Pourquoi faire ?

Pourquoi faire ?

Consolidation !

ELK

- Se compose de 3 produits différents
  - Elasticsearch: base NoSQL
  - Logstash: Parser de log
  - Kibana: Interface web

Elasticsearch

- L'entrepot de donnée d'ELK
  - Base NoSQL
  - Clusterisable
  - Optimiser pour la recherche de donnée

Logstash

- Permet de structurer les logs et les injecter dans Elasticsearch
  - Fonctionne sous le principe de plugin
    - Input (syslog / lumberjack)
    - Filter (grok)
    - Output (elasticsearch)

Logstash

- Permet de récupérer les logs
  - Syslog
  - Lumberjack
- Quelque soit la source
  - Logstash-forwarder

Logstash

- Permet de tagger les logs
  - Basé sur des expressions régulières
  - Peux utiliser des filtres externes pour rajouter de l'information (géolocalisation)

Logstash

- Injecte les informations dans Elasticsearch
  - Mais pas que !
    - Permet d'envoyer des mails d'alertes
    - Sur le principe similaire aux tags

QUELQUES EXEMPLES DE CONFIGURATION

Accepter les connexions syslog

```
input {
  tcp {
    port => 5000
    type => syslog
  }
  udp {
    port => 5000
    type => syslog
  }
}
```

Envoyer les fichiers apache a logstash

```
[root@syslog ~]# cat /etc/logstash-forwarder.conf
{
  "network": {
    "servers": [ "syslog.lal.in2p3.fr:5043" ],
    "ssl ca": "/etc/pki/tls/certs/logstash-forwarder.crt",
    "timeout": 15
  },

  "files": [
    {
      "paths": [
        "/var/log/httpd/*.log",
        "/var/log/httpd/*.err"
      ],
      "fields": { "type": "apache" }
    }
  ]
}
```

Ecrire ses filtres

```
grok {
  match => { "message[1]" => "Failed password for invalid user %{USERNAME:username} from %{IP:src_ip} port %{BASE10NUM:port} ssh2" }
  add_tag => [ "invalid_user", "failed", "auth" ]
  tag_on_failure => []
}
```

