從物聯網、雲端到巨量資料
– 談 SSD 的應用情境與挑戰

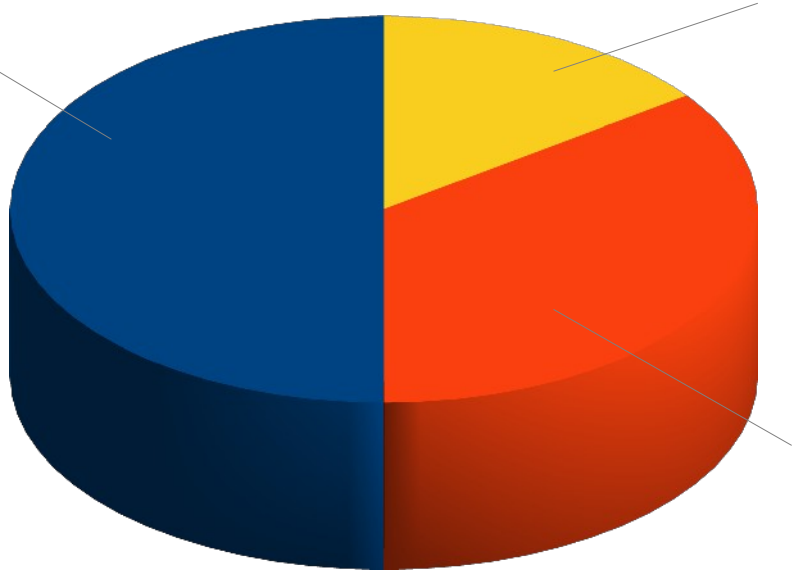
Life of Big Data :
Use Cases and Challenges for SSD Industry

Jazz Yao-Tsung Wang
Co-Founder and Evangelist of Hadoop.TW
National Center for High-performance Computing
jazzwang.tw@gmail.com

WHO AM I ? JAZZ ?

- About Speaker ：
  - Jazz Yao-Tsung Wang / Associate Researcher
  - Co-Founder and Evangelist of Hadoop.TW
  - jazzwang.tw@gmail.com
- All slides and training materials could be found at
  - http://trac.nchc.org.tw/cloud
  - Be Green ! Try not to print the slides. It changes frequently.

FOSS User
Debian/Ubutnu
Access Grid
Motion/VLC
Red5
Debian Router
DRBL/Clonezilla
Hadoop



FOSS Developer
TRTC WSU/
Haduzilla /
Hadop4Win / Ezilla

FOSS Evangelist
DRBL/Clonezilla
Partclone/Tuxboot
Hadoop Ecosystem

NARLabs

演講大綱 Agenda

Life of Big Data
From Internet of Things to Mobile Computing

Paradigm Shift of Cloud Computing
from Computing Center to Data Center

Evolution of Big Data Software Stacks
for Next Generation Data Center

Roles of *** in Big Data Supply Chain***

Conclusion

3 Buzzwords in 2013

NARLabs

三大年度熱門關鍵字

物聯網

Internet of Things

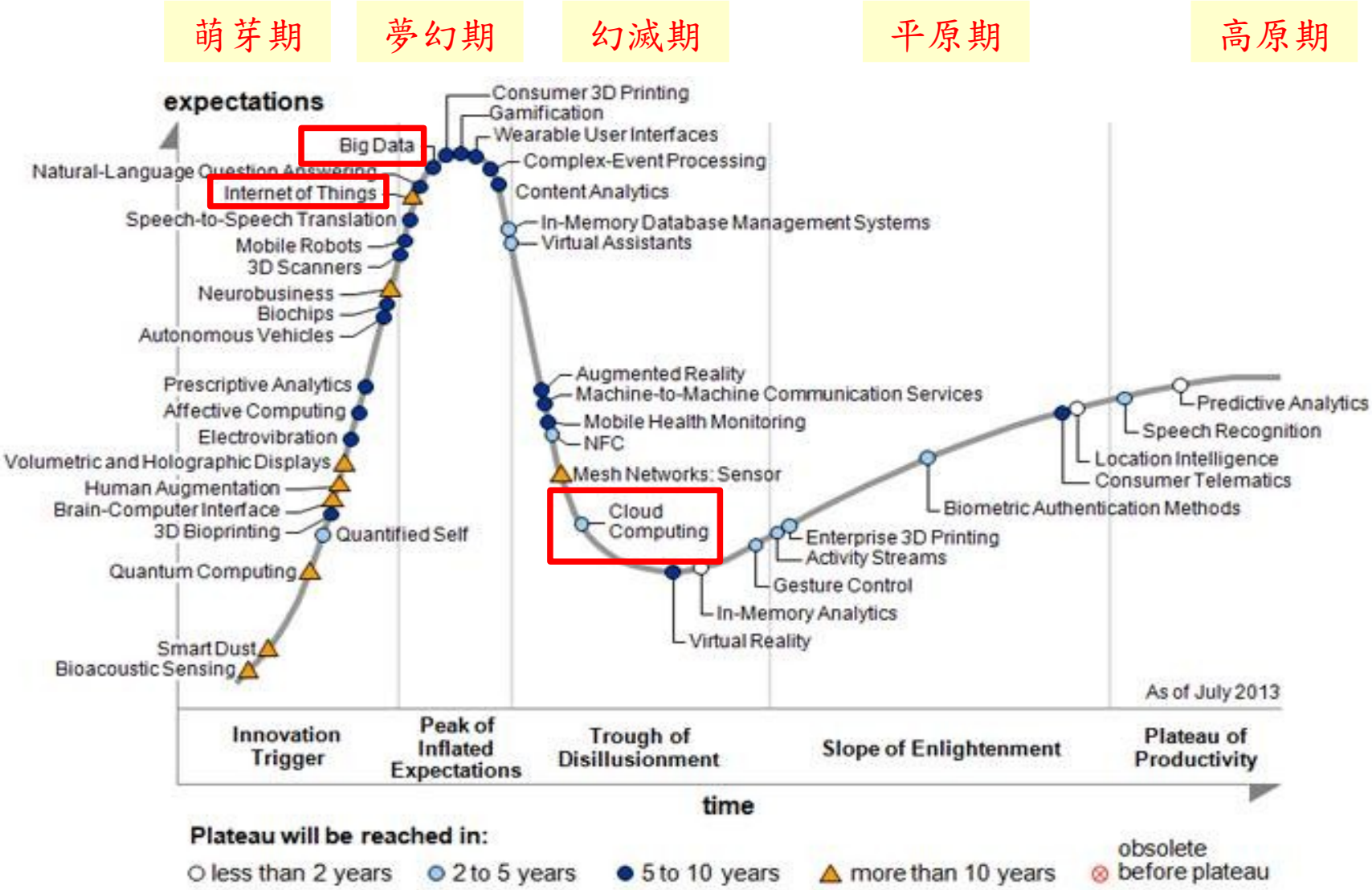
雲端運算

Cloud Computing

巨量資料

Big Data

市場現況： Gartner Hype Cycle 2013



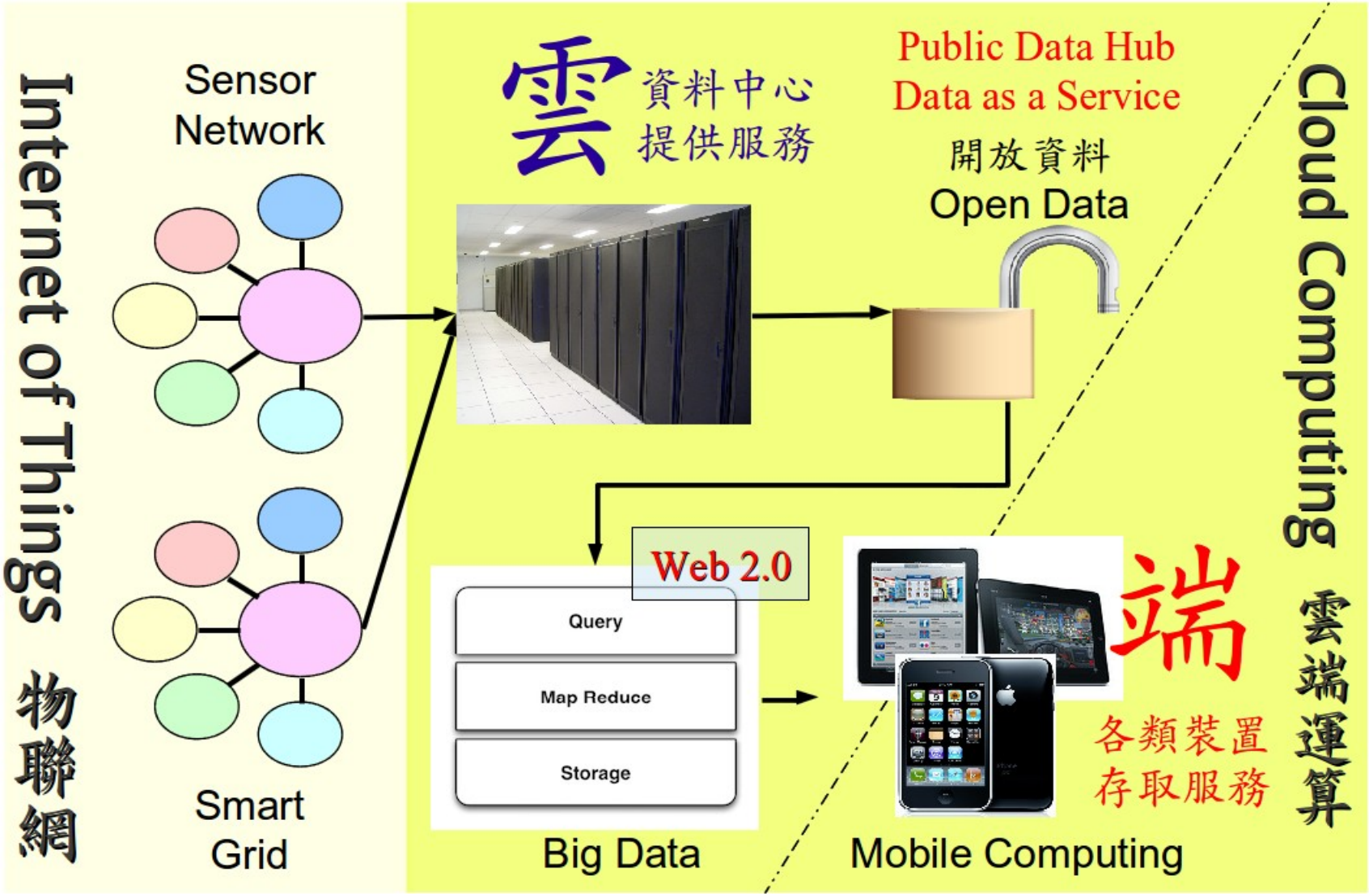
NARLabs

This is a **STORY** about **DATA FLOW**

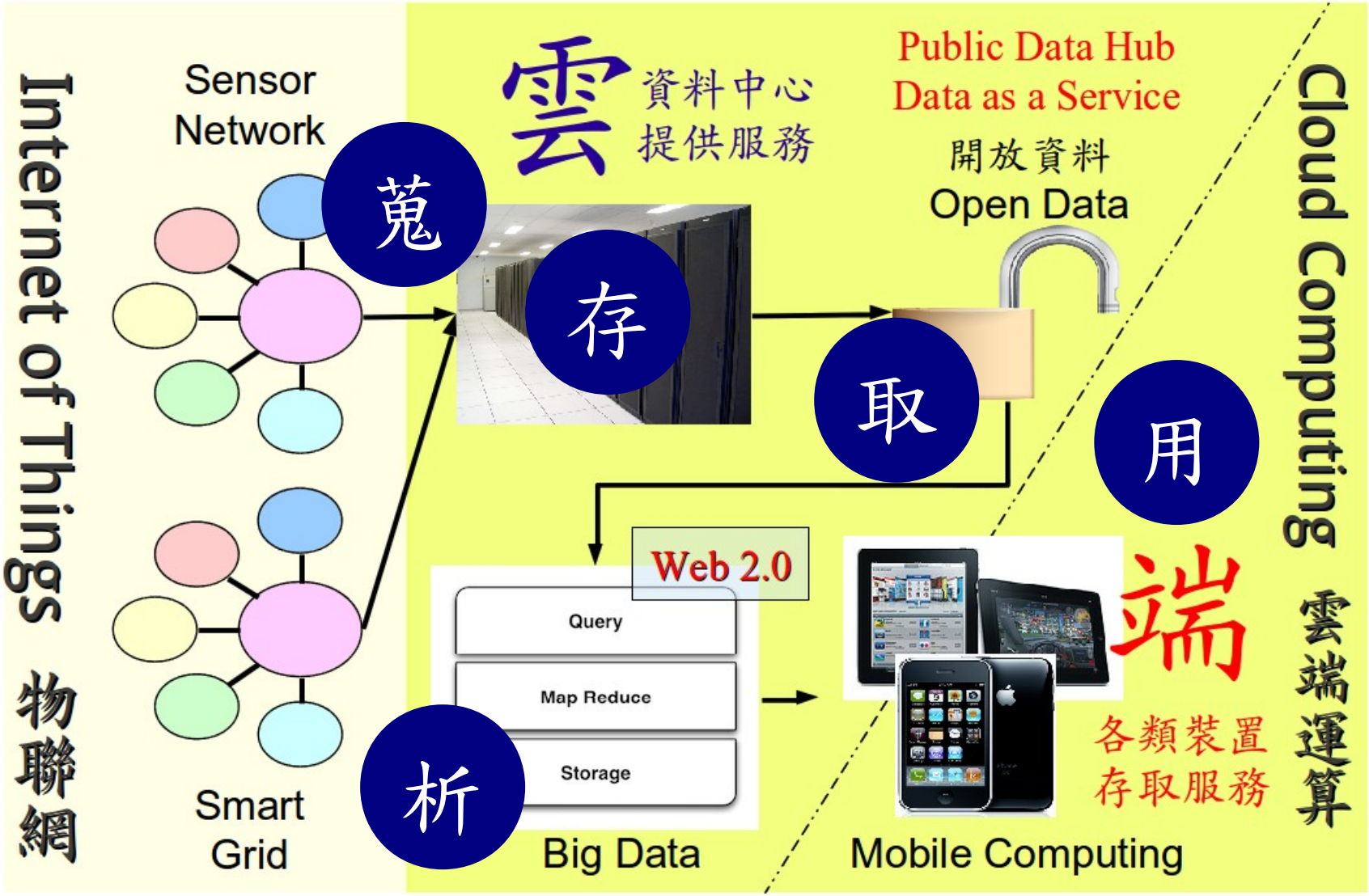
今天要講一個關於
資料的漫長旅程

(也是講者過去十一年的職涯縮影)

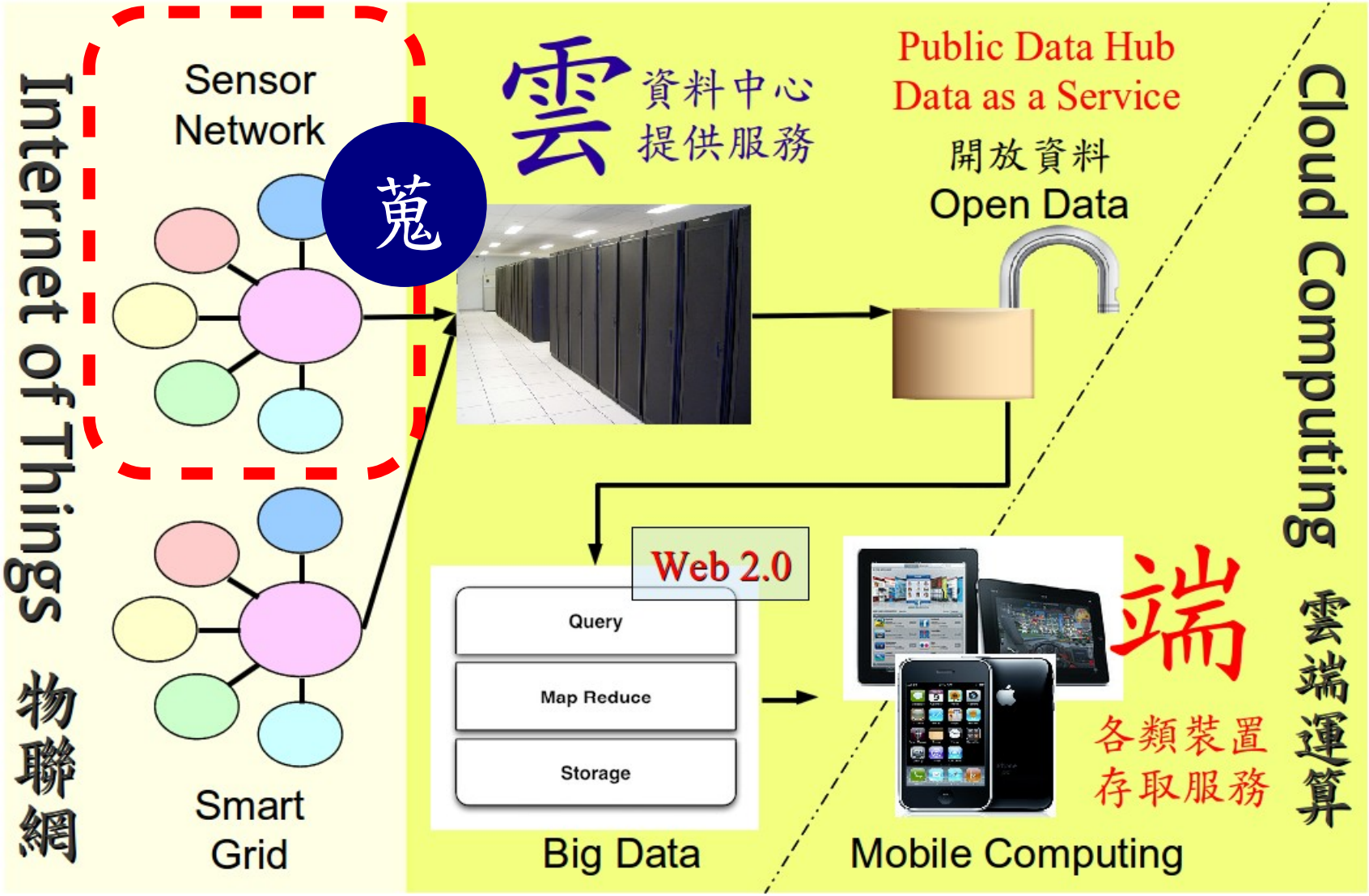
巨量資料的奇幻漂流
Life of Big Data



巨量資料的五大生命週期
5 Stage of Big Data Life Cycle

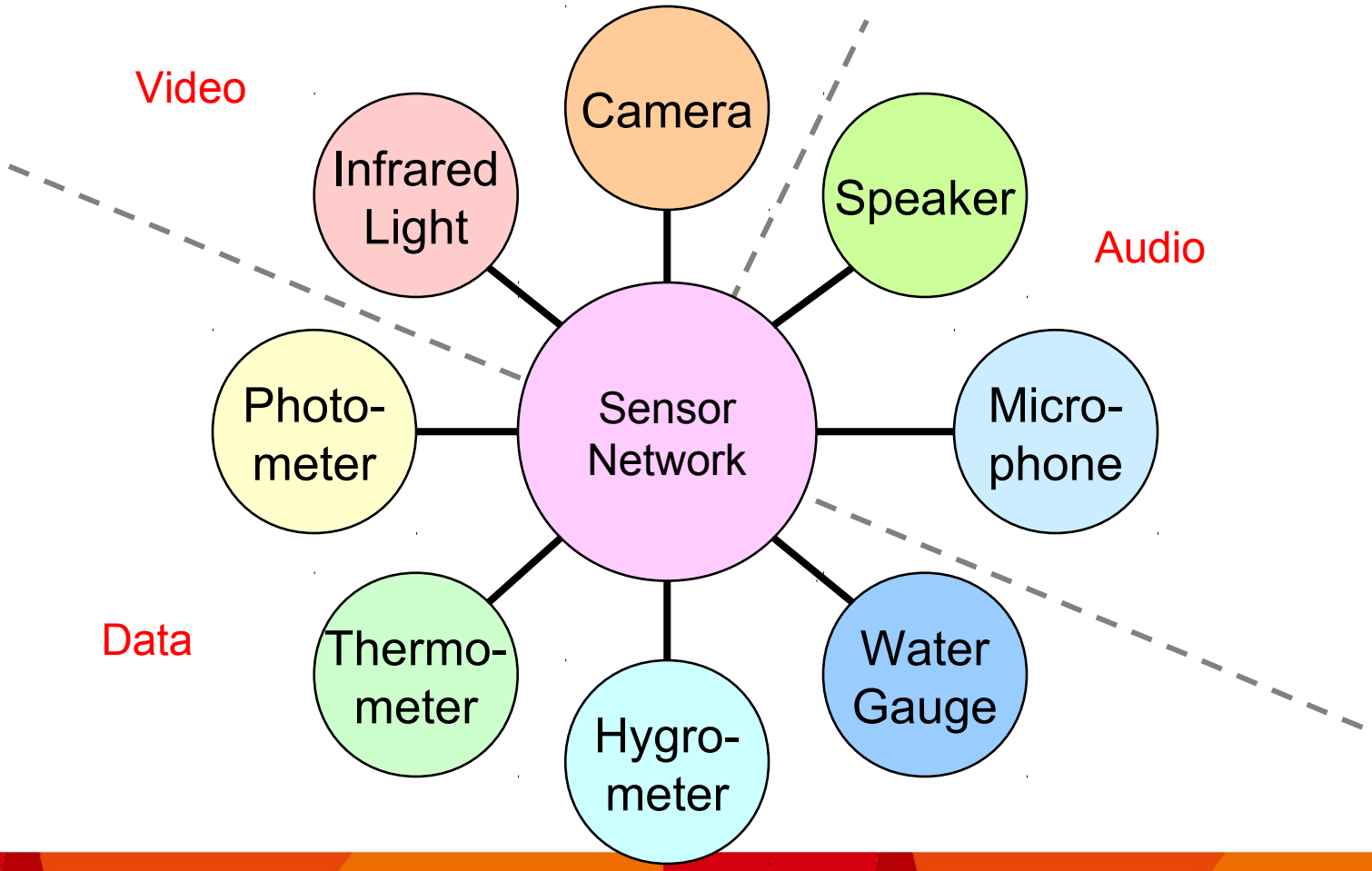


巨量資料的五大生命週期
5 Stage of Big Data Life Cycle



無線感測器網路
Wireless Sensor Network

感測器網路泛指由各式各樣不同的感測器所組成，透過網路鏈結成一大型的資訊蒐集網。



生態觀測格網

Ecology Grid : 2002~2008

NCHC

KING Knowledge Innovation National Grid

ECOGRID

Ecogrid Insight | Events | Ecosites | Ecogrid on Media | Data Query |

Ecogrid project is a joint effort among National Center for High-Performance Computing (NCHC), Taiwan Ecological Research Network (TERN) and Taiwan Forestry Research Institute (TFRI).

Using state-of-the-art grid technology, scientists at NCHC established a cyberinfrastructure for ecological research, the Ecogrid, which integrates geographically distributed sensors, computing power, and storage resources into a uniform and secure platform. Ecologists can conduct researches on this platform, including data acquisition, data analysis, and data sharing. The real time intervention capability of the Ecogrid system allows researchers to carry out interactive/control experiments that have been impossible until now.

Global Lake Observatories Network

With the support from PRAGMA, Ecogrid is now extended to international. By the cooperation with scientists from Center for Limnology of University of Wisconsin, San Diego Supercomputing Center (SDSC), and Center for Biodiversity of Academia Sinica, we deployed a lake observatories network connected LTER Trout Lake station in Wisconsin and TERN Yuan Yang Lake Station in Taiwan in April 2004. Through the ubiquitous web interface, scientists can easily access and compare data from both research stations in near real-time. The pioneer network was used as an example to call for broader partnership when scientists gathered at Scripp Institute of Oceanography in March, 2005 to discuss next move upon the global network.

Underwater Reef Monitoring

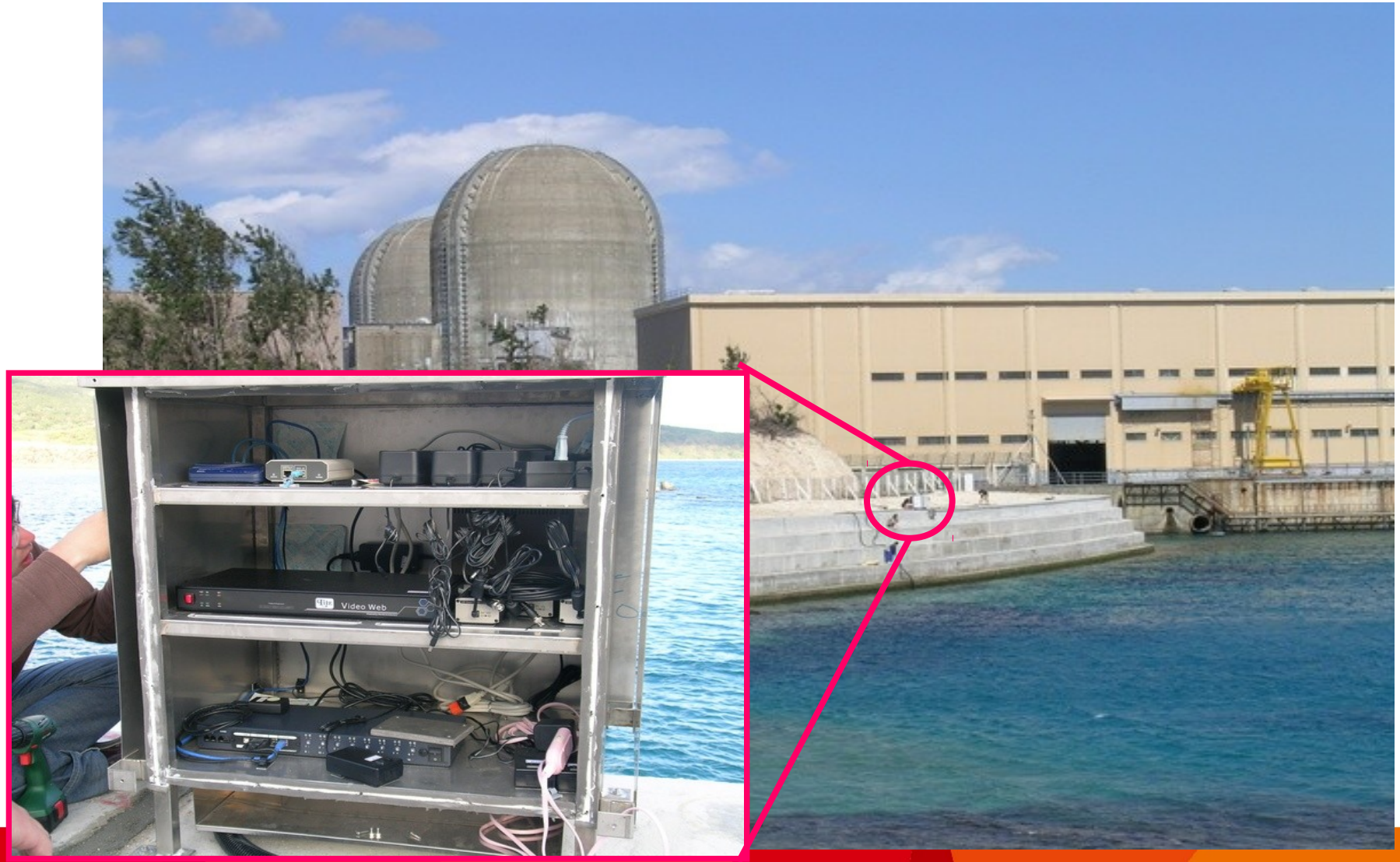
Coral reefs are precious because they are the most biologically diverse and productive marine ecosystems on earth. In addition, revenue from tourists attracted to the beauty of coral reefs can be a significant source of income for human communities in these areas. Because of the important ecological and economic roles coral reef communities fulfill, resources are pulling in to study the stresses and dangers to the reefs.

