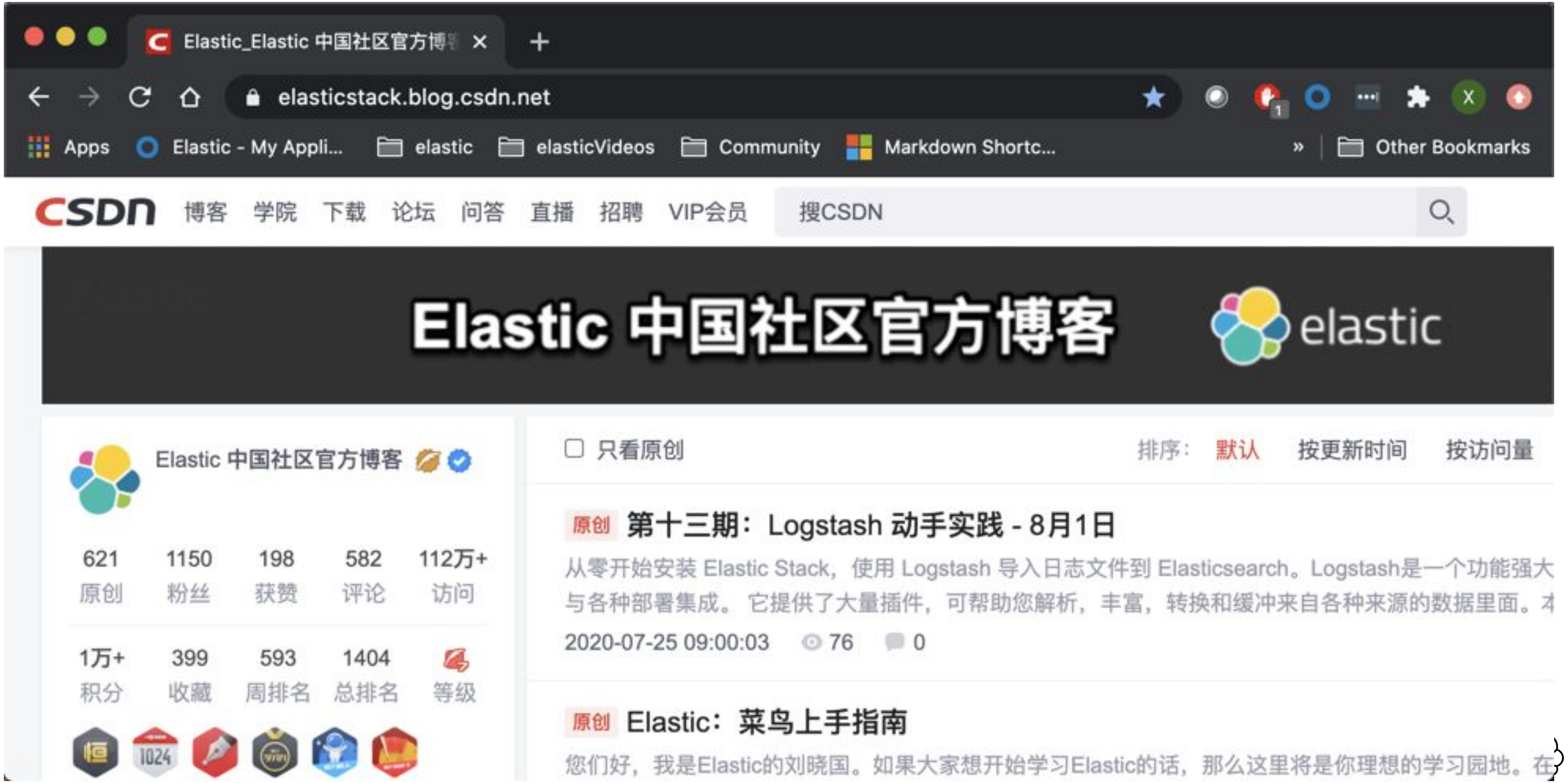
elastic

Logstash 动手实践

刘晓国
Elastic 社区布道师
2020年9月1日

elasticstack.blog.csdn.net



Elastic 中国社区官方博客

Elastic 中国社区官方博客

621	1150	198	582	112万+
原创	粉丝	获赞	评论	访问
1万+	399	593	1404	
积分	收藏	周排名	总排名	等级

□ 只看原创 排序： 默认 按更新时间 按访问量

原创 第十三期：Logstash 动手实践 - 8月1日

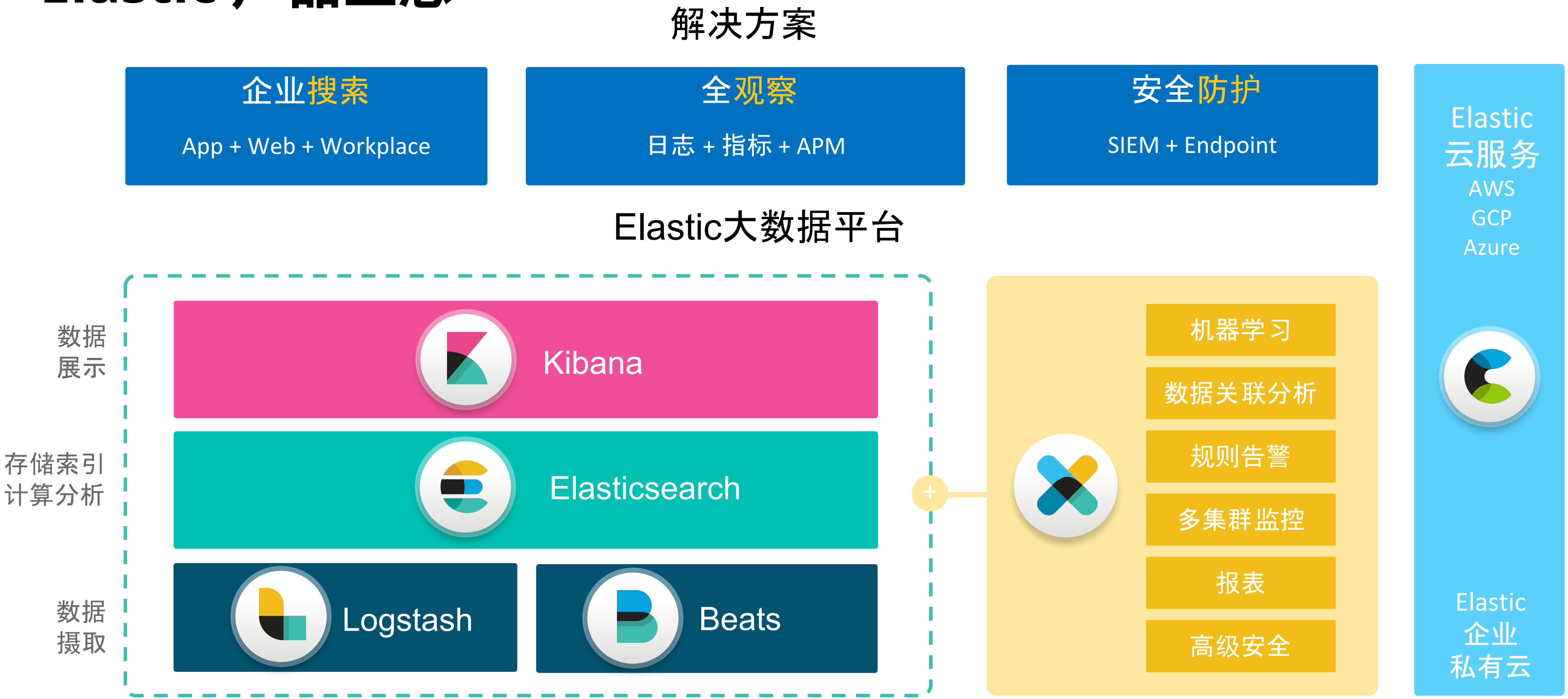
从零开始安装 Elastic Stack，使用 Logstash 导入日志文件到 Elasticsearch。Logstash是一个功能强大
与各种部署集成。 它提供了大量插件，可帮助您解析，丰富，转换和缓冲来自各种来源的数据里面。

2020-07-25 09:00:03 76 0

原创 Elastic：菜鸟上手指南

您们好，我是Elastic的刘晓国。如果大家想开始学习Elastic的话，那么这里将是你理想的学习园地。在

