elasticsearch 在學校的應用

國立虎尾科技大學
江季翰

ES 在學校的應用

學習、培訓和技術能力的提升

- 資料收集和日誌集中
- non-SQL 資料庫學習
- realtime dashboard 分析

大數據分析的基礎課程

- 日誌採集解析部件： Logstash
- 基於Lucene的全文搜索引擎：Elasticsearch
- 分析視覺化平台：Kibana

Logstash

ETL

Elasticsearch

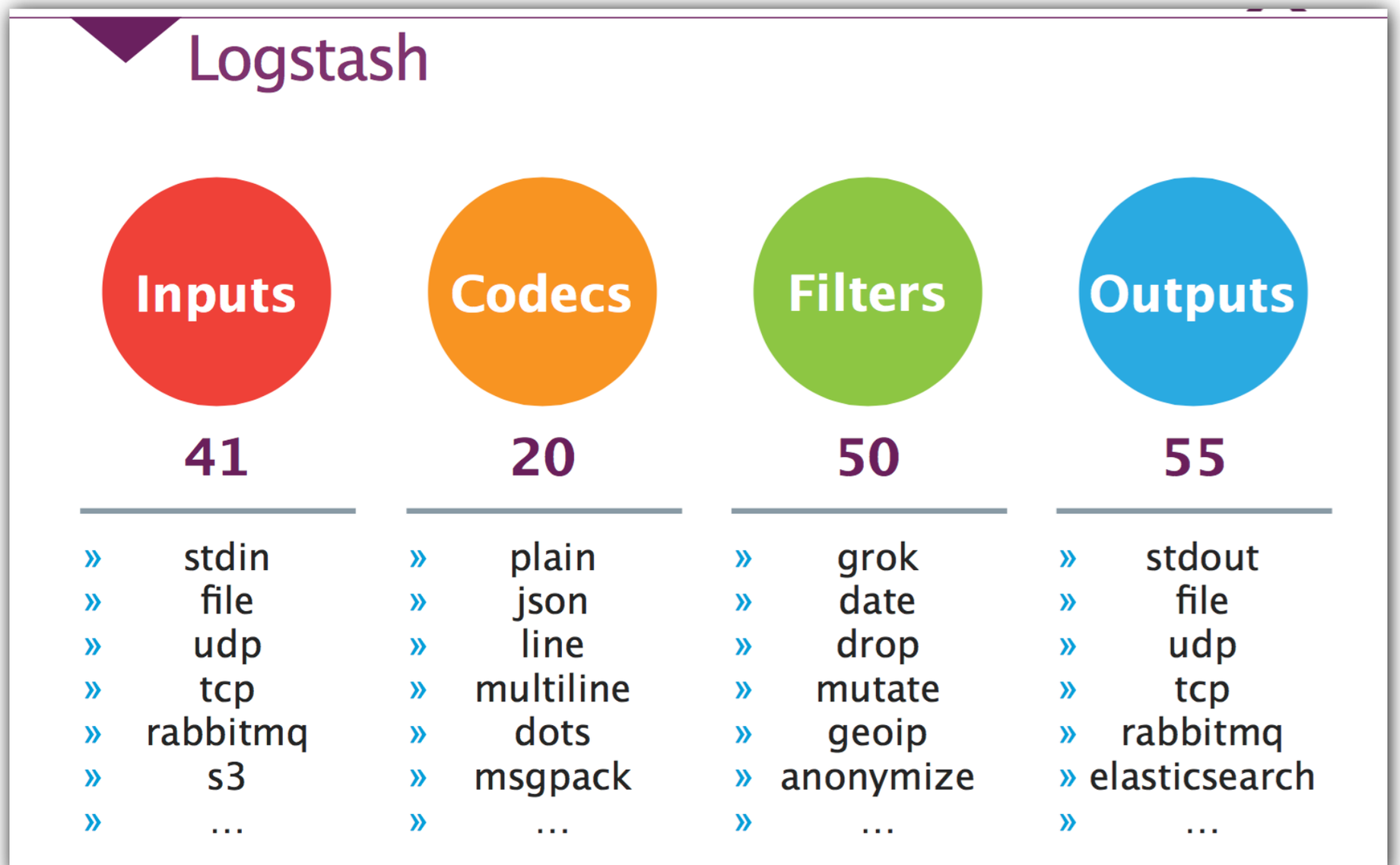
Stockage

Kibana

Visualisation

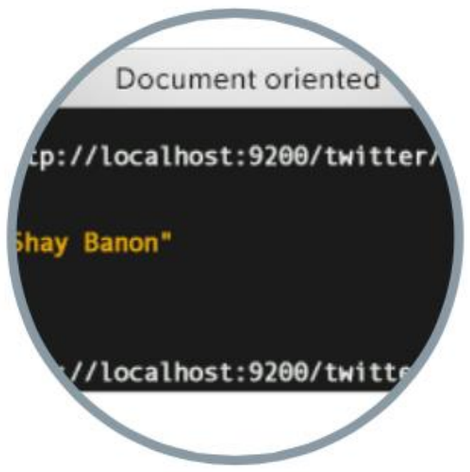
日誌的收集和解析工具

- 文本日誌／網路協議的收集
- 結構化/非結構化資料的解析
- 多種的輸入輸出插件和過濾語法

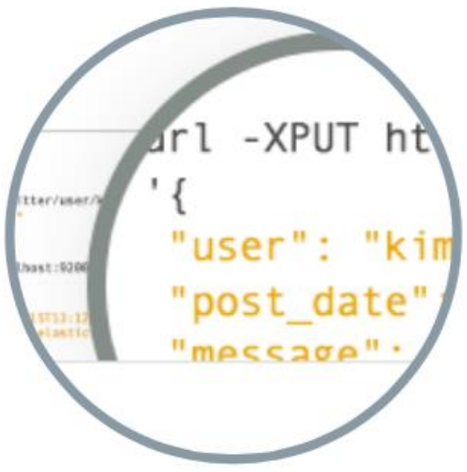


elasticsearch 基本功能特點

