elastic

Agenda

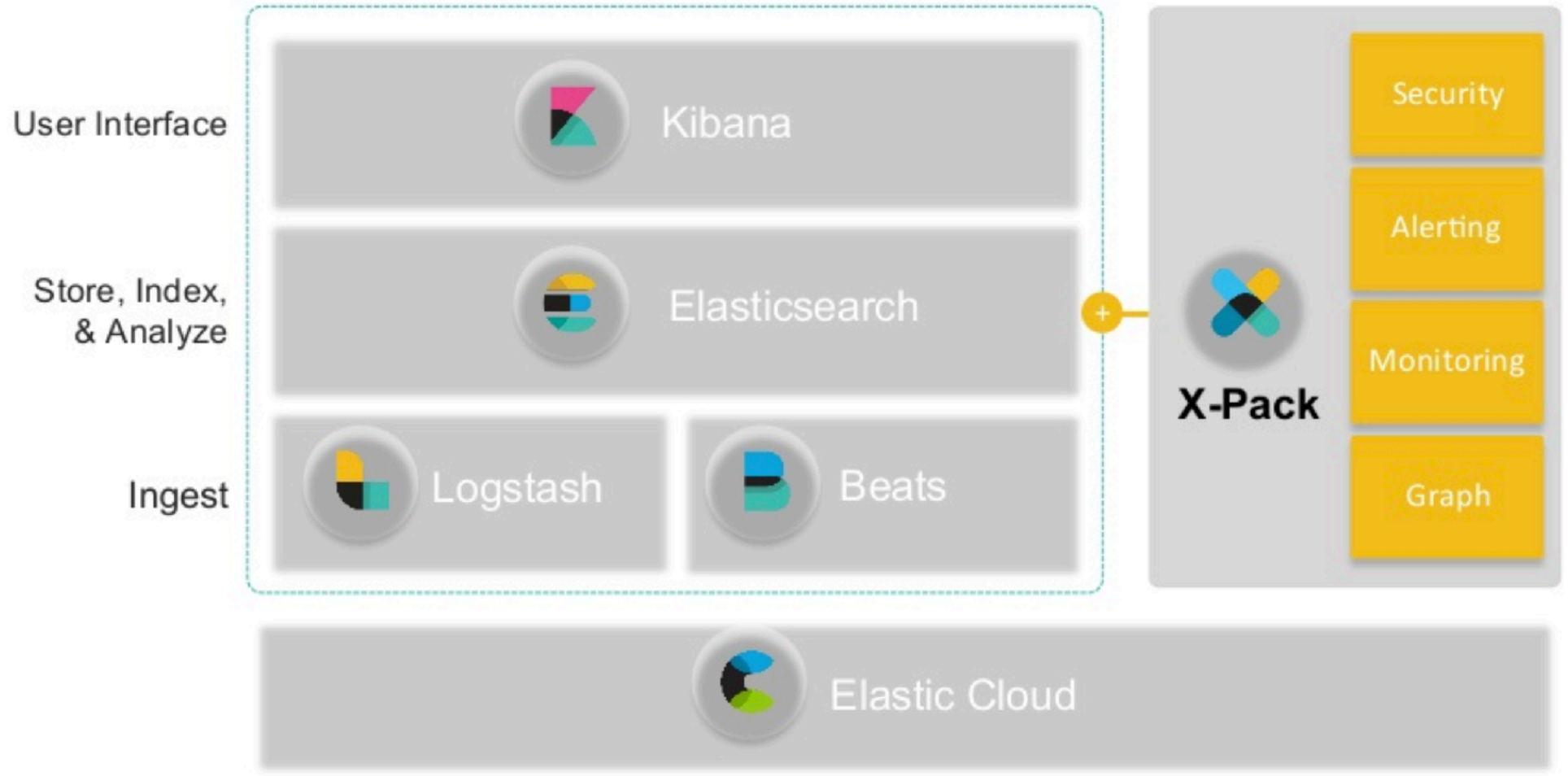
- Download Data
- Install Elasticsearch, Kibana and Logstash
- Introduction Elasticsearch, Kibana and Logstash
- Import Data to Elasticsearch using Logstash
- Create Kibana Dashboard

請別太期待 ...

沒什麼能分享給大家

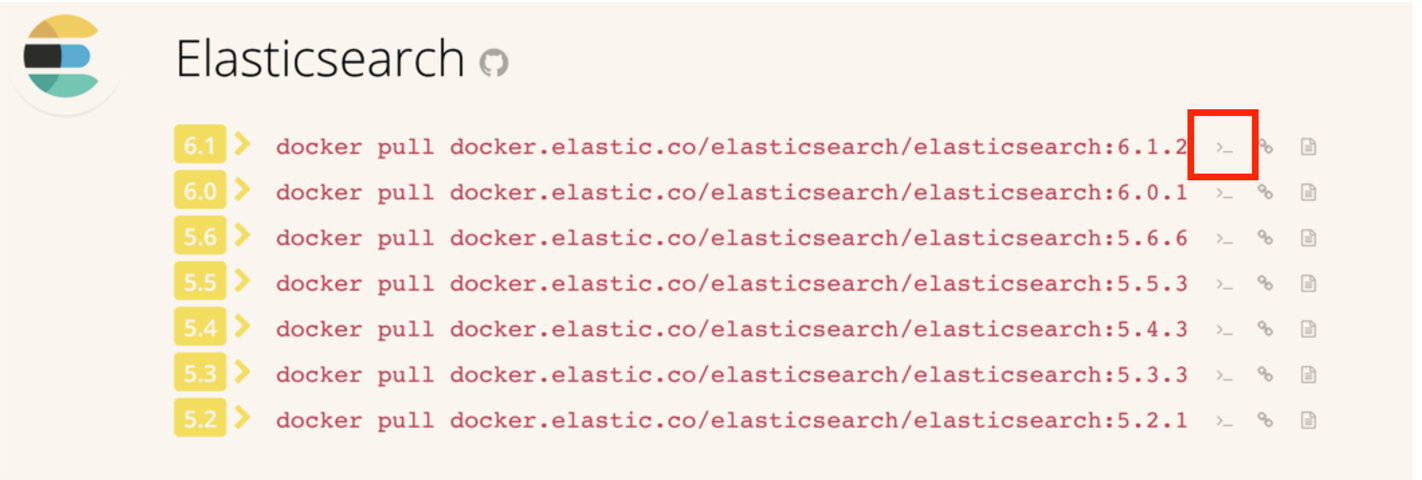
感謝 Square 和 Andy

The Elastic Stack



Step1: https://www.docker.elastic.co/#

Step2:



Step3: 啟動 Docker

Step4: 執行指令

docker run -p 9200:9200 -p 9300:9300 -e "discovery.type=single-node" docker.elastic.co/elasticsearch/elasticsearch:6.1.2

Step1: Mac: https://artifacts.elastic.co/downloads/kibana/kibana-6.1.2-darwin-x86_64.tar.gz

Window: https://www.elastic.co/guide/en/kibana/current/windows.html

Step2: 執行

Step3: vim config/kibana.yml

Step4: #server.port:5601

#server.host:“local host”

#server.name:“vickie”

#elasticsearch.url: “http://localhost:9200"

將 # 拿掉，server.name 設定為個人名稱

Step5: 執行指令

./bin/kibana

Step1: https://www.elastic.co/products/logstash

Step2: Download

Step3: 安裝 beats
./bin/logstash-plugin install logstash-input-beats

Step4: 執行指令
mkdir conf.d
vim conf.d/test1.conf
./bin/logstash -f /Users/vickieliu/Developer/
resources/logstash-6.1.2/conf.d/test1.conf

Version Compatibility with Elasticsearch

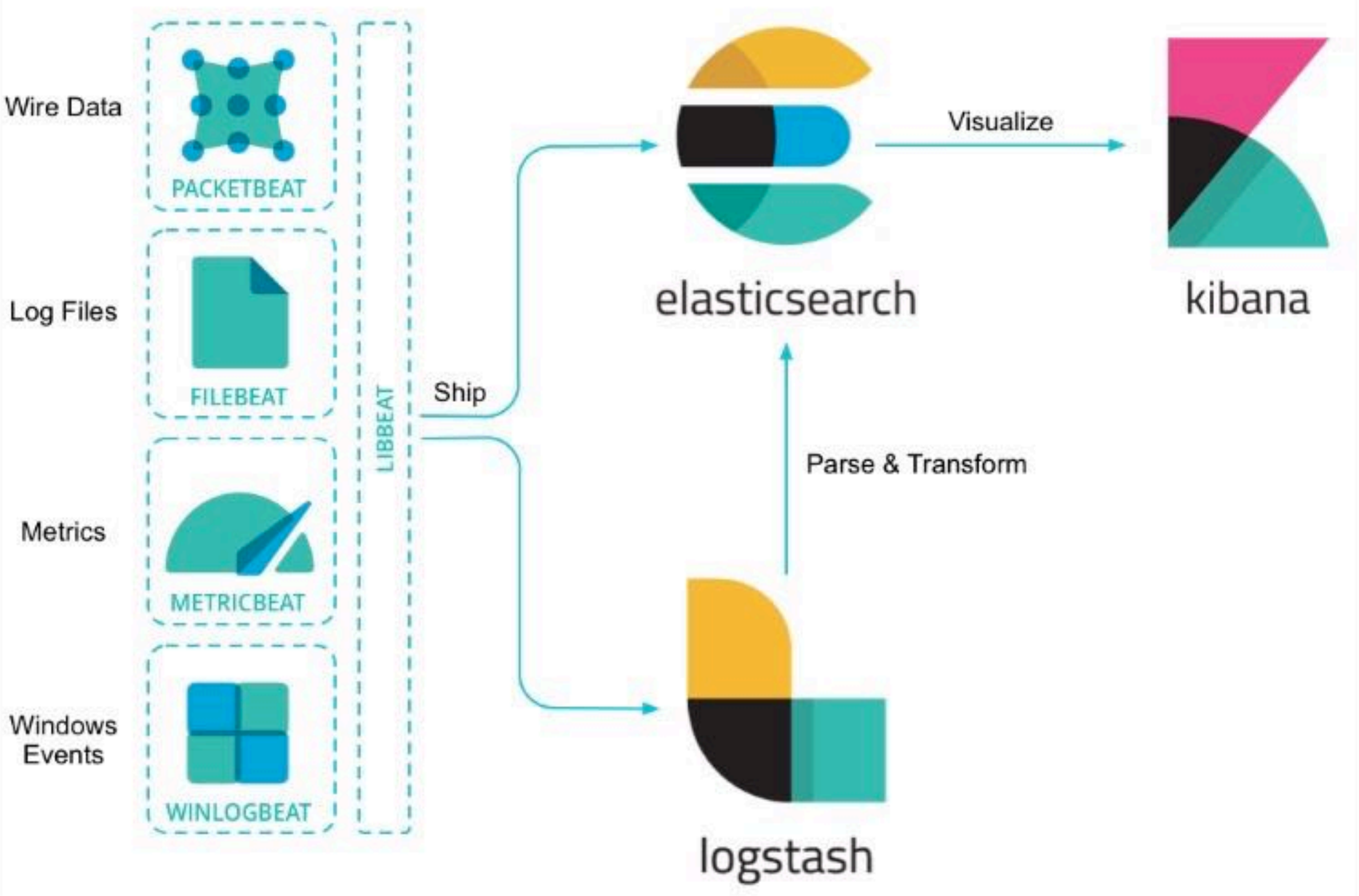
Situation	Example Kibana version	Example ES version	Outcome
Versions are the same.	5.1.2	5.1.2	💚 OK
ES patch number is newer.	5.1.**2**	5.1.**5**	⚠️ Logged warning
ES minor number is newer.	5.**1**.2	5.**5**.0	⚠️ Logged warning
ES major number is newer.	**5**.1.2	**6**.0.0	🚫 Fatal error
ES patch number is older.	5.1.**2**	5.1.**0**	⚠️ Logged warning
ES minor number is older.	5.**1**.2	5.**0**.0	🚫 Fatal error
ES major number is older.	**5**.1.2	**4**.0.0	🚫 Fatal error