Elastic 产品生态



解决方案

企业搜索

App + Web + Workplace

全观察

日志 + 指标 + APM

安全防护

SIEM + Endpoint

Elastic
云服务
AWS
GCP
Azure

Elastic大数据平台

数据
展示

Kibana

存储索引
计算分析

Elasticsearch

数据
摄取

Logstash

Beats

机器学习

数据关联分析

规则告警

多集群监控

报表

高级安全

Elastic
企业
私有云

议程

Getting Started with Logstash

1 **Logstash 产品简介**

2 Logstash 剖析

3 动手实践

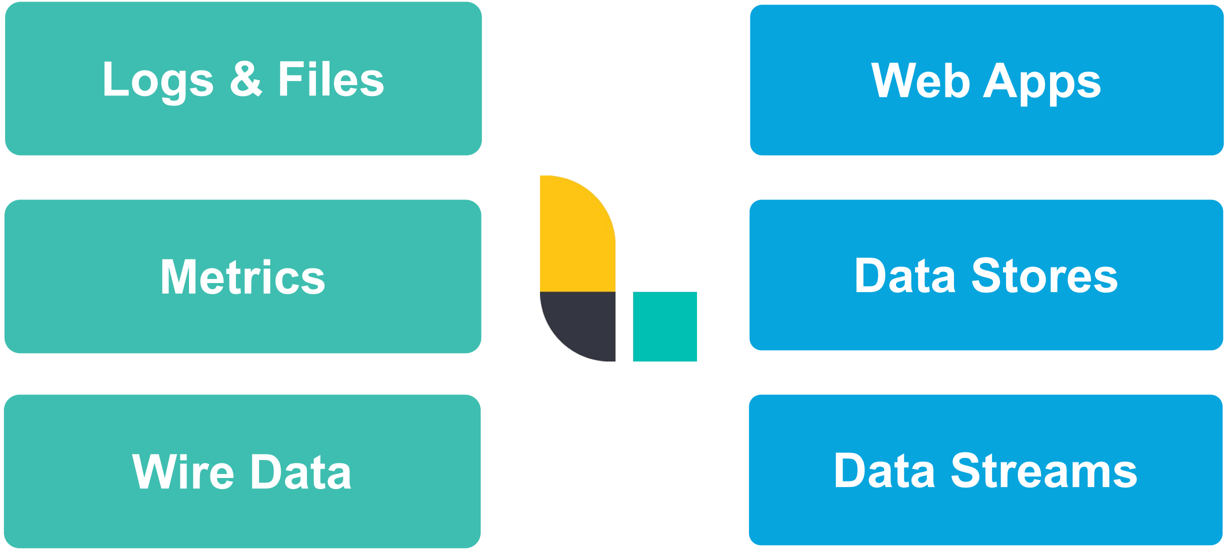
Logstash

The Dataflow Engine

- 它是用于数据物流的开源流式 ETL 引擎
- 在几分钟内建立数据流管道
- **具有水平可扩展及韧性且具有自适应缓冲**
- 不可知的数据源
- **具有200多个集成和处理器**的插件生态系统
- 使用 Elastic Stack **监视**和**管理**部署