Elasticsearch

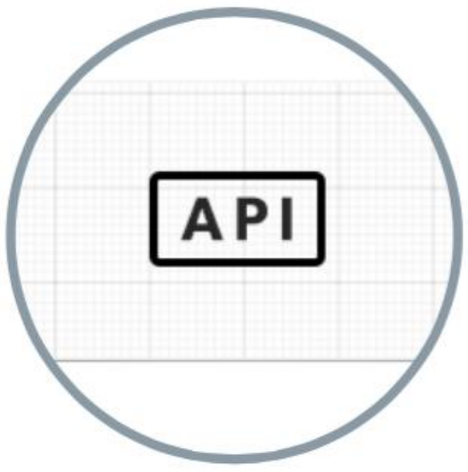


Document

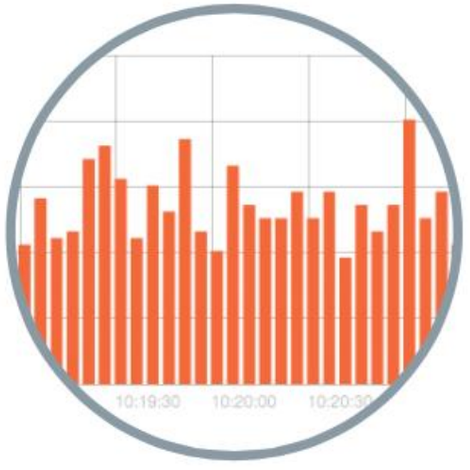


Schema Free

Full Text



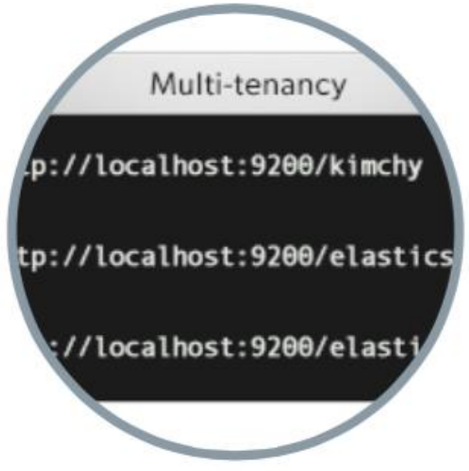
REST



Real Time

Distributed

HA



Multi-tenancy

Kibana 分析儀表板

Kibana

elasticsearch 資料流

日誌源 → 集中解析 → 全文檢索 → 資料可視化

非結構化日誌

APP日誌

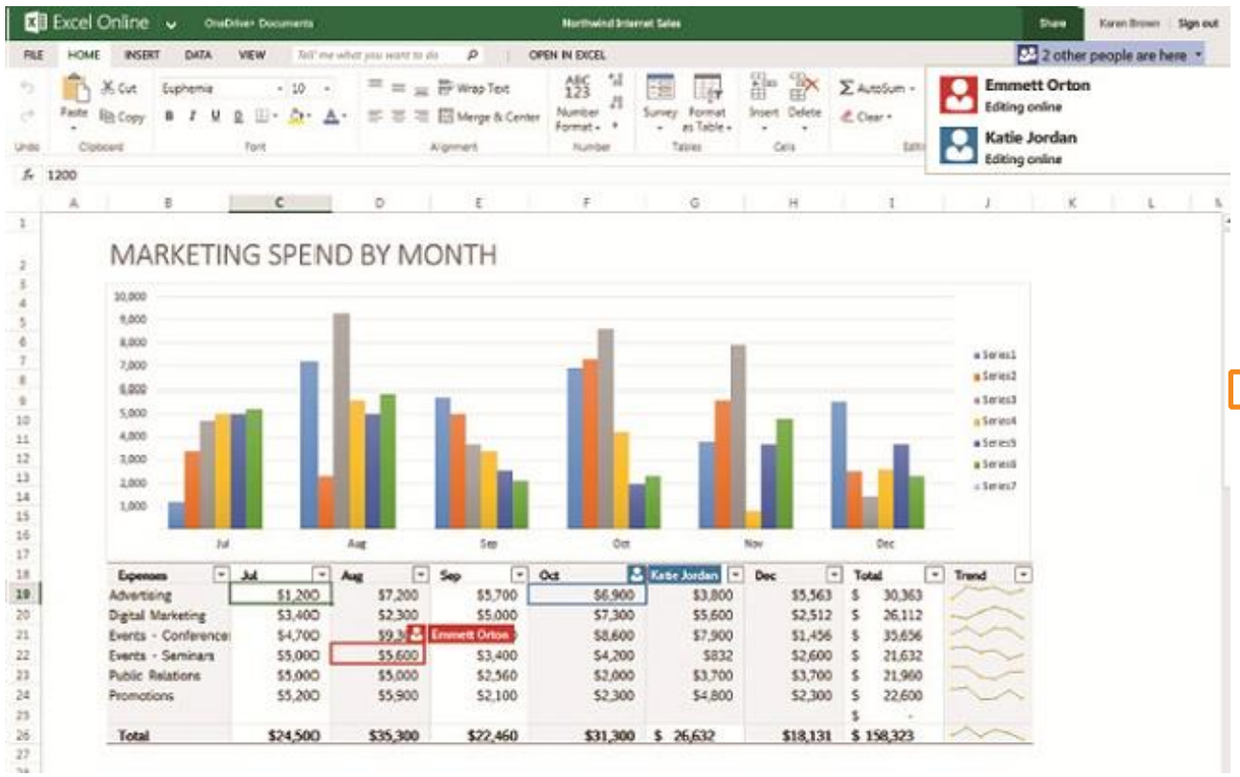
設備日誌

Elasticsearch
搜索引擎

資料分析展示方法的升級

- 從EXCEL 到 Kibana, 提升分析效率
- 日誌從幾10萬行到幾千萬行在ES 分析
- 可以做realtime展示和即時報表

EXCEL 報表



