ELK: a log files management framework

Giovanni Bechis
<g.bechis@snb.it>

LinuxCon Europe 2016

About Me

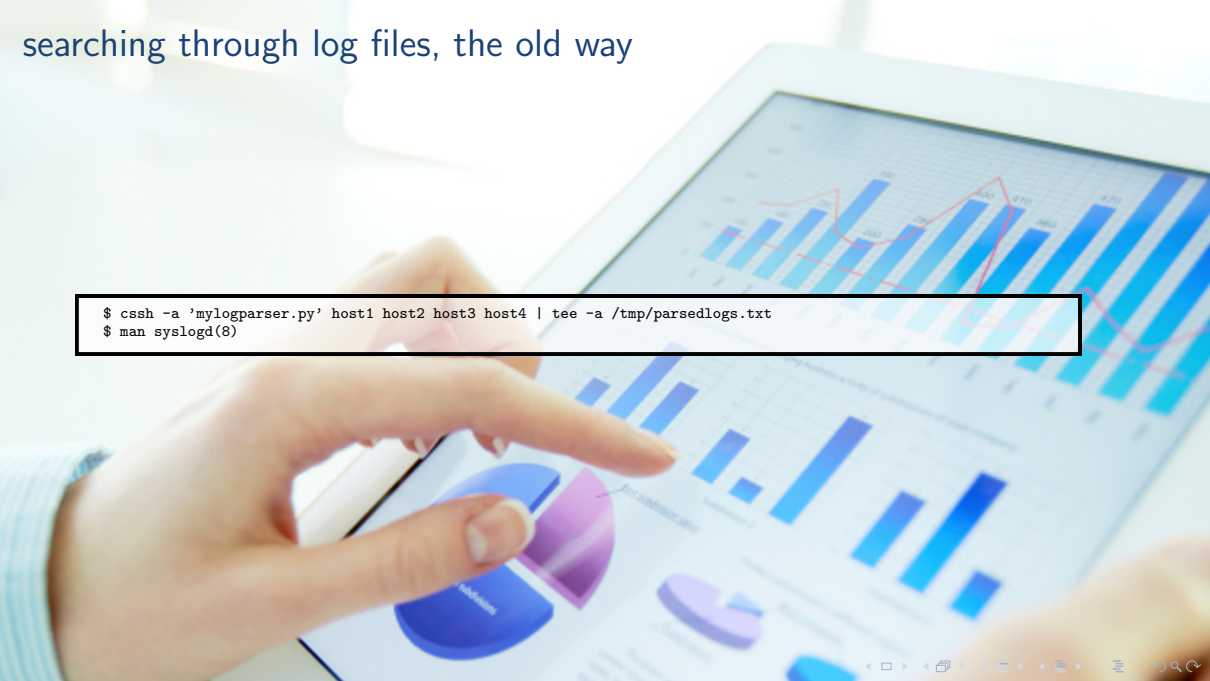
- sys admin and developer @SNB
- OpenBSD developer
- Open Source developer in several other projects

searching through log files, the old way

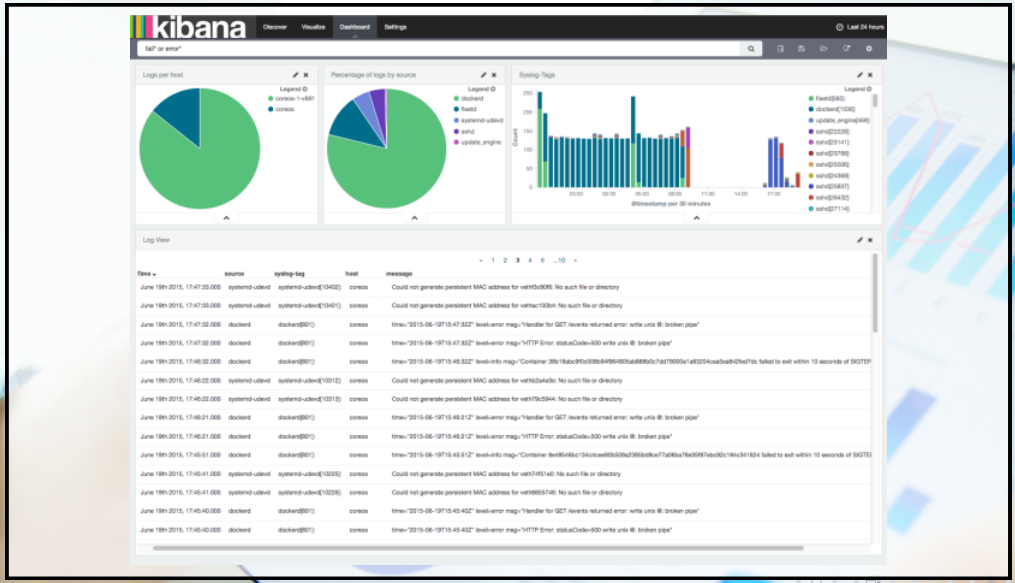
```
$ man 1 pflogsumm
$ grep spammer@baddomain.com /var/log/maillog | awk '{print $1 "-" $2 "-" $3;}'
$ grep -e 'from=.*@gmail\.com' /var/log/maillog | grep "550" \
| awk {'print $1 "-" $2 "-" $3 " " $7 " " $10 " " $11 " " $13;}'
$ vi logparser.sh

$ git clone https://github.com/random/parser_that_should_work

$ man 1 perltoc
$ man 1 python
```

