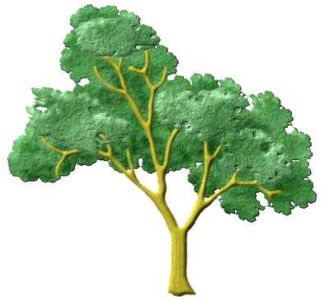
進階課程

Hadoop專案分享

透過 一些使用Hadoop的案例帶給
大家一些新的啟發

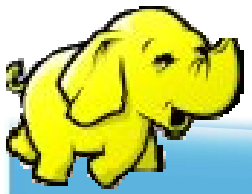
財團法人國家實驗研究院
國家高速網路與計算中心
NATIONAL CENTER FOR HIGH-PERFORMANCE COMPUTING



A：用Hadoop 打造 Location Plus

2009/11/17

- 「Location Plus！」服務，擊敗了其他38組作品獲得Yahoo開發競賽的優勝
- 從大量的批踢踢BBS文章中，找出臺灣17個城市的熱門話題
- 每個城市提供30個熱門話題和相關的參考詞
- 可以讓使用者用這些熱門話題來搜尋 Yahoo知識+、生活+、無名小站等內容
- 提供了手機版介面，讓使用者到任何地方就知道當地有哪些熱門話題

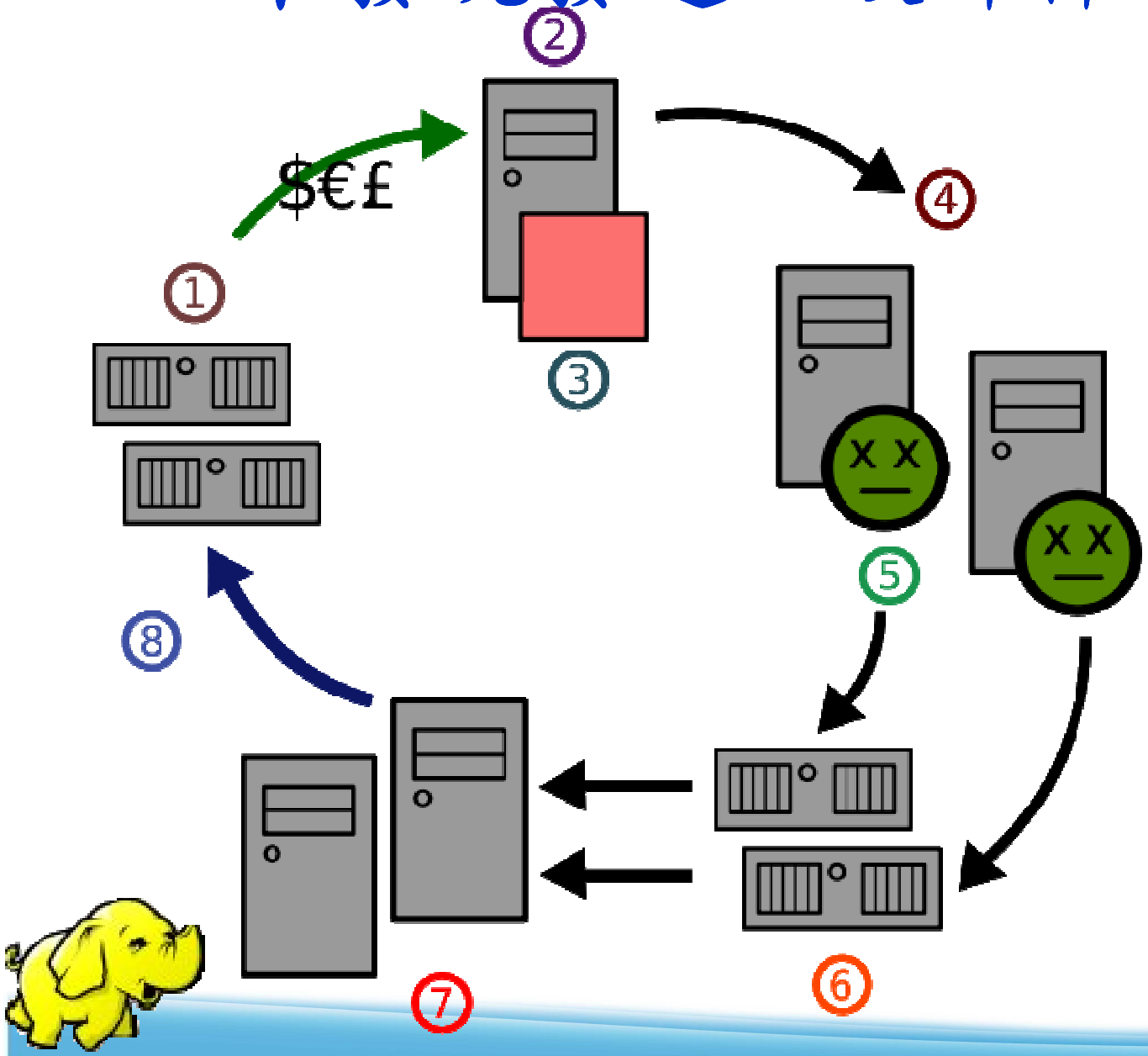


A：Location Plus

- 用RSS蒐集批踢踢BBS站的十大熱門看板的文章，約20萬筆文章記錄
- 維基百科提供的30萬筆詞條作為關鍵詞
- 將所有關鍵詞套用到所有文章紀錄，共需 600 億次比對（還不包含排序..）
- 這些都交給Hadoop吧！

http://www.ithome.com.tw/itadm/article.php?c=58008

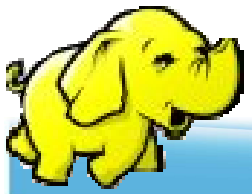
B：Yahoo 使用Hadoop平台
來發現發送垃圾郵件的殭屍網絡



（1） 垃圾郵件網站
（2） 垃圾郵件
（3） 垃圾郵件軟體
（4） 被感染的電腦
（5） 病毒或木馬
（6） 信件伺服器
（7） 用戶
（8） 網路流量

C：警訊整合系統

- 目的：
  - 將原本複雜難懂的警訊日誌整合成易於明瞭的報告
  - 透過“雲端”來運算大量資料
- 環境：
  - hadoop 0.20
  - Java 1.6
  - Apache 2



輸入資料

[**] [1:538:15] NETBIOS SMB IPC$ unicode share access [**]
[Classification: Generic Protocol Command Decode] [Priority: 3]
09/04-17:53:56.363811 168.150.177.165:1051 ->
