

NAR Labs

National Applied Research Laboratories

National Center for
High-performance Computing

高通量運算技術與平台

High Throughput Computing Technologies
and NCHC's Platform Service

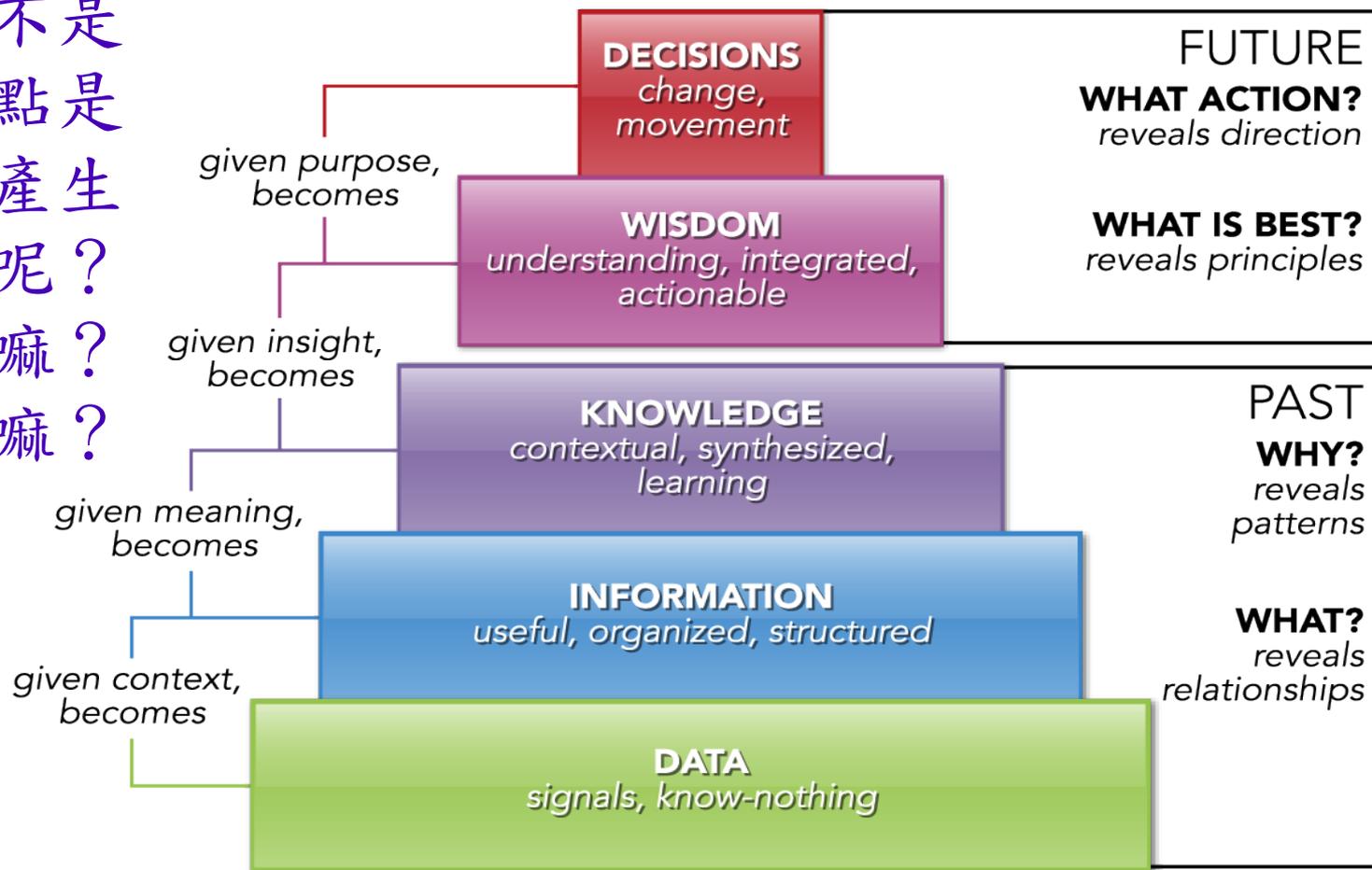
國家高速網路與計算中心

王耀聰 <jazz@narlabs.org.tw>

2013/09/13 - 2013 Big Data 前瞻論壇

知識源自彙整過去， 智慧在能預測未來

資料多寡不是
重點，重點是
我們想要產生
什麼價值呢？
時效合理嘛？
成本合理嘛？



大家都說「資料是金礦」，
那就讓我們拿採礦當類比吧！

國際金價

提供給客戶的價值

產品通路

開採成本

總擁有成本

軟硬體投資

提煉廠

分析平台與工具軟體

SMAQ

含金量

資料鑑價？

商業模式

開採權

分析資料的合法性

個資法

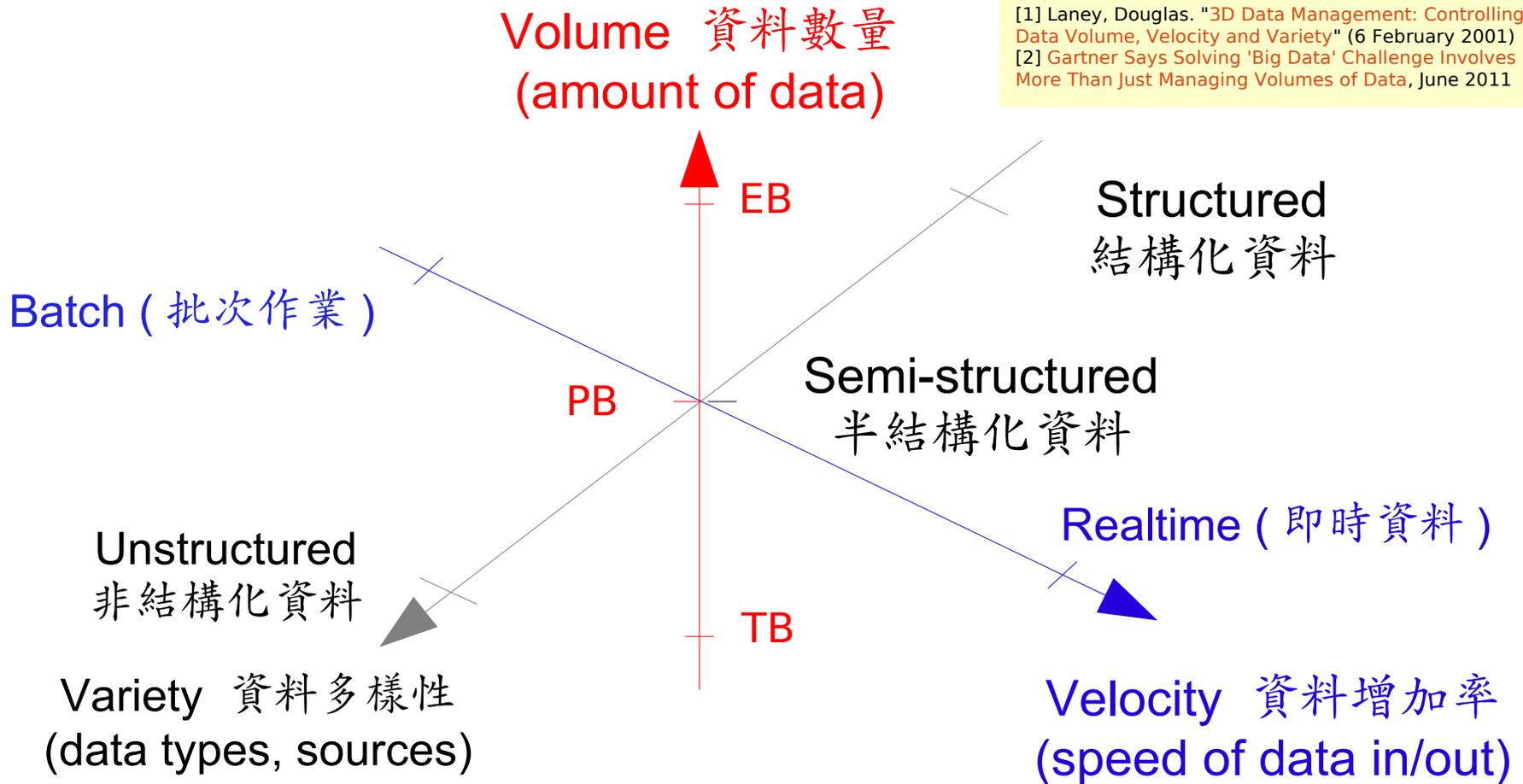
金礦

資料集

Open Data

巨量資料的三大挑戰