searching through log files, the old way

```
$ cssh -a 'mylogparser.py' host1 host2 host3 host4 | tee -a /tmp/parsedlogs.txt
$ man syslogd(8)
```



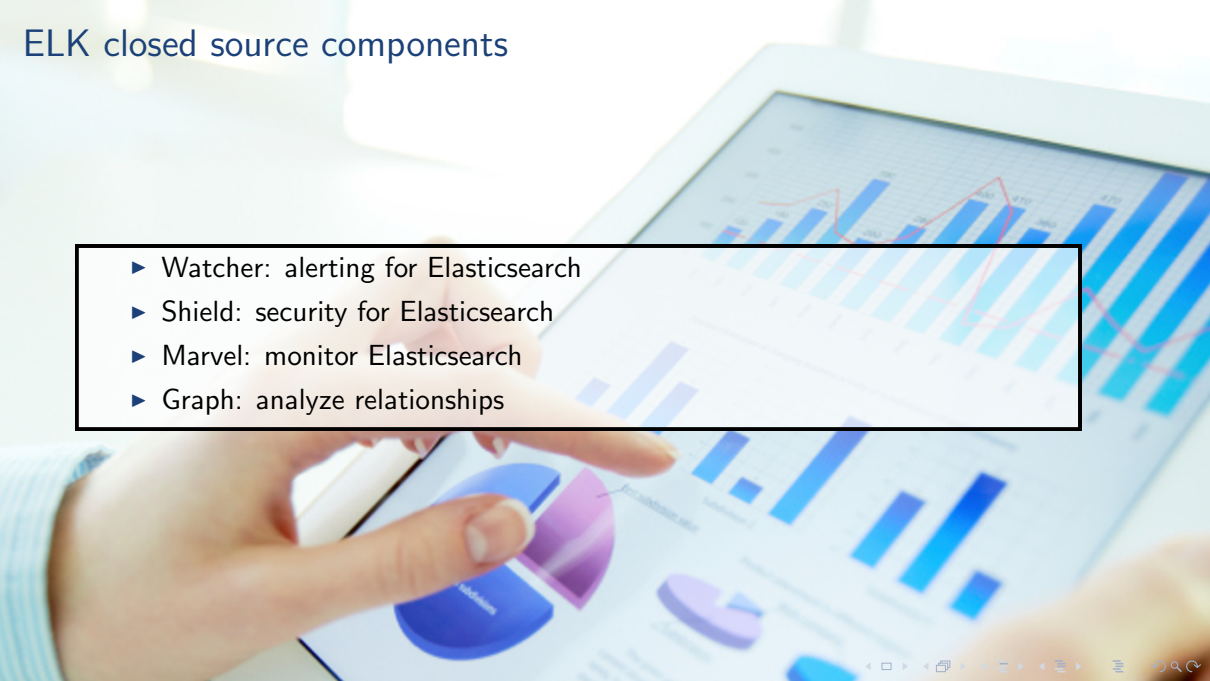
searching through log files, the new way