168.150.177.166:139
TCP TTL:128 TOS:0x0 ID:4000 IpLen:20 DgmLen:138 DF
AP Seq: 0x2E589B8 Ack: 0x642D47F9 Win: 0x4241
TcpLen: 20

[**] [1:1917:6] SCAN UPnP service discover attempt [**]
[Classification: Detection of a Network Scan] [Priority: 3]
09/04-17:53:56.385573 168.150.177.164:1032 ->
239.255.255.250:1900
UDP TTL:1 TOS:0x0 ID:80 IpLen:20 DgmLen:161
Len: 133

[**] [1:1917:6] SCAN UPnP service discover attempt [**]
[Classification: Detection of a Network Scan] [Priority: 3]
09/04-17:53:56.386910 168.150.177.164:1032 ->
239.255.255.250:1900
UDP TTL:1 TOS:0x0 ID:82 IpLen:20 DgmLen:161
Len: 133

[**] [1:1917:6] SCAN UPnP service discover attempt [**]
[Classification: Detection of a Network Scan] [Priority: 3]
09/04-17:53:56.388244 168.150.177.164:1032 ->
239.255.255.250:1900
UDP TTL:1 TOS:0x0 ID:84 IpLen:20 DgmLen:161
Len: 133

[**] [1:1917:6] SCAN UPnP service discover attempt [**]
[Classification: Detection of a Network Scan] [Priority: 3]
09/04-17:53:56.417045 168.150.177.164:45461 -> 168.150.177.1:1900
UDP TTL:1 TOS:0x0 ID:105 IpLen:20 DgmLen:161
Len: 133

