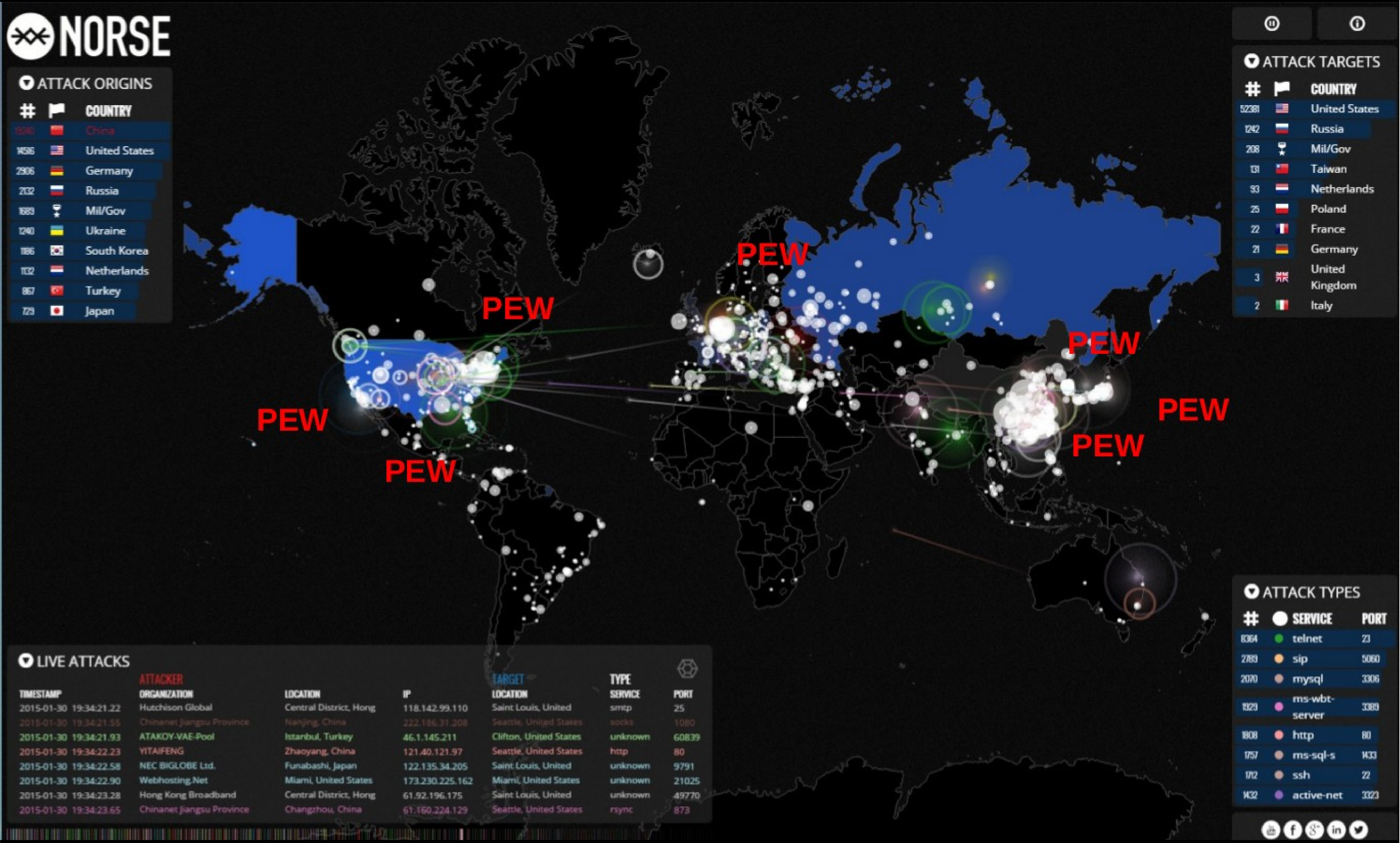
whoami

- my first PC was a VIC 20 (nobody could ever need more than 5k amiright?)

- current status: racking servers and scripting things at an awesome local ISP

- this talk is based around a project I did for my Ryerson compsec course, this field is new to me, so jump in – having a discussion around ML would be great since I'm primarily here to learn from you

So, This Is Happening



ATTACK ORIGINS

#		COUNTRY
[illegible]	🇨🇳	China
14586	🇺🇸	United States
2906	🇩🇪	Germany
2132	🇷🇺	Russia
1683		Mil/Gov
1240	🇺🇦	Ukraine
1186	🇰🇷	South Korea
1132	🇳🇱	Netherlands
867	🇹🇷	Turkey
729	🇯🇵	Japan

ATTACK TARGETS

#		COUNTRY
52381	🇺🇸	United States
1242	🇷🇺	Russia
208		Mil/Gov
131	🇹🇼	Taiwan
93	🇳🇱	Netherlands
25	🇵🇱	Poland
22	🇫🇷	France
21	🇩🇪	Germany
3	🇬🇧	United Kingdom
2	🇮🇹	Italy

LIVE ATTACKS

	ATTACKER			TARGET	TYPE	
TIMESTAMP	ORGANIZATION	LOCATION	IP	LOCATION	SERVICE	PORT
2015-01-30 19:34:21.22	Hutchison Global	Central District, Hong	118.142.99.110	Saint Louis, United	smtp	25
2015-01-30 19:34:21.55	Chinanet Jiangsu Province	Nanjing, China	222.186.31.208	Seattle, United States	socks	1080
2015-01-30 19:34:21.93	ATAKOY-VAE-Pool	Istanbul, Turkey	46.1.145.211	Clifton, United States	unknown	60839
2015-01-30 19:34:22.23	YITAIFENG	Zhaoyang, China	121.40.121.97	Seattle, United States	http	80
2015-01-30 19:34:22.58	NEC BIGLOBE Ltd.	Funabashi, Japan	122.135.34.205	Saint Louis, United	unknown	9791
2015-01-30 19:34:22.90	Webhosting.Net	Miami, United States	173.230.225.162	Miami, United States	unknown	21025
2015-01-30 19:34:23.28	Hong Kong Broadband	Central District, Hong	61.92.196.175	Saint Louis, United	unknown	49770
2015-01-30 19:34:23.65	Chinanet Jiangsu Province	Changzhou, China	61.160.224.129	Seattle, United States	rsync	873

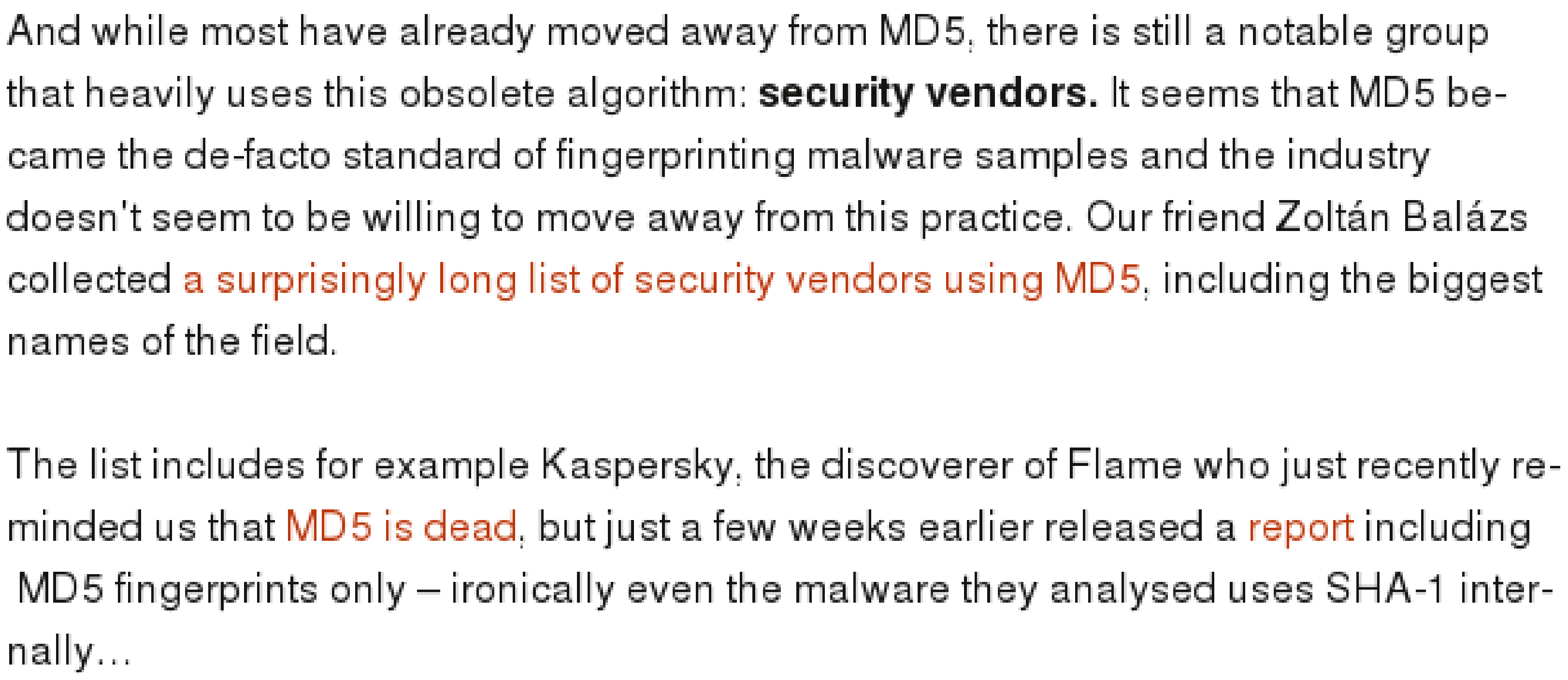
ATTACK TYPES

#	●	SERVICE	PORT
8364	🟢	telnet	23
2789	●	sip	5060
2070	●	mysql	3306
1929	●	ms-wbt-server	3389
1808	●	http	80
1757	●	ms-sql-s	1433
1712	●	ssh	22
1432	🟣	active-net	3323

Blue Team – Traditional IDS

Signature based IDS has served for some time, but suffers from many problems – including collision attacks

“

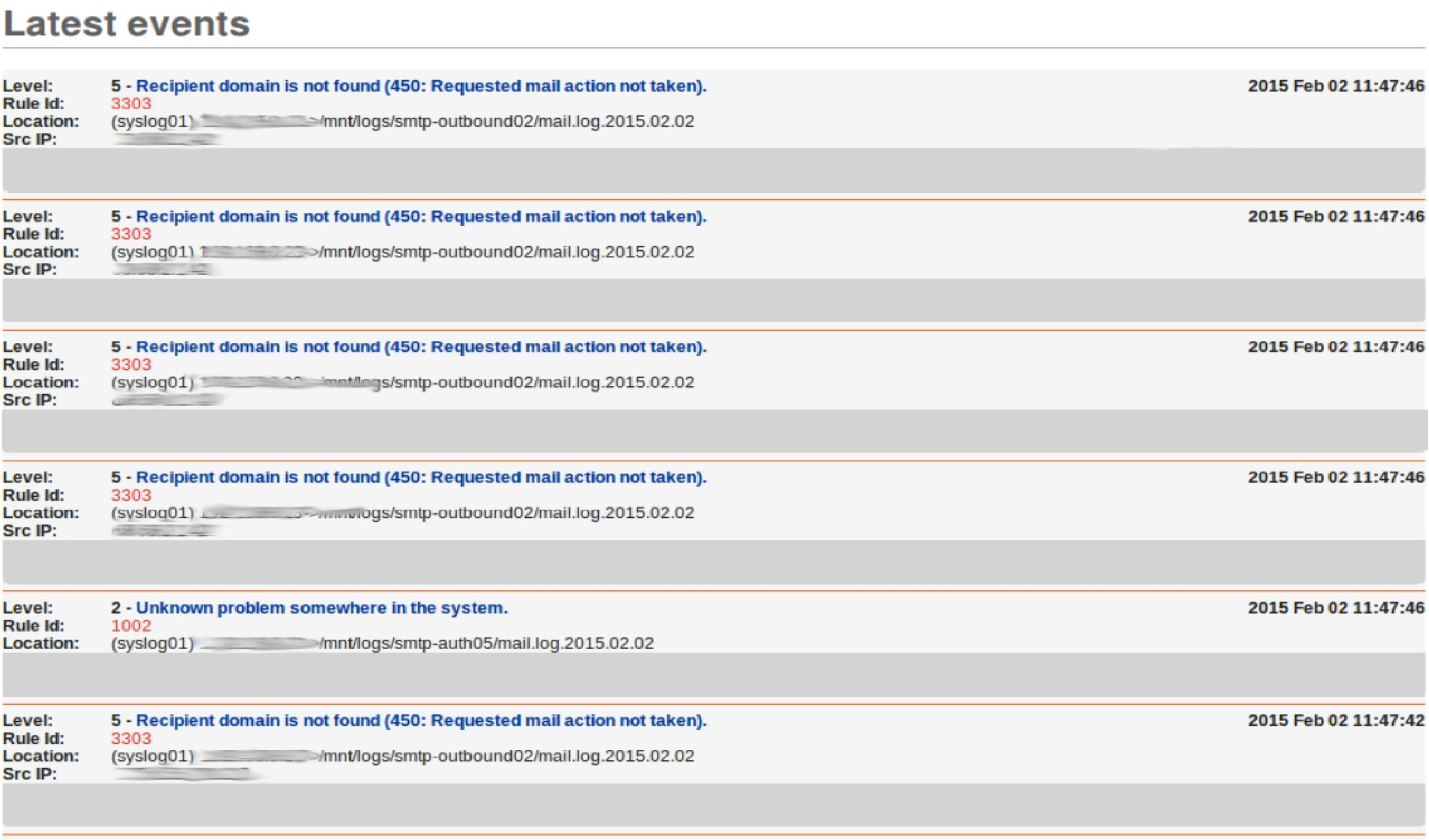


And while most have already moved away from MD5, there is still a notable group that heavily uses this obsolete algorithm: **security vendors.** It seems that MD5 became the de-facto standard of fingerprinting malware samples and the industry doesn't seem to be willing to move away from this practice. Our friend Zoltán Balázs collected a surprisingly long list of security vendors using MD5, including the biggest names of the field.