With the funding from Taipower, NCHC Ecogrid team worked together with Research for Biodiversity Center, Academic Sinica implemented an underwater reef monitoring system. The system composed with 9 cameras, located at 3 sites, lively streaming bioactivities undergoing in reef communities to marine scientists' desktop.

http://ecogrid.nchc.org.tw

水下珊瑚礁與熱帶魚類生態研究
Underwater Camera

墾丁核三廠入水口 (1)



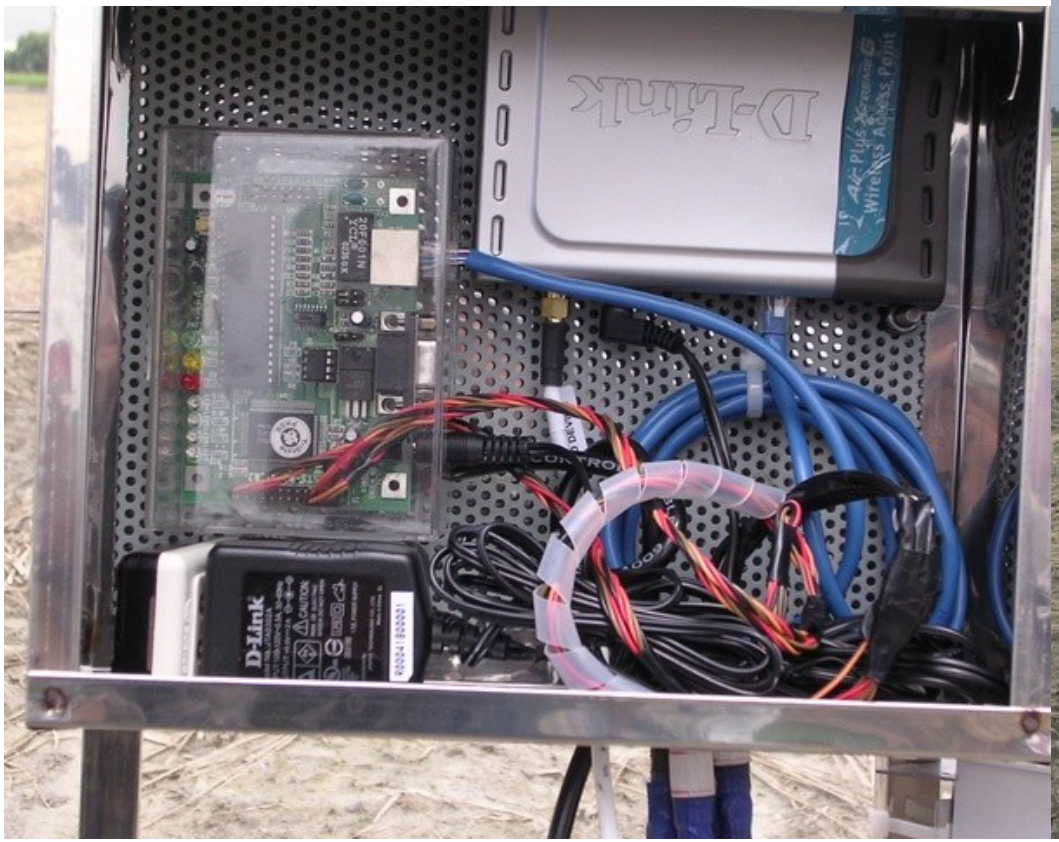
墾丁核三廠入水口 (2)

農業格網：台灣生產履歷的開端
Agriculture Grid

http://sensor.nchc.org.tw/tounan

架設在農田裡的感測網路

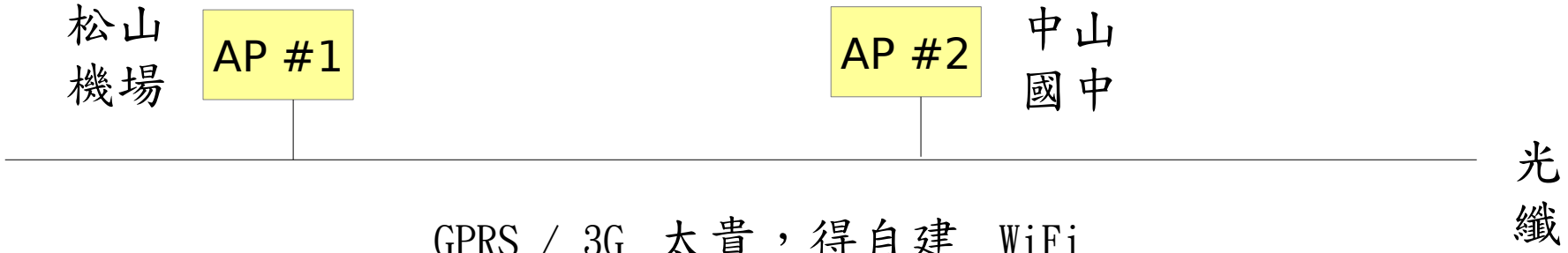
被動式的田間伺服器



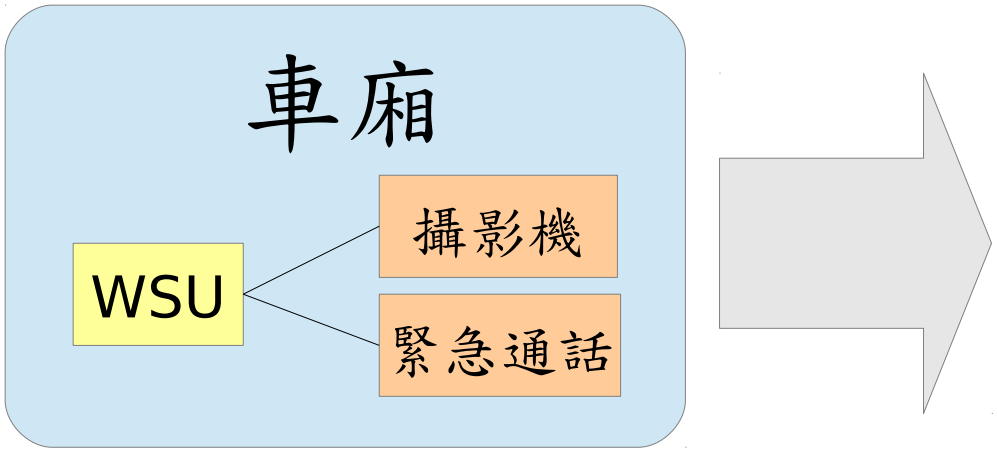
文湖線無線管理單元
TRTC Neihu WSU

Source: http://www.flickr.com/photos/rail02000/5292653579/lightbox/

無線網路快速漫遊
Wireless Fast Roaming



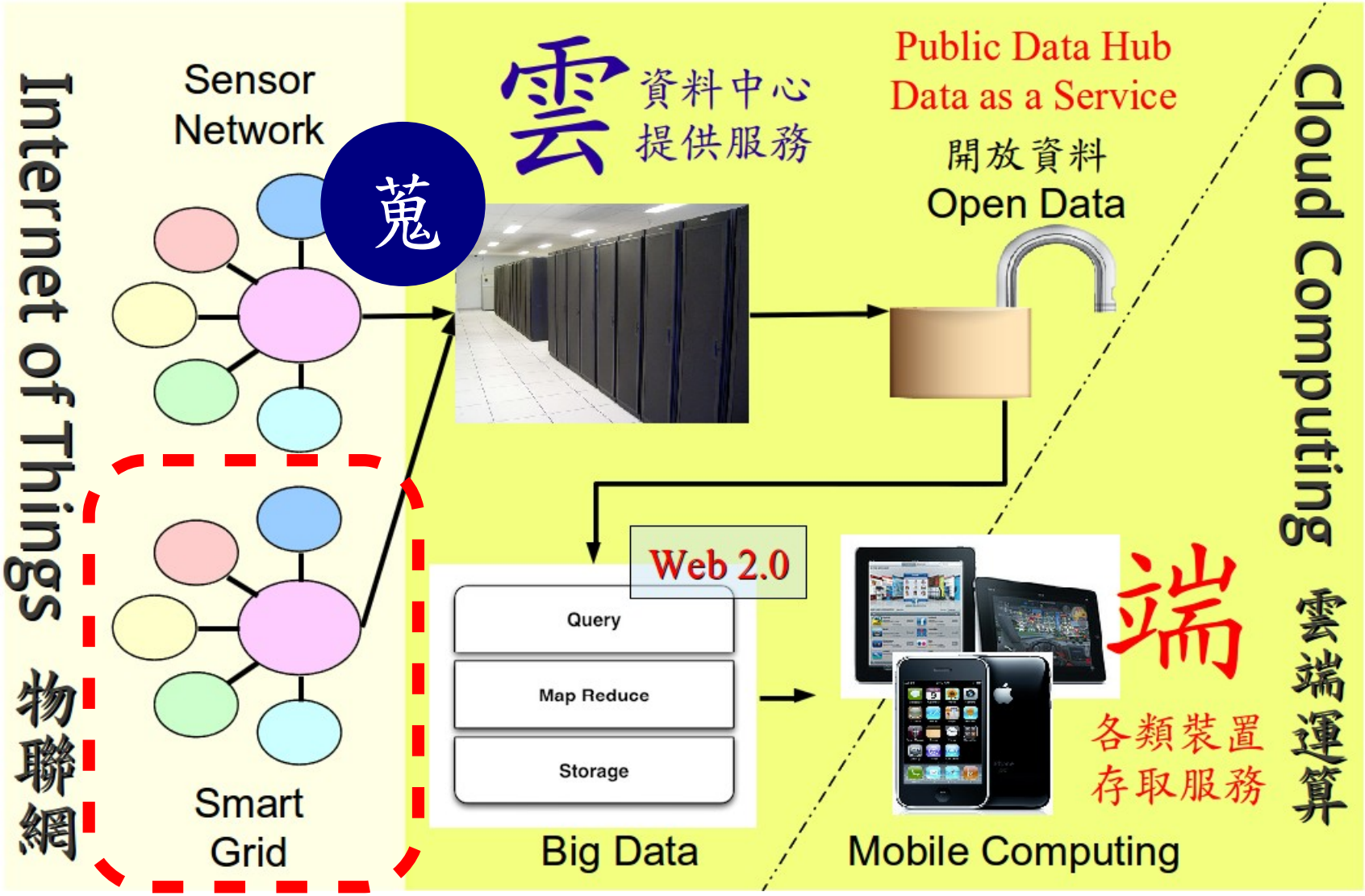
GPRS / 3G 太貴，得自建 WiFi
間隔 400 公尺，時速 60 公里，
如何快速地在無線基地台切換呢？



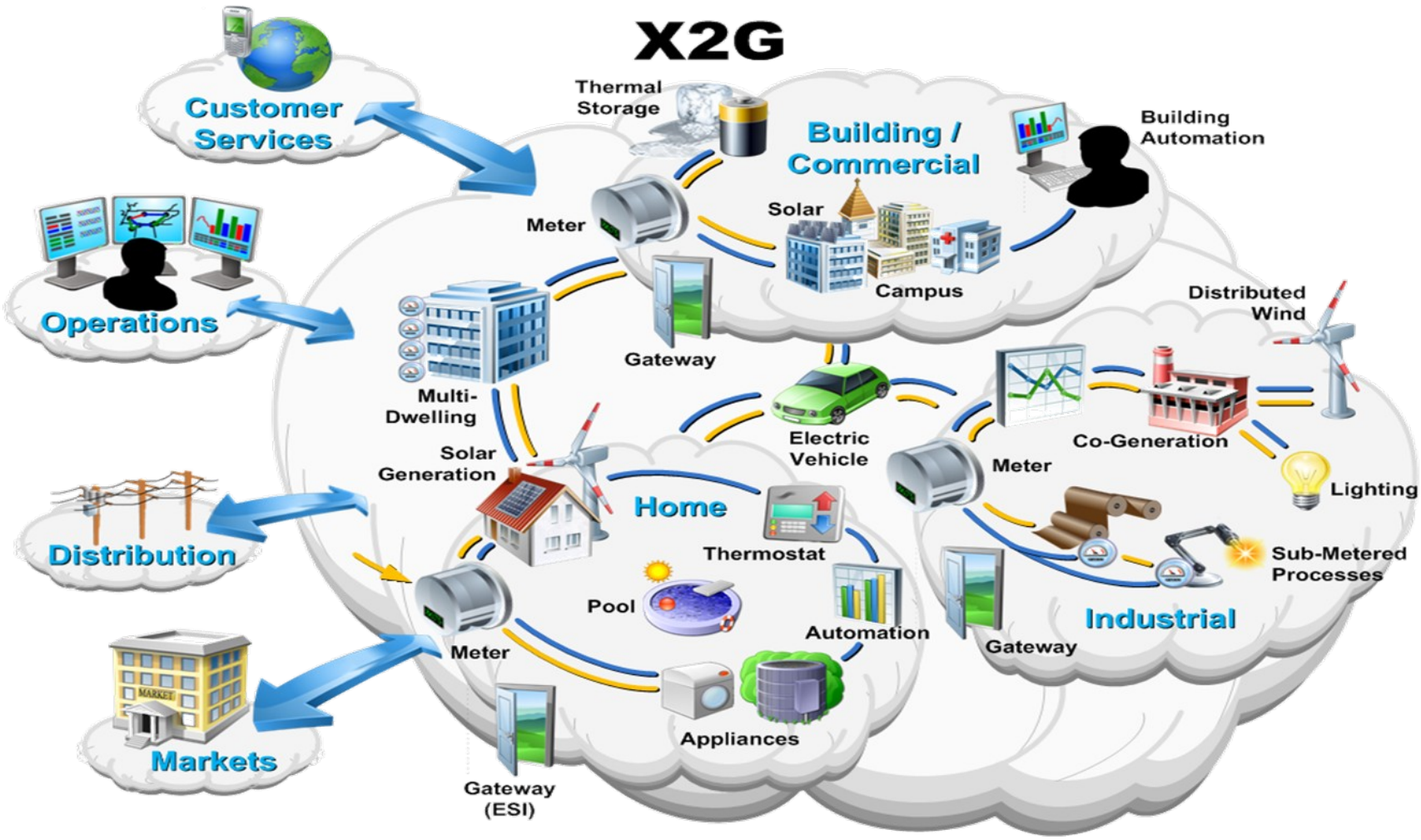
300ms 完成漫遊

切換時聲音跟影像資料不能中斷！

巨量資料的五大生命週期
5 Stage of Big Data Life Cycle



智慧電網的新政策
Smart Grid



圖片來源：NIST，美國國家標準局

物聯網的時代來臨
Internet of Things

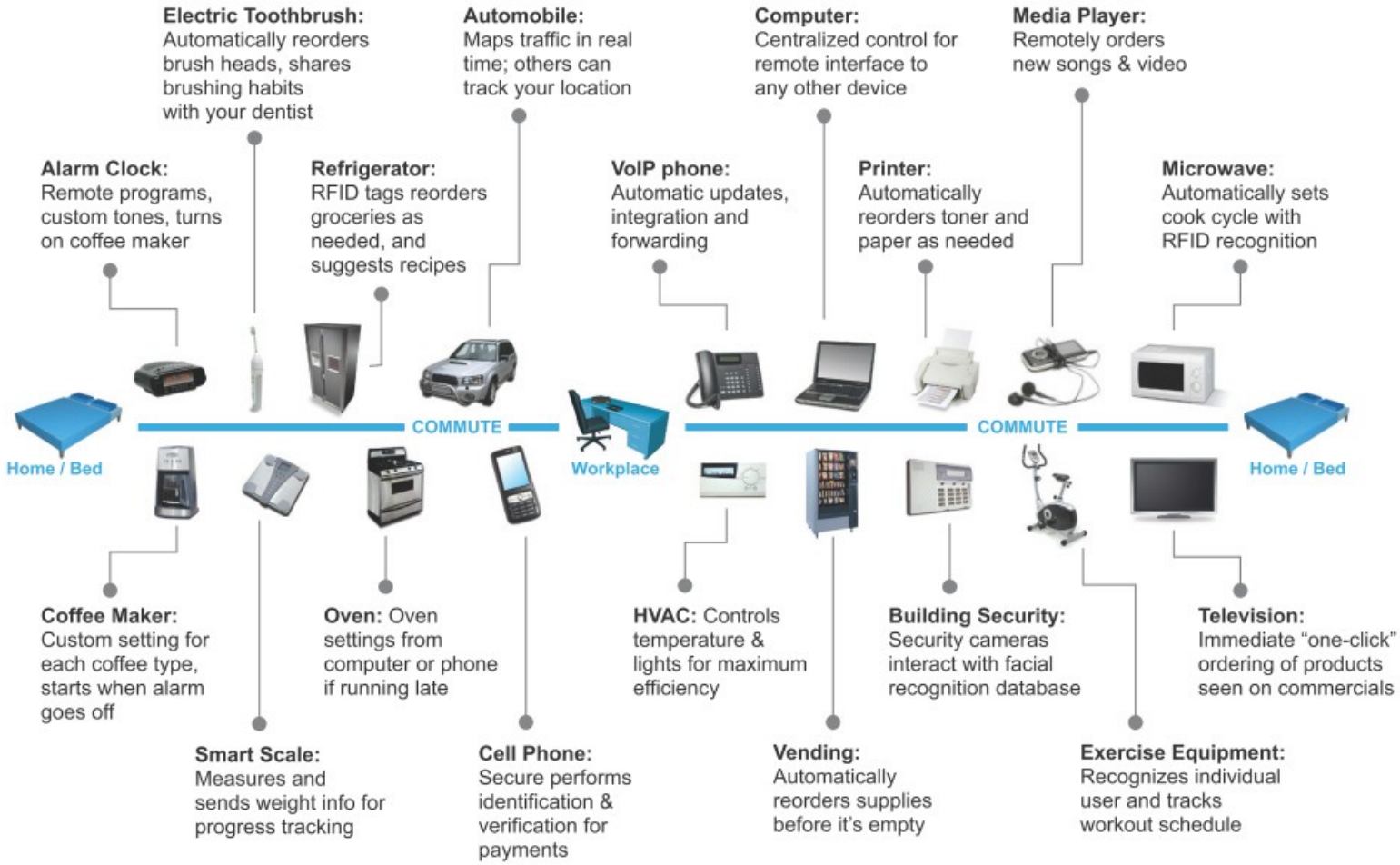


Figure 3. The Internet of Things

Source: "Attacks on Mobile and Embedded Systems: Current Trends" by Mocana

Wifi體重計：透過網路管理你的體重

無敵小恩恩 發表於 2009年10月05日 13:54 | 收藏此文

讚 0

+1 0

網路入侵到家電用品只是時間的問題，現在體重計也有內建無線網路，透過網路能將你每天量的體重都紀錄下來，不用筆記也能做好更準確的體重管理。有了這麼方便的上網體重計，這下沒有藉口懶得管理你的體重了。 懶人減肥專用機

不只是記錄體重，這台體重計還能算出你的BMI，不但能將資料傳到網站上去，也能透過Wifi傳到iPhone上，經過每天的測量之後，能在網路上看到曲線圖，進而管理你的體重，如果上升的太快就表示最近吃太多動太少，對於懶得記錄自己體重的人來說，這是一台不錯的監督機器，而一台要價159美金，約5000元新台幣，並不便宜，昂貴的價格要普及化還是有一段很長的時間，而且台灣也沒販賣，要買到也不容易。

Source: http://www.techbang.com/posts/1258-wifi-scales-through-the-internet-to-manage-your-weight

【另類解釋】

物聯網＝連結多個異質感測網路

Ex. 智慧家庭

氣象局＋鬧鐘＋窗簾＋咖啡機＋智慧車輛

M2M Protocol
(Machine To Machine)

哪一個才是跨感測裝置的通訊格式標準？

SSD 在物聯網的應用情境
本地端儲存 (Local Storage)

感測網路(物聯網)的挑戰：

1. 電力！！
2. 上傳網路頻寬？！
3. 暫存資料要保留多久？！ (容量上限)

解決方案：硬碟 → CF → SD → MMC？

演講大綱 Agenda

Life of Big Data
From Internet of Things to Mobile Computing

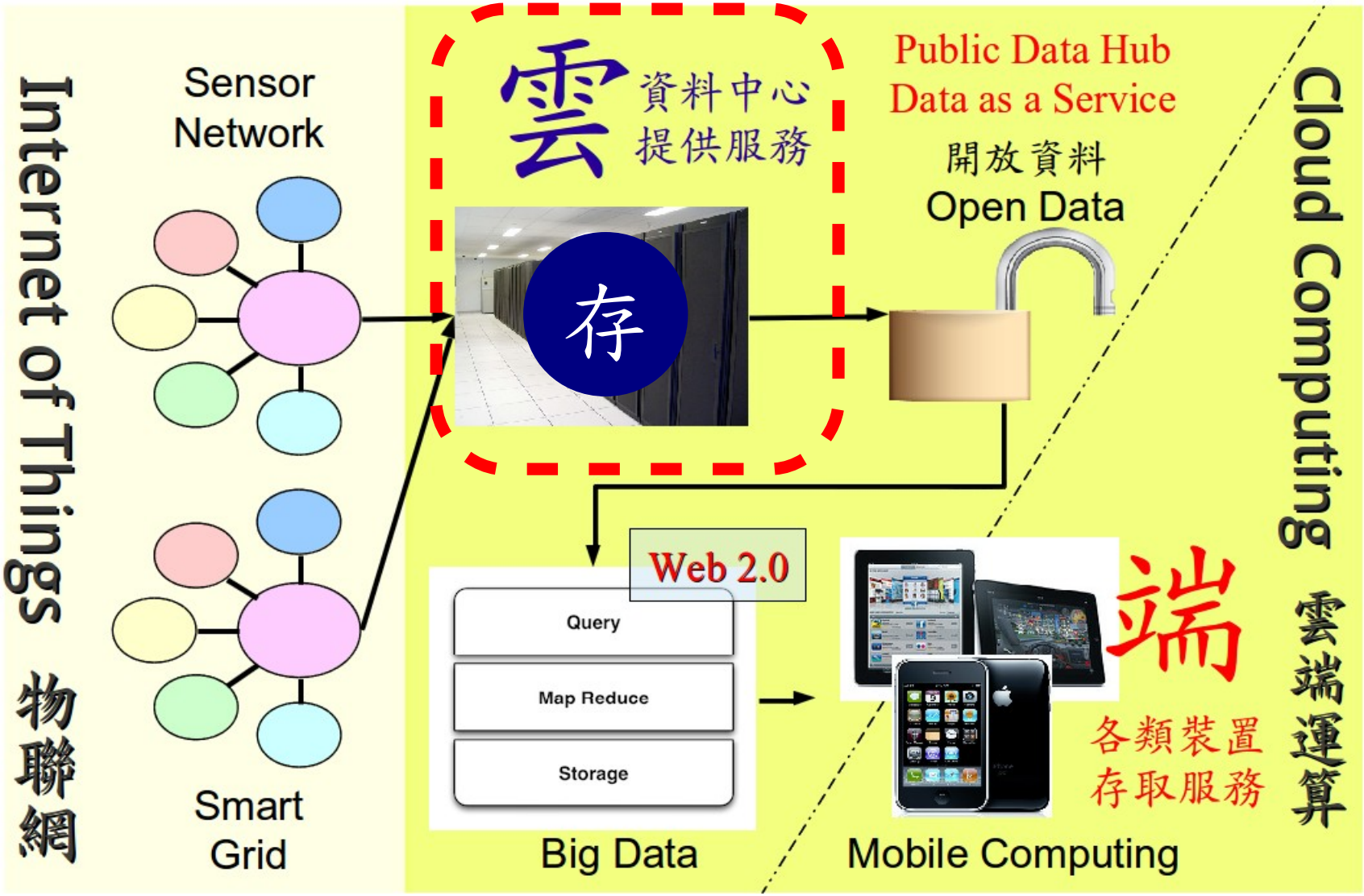
Paradigm Shift of Cloud Computing
from Computing Center to Data Center

Evolution of Big Data Software Stacks
for Next Generation Data Center

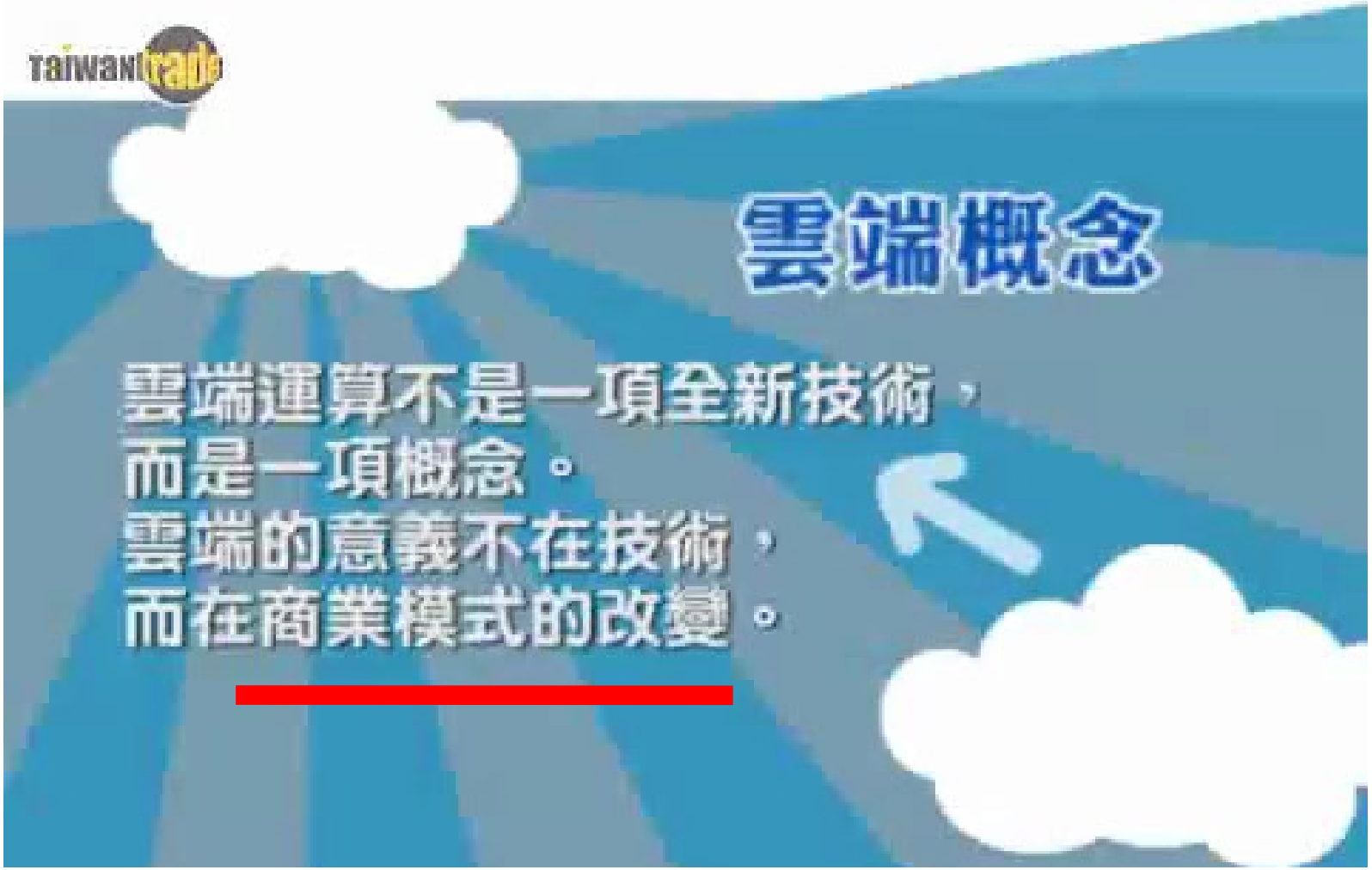
Roles of *** in Big Data Supply Chain***

Conclusion

巨量資料的五大生命週期
5 Stage of Big Data Life Cycle



什麼是雲端運算啊？
What is Cloud Computing ?



http://www.youtube.com/watch?v=bJLSAcU6O3U 當紅「雲端運算」 你瞭解了嗎？
http://www.youtube.com/watch?v=VIMtd3nfPqc 雲端產業 8 分鐘就上手

雲端商業模式的典範轉移
Paradigm Shift of Cloud Business Model !!

