



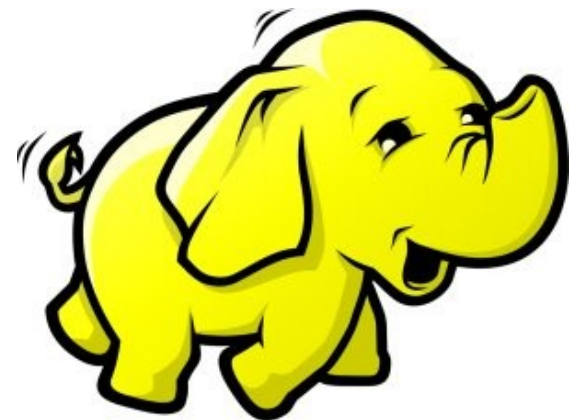
雲端運算 - 商業模式、核心技術與架構

Business Model and Core Technologies of Cloud Computing

Jazz Wang

Yao-Tsung Wang

jazz@nchc.org.tw



WHO AM I ? 這傢伙是誰啊? JAZZ ?

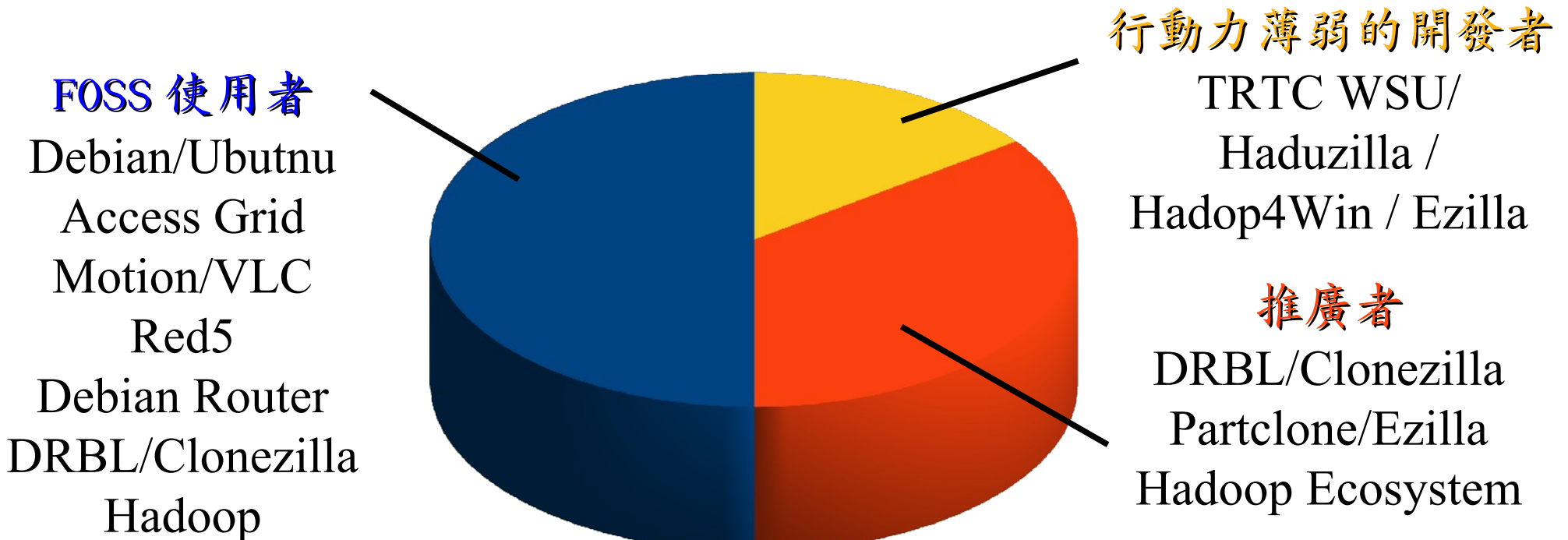
- 講者介紹：

- 國網中心 王耀聰 副研究員 / 交大電控八九級碩士
- jazz@nchc.org.tw



- 所有投影片、參考資料與操作步驟均在網路上

- <http://trac.nchc.org.tw/cloud>
- 由於雲端資訊變動太快，愛護地球，請減少不必要之列印。



Agenda 演講大綱

雲端商業模式及三大關鍵技術

雲端運算與鴻海產品線之關聯

架構即服務：虛擬化技術

平台即服務：資料分析技術

軟體即服務：網頁服務技術



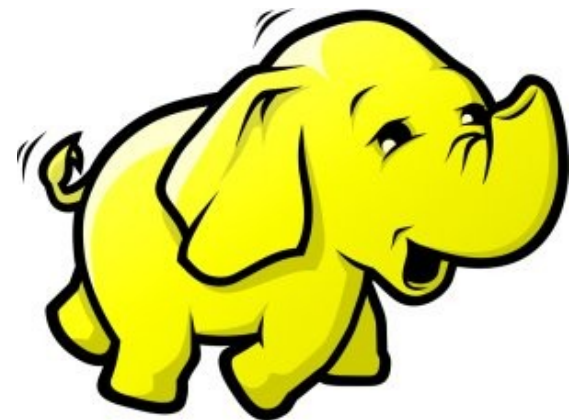
雲端商業模式及三大關鍵技術

Three Key Technologies of Cloud Computing

Jazz Wang

Yao-Tsung Wang

jazz@nchc.org.tw





什麼是雲端運算啊？

What is Cloud Computing ?

雲端概念

雲端運算不是一項全新技術，
而是一項概念。
雲端的意義不在技術，
而在商業模式的改變。

Paradigm Shift of Cloud Business Model !!

雲端商業模式的典範轉移

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

The wisdom of Clouds (Crowds)

雲端序曲：雲端的智慧始終來自於群眾的智慧

2006年8月9日

Google 執行長施密特 (Eric Schmidt) 於 SES'06 會議中首次使用「雲端運算 (Cloud Computing) 」來形容無所不在的網路服務

2006年8月24日

Amazon 以 Elastic Compute Cloud 命名其虛擬運算資源服務



New IT Architecture toward Cloud Computing !!

用雲掌握資料，加以分析，形成智能給端用



雲

資料中心
提供服務

雲端設計新思維：端的智能來自於雲的服務

Devices share the wisdom of Cloud

端



各類裝置
存取服務

National Definition of Cloud Computing

美國國家標準局 NIST 給雲端運算所下的定義

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

4 Deployment Models of Cloud Computing

雲端運算的四種佈署模型

Public Cloud
公用雲端

Target Market
is **S.M.B.**
主要客戶為
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**
為主要客戶
Enterprise is
key market