数据源

Ingest All the Things



Logs & Files

Metrics

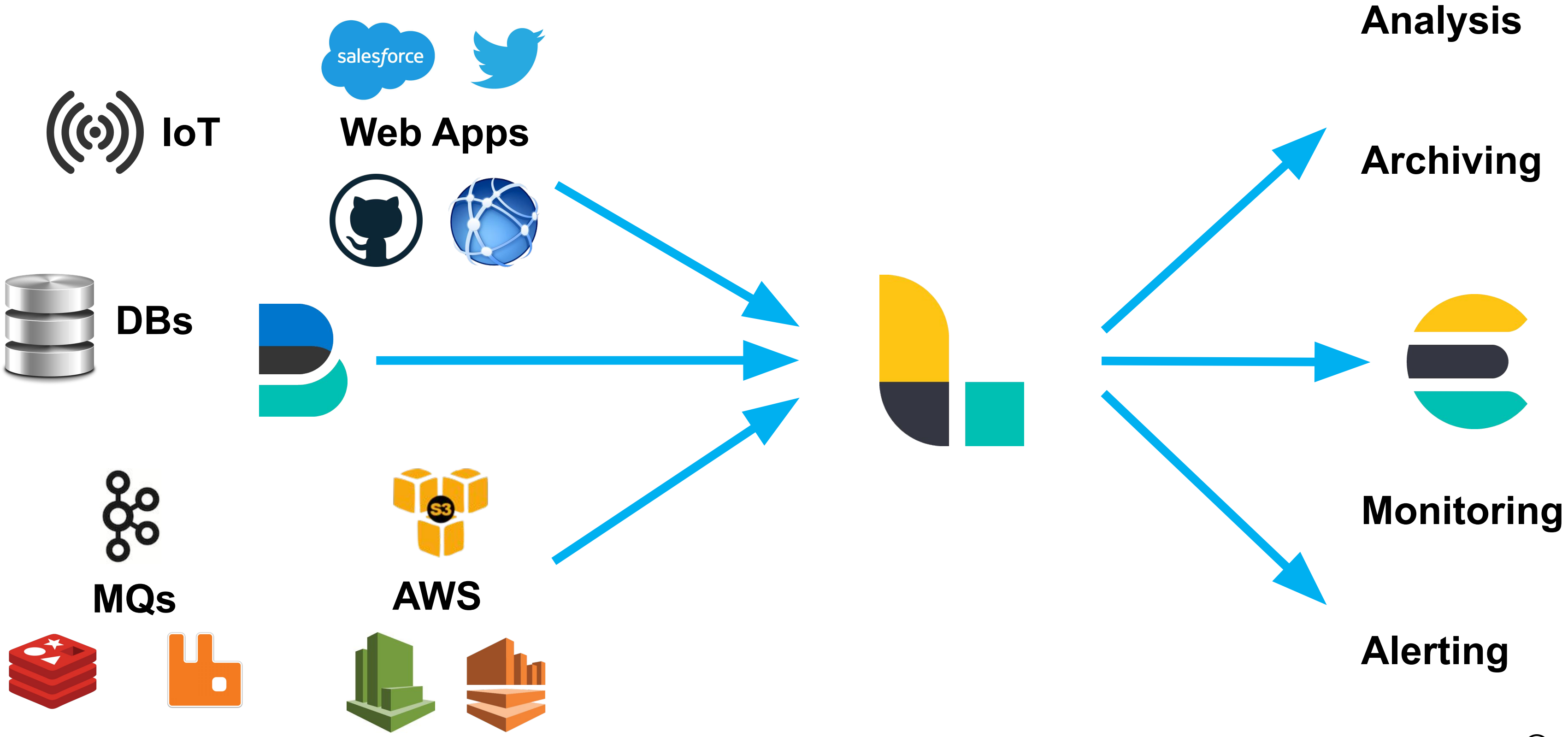
Wire Data

Web Apps

Data Stores

Data Streams

热门数据源



IoT

Web Apps

DBs

MQs

AWS

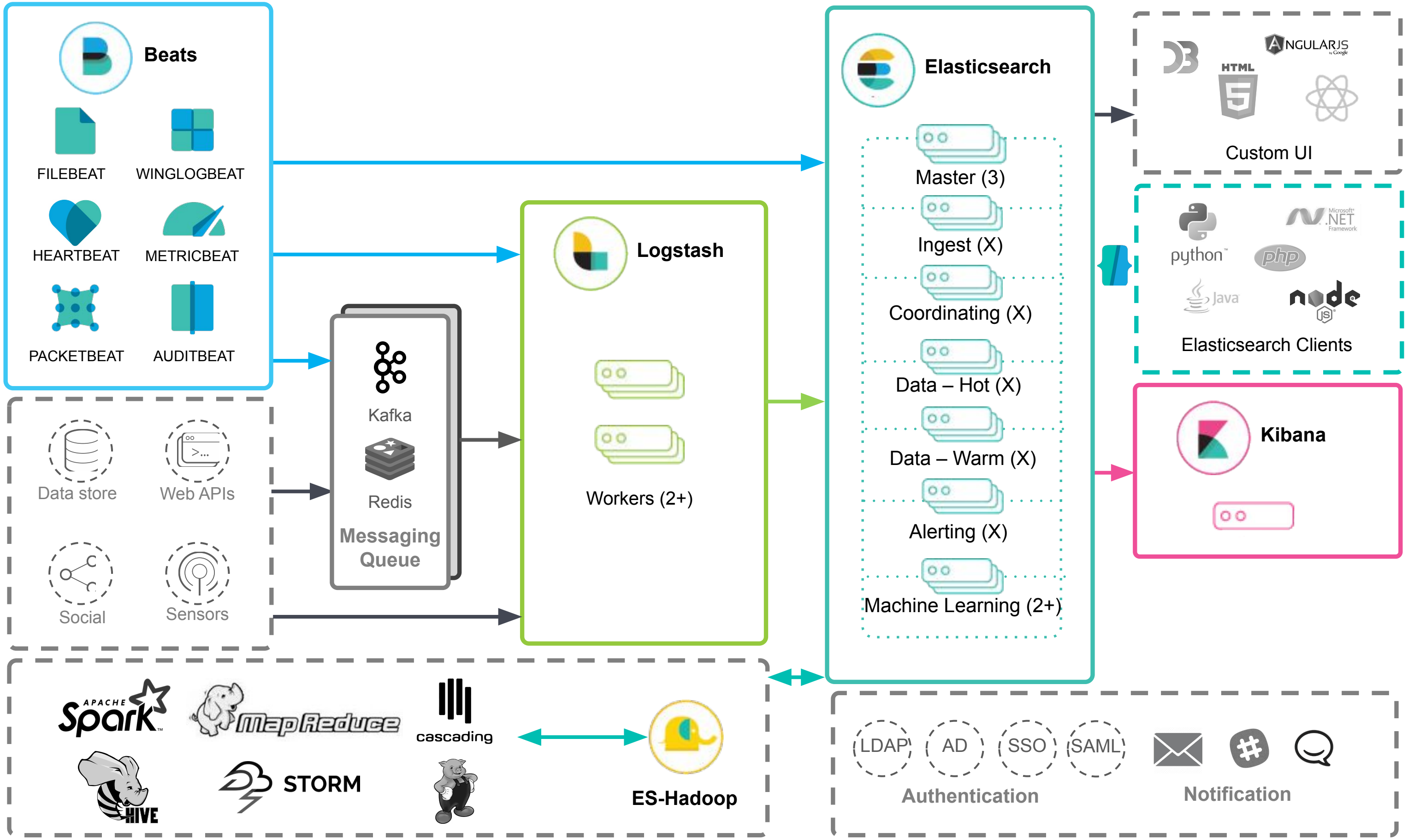
Analysis

Archiving

Monitoring

Alerting

Beats 是如何接入到Elasticsearch中的?