ELK open source components

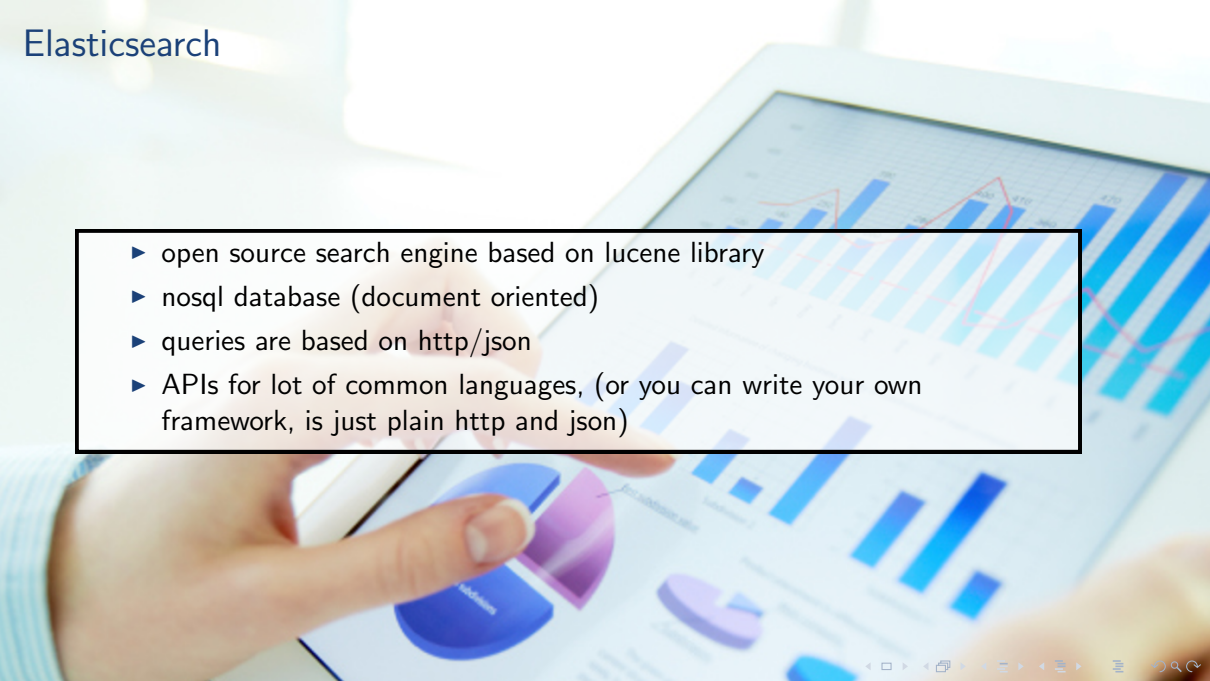
- Beats: collect, parse and ship
- Logstash: collect, enrich and transport data
- Elasticsearch: search and analyze data in real time
- Kibana: explore and visualize your data

ELK closed source components



- Watcher: alerting for Elasticsearch
- Shield: security for Elasticsearch
- Marvel: monitor Elasticsearch
- Graph: analyze relationships

Elasticsearch



- open source search engine based on lucene library
- nosql database (document oriented)
- queries are based on http/json
- APIs for lot of common languages, (or you can write your own framework, is just plain http and json)

Elasticsearch: security

- not available in open source version, you need Shield
- Elasticsearch should not be exposed on the wild, use firewalling to protect your instances
- manage security on your software, not in your backend (Elasticsearch)
- use .htaccess files to protect your Kibana instance