Community Cloud
社群雲端

Academia **學術**為主



私有雲端
Private Cloud

3 Service Models of Cloud Computing

三種服務模式（商業模式市場區隔）

IaaS

Infrastructure as a Service

架構即服務

PaaS

Platform as a Service

平台即服務

SaaS

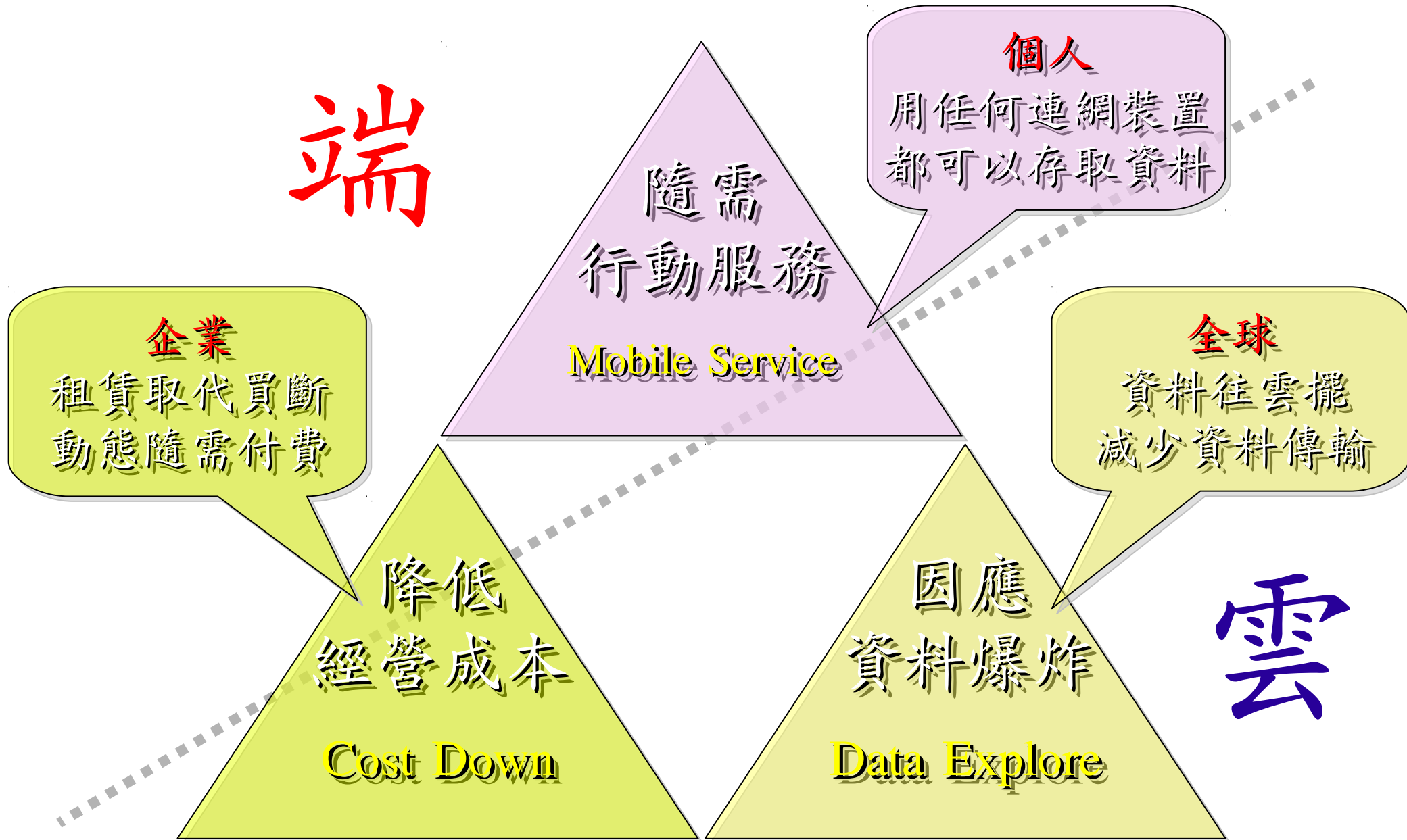
Software as a Service

軟體即服務



Key Driving Forces of Cloud Computing

雲端運算的關鍵驅動力

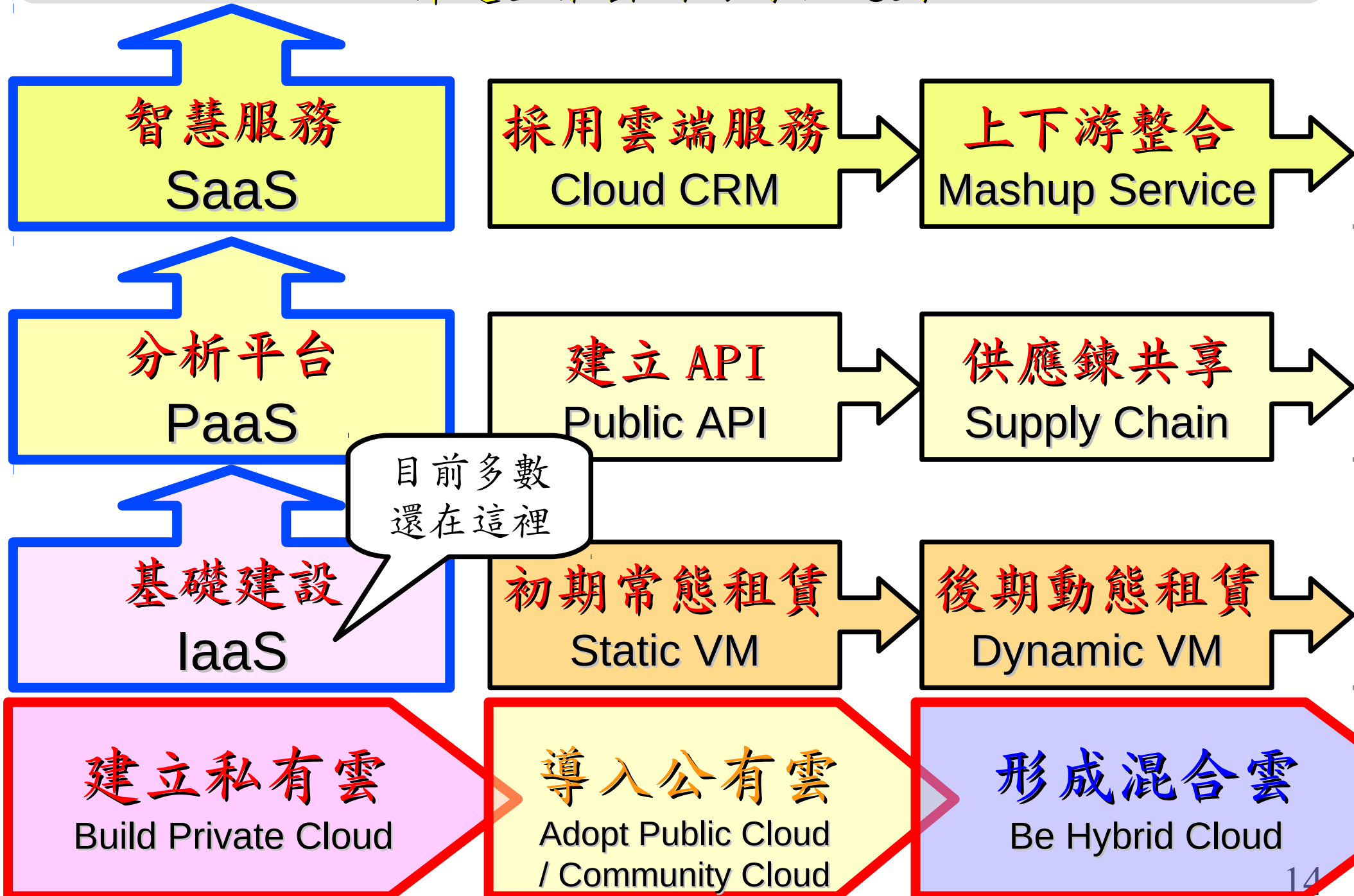


Three Key Technologies !! 三種服務模式 vs. 三大關鍵技術



Roadmap to build Your Enterprise Cloud !!

佈建企業雲端的時程規劃



Gartner Priority Matrix for Cloud Computing, 2010 !!

連 Gartner 也是這麼說的！

benefit	years to mainstream adoption			
	less than 2 years	2 to 5 years	5 to 10 years	m
transformational	Virtualization	"Big Data" and Extreme Information Processing and Management Cloud Advertising Cloud Computing Platform as a Service (PaaS) Public Cloud Computing/the Cloud	Community Cloud DevOps Hybrid Cloud Computing Real-Time Infrastructure	
high		Application PaaS (aPaaS) Cloud/Web Platforms Cloud-Enabled BPM Platforms Elasticity Enhanced Network Delivery Infrastructure as a Service (IaaS) Multitenancy Private Cloud Computing	Cloud Application Development Services Cloud Parallel Processing Cloud Security Cloud Security and Risk Standards Cloud Services Brokerage Cloudbursting Private PaaS	

兩年內應該
導入虛擬化技術

五年內應該
導入 Big Data

Source: "Hype Cycle for Cloud Computing, 2011", David Mitchell Smith, Gartner
<http://www.gartner.com/id=1753115>



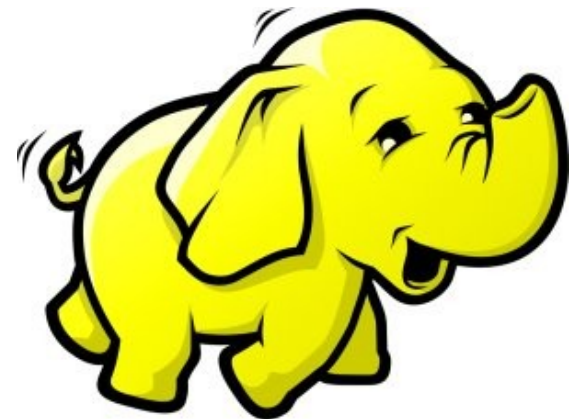
雲端運算與鴻海產品線之關聯

The Relation between Cloud Computing and Foxconn

Jazz Wang

Yao-Tsung Wang

jazz@nchc.org.tw



Cloud Market Segmentation of Foxconn !! 鴻海集團的雲端定位

應用軟體
供應商



端

行動裝置
共通平台



各類裝置
存取服務

軟體服務
供應商



雲

資料中心
機房維運



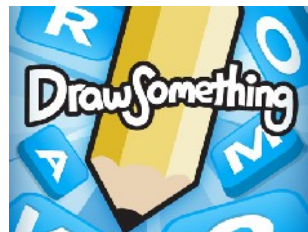
資料中心
提供服務

基本硬體
建設組件



Let's focus on the product of Foxconn !! 鴻海產品線的雲端定位

應用軟體
供應商



物聯網

屠家照護

行動裝置
共通平台

htc
quietly brilliant



手機

機上盒

軟體服務
供應商

Google™

amazon
webservices™

準系統

資料中心
機房維運



acer

主機板

伺服器

基本硬體
建設組件



FOXCONN
鴻海科技集團

交換器

路由器

儲存器

光纖卡

Relation between Cloud Technologies and Foxconn Products!!