NARLabs

軟體從買斷變成租賃

Office 2007 →
Google Docs / Office 365

硬體從買斷變成租賃

PC / Server →
Hosting / Colocation →
Amazon EC2 / S3

軟體從單機變成行動

Outlook → Webmail →
Mail Web Apps →
Mail Mobile Apps

硬體從固定變成行動

PC / Server →
NB / Tablet →
Pad / Mobile

雲端運算的標準定義
National Definition of Cloud Computing

NARLabs

5 Characteristics 五大基礎特徵

4 Deployment Models 四個佈署模型

3 Service Models 三個服務模式

1. On-demand self-service.
隨需自助服務

2. Broad network access
隨時隨地用任何網路裝置存取

3. Resource pooling
多人共享資源池

4. Rapid elasticity
快速重新佈署靈活度

5. Measured Service
可被監控與量測的服務

雲端運算的四種佈署模型
4 Deployment Models of Cloud Computing

NARLabs

Public Cloud
公用雲端

amazon web services™
Microsoft
Google

Target Market is S.M.B.
主要客户爲
中小企業

Dynamic Resource Provisioning between public and private cloud
私有雲端動態根據計算需求
調用公用雲端的資源

Hybrid Cloud

以大型企業
爲主要客户
Enterprise is key market

IBM

私有雲端
Private Cloud

Community Cloud
社群雲端

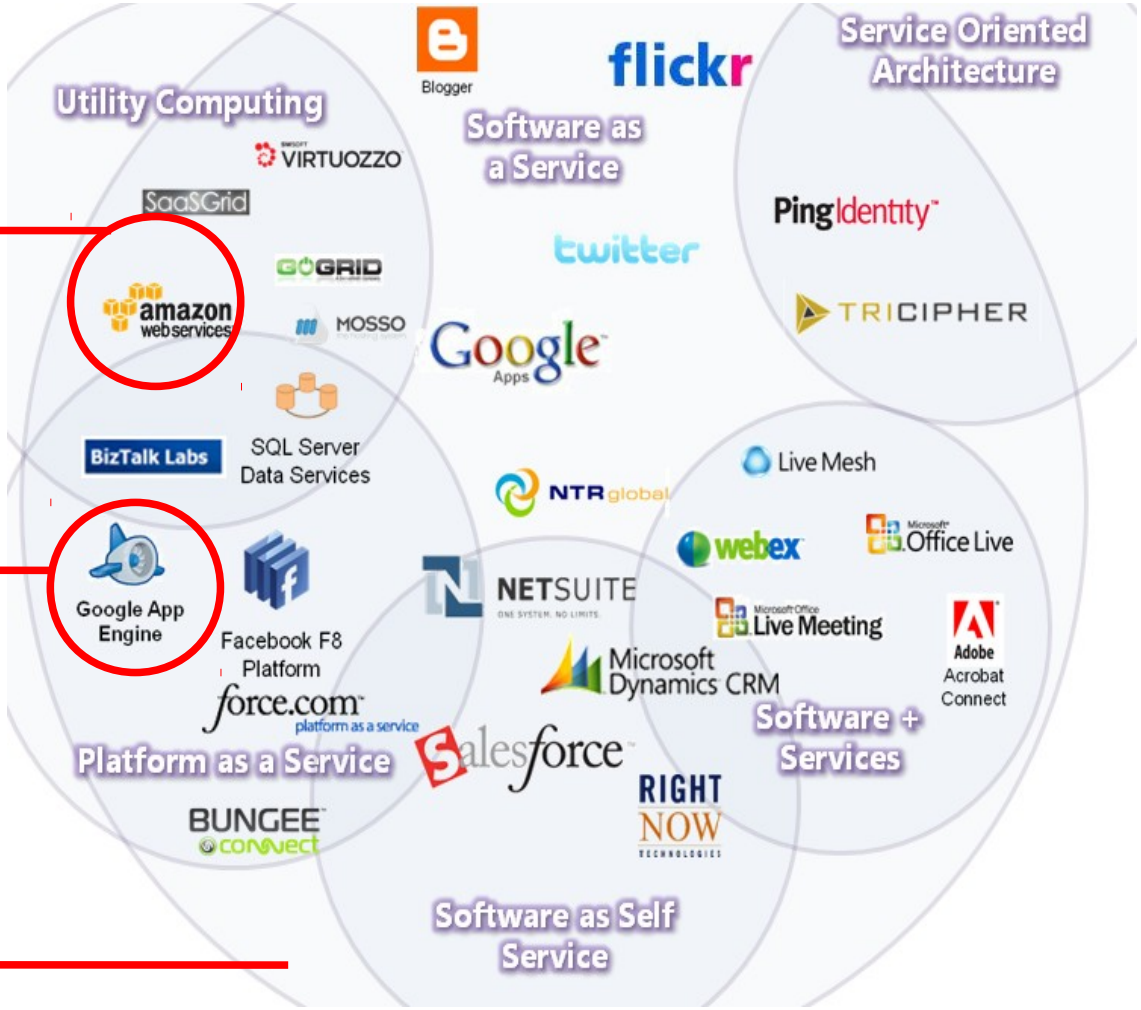
Academia 學術爲主

三種服務模式（商業模式市場區隔）
3 Service Models of Cloud Computing

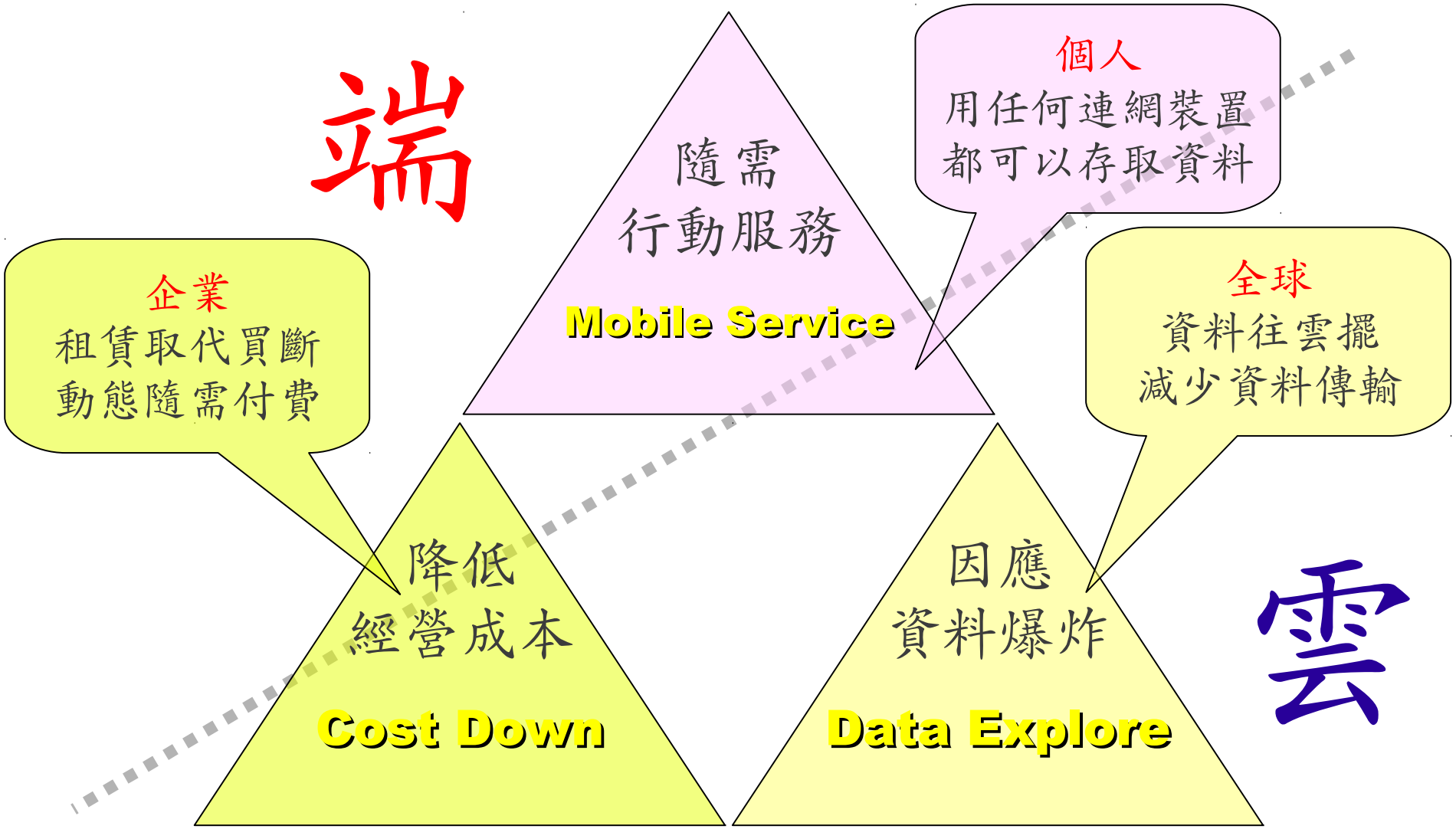
IaaS
Infrastructure as a Service
架構即服務

PaaS
Platform as a Service
平台即服務

SaaS
Software as a Service
軟體即服務



雲端運算的關鍵驅動力
Key Driving Forces of Cloud Computing



三種服務模式 vs. 三大關鍵技術
Three Key Technologies !!

NARLabs

隨需
行動服務
Mobile Service

因應
資料爆炸
Data Explore

降低
經營成本
Cost Down

SaaS
Software as a Service
軟體即服務

PaaS
Platform as a Service
平台即服務

IaaS
Infrastructure as a Service
架構即服務

Web 2.0
網頁服務

Big Data
資料分析

Virtualization
虛擬化技術

It's the real story of NCHC

NARLabs

國家高速電腦中心　　（計算為王）

- 1992 年　經行政院核准成立，進駐新竹科學園區
- 1993 年　正式對外運轉提供高速計算服務

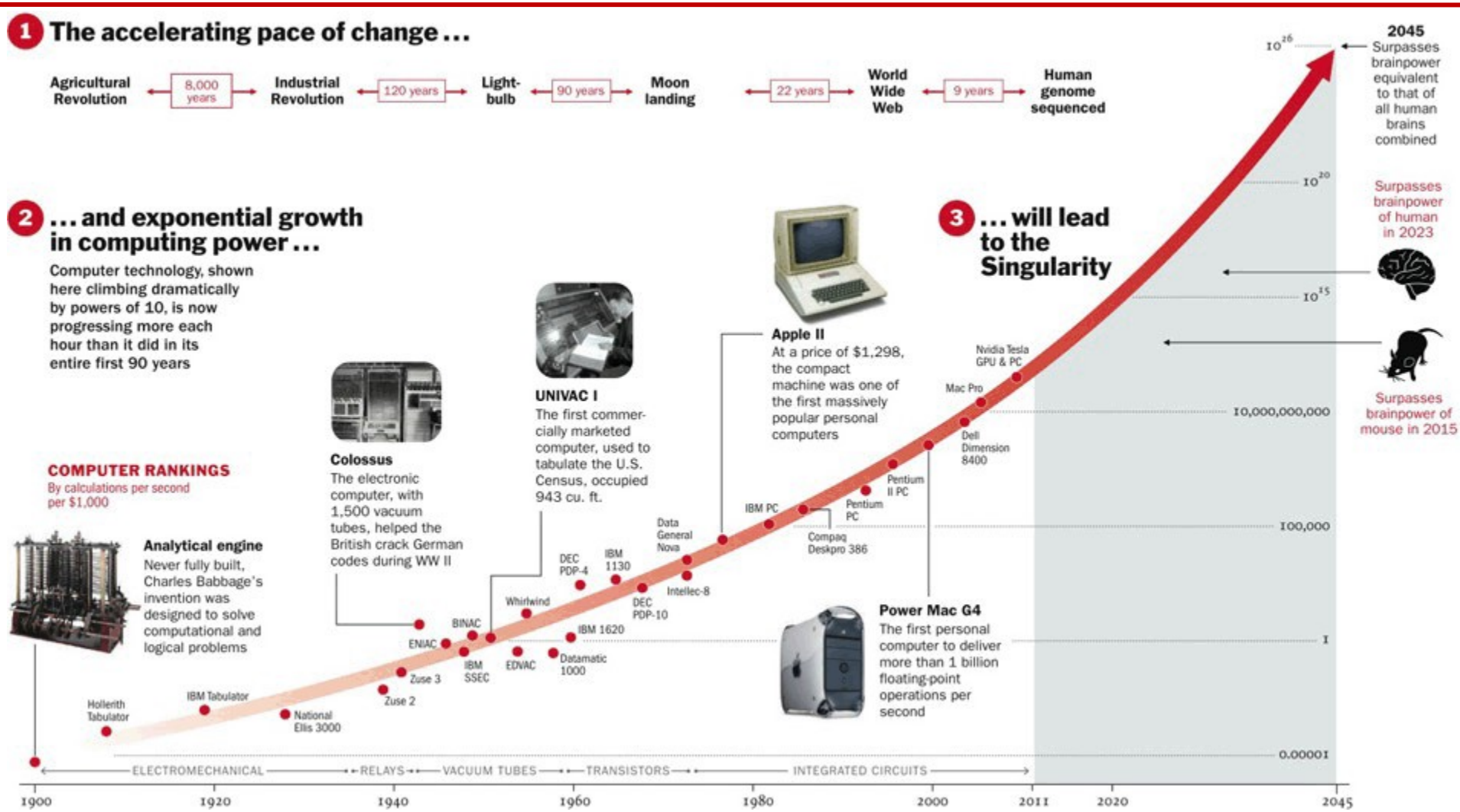
國家高速網路與計算中心　（網路為王）

- 1999 年　正式提供研究網路連線服務（ TANet2 ）
- 2002 年　『挑戰 2008 』亞太中樞計畫（ TWAREN ）　← 到職
- 2003 年　轉制財團法人，正式更名

國家高速網路與計算中心　（資料為王）

- 2008 年　雲端運算研究小組
- 2009 年　提供 Hadoop 實驗叢集服務
- 2013 年　宣示成為國際級巨量資料中心

典範轉移的時間間距據愈來愈短



Source : TIME Magazine, "2045: The Year Man Becomes Immortal", Feb. 10, 2011
http://content.time.com/time/magazine/article/0,9171,2048299,00.html
Image Source : http://trickvilla.com/wp-content/uploads/Moores-law-graph.gif

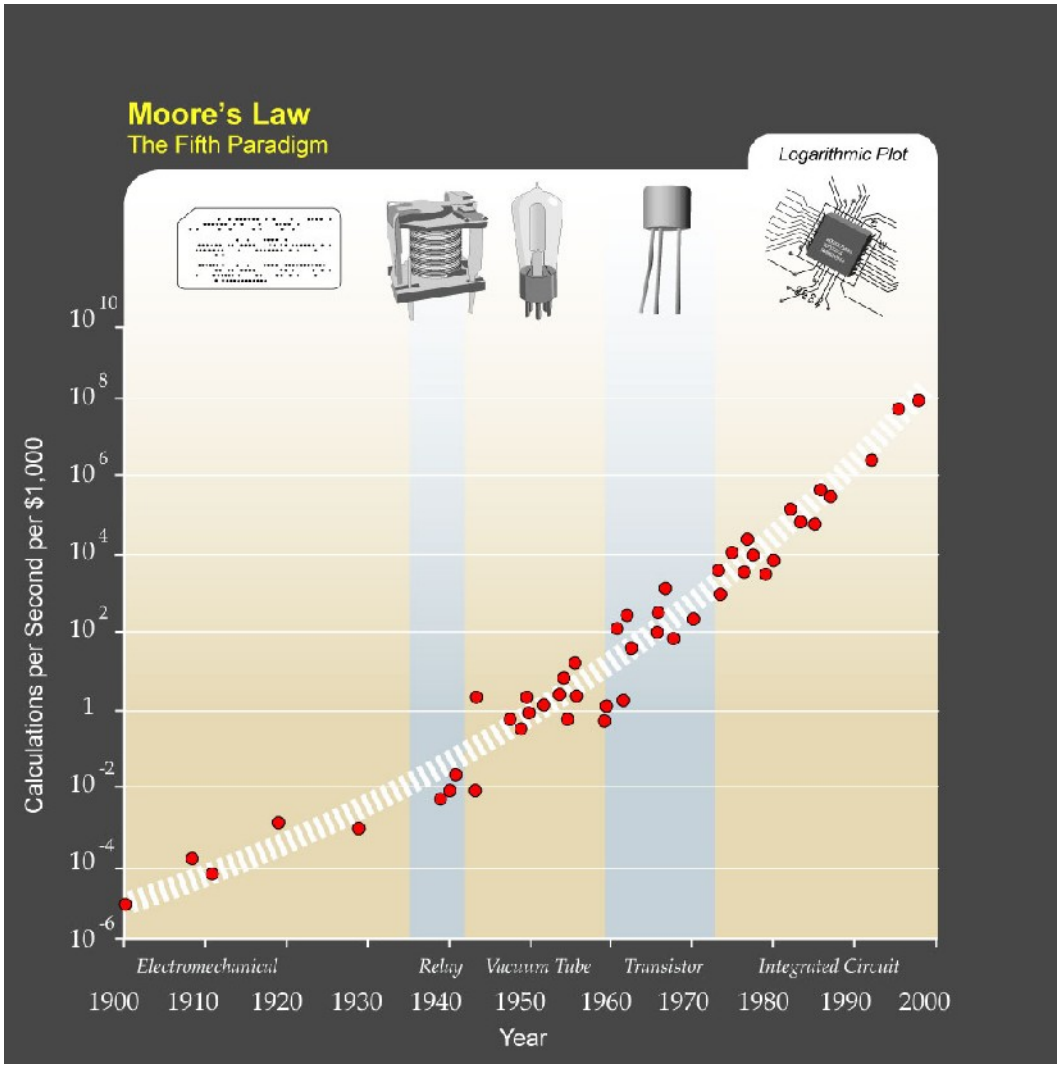
Trend of Computing – Moore's Law

NARLabs

摩爾定律是1965年由英特爾創始人之一戈登·摩爾提出來的。

積體電路上可容納的電晶體數目，約每隔24個月便會增加一倍。

英特爾執行長David House所說：每隔18個月晶片的效能提高一倍。



Source: Moore's Law, Wikipedia
http://upload.wikimedia.org/wikipedia/commons/c/c5/PPTMooresLawai.jpg

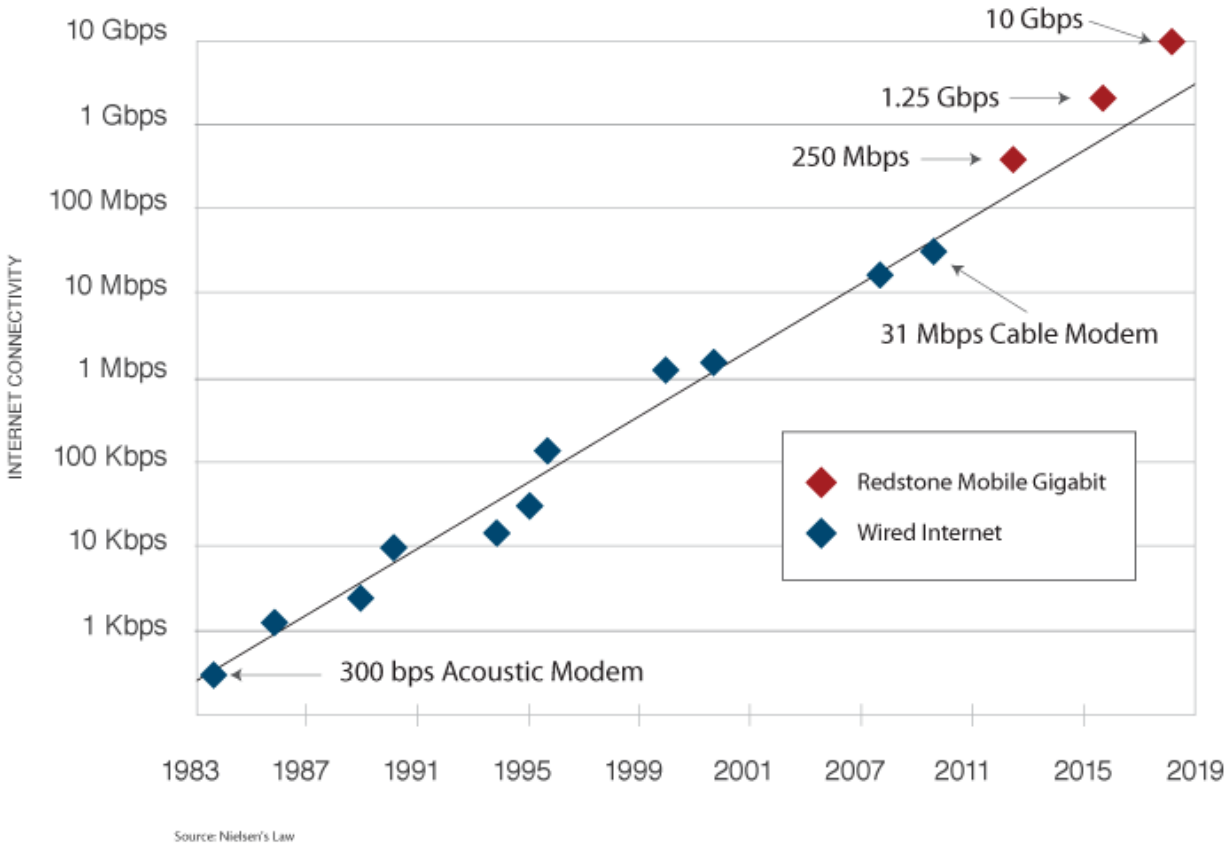
Trend of Network

NARLabs

– Nielsen's Law of Internet Bandwidth

尼爾森定律是 1998 年由 Jakob Nielsen 提出。

每隔 20 個月，網際網路頻寬會增加一倍。



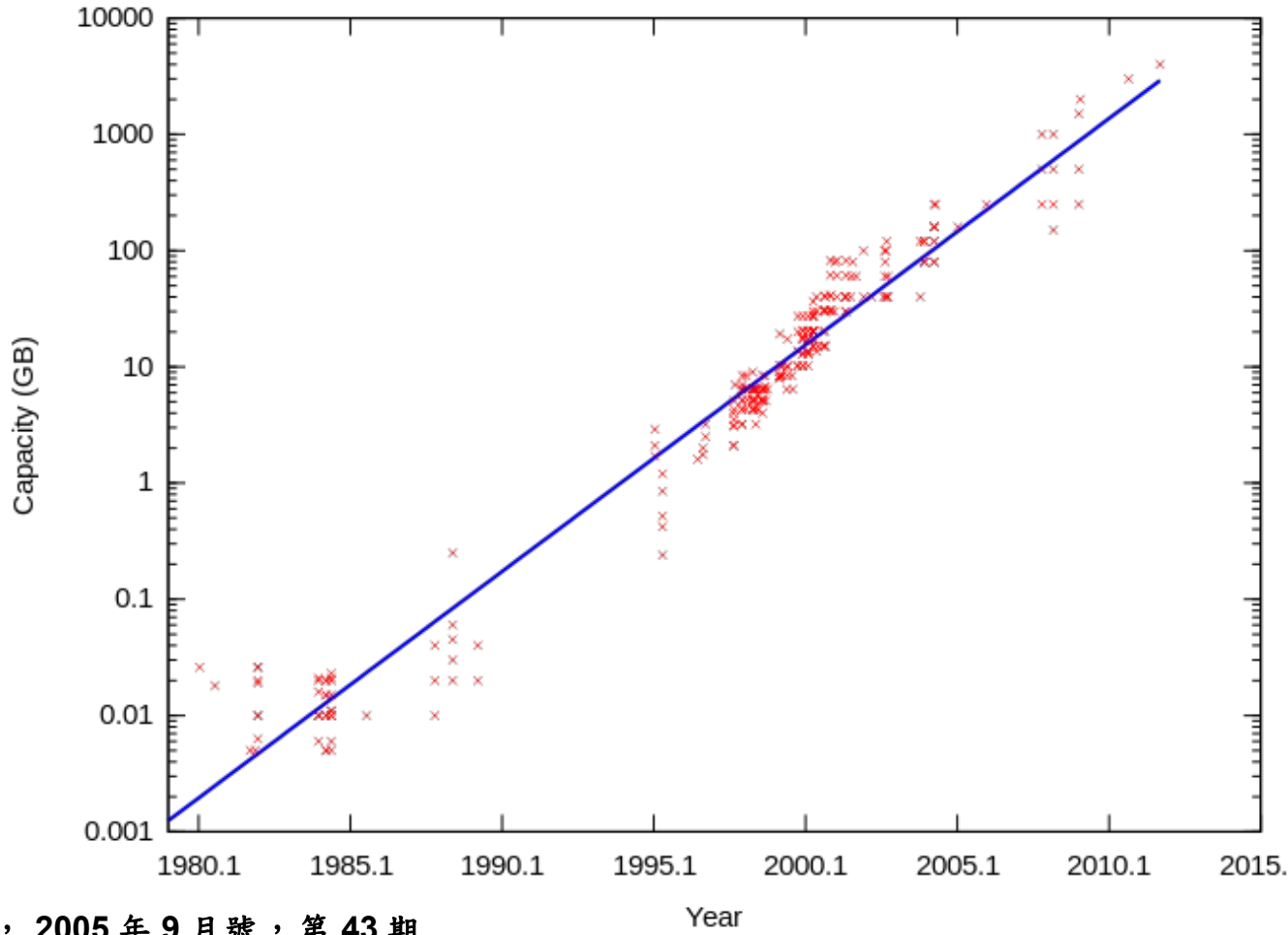
Source: "Nielsen's Law of Internet Bandwidth", April 5, 1998
http://www.nngroup.com/articles/law-of-bandwidth/
Image Source: "Nielsen's Law", May 31, 2013
http://redstone.us.com/2013/05/31/subject-nielsens-law/