ES 資料即時檢索分析

Formosa雲端租屋生活網 - http://house.nfu.edu.tw/

- 利用ES 做雲端租屋生活網的大數據分析

Apache web 租屋網的運行的大數據分析

24小時內

日誌筆數

nfu-apache-筆數

17,937

Count

login 人數

nfu-apache-onlin

561

Unique count of geoip.ip

狀態404數量

nfu-apach

1,926

Count

狀態500數量

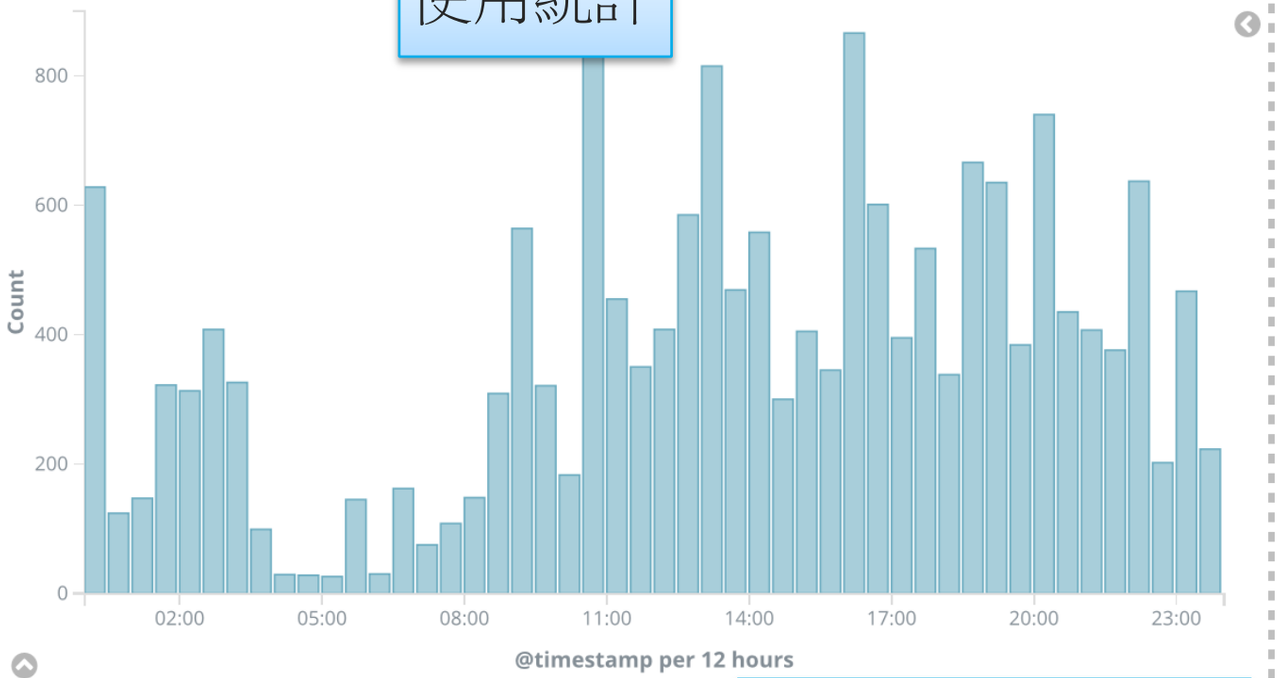
nfu-apac

123

Count

使用統計

nfu-apache-筆數統計



城市統計

nfu-apache-location-map

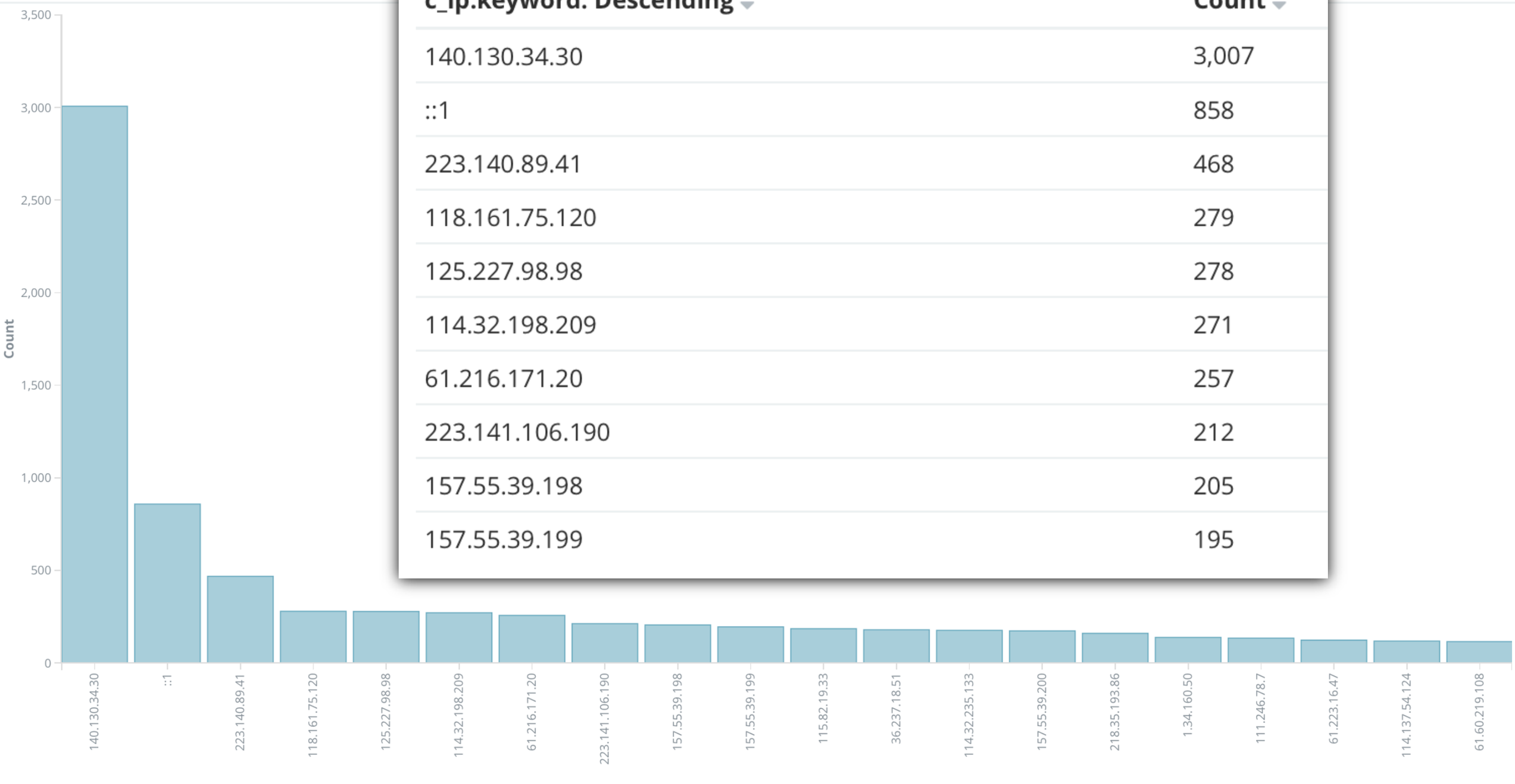
The container is too small to . Tags might be cropped or omitted.