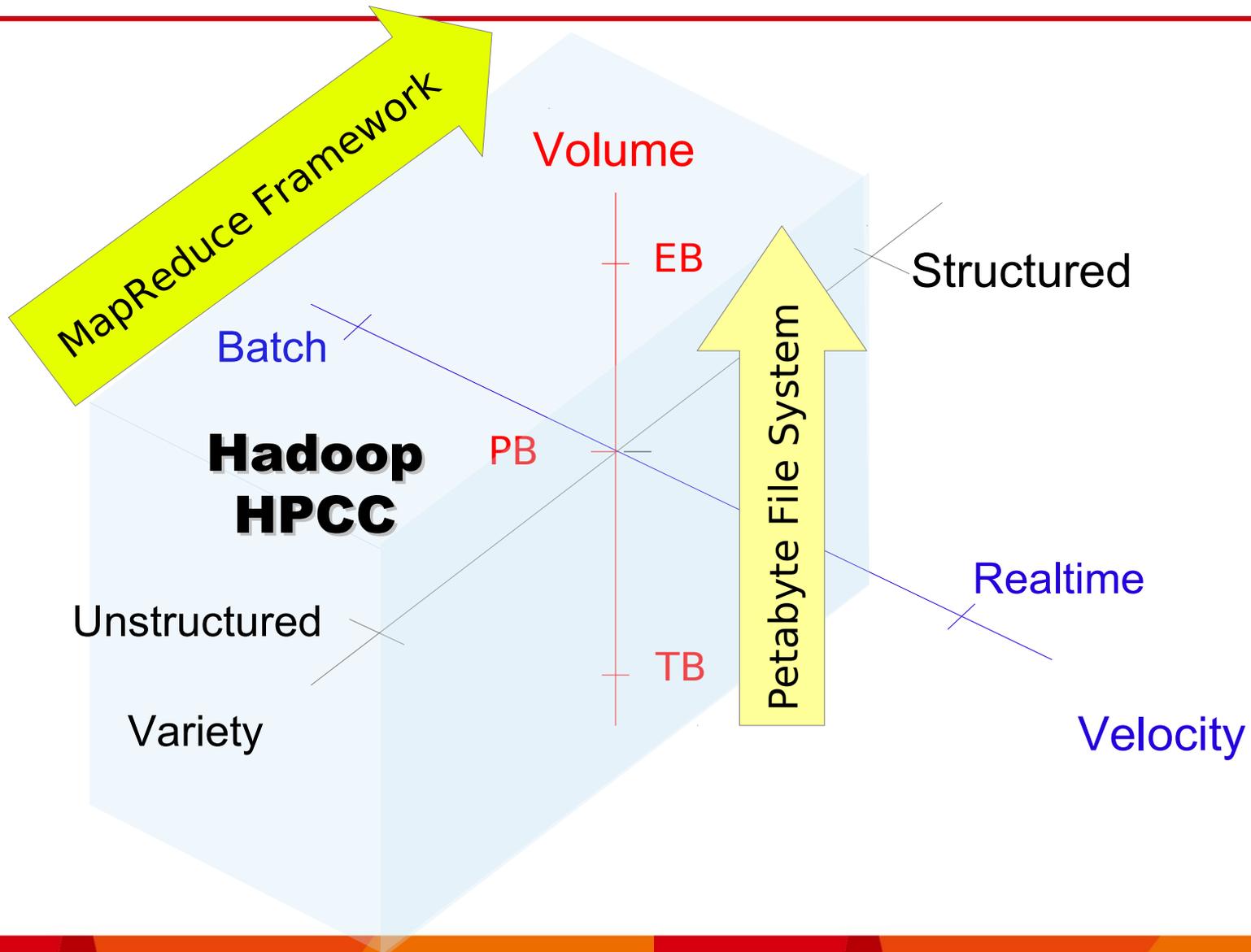
3 Vs of Big Data



巨量資料的挑戰在於如何管理「數量」、「增加率」與「多樣性」

處理巨量資料的三類技術 (1)

Data at Rest – MapReduce Framework



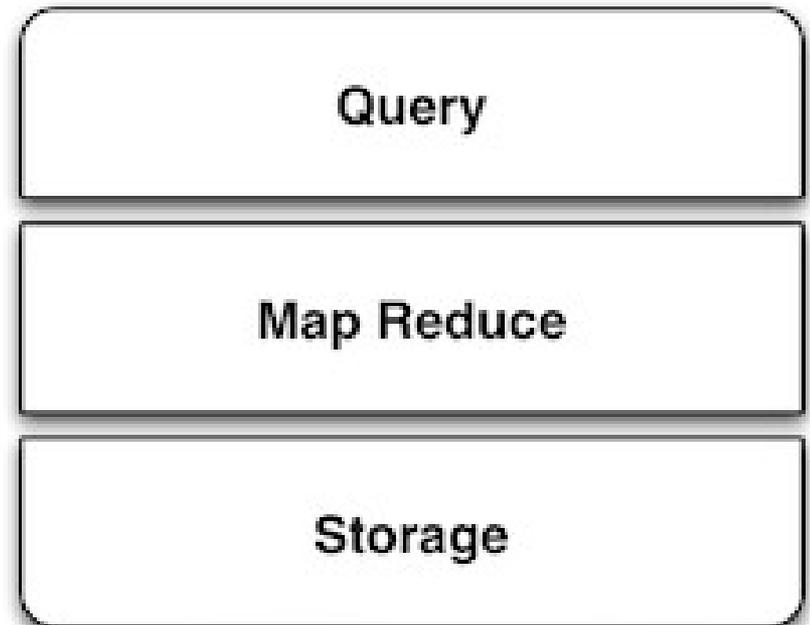
巨量資料處理的資訊架構

The SMAQ stack for big data

做網頁相關的人可能聽過 LAMP



未來處理海量資料的人必需知道
SMAQ (Storage, MapReduce and Query)



參考來源：The SMAQ stack for big data，Edd Dumbill，22 September 2010，

<http://radar.oreilly.com/2010/09/the-smaq-stack-for-big-data.html>

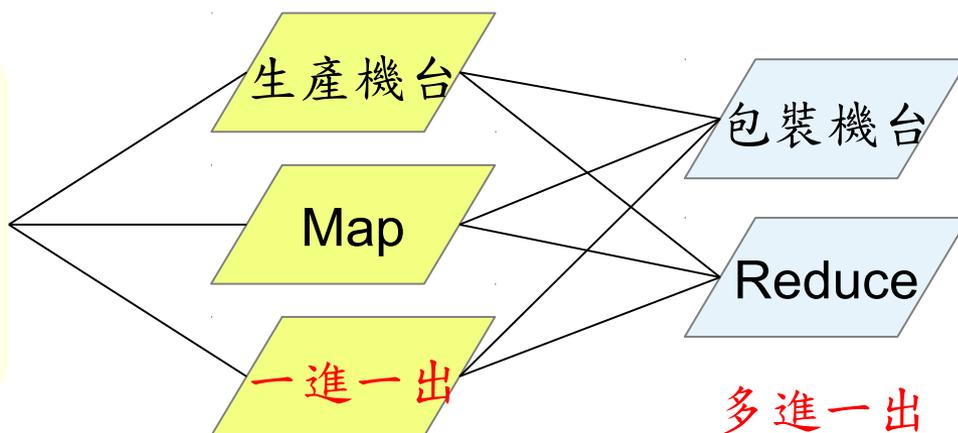
圖片來源：<http://smashingweb.ge6.org/wp-content/uploads/2011/10/apache-php-mysql-ubuntu.png>