Trend of Storage – Kryder's Law

NARLabs

奎德定律是 2005 年，由希捷資深研發副總馬克奎德提出的。

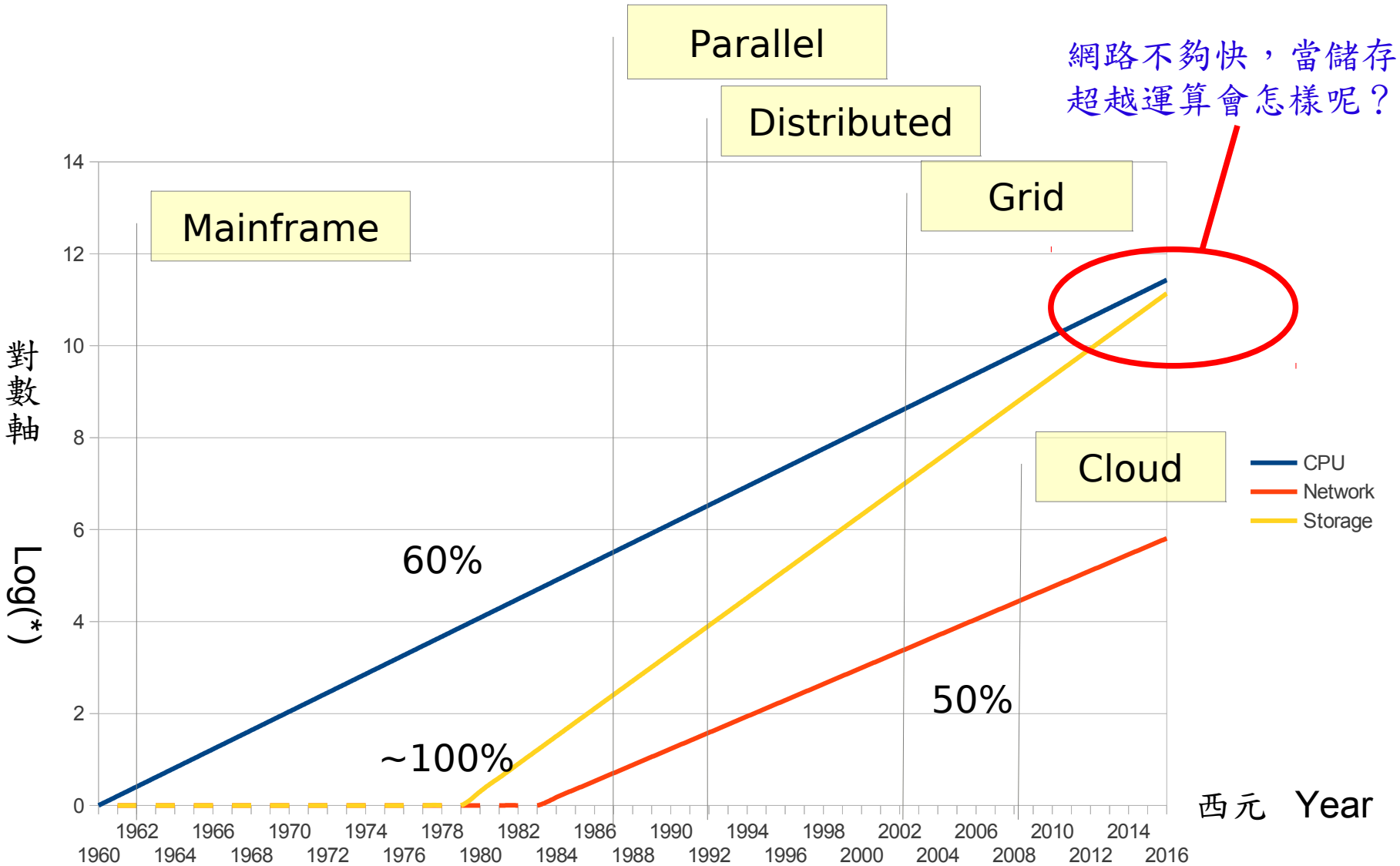
每隔 13 個月相同價格的儲存容量就會增加一倍。



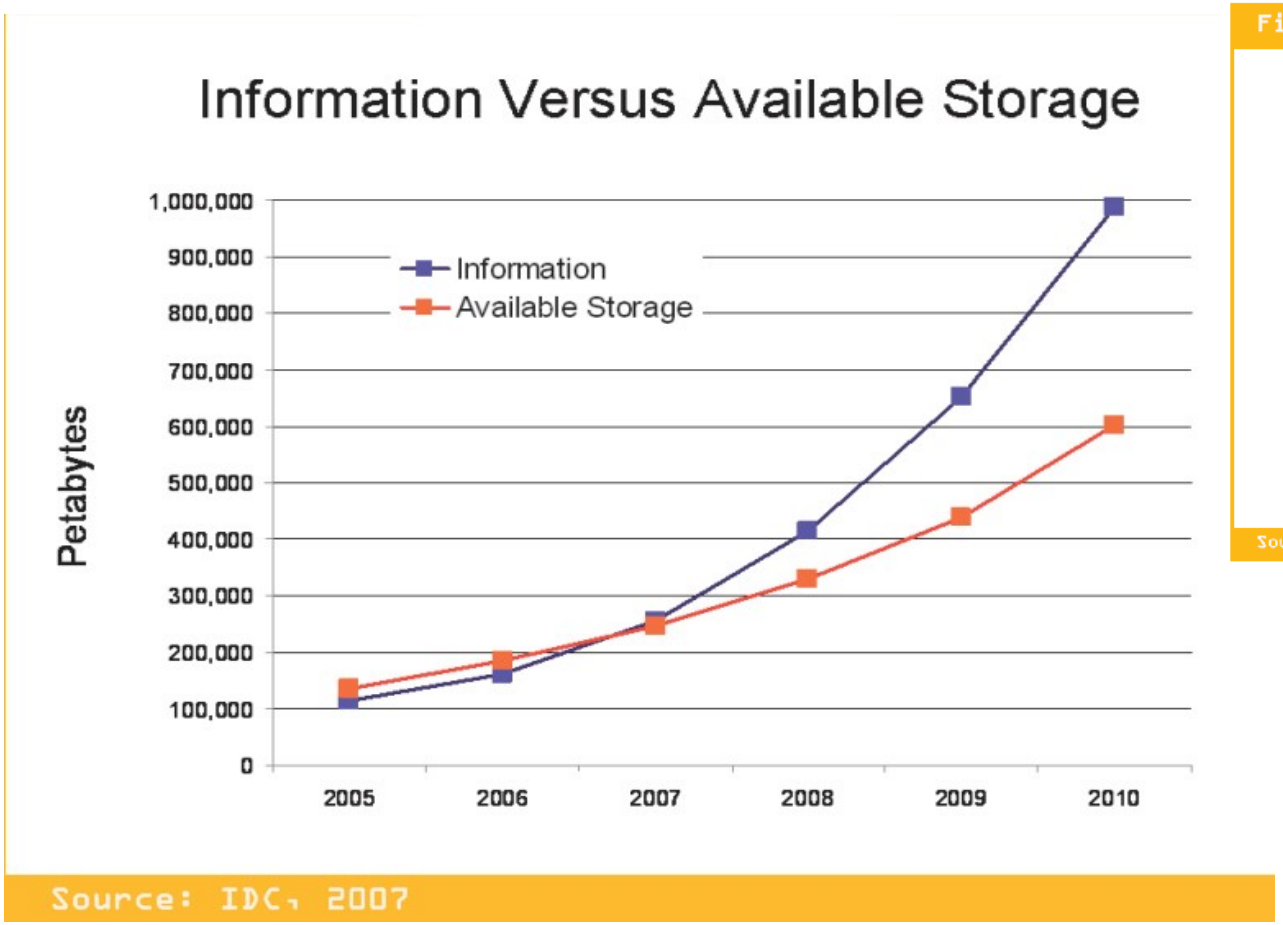
Source: 奎德定律，科學人雜誌， **2005** 年 **9** 月號，第 **43** 期
http://203.68.243.199/saweb/read.asp?docsn=2005092489
Image Source: "Hard drive capacity over time following Kryder's Law (1980-2011)", Wikepedia
http://en.wikipedia.org/wiki/File:Hard_drive_capacity_over_time.svg

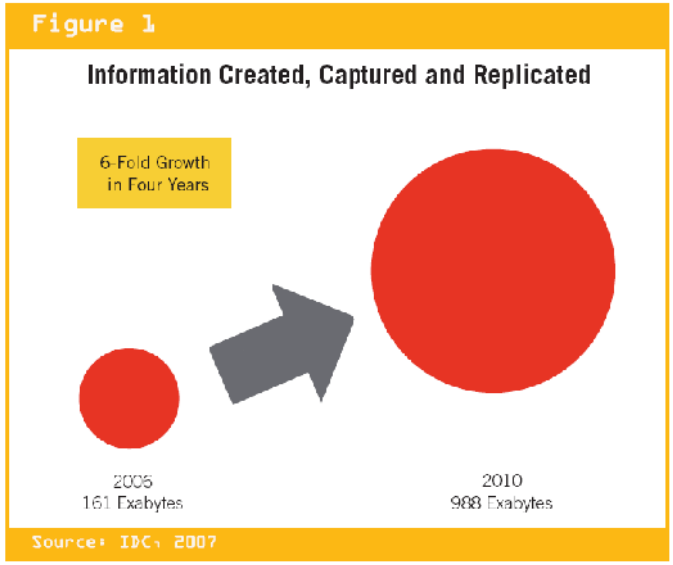
Moore's Law , Nielsen's Law , Kryder's Law

超級電腦→平行運算→分散運算→格網運算→雲端運算



始於 2007 的「資料大爆炸」時代
Data Explosion!!



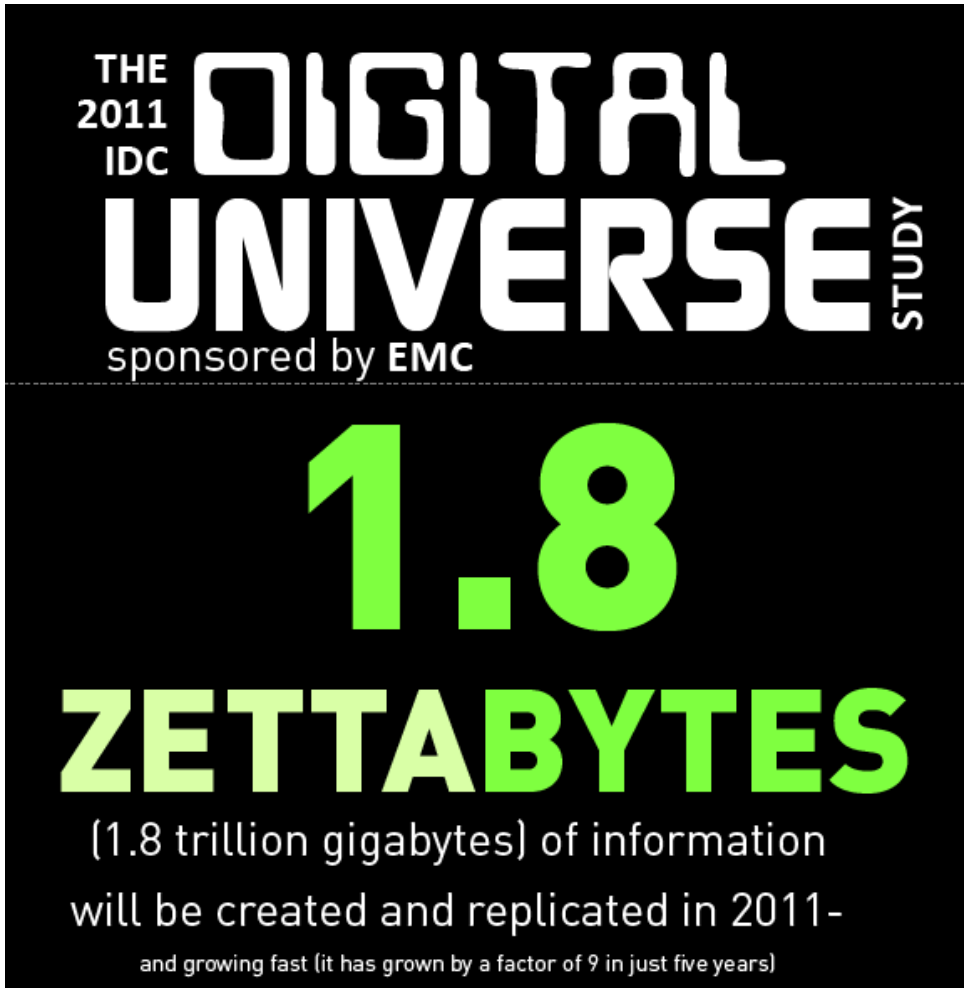


2007 年， IDC 預估
2010 年會成長六倍！
（相較 2006 年）

2006 161 EB
2010 988 EB（預測）

出處： The Expanding Digital Universe,
A Forecast of Worldwide Information Growth Through 2010,
March 2007, An IDC White Paper - sponsored by EMC
http://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf

數位宇宙以每年 1.5 倍速度成長



追蹤歷年的 IDC 數據：

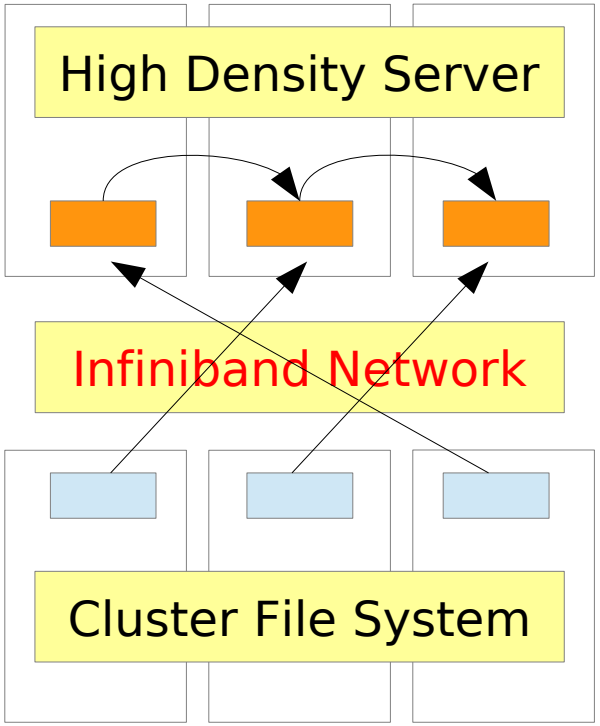
2006	161 EB	
2007	281 EB	
2008	487 EB	
2009	800 EB	(0.8 ZB)
2010	988 EB	(預測)
2010	1227 EB	(1.2 ZB)
2011	1773 EB	(預測)
2011	1800 EB	(1.8 ZB)
2012	2837 EB	(2.8 ZB)

景氣差而成長趨緩？
或受新技術抑制？

出處：Extracting Value from Chaos,
June 2011, An IDC White Paper - sponsored by EMC
http://www.emc.com/collateral/about/news/idc-emc-digital-universe-2011-infographic.pdf

Paradigm Shift in Architecture from Computing Center to Data Center

NARLabs



Computing Center
Move Data
To Compute
Message Passing

減少資料搬運

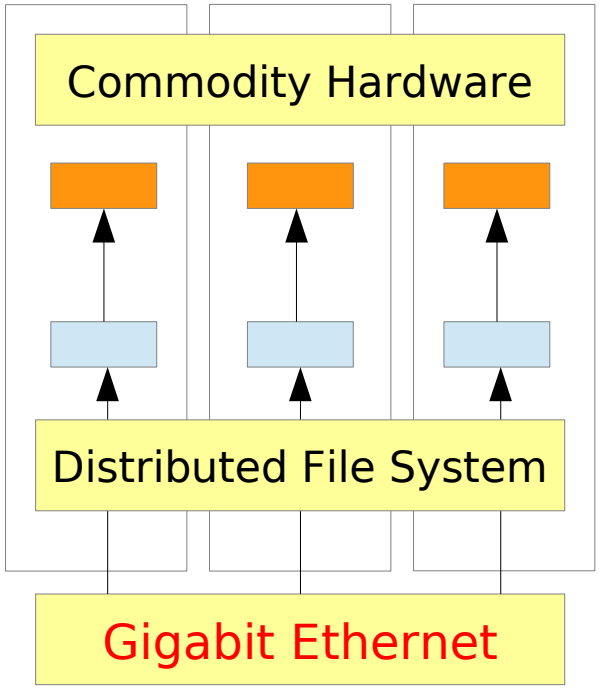
Reduce
Data Transfer

強調能源效率

Energy-
Efficiency

易於橫向擴充

High-
Scalability



Data Center
Move Compute
To Data
Share Noting

Business Model of Cloud Computing

NARLabs

規模經濟 **(Economies of Scale)**

眾人共用資料中心的軟硬體資源，降低總持有成本

資料即服務 **(Data as a Service)**

綁架你的資料，當資料越來越大，網路傳不動，你就付錢吧！

網路即通路 **(Network as a Channel)**

一雲多螢，雲端的成功關鍵在於網路頻寬普及率

運算即價值 **(Compute as a Value)**

當資料集中，連結愈多，愈能透過運算的手段，
找出群眾的智慧，就是提供給客戶最好的價值！

SSD 在資料中心(雲)的應用情境
From HPC to HTC (High Throughput Computing)

NARLabs

SSD 在高速運算領域 (High Performance Computing) 與雲端運算領域開始出現需求。關鍵因素：速度 (IOPS) ！

SOFA Software Orchestrated Flash Array

- Home
- Software
- Hardware Reference Design
- Performance
- Demo
- News
- Company
- Contact US

For more information
Email:yf.juan@itri.org.tw

Fast IOPS and High Capacity Flash Storage Appliance using Only Industry Standard Hardware

SOFA is a revolutionary flash storage management software technology that transforms standard x86 servers and solid state disks (SSD) into enterprise-class flash-based iSCSI storage servers with both large capacity and high IOPS.

ITRI CCMA SOFA storage server

http://ssd.cloud.itri.org.tw/ , http://www.youtube.com/watch?v=-sAvapcrzKI

演講大綱 Agenda

Life of Big Data
From Internet of Things to Mobile Computing

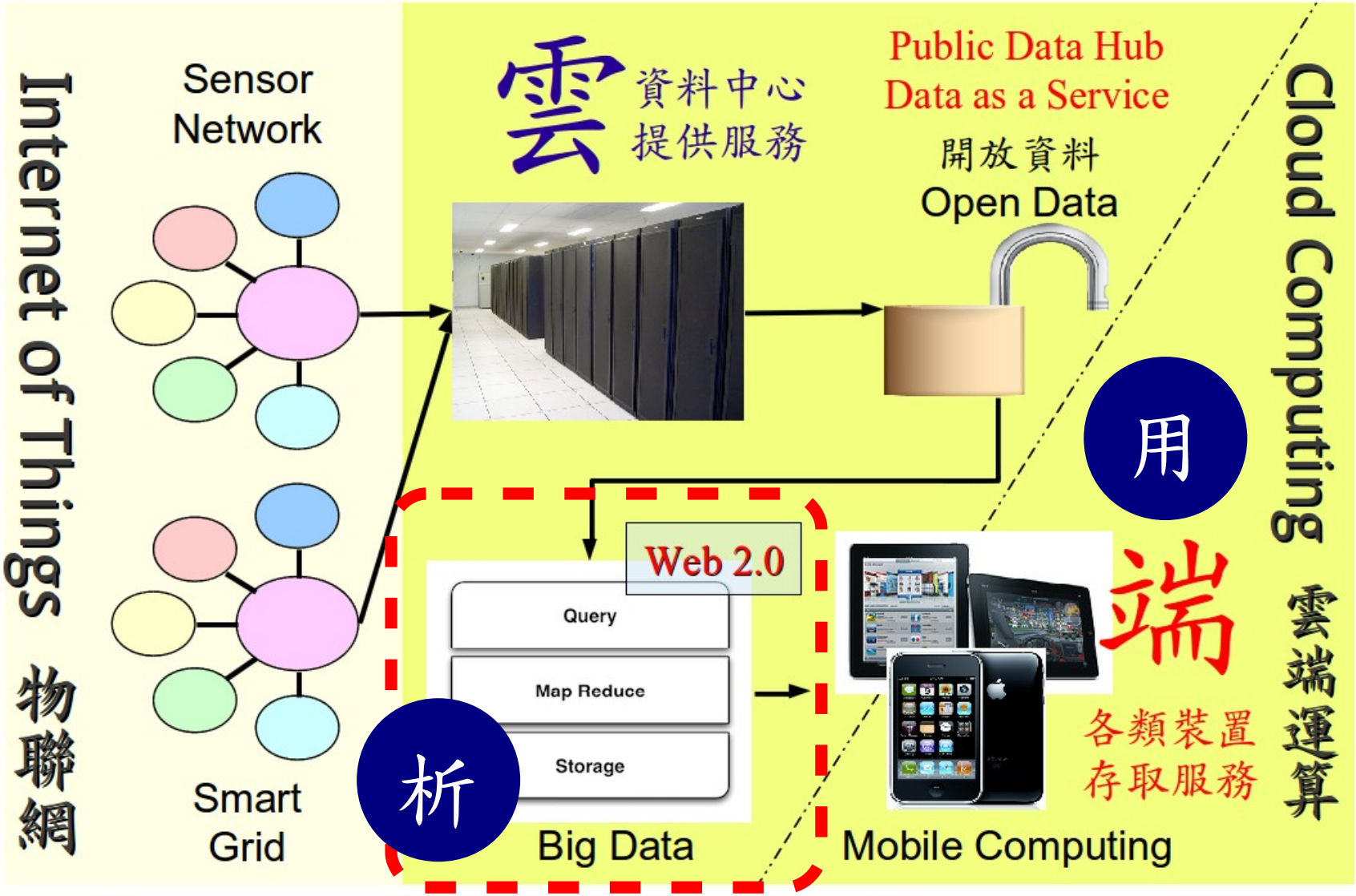
Paradigm Shift of Cloud Computing
from Computing Center to Data Center

Evolution of Big Data Software Stacks
for Next Generation Data Center

Roles of *** in Big Data Supply Chain***

Conclusion

巨量資料的五大生命週期
5 Stage of Big Data Life Cycle



Current Status of Big Data .....

Big data is like **teenage sex**:
everyone **talks about** it,
nobody really knows **how to do** it,
everyone **thinks** everyone else is doing it,
so everyone **claims** they are doing it ..

– Dan Ariely, Professor at Duke University
and Professor at Center for Advanced Hindsight

What is Big Data?!

海量資料泛指資料大小已無法用一般軟體擷取、管理與處理；
單一資料集大小介於數十 TB 至數 PB 的資料。
'Big Data' = few dozen TeraBytes to PetaBytes in single data set.

Definition [edit]

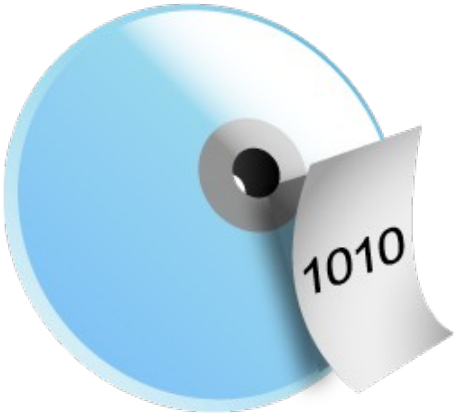
Big data is a term applied to data sets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time. Big data sizes are a constantly moving target currently ranging from a few dozen terabytes to many petabytes of data in a single data set.

