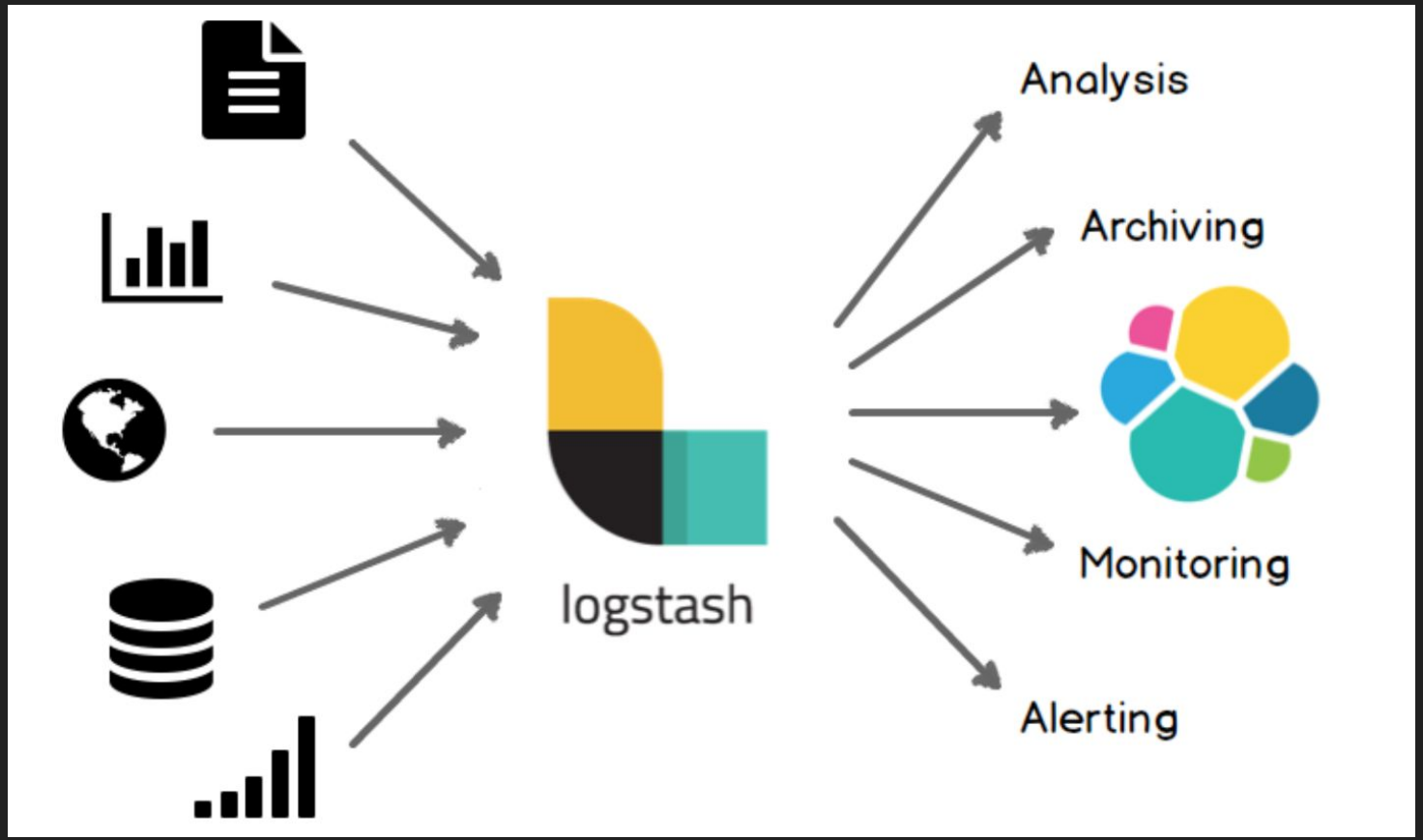
Logstash 資料收集

蔡宗憲 (Andrew)

Logstash 簡介


Analysis
Archiving
Monitoring
Alerting
logstash

Logstash 特點

- ELK stack
  - Elastic search
  - Logstash
  - Kibana
- 多種 plugin (擴充)
  - Input
  - Filter
  - Output
  - Codec
- 異質資料間的傳輸與接收