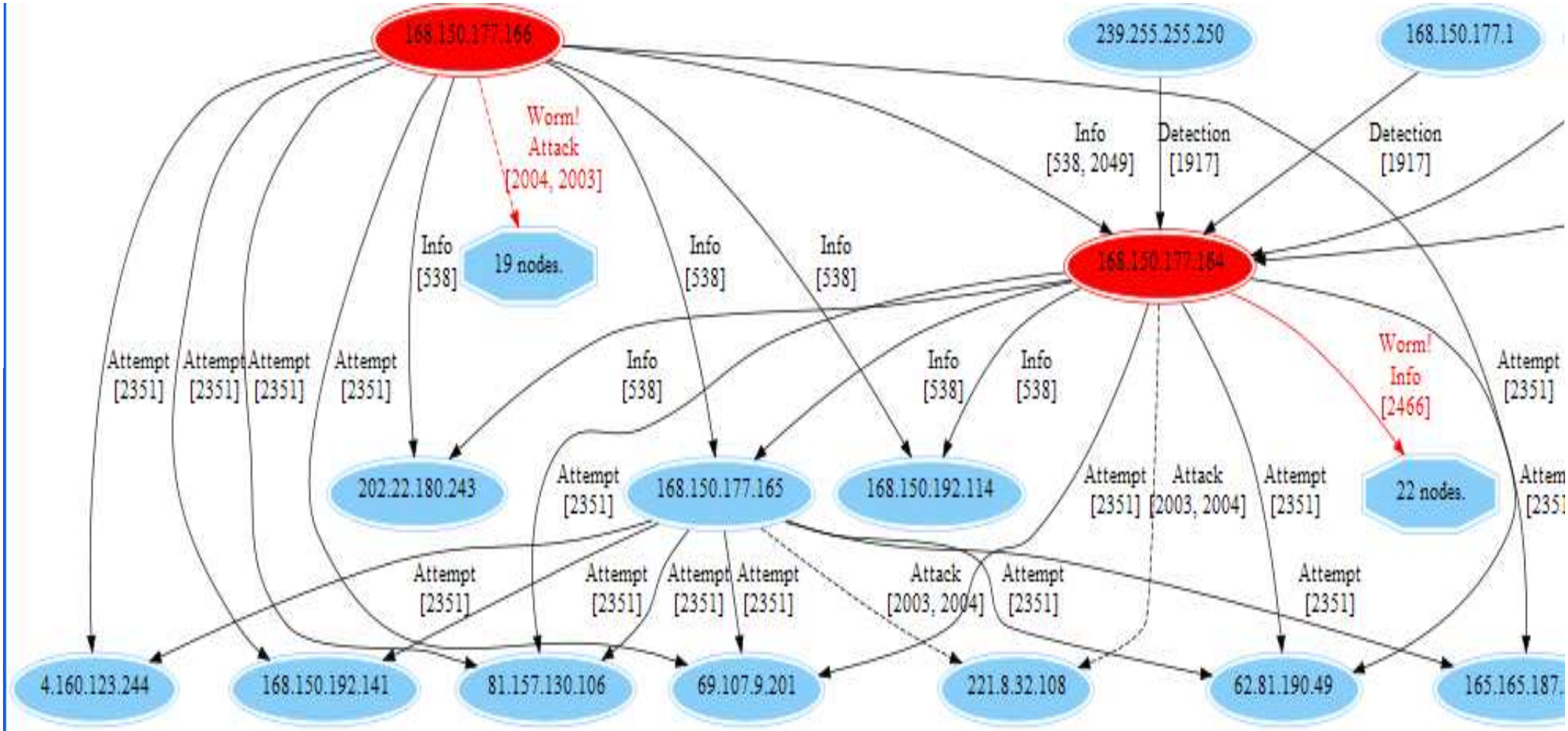
[**] [1:1917:6] SCAN UPnP service discover attempt [**]
[Classification: Detection of a Network Scan] [Priority: 3]
09/04-17:53:56.420759 168.150.177.164:45461 -> 168.150.177.1:1900
UDP TTL:1 TOS:0x0 ID:117 IpLen:20 DgmLen:160
Len: 132

[**] [1:1917:6] SCAN UPnP service discover attempt [**]
[Classification: Detection of a Network Scan] [Priority: 3]
09/04-17:53:56.422095 168.150.177.164:45461 -> 168.150.177.1:1900
UDP TTL:1 TOS:0x0 ID:118 IpLen:20 DgmLen:161
Len: 133

