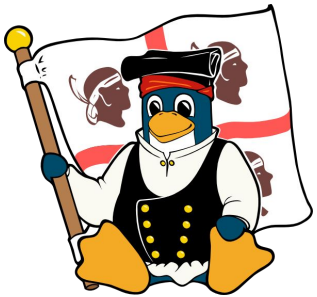
Log management con Logstash/Elasticsearch



Matteo Dessalvi

Linux Day 22 Ottobre 2016

About me

- Sysadmin da circa 12 anni
- ~8 anni presso il Dip. di Fisica (UniCA)
- Attualmente sysadm per il gruppo HPC del GSI Helmholtz Centre for Heavy Ion Research (www.gsi.de)
- Proud member of GULCh since 2000

Sommario

- Manipolare i logs: “the UNIX way”
- Logstash: cosa offre / come usarlo
- Analisi dei logs: Elasticsearch + Kibana
- Vantaggi e svantaggi
- Riferimenti

Logs & Linux/UNIX

Qualcosa non va con il server xyz e/o un servizio non risponde?
Prima cosa da fare: controllare i logs.

Tools tradizionali + UNIX pipes + regular expressions:

```
cat / sed / awk / {e}grep / tail / sort / head /
shell/perl/python/ruby scripts
```

Esempio tipico: “`tail -f /var/log/auth.log | grep denied`”

Logs come file di testo

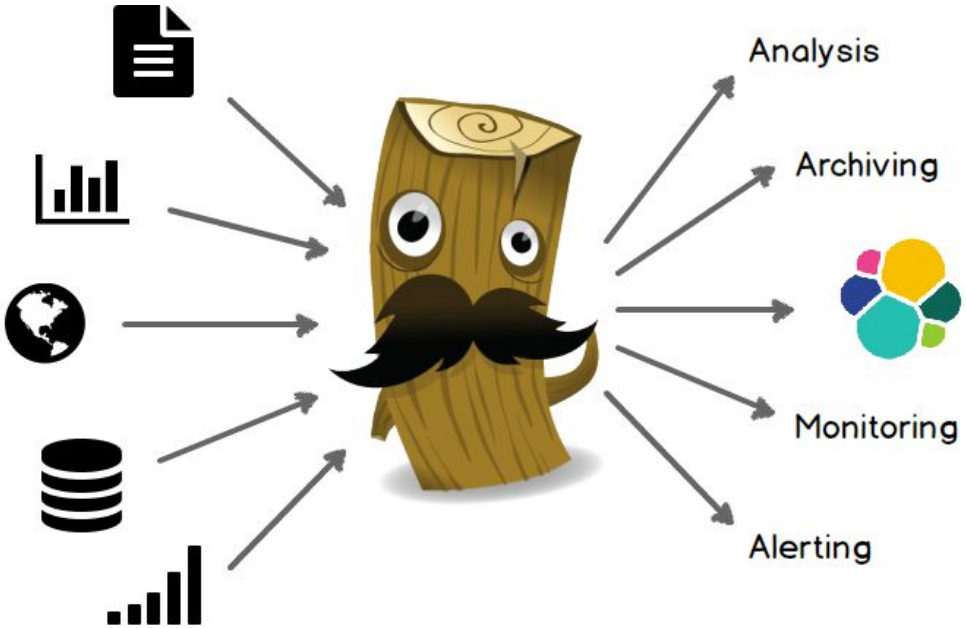
Vantaggi:

- Facilita' di manipolazione
- "Quick & Dirty" ma con risultati quasi immediati

Svantaggi:

- Non sempre e' immediato rispondere alla domanda: "quante volte si e' verificato l'evento X in un determinato arco di tempo?".
- E' complicato avere una visione di insieme, tale da poter rispondere: "gli eventi Y e Z si ripetono con costanza nel tempo".