用戶使用軌跡列表

nfu-apache-status清單

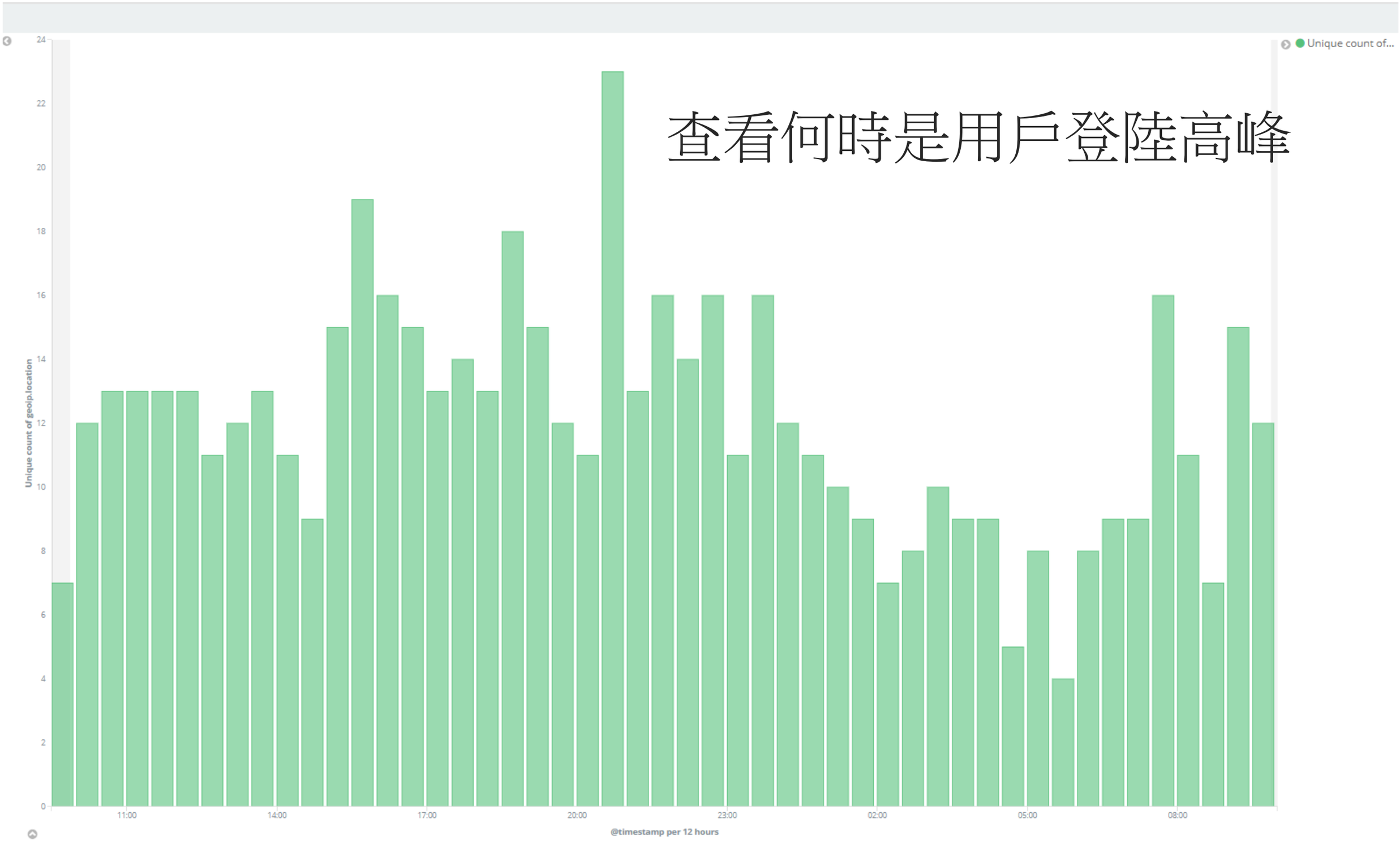
Time	cs_uri_stem_name	oip.ip	geoip.city_name	cs_status	cs_method
January 24th 2018, 23:59:37.000	/data/NHU/images_house/2228x2.jpg	111.248.5.234	Daxi	404	GET
January 24th 2018, 23:59:37.000	/data/NHU/images_house/2228x2.jpg	111.248.5.234	Daxi	404	GET
January 24th 2018, 23:58:55.000	/?page=viewnews&school=NTNU	157.55.39.199	Redmond	200	GET
January 24th 2018, 23:58:40.000	/house/CTUST/info/2785.html	157.55.39.199	Redmond	404	GET