In a 2001 research report[14] and related conference presentations, then META Group (now Gartner) analyst, Doug Laney, defined data growth challenges (and opportunities) as being three-dimensional, i.e. increasing volume (amount of data), velocity (speed of data in/out), and variety (range of data types, sources). Gartner continues to use this model for describing big data.[15]

出處：http://en.wikipedia.org/wiki/Big_data

多個檔案，容量 100TB

一個資料庫，容量 100TB

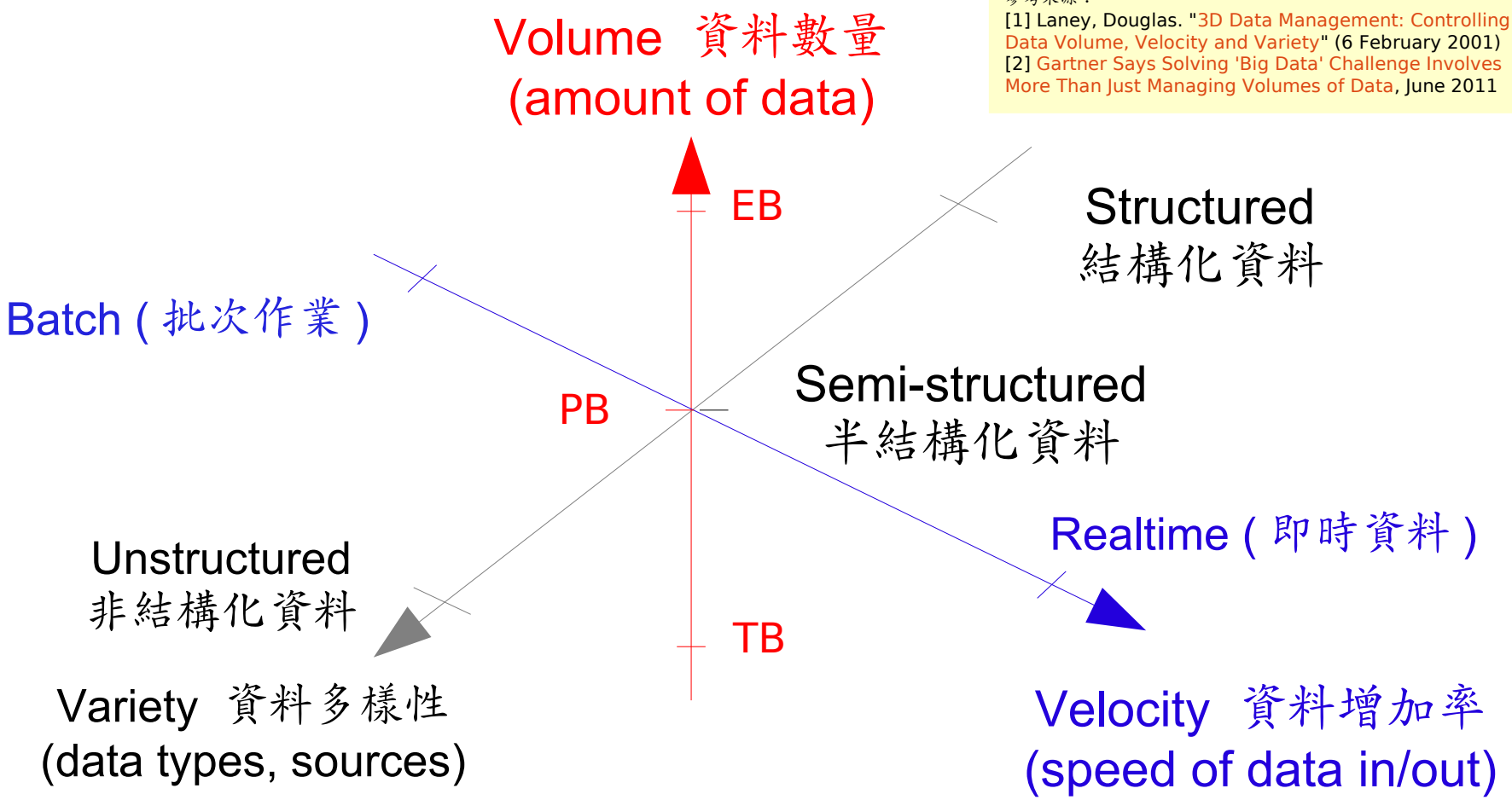


一個檔案，容量 100TB

巨量資料的三大挑戰
Challenges - 3 Vs of Big Data

NARLabs

參考來源：
[1] Laney, Douglas. "3D Data Management: Controlling Data Volume, Velocity and Variety" (6 February 2001)
[2] Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data, June 2011

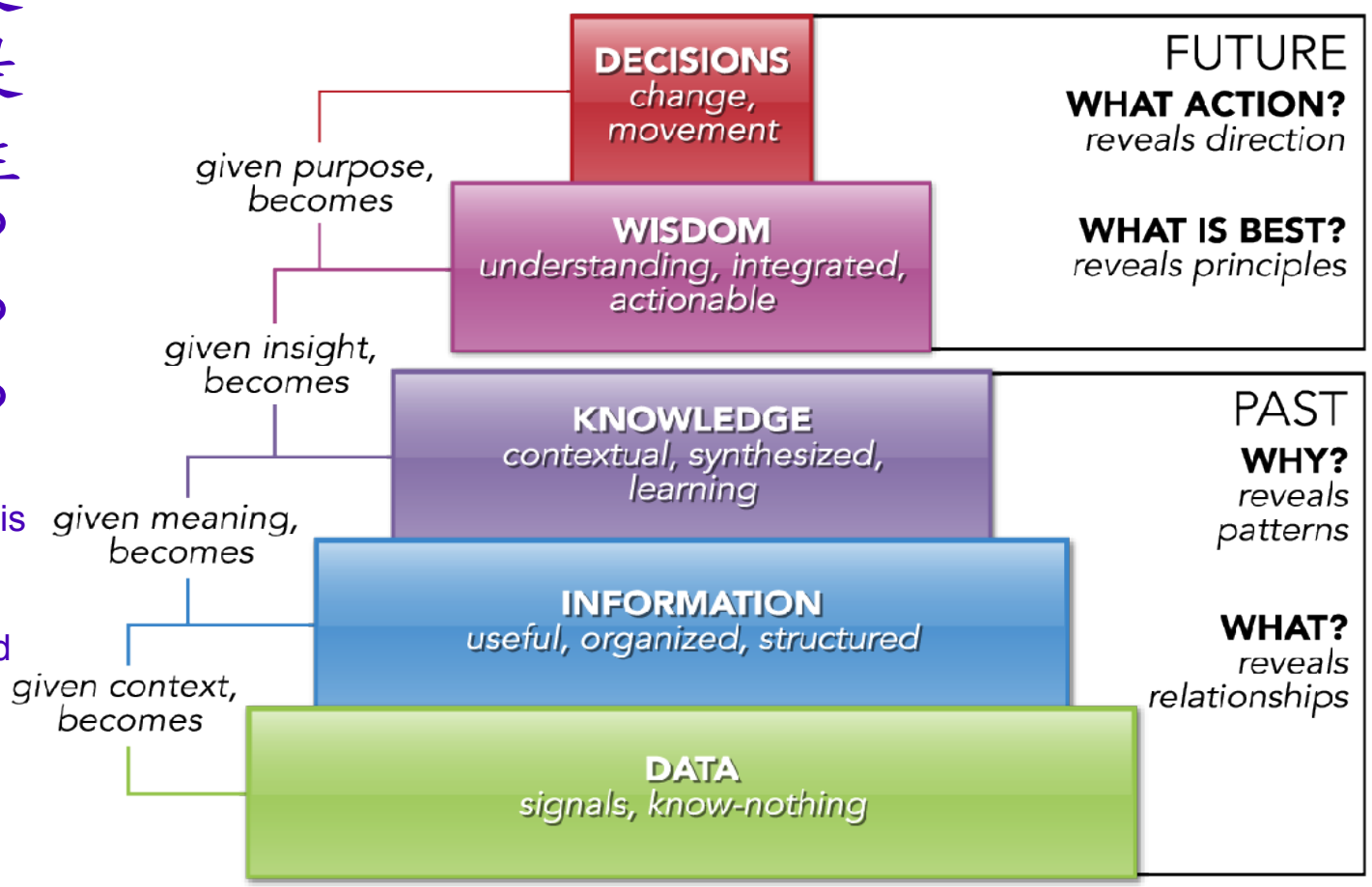


巨量資料的挑戰在於如何管理「數量」、「增加率」與「多樣性」

知識源自彙整過去，智慧在能預測未來
Knowledge is from the PAST, Wisdom is for the FUTURE.

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？

It does not matter how big is your data. The goal is to create VALUE within reasonable time period and total cost of ownership.



http://www.pursuantgroup.com/blog/tag/dikw-model/

大家都說「資料是金礦」，
那就讓我們拿採礦當類比吧！

NARLabs

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**
含金度	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**

今天我們能談的是該怎麼蓋「提煉廠」

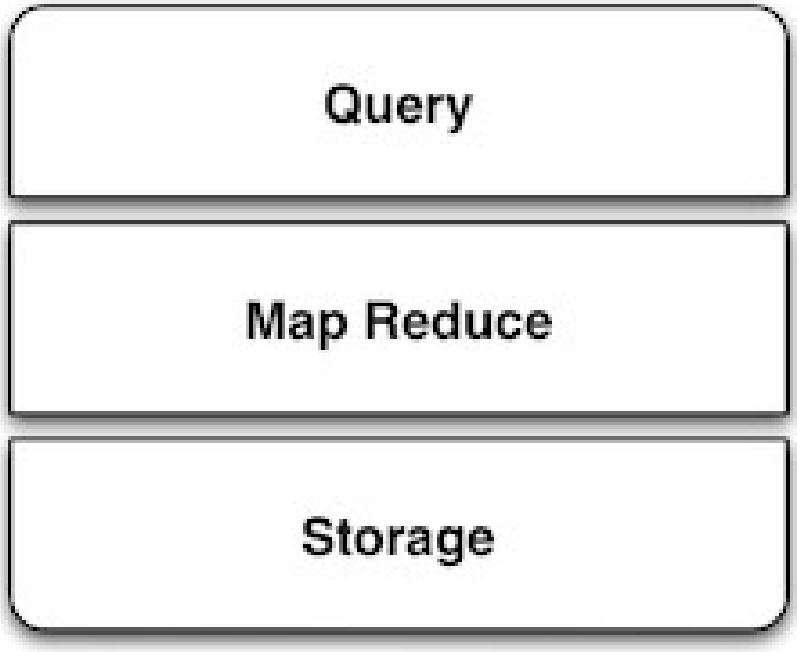
國際金價	提供給客戶的價值	產品通路
開採成本	總擁有成本	軟硬體投資
提煉廠	分析平台與工具軟體	**SMAQ**
含金度	資料鑑價？	商業模式
開採權	分析資料的合法性	個資法
金礦	資料集	**Open Data**

巨量資料處理的資訊架構
The SMAQ stack for big data

NARLabs

LAMP is for Web Services.

未來處理海量資料的人必需知道
You will need a big data stack called 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

PAST: Big Data at Rest

Can gigabytes predict the next Lady Gaga?

By Stacey Higginbotham

Want to know how playing on Jimmy Kimmel Live will boost the sales of an artist's album? Or how about figuring out where fans go to find artists after they hit the evening news? What about the effect Whitney Houston's death had on her YouTube and Vevo plays? They shot up 4,525 percent, by the way.

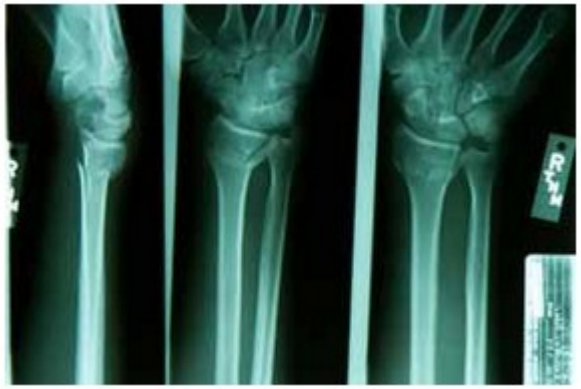
http://nextbigsound.com/

How big data can curb the world's energy consumption

http://www.openpdc.com/

fakebook

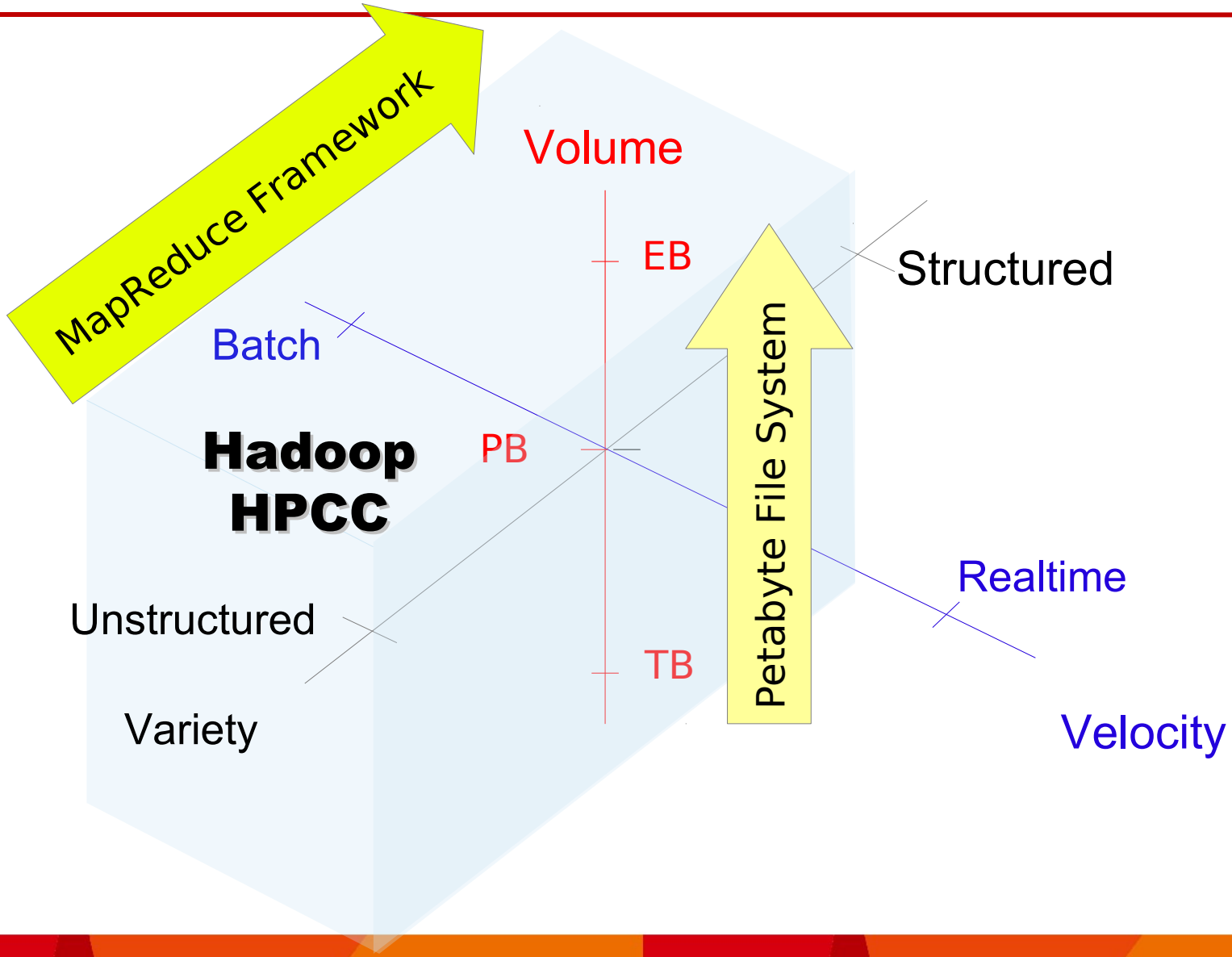
NETFLIX



One hospital's embrace of big data

Source: 10 ways big data changes everything,
http://gigaom.com/2012/03/11/10-ways-big-data-is-changing-everything

處理巨量資料的三類技術 (1)
Data at Rest – MapReduce Framework



高資料通量處理平台 Hadoop
Key Concept : Data Locality

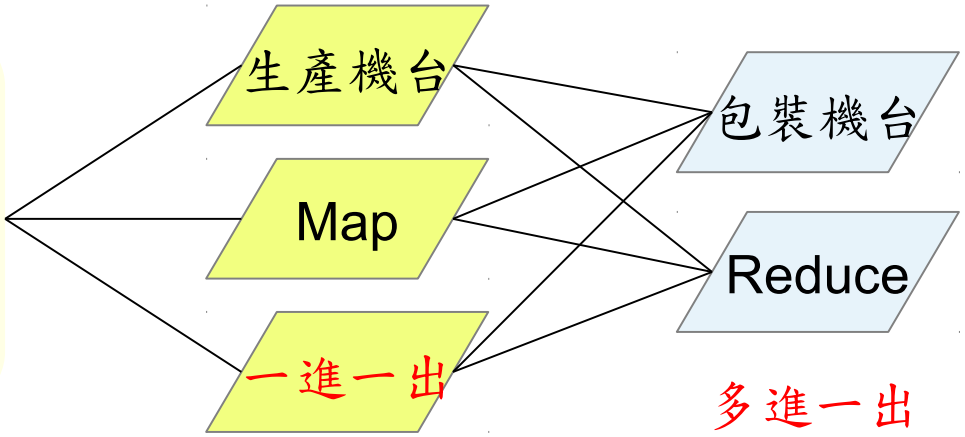
NARLabs

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。

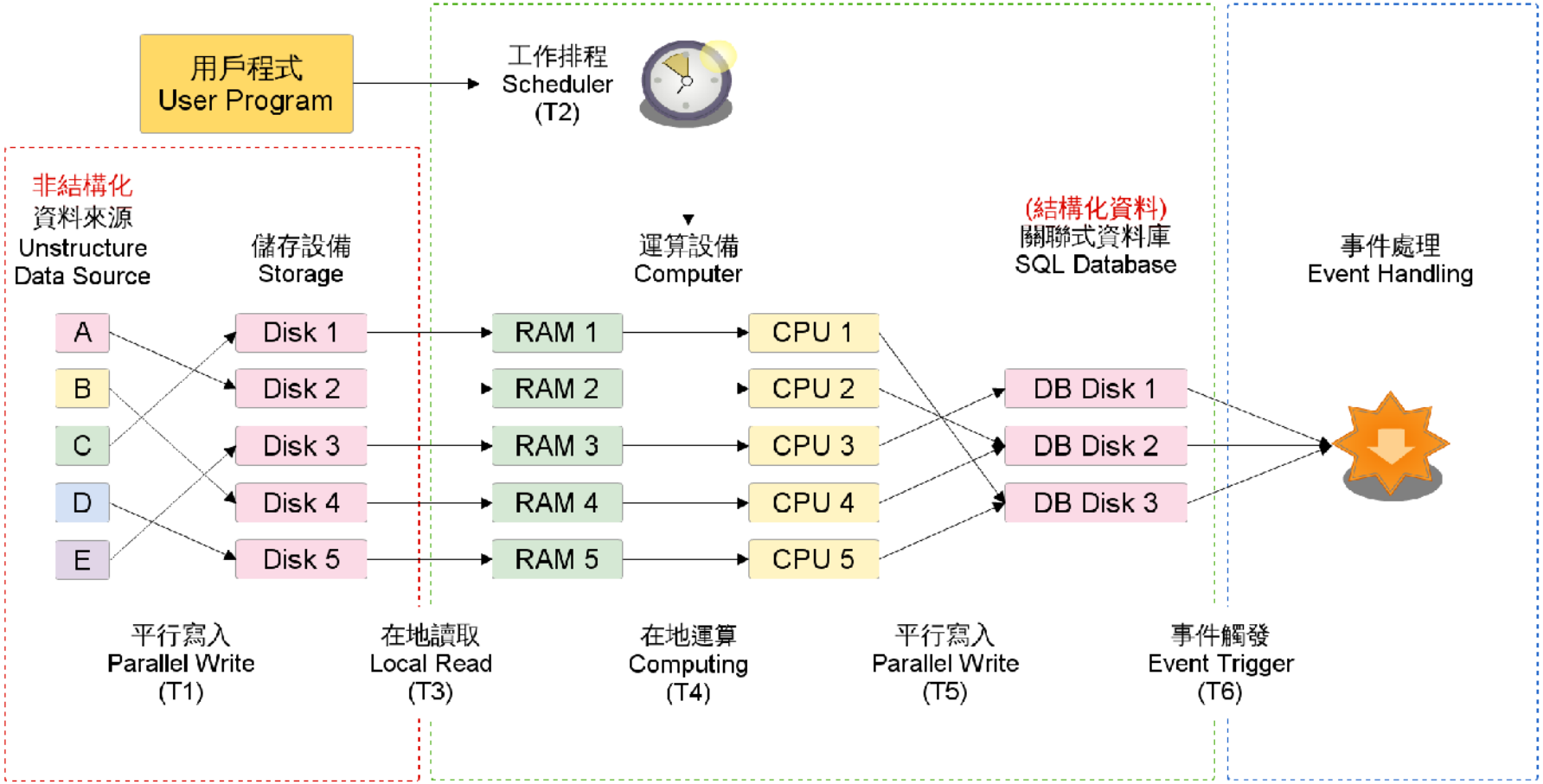
亦可以想像成一個處理海量資料的生產線，只須學會定義 **map** 跟 **reduce** 工作站該做哪些事情。

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



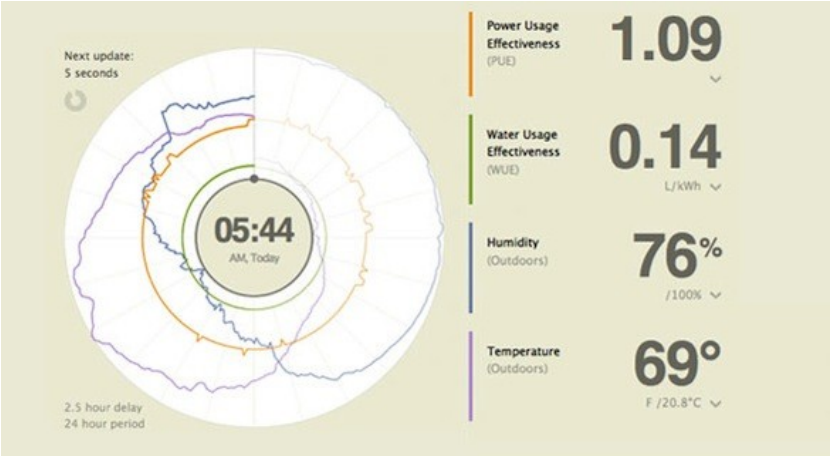
資料蒐集階段
Phase 1 : Data Collection

資料處理階段
Phase 2 : Data Processing

事件處理階段
Phase 3 : Event Handling

NOW: Big Data in Motion

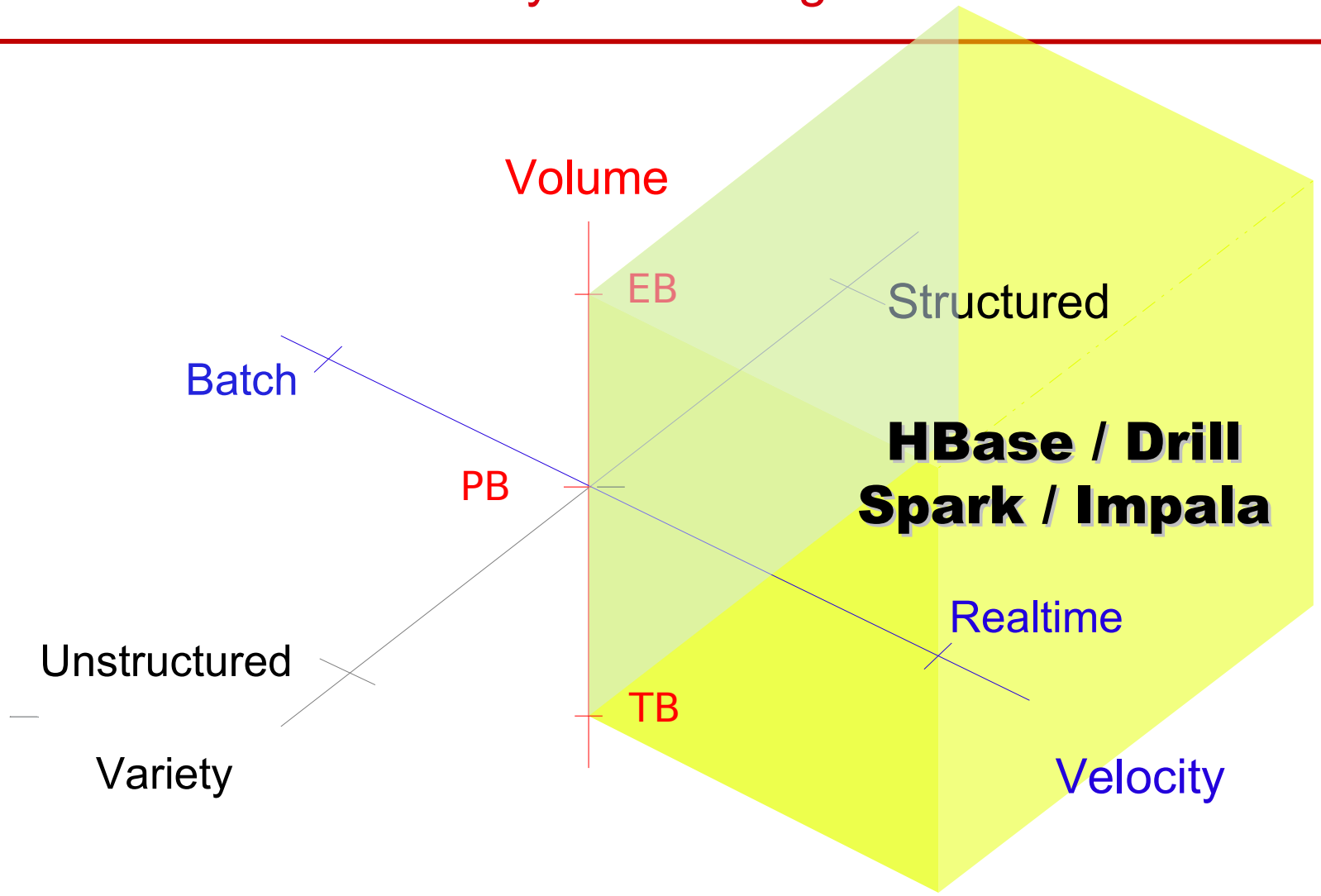
[金融] **Trading Robot**



[災防] 海嘯、土石流
Disaster Prevention
Tsunami Forecast

[資訊] 機房即時用電資訊監控、警訊
Realtime Data Center Power Usage and related notifications
http://www.newmobilelife.com/2013/04/21/facebook-pue-real-time-charts/