ELK 主要由三個開源套件所組成，用於

收集日誌資料 (Logstash)、

檢索資料 (Elasticsearch) 並將

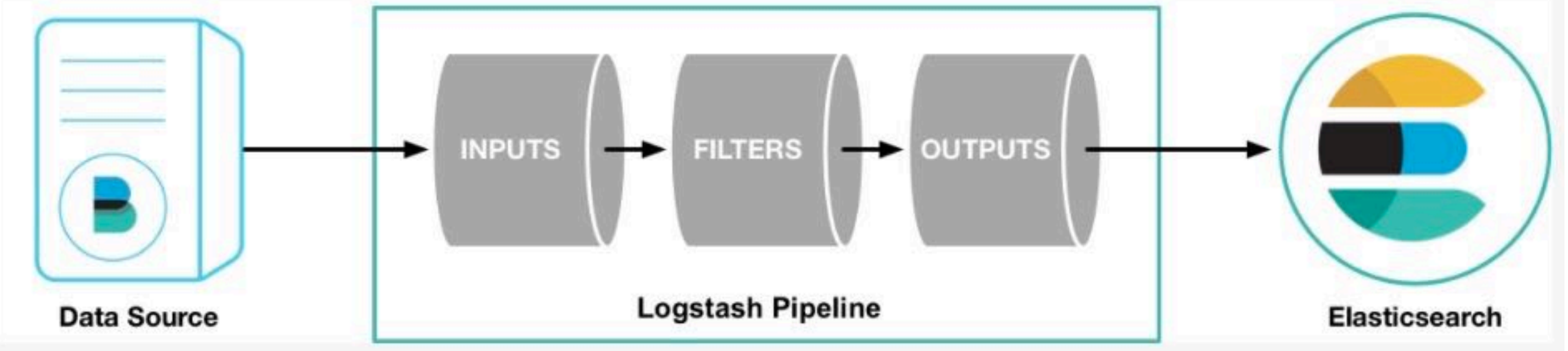
資料做視覺化呈現 (Kibana)。

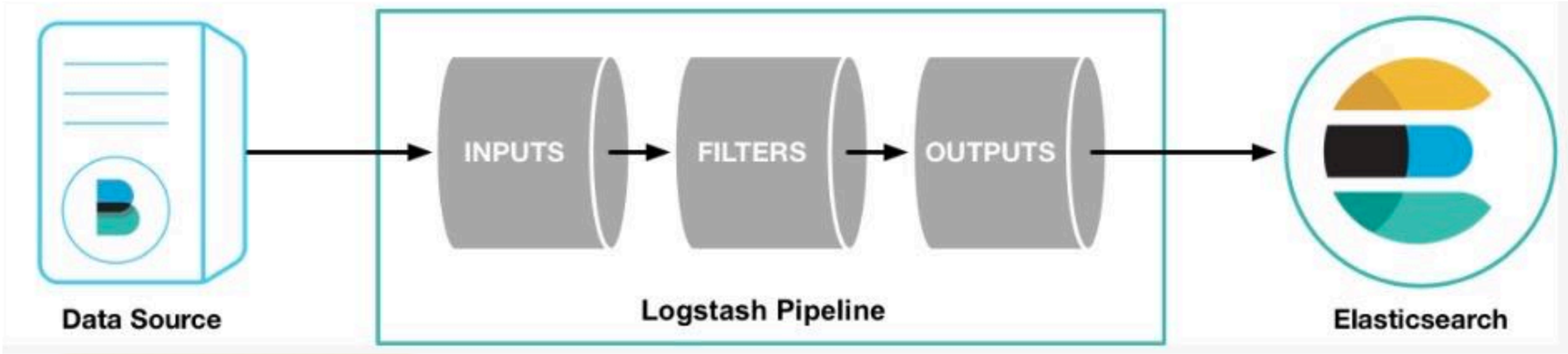

Wire Data
PACKETBEAT
Log Files
FILEBEAT
Metrics
METRICBEAT
Windows Events
WINLOGBEAT
LIBBEAT
Ship
elasticsearch
Visualize
kibana
Parse & Transform
logstash

- 提供各式各樣的日誌收集及輸出的 Plugin
- 可用於串接資料庫或日誌檔案
- 將資料收集處理後，輸出至 Elasticsearch

- 分散的索引搜尋系統
- 可用於全文檢索
- 提供 REST API 串接
- NoSQL 資料庫的一種，資料以 JSON 進行存取
- 與 MongoDB 一樣都是 Document DB，最大的不同是所有欄位都可以建立索引進行全文搜尋

- 透過網頁介面來呈現資料、製作圖表
- 透過開發工具對 elasticsearch 執行資料操作變得更簡單


INPUTS
FILTERS
OUTPUTS
Data Source
Logstash Pipeline
Elasticsearch



INPUTS

由各種地方抓取相關的資料，不論是標準輸入 stdin，檔案 (如: csv) 或是監聽一個連線 (如: kafka) 都可以，並且可以把不同的資料來區隔開來作為之後用途

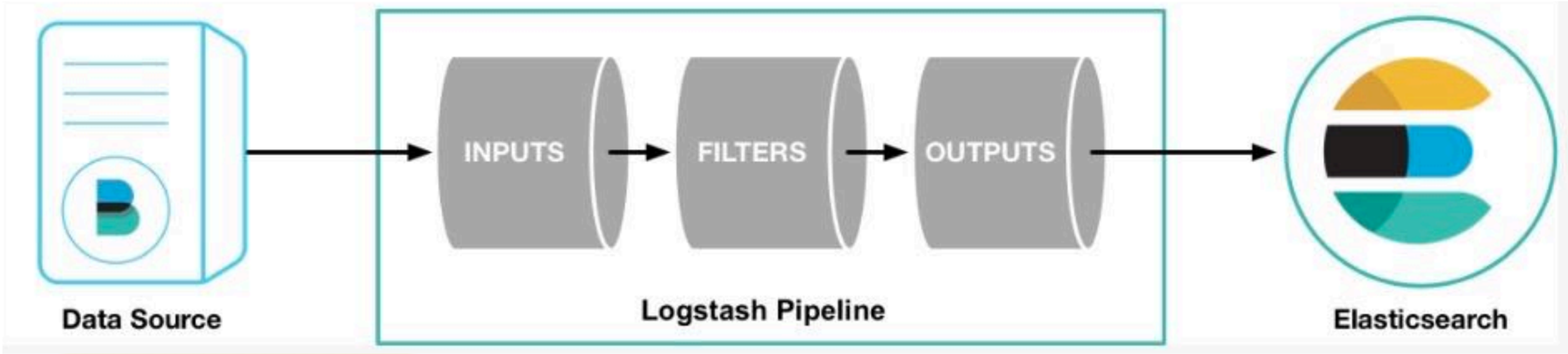
```
input {
    file {
        path => “/.csv"
    }
}
```

怎麼讓 Logstash 從頭讀文件？

```
input {
    file {
        path => "./csv"
        start_position => "beginning"
        sincedb_path => "/dev/null"
    }
}
```