The list includes for example Kaspersky, the discoverer of Flame who just recently reminded us that MD5 is dead, but just a few weeks earlier released a report including MD5 fingerprints only – ironically even the malware they analysed uses SHA-1 internally…

“Poisonous MD5 – Wolves Among the Sheep”
blog.silentsignal.eu

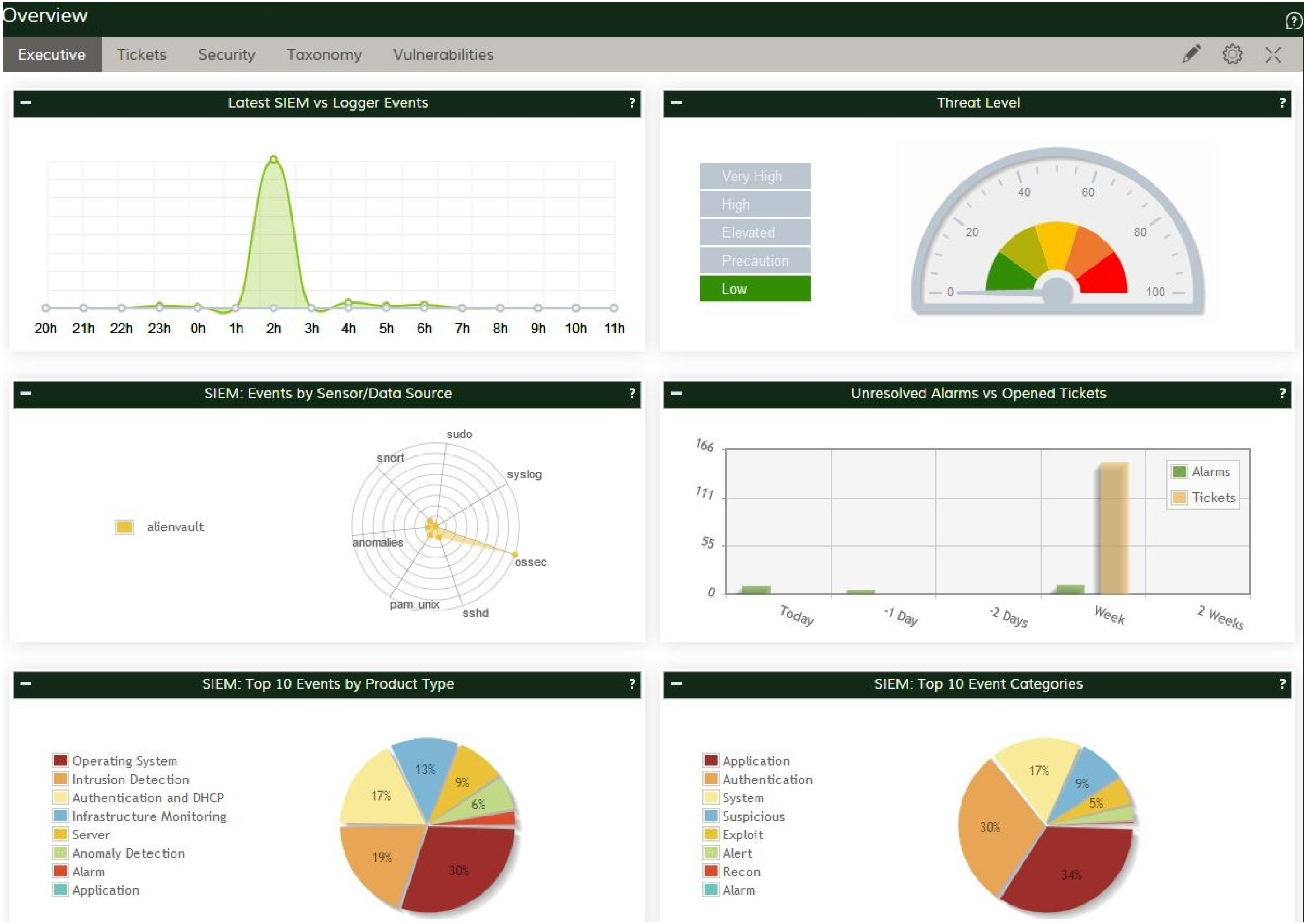
Insight?



Latest events

```
Level:      5 - Recipient domain is not found (450: Requested mail action not taken).      2015 Feb 02 11:47:46
Rule Id:    3303
Location:   (syslog01) [redacted]->/mnt/logs/smtp-outbound02/mail.log.2015.02.02
Src IP:     [redacted]

Level:      5 - Recipient domain is not found (450: Requested mail action not taken).      2015 Feb 02 11:47:46
Rule Id:    3303
Location:   (syslog01) [redacted]->/mnt/logs/smtp-outbound02/mail.log.2015.02.02
Src IP:     [redacted]

Level:      5 - Recipient domain is not found (450: Requested mail action not taken).      2015 Feb 02 11:47:46
Rule Id:    3303
Location:   (syslog01) [redacted]gs/smtp-outbound02/mail.log.2015.02.02
Src IP:     [redacted]

Level:      5 - Recipient domain is not found (450: Requested mail action not taken).      2015 Feb 02 11:47:46
Rule Id:    3303
Location:   (syslog01) [redacted]ogs/smtp-outbound02/mail.log.2015.02.02
Src IP:     [redacted]

Level:      2 - Unknown problem somewhere in the system.                                   2015 Feb 02 11:47:46
Rule Id:    1002
Location:   (syslog01) [redacted]->/mnt/logs/smtp-auth05/mail.log.2015.02.02

Level:      5 - Recipient domain is not found (450: Requested mail action not taken).      2015 Feb 02 11:47:42
Rule Id:    3303
Location:   (syslog01) [redacted]->/mnt/logs/smtp-outbound02/mail.log.2015.02.02
Src IP:     [redacted]
```

Insight?


Overview
Executive
Tickets
Security
Taxonomy
Vulnerabilities
Latest SIEM vs Logger Events
20h
21h
22h
23h
0h
1h
2h
3h
4h
5h
6h
7h
8h
9h
10h
11h
Threat Level
Very High
High
Elevated
Precaution
Low
0
20
40
60
80
100
SIEM: Events by Sensor/Data Source
alienvault
sudo
snort
syslog
anomalies
ossec
pam_unix
sshd
Unresolved Alarms vs Opened Tickets
166
111
55
0
Today
-1 Day
-2 Days
Week
2 Weeks
Alarms
Tickets
SIEM: Top 10 Events by Product Type
Operating System
Intrusion Detection
Authentication and DHCP
Infrastructure Monitoring
Server
Anomaly Detection
Alarm
Application
13%
9%
6%
17%
19%
30%
SIEM: Top 10 Event Categories
Application
Authentication
System
Suspicious
Exploit
Alert
Recon
Alarm
17%
9%
5%
30%
34%

Let's Make Our Own

But designing SIEM is not trivial, there are two problems to solve here

- Getting relevant intrusion information
  - Analysts overwhelmed by IDS alerting, false positives
- Presenting it in a useful way
  - Stock visualizations are full of **“chart junk”**, seldom match the way people process information or the context of your network
  - Always start with a question: What problem are you solving?

Dashboards: First Principles

Information Design – Tufte, Few – The practice of presenting information in a way that fosters efficient and effective understanding of it – Wikipedia

“...working memory is limited to three or four simultaneous chunks of information at a time”

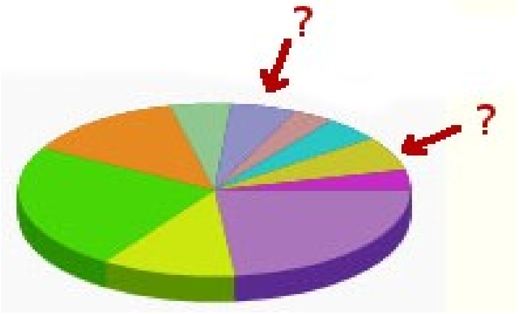
– Stephen Few

(Why Do We Visualize Quantitative Data?, http://www.perceptualedge.com/blog/?p=1897)

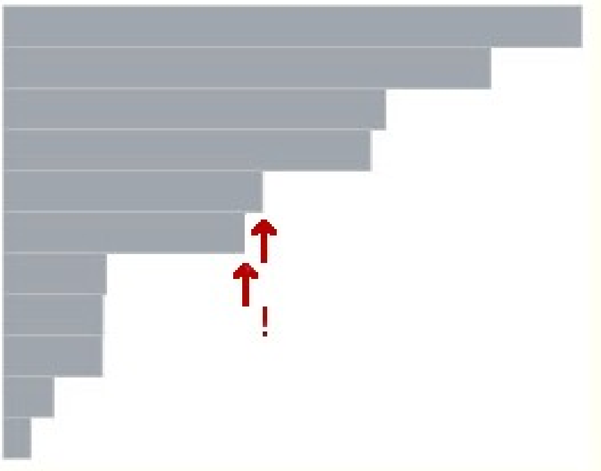
Design: Pie Charts Are Evil

Things people are bad at (not exhaustive)

- Estimating 2D areas



<- compare ->

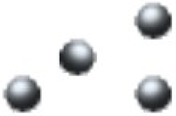


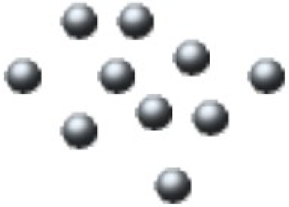
- Holding lists in working memory

Less is More: The Crow Epistemology

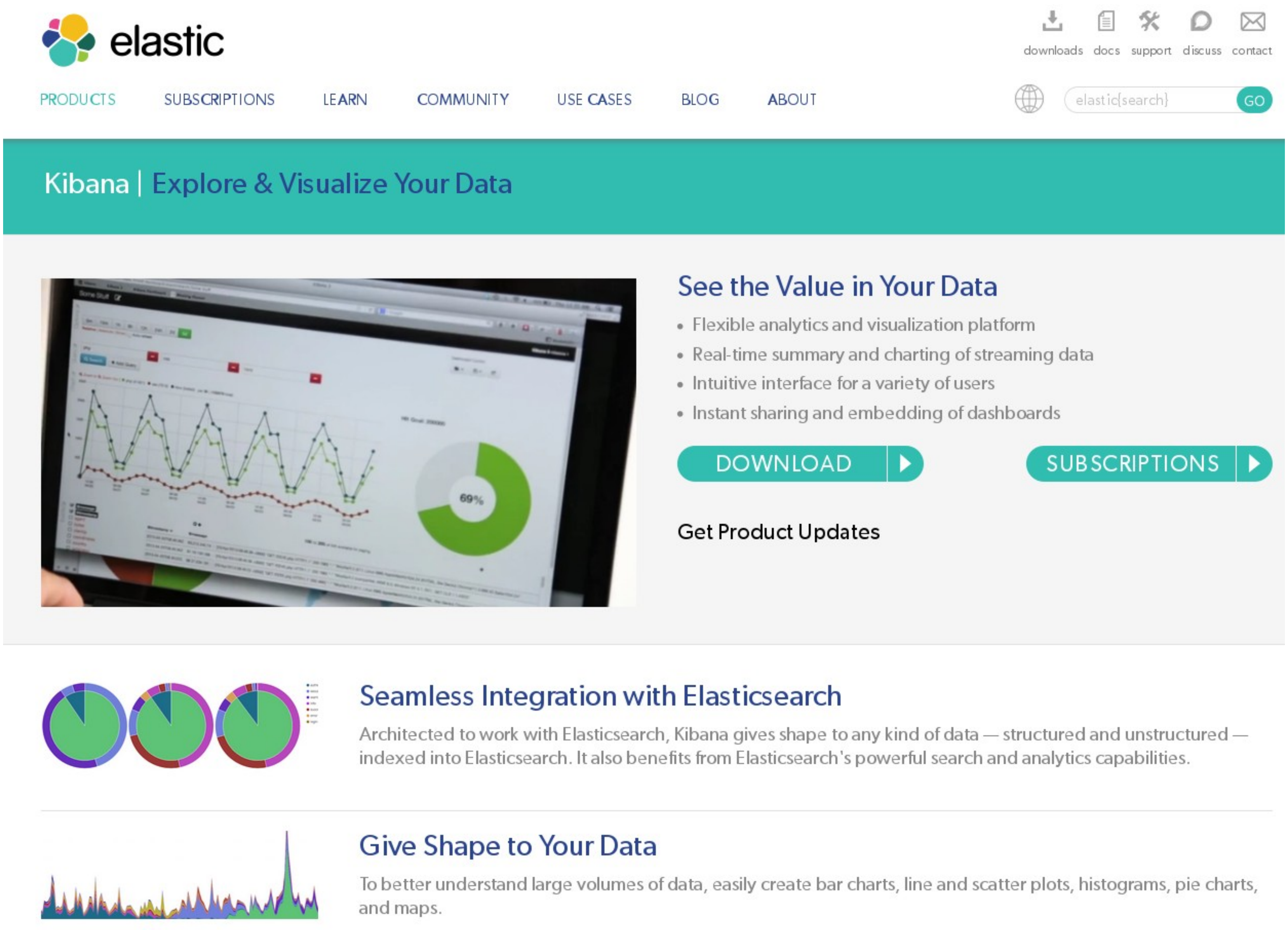
Quick, how many dots?

How about here?