處理巨量資料的三類技術 (2)
Data in Motion – In-Memory Processing

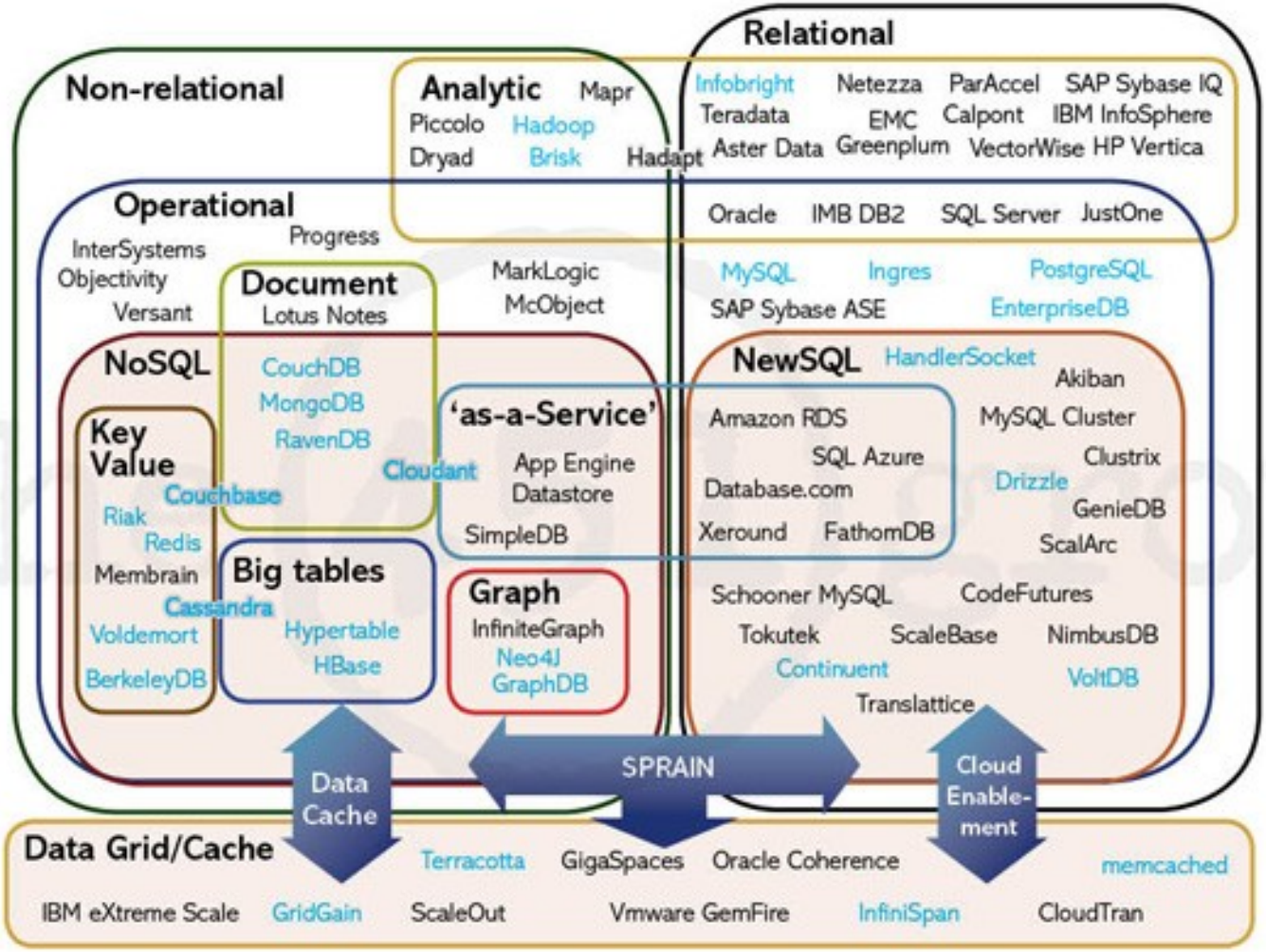


Google 的技術演進 vs Apache 專案

NARLabs

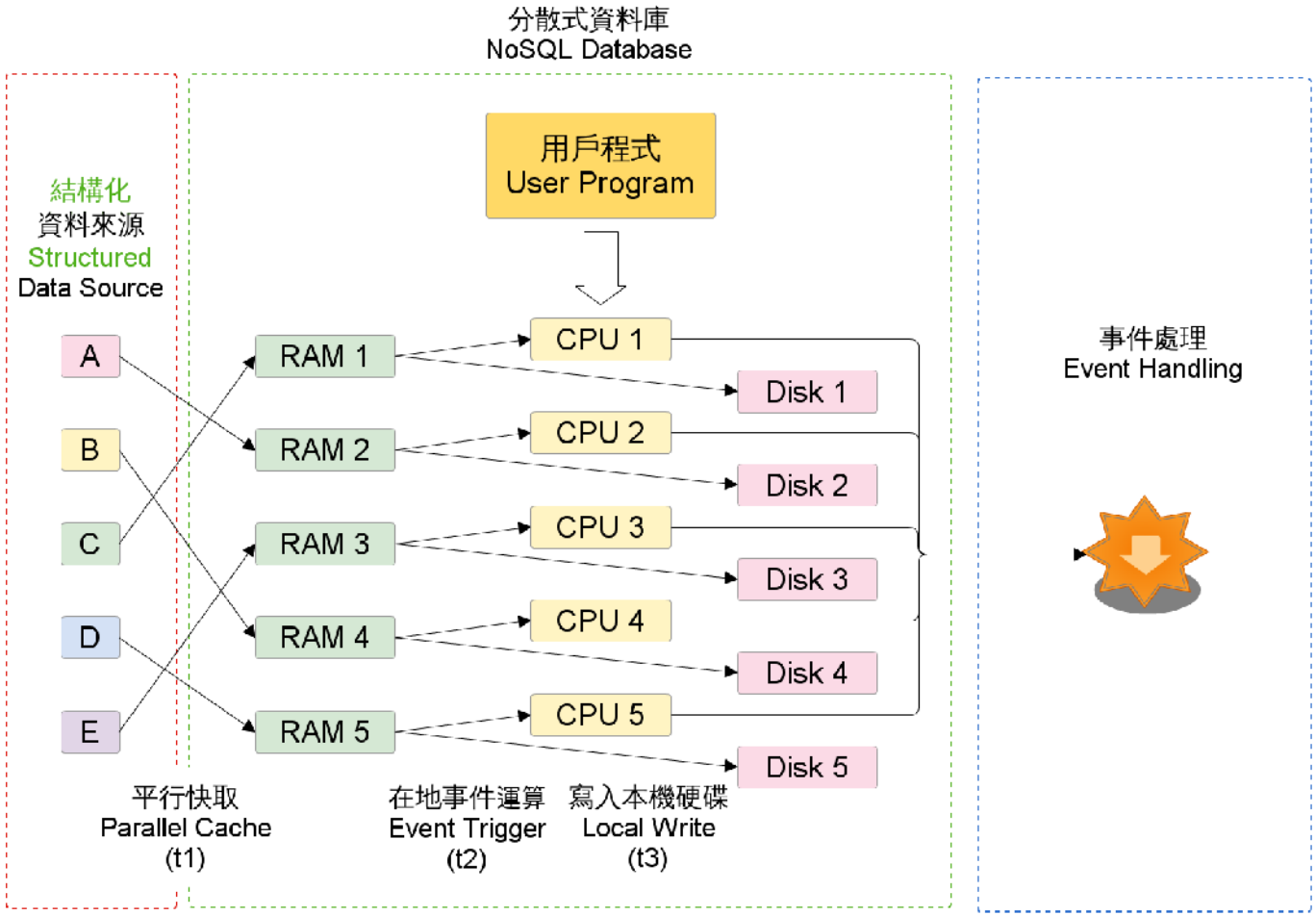
Big Query (JSON, SQL-like)	Dremel (2010)	Apache Drill (2012)
Incremental Index Update (Caffeine)	Percolator (2010)	
Graph Database	Pregel (2009)	Apache Giraph (2011)
Query	BigTable (2006)	Apache HBase (2007)
Map Reduce	MapReduce (2004)	Hadoop MapReduce (2006)
Storage	Google File System (2003)	HDFS (2006)

令人眼花撩亂的多樣化資料庫選擇
NoSQL vs NewSQL



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

In-Memory Processing 的運算時間
以 HBase 為例

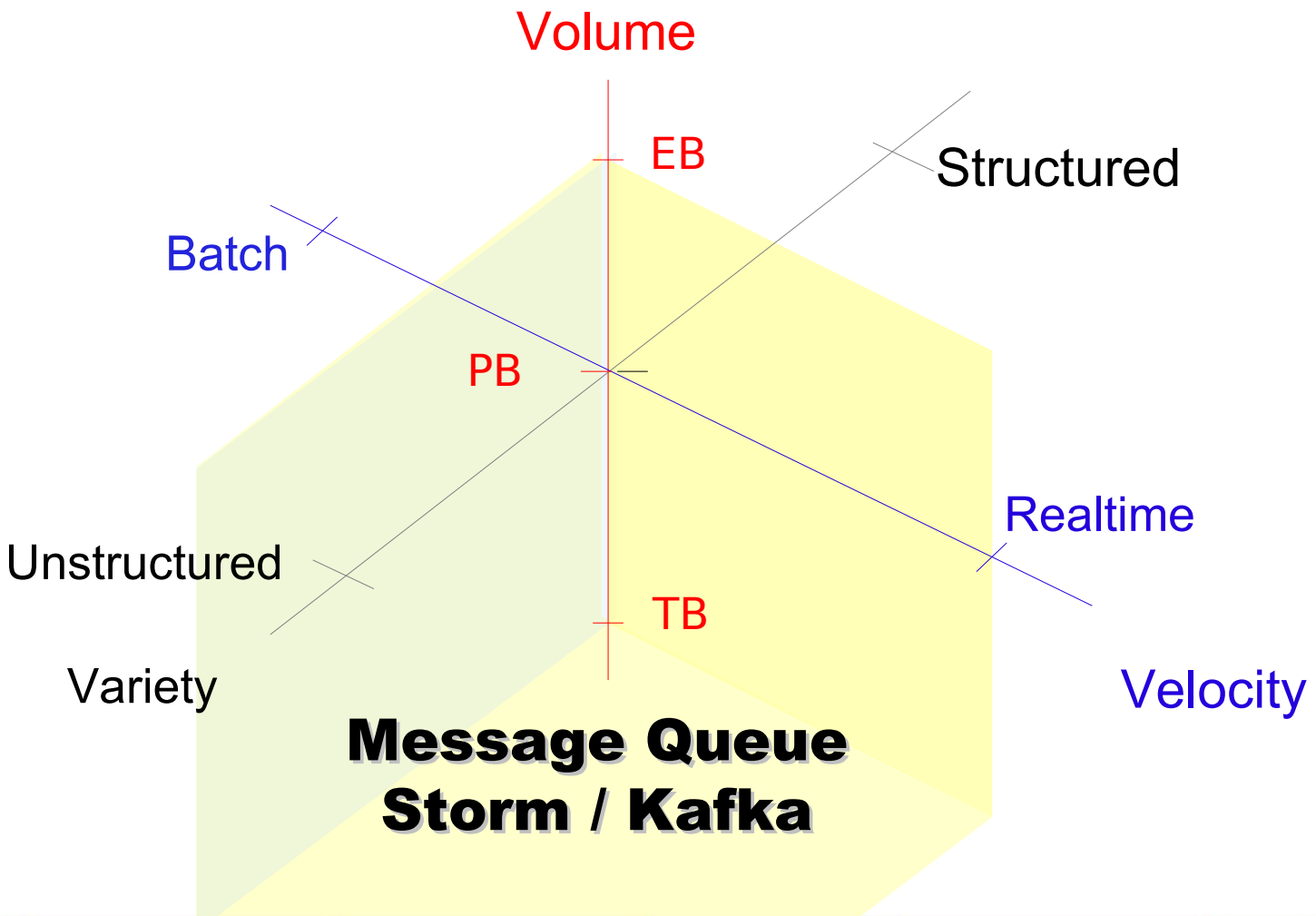


資料蒐集階段
Phase 1 : Data Collection

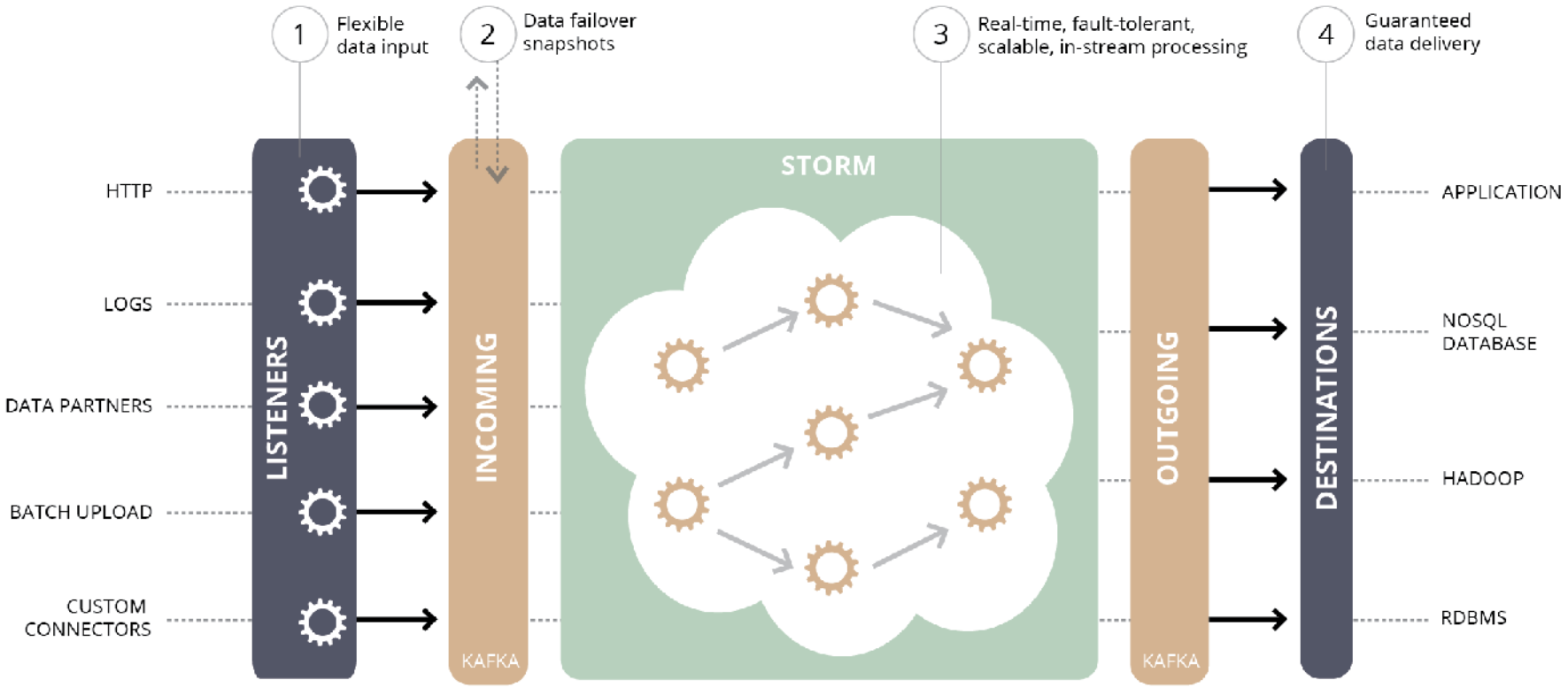
資料處理階段
Phase 2 : Data Processing

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 , Data Cleaning



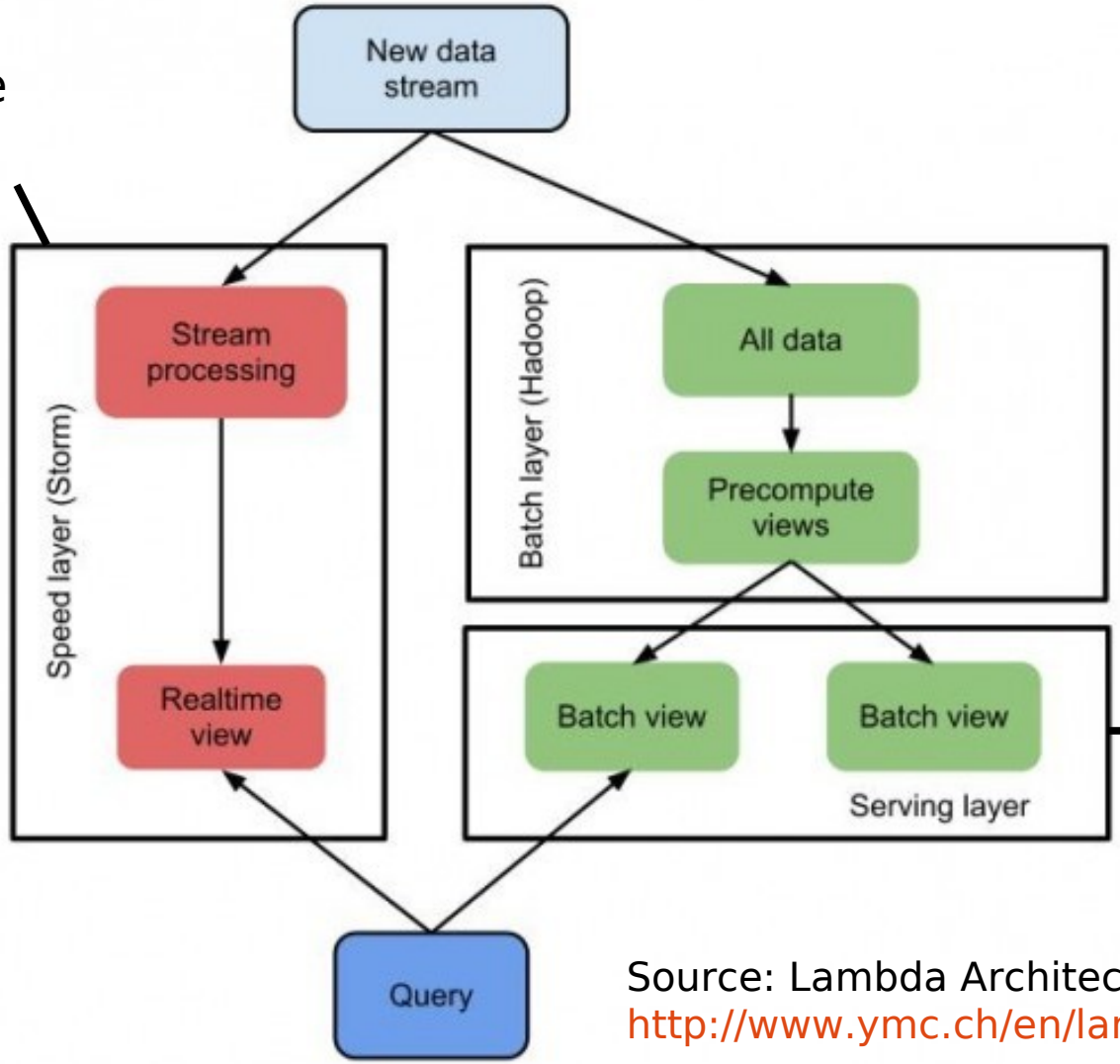
Twitter Storm + Apache Kafka



http://blog.infochimps.com/2012/10/30/next-gen-real-time-streaming-storm-kafka-integration/

混合模式的巨量資料處理架構
Lambda Architecture for Big Data after 2013

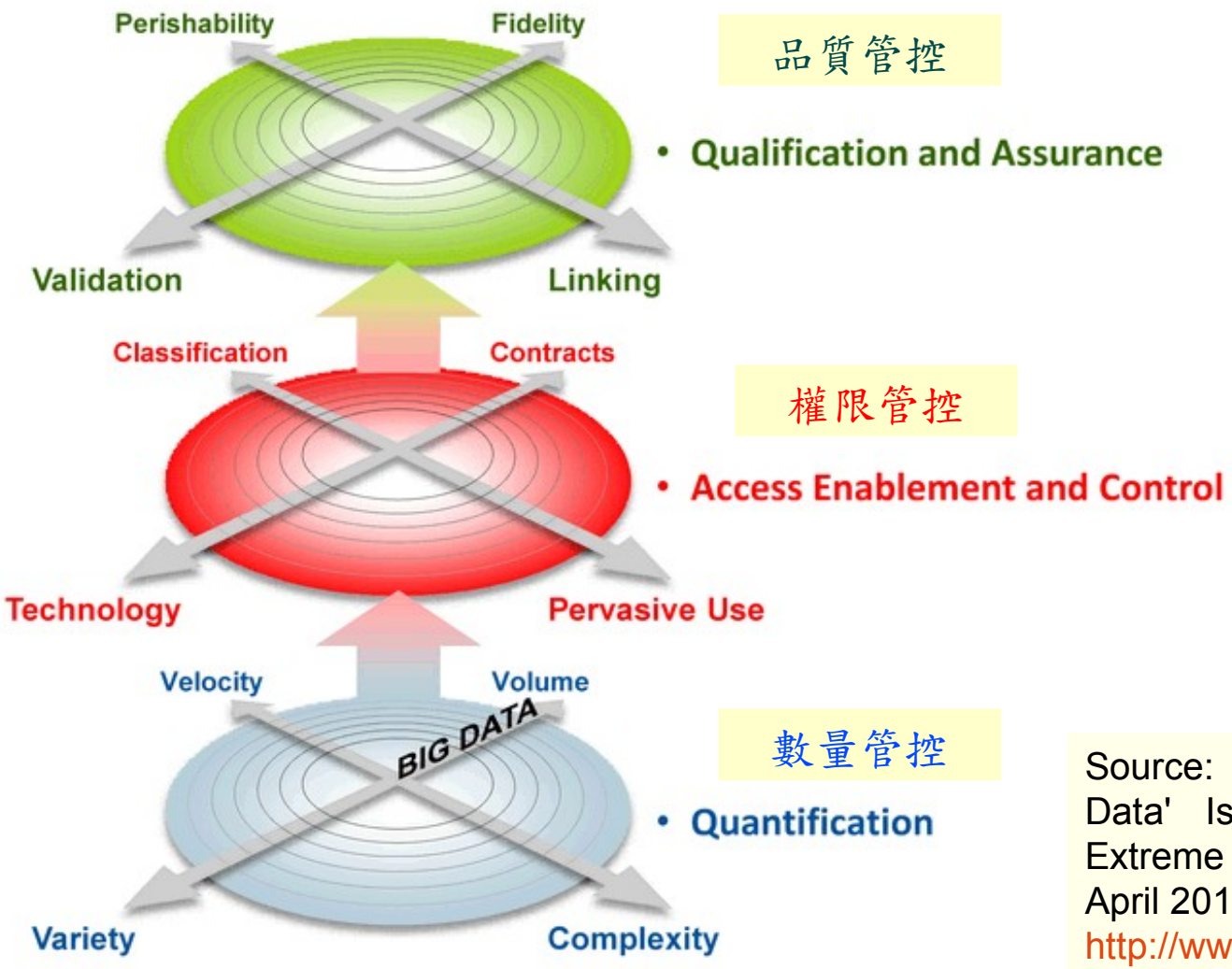
HBase
Storm



ElephantDB
Or
Voldemort

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

Future: Big Data Security



當我們緊密相連 .....