Hadoop 是一個讓使用者簡易撰寫並執行處理海量資料應用程式的軟體平台。

亦可以想像成一個處理海量資料的生產線，只須學會定義 **map** 跟 **reduce** 工作站該做哪些事情。

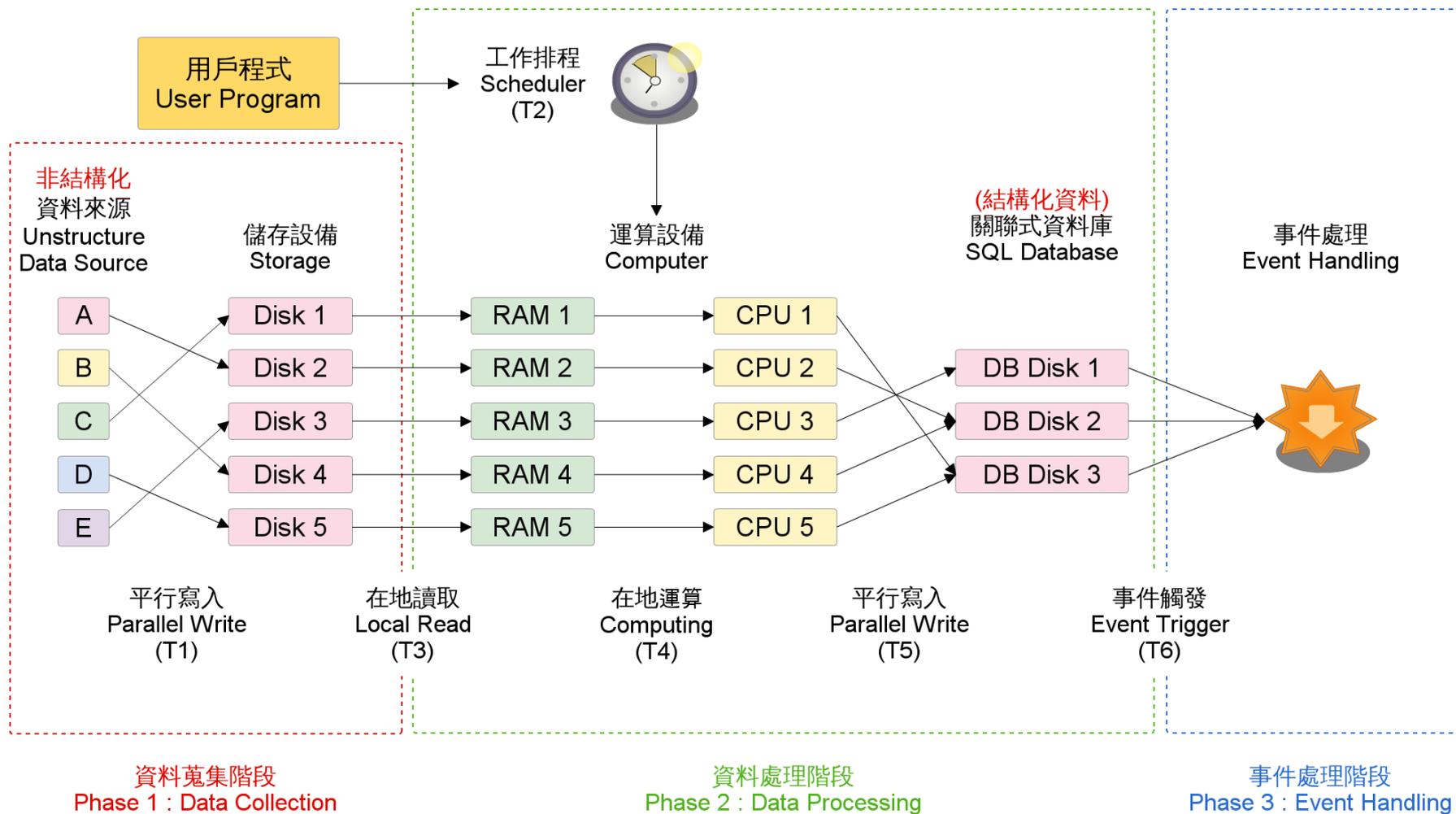
就像工廠的倉庫
存放生產原料跟待售貨物

HDFS 存放
待處理的**非結構化**資料
與處理後的**結構化**資料



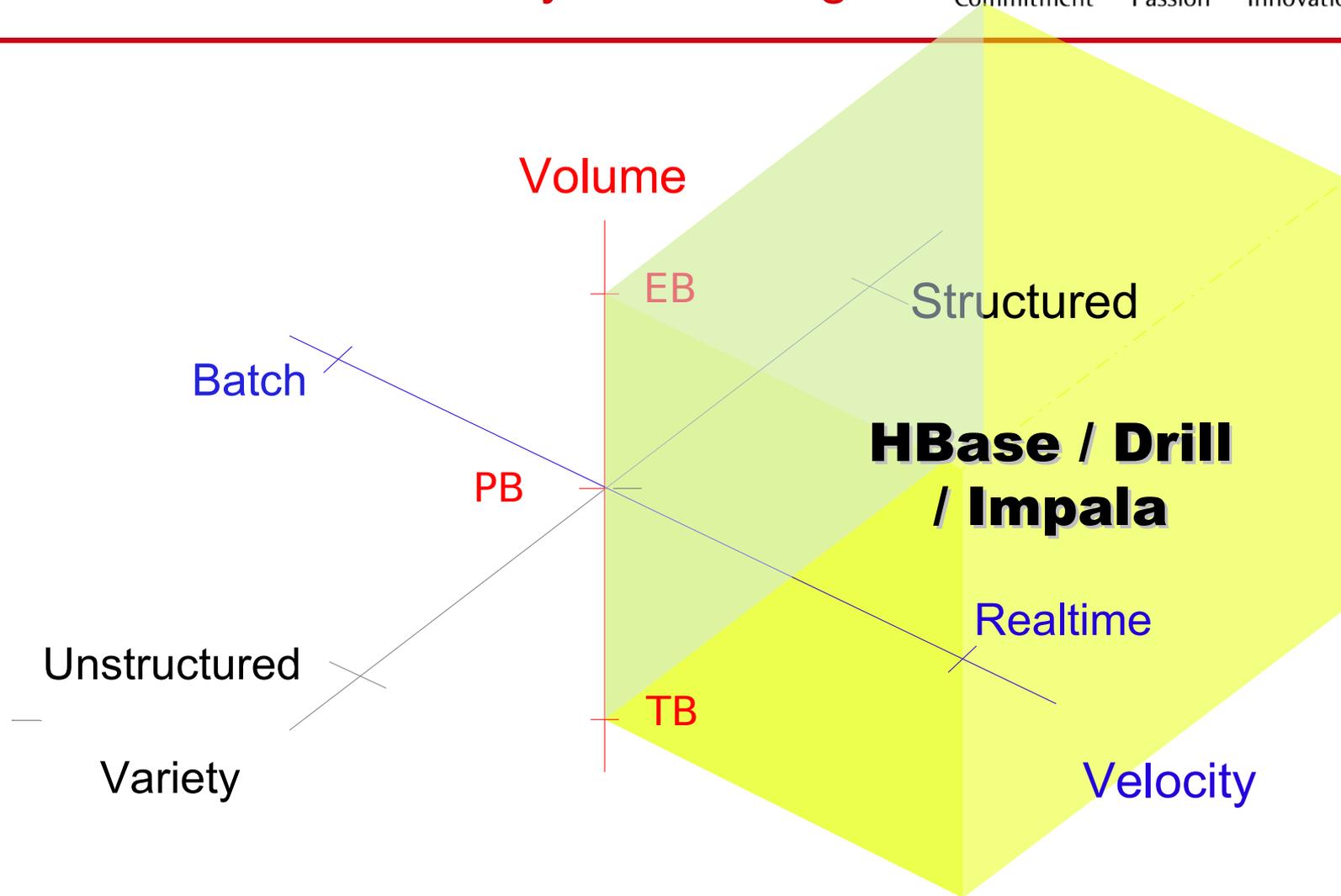
批次作業的運算時間

Processing Time of Batch Jobs



處理巨量資料的三類技術 (2)

Data in Motion – In-Memory Processing



Google 的技術演進 VS Apache 專案

Big Query
(JSON, SQL-like)

Dremel
(2010)

Apache Drill
(2012)

Incremental Index Update
(Caffeine)

Percolator
(2010)

Graph Database

Pregel
(2009)

Apache Giraph
(2011)

Query

BigTable
(2006)

Apache HBase
(2007)

Map Reduce

MapReduce
(2004)

Hadoop MapReduce
(2006)

Storage

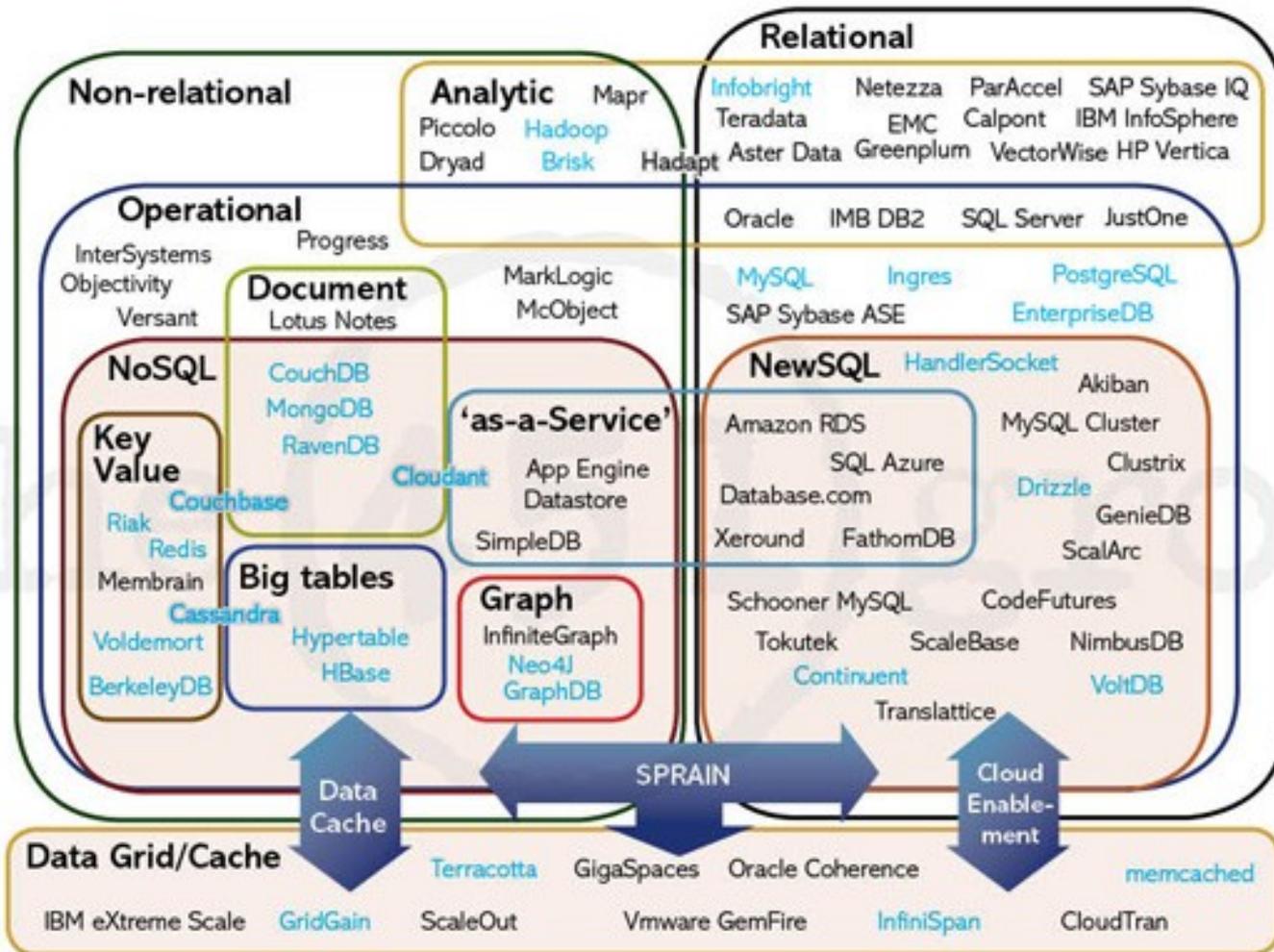
Google File System
(2003)