ELK Stack – Open Source and Flexible



Kibana | Explore & Visualize Your Data

See the Value in Your Data

- Flexible analytics and visualization platform
- Real-time summary and charting of streaming data
- Intuitive interface for a variety of users
- Instant sharing and embedding of dashboards

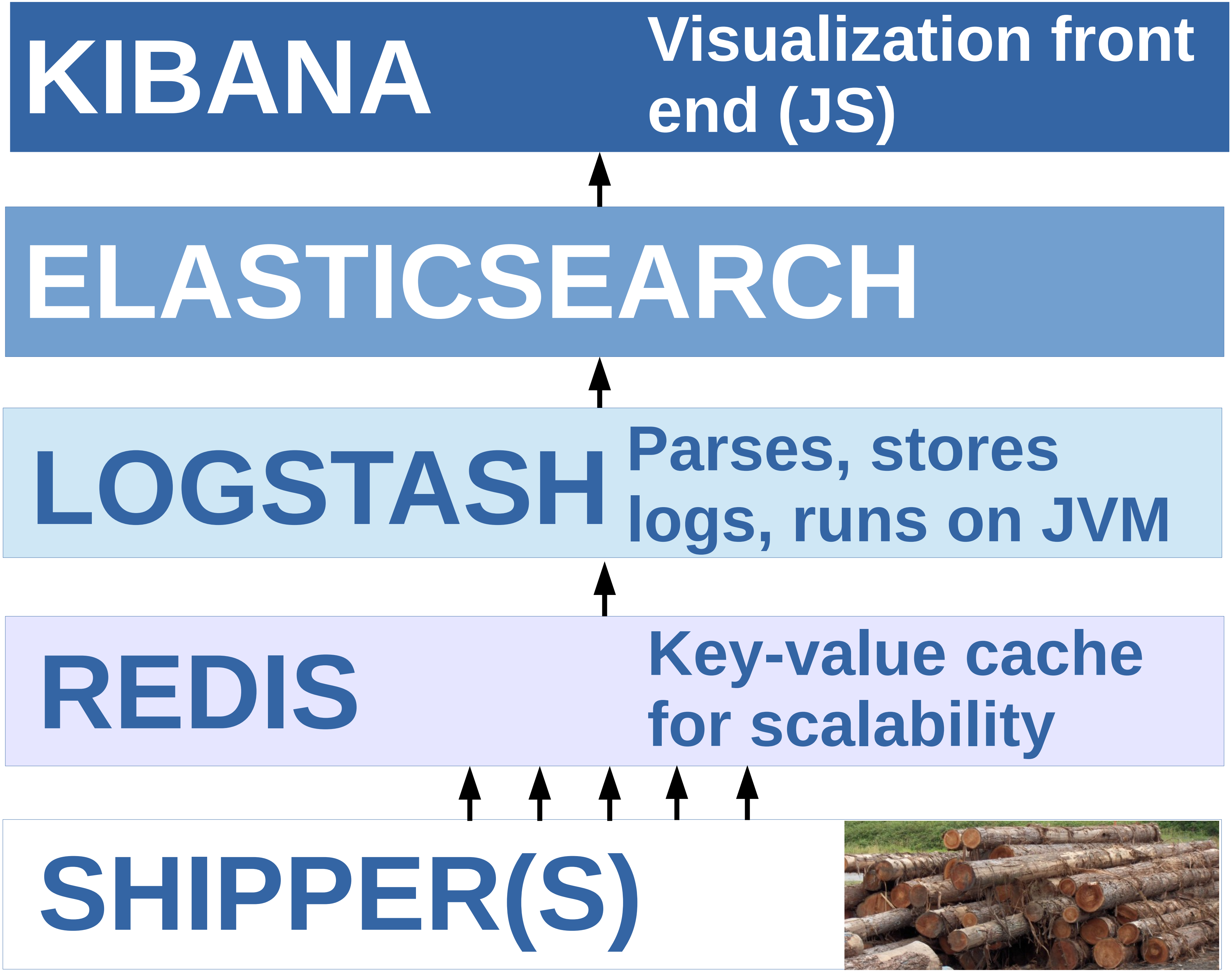
Get Product Updates

Seamless Integration with Elasticsearch

Architected to work with Elasticsearch, Kibana gives shape to any kind of data — structured and unstructured — indexed into Elasticsearch. It also benefits from Elasticsearch's powerful search and analytics capabilities.

Give Shape to Your Data

To better understand large volumes of data, easily create bar charts, line and scatter plots, histograms, pie charts, and maps.



KIBANA

Visualization front end (JS)

ELASTICSEARCH

LOGSTASH

Parses, stores logs, runs on JVM

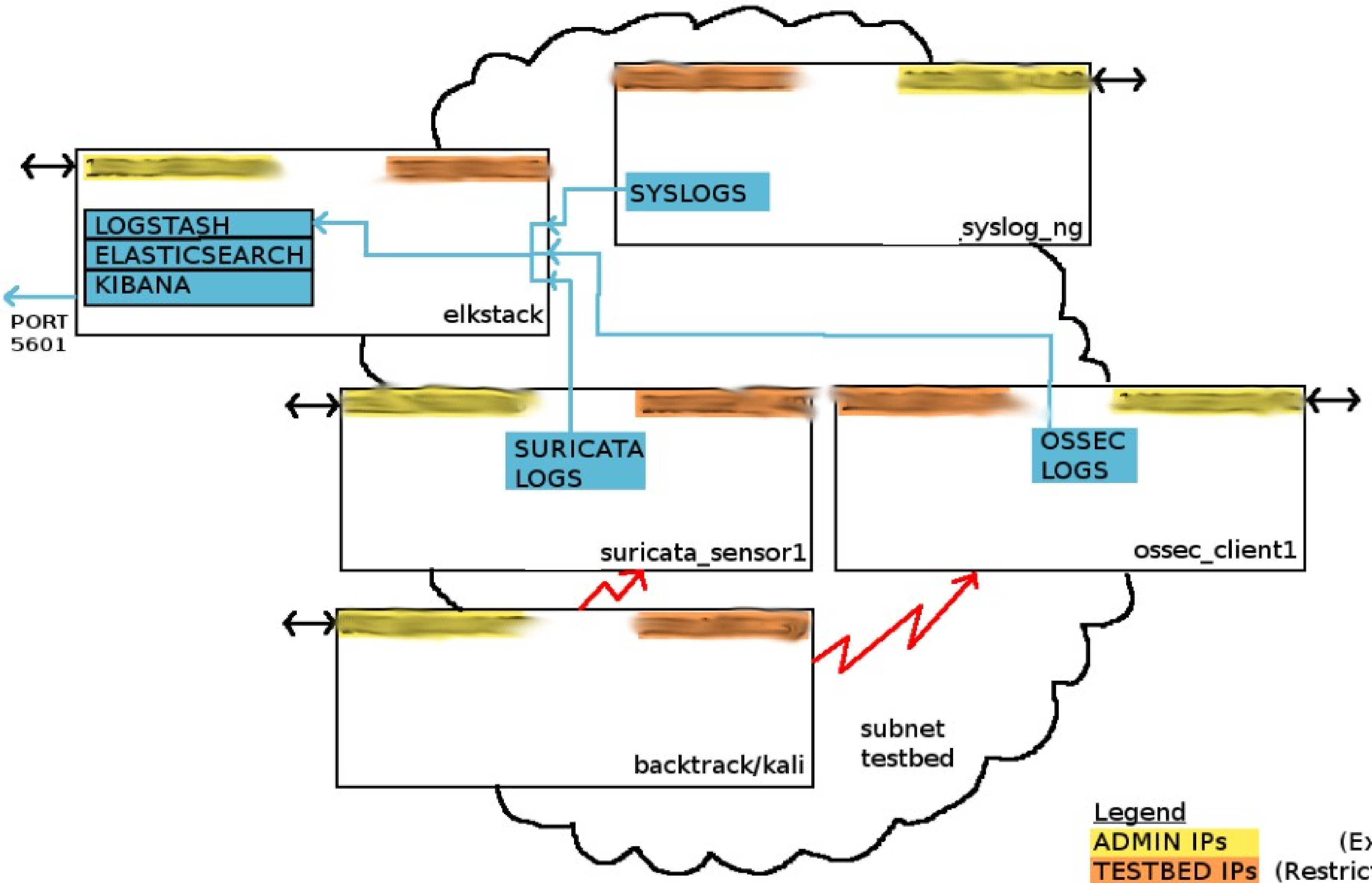
REDIS

Key-value cache for scalability

SHIPPER(S)

COLLECT
ALL THE THINGS!

Sample Testbed

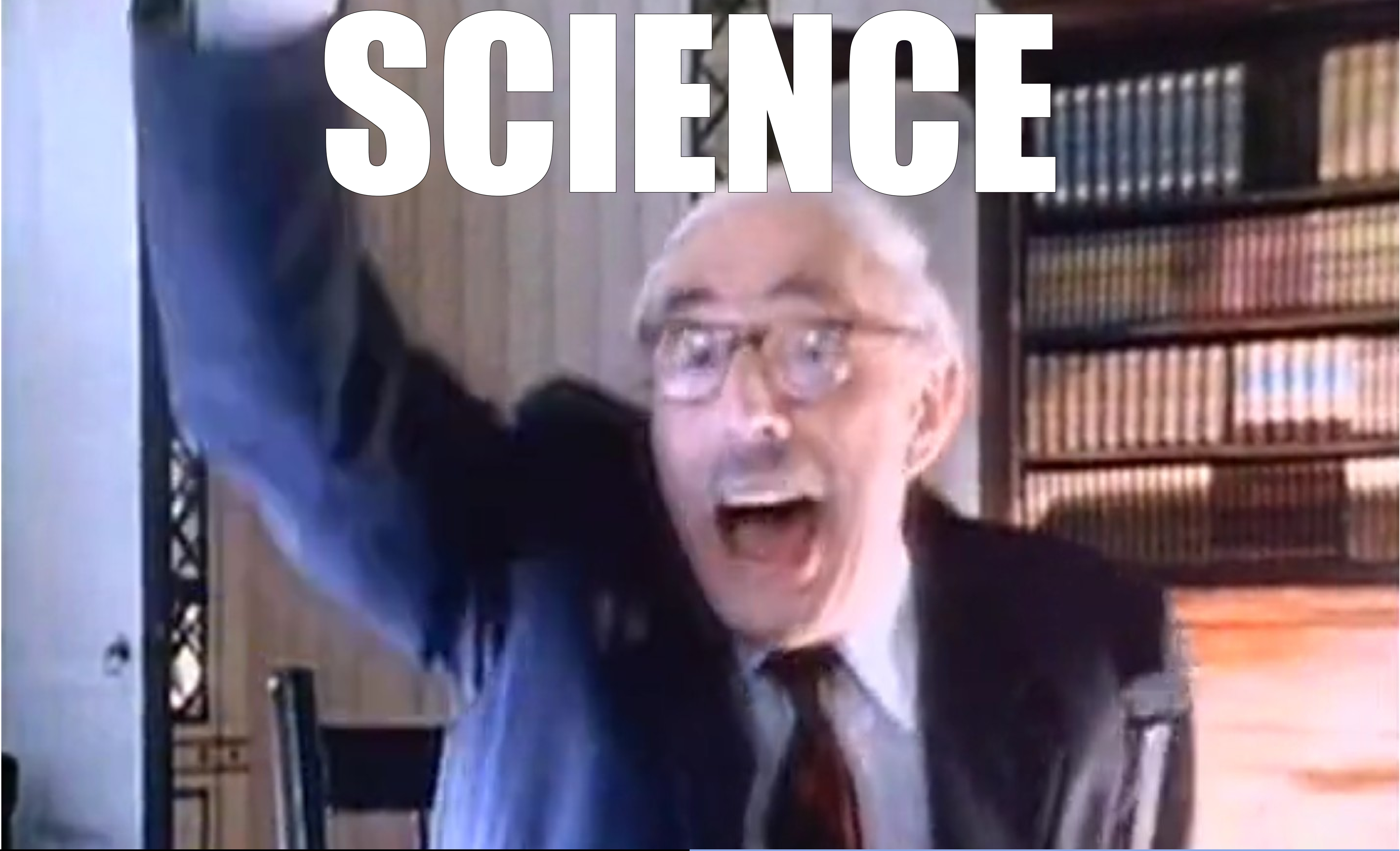


Legend

ADMIN IPs	(External Facing)
TESTBED IPs	(Restricted to subnet)
LOG DATA	(Arrows indicate flow)

NOW WHAT?