elastic

Agenda

Getting Started with Logstash

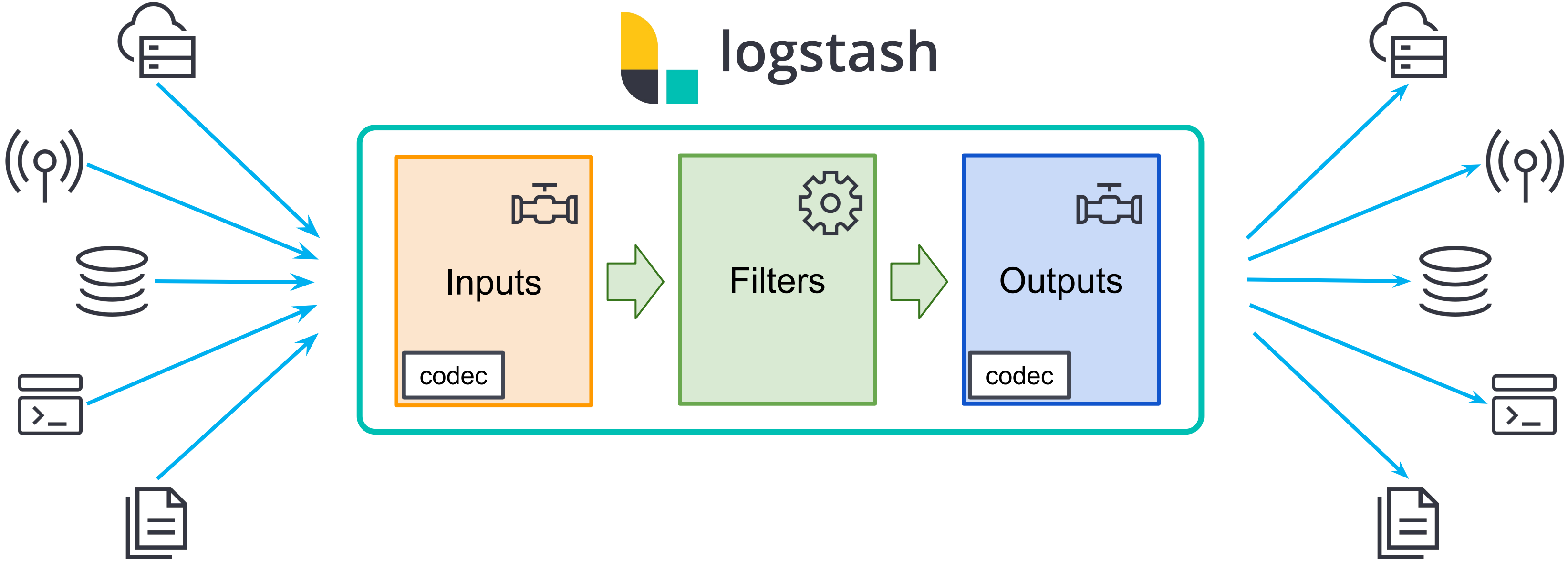
1 Logstash Product Overview

2 **Logstash 剖析**

3 动手实践

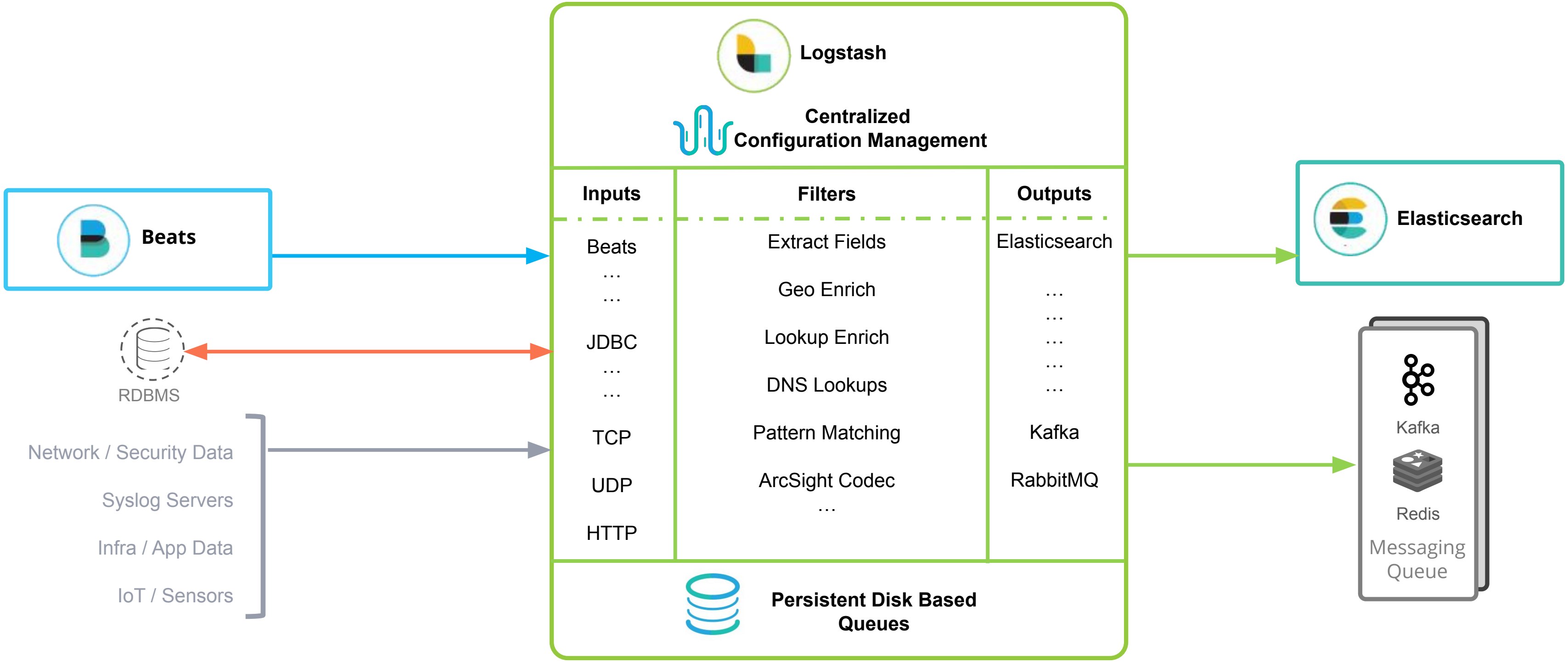
Logstash 内部 1/2

Inputs, Filters and Outputs



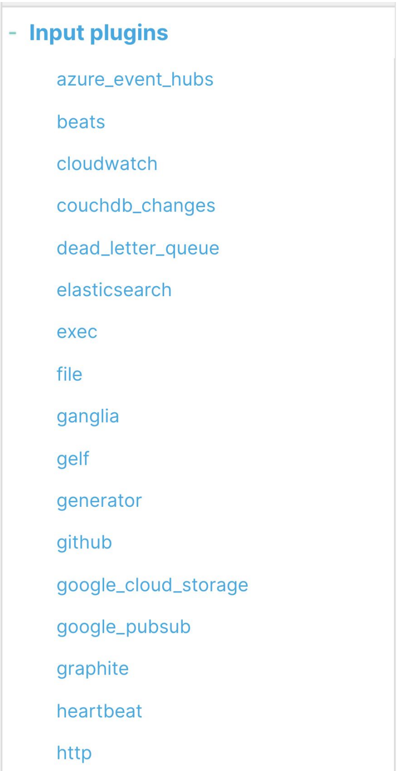
Logstash 内部 2/2

Inputs, Filters and Outputs

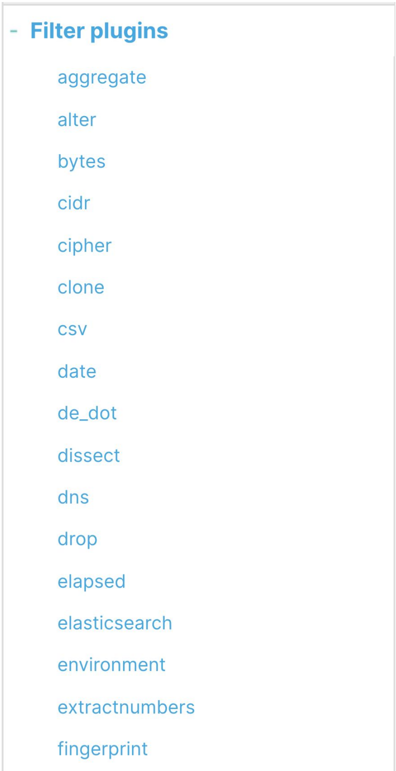


Logstash Reference

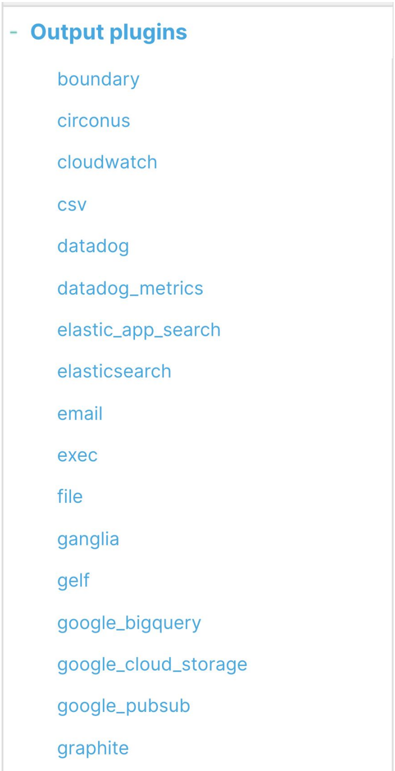
https://www.elastic.co/guide/en/logstash/current/index.html



- Input plugins
  - azure_event_hubs
  - beats
  - cloudwatch
  - couchdb_changes
  - dead_letter_queue
  - elasticsearch
  - exec
  - file
  - ganglia
  - gelf
  - generator
  - github
  - google_cloud_storage
  - google_pubsub
  - graphite
  - heartbeat
  - http



- Filter plugins
  - aggregate
  - alter
  - bytes
  - cidr
  - cipher
  - clone
  - csv
  - date
  - de_dot
  - dissect
  - dns
  - drop
  - elapsed
  - elasticsearch
  - environment
  - extractnumbers
  - fingerprint