Managing Elasticsearch: backups

- backup with snapshots

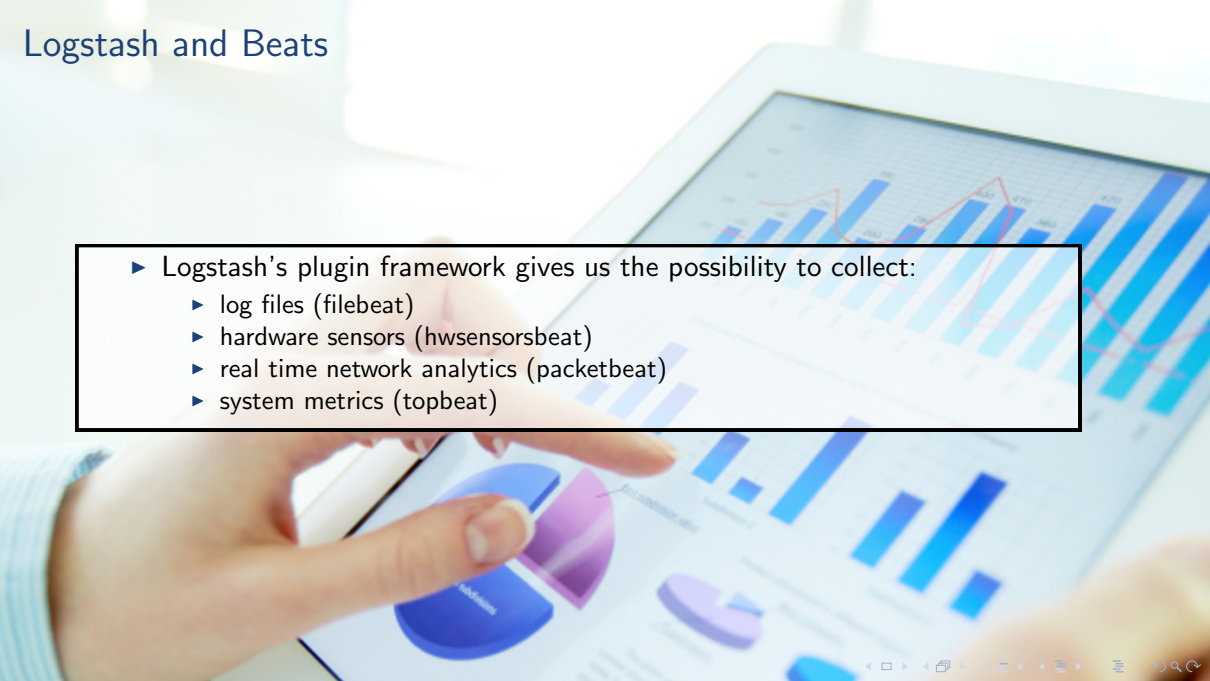
```
curl -XPUT "http://localhost:9200/_snapshot/es_backup" -d '{
   "type": "fs",
   "settings": {
       "location": "/mnt/backup/es",
       "compress": true
   }
}'

SNAP=$(date "+%Y-%m-%d")
/bin/curl -XPUT "http://localhost:9200/_snapshot/es_backup/snapshot_$SNAP"
```

- "curator" to manage indices and snapshots, actions set with a yaml config file

Logstash and Beats

- log files collector, "beats" reads log files and send them over the network to Logstash which parses and saves them in Elasticsearch
- grok and ruby based parser
- possibility to use redis to accelerate processing

Logstash and Beats



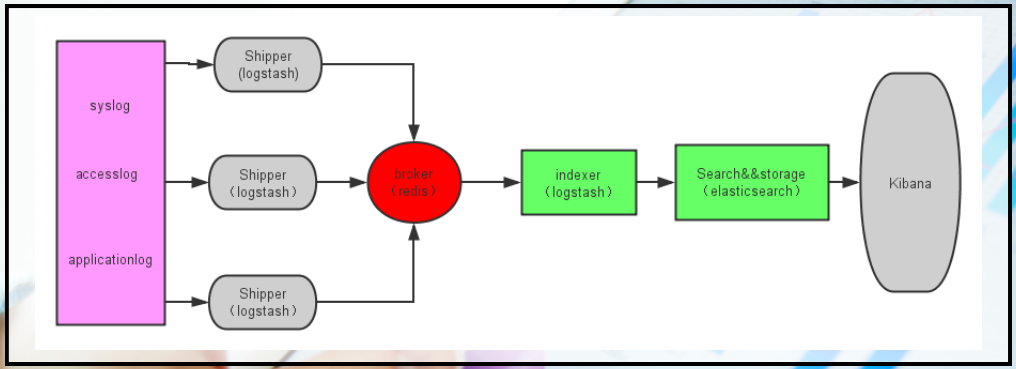
- Logstash's plugin framework gives us the possibility to collect:
  - log files (filebeat)
  - hardware sensors (hwsensorsbeat)
  - real time network analytics (packetbeat)
  - system metrics (topbeat)

Logstash and Beats

other plugins available:

- drupal_dblog
- exec
- (Windows) eventlog
- github (webhook)
- imap
- jdbc
- puppet_facter
- salesforce
- snmptrap
- twitter
- varnishlog

ELK flow


syslog
accesslog
applicationlog
Shipper (logstash)
Shipper (logstash)
Shipper (logstash)
broker (redis)
indexer (logstash)
Search&&storage (elasticsearch)
Kibana

filebeat.yml

```
filebeat:
  prospectors:
    -
      paths:
        - "/var/log/maillog"
      document_type: postfix
    -
      paths:
        - "/var/www/*/log/access.log"
      document_type: apache

  registry_file: /var/lib/filebeat/registry

output:

  logstash:
    # The Logstash hosts
    hosts: ["10.0.0.203:5001"]
```