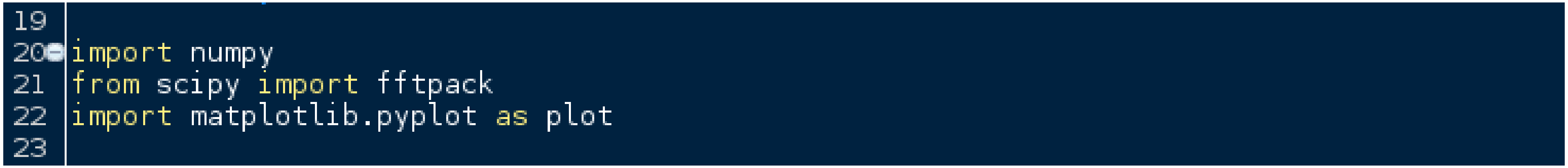
PATTERN RECOGNITION



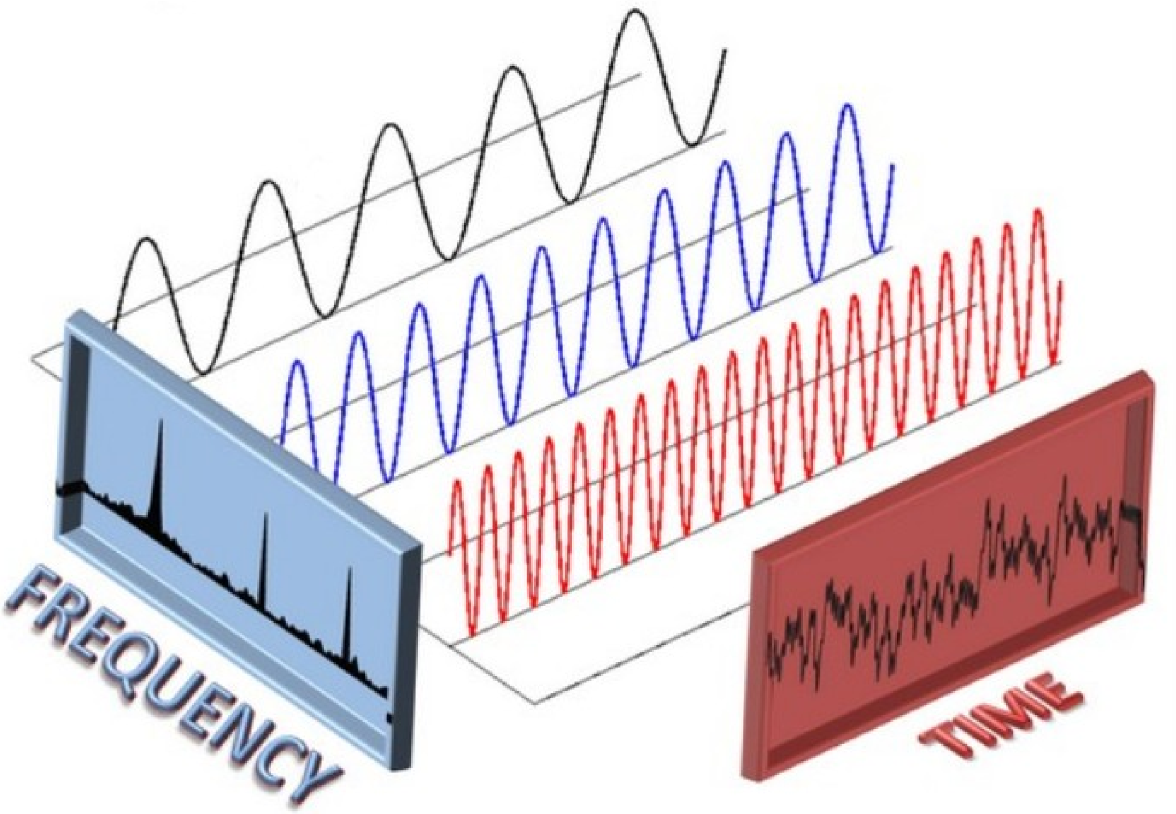
python™

SciPy.org

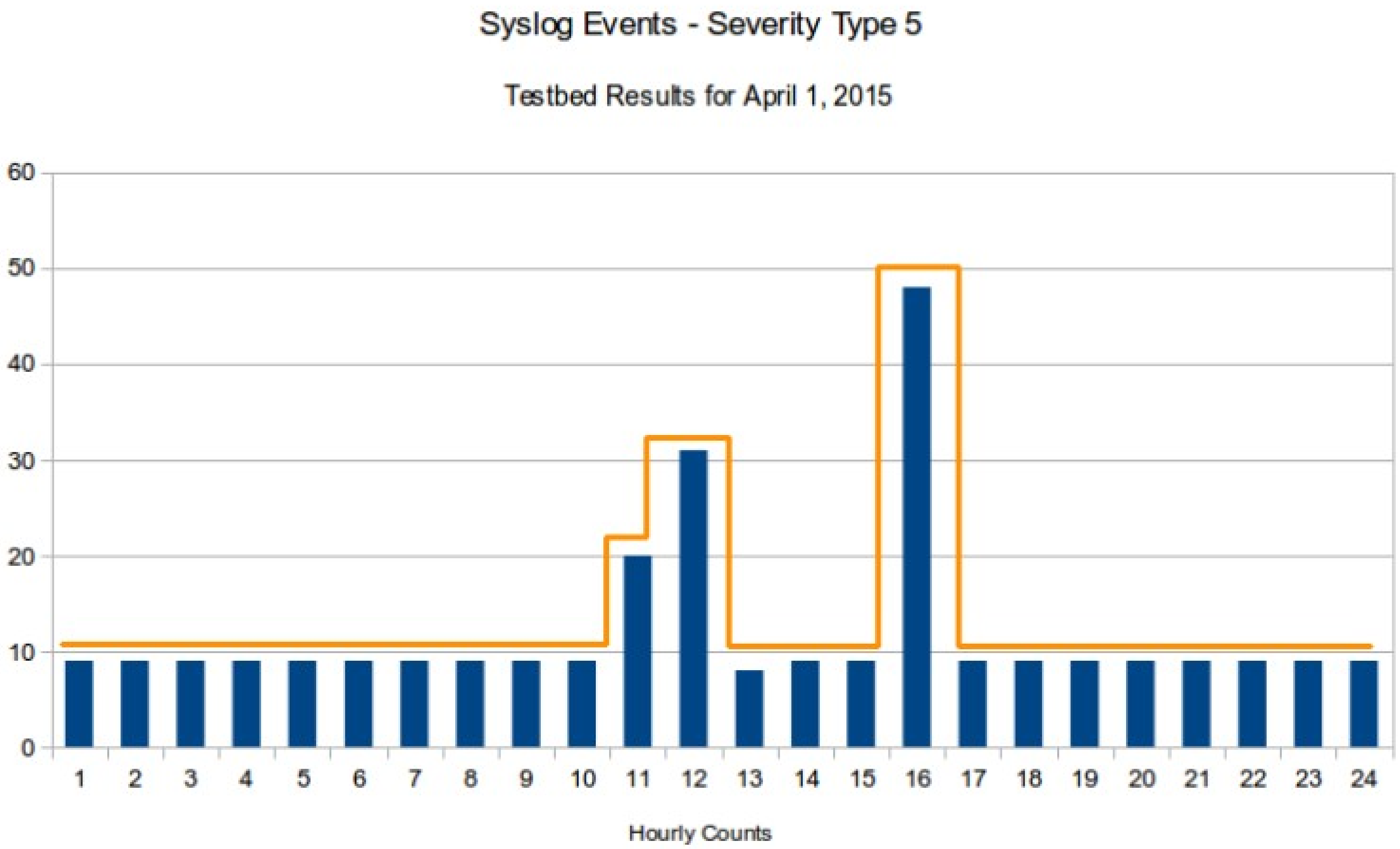
One Approach - Fast Fourier Transform



```python

import numpy
from scipy import fftpack
import matplotlib.pyplot as plot

```



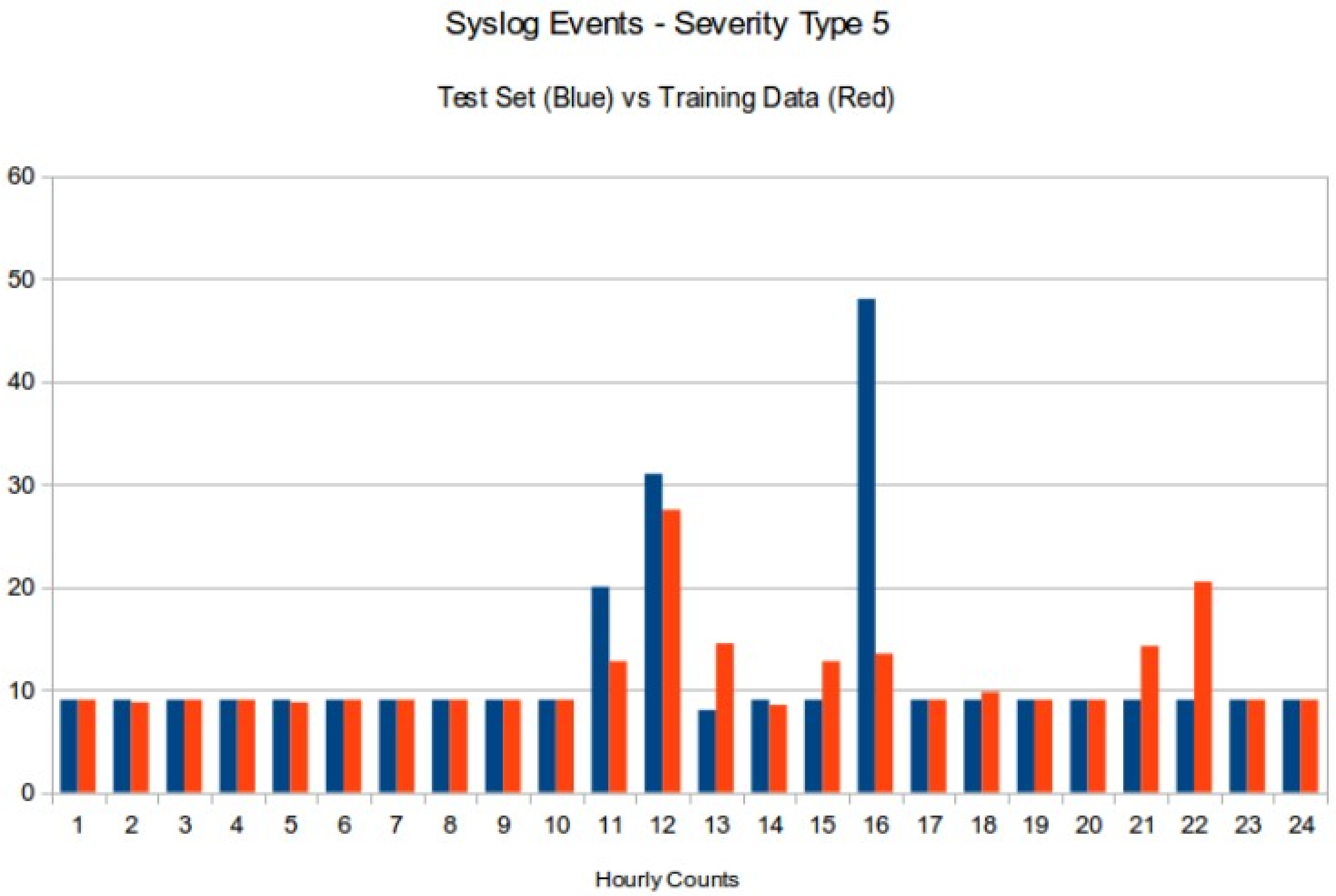
Time Series Data as a Waveform



Anomaly Detection

- Any time series (logs) may be viewed as a waveform
- Complex waveforms may be decomposed into simpler components, fingerprinted
- It's not hard to use off the shelf machine learning libraries to experiment with creating fingerprints of your own network traffic
- Machine learning in its simplest form is the comparison of test data to a set of training data
- In this case, I used a smoothed 30-day history of traffic by day and alert type to create a training dataset
- Any significant differences get flagged, tagged and reinserted into logstash

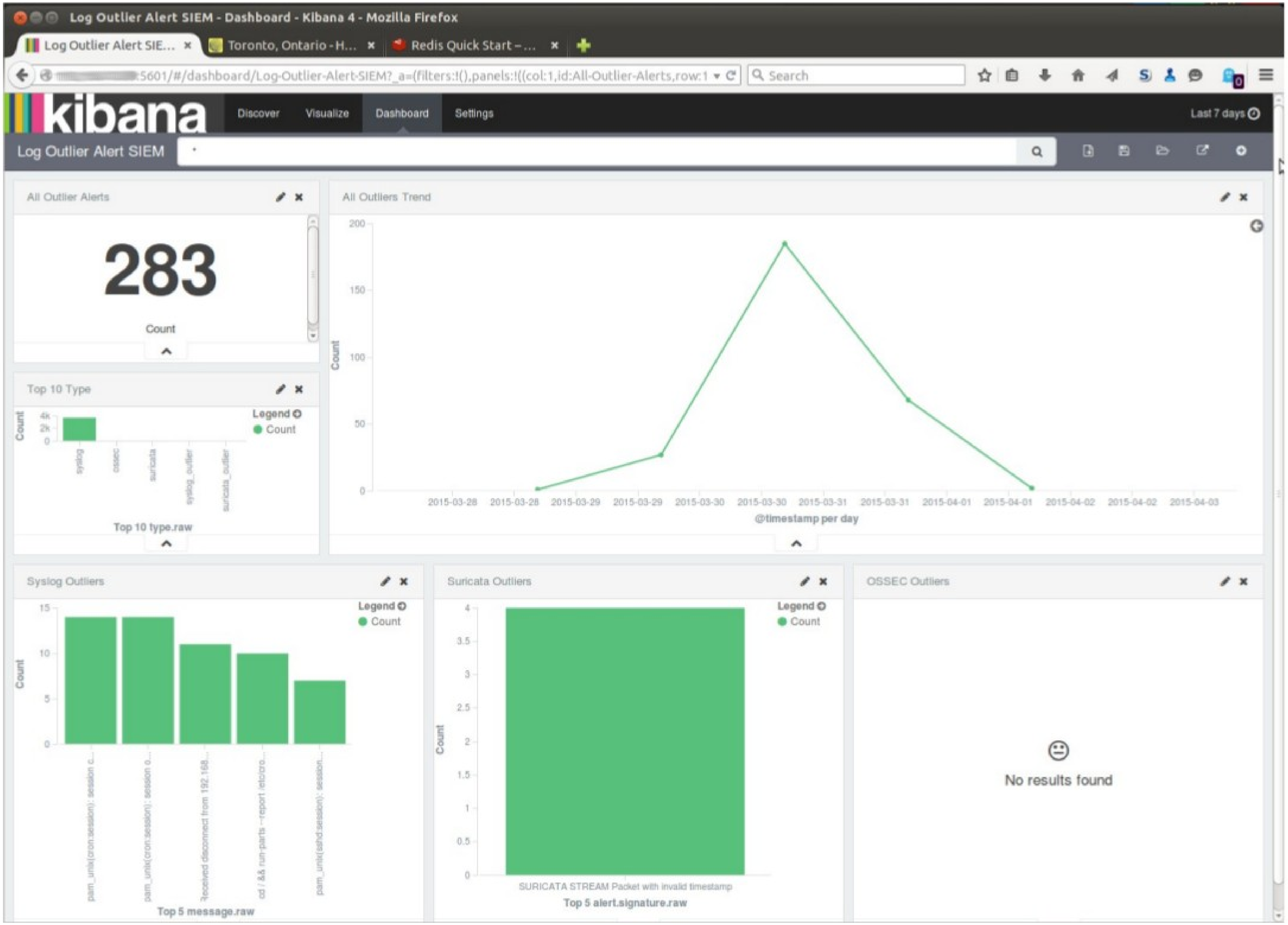
Test Set Data vs Training Set Data



Smooth Data

```
ALERT on test data:
[ 9   9   9   9   9   9   9   9   9   9 20 31  8   9   9 48   9   9   9   9   9
9   9   9]
vs training data
[ 9. 8.75 9. 9. 8.75 9. 9. 9. 9. 9. 12.75 27.5 14.5 8.5 12.75
13.5 9. 9.75 9. 9. 14.25 20.5 9. 9. ]
For logtype syslog, subtype 5 for date 2015-04-01, the value 20
at hour 10 may be an outlier!
Test/Train difference: 0.538976326886
Observation variance: 0.252585149809

ALERT on test data:
[ 9   9   9   9   9   9   9   9   9   9 20 31  8   9   9 48   9   9   9   9   9
9   9   9]
vs training data
[ 9. 8.75 9. 9. 8.75 9. 9. 9. 9. 9. 12.75 27.5 14.5 8.5 12.75
13.5 9. 9.75 9. 9. 14.25 20.5 9. 9. ]
For logtype syslog, subtype 5 for date 2015-04-01, the value 48
at hour 15 may be an outlier!
Test/Train difference: 0.538976326886
Observation variance: 0.44736895924

OUTLIER RESULTS found for 2015-04-01 by type, subtype in the
following hours:
{'syslog': {'5': [10, 15]}}

Results of reindexing outlying records
--------------------------------------
A total of 68 records were updated and saved as file
'records2015-04-01'
```

Display



Log Outlier Alert SIEM - Dashboard - Kibana 4 - Mozilla Firefox

[redacted]:5601/#/dashboard/Log-Outlier-Alert-SIEM?_a=(filters:!(),panels:!((col:1,id:All-Outlier-Alerts,row:1