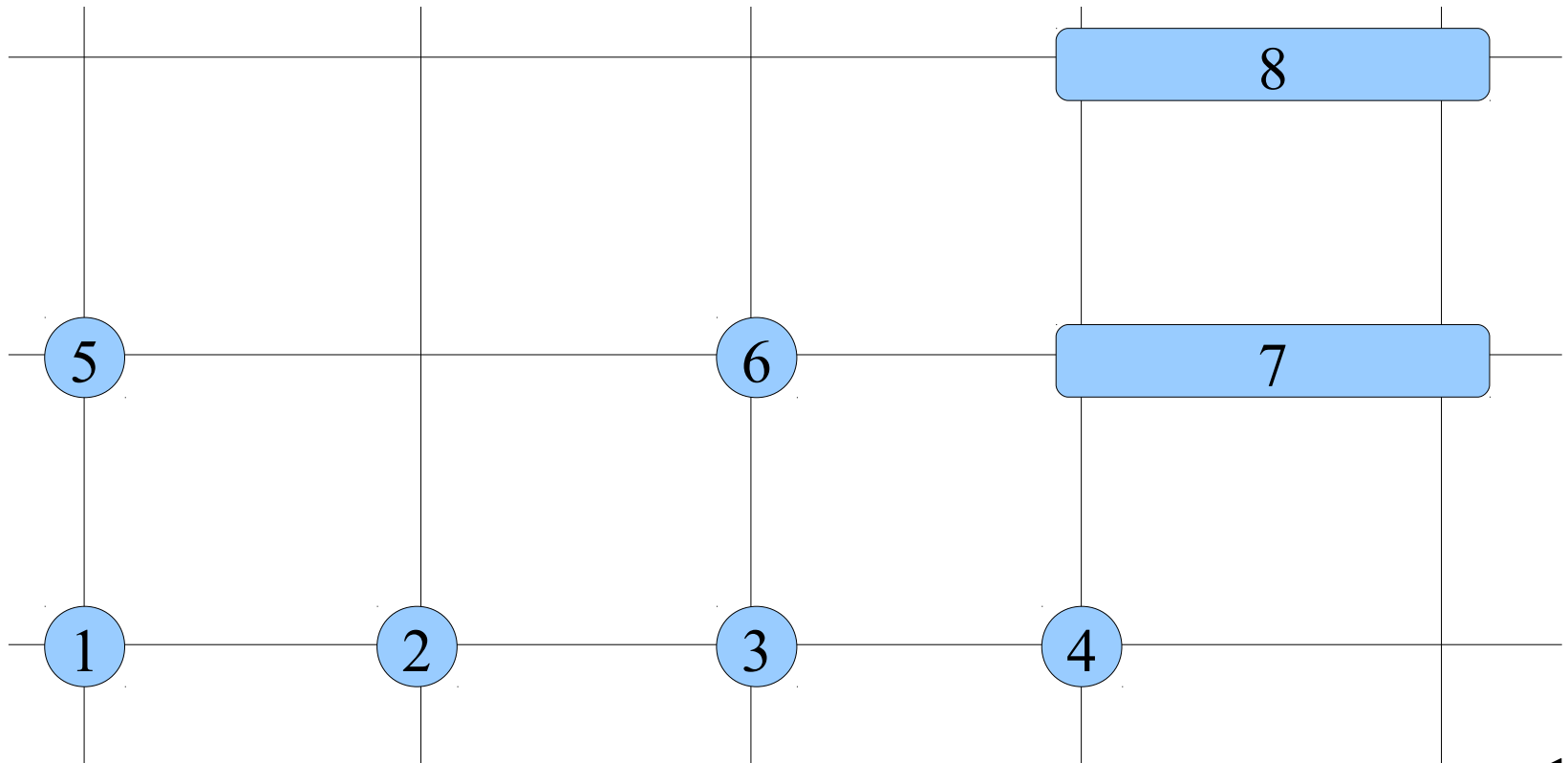
給鴻海產品線與雲端關鍵技術的八個建議



Web 2.0
網頁服務

Big Data
資料分析

Virtualization
虛擬化技術





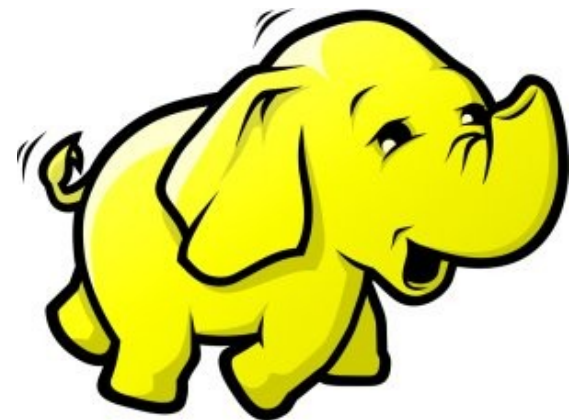
架構即服務：虛擬化技術

Introduction to Virtualization Technologies

Jazz Wang

Yao-Tsung Wang

jazz@nchc.org.tw



CIO 2010 : Virtualization, Cloud and Web 2.0

虛擬化技術是 2010 年 CIO 資訊長認為最重要的資訊投資

CIO strategic technologies reflect increased interest in “lighter-weight” solutions

CIO technologies

Ranking of technologies CIOs selected as one of their top 5 priorities in 2010

Ranking	2010		2009	2008	2007
Virtualization	1	↑	3	3	5
Cloud computing	2	↑	16	*	*
Web 2.0	3	↑	15	15	*
Networking, voice and data communications	4	↑	6	7	4
Business intelligence (BI)	5	↓	1	1	1
Mobile technologies	6	↑	12	12	11
Data/document management and storage	7	↑	10	9	9
Service-oriented applications and architecture	8	↑	9	10	7
Security technologies	9	↓	8	5	6
IT management	10		*	*	*
Enterprise applications	11	↓	2	2	2

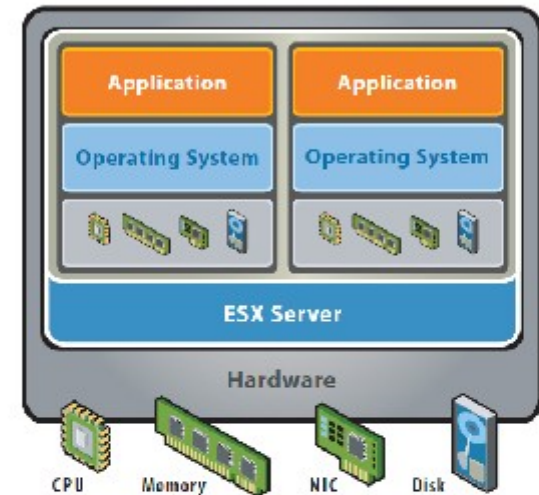
* New question for that year

Source: *Gartner Executive Programs* : “Leading in Times of Transition: The 2010 CIO Agenda”

First Wave of Cloud : Virtualization

企業導入雲端第一波主打：虛擬化！！

- 導入虛擬化的效益高於雲端運算
- 導入評量指標：
 - 總硬體使用率 (100% CPU 使用 ?)
 - 總電源開銷、空調、管理人力
- 商業解決方案：
 - **VMWare** ESXi / vSphere
 - **Microsoft** Hyper-V
 - **Citrix** XenServer
- 預期效益：
 - 減少伺服器採購成本與營運成本
 - **(Server Consolidation)**
 - 增加管理彈性與災害復原機制
 - (Ex. 異常斷電造成的服務修復)



VMware ESX Server virtualizes server storage and networking, allowing multiple applications to run in virtual machines on the same physical server.



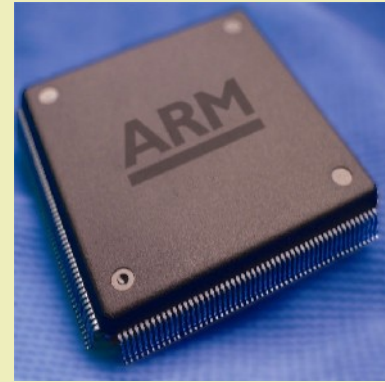
Windows Server™ 2008
Hyper-V™



Virtualization ?? Emulator ??

模擬器，最早的虛擬化技術 ??

Virtual Hardware / OS 模擬硬體 / 作業系統



mame4iphone

Latest Version:

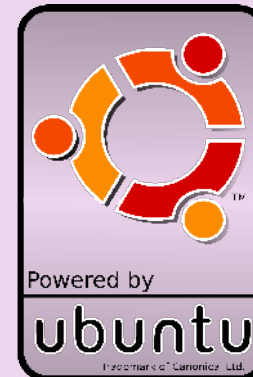
MAME 0.138 15 May 10

— DOWNLOAD NOW —

[Download source updates to MAME 0.138](#)

Mac4Lin

QEMU



Physical Hardware / OS 實際硬體 / 作業系統

What is Virtualization ?? 虛擬化技術有哪些呢 ??

Application Virtualization 應用程式虛擬化

Desktop Virtualization
Client Virtualization 桌面虛擬化

Presentation Virtualization 顯示虛擬化

OS-level Virtualization 作業系統虛擬化

Network Virtualization 網路虛擬化

Storage Virtualization 儲存虛擬化

資料庫虛擬化

資料虛擬化

Database Virtualization

Data Virtualization

Open Source for Virtualization

虛擬化技術對應的自由軟體

Application Virtualization

Ex. VMWare ThinApp

Desktop Virtualization
Client Virtualization

Redhat SPICE

Presentation Virtualization

VNC, FreeNX,
FreeRDP

OS-level Virtualization

Xen, KVM, OpenVZ
QEMU, VirtualBox

Network Virtualization

OpenFlow vSwitch

Storage Virtualization

Lessfs, SDFS

NIST Mapping of Cloud Technologies

美國國家標準局的定義主要鎖定虛擬化技術

5. Measured Service
可被監控與量測的服務

Monitoring / AAA
狀態監控與認證收費機制

1. On-demand self-service.
隨需自助服務

VM Management Tool
虛擬機器管理平台

2. Broad network access
隨時隨地用任何網路裝置存取

OS-level Virtualization
作業系統虛擬化

3. Resource pooling
多人共享資源池

Network Virtualization
網路虛擬化

4. Rapid elasticity
快速重新佈署靈活度

Storage Virtualization
儲存虛擬化

NIST Mapping of Open Source Cloud

美國國家標準局的定義對應自由軟體技術

Monitoring / AAA
狀態監控與認證收費機制

Ganglia, Nagios
/ OpenID, SAML

VM Management Tool
虛擬機器管理平台

OpenStack (美)
OpenNebula (歐)