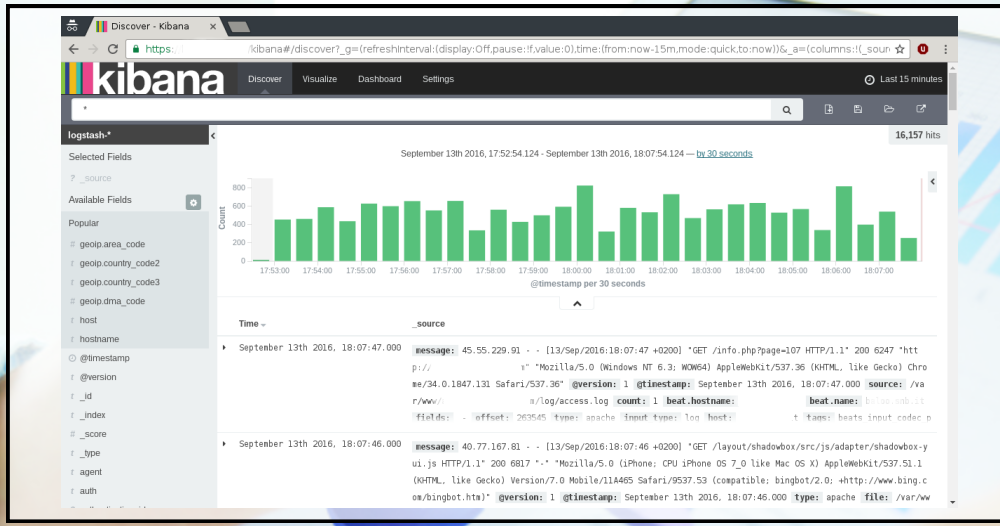
logstash.conf

```
input {
  beats {
    port => 5001
    type => "logs"
  }
}

filter {
  if [type] == "syslog" {
    grok {
      match => { "message" => "%{SYSLOGTIMESTAMP:syslog_timestamp} %{SYSLOGHOST:syslog_hostname} \
%{DATA:syslog_program}(?:\[%{POSINT:syslog_pid}\])?: %{GREEDYDATA:syslog_message}" }
      add_field => [ "received_at", "%{@timestamp}" ]
      add_field => [ "received_from", "%{host}" ]
    }
    syslog_pri { }
    date {
      match => [ "syslog_timestamp", "MMM  d HH:mm:ss", "MMM dd HH:mm:ss" ]
    }
  }
}

output {
  elasticsearch {
    hosts => ["127.0.0.1:9200"]
  }
  stdout { codec => rubydebug }
}
```

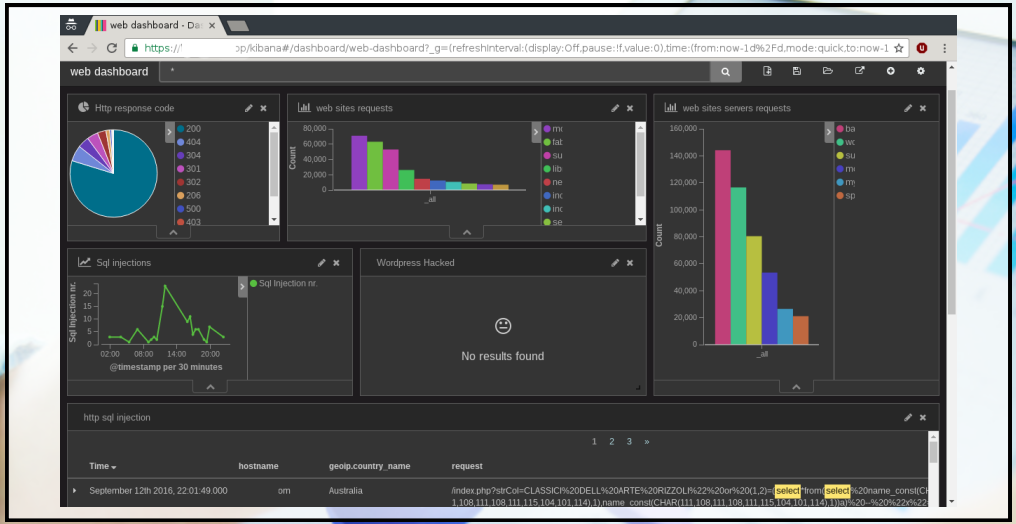
logstash.conf - filters

```
filter {
  if [type] == "postfix" {
...
    if [message] =~ /=/ {
      kv { source => "message" trim => "<>," }
    }
    grok {
      match => [ "message", "Accepted authentication for user %{DATA:sasl_username} on session" ]
    }
    geoip {
        source => "[ip]"
        add_field => [ "[geoip][location]", "%{[geoip][longitude]}" ]
        add_field => [ "[geoip][location]", "%{[geoip][latitude]}" ]
    }
    ruby {
      code => "
        event.to_hash.keys.each { |k|
          if k.start_with?('<')
            event.remove(k)
          end
        }
      "
    }
    mutate { remove_field => [ "_syslog_payload" ] }
  }
  de_dot {
  }
}
```