- Output plugins
  - boundary
  - circonus
  - cloudwatch
  - csv
  - datadog
  - datadog_metrics
  - elastic_app_search
  - elasticsearch
  - email
  - exec
  - file
  - ganglia
  - gelf
  - google_bigquery
  - google_cloud_storage
  - google_pubsub
  - graphite

Agenda

Getting Started with Logstash

1 Logstash Product Overview

2 The Anatomy of Logstash

3 **Hands On Workshop**

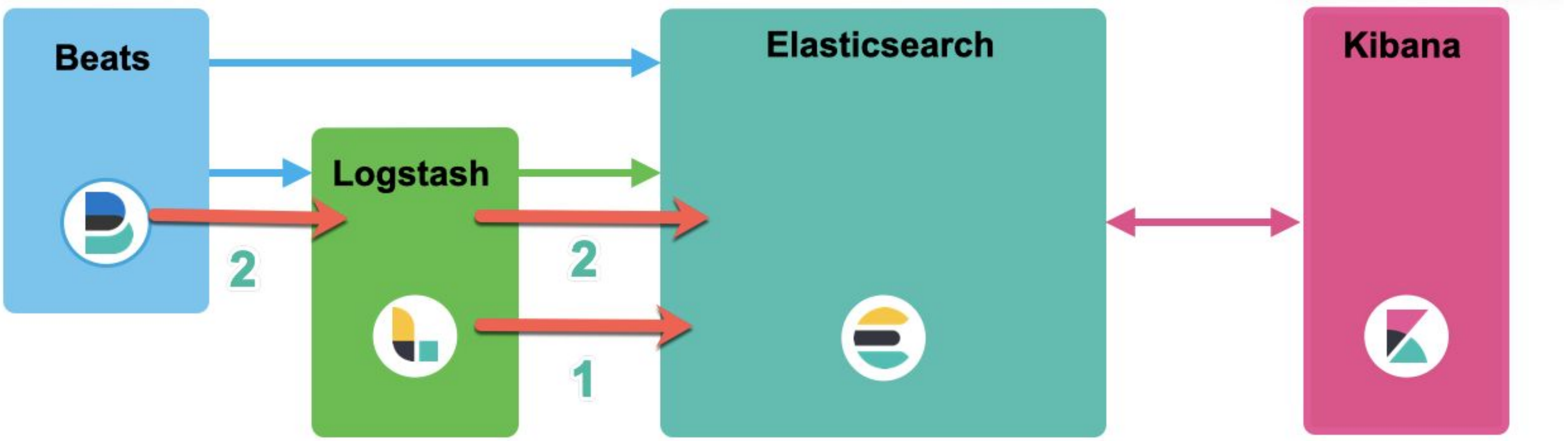
elastic

动手实践包括

- 下载及运行 Logstash
- 使用 Apache Weblog 作为输入
- 使用如下的过滤器来丰富 Apache Weblog
  - Grok
  - Geoip
  - Useragent
  - Date
  - Mutate
- 把丰富后的日志导入到 Elasticsearch

https://github.com/liu-xiao-guo/logstash_getting_started

Demo 情形



- Logstash => Elasticsearch
- Filebeat => Logstash => Elasticsearch

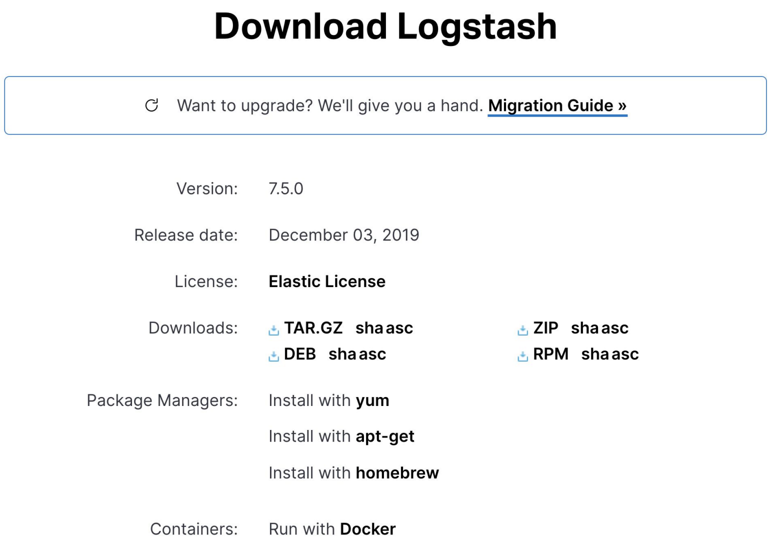
有用的 Logstash 链接

1. Logstash 频道博客文章
   - https://blog.csdn.net/ubuntutouch/category_9335275.html
2. 如何安装 Elastic 栈中的 Logstash
   - https://blog.csdn.net/UbuntuTouch/article/details/99655350
3. Logstash：Logstash 入门教程（一）
   - https://elasticstack.blog.csdn.net/article/details/105973985
4. Logstash：Logstash 入门教程（二）
   - https://elasticstack.blog.csdn.net/article/details/105979677
5. Logstash：Data转换，分析，提取，丰富及核心操作
   - https://blog.csdn.net/UbuntuTouch/article/details/100770828
6. Logstash：把Apache日志导入到 Elasticsearch
   - https://blog.csdn.net/UbuntuTouch/article/details/100727051
7. Logstash: 启动监控及集中管理
   - https://blog.csdn.net/UbuntuTouch/article/details/103767088
8. Logstash 培训视频
   - https://www.elastic.co/cn/webinars/getting-started-logstash