OS-level Virtualization
作業系統虛擬化

Xen, KVM, OpenVZ

Network Virtualization
網路虛擬化

OpenFlow vSwitch

Storage Virtualization
儲存虛擬化

Lessfs, SDIFS, ZFS

光纖卡

伺服器

機上盒

居家照護

儲存器

交換器

準系統

手機

物聯網

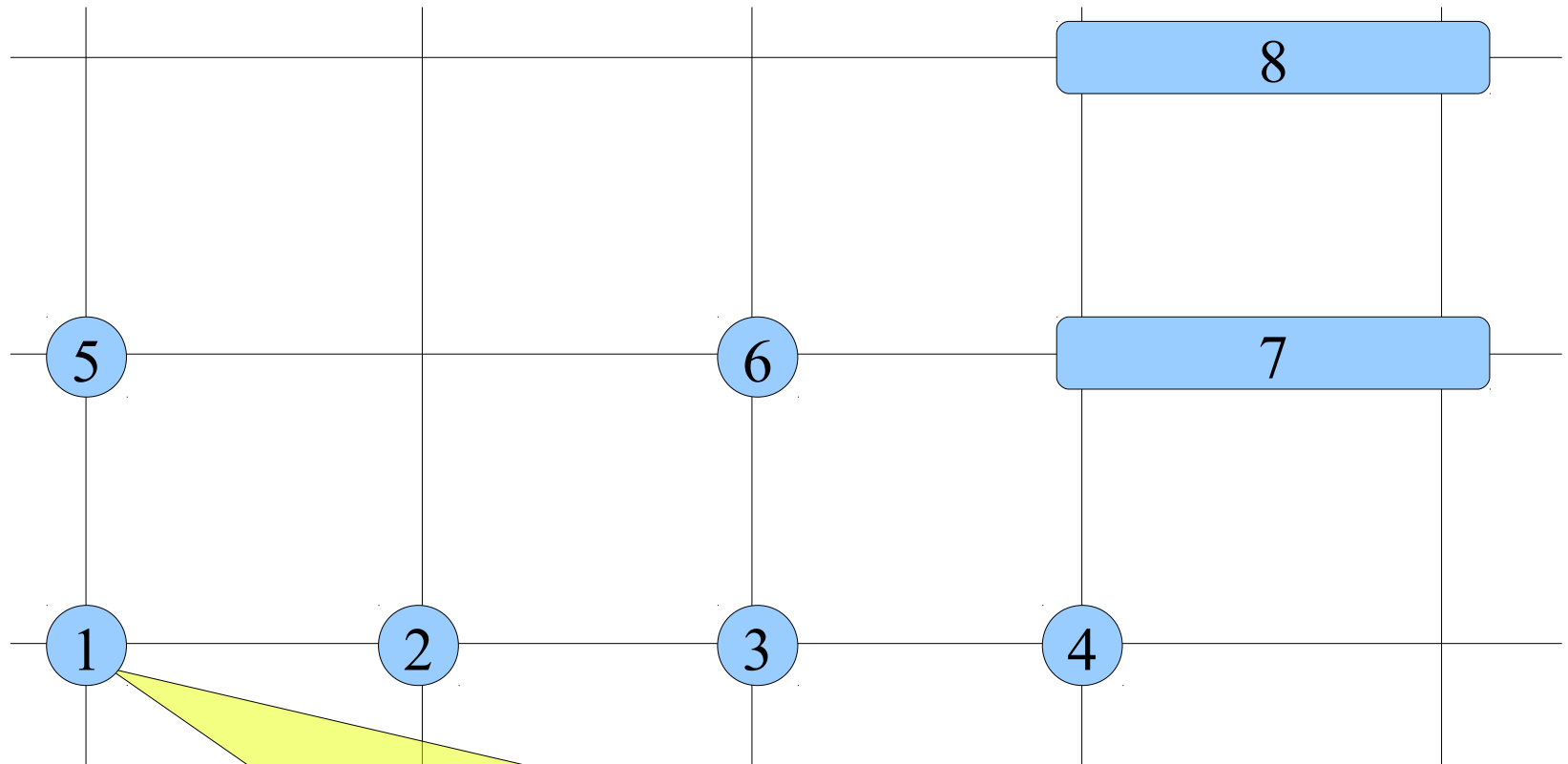
路由

主機板

Web 2.0
網頁服務

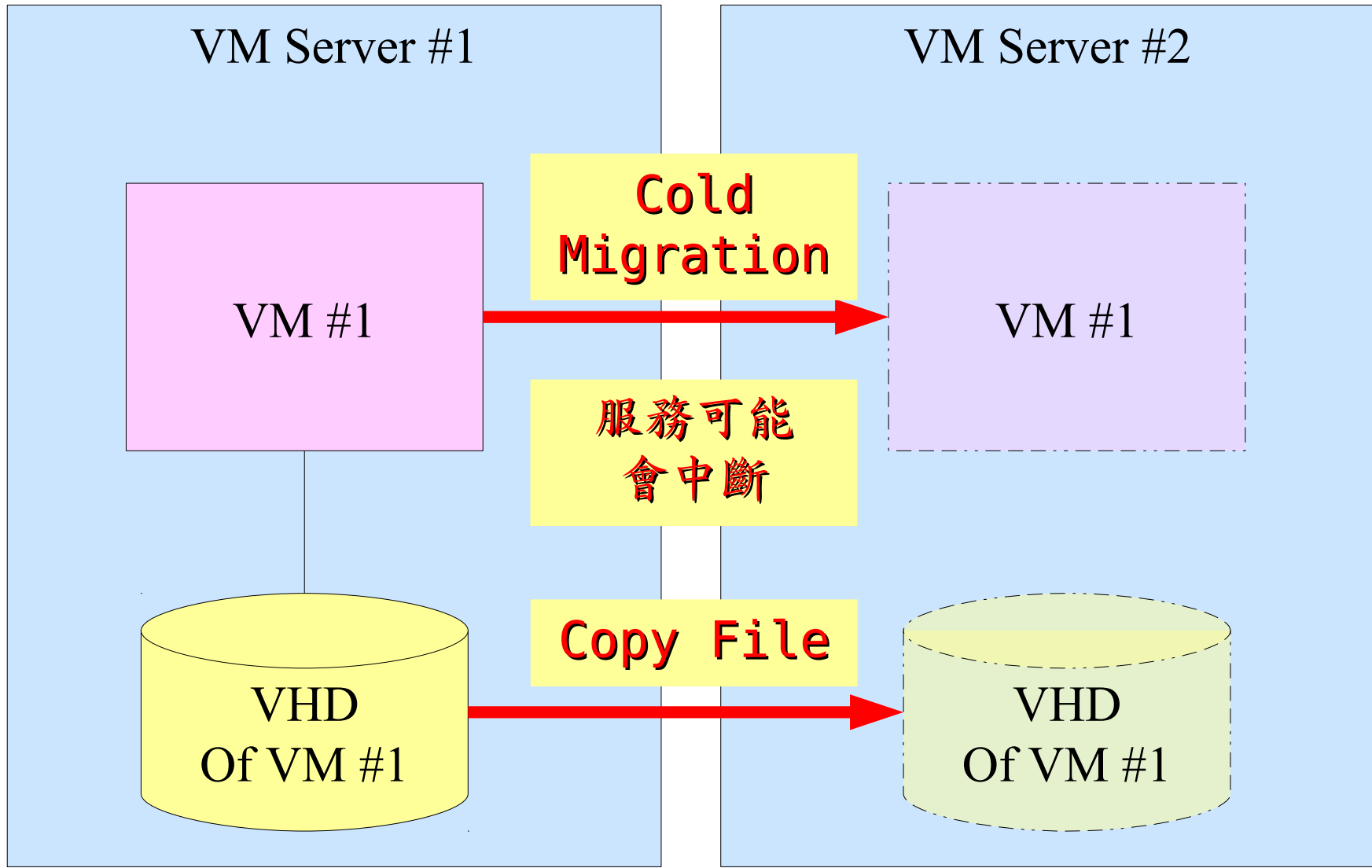
Big Data
資料分析

Virtualization
虛擬化技術