HDFS
(2006)

令人眼花撩亂的多樣化資料庫選擇

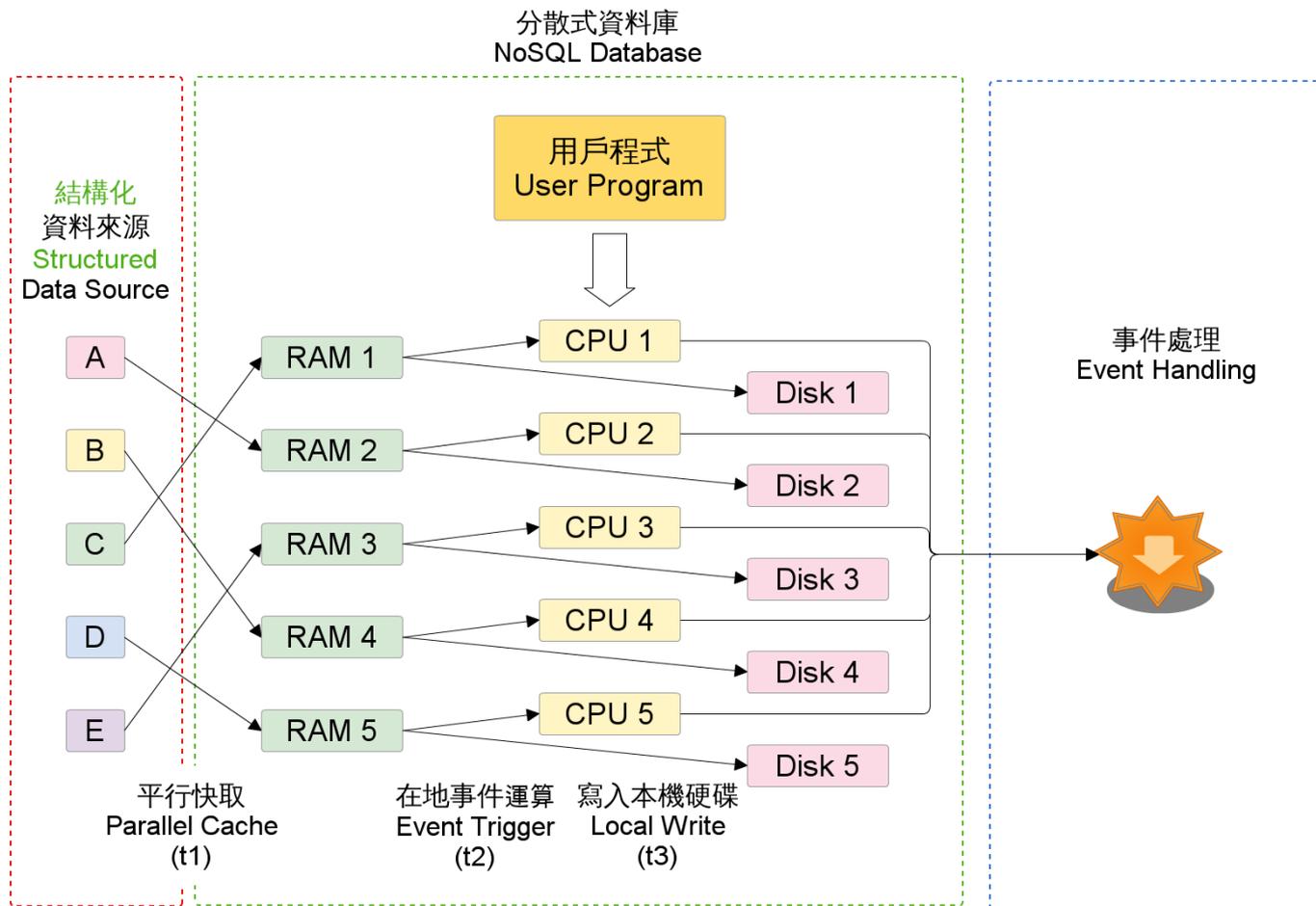
NoSQL vs NewSQL

Commitment · Passion · Innovation



<http://www.infoq.com/news/2011/04/newsql>

In-Memory Processing 的運算時間 以 HBase 為例



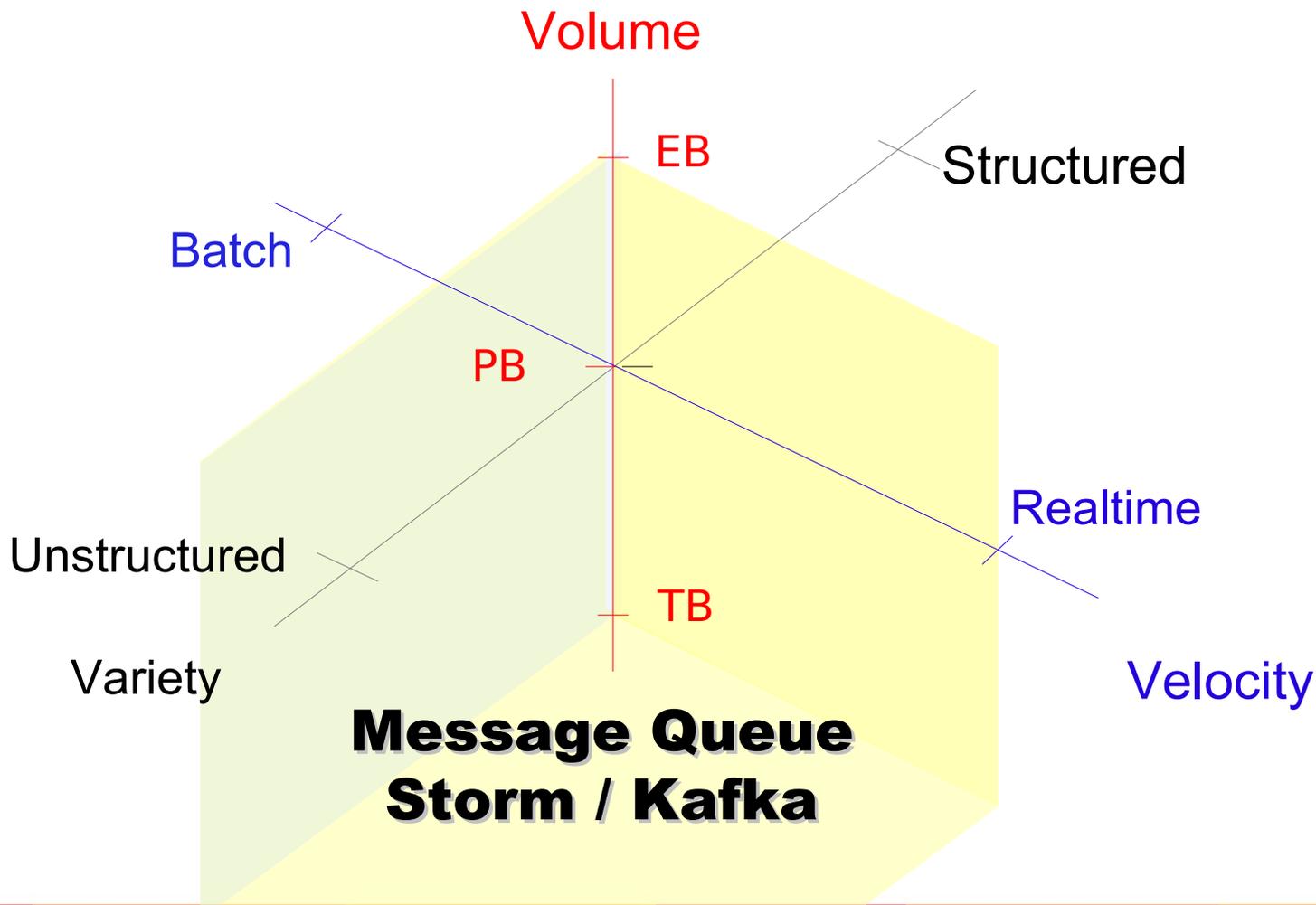
資料蒐集階段
Phase 1 : Data Collection

資料處理階段
Phase 2 : Data Processing

事件處理階段
Phase 3 : Event Handling

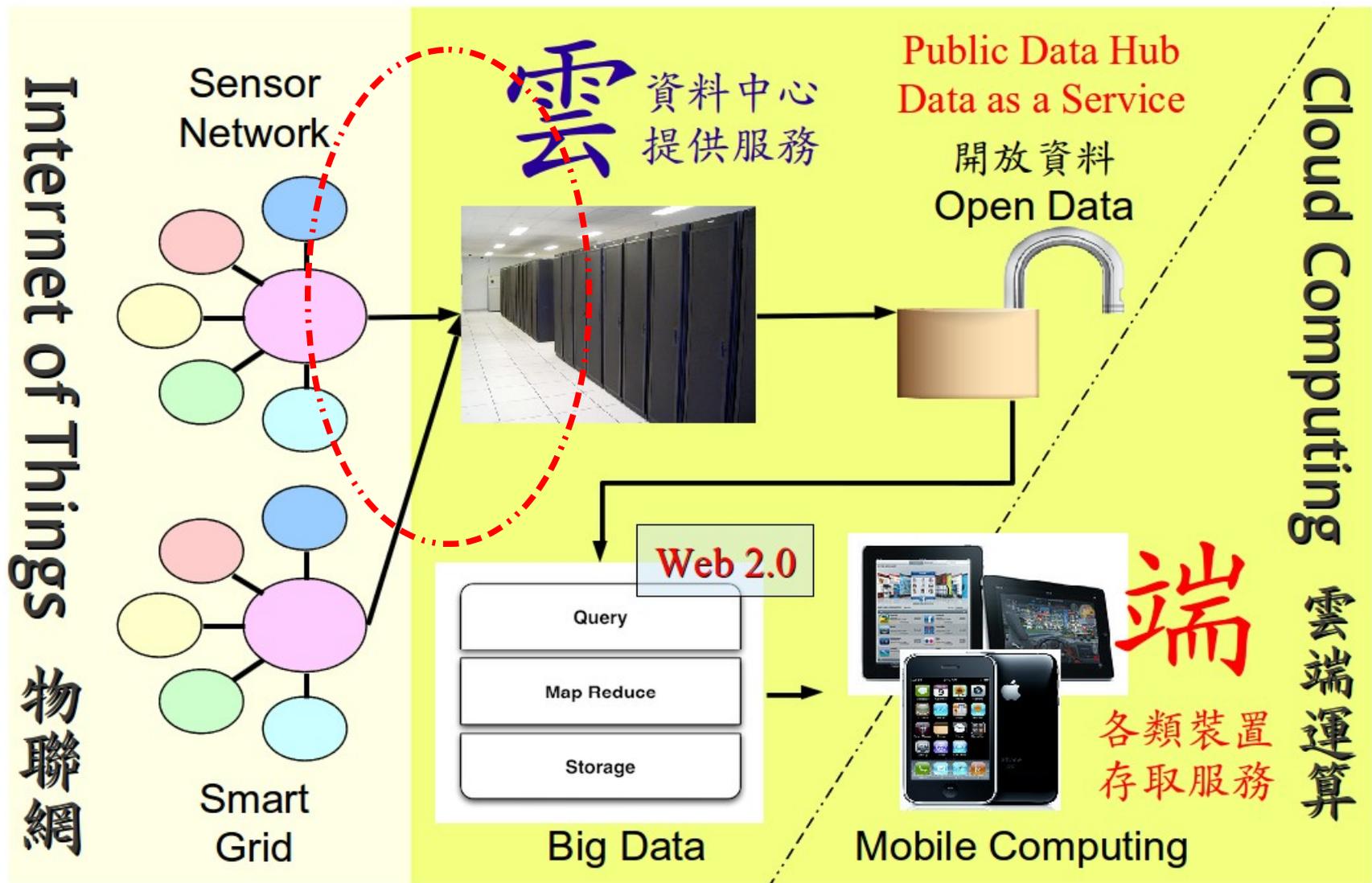
處理巨量資料的三類技術 (3)