也可以 key-in json 格式文件

```
input {
    stdin {
      codec => json
    }
}
```

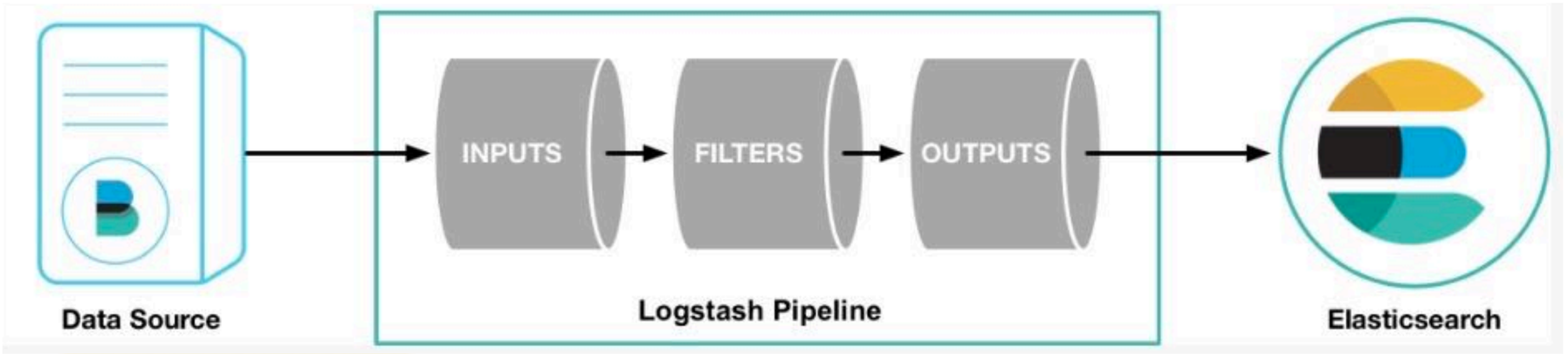


FILTERS

抓取到資料之後，這時候可以選擇使用 grok 去拆解資料，比如說把一串非常長但是沒人看得懂的 log 拆解成各個相對應得資料欄位

```
filter {

    mutate {  #資料型態轉換

        convert => ["request_time", "float"]

    }

}
```

Logstash 在執行 FILTERS 的時候是照順序執行的，有時候依照不同需求會有同樣 Data 需做不同的篩選條件或輸出條件，而在執行多個配置文件可參考http://www.cnblogs.com/licongyu/p/5442652.html



OUTPUTS

可以輸出到各種地方，不論是輸出到 stdout 印出來，或是直接丟給 Elasticsearch 拿來做分析用甚至可以直接輸出到資料庫

```
output {
    stdout {
        codec => rubydebug{ }
        # codec => json{ }
    }
}
```

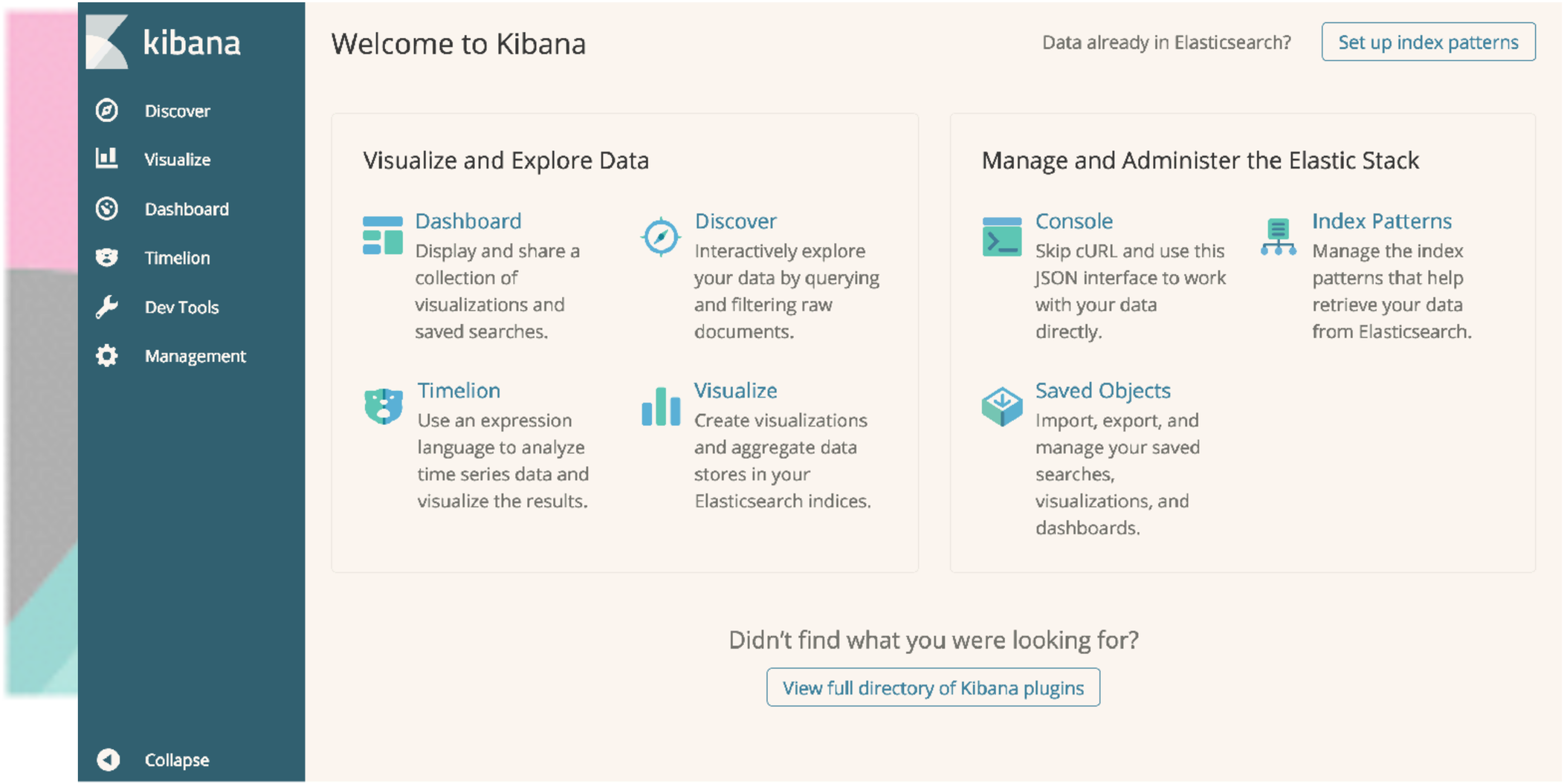
怎麼判斷 Elasticsearch 能不能用？

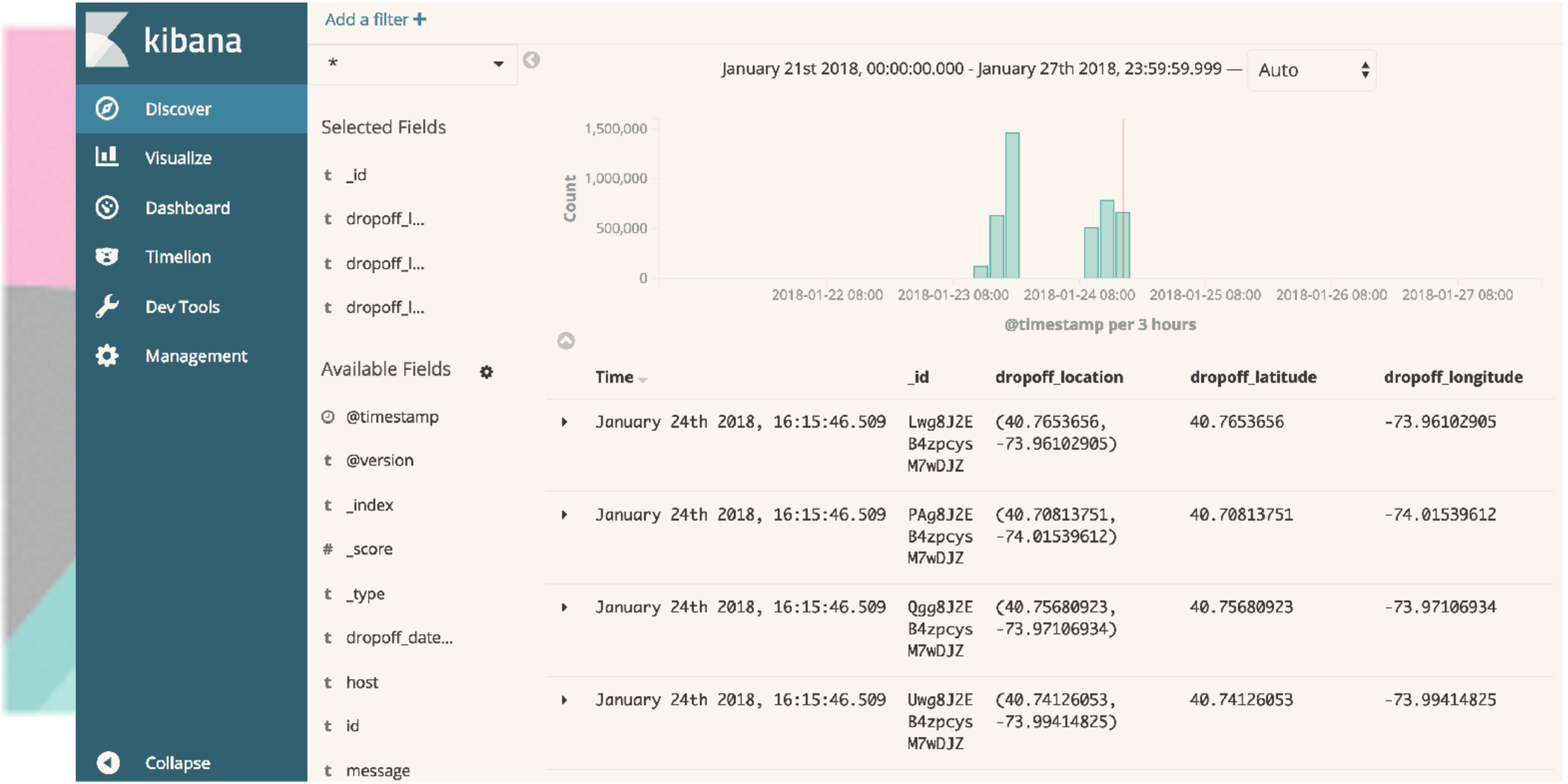
http://127.0.0.1:9200/


```
{
  "name" : "m8SVP2q",
  "cluster_name" : "docker-cluster",
  "cluster_uuid" : "1VBLm1iaQLC3frF1hksB0w",
  "version" : {
    "number" : "6.1.2",
    "build_hash" : "5b1fea5",
    "build_date" : "2018-01-10T02:35:59.208Z",
    "build_snapshot" : false,
    "lucene_version" : "7.1.0",
    "minimum_wire_compatibility_version" : "5.6.0",
    "minimum_index_compatibility_version" : "5.0.0"
  },
  "tagline" : "You Know, for Search"
}
```