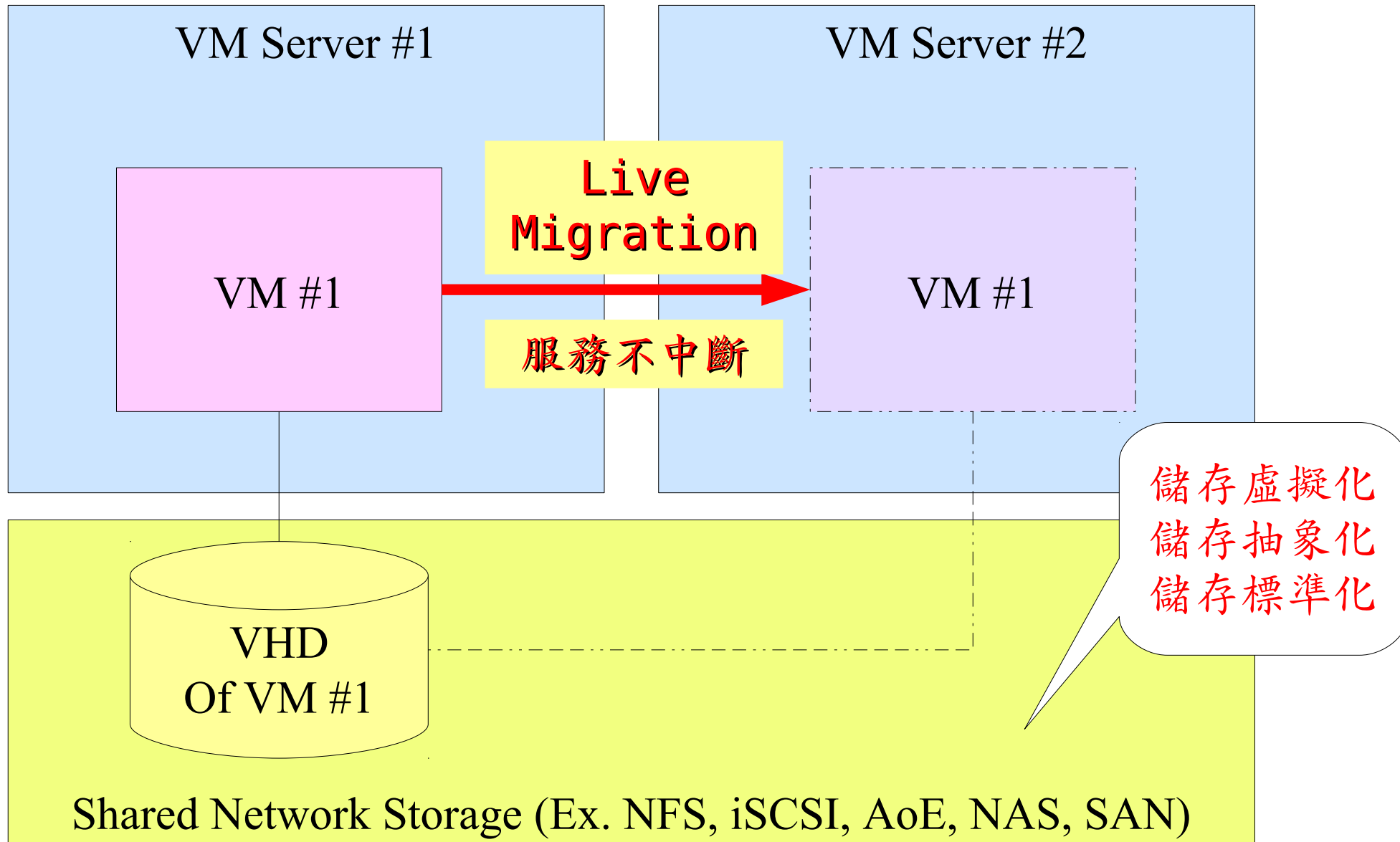


趨勢 (1) 為了建構虛擬化架構，需要共通儲存設施！
 建議 (1) 進行對儲存虛擬化支援性的測試及策略聯盟
 行動 (1) 從 RAID 到分散式檔案系統的效能測試

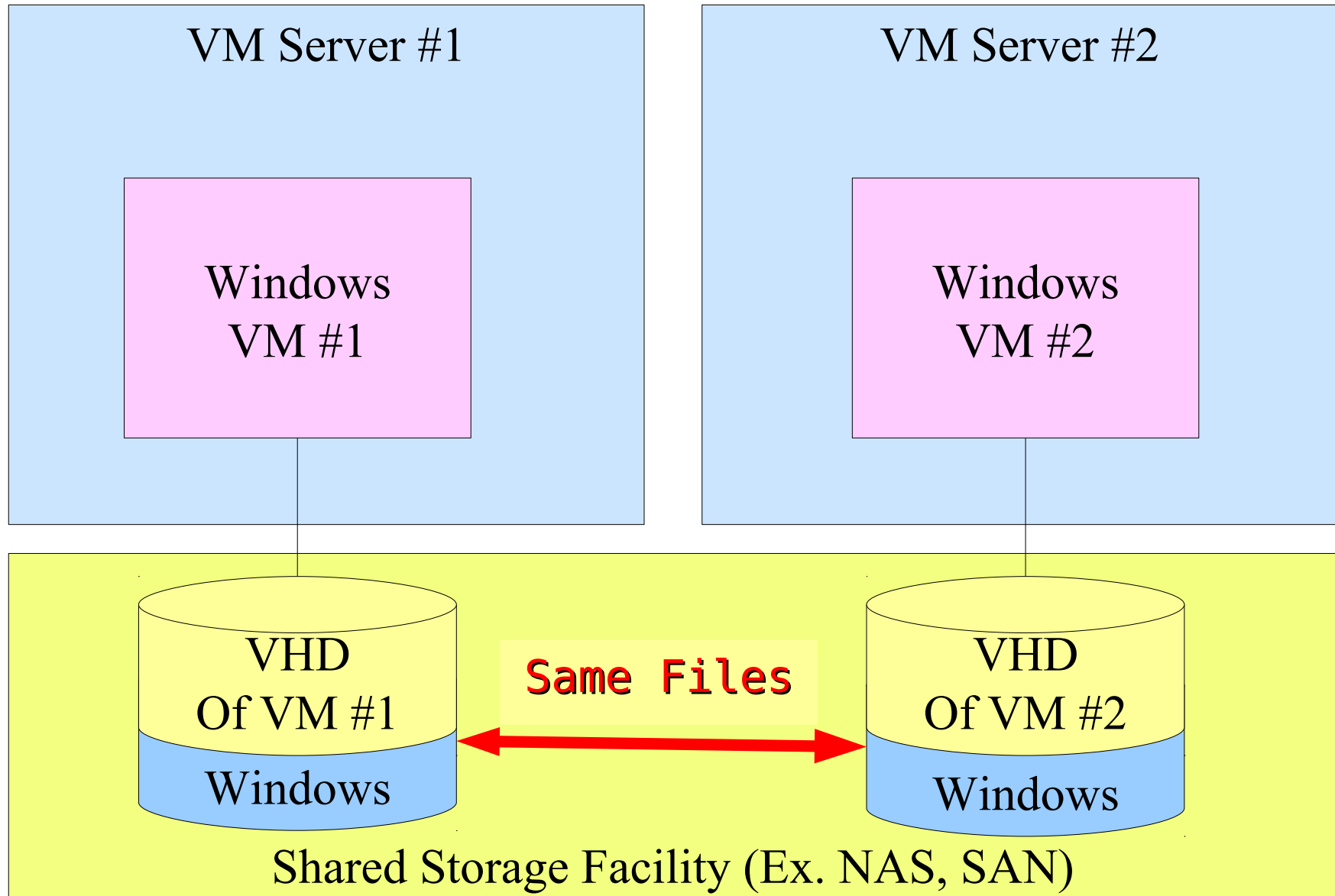
Why Storage Virtualization !! 為何需要儲存虛擬化？



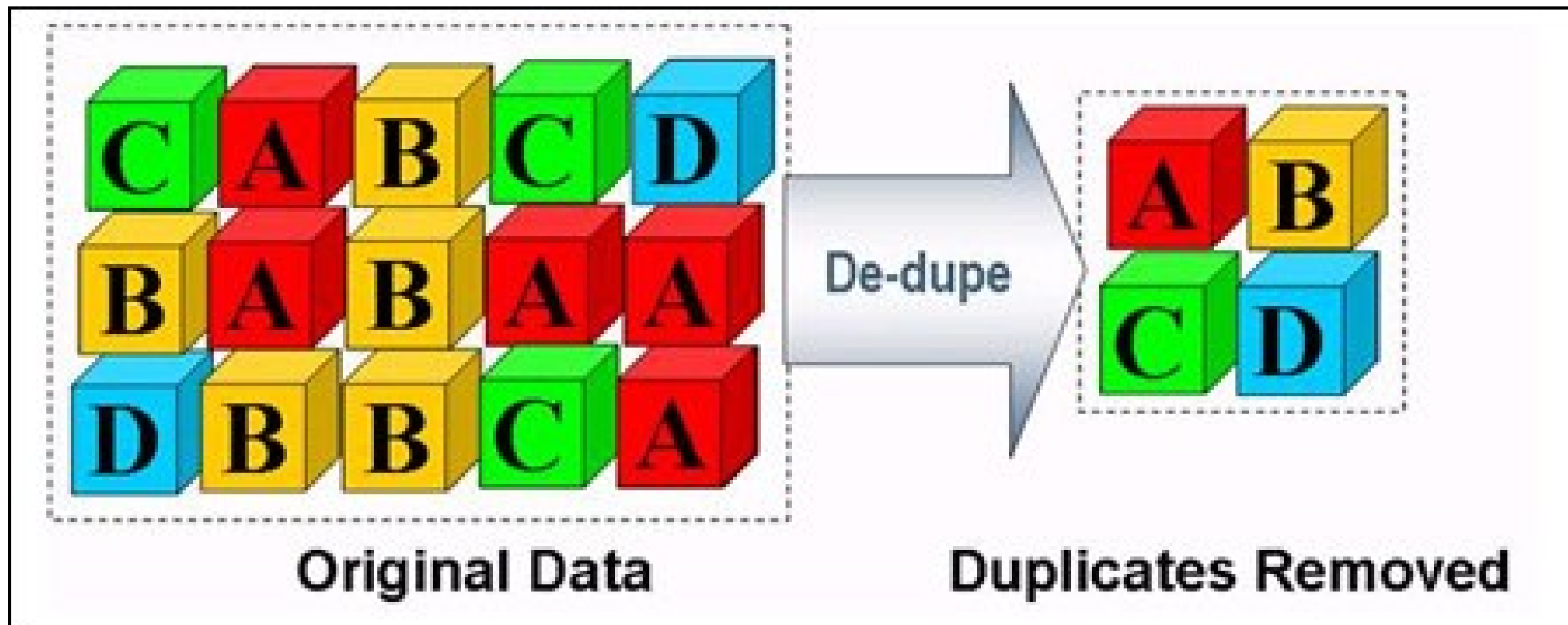
Why Storage Virtualization !! 為何需要儲存虛擬化？



Why Storage Virtualization !! 為何需要儲存虛擬化？



Deduplication? 去除重複儲存的資料?