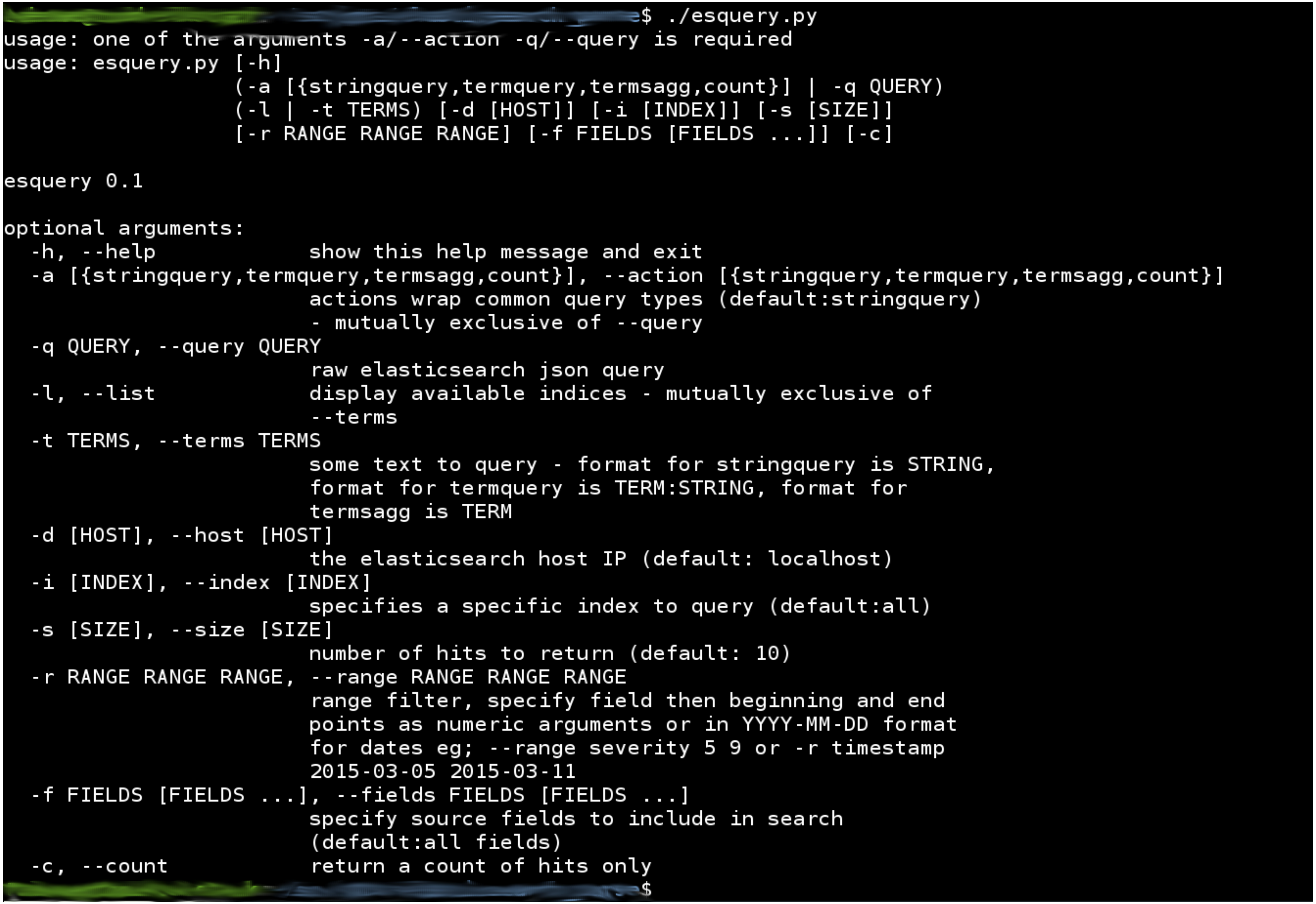
No results found

Demo(?)

Let us propitiate the demo gods

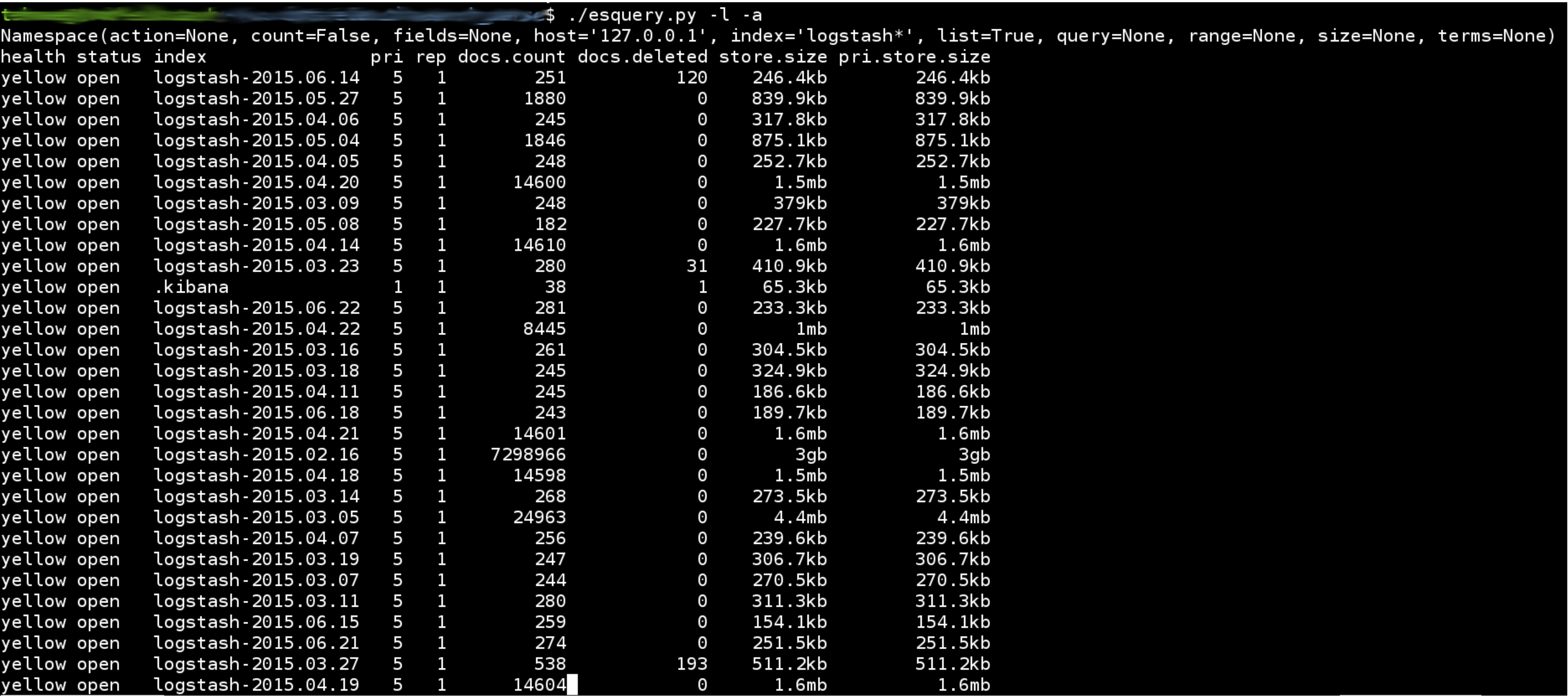
May they be merciful

esquery.py



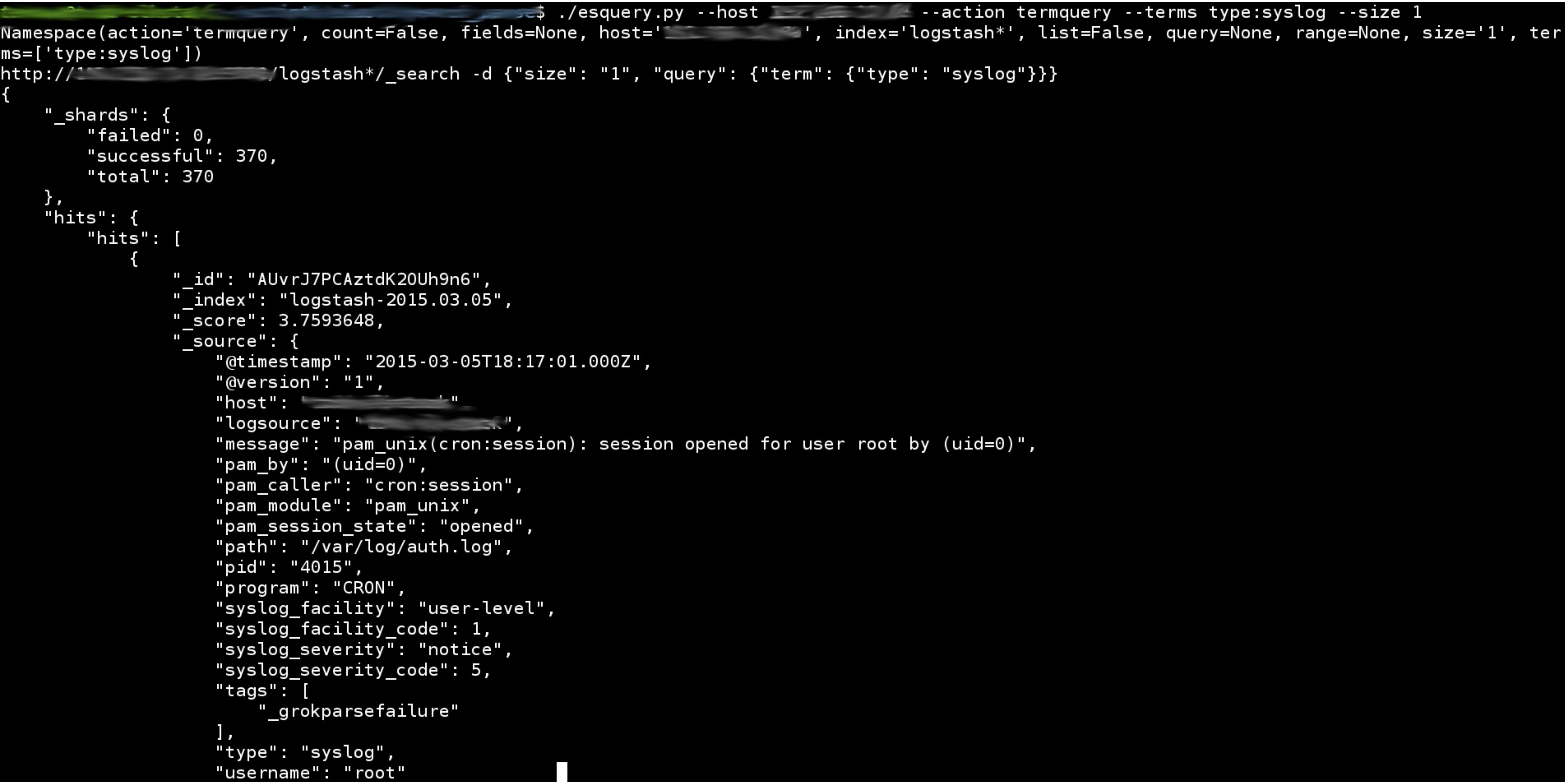
```
[redacted]$ ./esquery.py
usage: one of the arguments -a/--action -q/--query is required
usage: esquery.py [-h]
                  (-a [{stringquery,termquery,termsagg,count}] | -q QUERY)
                  (-l | -t TERMS) [-d [HOST]] [-i [INDEX]] [-s [SIZE]]
                  [-r RANGE RANGE RANGE] [-f FIELDS [FIELDS ...]] [-c]

esquery 0.1

optional arguments:
  -h, --help            show this help message and exit
  -a [{stringquery,termquery,termsagg,count}], --action [{stringquery,termquery,termsagg,count}]
                        actions wrap common query types (default:stringquery)
                        - mutually exclusive of --query
  -q QUERY, --query QUERY
                        raw elasticsearch json query
  -l, --list            display available indices - mutually exclusive of
                        --terms
  -t TERMS, --terms TERMS
                        some text to query - format for stringquery is STRING,
                        format for termquery is TERM:STRING, format for
                        termsagg is TERM
  -d [HOST], --host [HOST]
                        the elasticsearch host IP (default: localhost)
  -i [INDEX], --index [INDEX]
                        specifies a specific index to query (default:all)
  -s [SIZE], --size [SIZE]
                        number of hits to return (default: 10)
  -r RANGE RANGE RANGE, --range RANGE RANGE RANGE
                        range filter, specify field then beginning and end
                        points as numeric arguments or in YYYY-MM-DD format
                        for dates eg; --range severity 5 9 or -r timestamp
                        2015-03-05 2015-03-11
  -f FIELDS [FIELDS ...], --fields FIELDS [FIELDS ...]
                        specify source fields to include in search
                        (default:all fields)
  -c, --count           return a count of hits only
[redacted]$
```