Logstash Download

https://www.elastic.co/downloads/logstash

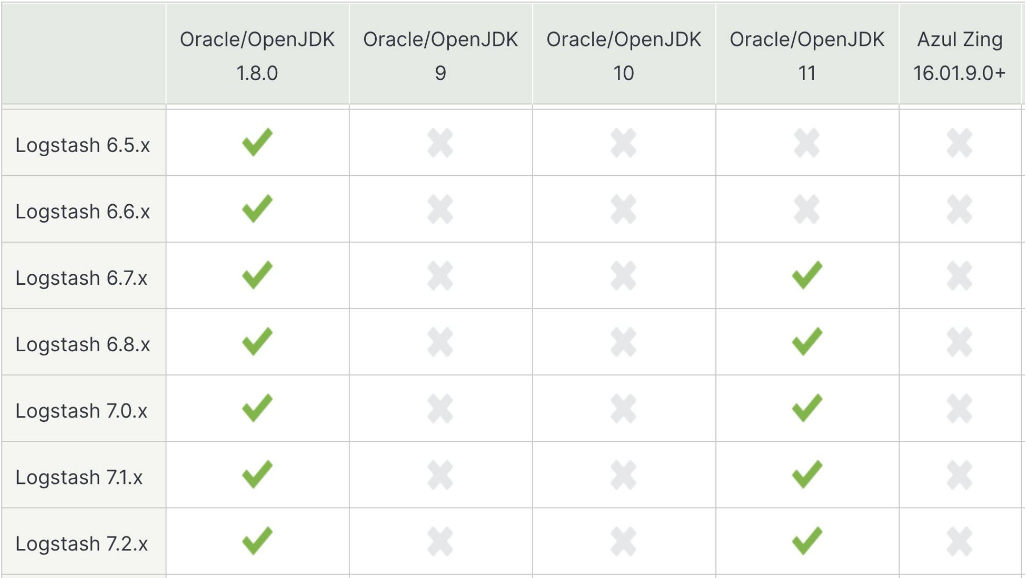
https://www.elastic.co/support/matrix#matrix_jvm



Download Logstash

Want to upgrade? We'll give you a hand. Migration Guide »

Version:	7.5.0
Release date:	December 03, 2019
License:	**Elastic License**
Downloads:	**TAR.GZ** sha asc **DEB** sha asc **ZIP** sha asc **RPM** sha asc
Package Managers:	Install with **yum** Install with **apt-get** Install with **homebrew**
Containers:	Run with **Docker**



	Oracle/OpenJDK 1.8.0	Oracle/OpenJDK 9	Oracle/OpenJDK 10	Oracle/OpenJDK 11	Azul Zing 16.01.9.0+
Logstash 6.5.x	✓	✗	✗	✗	✗
Logstash 6.6.x	✓	✗	✗	✗	✗
Logstash 6.7.x	✓	✗	✗	✓	✗
Logstash 6.8.x	✓	✗	✗	✓	✗
Logstash 7.0.x	✓	✗	✗	✓	✗
Logstash 7.1.x	✓	✗	✗	✓	✗
Logstash 7.2.x	✓	✗	✗	✓	✗

Run Logstash from the command line

- Mac, Unix & Linux

```
bin/logstash [options]
```

- Windows

```
bin/logstash.bat [options]
```

Pipeline configurations

Input, Filter and Output configurations must be defined

```
input {
 …
}

filter {
 …
}

output {
 …
}
```

2 ways of running logstash with configurations

- -e : Set configurations in command line

```
bin/logstash -e 'input { stdin { } } output { stdout { } }'
```

- -f : If configurations are set in a file (ex. pipeline.conf)

```
bin/logstash -f pipeline.conf
```

Before we start

- Run Elasticsearch and Kibana at local
- Edit config/logstash.yml
    - Recommendation - uncomment and set `config.reload.automatic` to `true` to avoid restarting logstash every time when we change configurations.

```
config.reload.automatic : true
```

- Create weblog.conf file, set input and output

```
input {
     tcp {
          port => 9900
     }
}

output {
     stdout { }
}
```

Run logstash

```
$ bin/logstash -f weblog.conf

…

"pipeline.sources"=>["/Users/elastic/logstash-7.5.0/weblog.conf"],
:thread=>"#<Thread:0xcf50672 run>"}
[2019-12-05T15:47:34,254][INFO ][logstash.javapipeline    ][main] Pipeline started
{"pipeline.id"=>"main"}
[2019-12-05T15:47:34,265][INFO ][logstash.inputs.tcp      ][main] Starting tcp
input listener {:address=>"0.0.0.0:9900", :ssl_enable=>"false"}
[2019-12-05T15:47:34,307][INFO ][logstash.agent           ] Pipelines running
{:count=>1, :running_pipelines=>[:main], :non_running_pipelines=>[]}
[2019-12-05T15:47:34,522][INFO ][logstash.agent           ] Successfully started
Logstash API endpoint {:port=>9600}
```

Send simple message to logstash

- Send 'hello logstash' text to tcp 9900 port use netcat

```
$ echo 'hello logstash' | nc localhost 9900
```

- Check logstash output console log. You will see sent text in message field

```
{
      "@version" => "1",
       "message" => "hello logstash",
          "host" => "localhost",
          "port" => 61403,
    "@timestamp" => 2019-12-05T06:54:40.767Z
}
```

Read weblog file and sent to logstash

- Download weblog-sample.log file : https://ela.st/weblog-sample
- Read first line of file and send logstash

```
$ head -n 1 weblog-sample.log | nc localhost 9900
```

- Check logstash console.

```
{
    "@version" => "1",
     "message" => "14.49.42.25 - - [12/May/2019:01:24:44 +0000] \"GET
/articles/ppp-over-ssh/ HTTP/1.1\" 200 18586 \"-\" \"Mozilla/5.0 (Windows; U;
Windows NT 6.1; en-US; rv:1.9.2b1) Gecko/20091014 Firefox/3.6b1 GTB5\"",
        "host" => "localhost",
        "port" => 61639,
  "@timestamp" => 2019-12-05T07:15:33.105Z
}
```

Set Filter - grok

- message field can be parsed with **grok** filter
- Edit weblog.conf file - add **grok** filter

```
input {
      tcp {
            port => 9900
      }
}

filter {
   grok {
      match => { "message" => "%{COMBINEDAPACHELOG}" }
   }
}

output {
      stdout { }
}
```