Utiliser les variables

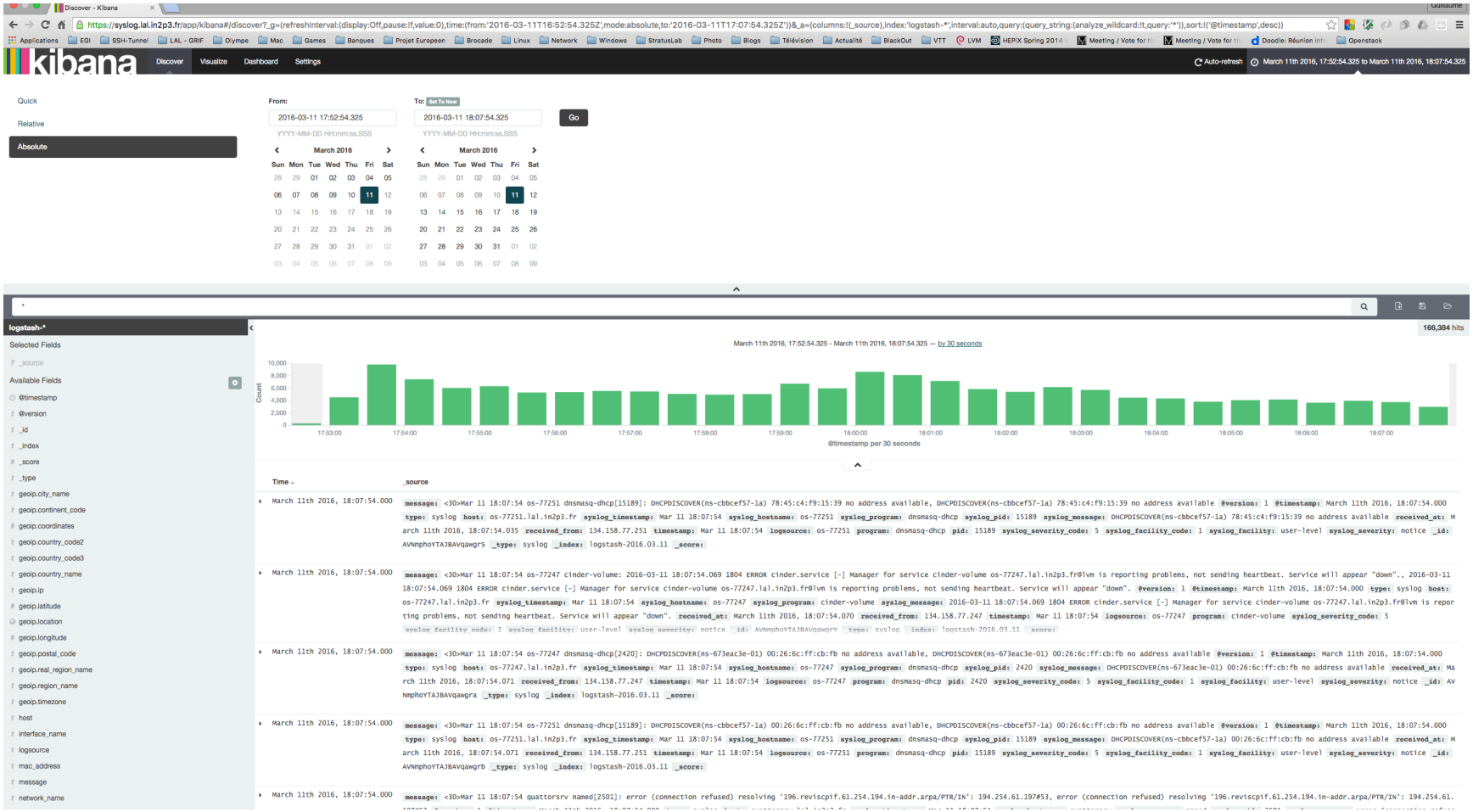
```
if [sender] {
  mutate {
    lowercase => [ "sender" ]
  }
  if "outcomming" in [tags] {
    grok {
      match => { "sender" => "%{USERNAME:username}@.*" }
      tag_on_failure => []
    }
  }
}
```