Lets Look at the Logs We've Collected



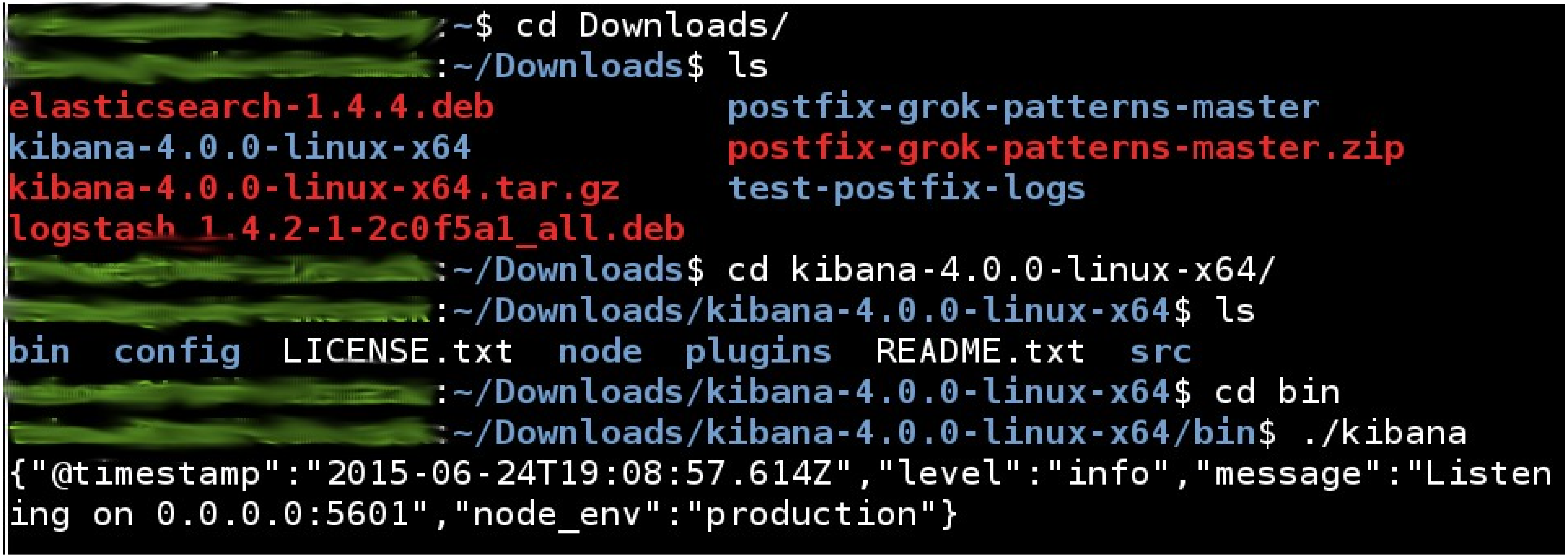
```
[redacted]$ ./esquery.py -l -a
Namespace(action=None, count=False, fields=None, host='127.0.0.1', index='logstash*', list=True, query=None, range=None, size=None, terms=None)
health status index               pri rep docs.count docs.deleted store.size pri.store.size
yellow open   logstash-2015.06.14   5   1        251          120    246.4kb        246.4kb
yellow open   logstash-2015.05.27   5   1       1880            0    839.9kb        839.9kb
yellow open   logstash-2015.04.06   5   1        245            0    317.8kb        317.8kb
yellow open   logstash-2015.05.04   5   1       1846            0    875.1kb        875.1kb
yellow open   logstash-2015.04.05   5   1        248            0    252.7kb        252.7kb
yellow open   logstash-2015.04.20   5   1      14600            0      1.5mb          1.5mb
yellow open   logstash-2015.03.09   5   1        248            0      379kb          379kb
yellow open   logstash-2015.05.08   5   1        182            0    227.7kb        227.7kb
yellow open   logstash-2015.04.14   5   1      14610            0      1.6mb          1.6mb
yellow open   logstash-2015.03.23   5   1        280           31    410.9kb        410.9kb
yellow open   .kibana               1   1         38            1     65.3kb         65.3kb
yellow open   logstash-2015.06.22   5   1        281            0    233.3kb        233.3kb
yellow open   logstash-2015.04.22   5   1       8445            0        1mb            1mb
yellow open   logstash-2015.03.16   5   1        261            0    304.5kb        304.5kb
yellow open   logstash-2015.03.18   5   1        245            0    324.9kb        324.9kb
yellow open   logstash-2015.04.11   5   1        245            0    186.6kb        186.6kb
yellow open   logstash-2015.06.18   5   1        243            0    189.7kb        189.7kb
yellow open   logstash-2015.04.21   5   1      14601            0      1.6mb          1.6mb
yellow open   logstash-2015.02.16   5   1    7298966            0        3gb            3gb
yellow open   logstash-2015.04.18   5   1      14598            0      1.5mb          1.5mb
yellow open   logstash-2015.03.14   5   1        268            0    273.5kb        273.5kb
yellow open   logstash-2015.03.05   5   1      24963            0      4.4mb          4.4mb
yellow open   logstash-2015.04.07   5   1        256            0    239.6kb        239.6kb
yellow open   logstash-2015.03.19   5   1        247            0    306.7kb        306.7kb
yellow open   logstash-2015.03.07   5   1        244            0    270.5kb        270.5kb
yellow open   logstash-2015.03.11   5   1        280            0    311.3kb        311.3kb
yellow open   logstash-2015.06.15   5   1        259            0    154.1kb        154.1kb
yellow open   logstash-2015.06.21   5   1        274            0    251.5kb        251.5kb
yellow open   logstash-2015.03.27   5   1        538          193    511.2kb        511.2kb
yellow open   logstash-2015.04.19   5   1      14604            0      1.6mb          1.6mb
```

Query a Single Syslog?



```
[redacted]$ ./esquery.py --host [redacted] --action termquery --terms type:syslog --size 1
Namespace(action='termquery', count=False, fields=None, host='[redacted]', index='logstash*', list=False, query=None, range=None, size='1', ter
ms=['type:syslog'])
http://[redacted]/logstash*/_search -d {"size": "1", "query": {"term": {"type": "syslog"}}}
{
    "_shards": {
        "failed": 0,
        "successful": 370,
        "total": 370
    },
    "hits": {
        "hits": [
            {
                "_id": "AUvrJ7PCAztdK20Uh9n6",
                "_index": "logstash-2015.03.05",
                "_score": 3.7593648,
                "_source": {
                    "@timestamp": "2015-03-05T18:17:01.000Z",
                    "@version": "1",
                    "host": [redacted]"
                    "logsource": [redacted]",
                    "message": "pam_unix(cron:session): session opened for user root by (uid=0)",
                    "pam_by": "(uid=0)",
                    "pam_caller": "cron:session",
                    "pam_module": "pam_unix",
                    "pam_session_state": "opened",
                    "path": "/var/log/auth.log",
                    "pid": "4015",
                    "program": "CRON",
                    "syslog_facility": "user-level",
                    "syslog_facility_code": 1,
                    "syslog_severity": "notice",
                    "syslog_severity_code": 5,
                    "tags": [
                        "_grokparsefailure"
                    ],
                    "type": "syslog",
                    "username": "root"
```

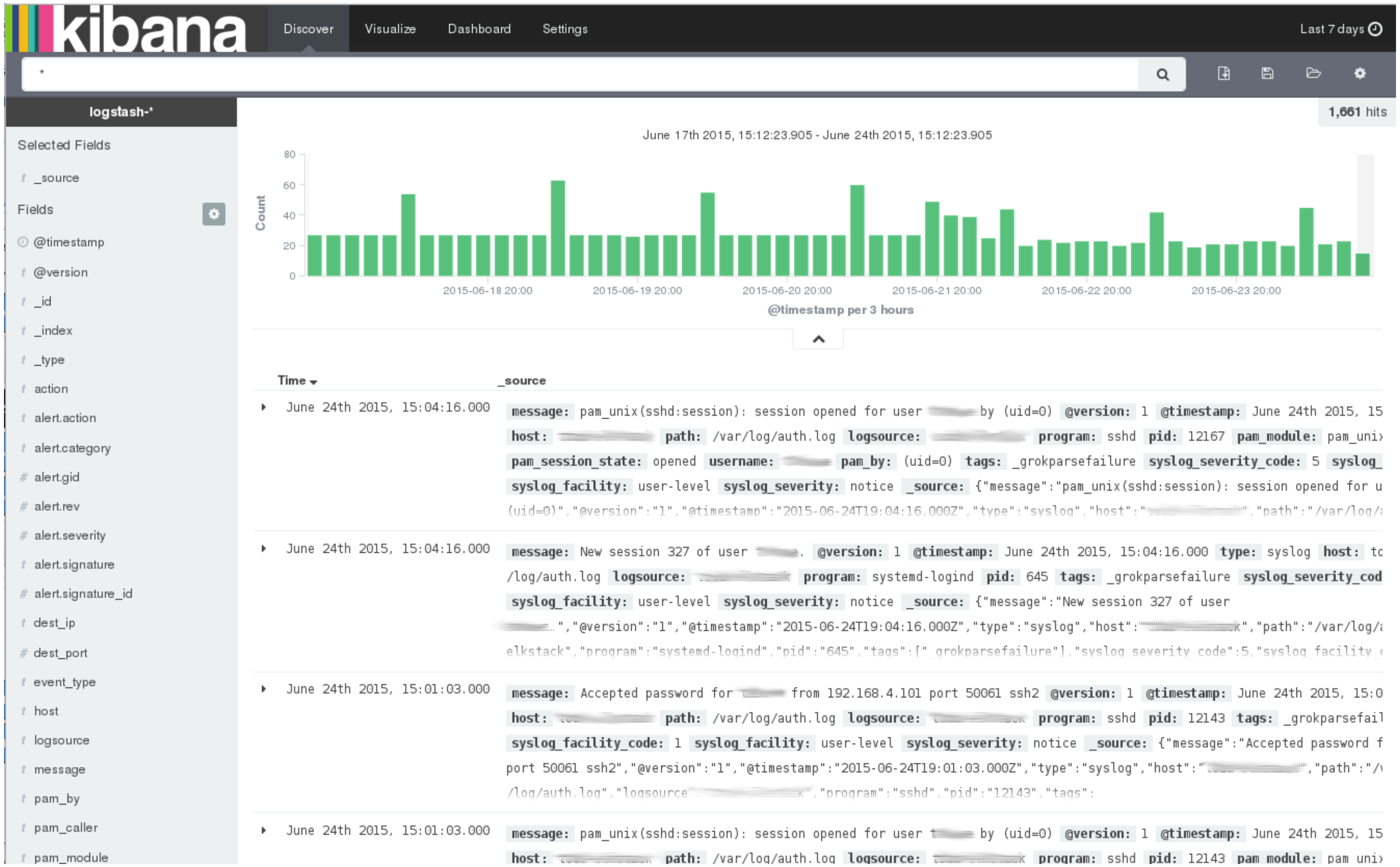
Start Kibana



```
[redacted]:~$ cd Downloads/
[redacted]:~/Downloads$ ls
elasticsearch-1.4.4.deb           postfix-grok-patterns-master
kibana-4.0.0-linux-x64            postfix-grok-patterns-master.zip
kibana-4.0.0-linux-x64.tar.gz     test-postfix-logs
logstash_1.4.2-1-2c0f5a1_all.deb
[redacted]:~/Downloads$ cd kibana-4.0.0-linux-x64/
[redacted]:~/Downloads/kibana-4.0.0-linux-x64$ ls
bin  config  LICENSE.txt  node  plugins  README.txt  src
[redacted]:~/Downloads/kibana-4.0.0-linux-x64$ cd bin
[redacted]:~/Downloads/kibana-4.0.0-linux-x64/bin$ ./kibana
{"@timestamp":"2015-06-24T19:08:57.614Z","level":"info","message":"Listen
ing on 0.0.0.0:5601","node_env":"production"}
```

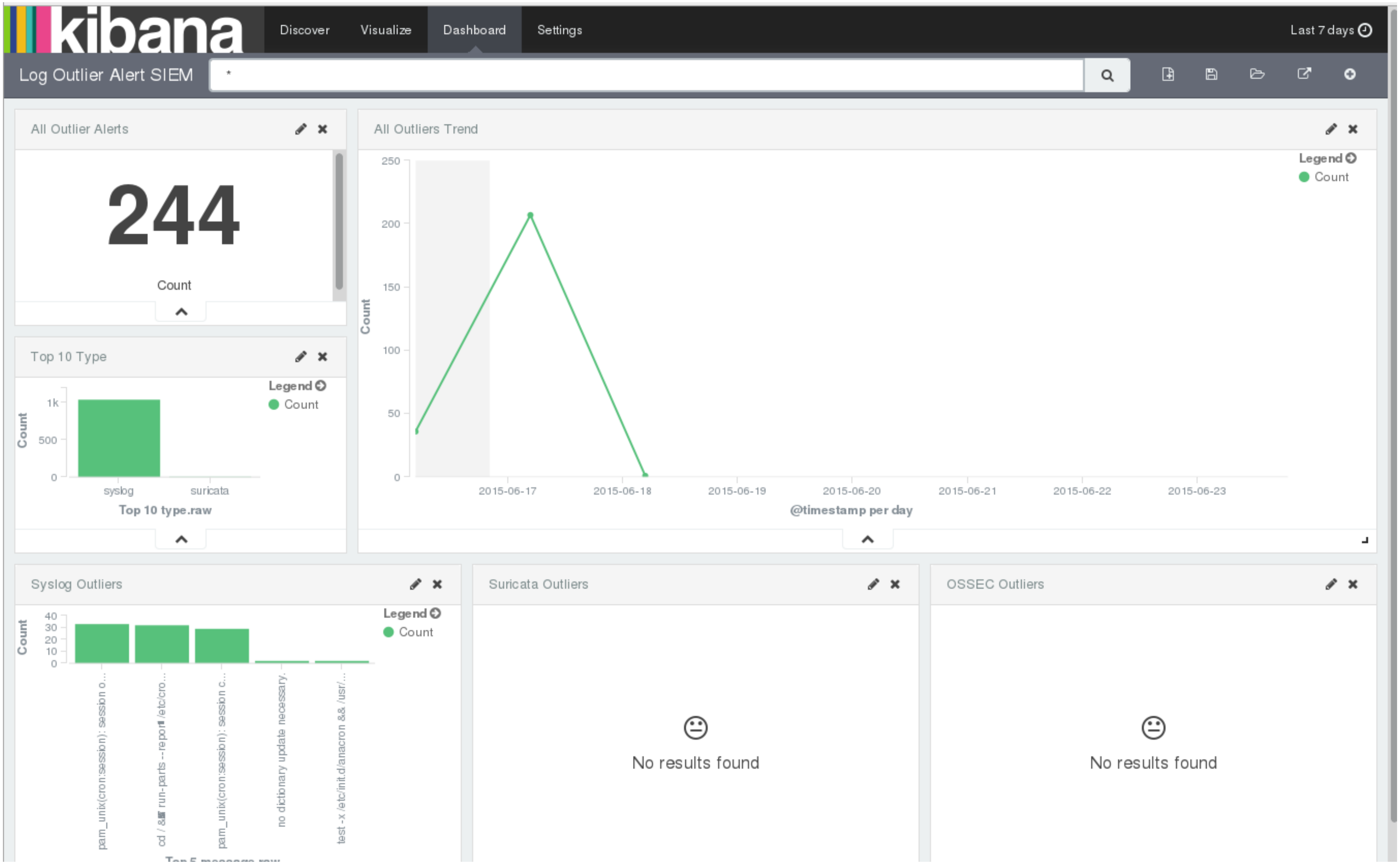
Launch the Kibana GUI

- Connect on port 5601
- Kibana spews a bunch of JSON to the console for debugging, and the browser displays the Discover view



```
June 24th 2015, 15:04:16.000   message: pam_unix(sshd:session): session opened for user [redacted] by (uid=0) @version: 1 @timestamp: June 24th 2015, 15
                               host: [redacted] path: /var/log/auth.log logsource: [redacted] program: sshd pid: 12167 pam_module: pam_uni
                               pam_session_state: opened username: [redacted] pam_by: (uid=0) tags: _grokparsefailure syslog_severity_code: 5 syslog_
                               syslog_facility: user-level syslog_severity: notice _source: {"message":"pam_unix(sshd:session): session opened for u
                               (uid=0)","@version":"1","@timestamp":"2015-06-24T19:04:16.000Z","type":"syslog","host":"[redacted]","path":"/var/log/a

June 24th 2015, 15:04:16.000   message: New session 327 of user [redacted]. @version: 1 @timestamp: June 24th 2015, 15:04:16.000 type: syslog host: tc
                               /log/auth.log logsource: [redacted] program: systemd-logind pid: 645 tags: _grokparsefailure syslog_severity_cod
                               syslog_facility: user-level syslog_severity: notice _source: {"message":"New session 327 of user
                               [redacted]","@version":"1","@timestamp":"2015-06-24T19:04:16.000Z","type":"syslog","host":"[redacted]","path":"/var/log/a
                               elkstack","program":"systemd-logind","pid":"645","tags":["_grokparsefailure"],"syslog_severity_code":5,"syslog_facility_c

June 24th 2015, 15:01:03.000   message: Accepted password for [redacted] from 192.168.4.101 port 50061 ssh2 @version: 1 @timestamp: June 24th 2015, 15:0
                               host: [redacted] path: /var/log/auth.log logsource: [redacted] program: sshd pid: 12143 tags: _grokparsefail
                               syslog_facility_code: 1 syslog_facility: user-level syslog_severity: notice _source: {"message":"Accepted password f
                               port 50061 ssh2","@version":"1","@timestamp":"2015-06-24T19:01:03.000Z","type":"syslog","host":"[redacted]","path":"/v
                               /log/auth.log","logsource[redacted]","program":"sshd","pid":"12143","tags":

June 24th 2015, 15:01:03.000   message: pam_unix(sshd:session): session opened for user [redacted] by (uid=0) @version: 1 @timestamp: June 24th 2015, 15
                               host: [redacted] path: /var/log/auth.log logsource: [redacted] program: sshd pid: 12143 pam_module: pam_uni
```