- 資料整合為跨單位整合的第一步 !!
- 商業硬體方案：EMC、NetApp
- 自由軟體方案：
 - ZFS、Lessfs、SDFS...

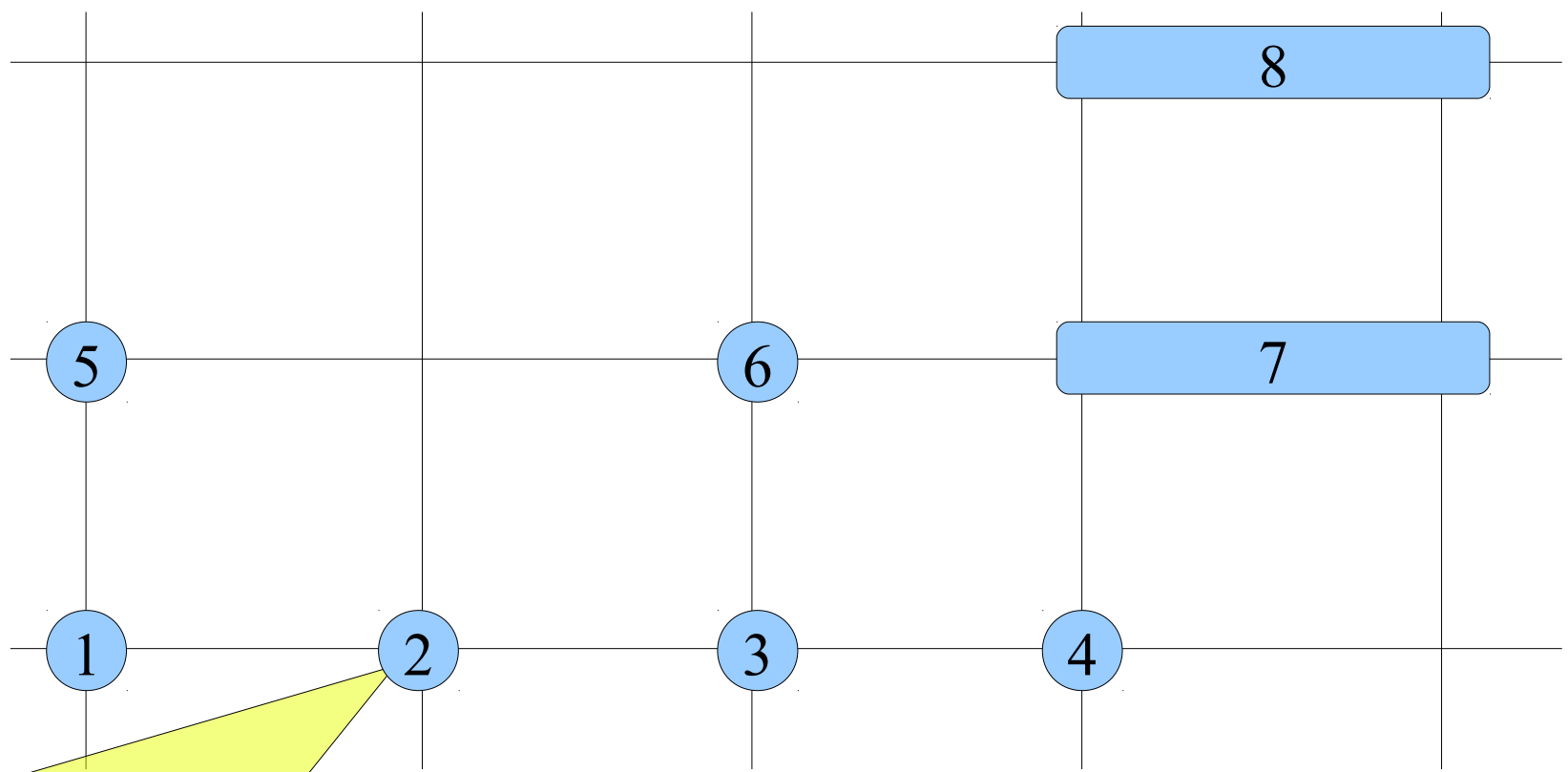




Web 2.0
網頁服務

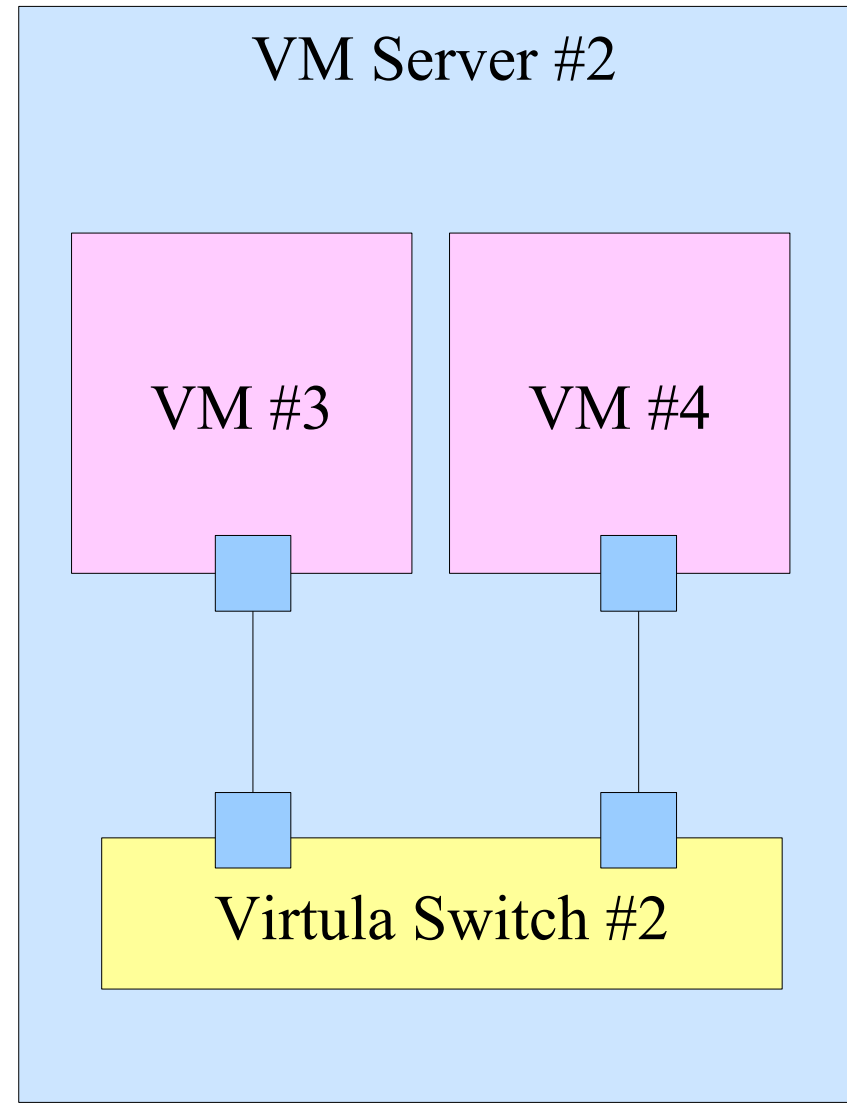
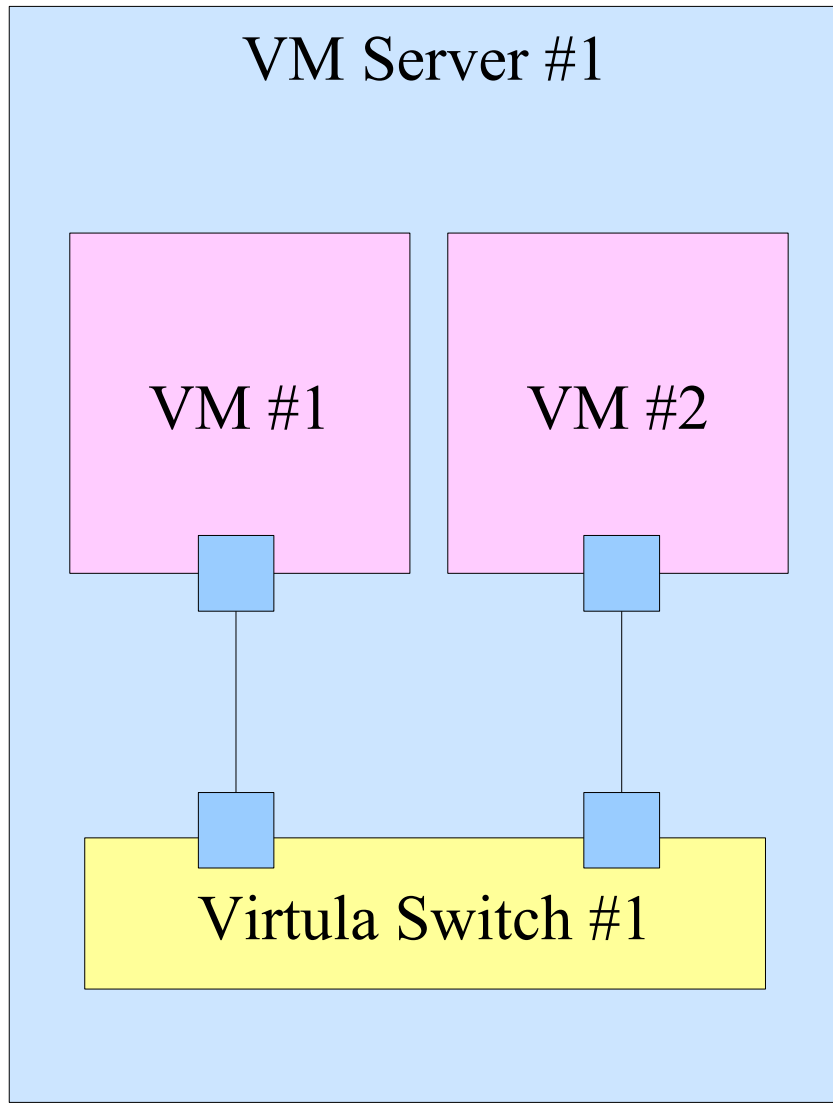
Big Data
資料分析