訪問IP 的統計分析



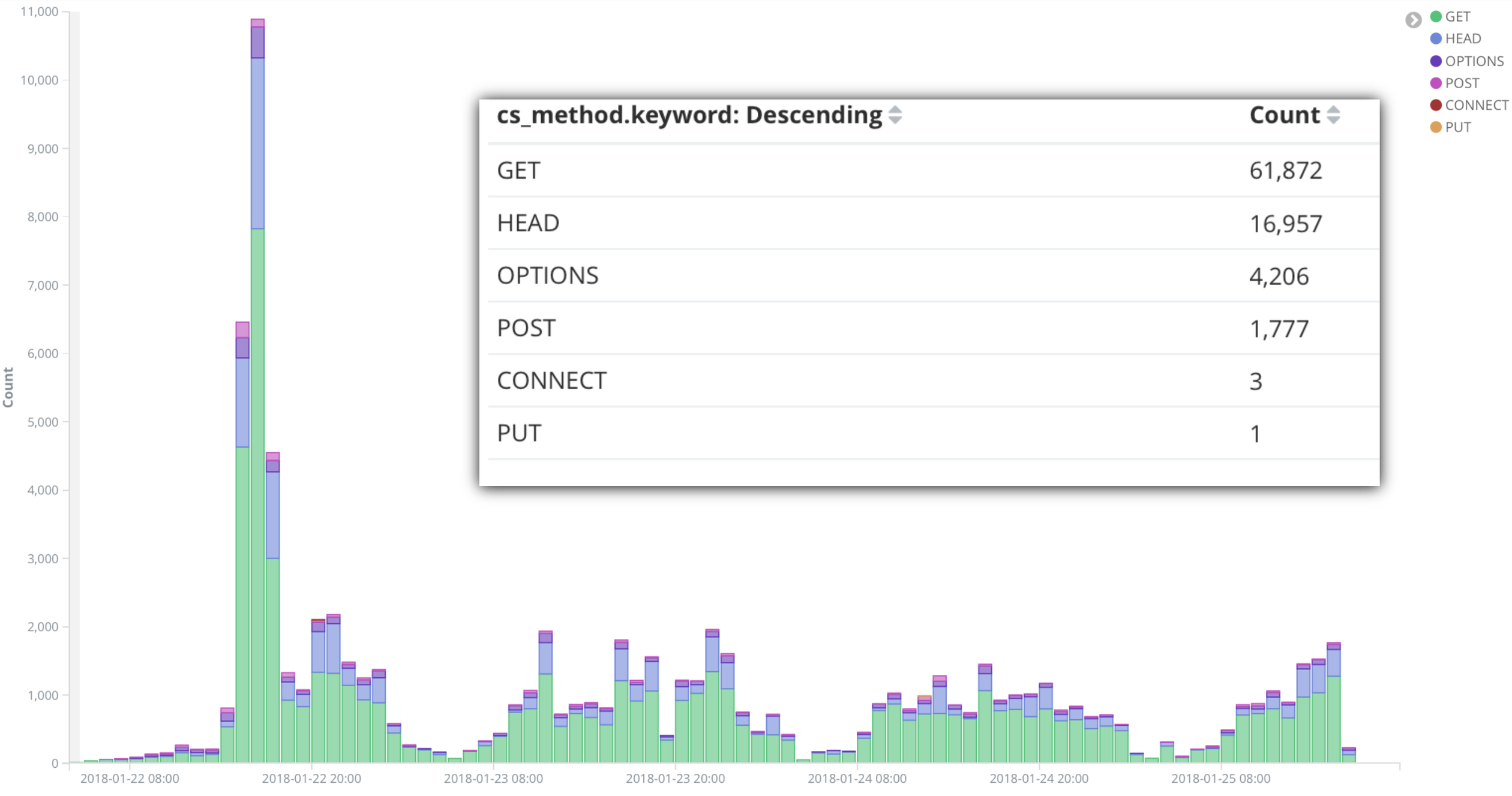
c_ip.keyword: Descending	Count
140.130.34.30	3,007
::1	858
223.140.89.41	468
118.161.75.120	279
125.227.98.98	278
114.32.198.209	271
61.216.171.20	257
223.141.106.190	212
157.55.39.198	205
157.55.39.199	195

Apache web 租屋網上線人數總數



訪問類別分析

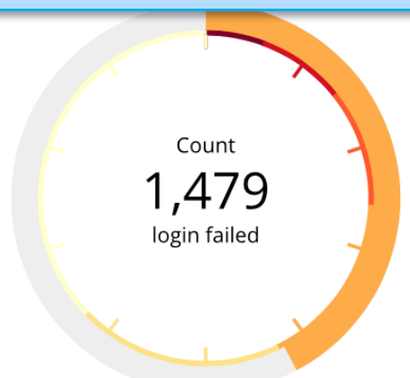
- GET 是正常的訪問行為
- HEAD/OPTION/POST/CONNECT/PUT 是不正常(特殊)的訪問



cs_method.keyword: Descending	Count
GET	61,872
HEAD	16,957
OPTIONS	4,206
POST	1,777
CONNECT	3
PUT	1