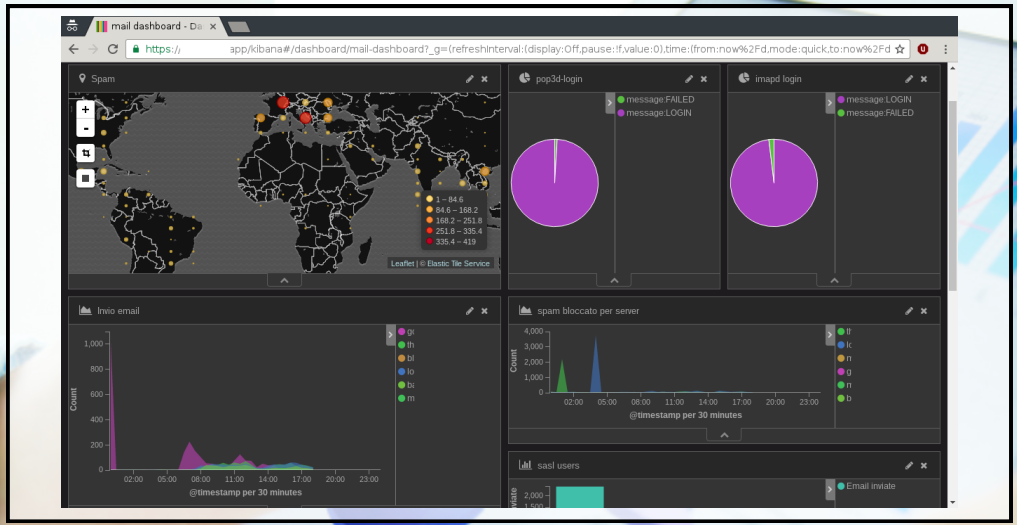
Kibana


Discover - Kibana
kibana
Discover
Visualize
Dashboard
Settings
Last 15 minutes
logstash-*
Selected Fields
_source
Available Fields
Popular
geoip.area_code
geoip.country_code2
geoip.country_code3
geoip.dma_code
host
hostname
@timestamp
@version
_id
_index
_score
_type
agent
auth
16,157 hits
September 13th 2016, 17:52:54.124 - September 13th 2016, 18:07:54.124 — by 30 seconds
Count
800
600
400
200
0
17:53:00
17:54:00
17:55:00
17:56:00
17:57:00
17:58:00
17:59:00
18:00:00
18:01:00
18:02:00
18:03:00
18:04:00
18:05:00
18:06:00
18:07:00
@timestamp per 30 seconds
Time
_source
September 13th 2016, 18:07:47.000
message: 45.55.229.91 - - [13/Sep/2016:18:07:47 +0200] "GET /info.php?page=107 HTTP/1.1" 200 6247 "http://
" "Mozilla/5.0 (Windows NT 6.3; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/34.0.1847.131 Safari/537.36" @version: 1 @timestamp: September 13th 2016, 18:07:47.000 source: /var/ww
a/log/access.log count: 1 beat.hostname: beat.name:
fields: - offset: 263545 type: apache input_type: log host: t tags: beats_input_codec_p
September 13th 2016, 18:07:46.000
message: 40.77.167.81 - - [13/Sep/2016:18:07:46 +0200] "GET /layout/shadowbox/src/js/adapter/shadowbox-yui.js HTTP/1.1" 200 6817 "-" "Mozilla/5.0 (iPhone; CPU iPhone OS 7_0 like Mac OS X) AppleWebKit/537.51.1 (KHTML, like Gecko) Version/7.0 Mobile/11A465 Safari/9537.53 (compatible; bingbot/2.0; +http://www.bing.com/bingbot.htm)" @version: 1 @timestamp: September 13th 2016, 18:07:46.000 type: apache file: /var/ww