Virtualization
虛擬化技術

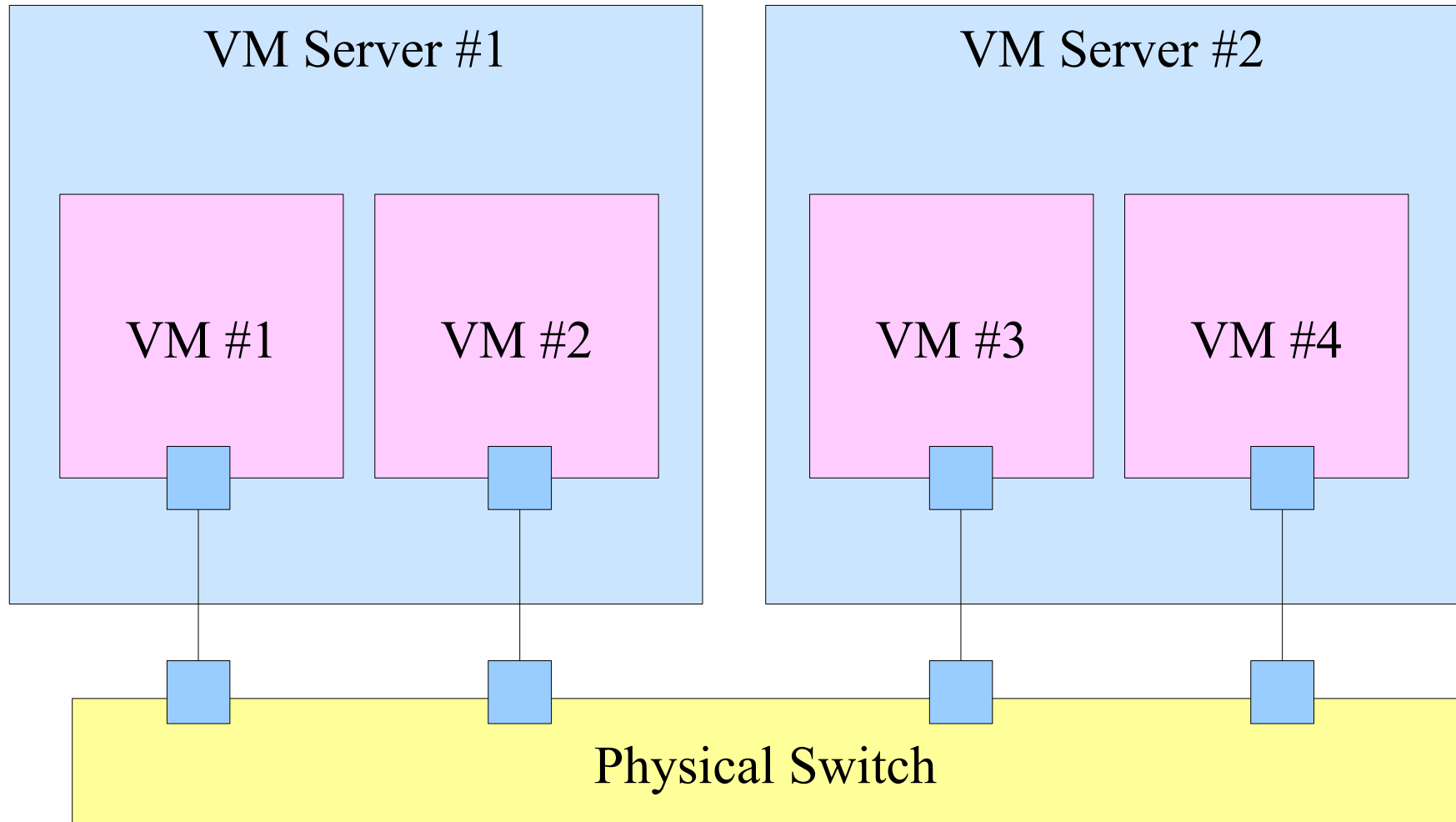


趨勢 (2) 為了建構虛擬化架構，需要新的網路拓樸！
 建議 (2) 進行對網路虛擬化支援性的測試及策略聯盟
 行動 (2) 導入 OpenFlow / OpenVSwitch 標準

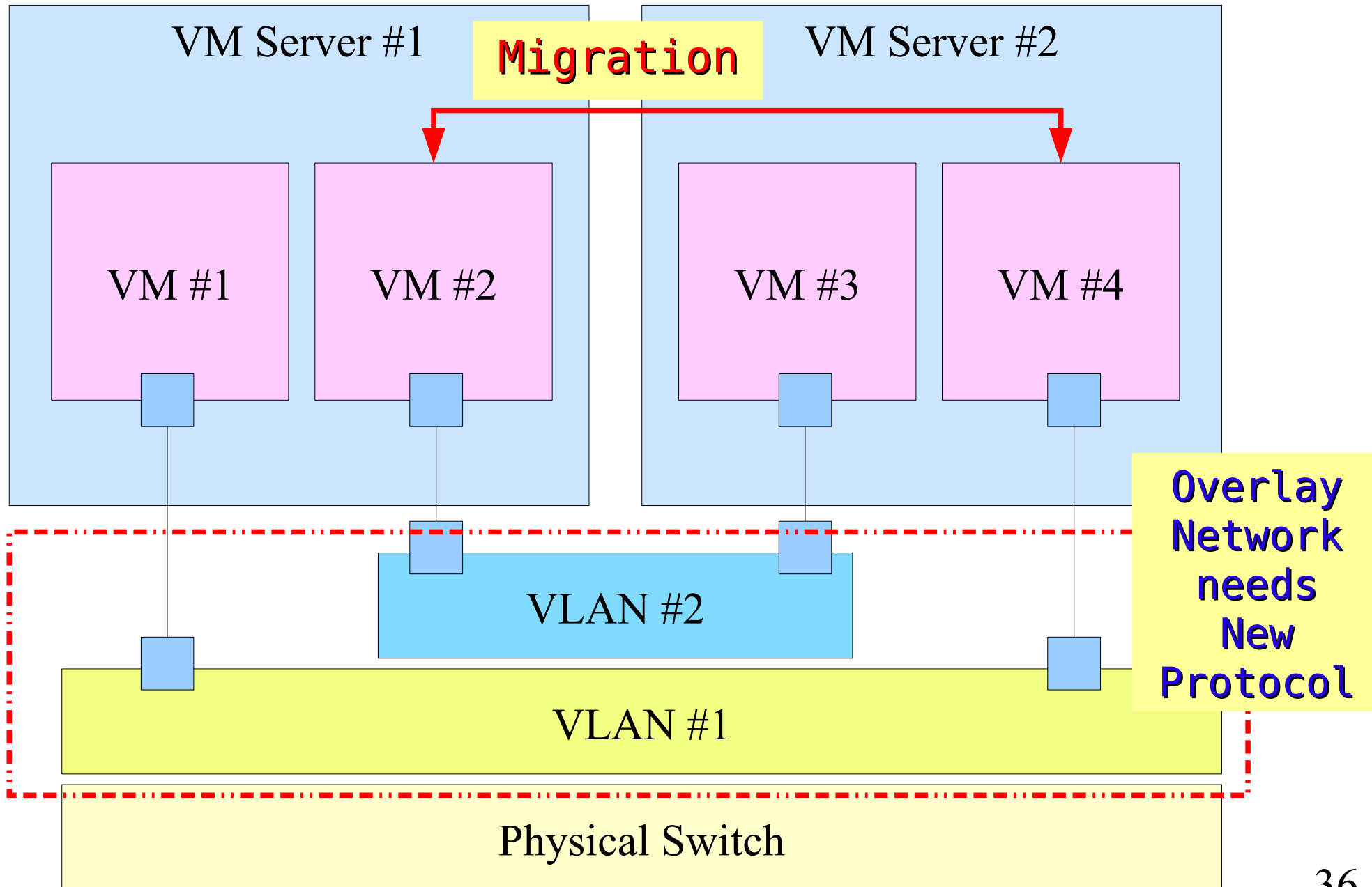
Why Network Virtualization !! 為何需要網路虛擬化？