世界政經：歐盟想分 **Tweeter**
找出經濟、政治的脈動

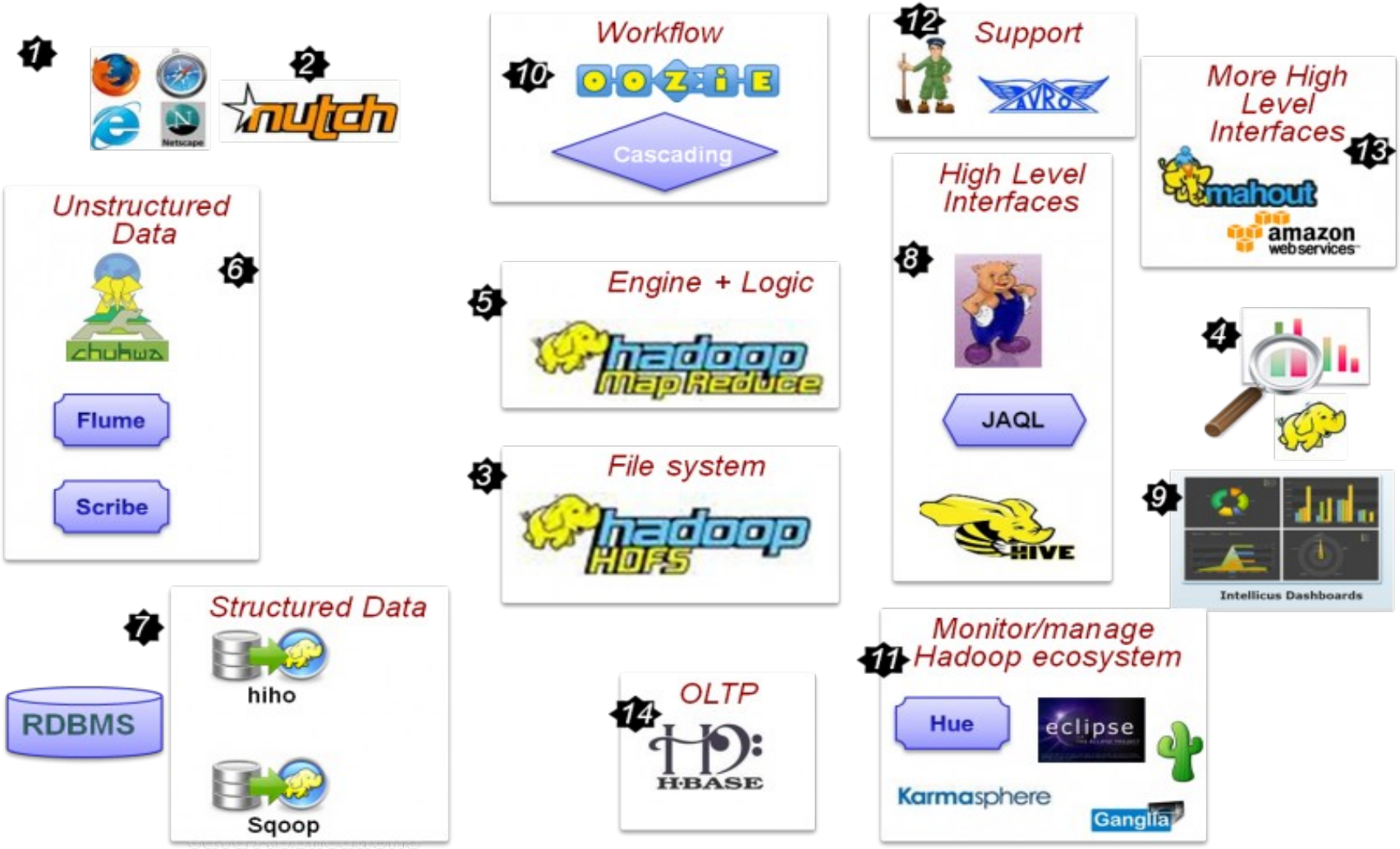
國家安全：美國 **PRISM** 計劃
(網軍 **!** 終極警探 **4.0)**

組織如何因應 **APT ?**
Big Data 平台本身的安全性 **?**

有太多安全的問題等待解決！

Source: Gartner (March 2011), 'Big Data' Is Only the Beginning of Extreme Information Management, April 2011, http://www.gartner.com/id=1622715

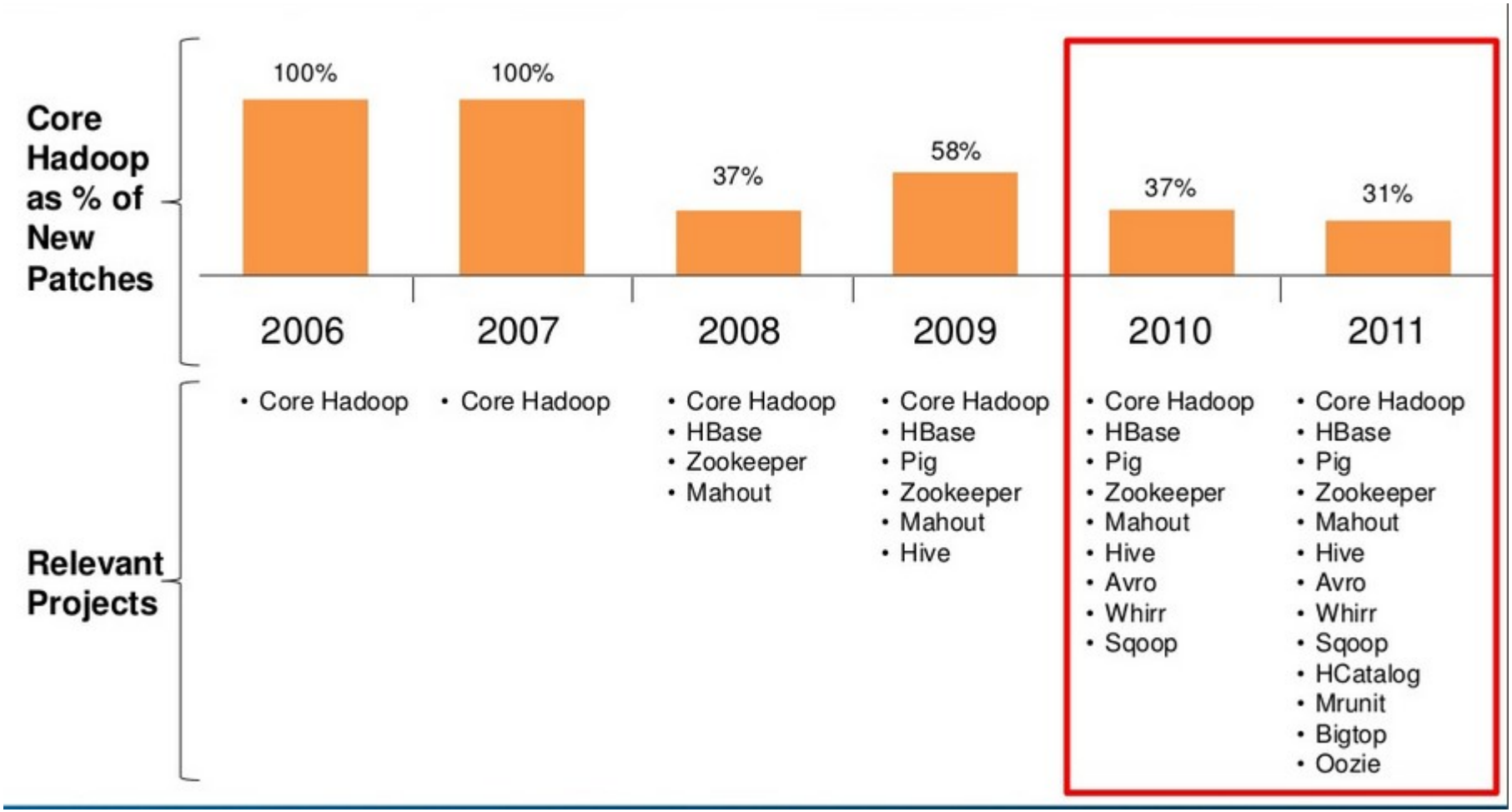
Future:
Better Hadoop Ecosystem



http://rationalintelligence.com/wp_log/?p=104

Evolution of Apache Hadoop Ecosystem

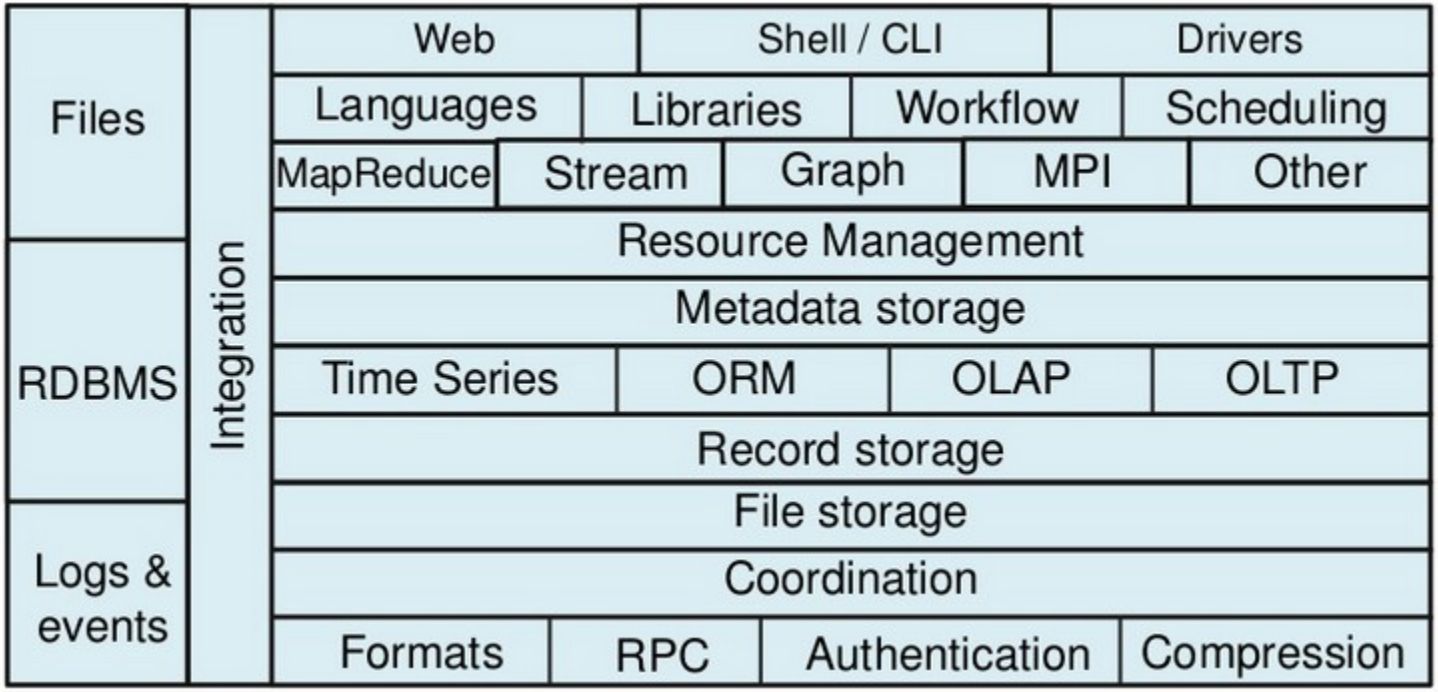
NARLabs



Hadoop World 2011: The Hadoop Stack - Then, Now and in the Future
http://www.slideshare.net/slideshow/embed_code/10110006

Complexity of Apache Big Data Stack

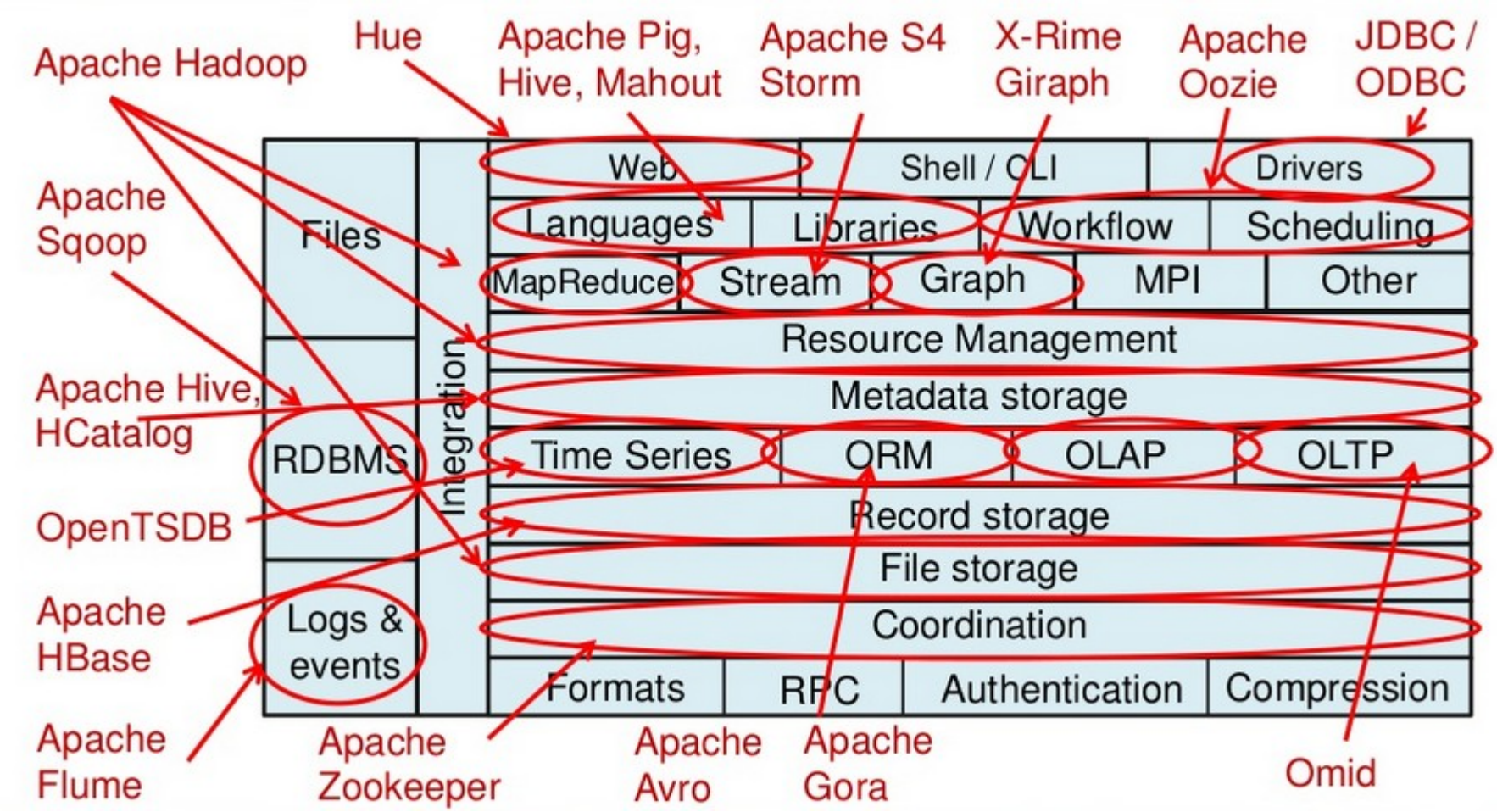
NARLabs



Hadoop World 2011: The Hadoop Stack - Then, Now and in the Future
http://www.slideshare.net/slideshow/embed_code/10110006

Complexity of Apache Big Data Stack

NARLabs

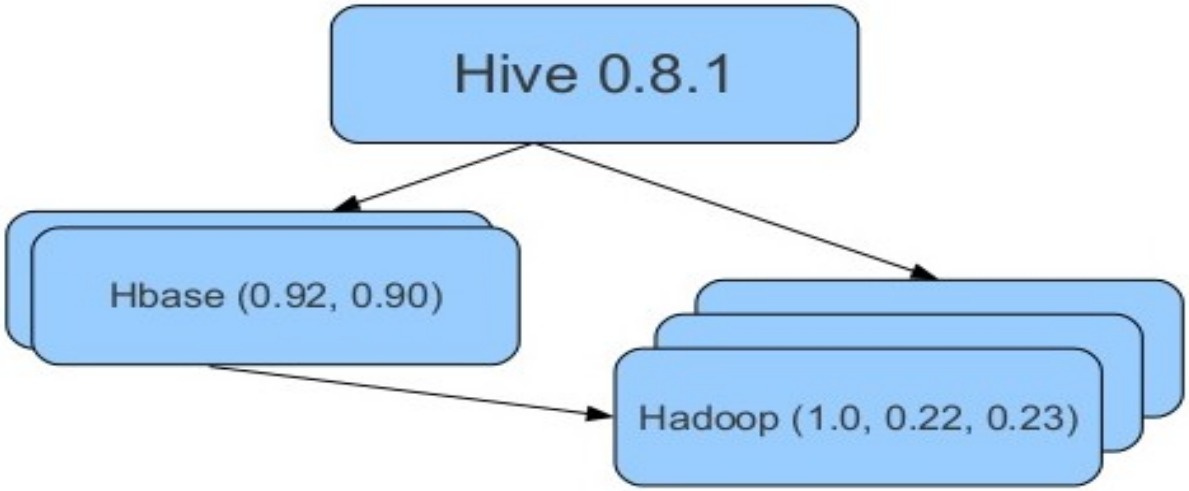


Hadoop World 2011: The Hadoop Stack - Then, Now and in the Future
http://www.slideshare.net/slideshow/embed_code/10110006

Big Problem: Package dependency

NARLabs

Dependencies Inferno:

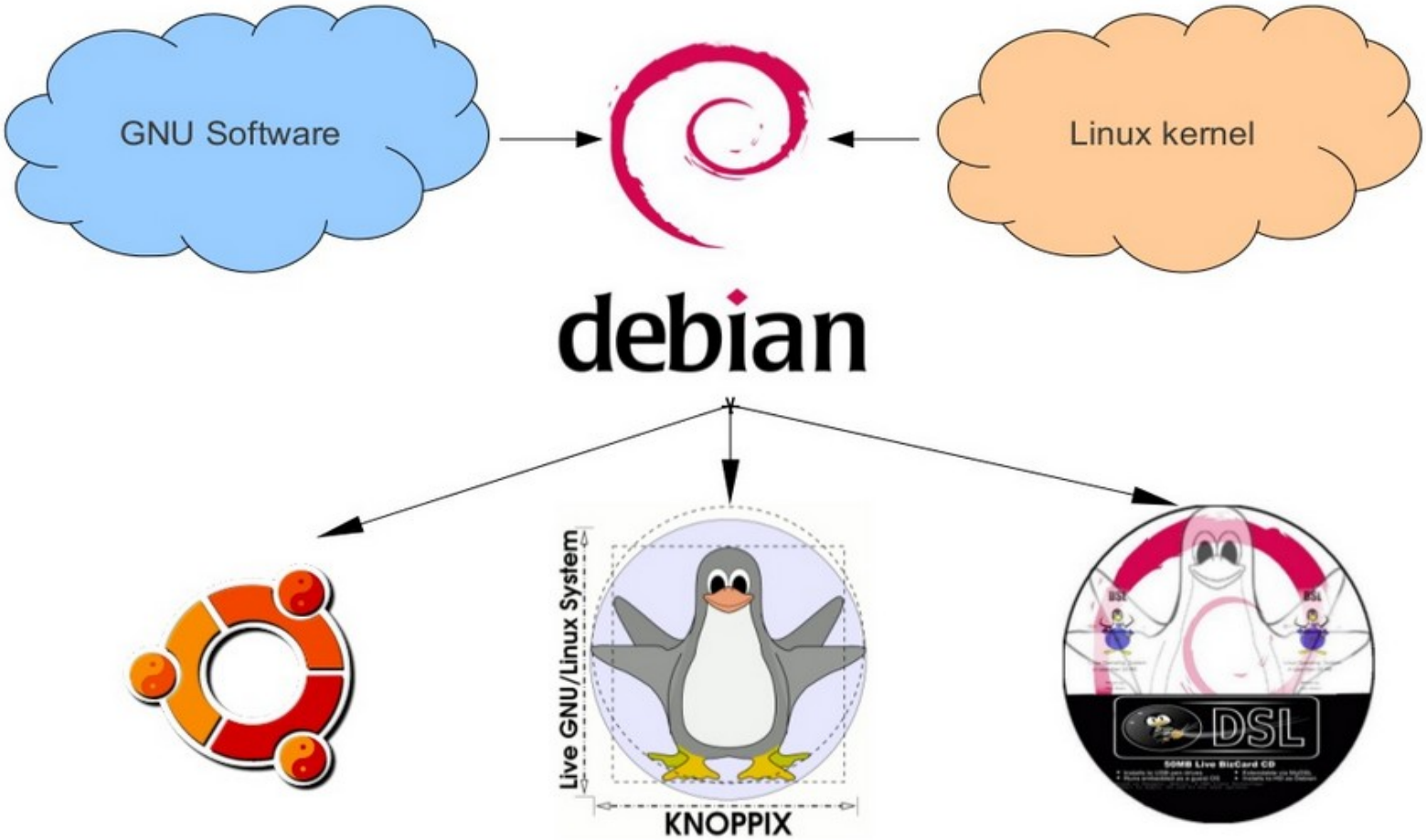


A million dollar question:

```
$ tar xzvf hive-0.8.1.tar.gz
$ ls hive-0.8.1/lib
hbase-0.89.jar log4j-1.2.15.jar log4j-1.2.16.jar
```

Deploying Hadoop-based Bigdata Environments,
http://www.slideshare.net/slideshow/embed_code/16370152

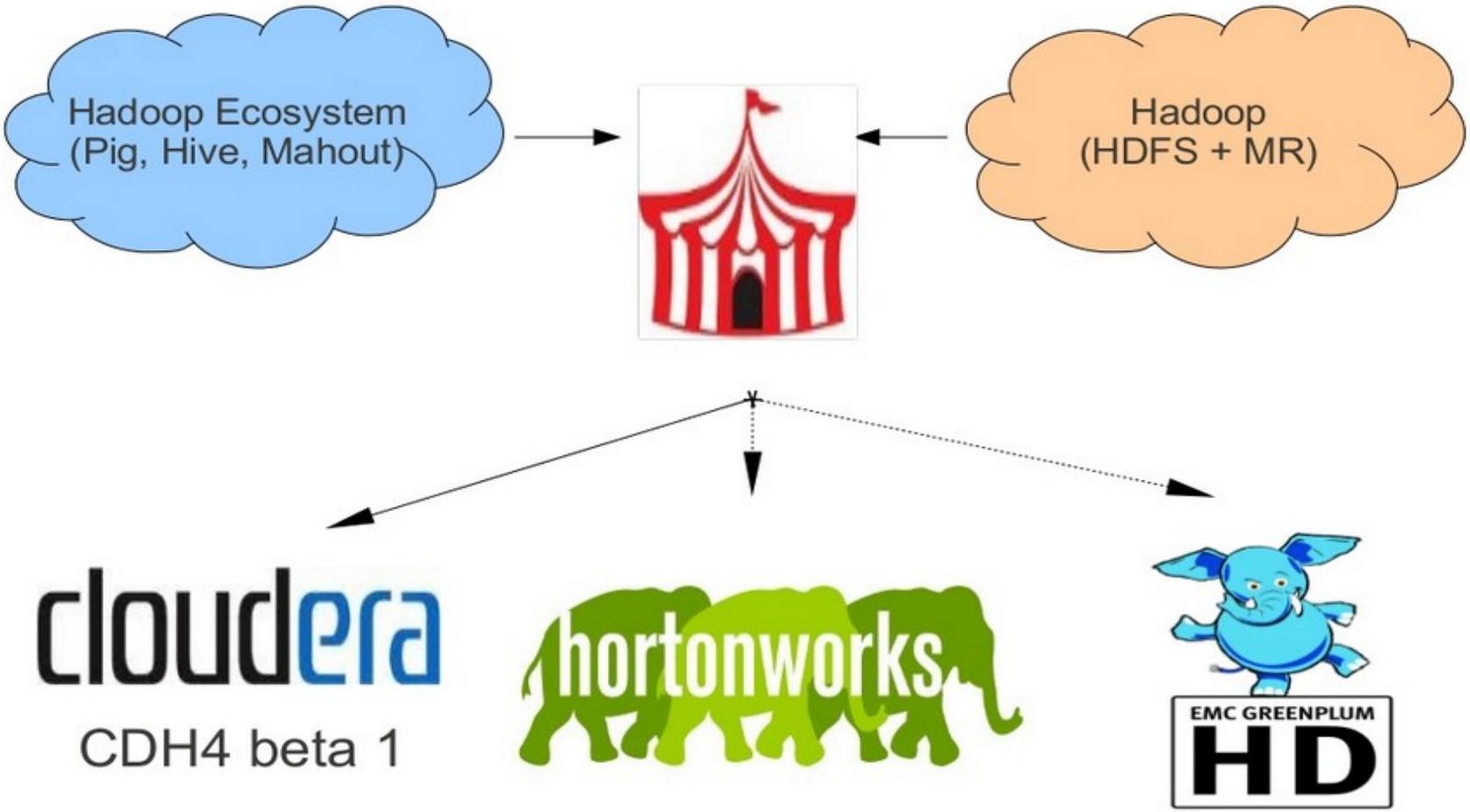
What Debian did to Linux?