Kibana

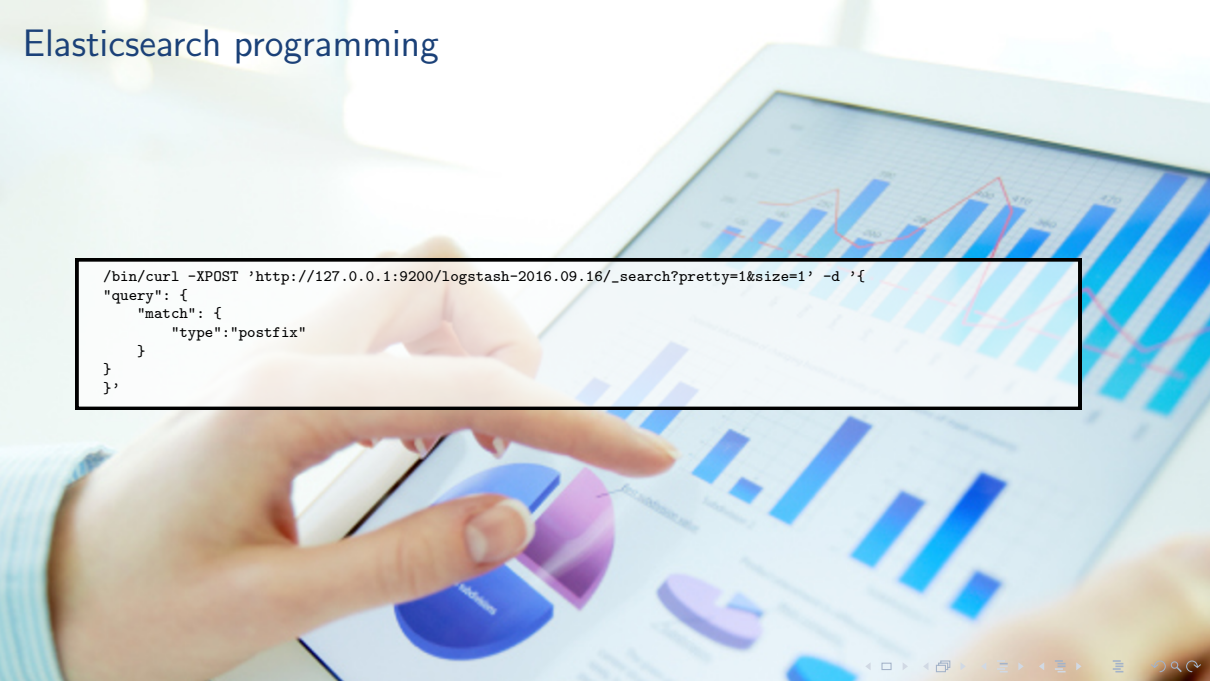

web dashboard - Da
https://
web dashboard
Http response code
200
404
304
301
302
206
500
403
web sites requests
Count
80,000
60,000
40,000
20,000
0
_all
web sites servers requests
160,000
140,000
120,000
100,000
80,000
60,000
40,000
20,000
0
Sql injections
Sql Injection nr.
Sql injection nr.
20
15
10
5
0
02:00
08:00
14:00
20:00
@timestamp per 30 minutes
Wordpress Hacked
No results found
http sql injection
1 2 3 »
Time
hostname
geoip.country_name
request
September 12th 2016, 22:01:49.000
Australia
/index.php?strCol=CLASSICI%20DELL%20ARTE%20RIZZOLI%22%20or%20(1,2)=(select from(select %20name_const(CH
1,108,111,108,111,115,104,101,114),1),name_const(CHAR(111,108,111,108,111,115,104,101,114),1))a)%20--%20%22x%22

Kibana


mail dashboard - Da
https://
app/kibana#/dashboard/mail-dashboard?_g=(refreshInterval:(display:Off,pause:!f,value:0),time:(from:now%2Fd,mode:quick,to:now%2Fd
Spam
1 – 84.6
84.6 – 168.2
168.2 – 251.8
251.8 – 335.4
335.4 – 419
Leaflet | © Elastic Tile Service
pop3d-login
message:FAILED
message:LOGIN
imapd login
message:LOGIN
message:FAILED
Invio email
Count
@timestamp per 30 minutes
spam bloccato per server
Count
@timestamp per 30 minutes
sasl users
Email inviate

Elasticsearch programming

```
/bin/curl -XPOST 'http://127.0.0.1:9200/logstash-2016.09.16/_search?pretty=1&size=1' -d '{
"query": {
    "match": {
        "type":"postfix"
    }
}
}'
```