恭喜你！離 Kibana 更近一步了！

http://127.0.0.1:5601/

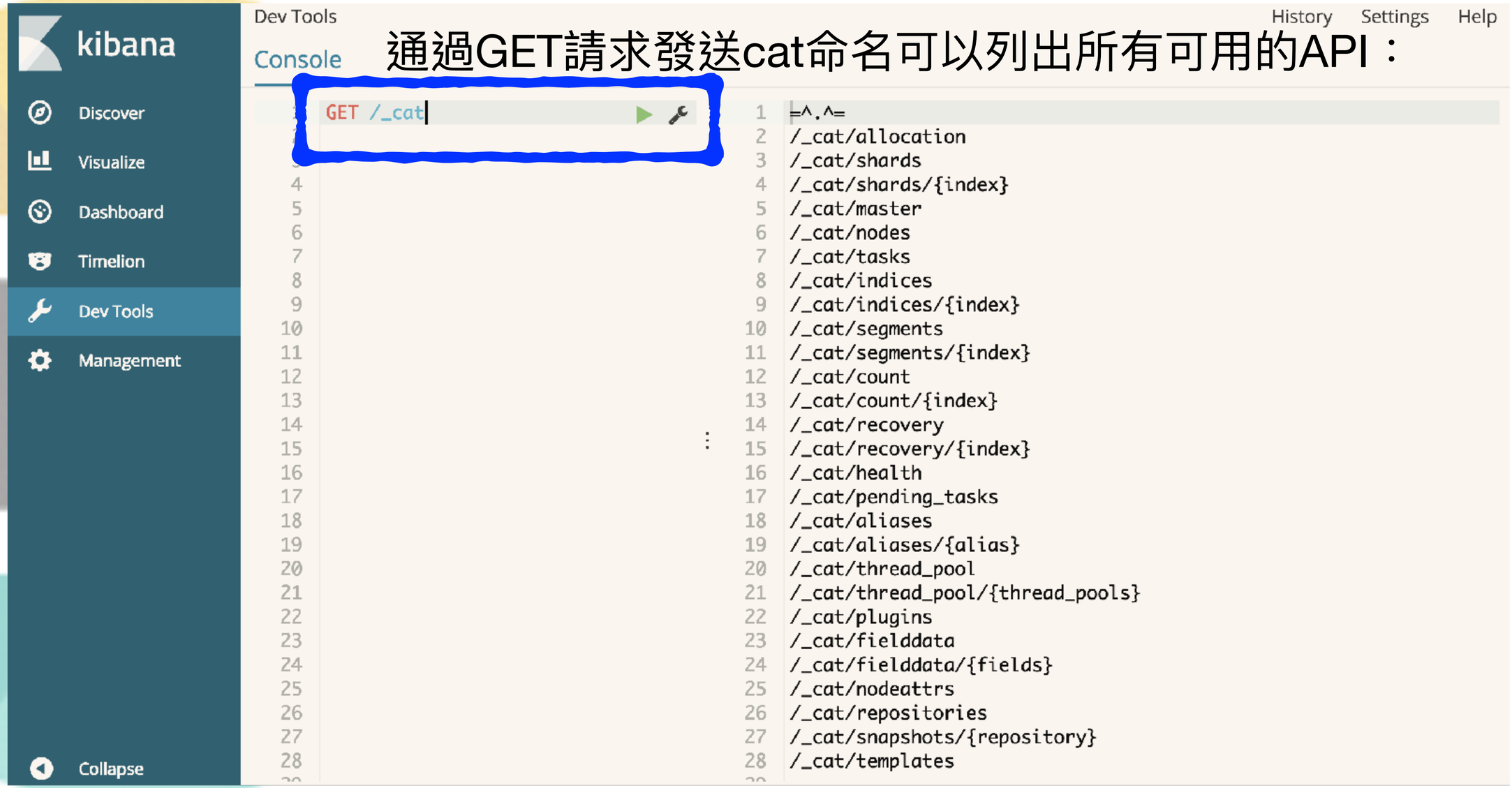



kibana
Discover
Visualize
Dashboard
Timelion
Dev Tools
Management
Collapse
Add a filter +
*
Selected Fields
t _id
t dropoff_l...
t dropoff_l...
t dropoff_l...
Available Fields
@timestamp
t @version
t _index
_score
t _type
t dropoff_date...
t host
t id
t message
January 21st 2018, 00:00:00.000 - January 27th 2018, 23:59:59.999 —
Auto
1,500,000
1,000,000
500,000
0
Count
2018-01-22 08:00
2018-01-23 08:00
2018-01-24 08:00
2018-01-25 08:00
2018-01-26 08:00
2018-01-27 08:00
@timestamp per 3 hours
Time
_id
dropoff_location
dropoff_latitude
dropoff_longitude
January 24th 2018, 16:15:46.509 Lwg8J2EB4zpcysM7wDJZ (40.7653656, -73.96102905) 40.7653656 -73.96102905
January 24th 2018, 16:15:46.509 PAg8J2EB4zpcysM7wDJZ (40.70813751, -74.01539612) 40.70813751 -74.01539612
January 24th 2018, 16:15:46.509 Qgg8J2EB4zpcysM7wDJZ (40.75680923, -73.97106934) 40.75680923 -73.97106934
January 24th 2018, 16:15:46.509 Uwg8J2EB4zpcysM7wDJZ (40.74126053, -73.99414825) 40.74126053 -73.99414825

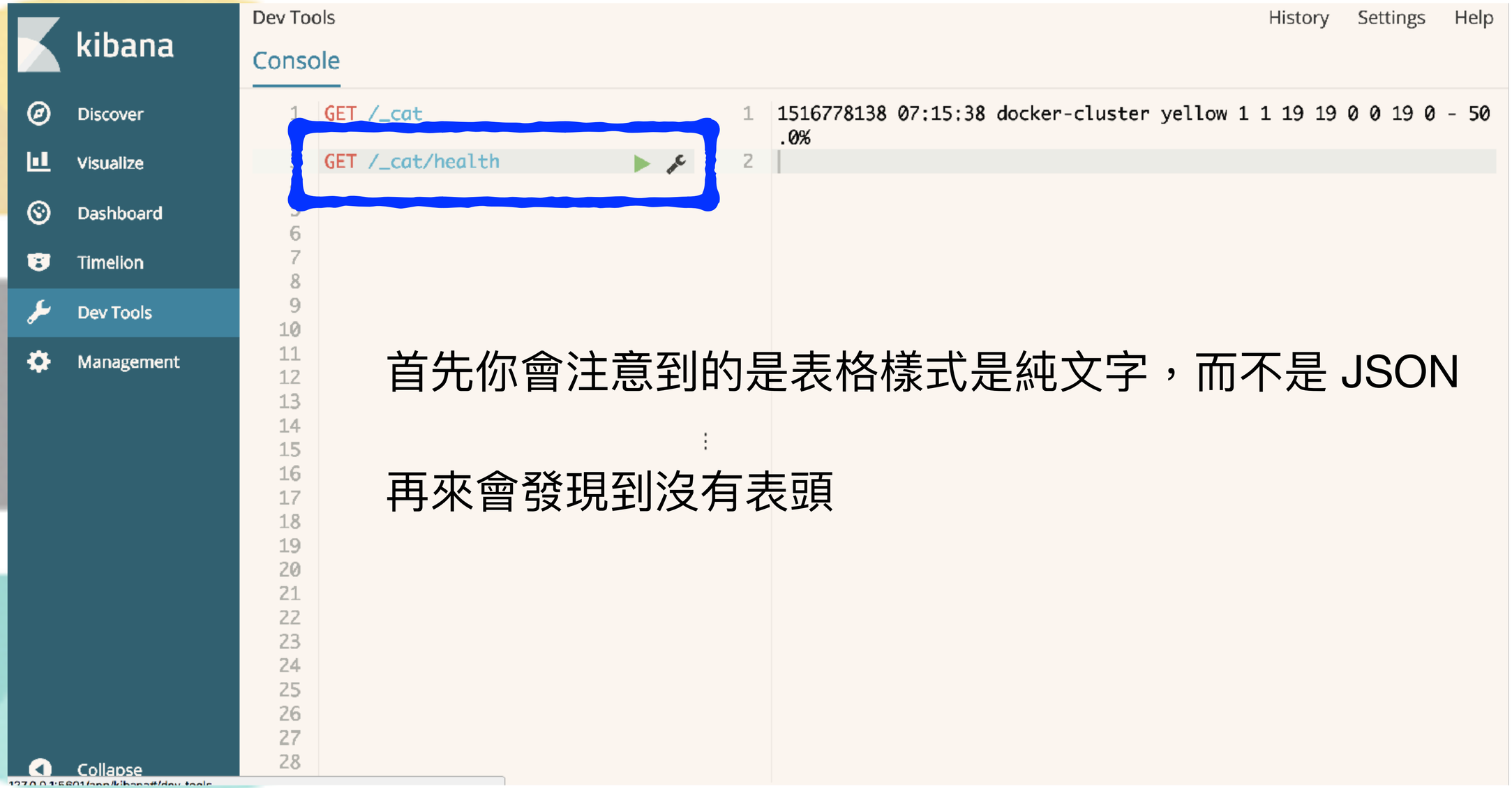

kibana
Discover
Visualize
Dashboard
Timelion
Dev Tools
Management
Status: Red
vickie
Heap Total
71.63 MB
Heap Used
62.25 MB
Load
5.03, 7.30, 7.61
重新投胎！啊不對~重新啟動
Response Time Avg
30527.00 ms
Response Time Max
30527.00 ms
Requests Per Second
0.00
Status Breakdown
ID
Status
plugin:kibana@6.1.2
Ready
plugin:elasticsearch@6.1.2
Request Timeout after 3000ms
plugin:console@6.1.2
Ready