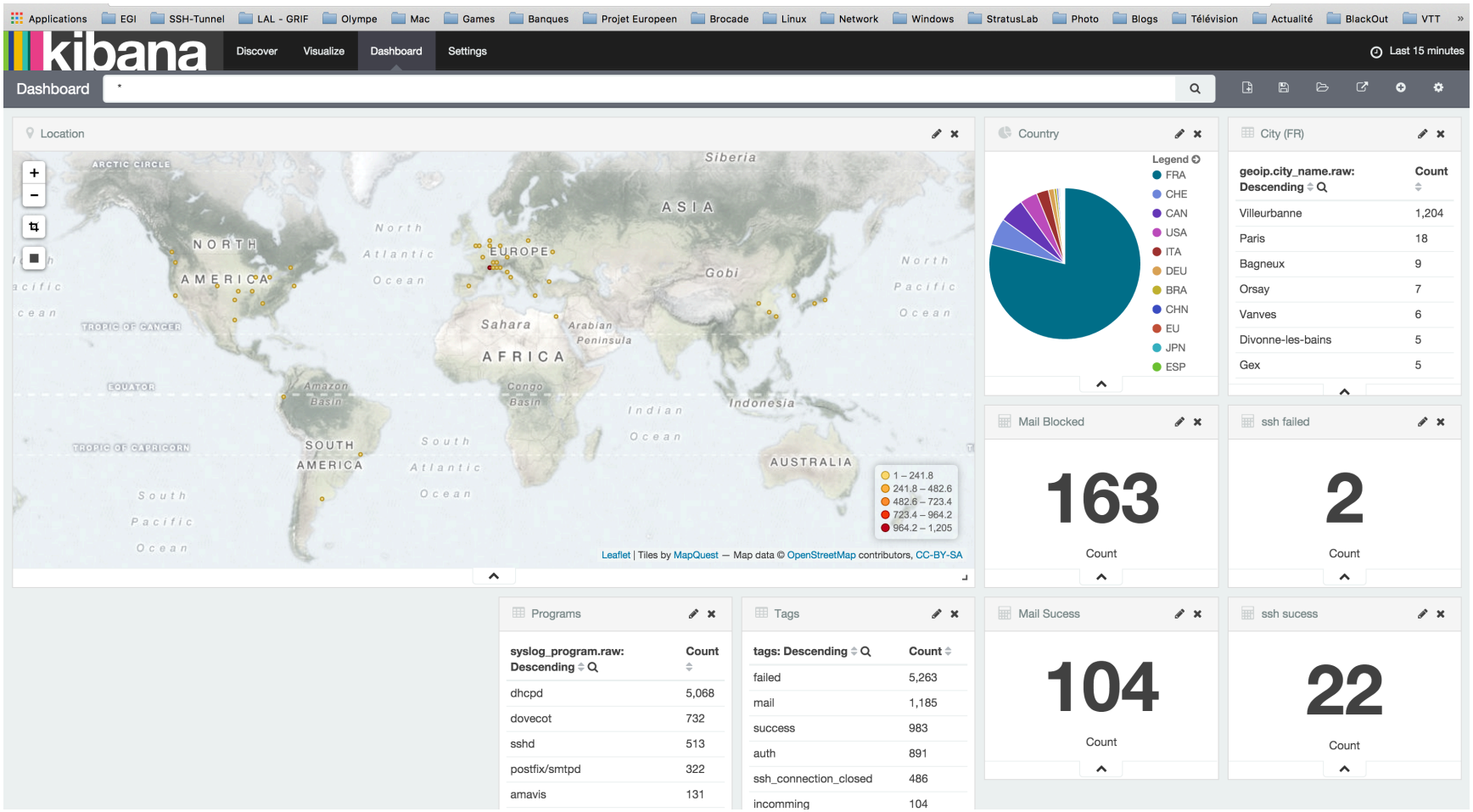
Geolocaliser une IP

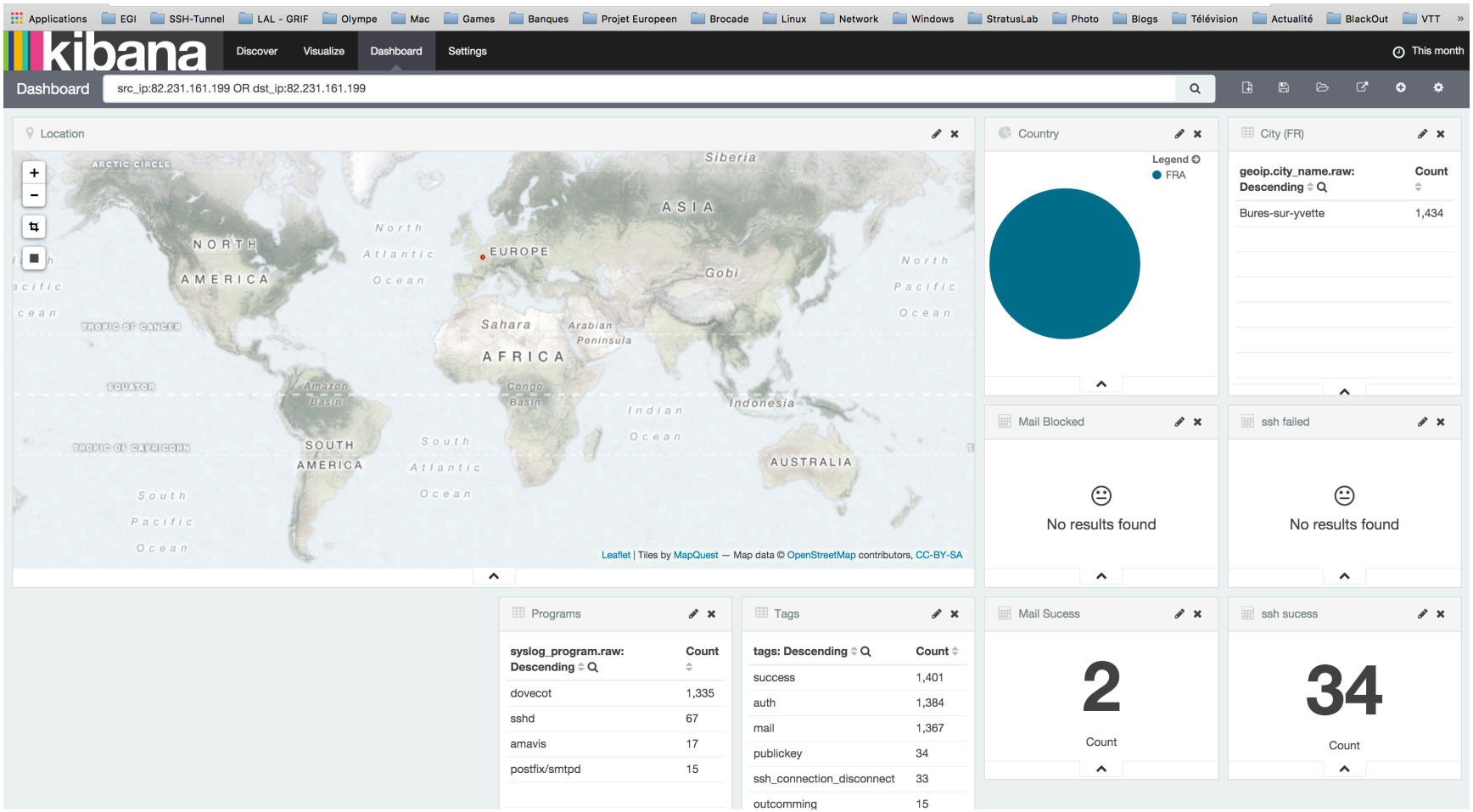
```
if [src_ip]  {
  geoip {
    source => "src_ip"
    target => "geoip"
    database => "/opt/logstash/vendor/geoip/GeoLiteCity.dat"
    add_field => [ "[geoip][coordinates]", "%{[geoip][longitude]}" ]
    add_field => [ "[geoip][coordinates]", "%{[geoip][latitude]}"  ]
  }
  mutate {
    convert => [ "[geoip][coordinates]", "float" ]
  }
}
```

Kibana

- Interface (web) d'affichage d'Elasticsearch
  - Permet de faire des requêtes de recherche complexe plain text ou basé sur les index
  - Permet de réaliser des dashboard pour afficher les informations de façon synthétique


Discover - Kibana
https://syslog.lal.in2p3.fr/app/kibana#/discover?_g=(refreshInterval:(display:Off,pause:!f,value:0),time:(from:'2016-03-11T16:52:54.325Z',mode:absolute,to:'2016-03-11T17:07:54.325Z'))&_a=(columns:!(_source),index:'logstash-*',interval:auto,query:(query_string:(analyze_wildcard:!t,query:'*')),sort:!('@timestamp',desc))
kibana
Discover
Visualize
Dashboard
Settings
Auto-refresh
March 11th 2016, 17:52:54.325 to March 11th 2016, 18:07:54.325
Quick
Relative
Absolute
From:
2016-03-11 17:52:54.325
To:
2016-03-11 18:07:54.325
Go
March 2016
logstash-*
Selected Fields
Available Fields
@timestamp
@version
_id
_index
_score
_type
geoip.city_name
geoip.continent_code
geoip.coordinates
geoip.country_code2
geoip.country_code3
geoip.country_name
geoip.ip
geoip.latitude
geoip.location
geoip.longitude
geoip.postal_code
geoip.real_region_name
geoip.region_name
geoip.timezone
host
interface_name
logsource
mac_address
message
network_name
166,384 hits
March 11th 2016, 17:52:54.325 - March 11th 2016, 18:07:54.325 — by 30 seconds
Count
@timestamp per 30 seconds
Time
_source
March 11th 2016, 18:07:54.000
March 11th 2016, 18:07:54.000
March 11th 2016, 18:07:54.000
March 11th 2016, 18:07:54.000
March 11th 2016, 18:07:54.000