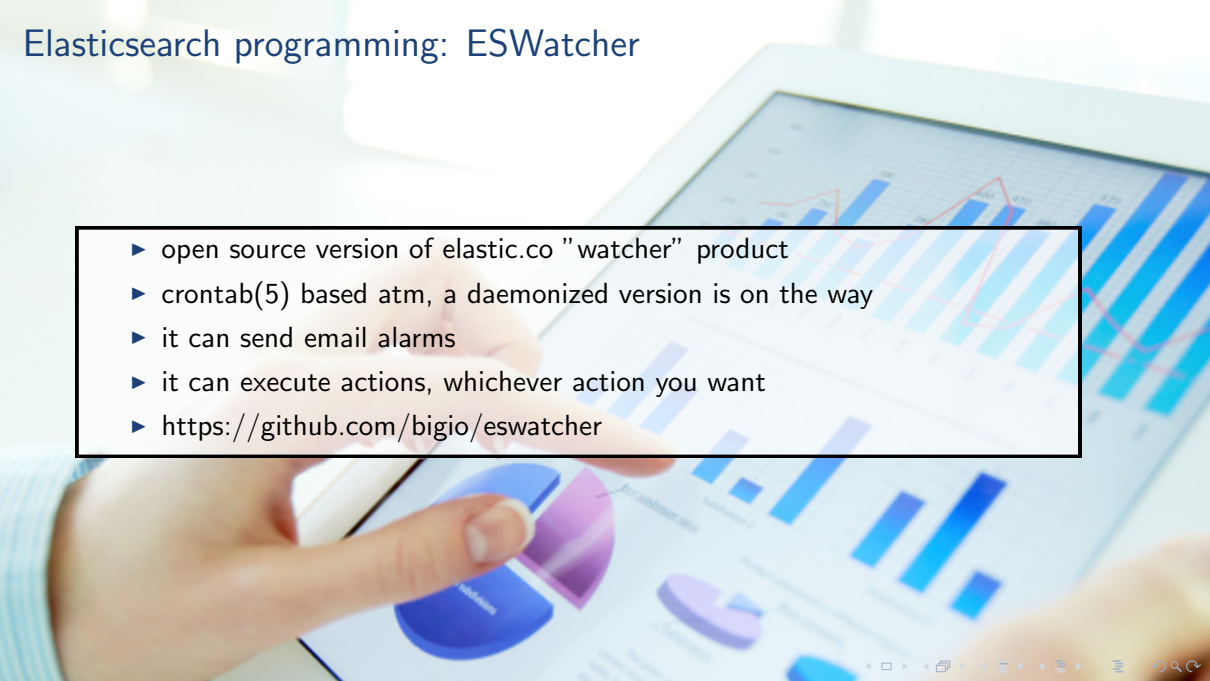
Elasticsearch programming


```
{
  "took" : 10,
  "timed_out" : false,
  "_shards" : {
    "total" : 5,
    "successful" : 5,
    "failed" : 0
  },
  "hits" : {
    "total" : 540467,
    "max_score" : 1.3722948,
    "hits" : [ {
      "_index" : "logstash-2016.09.16",
      "_type" : "postfix",
      "_id" : "AVcxC6_ujEbIPCEOvhkb",
      "_score" : 1.3722948,
      "_source" : {
        "message" : "Sep 16 05:30:22 srv postfix/smtpd[7815]: lost connection after AUTH from client.host.com[97.64.239.154]",
        "@version" : "1",
        "@timestamp" : "2016-09-16T03:30:22.000Z",
        "type" : "postfix",
        "file" : "/var/log/maillog",
        "host" : "srv.domain.tld",
        "program" : "postfix/smtpd",
        "tags" : [ "_grokparsefailure" ],
        "geoip" : {
          "ip" : "97.64.239.154",
          "country_code2" : "US",
          "country_name" : "United States",
          "latitude" : 41.1987,
          "longitude" : -90.7219,
          [...]
        }
      }
    } ]
  }
}
```

Elasticsearch programming

```
use Search::Elasticsearch;

# Connect to localhost:9200:
my $e = Search::Elasticsearch->new();

my $results = $e->search(
    index => 'my_app',
    body  => {
        query => {
            match => { title => 'LinuxCon' }
        }
    }
);
```

Elasticsearch programming: ESWatcher



- open source version of elastic.co "watcher" product
- crontab(5) based atm, a daemonized version is on the way
- it can send email alarms
- it can execute actions, whichever action you want
- https://github.com/bigio/eswatcher

Questions ?