Cat API

用 Linux 的 cat 命令命名，輸出以表格的形式提供，而不是JSON。當只是想瀏覽一遍集群或者找出內存使用偏高的節點時，對於系統管理員來說這是非常方便的



通過GET請求發送cat命名可以列出所有可用的API：


kibana
Discover
Visualize
Dashboard
Timelion
Dev Tools
Management
Collapse
Dev Tools
History
Settings
Help
Console
GET /_cat
GET /_cat/health
1516778138 07:15:38 docker-cluster yellow 1 1 19 19 0 0 19 0 - 50.0%
首先你會注意到的是表格樣式是純文字，而不是 JSON
再來會發現到沒有表頭

要啟用表頭，加上 ?v 即可


kibana
Discover
Visualize
Dashboard
Timelion
Dev Tools
Management
Collapse
Dev Tools
History
Settings
Help
Console
GET /_cat
GET /_cat/health
GET /_cat/health?v
epoch timestamp cluster status node.total node.data shards pri relo init unassign pending_tasks max_task_wait_time active_shards_percent
1516778268 07:17:48 docker-cluster yellow 1 1 19 19 0 0 19 0 - 50.0%

看現有的 Logstash 有哪些

檢索 Data


kibana
Discover
Visualize
Dashboard
Timelion
Dev Tools
Management
Collapse
Dev Tools
History
Settings
Help
Console
GET /_cat
GET /_cat/health
GET /_cat/health?v
GET /_cat/nodes?v
GET /_cat/nodes?help
GET /_cat/nodes?v&h=ip,port,heapPercent,heapMax
GET /_cat/indices?v
GET _search
{
"query": {
"match_all": {}
}
}
{
"took": 23,
"timed_out": false,
"_shards": {
"total": 19,
"successful": 19,
"skipped": 0,
"failed": 0
},
"hits": {
"total": 3556449,
"max_score": 1,
"hits": [
{
"_index": ".kibana",
"_type": "doc",
"_id": "config:6.1.2",
"_score": 1,
"_source": {
"type": "config",
"updated_at": "2018-01-23T06:50:26.338Z",
"config": {
"buildNum": 16363,
"defaultIndex": "afee8430-0009-11e8-a196-cfbf7ba467d8"
}
}
},

定義檢索 Data 方式

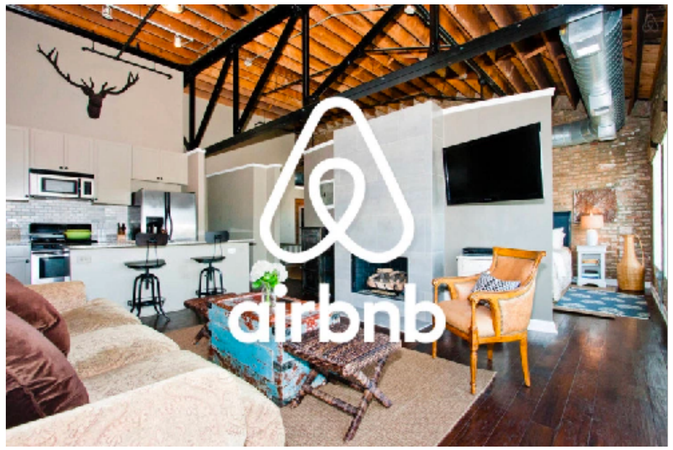

kibana
Discover
Visualize
Dashboard
Timelion
Dev Tools
Management
Collapse
Dev Tools
History
Settings
Help
Console
GET /_cat
GET /_cat/health
GET /_cat/health?v
GET /_cat/nodes?v
GET /_cat/nodes?help
GET /_cat/nodes?v&h=ip,port,heapPercent,heapMax
GET /_cat/indices?v
GET _search
{
"query": {
"match_all": {}
}
}
GET /logstash-2018.01.23/_search
{
"query": {"match":{
"id":"gxn3p5htnn"
}}
}
{
"took": 53,
"timed_out": false,
"_shards": {
"total": 5,
"successful": 5,
"skipped": 0,
"failed": 0
},
"hits": {
"total": 8,
"max_score": 11.518404,
"hits": [
{
"_index": "logstash-2018.01.23",
"_type": "doc",
"_id": "LdFjImEB4zpcysM7mb92",
"_score": 11.518404,
"_source": {
"path": "/Users/vickieliu/Downloads/train_users_2.csv",
"signup_flow": "0",
"date_account_created": "2010/6/28",
"affiliate_provider": "direct",
"gender": "-unknown-",
"date_first_booking": null,
"@timestamp": "2018-01-23T09:40:04.935Z",
"id": "gxn3p5htnn",

藉由查找出的資訊反檢索 Data

尋找特定 Data


kibana
Discover
Visualize
Dashboard
Timelion
Dev Tools
Management
Collapse
Dev Tools
History
Settings
Help
Console
GET _search
{
"query": {
"match_all": {}
}
}
GET /logstash-2018.01.23/_search
{
"query": {"match":{
"id":"gxn3p5htnn"
}}
}
GET /nyc/_search
{
"query": {
"match_all": {}
}
}
GET /nyc/doc/8tW6ImEB4zpcysM75k1c
{
"_index": "nyc",
"_type": "doc",
"_id": "8tW6ImEB4zpcysM75k1c",
"_version": 1,
"found": true,
"_source": {
"path": "/Users/vickieliu/Downloads/ny_taxi/train.csv",
"@version": "1",
"@timestamp": "2018-01-23T11:15:21.777Z",
"vendor_id": "2",
"pickup_longitude": "-73.95611572265625",
"id": "id3491576",
"store_and_fwd_flag": "N",
"dropoff_datetime": "2016-01-14 10:10:49",
"dropoff_latitude": "40.760963439941406",
"host": "liupinrude-MacBook-Pro.local",
"passenger_count": "1",
"message": "id3491576,2,2016-01-14 09:54:51,2016-01-14 10:10:49,1,-73.95611572265625,40.767623901367187,-73.970970153808594,40.760963439941406,N,958",
"pickup_datetime": "2016-01-14 09:54:51",
"dropoff_longitude": "-73.970970153808594",
"pickup_latitude": "40.767623901367187",
"trip_duration": "958"
}
}

/train_users.csv



Airbnb New User Bookings

Where will a new guest book their first travel experience?

在這個挑戰中，會得到一個用戶列表以及他們的人口統計信息，網絡會話記錄和一些匯總統計信息。 被要求預測哪個國家是新用戶的第一個預訂目的地。 這個數據集中的所有用戶都來自美國。

目的地國有12種可能的結果：'US', 'FR', 'CA', 'GB', 'ES', 'IT', 'PT', 'NL','DE', 'AU', ‘NDF’ (no destination found), and 'other'.

沒有預定

/train_users.csv

id	date_account_create	timestamp_first_activ	date_first_booking	gender	age	signup_method	signup_flow
gxn3p5htnn	2010/6/28	2.01E+13		-unknown-		facebook	0
820tgsjxq7	2011/5/25	2.01E+13		MALE	38	facebook	0
4ft3gnwmtx	2010/9/28	2.01E+13	2010/8/2	FEMALE	56	basic	3
bjjt8pjhuk	2011/12/5	2.01E+13	2012/9/8	FEMALE	42	facebook	0
87mebub9p4	2010/9/14	2.01E+13	2010/2/18	-unknown-	41	basic	0
osr2jwljor	2010/1/1	2.01E+13	2010/1/2	-unknown-		basic	0
lsw9q7uk0j	2010/1/2	2.01E+13	2010/1/5	FEMALE	46	basic	0