用 docker 來玩 logstash

- docker pull logstash (非官方所提供之映像檔)
- docker run --name logstash -itd logstash:5.5.0 bash

指令

- logstash
  - -e: 使用設定字串
  - -f: 使用設定檔
  - -t: 測試設定檔文法是否正確

練習1:牛刀小試

- logstash -e “input { stdin {}} output { stdout {}}”
- 等待出現成功訊息: Successfully started Logstash API endpoint
- 輸入任意訊息

設定檔: logstash -f <設定檔>

```
input {
  ...
}
filter {
  ...
}
output {
  ...
}
```

練習2: 將練習1 改成用設定檔再執行一次

- 將練習1的字串放入檔案 demo1.conf
- logstash -t demo1.conf
- logstash -f demo1.conf

電腦如何表示 / 儲存 / 傳送資料?

- Bob，男性，22 歲，身高 170.2 cm，體重 60.3 kg，興趣：籃球, 棒球，未婚
- Amy，女性，18 歲，身高 163.1 cm，體重 58.5 kg，興趣：音樂，已婚，丈夫名字：John

常用資料格式

- JSON
- ~~XML~~
- CSV

JSON

- 物件 (object): {}
- 陣列 (Array): []
- key-value 方式儲存
  - 字串: “”
  - 布林值: true, false
  - 數字: 1234, …
- 例子:
  - { “name”: “Bob”, “age”: 22, “height”: 170.2, “weight”: 60.3, “interest”: [“baseketball” , “baseball”], “married”: false }
  - { “name”: “Amy”, “age”: 18, “height”: 163.1, “weight”: 58.5, “interest”: [“music”], married: true, “couple”: “John” }

XML

```
<people>
    <name>Bob</name>
    <age>22</age>
    <height>170.2</height>
    <weight>60.3</weight>
    <interest>baseketball</interest>
    <interest>baseball</interest>
    <married>false</married>
</people>
```

CSV

name	age	height	weight	interest	married	couple
Bob	22	170.2	60.3	basketball|baseball	false	
Amy	18	163.1	58.5	music	true	John

name,age,height,weight,interest,married,couple
Bob,22,170.2,60.3,baseketball|baseball,false,
Amy,18,163.1,58.5,music,true,John

Schema:
name:string, age:int, height:int, interest: array<string>,...

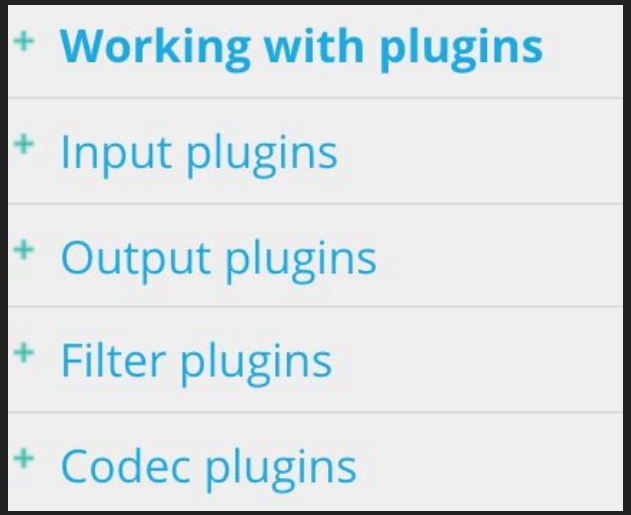
JSON 與 CSV 的比較

- JSON: 描述能力強
- CSV: 節省空間

Logstash 相關擴充介紹

完整的擴充內容

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



練習3: JSON 輸出

```
output {
  stdout {
    codec => json_lines
  }
}
```

練習4: JSON 輸入

練習5: 刪掉不必要的欄位

- 有些擴充可能會用到 @timestamp 之類的欄位, 移除要小心

練習6: 使用檔案輸入 (file input)

- Checkpoint
- Never stop
- 檔案時間限制

Reference

- 瞭解 JSON 格式
- Logstash 官方網站