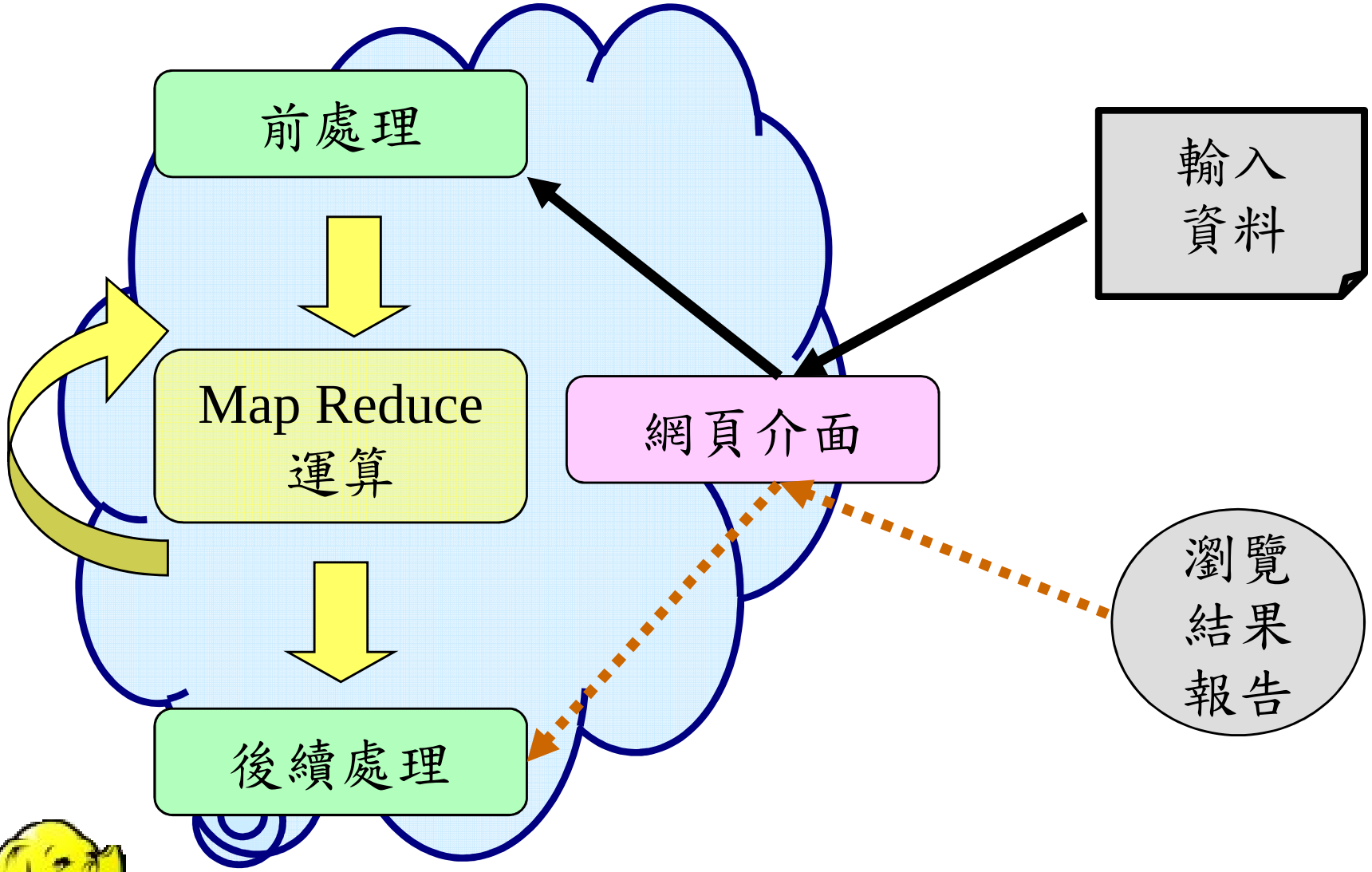
[**] [1:2351:10] NETBIOS DCERPC ISystemActivator path overflow
attempt little endian unicode [**]
[Classification: Attempted Administrator Privilege Gain] [Priority: 1]
09/04-17:53:56.442445 198.8.16.1:10179 -> 168.150.177.164:135
TCP TTL:105 TOS:0x0 ID:49809 IpLen:20 DgmLen:1420 DF
A* Seq: 0xF9589BBF Ack: 0x82CCF5B7 Win: 0xFFFF
TcpLen: 20
[Xref => http://www.microsoft.com/technet/security/bulletin/MS03-
026.mspx][Xref =>
http://cgi.nessus.org/plugins/dump.php3?id=11808][Xref =>
http://cve.mitre.org/cgi-bin/cvename.cgi?name=2003-0352][Xref =>
http://www.securityfocus.com/bid/8205]