/train_users.csv

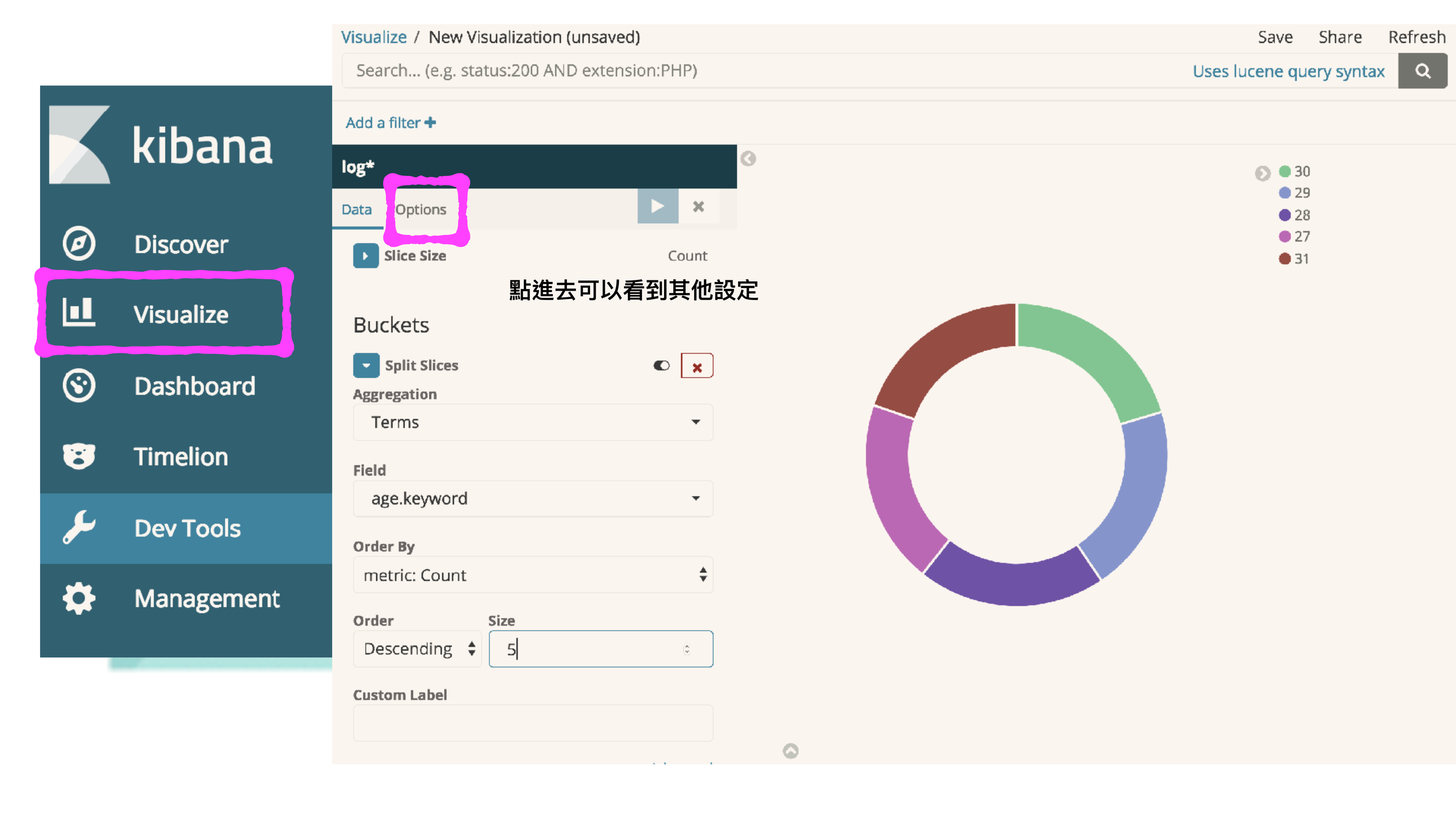
language	affiliate_channel	affiliate_provider	first_affiliate_tracked	signup_app	first_device_type	first_browser	country_destination
en	direct	direct	untracked	Web	Mac Desktop	Chrome	NDF
en	seo	google	untracked	Web	Mac Desktop	Chrome	NDF
en	direct	direct	untracked	Web	Windows Desktop	IE	US
en	direct	direct	untracked	Web	Mac Desktop	Firefox	other
en	direct	direct	untracked	Web	Mac Desktop	Chrome	US
en	other	other	omg	Web	Mac Desktop	Chrome	US
en	other	craigslist	untracked	Web	Mac Desktop	Safari	US

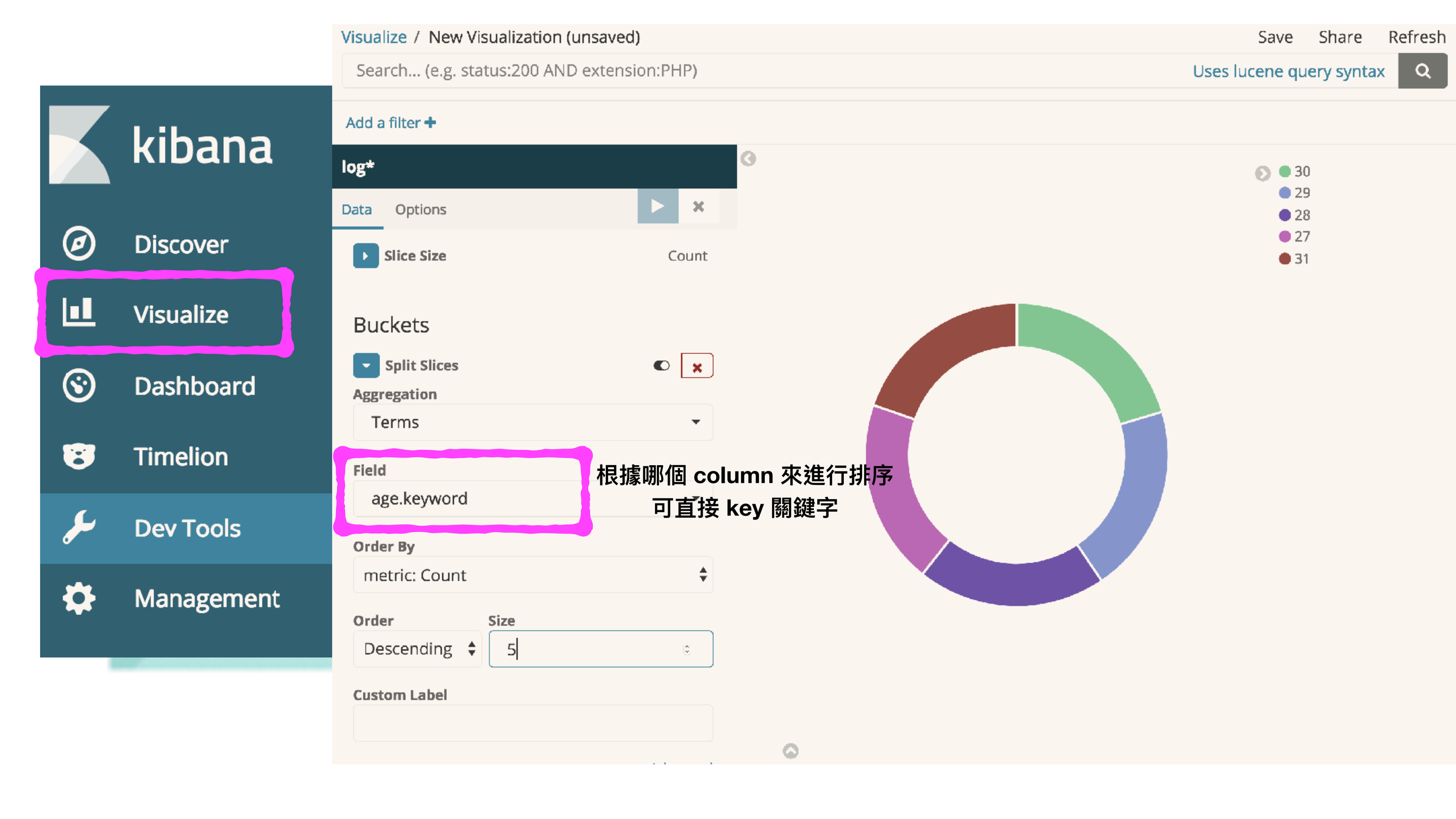
/test1.conf

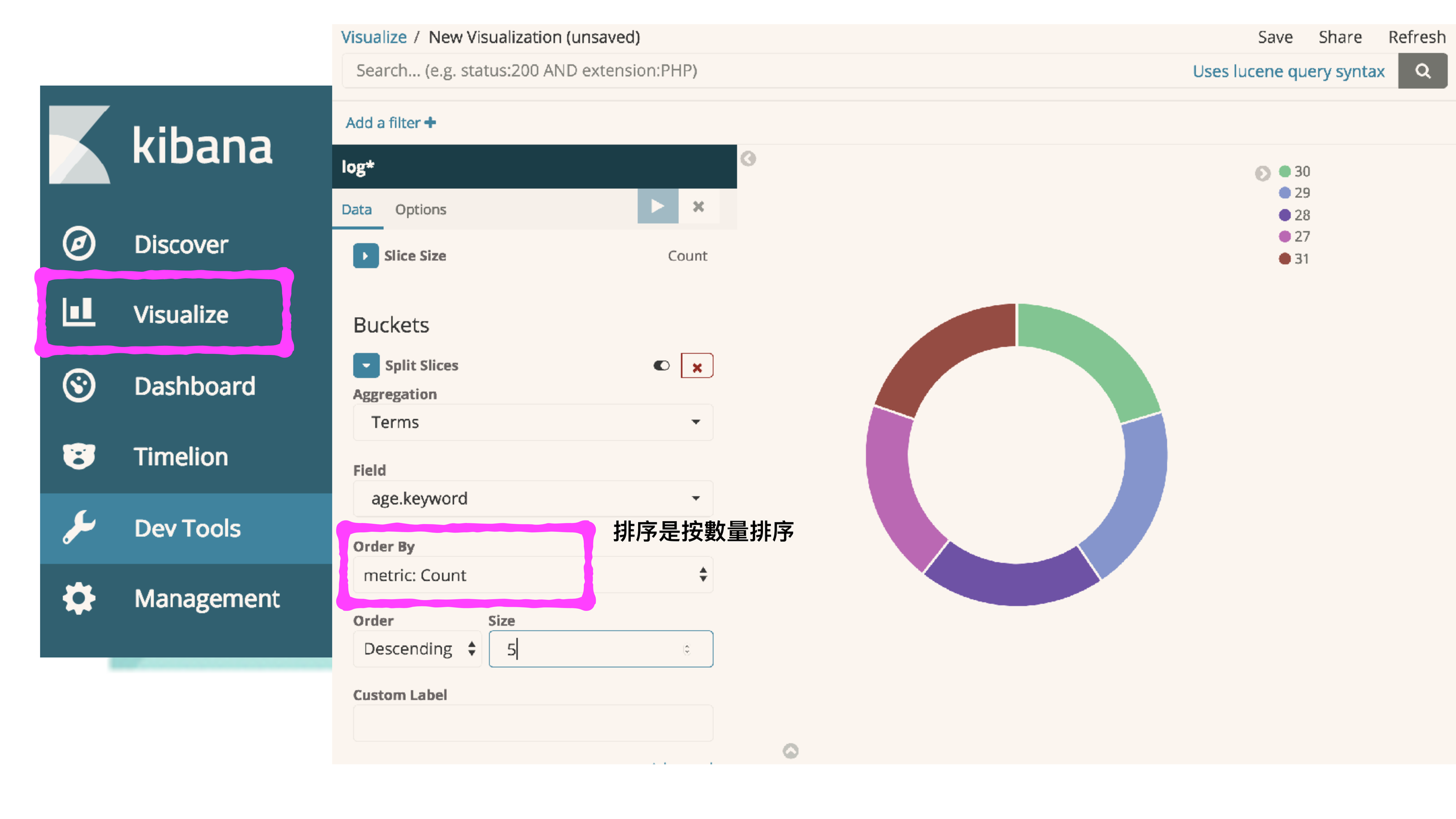
```
input {
    file {
        path => "/Users/vickieliu/Downloads/train_users.csv"
        start_position => "beginning"
        sincedb_path => "/dev/null"
    }
}
filter {
    csv {
        separator => ","
        columns => ["id", "date_account_created", "timestamp_first_active", "date_first_booking", "gender",
        "age", "signup_method", "signup_flow", "language", "affiliate_channel", "affiliate_provider",
        "first_affiliate_tracked", "signup_app", "first_device_type", "first_browser", "country_destination"]
  }

}
output {
    elasticsearch {
            hosts => ["localhost:9200"]  #"http://localhost:9200"
    }
        stdout {
            codec => rubydebug{ }
    }
}