Set Filter - geoip

- clientip field contains ip address. This field can be enriched with **geoip** filter
- Edit weblog.conf file - add **geoip** filter
- This filter must be set after **grok**

```
filter {
  grok {
    match => { "message" => "%{COMBINEDAPACHELOG}"  }
  }

  geoip {
    source => "clientip"
  }
```

Set Filter - useragent

- agent field contains client’s OS and device, browser information. This field can be enriched with **useragent** filter
- Edit weblog.conf file - add **useragent** filter
- This filter must be set after **grok**

```
filter {
…

  useragent {
    source => "agent"
    target => "useragent"
  }
```

Set Filter - mutate : convert

- You might noticed bytes fields format is text. This field can be converted to number with **mutate : convert** filter.
- Edit weblog.conf file - add **mutate : convert** filter.
- This filter must be set after **grok**

```
filter {
…
  mutate {
    convert => {
      "bytes" => "integer"
    }
  }
```

Set Filter - date

- Logstash stores it’s event time in @timestamp field. But actual log created time is in timestamp field (without @). This field’s format is not ISO8601, so stored as text. We can use **date** filter to convert this field to date type.
- Edit weblog.conf file - add **date** filter.
- This filter must be set after **grok**

```
filter {
…
  date {
    match => ["timestamp", "dd/MMM/yyyy:HH:mm:ss Z"]
  }
```

Set Output - elasticsearch

- Current output is stdout. Comment or remove stdout.
- Add elasticsearch output. Set **hosts** to living Elasticsearch cluster.

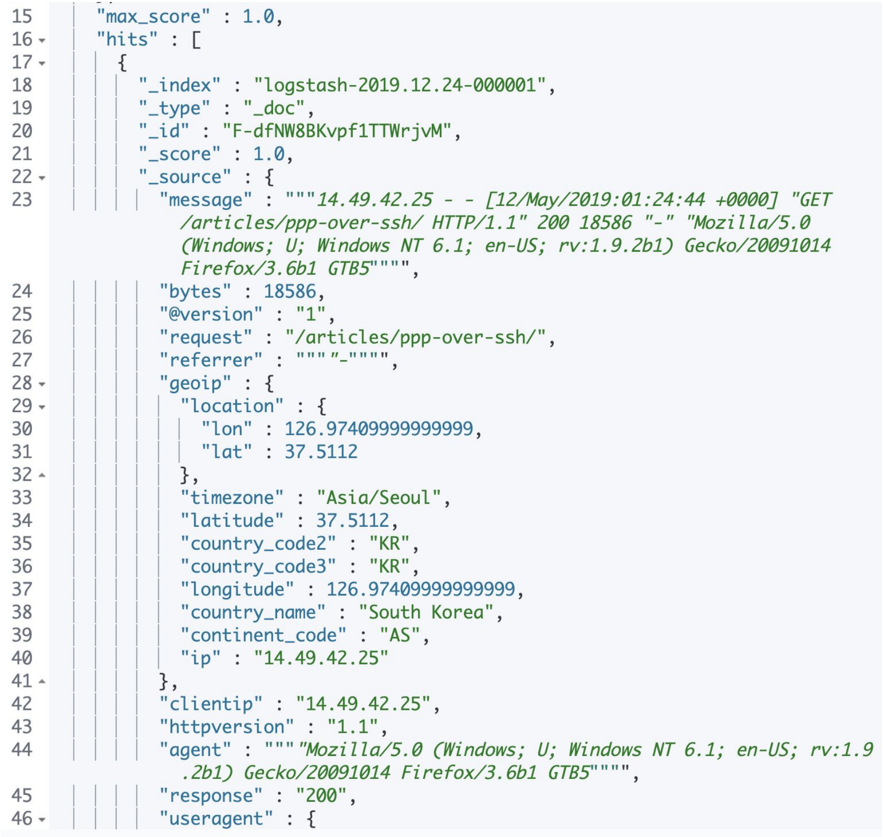
```
filter {
…
  date {
    match => ["timestamp", "dd/MMM/yyyy:HH:mm:ss Z"]
  }
}

output {
#    stdout { }
     elasticsearch {
          hosts => ["localhost:9200"]
          user => "elastic"
          password => "changeme"
     }
}
```

Set Output - elasticsearch

- Save weblog.conf file
- Restart Logstash and send the same weblog to logstash again.
- Data is indexed into elasticsearch.
- Search **logstash-*** index

```
GET logstash-*/_search
```



```
    "max_score" : 1.0,
    "hits" : [
      {
        "_index" : "logstash-2019.12.24-000001",
        "_type" : "_doc",
        "_id" : "F-dfNW8BKvpf1TTWrjvM",
        "_score" : 1.0,
        "_source" : {
          "message" : """14.49.42.25 - - [12/May/2019:01:24:44 +0000] "GET
            /articles/ppp-over-ssh/ HTTP/1.1" 200 18586 "-" "Mozilla/5.0
            (Windows; U; Windows NT 6.1; en-US; rv:1.9.2b1) Gecko/20091014
            Firefox/3.6b1 GTB5"""",
          "bytes" : 18586,
          "@version" : "1",
          "request" : "/articles/ppp-over-ssh/",
          "referrer" : """"-"""",
          "geoip" : {
            "location" : {
              "lon" : 126.97409999999999,
              "lat" : 37.5112
            },
            "timezone" : "Asia/Seoul",
            "latitude" : 37.5112,
            "country_code2" : "KR",
            "country_code3" : "KR",
            "longitude" : 126.97409999999999,
            "country_name" : "South Korea",
            "continent_code" : "AS",
            "ip" : "14.49.42.25"
          },
          "clientip" : "14.49.42.25",
          "httpversion" : "1.1",
          "agent" : """"Mozilla/5.0 (Windows; U; Windows NT 6.1; en-US; rv:1.9
            .2b1) Gecko/20091014 Firefox/3.6b1 GTB5"""",
          "response" : "200",
          "useragent" : {
```

elastic

THANK YOU