輸出資料

Generate dot graph format

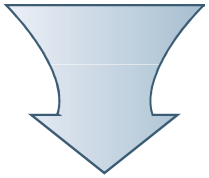


系統分析



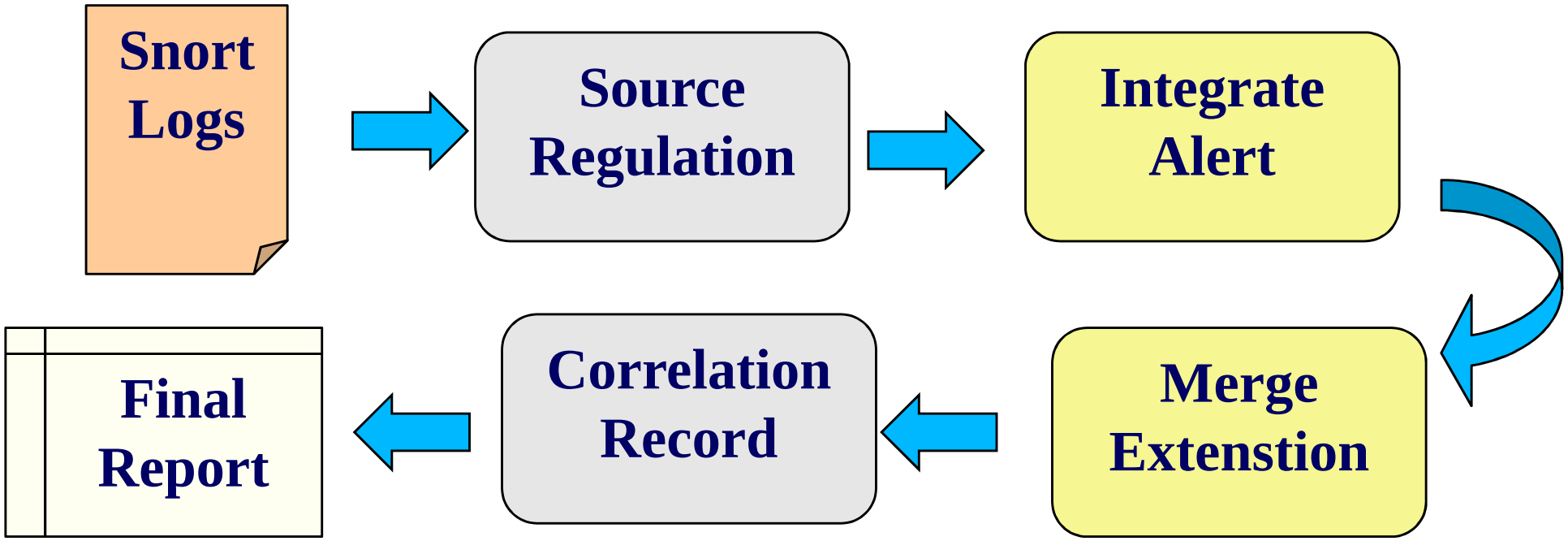
Alert Merge Example

Destination IP	Attack Signature	Source IP	Destination Port	Source Port	Packet Protocol	Timestamp
Host_1	Trojan	Sip1	80	4077	tcp	T1
Host_1	Trojan	Sip2	80	4077	tcp	T2
Host_1	Trojan	Sip1	443	5002	tcp	T3
Host_2	Trojan	Sip1	443	5002	tcp	T4
Host_3	D.D.O.S	Sip3	53	6007	udp	T5
Host_3	D.D.O.S	Sip4	53	6008	tcp	T5
Host_3	D.D.O.S	Sip5	53	6007	udp	T5
Host_3	D.D.O.S	Sip6	53	6008	tcp	T5



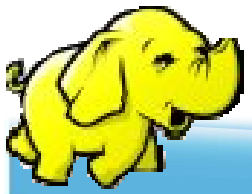
Key		Values				
Host_1	Trojan	Sip1,Sip2	80,443	4077,5002	tcp	T1,T2,T3
Host_2	Trojan	Sip1	443	5002	tcp	T4
Host_3	D.D.O.S.	Sip3,Sip4,Sip5 ,Sip6	53	6007,6008	tcp, udp	T5

程式流程圖


Snort Logs
Source Regulation
Integrate Alert
Merge Extenstion
Correlation Record
Final Report

Conclusions

- 評估
- 系統分析
  - 輸入輸出
  - 系統元件
  - 各元件參數與串流
- 實做
  - 多次運算
  - 前處理與後處理



進階課程

QUESTIONS
&
THANKS

財團法人國家實驗研究院
國家高速網路與計算中心
NATIONAL CENTER FOR HIGH-PERFORMANCE COMPUTING