```

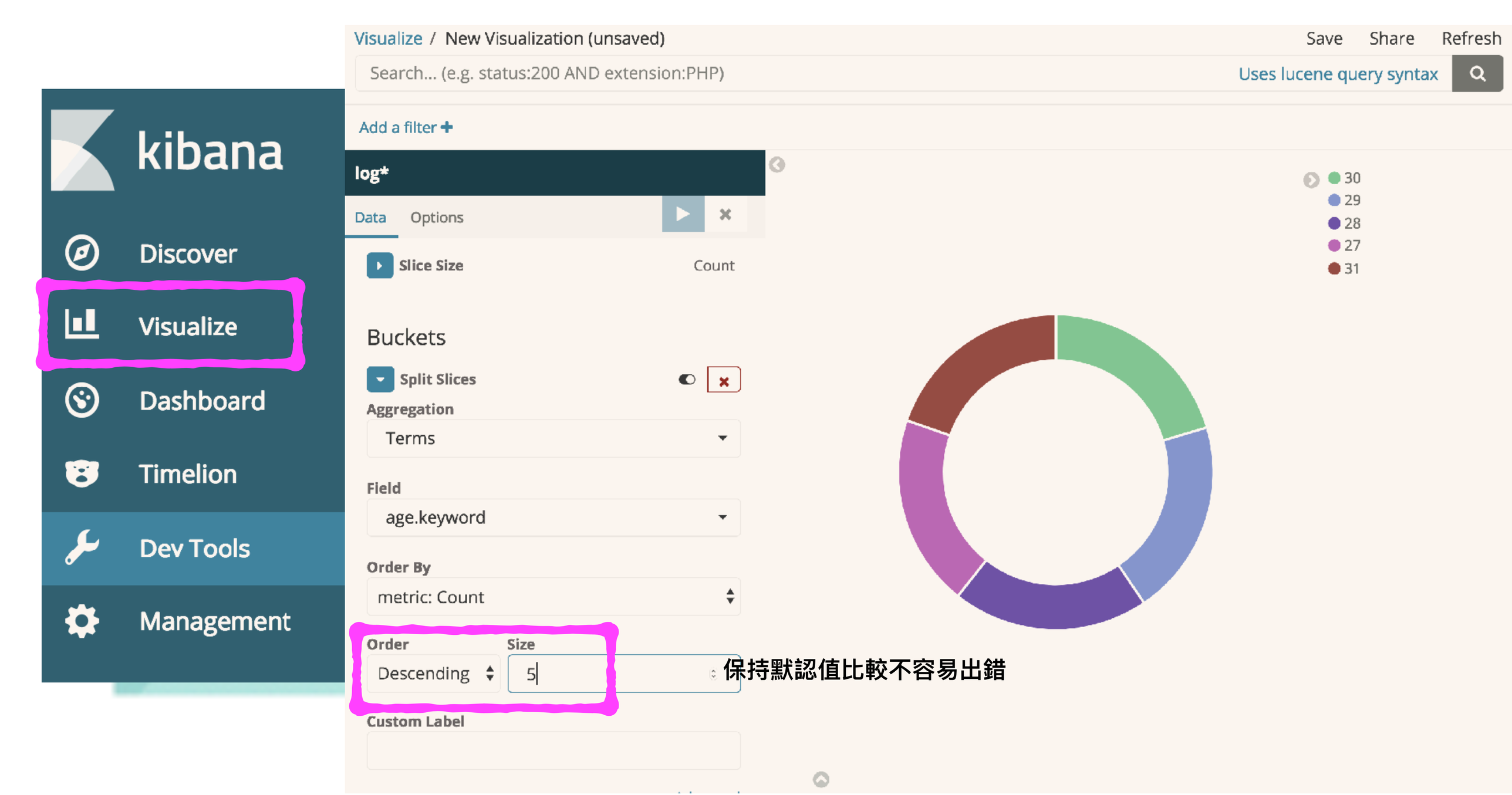
Shutdown Logstash

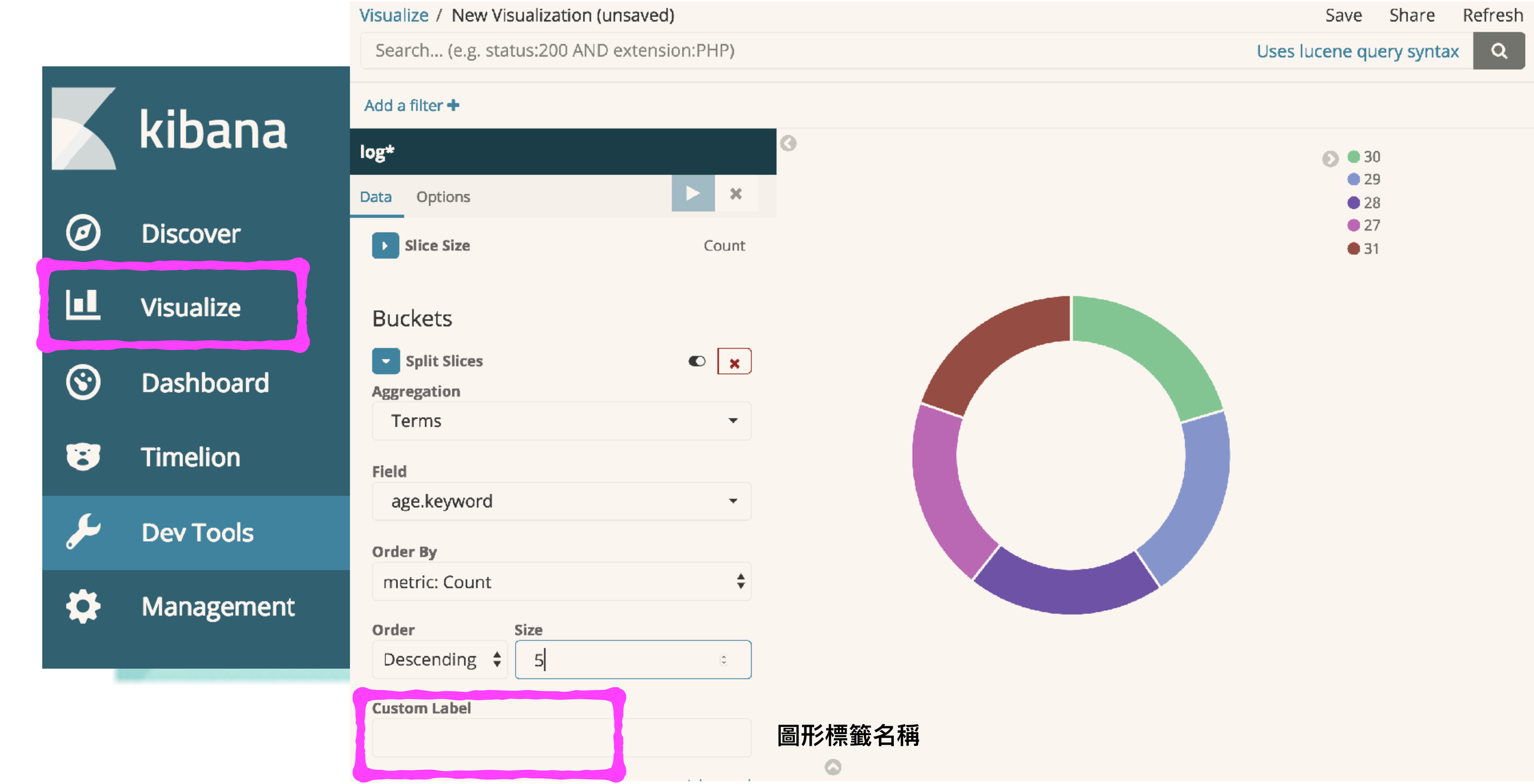
```
liupinrude-MacBook-Pro:logstash-6.1.2 vickieliu$ ps aux | grep logstash
vickieliu         3856  10.0  5.5  6965588 457852 s003  S+    2:38下午   1:28.47 /usr/bin/java -XX:+UseParNewGC -
XX:+UseConcMarkSweepGC -XX:CMSInitiatingOccupancyFraction=75 -XX:+UseCMSInitiatingOccupancyOnly -Djava.awt.headle
ss=true -Dfile.encoding=UTF-8 -Djruby.compile.invokedynamic=true -Djruby.jit.threshold=0 -XX:+HeapDumpOnOutOfMemo
ryError -Djava.security.egd=file:/dev/urandom -Xmx1g -Xms1g -Xss2048k -Djffi.boot.library.path=/Users/vickieliu/D
eveloper/resources/logstash-6.1.2/vendor/jruby/lib/jni -Dfile.encoding=UTF-8 -Xbootclasspath/a:/Users/vickieliu/D
eveloper/resources/logstash-6.1.2/vendor/jruby/lib/jruby.jar -classpath : -Djruby.home=/Users/vickieliu/Developer
/resources/logstash-6.1.2/vendor/jruby -Djruby.lib=/Users/vickieliu/Developer/resources/logstash-6.1.2/vendor/jru
by/lib -Djruby.script=jruby -Djruby.shell=/bin/sh org.jruby.Main /Users/vickieliu/Developer/resources/logstash-6.
1.2/lib/bootstrap/environment.rb logstash/runner.rb -f /Users/vickieliu/Developer/resources/logstash-6.1.2/conf.d
/test8.conf
vickieliu         3885   0.0  0.0  4276968    884 s002  S+    2:40下午   0:00.01 grep logstash
liupinrude-MacBook-Pro:logstash-6.1.2 vickieliu$ kill -9 3856
liupinrude-MacBook-Pro:logstash-6.1.2 vickieliu$
```