Streaming Data Collection

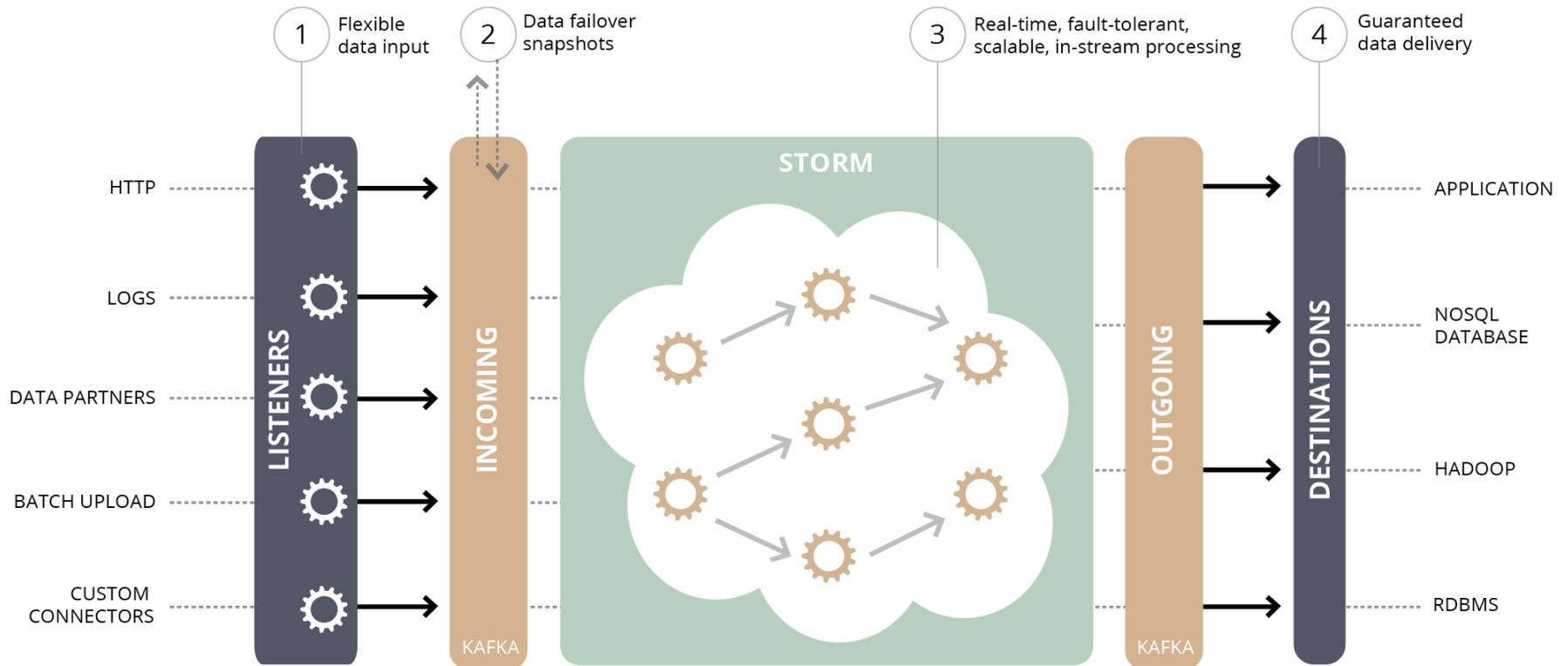


巨量資料的奇幻漂流

Life of Big Data

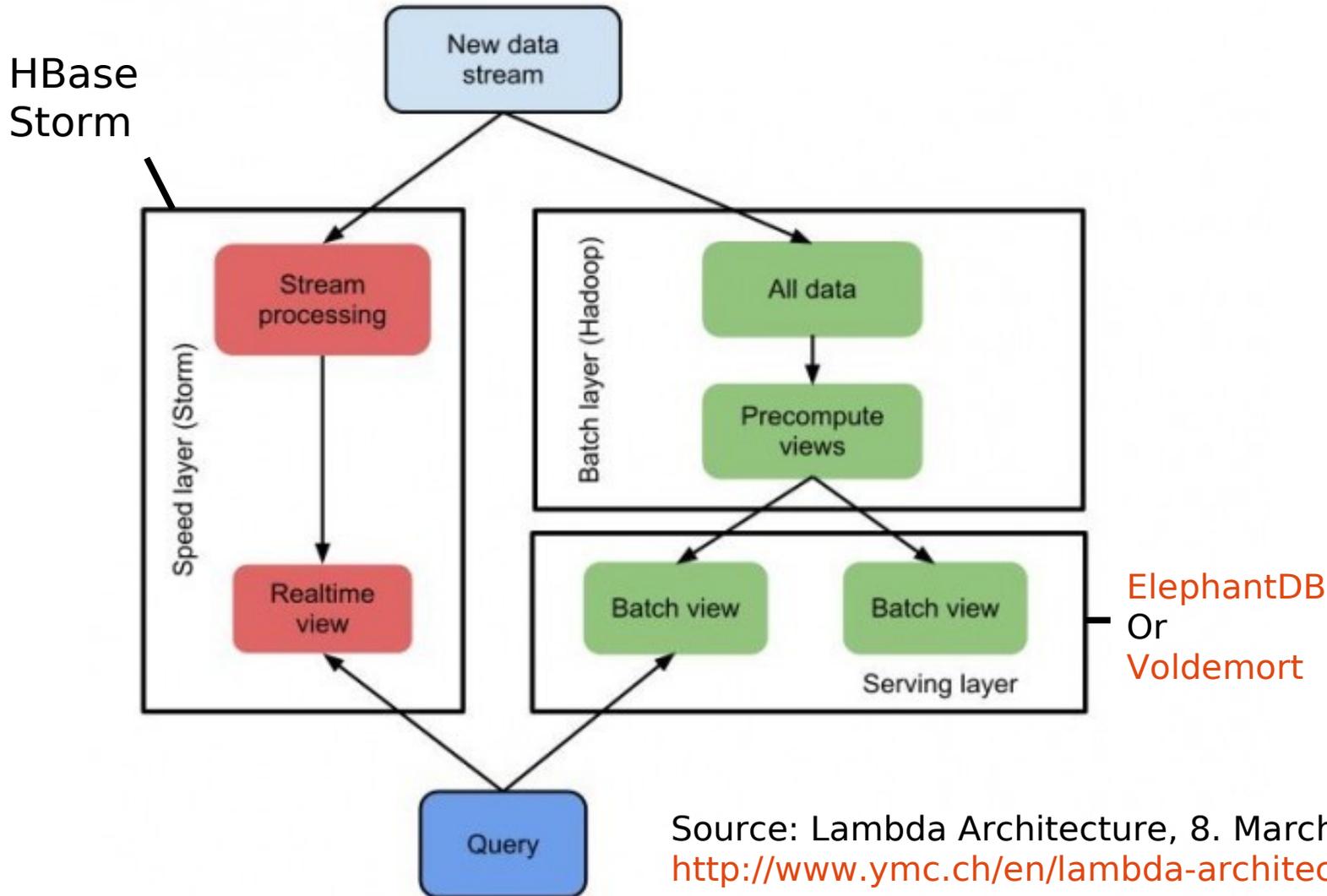


Twitter Storm + Apache Kafka



混合模式的巨量資料處理架構

Lambda Architecture



Source: Lambda Architecture, 8. March 2013
<http://www.ymc.ch/en/lambda-architecture-part-1>

NAR Labs

National Applied Research Laboratories

National Center for
High-performance Computing

高通量運算平台現況與未來規劃

國家高速網路與計算中心

王耀聰 <jazz@narlabs.org.tw>

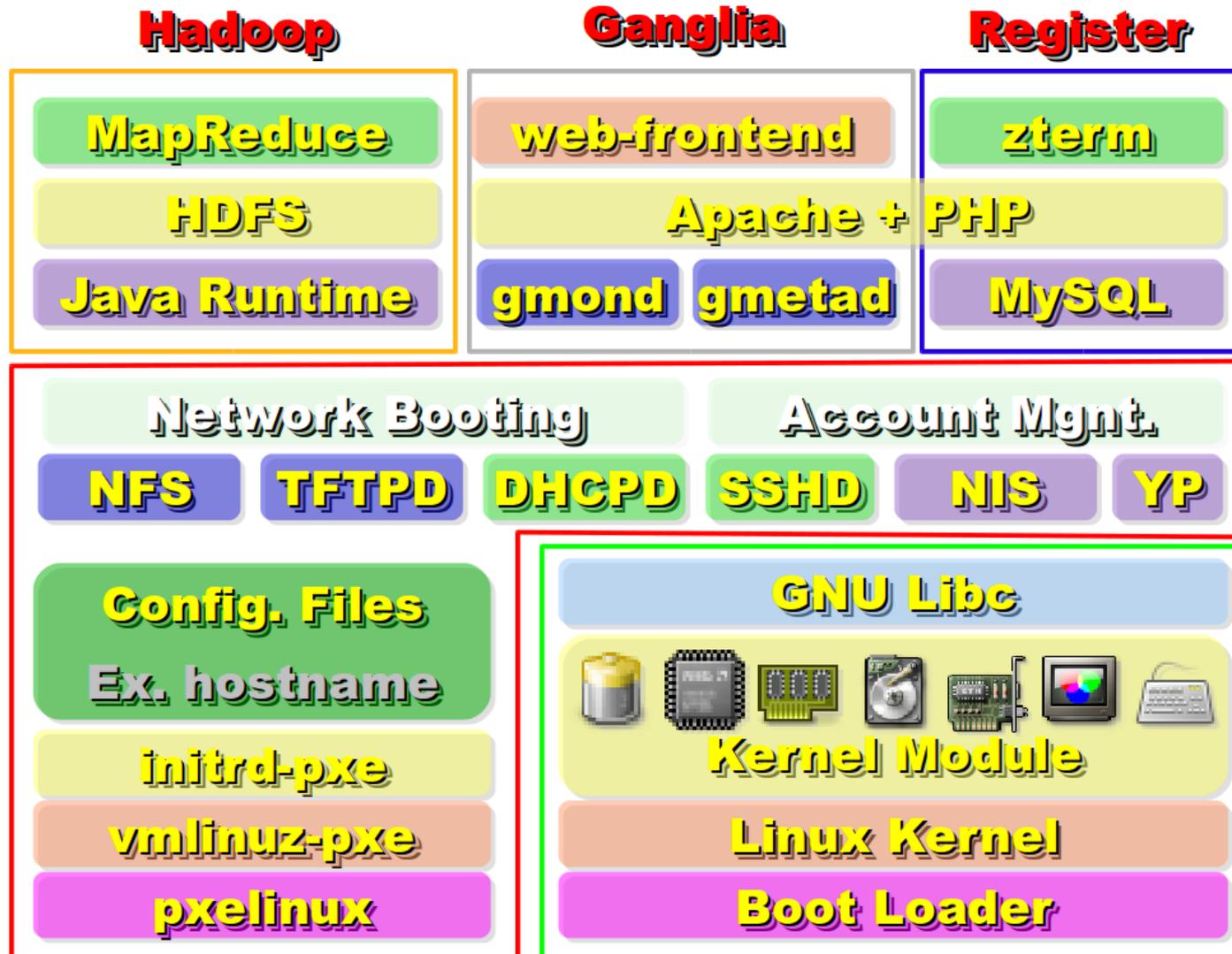
2013/09/13 - 2013 Big Data 前瞻論壇

hadoop.nchc.org.tw 現況

- 2009-04-13 對外開放申請帳號，12 台
- 2010-10-20 完成升級，21 台
- 截至 2013-09-10，共計 4012 人次申請
- 系統現況：6 台故障，15 台繼續服務中