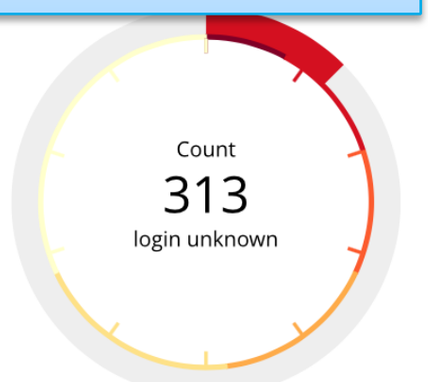
登入主機失敗分析報表

使用 root 帳號
登入失敗的次數



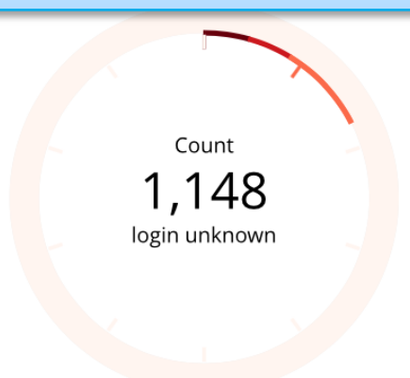
Count

使用存在帳號
登入失敗的次數



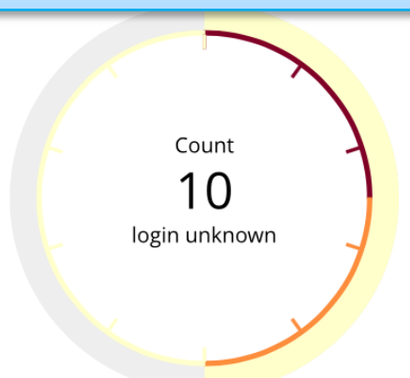
Count

使用不存在帳號
登入失敗的次數



Count

使用 SSH login 失敗
的次數



Count

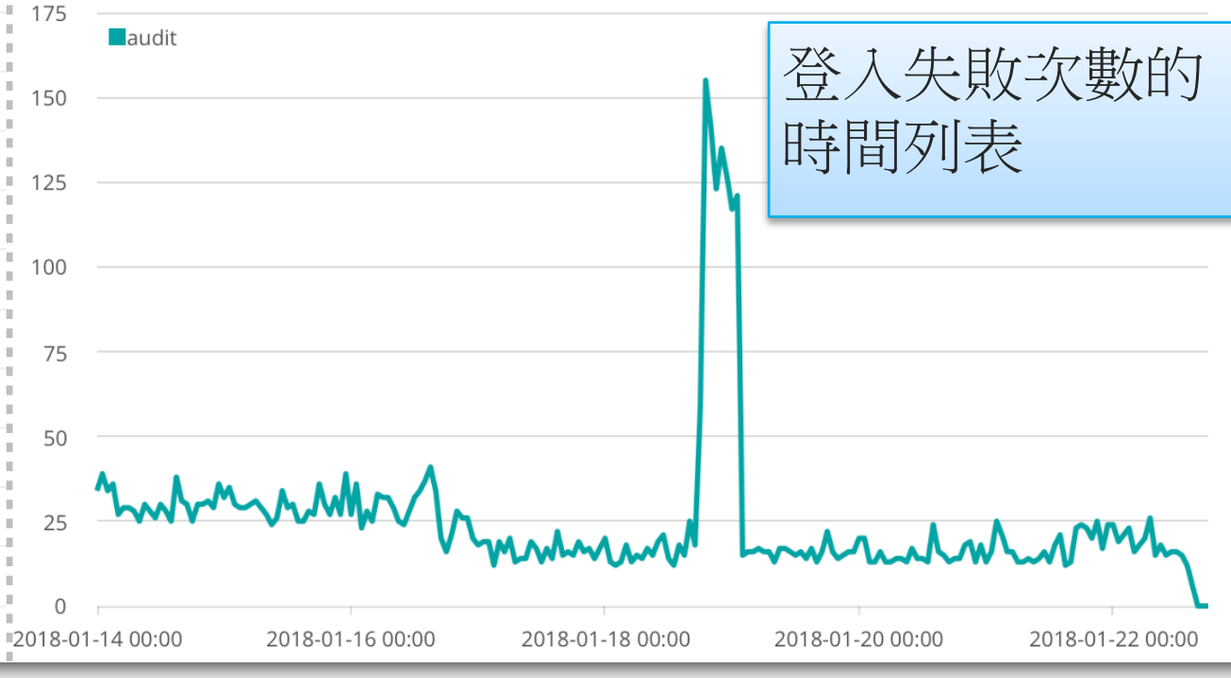
audit-failed-search

登入失敗次數的
軌跡列表

January 22nd 2018, 09:00:4			/usr/sbin/sshd
January 22nd 2018, 08:58:3			/usr/sbin/sshd
January 22nd 2018, 08:13:3			/usr/sbin/sshd
January 22nd 2018, 07:47:03.158	ipoc	?, (unknown)	/usr/sbin/sshd
January 22nd 2018, 07:46:36.699	ipoc	?, (unknown)	/usr/sbin/sshd
January 22nd 2018, 07:28:09.878	ipoc	?, (unknown)	/usr/sbin/sshd
January 22nd 2018, 07:15:32.129	ipoc	(unknown)	/usr/sbin/sshd
January 22nd 2018, 07:13:51.144	ipoc	(unknown)	/usr/sbin/sshd
January 22nd 2018, 07:13:24.959	ipoc	(unknown), ?	/usr/sbin/sshd
January 22nd 2018, 07:09:37.924	ipoc	?, (unknown)	/usr/sbin/sshd
January 22nd 2018, 07:07:27.774	ipoc	(unknown)	/usr/sbin/sshd

audit-all-timelion

登入失敗次數的
時間列表



防火牆Netflow 協議網路流量分析

總流量

asa total bytes

62,578,613,640

Sum of bytes

網路流量分析圖

asa 流量統計

asa top20 流量 interface

src_interface.keyword: Descending	Sum of bytes
OA	52,704,572,723
outside	4,234,470,625
FHN	3,146,392,318
DCN	1,558,684,916
	671,170,410
	263,322,648

asa top20 流量 dst interface

dst_interface.keyword: Descending	Sum of bytes
inside	27,780,786,742
DCN	17,536,878,992
outside	15,963,876,036
	501,593,187
	479,599,344
	280,694,498
	27,865,138
	7,319,703

不同網路區域的流量分析

流量最大的Top10 IP

src_ip.keyword: Descending	Sum of bytes
172.21.1.4	7,077,703,330
192.168.99.151	3,399,345,309
172.20.245.187	1,219,416,068
172.16.30.100	1,040,068,891
172.16.28.12	1,014,276,912
172.21.39.10	919,363,404
172.16.28.151	888,557,381
172.25.162.115	879,748,042
172.16.28.108	875,244,625
172.21.32.198	806,118,250