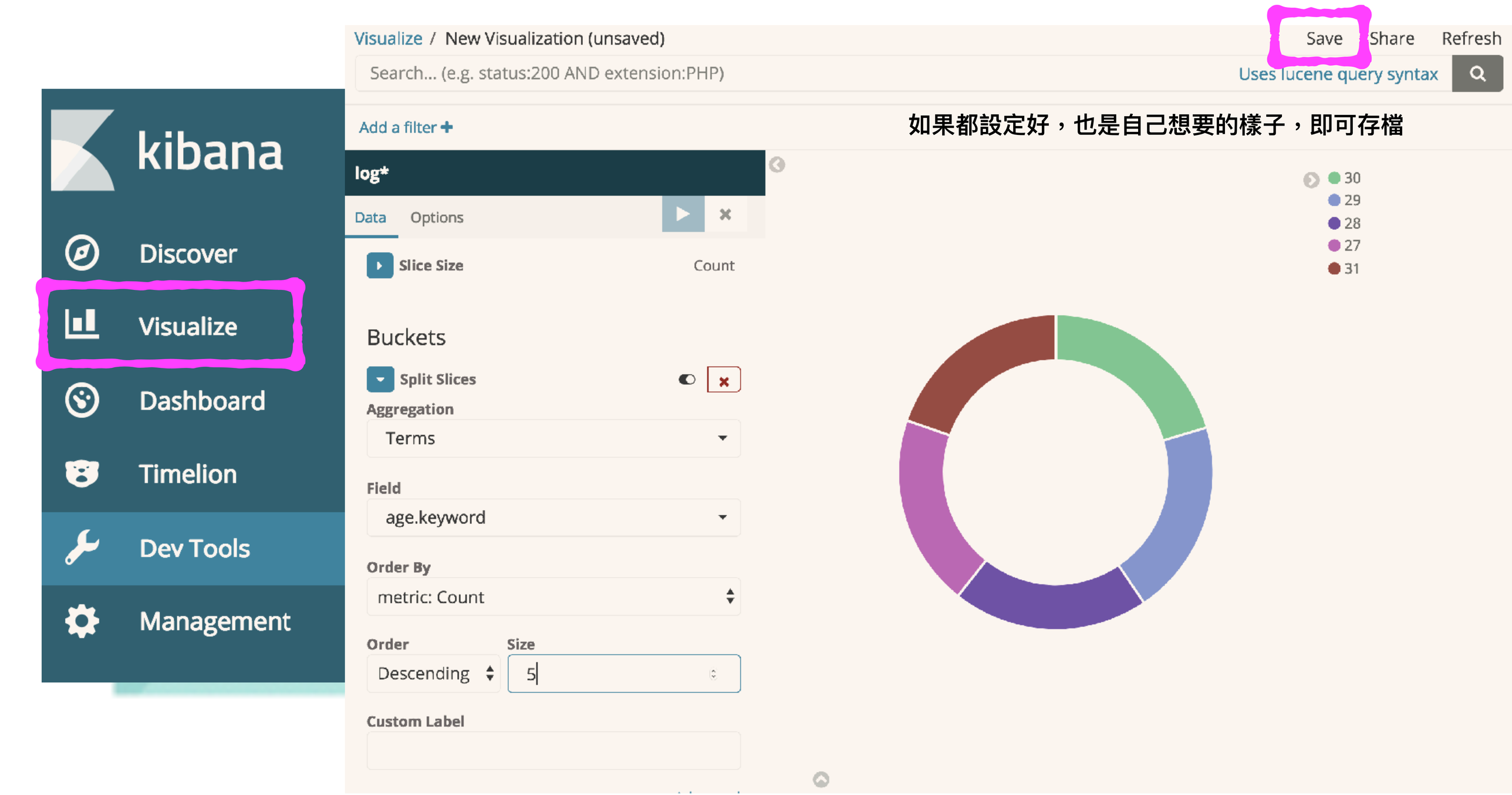

kibana
Discover
Visualize
Dashboard
Timelion
Dev Tools
Management
Visualize / New Visualization (unsaved)
Save
Share
Refresh
Search... (e.g. status:200 AND extension:PHP)
Uses lucene query syntax
Add a filter +
log*
Data
Options
Slice Size
Count
點進去可以看到其他設定
Buckets
Split Slices
Aggregation
Terms
Field
age.keyword
Order By
metric: Count
Order
Descending
Size
5
Custom Label
30
29
28
27
31


kibana
Discover
Visualize
Dashboard
Timelion
Dev Tools
Management
Visualize / New Visualization (unsaved)
Save
Share
Refresh
Search... (e.g. status:200 AND extension:PHP)
Uses lucene query syntax
Add a filter +
log*
Data
Options
Slice Size
Count
Buckets
Split Slices
Aggregation
Terms
Field
age.keyword
Order By
metric: Count
Order
Descending
Size
5
Custom Label
30
29
28
27
31
根據哪個 column 來進行排序
可直接 key 關鍵字


kibana
Discover
Visualize
Dashboard
Timelion
Dev Tools
Management
Visualize / New Visualization (unsaved)
Save
Share
Refresh
Search... (e.g. status:200 AND extension:PHP)
Uses lucene query syntax
Add a filter +
log*
Data
Options
Slice Size
Count
Buckets
Split Slices
Aggregation
Terms
Field
age.keyword
Order By
metric: Count
Order
Descending
Size
5
Custom Label
30
29
28
27
31
排序是按数量排序


kibana
Discover
Visualize
Dashboard
Timelion
Dev Tools
Management
Visualize / New Visualization (unsaved)
Save
Share
Refresh
Search... (e.g. status:200 AND extension:PHP)
Uses lucene query syntax
Add a filter +
log*
Data
Options
Slice Size
Count
Buckets
Split Slices
Aggregation
Terms
Field
age.keyword
Order By
metric: Count
Order
Descending
Size
5
Custom Label
30
29
28
27
31
保持默認值比較不容易出錯


Visualize / New Visualization (unsaved)
Save
Share
Refresh
Search... (e.g. status:200 AND extension:PHP)
Uses lucene query syntax
Add a filter +
kibana
Discover
Visualize
Dashboard
Timelion
Dev Tools
Management
log*
Data
Options
Slice Size
Count
Buckets
Split Slices
Aggregation
Terms
Field
age.keyword
Order By
metric: Count
Order
Descending
Size
5
Custom Label
30
29
28
27
31
圖形標籤名稱


Visualize / New Visualization (unsaved)
Save
Share
Refresh
Search... (e.g. status:200 AND extension:PHP)
Uses lucene query syntax
Add a filter +
如果都設定好，也是自己想要的樣子，即可存檔
kibana
Discover
Visualize
Dashboard
Timelion
Dev Tools
Management
log*
Data
Options
Slice Size
Count
Buckets
Split Slices
Aggregation
Terms
Field
age.keyword
Order By
metric: Count
Order
Descending
Size
5
Custom Label
30
29
28
27
31

ELK Stack (Elasticsearch + Logstash + Kibana)

可以幫你解決一些問題如下:

- 太長的 log 卻沒有一個方法可以整理跟分析它
- 太多來源的 log，每次查詢一個問題需要打開三四個檔案，但是往往還是不知道原因，得要人工查詢
- 想要透過系統的 log 去做一些商業智慧的分析，但是無從下手

Hands on！

建立一個路徑並將檔案放入

小提醒: 檔名與路徑盡可能都以英文小寫與 _ 符號作為間隔，
在讀取的時候比較不容易出現檔名格式不合的問題

Reference

ELK 介紹
https://oranwind.org/dv-elk-an-zhuang-ji-she-ding-jiao-xue/
讓 Logstash 從頭讀文件
https://elasticsearch.cn/article/11
Logstash Grok
https://www.elastic.co/guide/en/logstash/current/plugins-filters-grok.html
Kaggle - New York City Taxi Trip Duration
https://www.kaggle.com/c/nyc-taxi-trip-duration
Version Compatibility with Elasticsearch
https://github.com/elastic/kibana
Elasticsearch 簡介
https://www.slideshare.net/rueian3/elasticsearch-45855699
ELK教學
https://blog.johnwu.cc/article/how-to-install-elasticsearch-logstash-and-kibana-elk-stack-on-centos-red-hat.html
Docker @ Elastic
https://www.docker.elastic.co/#
Visualizing Logs Using ElasticSearch, Logstash and Kibana
https://www.youtube.com/watch?v=Kqs7UcCJquM
利用 Logstash，Elasticsearch 與 Kibana 來分析 log
http://www.evanlin.com/using-logstash-elsticsearch-and-kibana/

Reference

Hands on tutorial to perform Data Exploration using Elastic Search and Kibana (using Python)
https://www.analyticsvidhya.com/blog/2017/05/beginners-guide-to-data-exploration-using-elastic-search-and-kibana/
Elasticsearch 權威指南
https://es.xiaoleilu.com/index.html
Kibana + timelion: time series with the elastic stack
https://www.slideshare.net/swallez/kibana-timelion-time-series-with-the-elastic-stack
Use Logstash to load CSV into Elasticsearch
https://www.youtube.com/watch?v=rKy4sFbIZ3U
Logstash 最佳實踐
https://doc.yonyoucloud.com/doc/logstash-best-practice-cn/index.html
cat API
https://www.elastic.co/guide/cn/elasticsearch/guide/current/_cat_api.html

.
.
.