Logstash



Logstash: caratteristiche

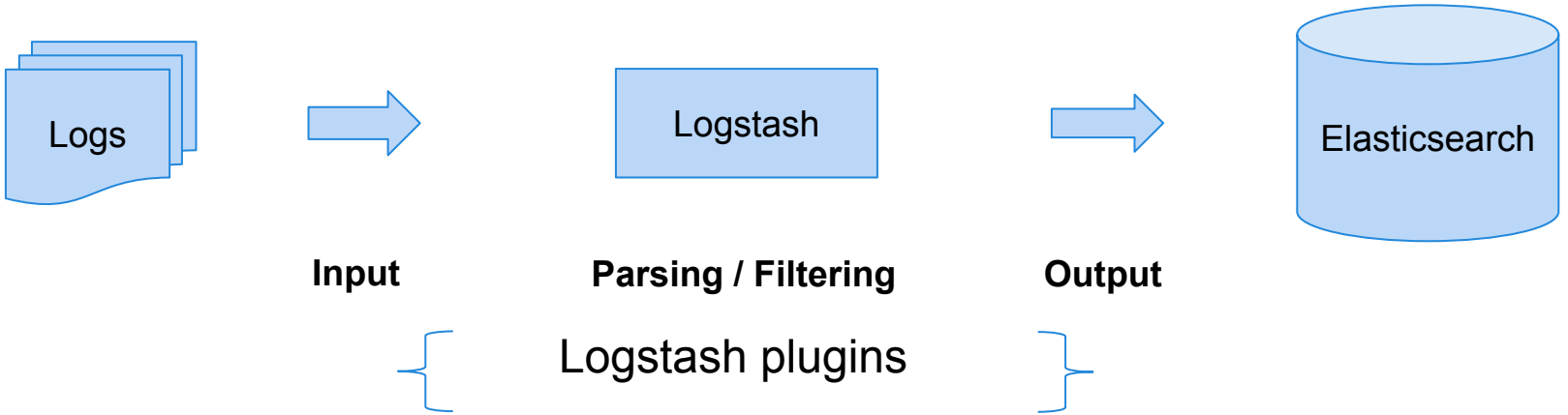
- Logstash e' uno strumento per manipolare logs.
- E' scritto in JRuby.

Consente di creare una '*event pipeline*', divisa in tre stadi:

- `Inputs:  come raccogliere gli eventi`
- `Filters: come manipolarli`
- `Outputs: dove indirizzare i risultati`

Un paragone con le UNIX pipes

```
$ log_producer | grep ...| sed ... |
awk ... | tee output | sort | uniq -c | ...
```



Plugins

Inputs	*Filters*	*Outputs*
file	CSV	Elasticsearch
http	grok	email
log4j	dns	Ganglia
kafka	multiline	Graphite
syslog	JSON	MongoDB
TCP / UDP	kv (key/value)	Nagios
jdbc	mutate	Jira

Logstash ‘hello world’

Un esempio da linea di comando (linea unica):

```
$ logstash -e 'input  { stdin  {} }
               output { stdout {} }'
```

```
hello world
2016-10-09T14:54:14.405+0000 127.0.0.1 hello
world (output)
```

Analizzare SSH logs

```
input {
    file {
            type => "auth-log"
            path => [/var/log/auth.log" ]
            start_position => "beginning"
            codec => plain
          }
}
output { stdout { codec => rubydebug } }
```

Filtrare i logs con Grok

Note: l'espressione seguita da "match" deve essere su una sola linea.

```
filter {
  if [type] == "auth-log" {
      grok {
        match => {
            "message" => "%{SYSLOGBASE} Accepted
             %{WORD:auth_method} for %{USER:username}
             from %{IP:src_ip}
             port %{INT:src_port} ssh2" }
      }
    }
}
```

Grok

- Grok e' il "motore" di filtering di Logstash.
- Consente di mappare una linea di log con una espressione regolare.

```
Esempio: %{PATTERN: FieldName}
```

In questo modo e' possibile mappare specifiche parti del log con un determinato nome.

Elasticsearch

"*You know, for search*"

- Si tratta di un motore di ricerca (open source) basato su Apache Lucene.
- Apache Lucene e' una libreria per motori di ricerca *full-text*.
- Sia Elasticsearch che Lucene sono scritti in Java.
- Elasticsearch consente di indicizzare i logs raccolti e preventivamente analizzati tramite Logstash ed offre funzioni di ricerca tramite REST API.

Kibana

- Kibana e’ una piattaforma di analisi/visualizzazione dati, progettata appositamente per Elasticsearch.

Lo sviluppo ha attraversato diverse fasi:

- Prima versione: applicazione standalone in Ruby
- Seconda versione: Javascript (servita via server web)
- Attualmente: NodeJS (nuovamente standalone

Elasticsearch / Kibana

It’s demo time!

Vantaggi

- Logstash e’ un tool molto versatile: consente di suddividere la pipeline di analisi dei logs in piu’ stadi.
- Plugins di input/output per (quasi) qualunque sorgente.
- Ecosistema in forte sviluppo.

Svantaggi

- Logstash ha un discreto ‘memory footprint’: non e’ indicato distribuirlo su ogni macchina.
- Elasticsearch non e’ un tool per la memorizzazione dei dati su lungo periodo.
- Deployment puo’ essere complicato: e’ consigliato avere uno strato intermedio fra Logstash ed Elasticsearch (Redis or RabbitMQ).

References

- https://www.elastic.co/products/logstash
- https://www.elastic.co/products/elasticsearch
- https://www.elastic.co/products/kibana

- Documentazione: https://www.elastic.co/guide/index.html

- "*The Logstash Book*": Logstash Book (by James Turnbull)

Grok Testing / Beats

Test per espressioni Grok:
https://grokdebug.herokuapp.com/

Logstash troppo pesante? I *Beats* possono aiutare!

https://www.elastic.co/products/beats

Thank you!

Domande?