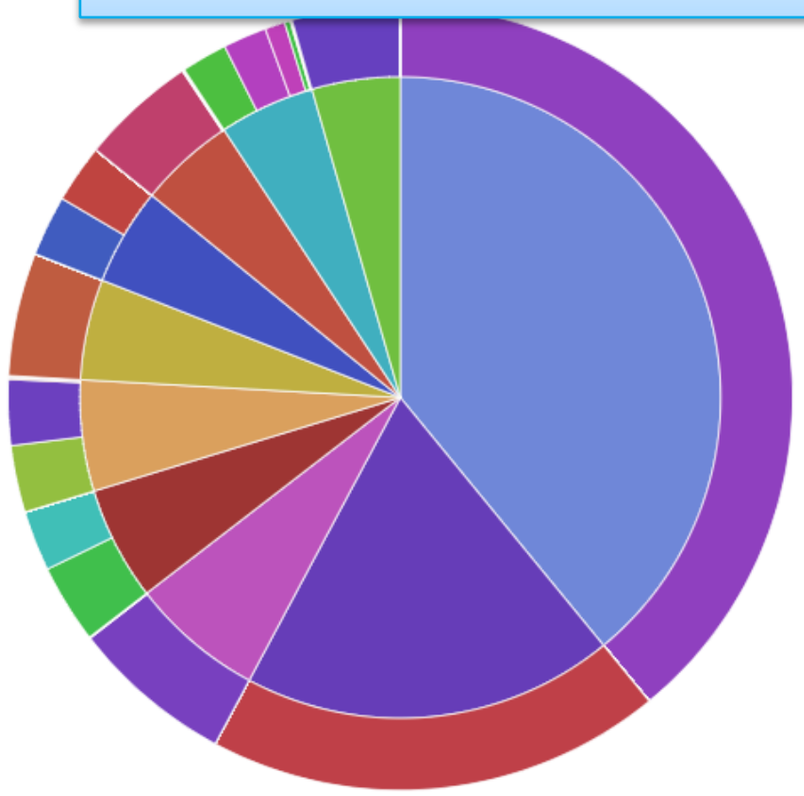
流量最大的Top10 USER

asa top20 流量 user

src_fwuser.keyword: Descending	Sum of bytes
LOCAL\ni	825,552,710
LOCAL\sa	92,653,761
LOCAL\al	91,257,183
LOCAL\ki	82,121,626
LOCAL\ar	72,047,632
LOCAL\xi	71,580,359
LOCAL\er	70,575,077
LOCAL\ar	67,176,252
LOCAL\le	65,635,328
LOCAL\ra	52,844,433