Applications
EGI
SSH-Tunnel
LAL - GRIF
Olympe
Mac
Games
Banques
Projet Europeen
Brocade
Linux
Network
Windows
StratusLab
Photo
Blogs
Télévision
Actualité
BlackOut
VTT
kibana
Discover
Visualize
Dashboard
Settings
Last 15 minutes
Dashboard
Location
ARCTIC CIRCLE
Siberia
ASIA
NORTH AMERICA
North Atlantic Ocean
EUROPE
Gobi
North Pacific Ocean
TROPIC OF CANCER
Sahara
Arabian Peninsula
AFRICA
EQUATOR
Amazon Basin
Congo Basin
Indian Ocean
Indonesia
TROPIC OF CAPRICORN
SOUTH AMERICA
South Atlantic Ocean
AUSTRALIA
South Pacific Ocean
1 – 241.8
241.8 – 482.6
482.6 – 723.4
723.4 – 964.2
964.2 – 1,205
Leaflet | Tiles by MapQuest — Map data © OpenStreetMap contributors, CC-BY-SA
Country
Legend
FRA
CHE
CAN
USA
ITA
DEU
BRA
CHN
EU
JPN
ESP
City (FR)
geoip.city_name.raw: Descending
Count
Villeurbanne 1,204
Paris 18
Bagneux 9
Orsay 7
Vanves 6
Divonne-les-bains 5
Gex 5
Mail Blocked
163
Count
ssh failed
2
Count
Programs
syslog_program.raw: Descending
Count
dhcpd 5,068
dovecot 732
sshd 513
postfix/smtpd 322
amavis 131
Tags
tags: Descending
Count
failed 5,263
mail 1,185
success 983
auth 891
ssh_connection_closed 486
incomming 104
Mail Sucess
104
Count
ssh sucess
22
Count


Applications EGI SSH-Tunnel LAL - GRIF Olympe Mac Games Banques Projet Europeen Brocade Linux Network Windows StratusLab Photo Blogs Télévision Actualité BlackOut VTT
kibana
Discover
Visualize
Dashboard
Settings
This month
Dashboard
src_ip:82.231.161.199 OR dst_ip:82.231.161.199
Location
ARCTIC CIRCLE
Siberia
ASIA
NORTH AMERICA
North Atlantic Ocean
EUROPE
Gobi
North Pacific Ocean
TROPIC OF CANCER
Sahara
Arabian Peninsula
AFRICA
EQUATOR
Amazon Basin
Congo Basin
Indian Ocean
Indonesia
TROPIC OF CAPRICORN
SOUTH AMERICA
South Atlantic Ocean
AUSTRALIA
South Pacific Ocean
Leaflet | Tiles by MapQuest — Map data © OpenStreetMap contributors, CC-BY-SA
Country
Legend
FRA
City (FR)
geoip.city_name.raw: Descending
Count
Bures-sur-yvette 1,434
Mail Blocked
No results found
ssh failed
No results found
Programs
syslog_program.raw: Descending
Count
dovecot 1,335
sshd 67
amavis 17
postfix/smtpd 15
Tags
tags: Descending
Count
success 1,401
auth 1,384
mail 1,367
publickey 34
ssh_connection_disconnect 33
outcomming 15
Mail Sucess
2
Count
ssh sucess
34
Count

Quelques retours

- Elasticsearch
  - Base de donnée qui travaille énormement avec la mémoire de la machine
    - 32 Giga minimum
    - 100 Giga de log/jour
    - Deteste le swap (swappiness, …)
- Logstash
  - 100 règles différentes pour 3 services (mail / ssh / dhcp)
  - Véritable casse-tête organisationnel
    - Toute l'utilisation (corrélation de log, …) découle de la façon dont on tag les logs
    - Sinon on retrouve les mêmes travers de la gestion des logs centralisés

Reflexions en cours

- L'utilité des logs
  - Analyse post-mortem
  - Monitoring de service
  - Monitoring d'utilisation
- Aujourd'hui fait avec 3 outils différents
  - ELK
  - Nagios
  - Ganglia/Cacti