- 累計服務對象數：（根據註冊資料整理結果）
 - 94 所大學
 - 33 間民間公司
 - 3 所醫院（國泰 / 童綜合 / 龍泉榮民醫院）



目前系統架構 Current Architecture



DRBL

Linux

全台首座公用 Hadoop 實驗叢集 On-Demand Self Service

NAR Labs

Commitment · Passion · Innovation

Hadoop 網頁介面登入

帳號： 密碼：

[新增帳號](#) [忘記密碼](#) [操作問題回報](#)

1. 歡迎至 forum.hadoop.tw 或 [臉書粉絲團](#) 進行討論
2. 歡迎加入 [公告群組](#)，以利接收即時公告事宜
3. 初學者請參閱[線上教學: Hadoop 觀念篇](#)，實作步驟請參閱以下三個連結：
[[帳號申請](#) | [HDFS 練習](#) | [MapReduce 練習](#)]
4. 本叢集所採用的 Hadoop 版本是 0.20.1，寫程式時請參考 [javadoc](#)
5. 倘若貴單位不允許 SSH 連線，或有 PROXY 限制，請改連 <https://hadoop.nchc.org.tw> 並採用系統帳號密碼(hXXXX)登入。至於 HDFS 與 MapReduce 也請點選對應連結存取。
6. R 軟體使用者，可用 <http://hadoop.nchc.org.tw/rstudio> 介面 (帳號為 hXXXX)