來源IP／目標IP 流量對比圖

asa-traffic top10-IP

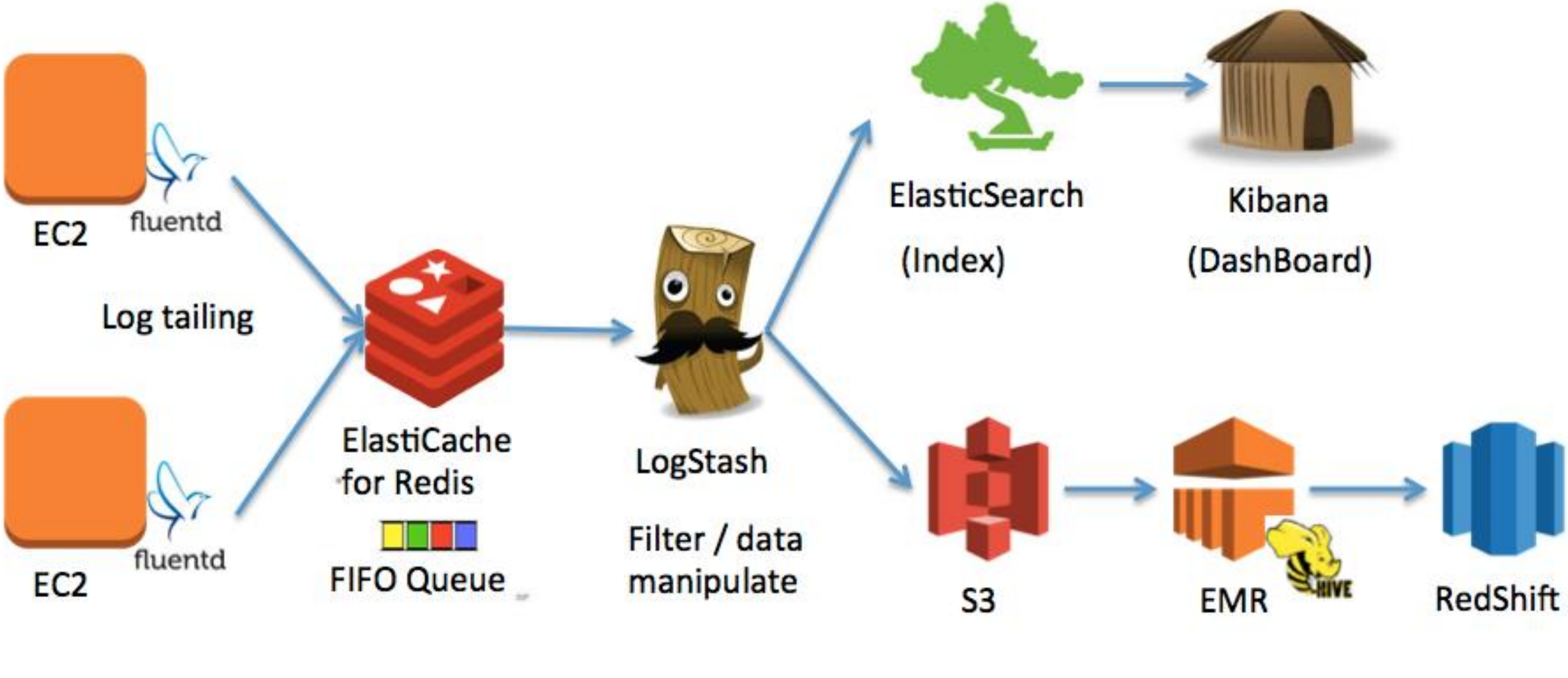


網路爬文和輿情分析

每日從社群軟體Facebook, tweets, blog 等收集從數百萬筆評論資料，保存到 Elasticsearch,進行查詢和大數據分析

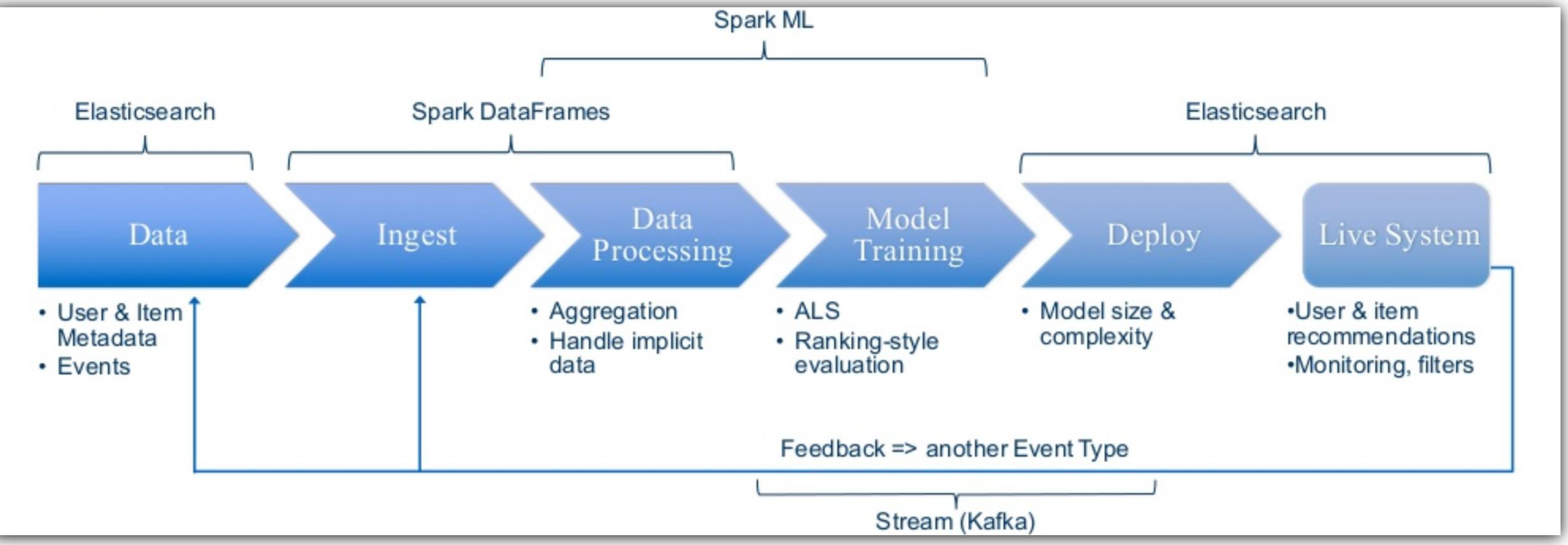
elasticsearch 結合大數據分析

- 日誌採集解析工具Logstash
- 基於Lucene的全文搜索引擎Elasticsearch
- 分析視覺化平台Kibana

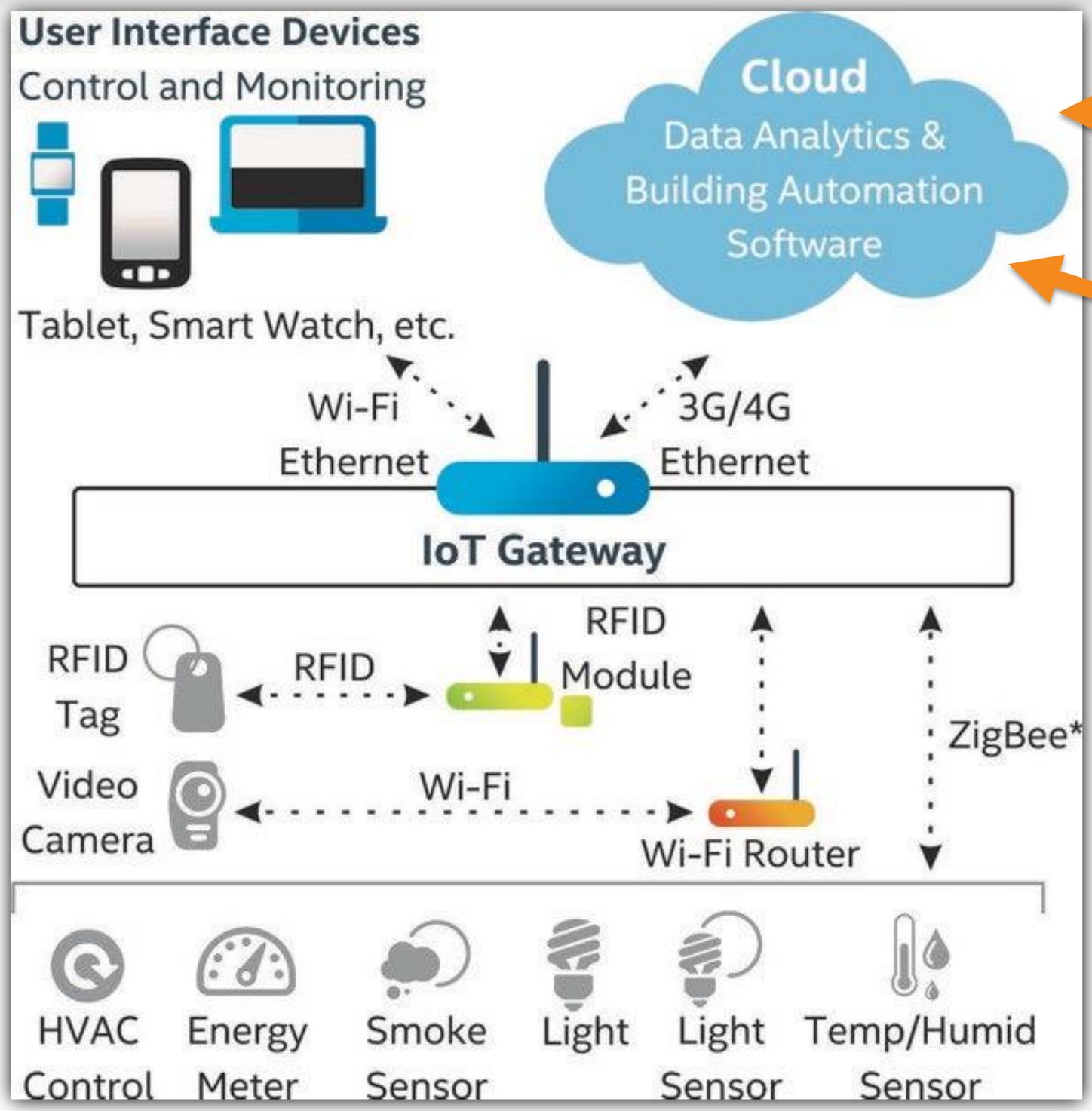


Machine Leaning 應用學習

- 整合Spark ML 和elasticsearch
  - 使用ES 進行大量資料的收集
  - 通過Spark 系統做資料處理，建模等



IoT 工業4.0 日誌收集和分析



- 資料收集
- 資料分析

elasticsearch.

elasticsearch.

- 利用ES 收集/分析／展示 IoT Gateway 上的所有 sensor 的所有log