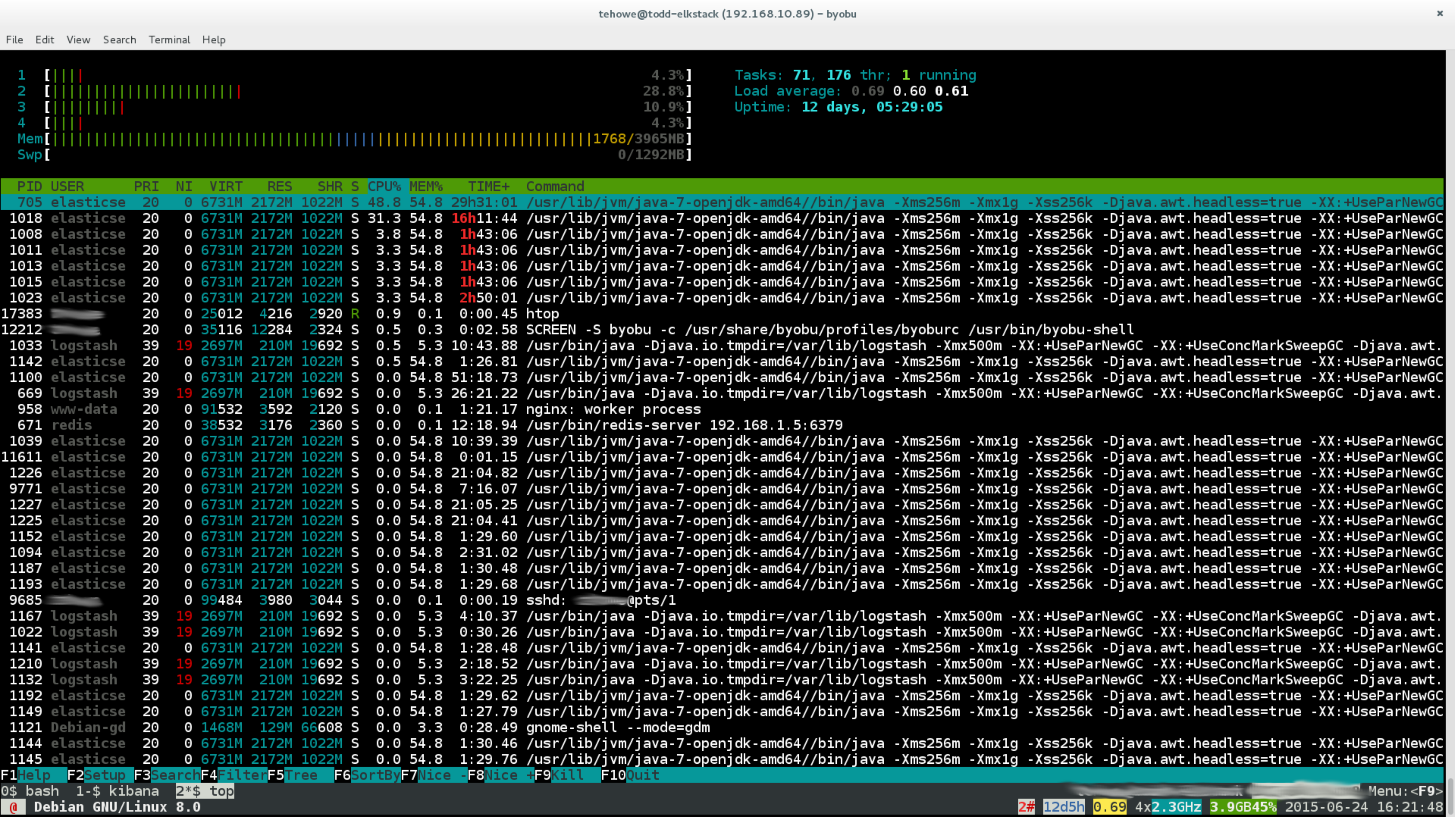
Dashboard View – Pre Analysis Jobs



Normalize, Vectorize, Analyze and Store

- set TEST_DAY in esconstants.py to the number of days ago the day you want to analyze was
- script currently only operates on an entire day
- then run ./escontrol.py
- uncaught exceptions (oops) mean sharding failed temporarily on Elasticsearch, try again or add a new node to the cluster

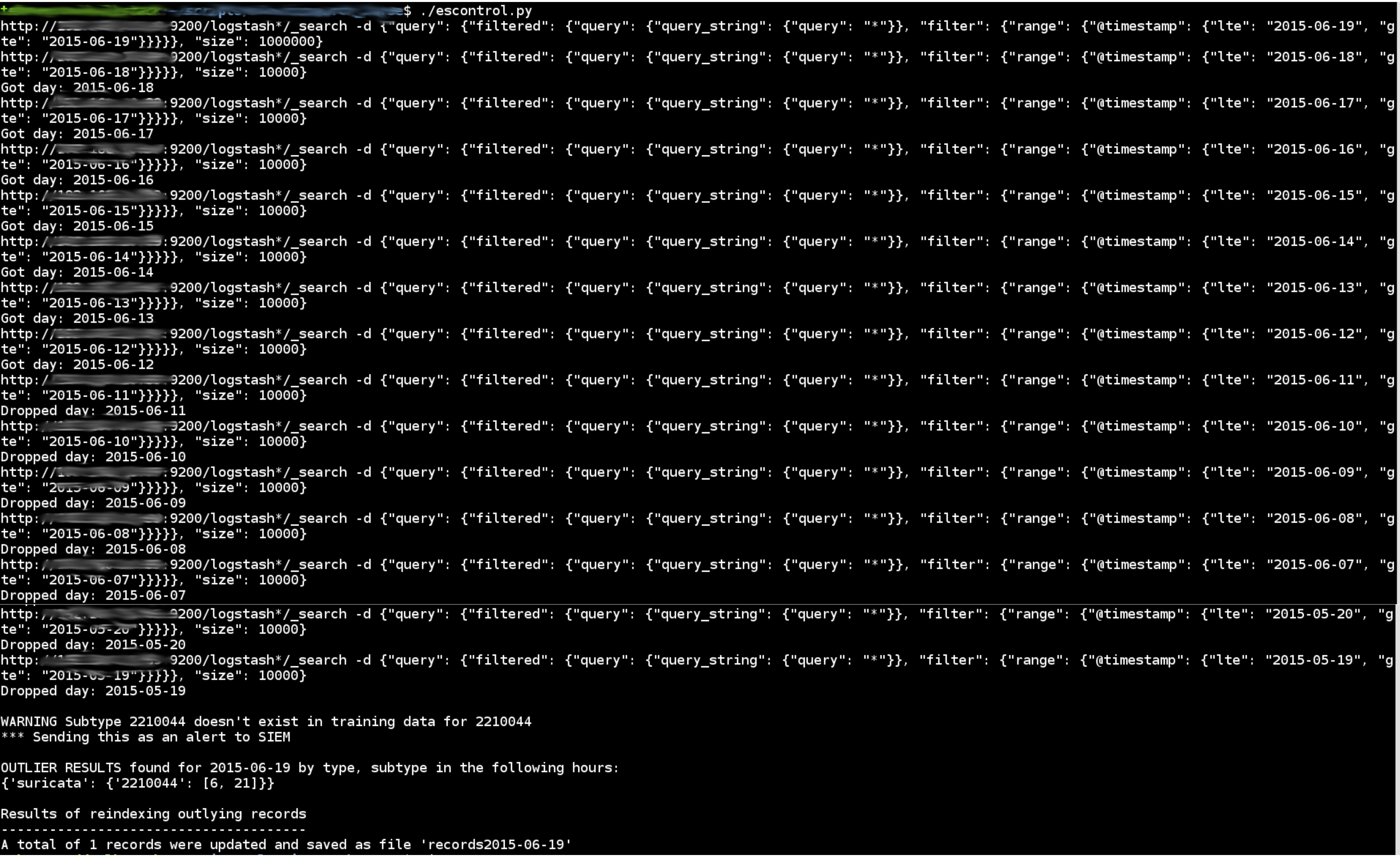
Busy Busy JVM



tehowe@todd-elkstack (192.168.10.89) - byobu
File Edit View Search Terminal Help

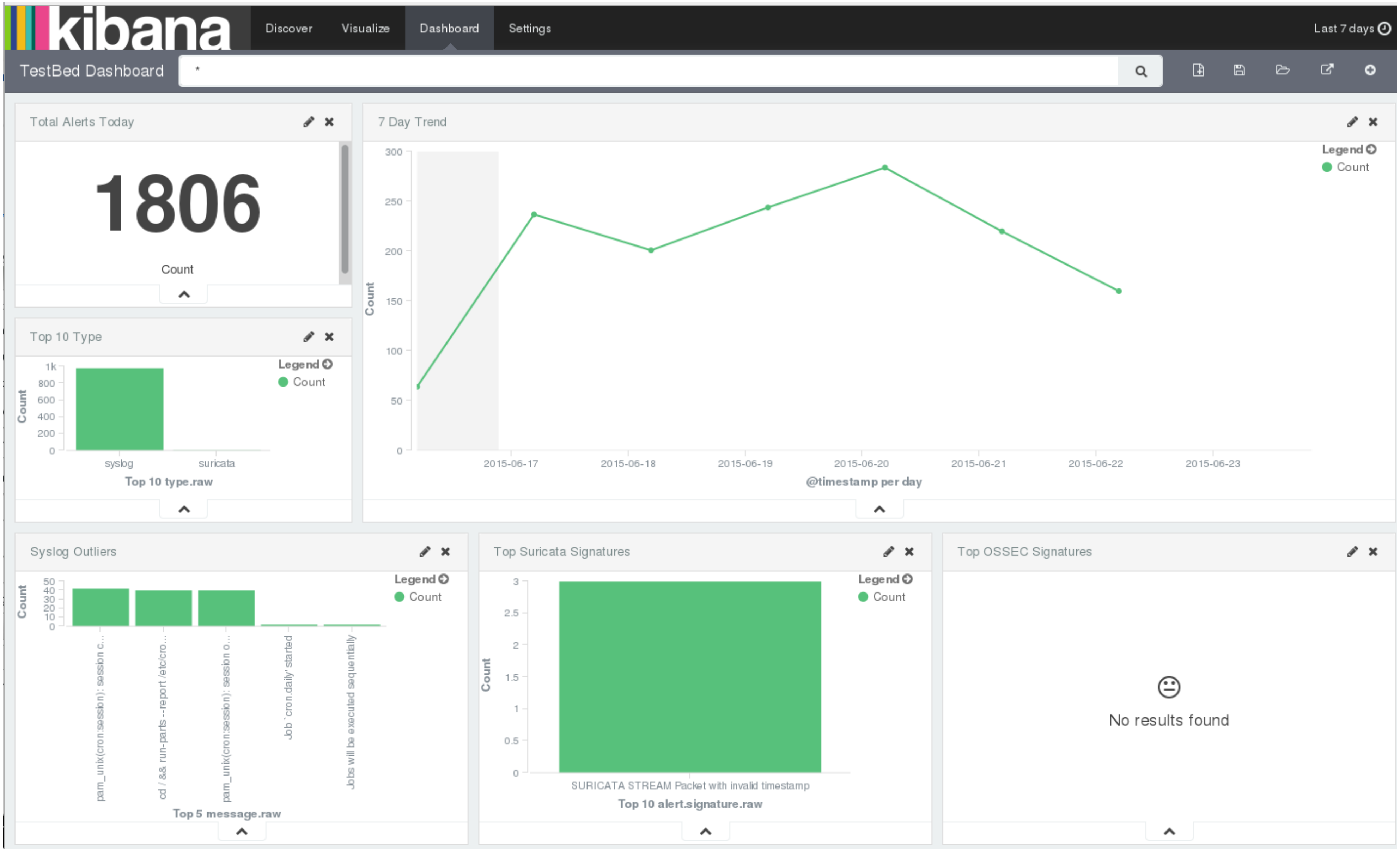
```
  1  [||||                                                                          4.3%]     Tasks: 71, 176 thr; 1 running
  2  [||||||||||||||||||||||                                                       28.8%]     Load average: 0.69 0.60 0.61
  3  [|||||||||                                                                    10.9%]     Uptime: 12 days, 05:29:05
  4  [||||                                                                          4.3%]
  Mem[|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||1768/3965MB]
  Swp[                                                                        0/1292MB]

  PID USER      PRI  NI  VIRT   RES   SHR S CPU% MEM%   TIME+  Command
  705 elasticse  20   0 6731M 2172M 1022M S 48.8 54.8 29h31:01 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
 1018 elasticse  20   0 6731M 2172M 1022M S 31.3 54.8 16h11:44 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
 1008 elasticse  20   0 6731M 2172M 1022M S  3.8 54.8  1h43:06 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
 1011 elasticse  20   0 6731M 2172M 1022M S  3.3 54.8  1h43:06 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
 1013 elasticse  20   0 6731M 2172M 1022M S  3.3 54.8  1h43:06 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
 1015 elasticse  20   0 6731M 2172M 1022M S  3.3 54.8  1h43:06 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
 1023 elasticse  20   0 6731M 2172M 1022M S  3.3 54.8  2h50:01 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
17383 [redacted]  20   0 25012  4216  2920 R  0.9  0.1  0:00.45 htop
12212 [redacted]  20   0 35116 12284  2324 S  0.5  0.3  0:02.58 SCREEN -S byobu -c /usr/share/byobu/profiles/byoburc /usr/bin/byobu-shell
 1033 logstash   39  19 2697M  210M 19692 S  0.5  5.3 10:43.88 /usr/bin/java -Djava.io.tmpdir=/var/lib/logstash -Xmx500m -XX:+UseParNewGC -XX:+UseConcMarkSweepGC -Djava.awt.
 1142 elasticse  20   0 6731M 2172M 1022M S  0.5 54.8  1:26.81 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
 1100 elasticse  20   0 6731M 2172M 1022M S  0.0 54.8 51:18.73 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
  669 logstash   39  19 2697M  210M 19692 S  0.0  5.3 26:21.22 /usr/bin/java -Djava.io.tmpdir=/var/lib/logstash -Xmx500m -XX:+UseParNewGC -XX:+UseConcMarkSweepGC -Djava.awt.
  958 www-data   20   0 91532  3592  2120 S  0.0  0.1  1:21.17 nginx: worker process
  671 redis      20   0 38532  3176  2360 S  0.0  0.1 12:18.94 /usr/bin/redis-server 192.168.1.5:6379
 1039 elasticse  20   0 6731M 2172M 1022M S  0.0 54.8 10:39.39 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
11611 elasticse  20   0 6731M 2172M 1022M S  0.0 54.8  0:01.15 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
 1226 elasticse  20   0 6731M 2172M 1022M S  0.0 54.8 21:04.82 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
 9771 elasticse  20   0 6731M 2172M 1022M S  0.0 54.8  7:16.07 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
 1227 elasticse  20   0 6731M 2172M 1022M S  0.0 54.8 21:05.25 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
 1225 elasticse  20   0 6731M 2172M 1022M S  0.0 54.8 21:04.41 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
 1152 elasticse  20   0 6731M 2172M 1022M S  0.0 54.8  1:29.60 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
 1094 elasticse  20   0 6731M 2172M 1022M S  0.0 54.8  2:31.02 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
 1187 elasticse  20   0 6731M 2172M 1022M S  0.0 54.8  1:30.48 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
 1193 elasticse  20   0 6731M 2172M 1022M S  0.0 54.8  1:29.68 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
 9685 [redacted]  20   0 99484  3980  3044 S  0.0  0.1  0:00.19 sshd: [redacted]@pts/1
 1167 logstash   39  19 2697M  210M 19692 S  0.0  5.3  4:10.37 /usr/bin/java -Djava.io.tmpdir=/var/lib/logstash -Xmx500m -XX:+UseParNewGC -XX:+UseConcMarkSweepGC -Djava.awt.
 1022 logstash   39  19 2697M  210M 19692 S  0.0  5.3  0:30.26 /usr/bin/java -Djava.io.tmpdir=/var/lib/logstash -Xmx500m -XX:+UseParNewGC -XX:+UseConcMarkSweepGC -Djava.awt.
 1141 elasticse  20   0 6731M 2172M 1022M S  0.0 54.8  1:28.48 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
 1210 logstash   39  19 2697M  210M 19692 S  0.0  5.3  2:18.52 /usr/bin/java -Djava.io.tmpdir=/var/lib/logstash -Xmx500m -XX:+UseParNewGC -XX:+UseConcMarkSweepGC -Djava.awt.
 1132 logstash   39  19 2697M  210M 19692 S  0.0  5.3  3:22.25 /usr/bin/java -Djava.io.tmpdir=/var/lib/logstash -Xmx500m -XX:+UseParNewGC -XX:+UseConcMarkSweepGC -Djava.awt.
 1192 elasticse  20   0 6731M 2172M 1022M S  0.0 54.8  1:29.62 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
 1149 elasticse  20   0 6731M 2172M 1022M S  0.0 54.8  1:27.79 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
 1121 Debian-gd  20   0 1468M  129M 66608 S  0.0  3.3  0:28.49 gnome-shell --mode=gdm
 1144 elasticse  20   0 6731M 2172M 1022M S  0.0 54.8  1:30.46 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
 1145 elasticse  20   0 6731M 2172M 1022M S  0.0 54.8  1:29.76 /usr/lib/jvm/java-7-openjdk-amd64//bin/java -Xms256m -Xmx1g -Xss256k -Djava.awt.headless=true -XX:+UseParNewGC
F1Help  F2Setup F3SearchF4FilterF5Tree  F6SortByF7Nice -F8Nice +F9Kill  F10Quit
0$ bash  1-$ kibana  2*$ top                                                          [redacted]  Menu:<F9>
 @  Debian GNU/Linux 8.0                                     2# 12d5h 0.69 4x2.3GHz 3.9GB45% 2015-06-24 16:21:48
```