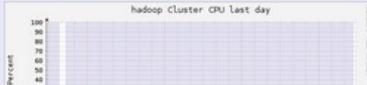
**本叢集純為實驗用途，無法保證7x24服務品質
重要數據資料請務必另行備份，謝謝！
重要運算工作亦請考慮使用其他付費平台**

家目錄空間吃緊中，請盡量上傳至HDFS後，清除家目錄檔案，謝謝！

若有公用資料集運算需求，請置於 <https://hadoop.nchc.org.tw/pub>，謝謝！

註冊人數：3967 / 3999 人

[MapReduce 狀態](#) | [HDFS 狀態](#)
過去 24 小時 CPU 負載 - [查詢完整系統負載](#)：



Running Jobs

[Quick Links](#)

none

Completed Jobs

none

Failed Jobs

none

Local L

網站帳號 jazzwang E-mail [redacted] 姓名 王耀聰 電話 0 單位 0 用途 0 主機帳號 h998 主機密碼 [redacted]

檔案(F) 編輯(E) 檢視(V) 歷史(Y) 工具(T) 說明(H)

Log directory, [1. hadoop.nchc.org.tw](#)

NameNode

Started:	Linux hadoop 2.6.32-5-amd64 #1 SMP Wed Jan 12 03:40:32 UTC 2011 x86_64
Version:	
Complexity:	
Upgrade:	

Browse the [NameNode](#)

Cluster Summary
4107529 files

WARNING: 1

Configuration	
DFS Use	
Non DFS	
DFS Replication	

```
Linux hadoop 2.6.32-5-amd64 #1 SMP Wed Jan 12 03:40:32 UTC 2011 x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

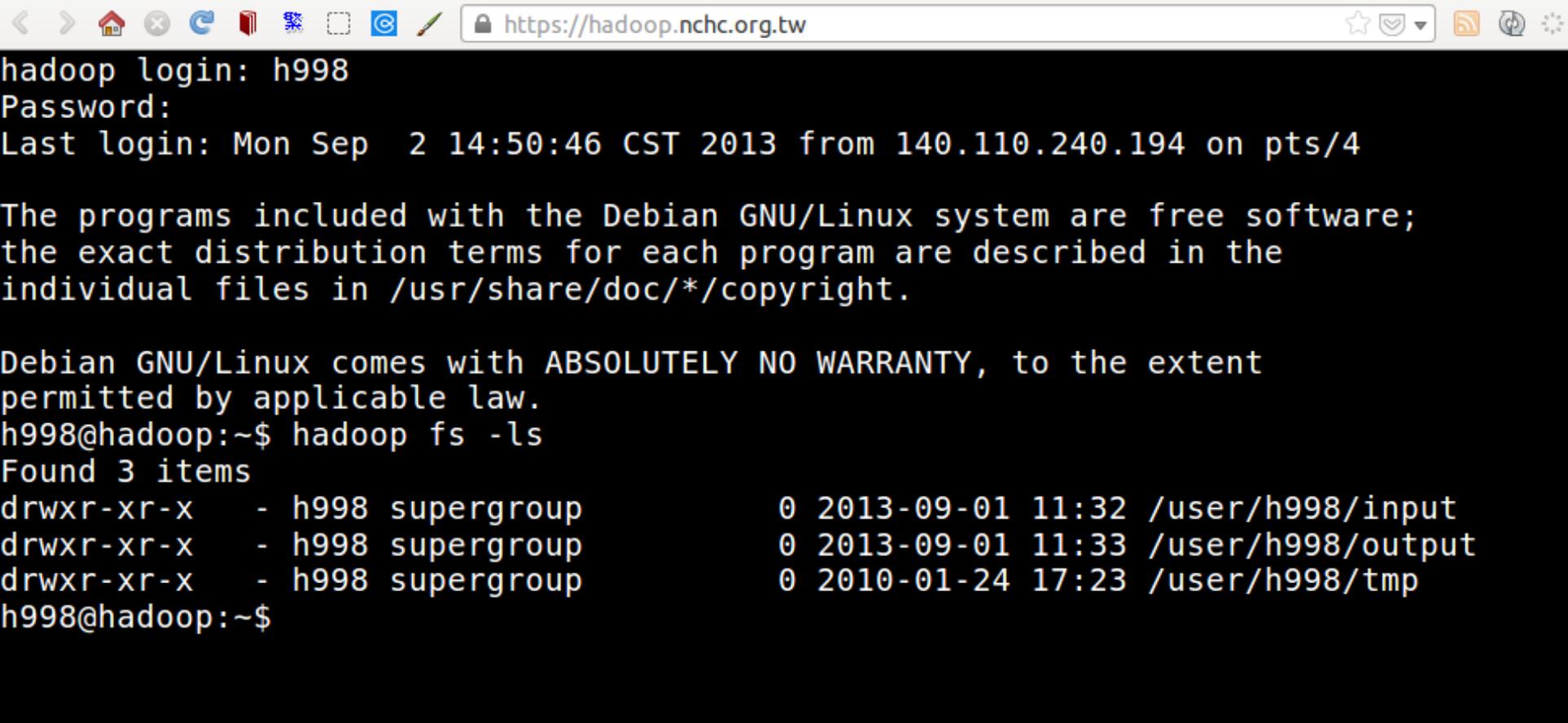
Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Tue Apr 26 15:45:44 2011 from nat235.dynamic.cs.nctu.edu.tw
h998@hadoop:~$
```

Powered by Zterm

<http://zhouer.org/ZTerm/>

讓網路受限的用戶更便利

- Web-based Console

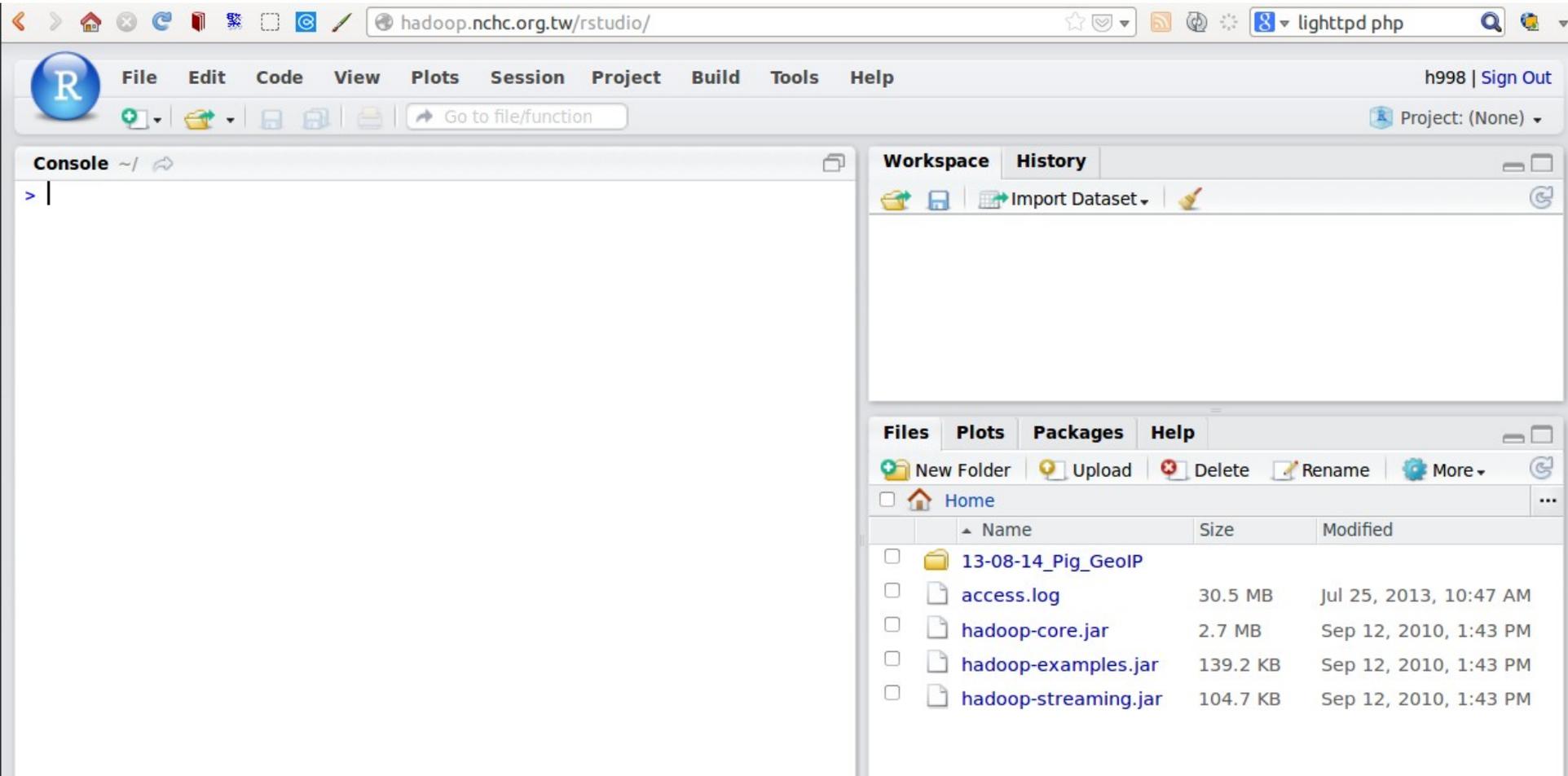


```
hadoop login: h998
Password:
Last login: Mon Sep  2 14:50:46 CST 2013 from 140.110.240.194 on pts/4

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
h998@hadoop:~$ hadoop fs -ls
Found 3 items
drwxr-xr-x  - h998 supergroup          0 2013-09-01 11:32 /user/h998/input
drwxr-xr-x  - h998 supergroup          0 2013-09-01 11:33 /user/h998/output
drwxr-xr-x  - h998 supergroup          0 2010-01-24 17:23 /user/h998/tmp
h998@hadoop:~$
```