Deploying Hadoop-based Bigdata Environments,
http://www.slideshare.net/slideshow/embed_code/16370152

Bigtop is trying to do it with Hadoop

NARLabs



Deploying Hadoop-based Bigdata Environments,
http://www.slideshare.net/slideshow/embed_code/16370152

演講大綱 Agenda

Life of Big Data
From Internet of Things to Mobile Computing

Paradigm Shift of Cloud Computing
from Computing Center to Data Center

Evolution of Big Data Software Stacks
for Next Generation Data Center

Roles of *** in Big Data Supply Chain***

Conclusion

Roles of ******
in Cloud Supply Chain

NARLabs

應用軟體供應商

行動裝置共通平台

htc quietly brilliant

端

各類裝置存取服務

軟體服務供應商

Google

amazon webservices™

資料中心機房維運

acer

雲

資料中心提供服務

基本硬體建設組件

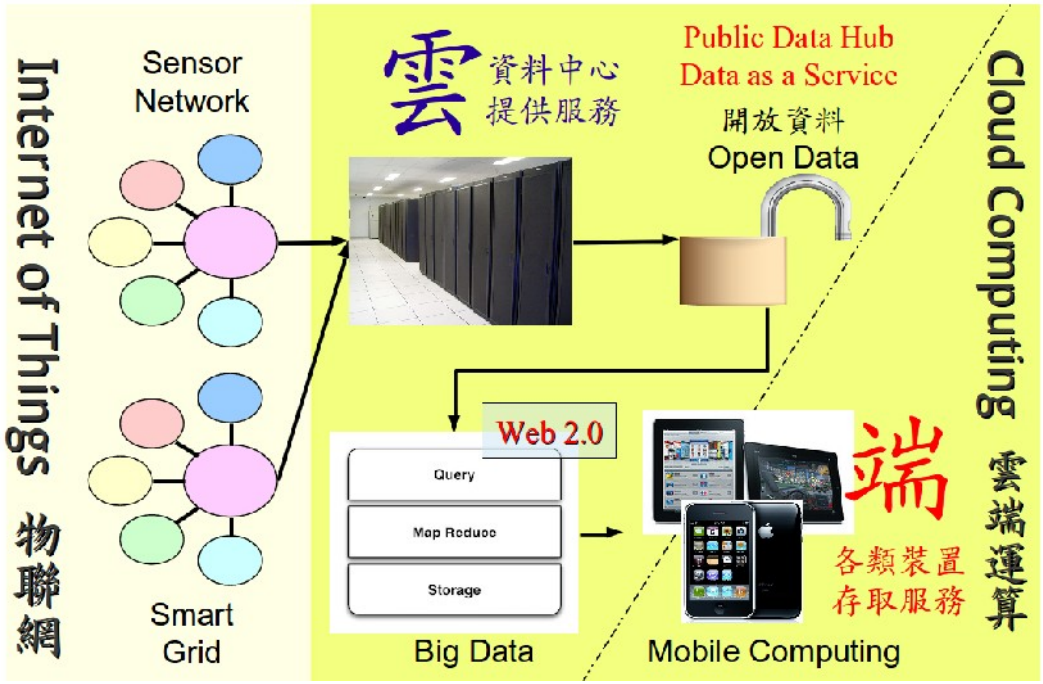
英業達集團 Inventec

廣達電腦 Quanta Computer

Supply Chain of Big Data Industry

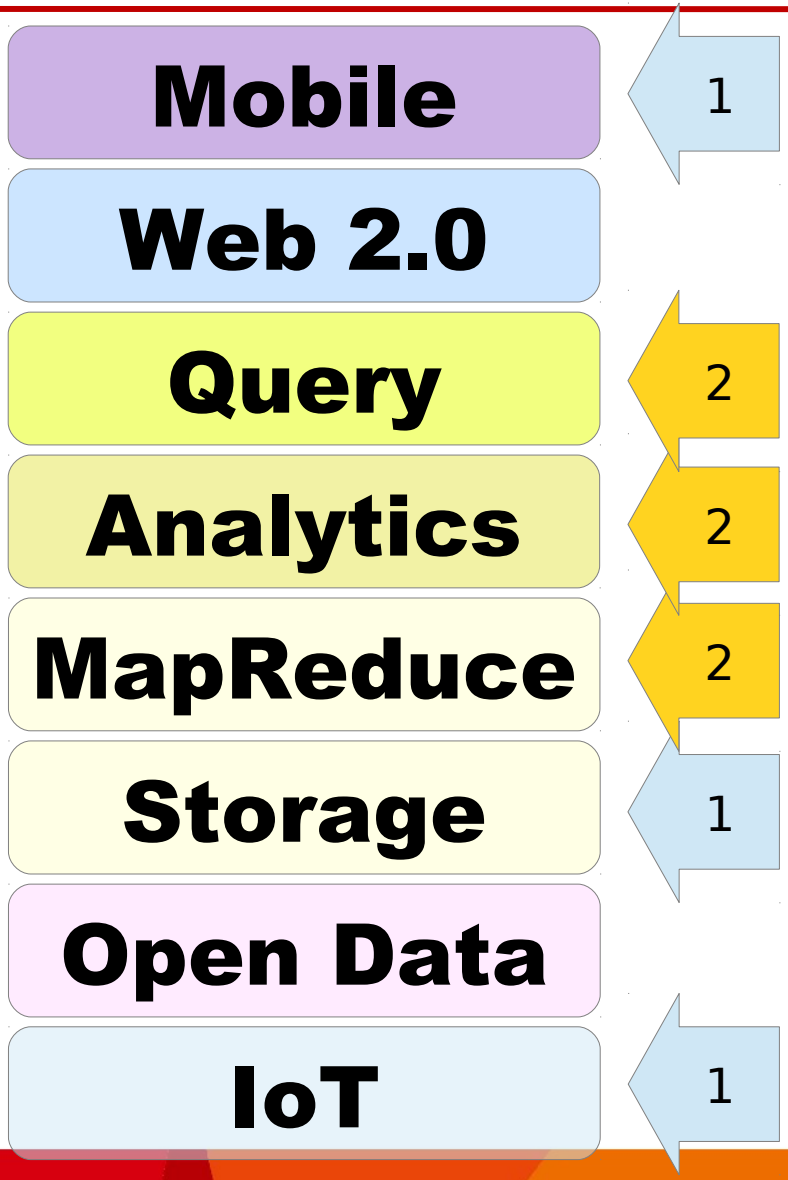
巨量資料產業的上下游供應鏈

- Mobile
- Web 2.0
- Query
- Analytics
- MapReduce
- Storage
- Open Data
- IoT



Roadmap of ******
for Big Data Industry

NARLabs

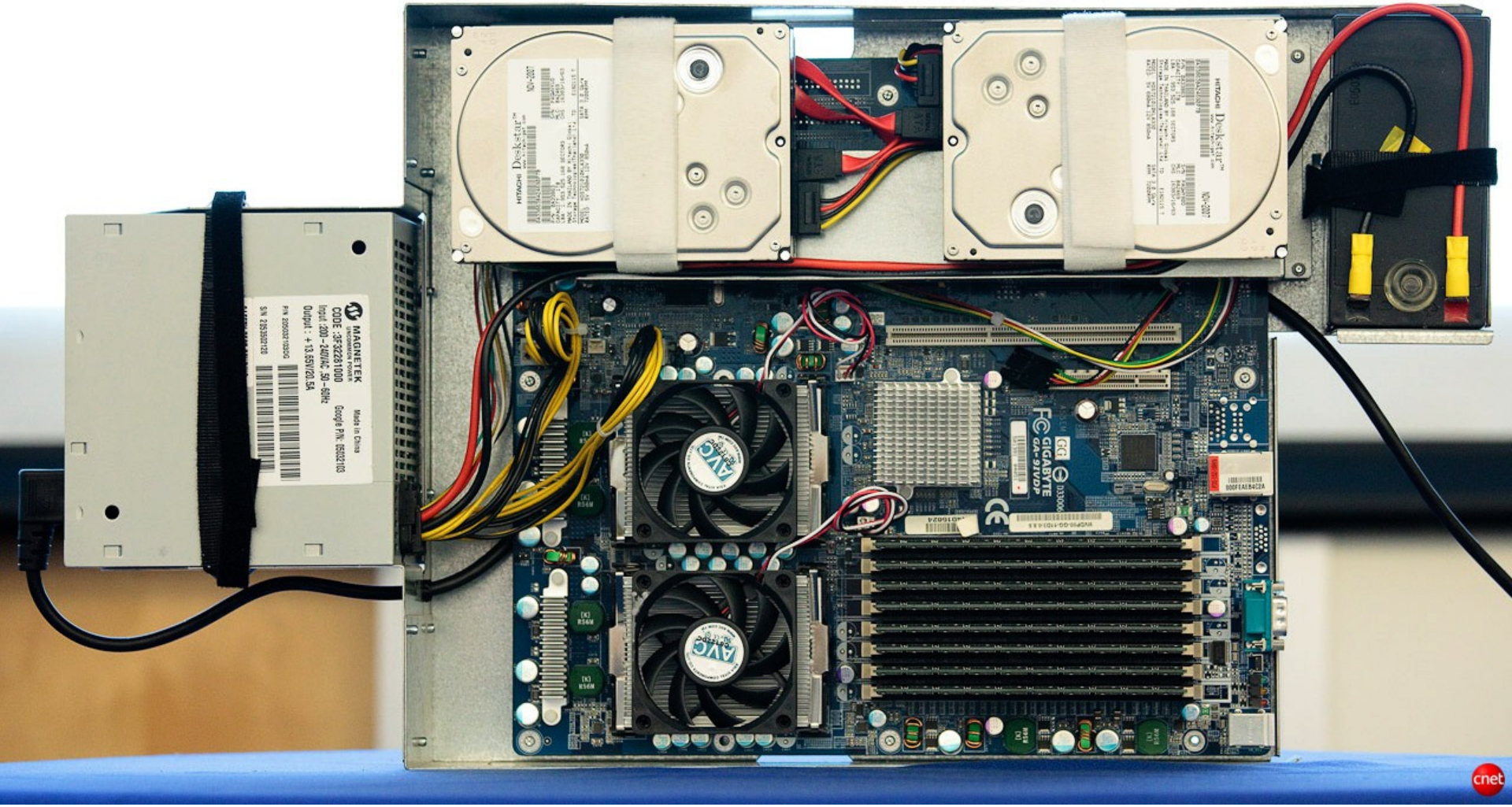


In-Memory Data Grid
(Ex. GridGain)

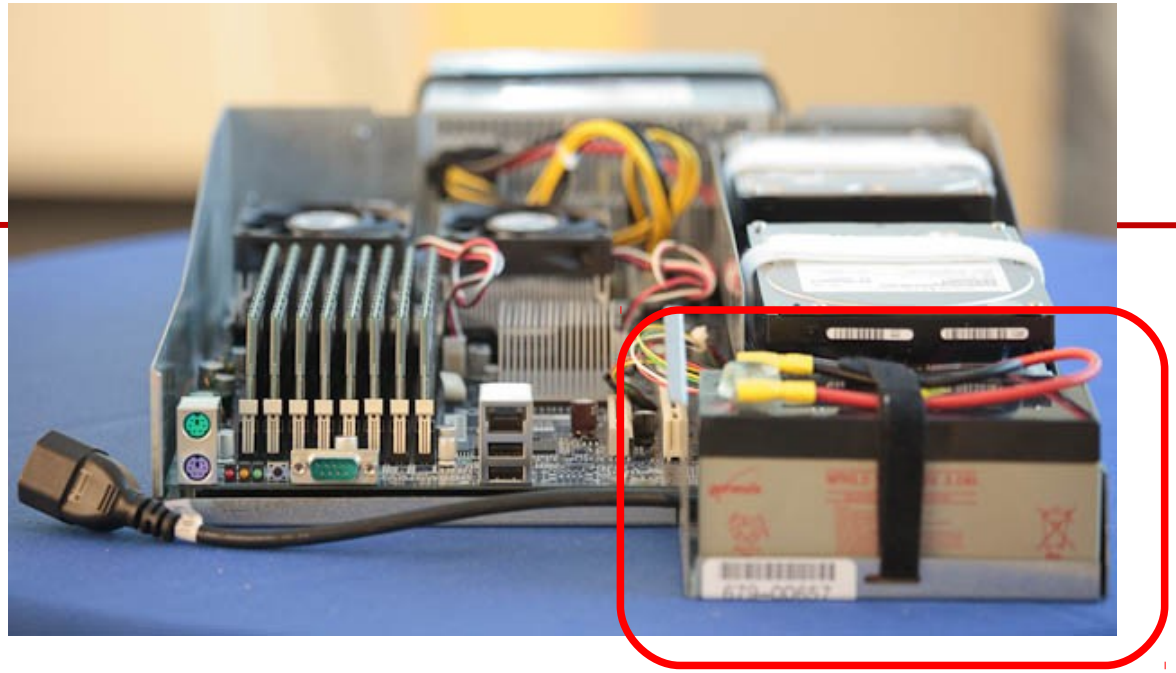
Realtime Predictive Analytics
(Ex. Revolution Analytics, R)

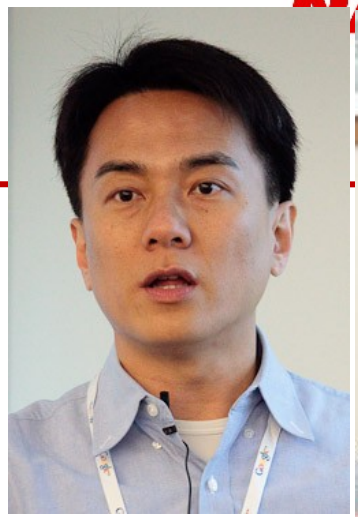
Parallel Data Warehouse
(Ex. Microsoft)

2009/04/01 谷歌公開神秘的伺服器設計（愚人節？）



Google Servers（2009-04-02 上傳）
http://www.youtube.com/watch?v=J139Aelaf0g





Google 伺服器設計師
Ben Jai（翟本喬）

1. 主機板透過直流供電（12V），電池當 UPS（電容防突波）
   → 所以後來演變成「貨櫃電腦」也是直流供電！
   → 直流供電的轉換效率較佳（少一級 AC/DC），降低廢熱
   → 谷歌甚至注意以 12 伏特經由銅線導電會比 5 伏特來有效率
2. 兩顆硬碟（為了 Big Data 運算特性）
3. 記憶體插滿（這依舊是為了 Big Data 運算特性）
4. 主機板拿掉不必要的介面晶片（光碟機、音效卡、顯示卡）
   → 省電！只留必要的介面（RS232, 鍵盤滑鼠，USB，網路）

Source: http://tfajang.pixnet.net/blog/post/24425368-google神秘伺服器大公開

Open Compute Project of Facebook

OPEN Compute Project

SEARCH

Log in

Specs & Designs View on Github Open Up

Hacking Conventional Computing Infrastructure

We started a project at Facebook a little over a year ago with a pretty big goal: to build one of the most efficient computing infrastructures at the lowest possible cost. We decided to honor our hacker roots and challenge convention by custom designing and building our software, servers and data centers from the ground up – and then share these technologies as they evolve.

The result is a data center full of vanity free servers which is 38% more efficient and 24% less expensive to build and run than other state-of-the-art data centers.

By releasing Open Compute Project technologies as open hardware, our goal is to develop servers and data centers following the model traditionally associated with open source software projects. That's where you come in.

SERVER designing vanity free, low cost compute nodes

DATA CENTER designing energy efficient data centers

Featured Projects

Intel Motherboard

The v1.0 Intel motherboard is a dual
Xeon® 5600 socket motherboard wit
we have a new v2.0 specification...

AMD Motherboard

The v1.0 AMD motherboard is a dual
socket motherboard with 24 DIMM sl
doubles the compute density, suppor

Power Supply

The Open Compute Project 450W pow
converter, single voltage 12.5VDC, cl
supply used in high-efficiency applica
includes...

Chassis

The Open Compute Project chassis is
the other components in a server, inc
motherboard and power supply. Over

http://opencompute.org/ , Since 2011 April

Open Compute Project：facebook的謀略

數位時代網站 | 撰文者：劉翰謙　發表日期：2011-04-08

讚　7 人說讚。快免費註冊來查看你的朋友對什麼說讚。　+1　0

你聽過開源軟體，但〝開源硬體〞聽起來就顯得陌生多了。The Open Compute Project，facebook的伺服器/資料中心硬體設計開放計畫，將要從底層科技開始改變整個業界。

當然，facebook
了相關的協定，
舉。以往科技公
由，根據不同的
了一大型的科技
Google、facebo
或特殊的需求，
甚至算一種一般

帶動開放雲端平台　華碩支持Open Compute Project(OCP)專案

2011/11/21-徐漢高

華碩領先支持由Facebook主導的OCP(Open Compute Project)—開放運算專案，此專案旨在提供自由和開放的伺服器技術，通過企業間密切合作、開放討論和知識共享等功能，制定大規模雲端運算的技術規格，為雲端產業提供高效、節能的合作平台。

華碩超級電腦和伺服器具備4大獨特設計，包含「高可用性」：獨特的電源設計與獨立熱插拔節點設計，使其具高可用性和易於維護；「高擴展性」：獨家研發PIKE卡

[1] Open Compute Project ： facebook 的謀略
http://www.bnext.com.tw/article/view/cid/0/id/17924
[2] 帶動開放雲端平台　華碩支持 Open Compute Project(OCP) 專案
http://www.digitimes.com.tw/tw/dt/n/shwnws.asp?id=0000259916_1YA8G7432BZVPB5RJ1HZX

SSD 的工作溫度限制？

網管人NetAdmin

2012/2/13

節省能源降低熱當風險 高環溫資料中心概念萌芽

余采霏

如何帶走IT設備在運轉過程中產生的熱能，一直都是資料中心必須面對的難題，尤其是在高密度機房下，伺服器透過堆疊提供高運算效率的同時，散熱不足卻也是造成伺服器可能當機的元兇之一，也因此，機櫃需要冷氣空調來帶走囤積的熱能，而這些冷卻成本卻也為企業帶來高昂的費用。根據數據統計，資料中心目前的用電量約佔全球總用電量的1.5%，換算為每年電費則可高達260億美元。

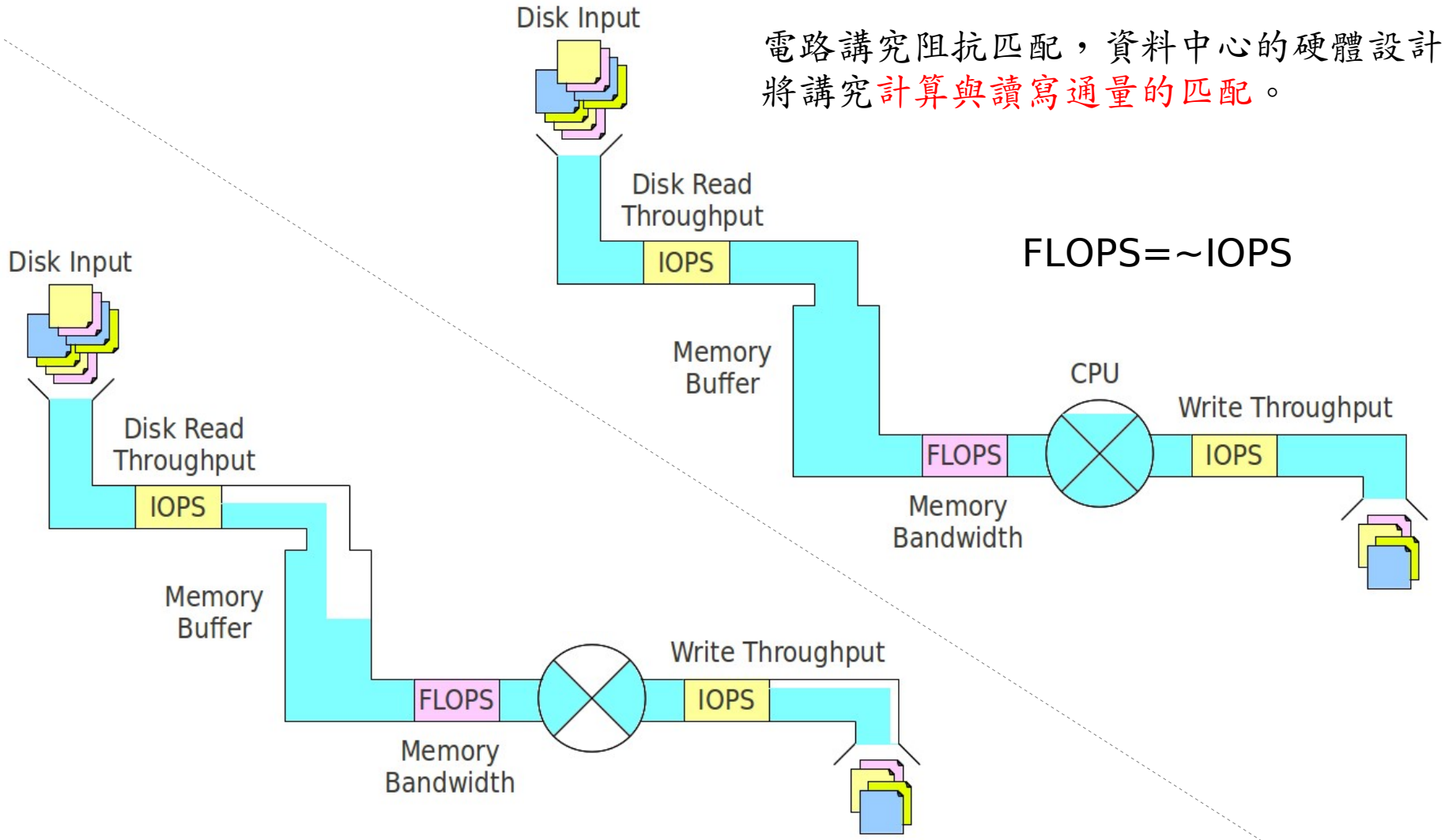
長期以來，機房設施業者想方設法要解決這個難題，從縮減機房空間、提高冷源效率、區分冷熱通道，甚至降低室內冷卻空調改用外氣的自然冷源，這些都是不錯的方案，也為企業解決不少挑戰，但卻還不夠。

http://www.netadmin.com.tw/article_content.aspx?sn=1202070005

Is There A Golden Ratio for Big Data Hardware Design

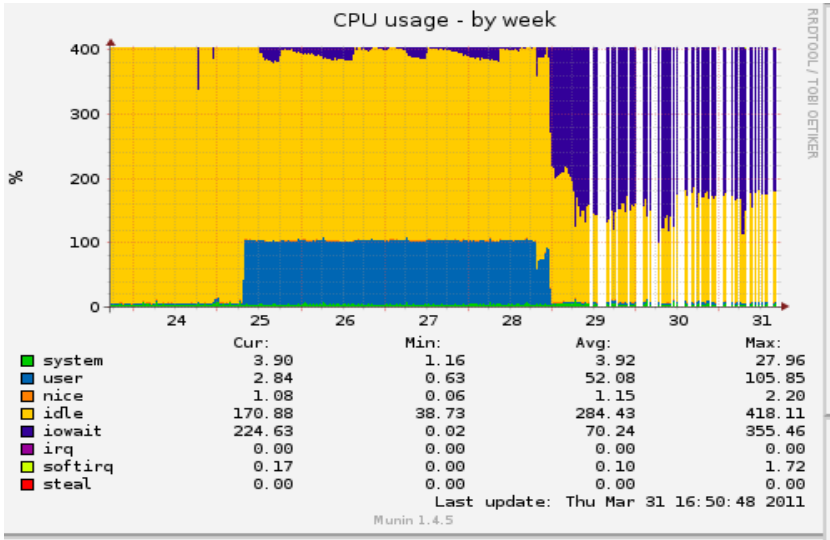
NARLabs

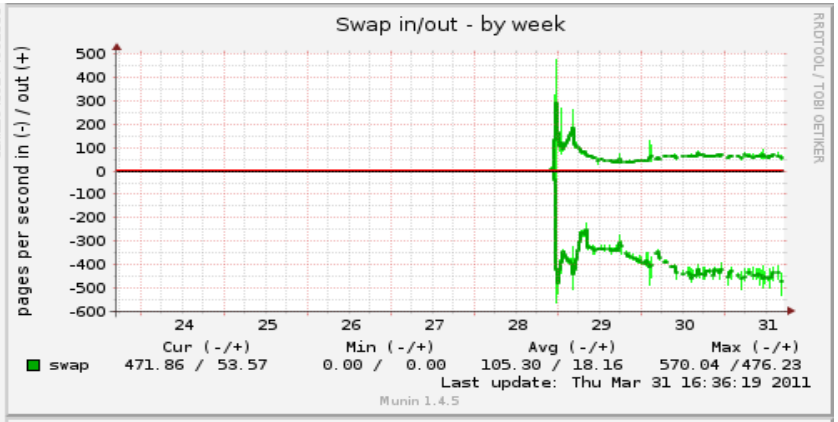
電路講究阻抗匹配，資料中心的硬體設計將講究計算與讀寫通量的匹配。

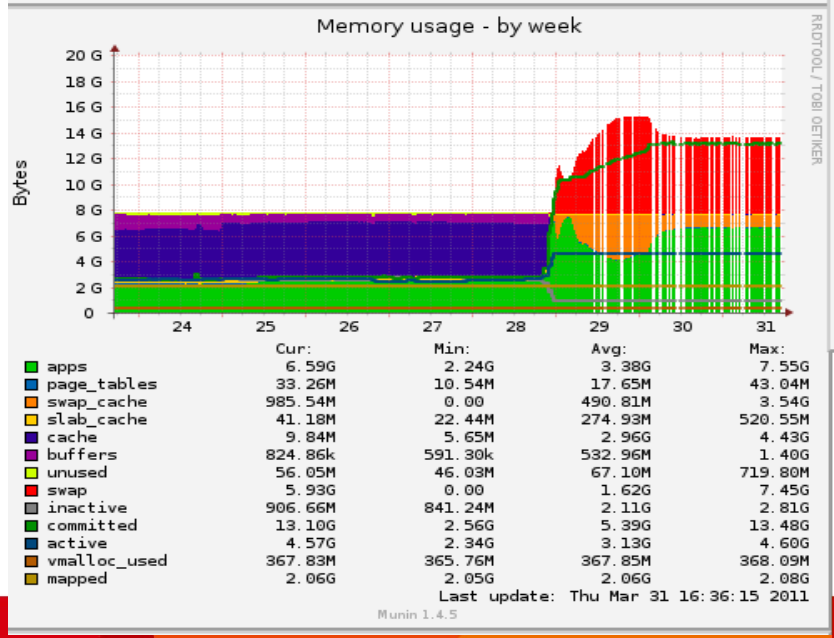


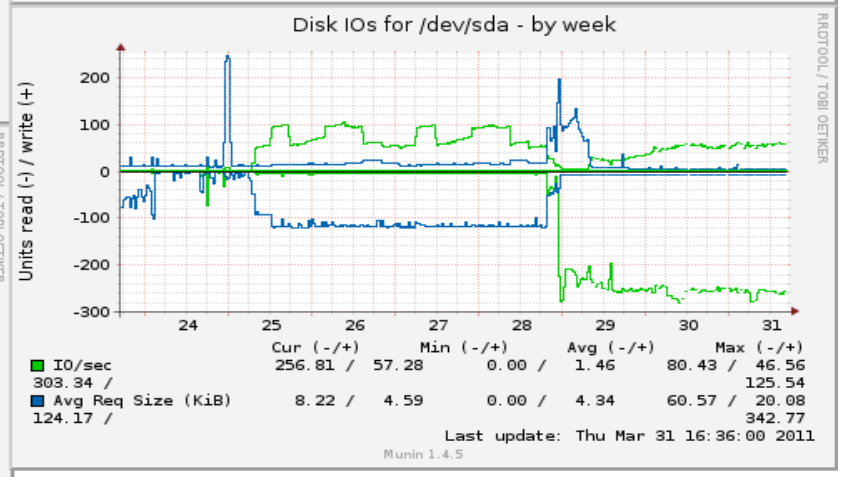
FLOPS=~IOPS

當讀寫 I/O 遠慢於 CPU 運算時，會發生 IOWAIT









程式執行與狀態監控形成回饋控制（Feedback Control）

Conclusion:

1. 雲端造成商業模式的改變，未來晶片設計廠與主機板製造商須面臨更嚴苛的 B2B 競爭，因為主要客戶將會是 IDC 。
2. 一雲多螢，通訊的需求將持續增加。行動通訊晶片前景可期！

   因應頻寬成長速度有限，資料成長卻急遽攀升，通訊晶片應內建壓縮標準。反思： Opera Mini 為何沒有成功？
3. 巨量資料的處理架構於 2013 年大致成型，未來企業應會採用混合模式的資訊架構，以因應 Big Data at Rest 與 Big Data in Motion 兩種問題，還有對不同查詢時間顆粒度的需求。
4. 因應 Big Data 時代來臨，建議晶片設計廠與主機板製造商應從計算機結構思考 I/O 阻抗匹配的最佳化，即使目前南北橋晶片的頻寬很大，但由於是多重路徑，很可能會是硬碟之外的效率瓶頸。
5. 以 ****** 目前的產品定位，宜佈局 Big Data in Motion ，特別是有 Realtime Predictive Analytics 需求的高附加價值產品。

實際案例：巨量資料未來的挑戰

基於保護洽談中、進行中的實際案例，以下言論純屬討論與臆測~

NARLabs

問題與討論

Questions?