Results!



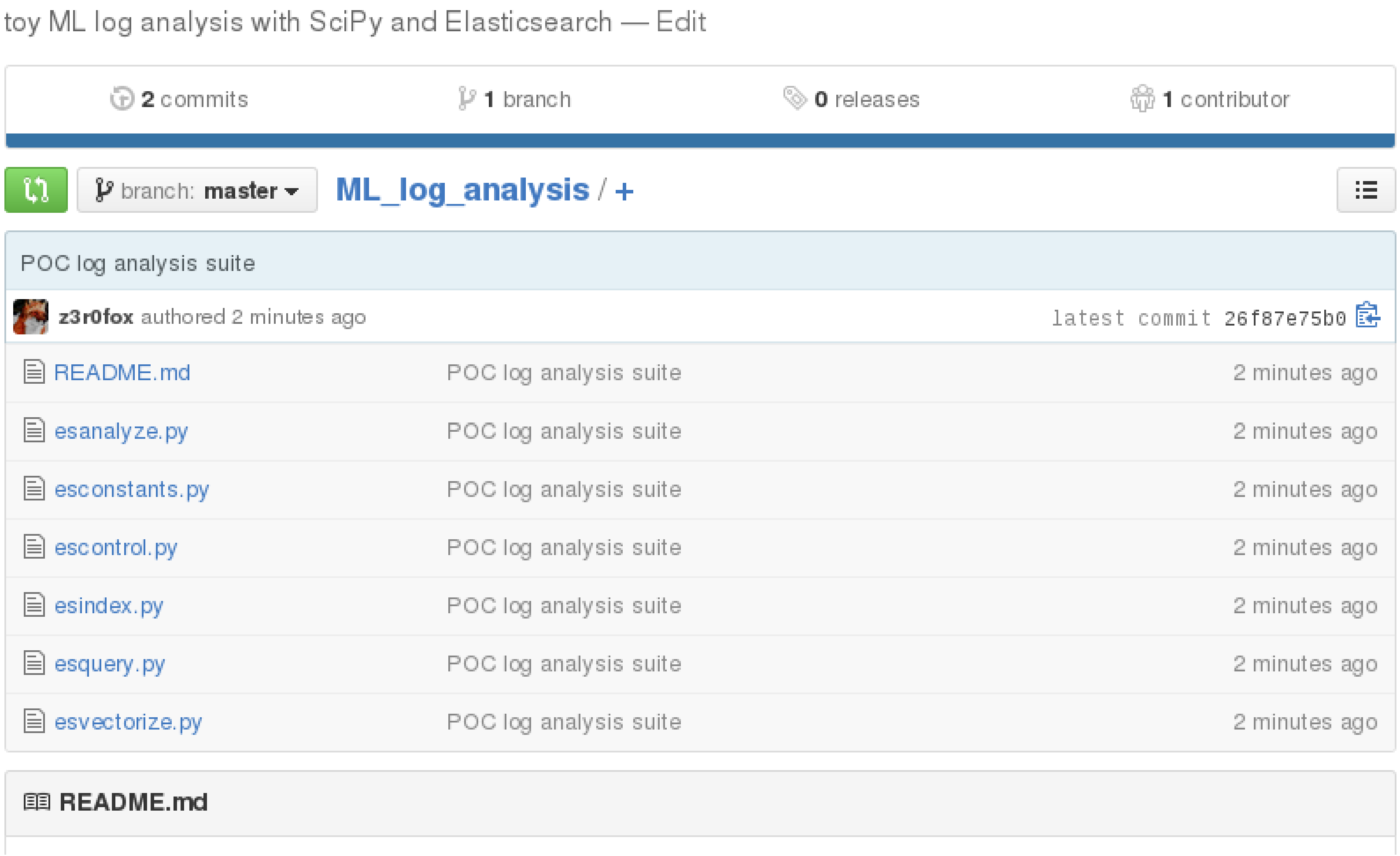
```
[redacted]$ ./escontrol.py
http://[redacted]9200/logstash*/_search -d {"query": {"filtered": {"query": {"query_string": {"query": "*"}}, "filter": {"range": {"@timestamp": {"lte": "2015-06-19", "g
te": "2015-06-19"}}}}}, "size": 1000000}
http://[redacted]9200/logstash*/_search -d {"query": {"filtered": {"query": {"query_string": {"query": "*"}}, "filter": {"range": {"@timestamp": {"lte": "2015-06-18", "g
te": "2015-06-18"}}}}}, "size": 10000}
Got day: 2015-06-18
http://[redacted]:9200/logstash*/_search -d {"query": {"filtered": {"query": {"query_string": {"query": "*"}}, "filter": {"range": {"@timestamp": {"lte": "2015-06-17", "g
te": "2015-06-17"}}}}}, "size": 10000}
Got day: 2015-06-17
http://[redacted]:9200/logstash*/_search -d {"query": {"filtered": {"query": {"query_string": {"query": "*"}}, "filter": {"range": {"@timestamp": {"lte": "2015-06-16", "g
te": "2015-06-16"}}}}}, "size": 10000}
Got day: 2015-06-16
http://[redacted]9200/logstash*/_search -d {"query": {"filtered": {"query": {"query_string": {"query": "*"}}, "filter": {"range": {"@timestamp": {"lte": "2015-06-15", "g
te": "2015-06-15"}}}}}, "size": 10000}
Got day: 2015-06-15
http://[redacted]:9200/logstash*/_search -d {"query": {"filtered": {"query": {"query_string": {"query": "*"}}, "filter": {"range": {"@timestamp": {"lte": "2015-06-14", "g
te": "2015-06-14"}}}}}, "size": 10000}
Got day: 2015-06-14
http://[redacted]9200/logstash*/_search -d {"query": {"filtered": {"query": {"query_string": {"query": "*"}}, "filter": {"range": {"@timestamp": {"lte": "2015-06-13", "g
te": "2015-06-13"}}}}}, "size": 10000}
Got day: 2015-06-13
http://[redacted]9200/logstash*/_search -d {"query": {"filtered": {"query": {"query_string": {"query": "*"}}, "filter": {"range": {"@timestamp": {"lte": "2015-06-12", "g
te": "2015-06-12"}}}}}, "size": 10000}
Got day: 2015-06-12
http://[redacted]9200/logstash*/_search -d {"query": {"filtered": {"query": {"query_string": {"query": "*"}}, "filter": {"range": {"@timestamp": {"lte": "2015-06-11", "g
te": "2015-06-11"}}}}}, "size": 10000}
Dropped day: 2015-06-11
http:[redacted]9200/logstash*/_search -d {"query": {"filtered": {"query": {"query_string": {"query": "*"}}, "filter": {"range": {"@timestamp": {"lte": "2015-06-10", "g
te": "2015-06-10"}}}}}, "size": 10000}
Dropped day: 2015-06-10
http://[redacted]:9200/logstash*/_search -d {"query": {"filtered": {"query": {"query_string": {"query": "*"}}, "filter": {"range": {"@timestamp": {"lte": "2015-06-09", "g
te": "2015-06-09"}}}}}, "size": 10000}
Dropped day: 2015-06-09
http:[redacted]:9200/logstash*/_search -d {"query": {"filtered": {"query": {"query_string": {"query": "*"}}, "filter": {"range": {"@timestamp": {"lte": "2015-06-08", "g
te": "2015-06-08"}}}}}, "size": 10000}
Dropped day: 2015-06-08
http:[redacted]9200/logstash*/_search -d {"query": {"filtered": {"query": {"query_string": {"query": "*"}}, "filter": {"range": {"@timestamp": {"lte": "2015-06-07", "g
te": "2015-06-07"}}}}}, "size": 10000}
Dropped day: 2015-06-07
http:[redacted]9200/logstash*/_search -d {"query": {"filtered": {"query": {"query_string": {"query": "*"}}, "filter": {"range": {"@timestamp": {"lte": "2015-05-20", "g
te": "2015-05-20"}}}}}, "size": 10000}
Dropped day: 2015-05-20
http:[redacted]9200/logstash*/_search -d {"query": {"filtered": {"query": {"query_string": {"query": "*"}}, "filter": {"range": {"@timestamp": {"lte": "2015-05-19", "g
te": "2015-05-19"}}}}}, "size": 10000}
Dropped day: 2015-05-19

WARNING Subtype 2210044 doesn't exist in training data for 2210044
*** Sending this as an alert to SIEM

OUTLIER RESULTS found for 2015-06-19 by type, subtype in the following hours:
{'suricata': {'2210044': [6, 21]}}

Results of reindexing outlying records
---------------------------------------
A total of 1 records were updated and saved as file 'records2015-06-19'
```

Dashboard View – Post Analysis Jobs



Please Enjoy Responsibly

https://github.com/z3r0fox/ML_log_analysis



toy ML log analysis with SciPy and Elasticsearch — Edit

POC log analysis suite

z3r0fox authored 2 minutes ago latest commit 26f87e75b0

README.md	POC log analysis suite	2 minutes ago
esanalyze.py	POC log analysis suite	2 minutes ago
esconstants.py	POC log analysis suite	2 minutes ago
escontrol.py	POC log analysis suite	2 minutes ago
esindex.py	POC log analysis suite	2 minutes ago
esquery.py	POC log analysis suite	2 minutes ago
esvectorize.py	POC log analysis suite	2 minutes ago

README.md

Interesting Reads

- Wikipedia: Information Design
  https://en.wikipedia.org/wiki/Information_design
- Stephen Few, Why Do We Visualize Quantitative Data?
  http://www.perceptualedge.com/blog/?p=1897
- Managing Logstash with the Redis Client
  http://www.nightbluefruit.com/blog/2014/03/managing-logstash-with-the-redis-client/
- An interactive introduction to FFT
  http://betterexplained.com/articles/an-interactive-guide-to-the-fourier-transform/
- Data Driven Security (Jay Jacobs & Bob Rudis)
  BONUS: Pan-fried gnocchi
  http://datadrivensecurity.info/

Thank You

@z3r0fox