便利熟悉高階語言的資料分析用戶 - RStudio 開發環境



<http://hadoop.nchc.org.tw/rstudio/>

Lesson Learned

Commitment · Passion · Innovation

- 善用 CDH 或 HDP2 的套件：好處是易於管理跟升級
- 大量帳號的管理：
 - 用 DRBL 內建指令大量建立帳號 `/opt/drbl/sbin/drbl-useradd`
 - 超過 5000+ 帳號該怎麼管理？LDAP + OpenID 帳號整合
 - 生命週期管理！（多人共用環境，由生到滅，要訂好遊戲規則）
- 使用者預設 HDFS 家目錄
 - 跑迴圈切換使用者，下 `hadoop fs -mkdir tmp`
- 安全性：設定使用者 HDFS 權限
 - 跑迴圈切換使用者，下 `hadoop dfs -chown $(id) /usr/$(id)`
 - 然後跑 `hadoop dfs -chmod -R 700 /usr/$(id)`

Lesson Learned

Commitment · Passion · Innovation

- 硬碟規劃
 - JBOD 架構，不用硬體 RAID。
 - I/O 分流：HDFS 一顆 (以上) 硬碟，MapReduce 一顆硬碟
- 記憶體規劃
 - 面對記憶體怪物，記得切 SWAP Partition
 - 未來面對 In-Memory Processing 的需求，記得多買記憶體
- 規劃黃金法則 @ 2013
 - 1 core : 2~8 GB RAM : 2 TB Disk

- 使用者不熟悉該如何使用我們提供的服務
 - 對外開辦教育訓練還不夠，直接深入各個需求單位會更好！
- 應用為王：資料庫、網頁服務與 Mobile App 整合需求頗高
 - 該挑選 NoSQL 還是 NewSQL 呢？端看 I/O 特性！
- 上游：Open Data，下游：統計分析應用
 - Data as a Service：資料集提供本身就是一種服務
 - 不要期待使用者改變寫程式的方法，儘量迎合他們熟悉的工具
- 整合虛擬化
 - 許多論文都需要跑不同節點數的效能比較
 - 個人研究資料如何保密？網路如何切割？透過虛擬化作隔離！

未來規劃 Future Plan

多
租
戶
帳
號
管
理
與
系
統
監
控

使用者介面 (Web-based GUI / IDE)

分析預測工具庫 (Ex. R, Mahout)

資料探勘工具庫 (Ex. Nutch, Lucene , Solr)

資料倉儲 (Ex. Hive)

分散式資料庫 (Ex. HBase)

高階語言
介面
(Ex. Pig)

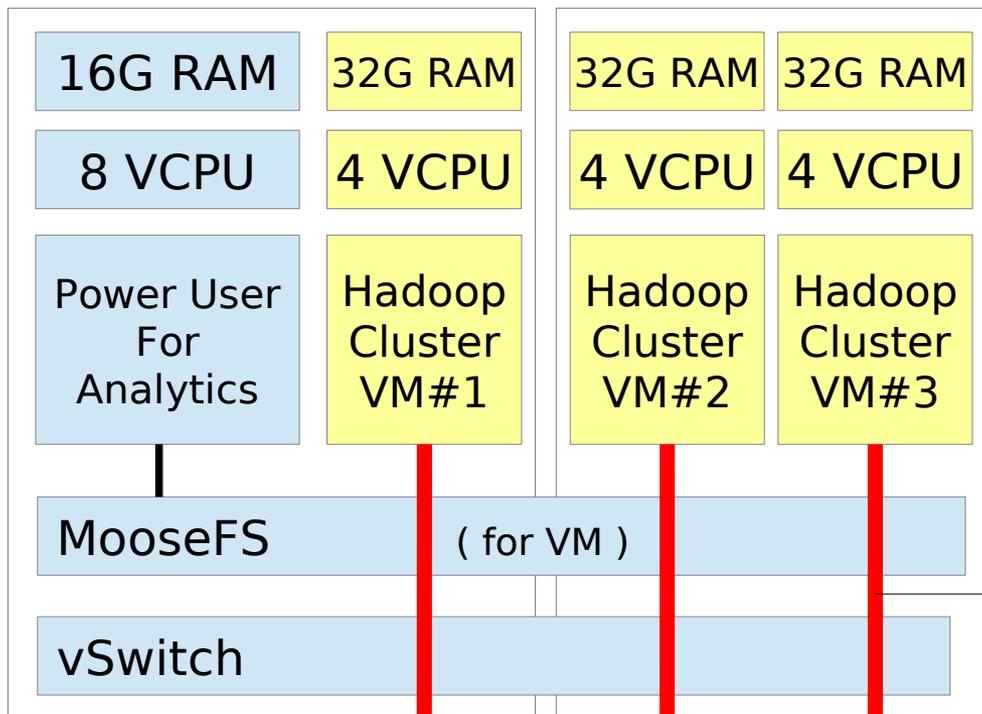
支援在地運算的工作排程 (MapReduce)

虛擬化管理軟體 (Ex. KVM + OpenNebula)

網路化虛擬 (Ex. Open vSwitch)

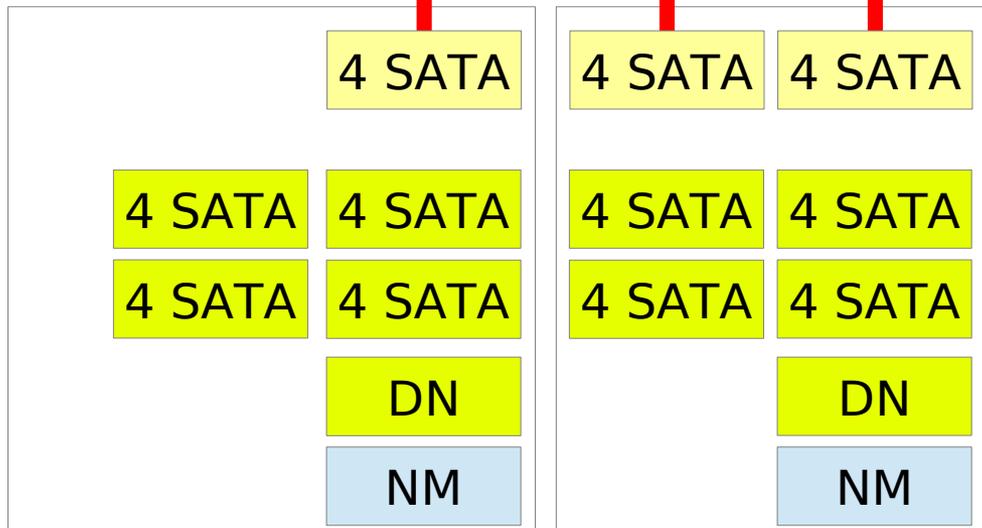
分散式儲存 (HDFS, MooseFS, 提升同時讀寫資料通量)

自
動
化
安
裝
佈
署

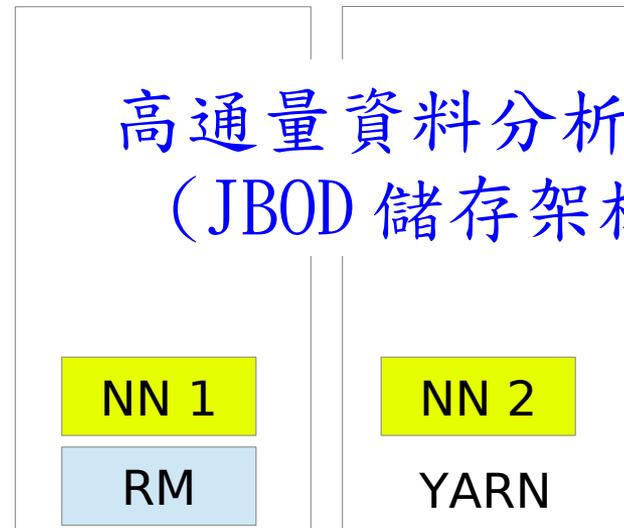


虛擬化雲端服務平台 (晶片組須支援虛擬化)

ATA over Ethernet
(AoE)



高通量資料分析平台 (JBOD 儲存架構)



問題與討論 Questions?