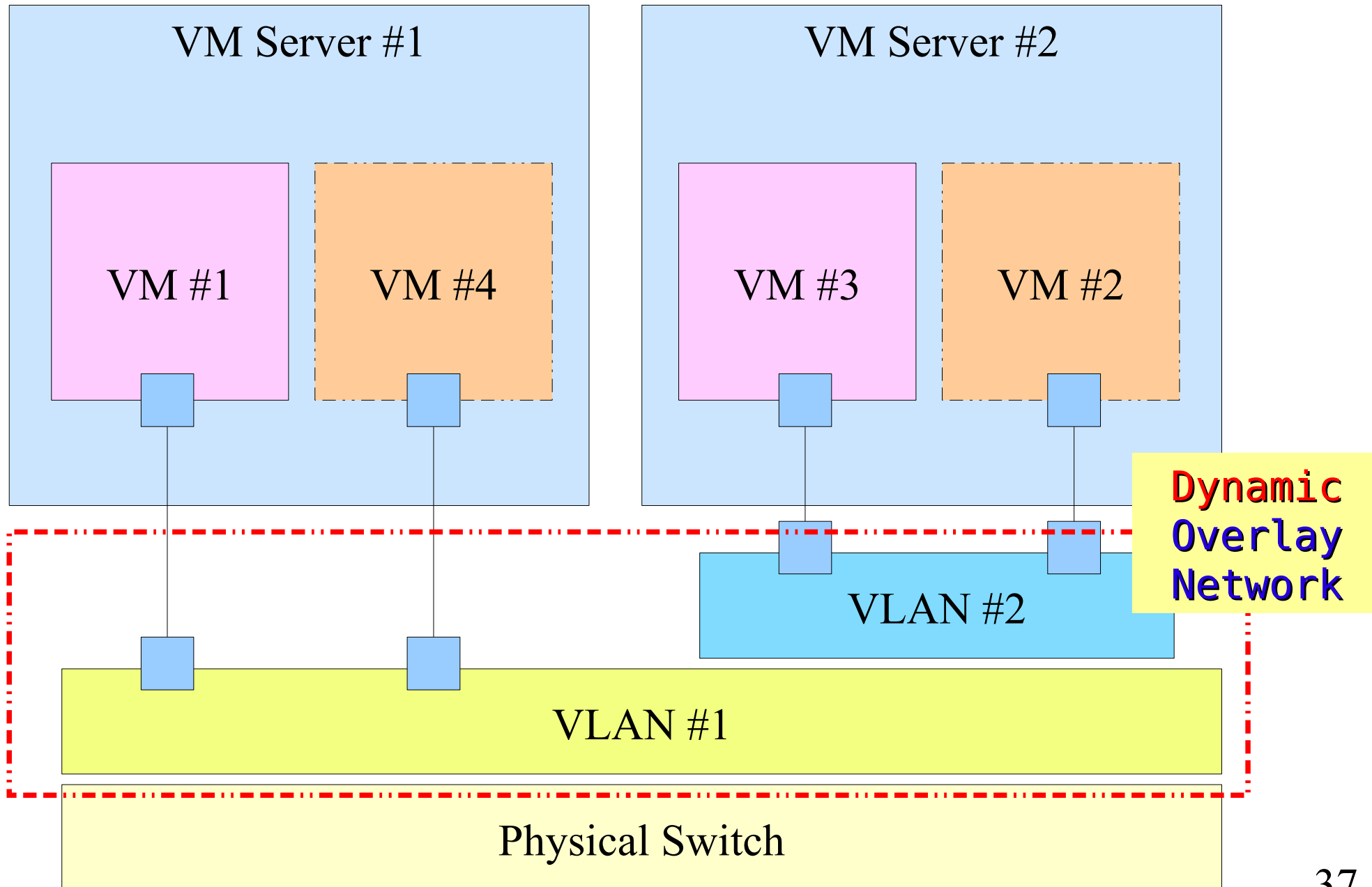
Why Network Virtualization !! 為何需要網路虛擬化？



Why Network Virtualization !! 為何需要網路虛擬化？



Why Network Virtualization !! 為何需要網路虛擬化？



OpenFlow Network Virtualization 網路虛擬化標準

- Open Networking Foundation 為非營利組織，致力推廣「軟體定義網路 (Software-Defined Networking; SDN)」的新網路方法
- SDN 包含幾個相當重要的元素，其中關鍵為 OpenFlow，其能為服務供應商提供更精確的網路基礎設施管理能力



The screenshot shows the OpenFlow website homepage. At the top left is the OpenFlow logo, which consists of a blue circle with a white arrow pointing right, followed by the text "OpenFlow". To the right of the logo is a search bar with the text "Google™ Custom Search" and a "Search" button. Below the logo and search bar is a navigation menu with buttons for "Home", "Videos", "Documents", "News", "Research", and "About". The main content area features a dark blue banner with the text "Innovate in Your Network" and a paragraph describing OpenFlow's capabilities. To the right of this banner are two buttons: "Learn More" with an information icon and "Develop" with an icon of two people and gears. Below the banner is a "Participate" section with text about the Open Networking Foundation (ONF). At the bottom left is an "OpenFlow News" section with a headline "Boston CIO Workshop on Expanding OpenFlow/SDN Success" and a sub-headline "July 13th, 2011 by jvanrej".

<http://www.openflow.org/>



The screenshot shows the Open Networking Foundation website homepage. At the top is the Open Networking Foundation logo, which consists of a blue circle with a white arrow pointing right, followed by the text "OpenNetworkingFoundation". Below the logo is a dark blue banner with the text "Welcome to the Open Networking Foundation!" and a paragraph describing the foundation's mission. To the right of the banner is a photograph of network cables plugged into a switch.

光纖卡

伺服器

機上盒

屠家照護

儲存器

交換器

準系統

手機

物聯網

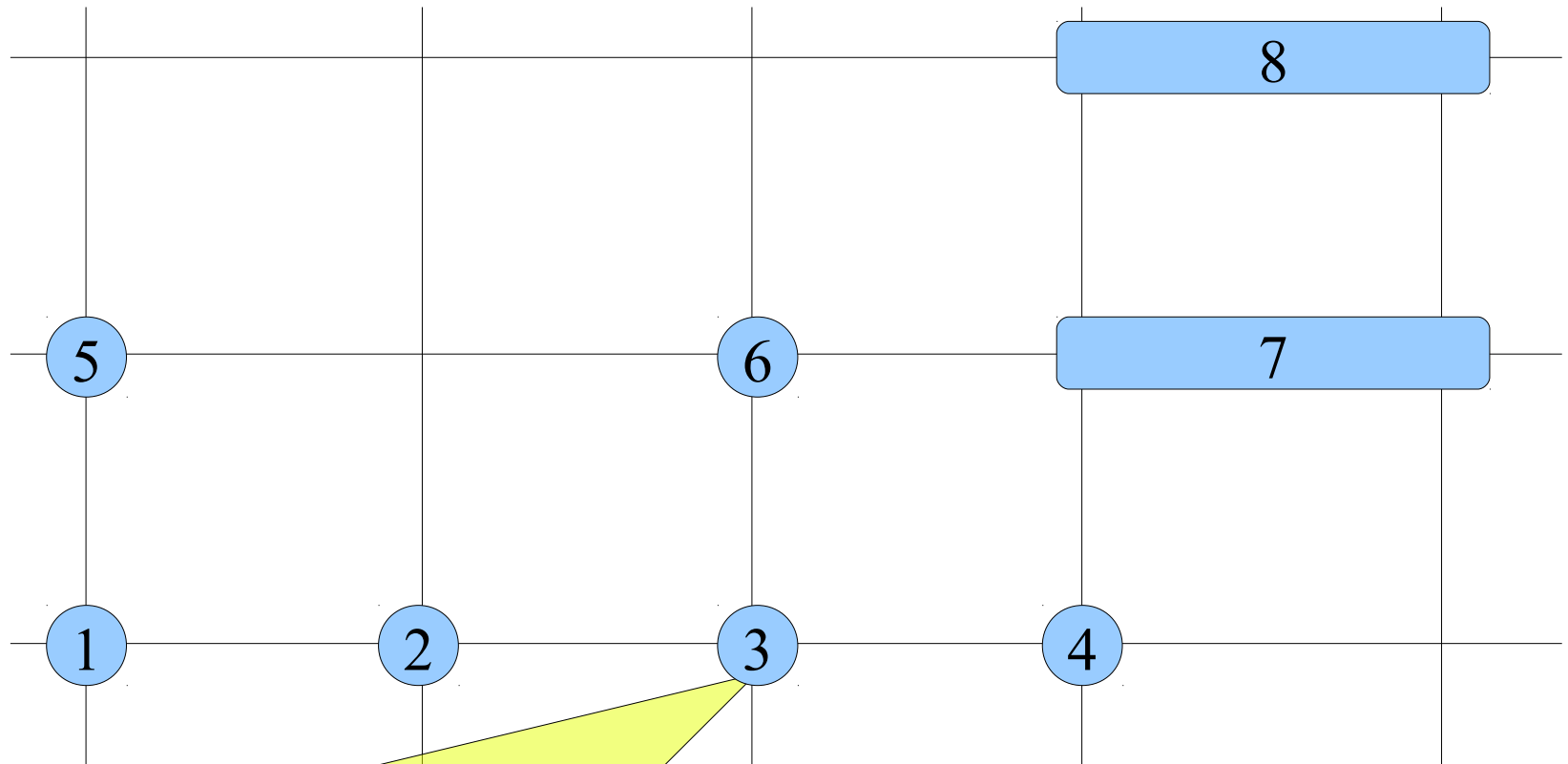
路由

主機板

Web 2.0
網頁服務

Big Data
資料分析

Virtualization
虛擬化技術

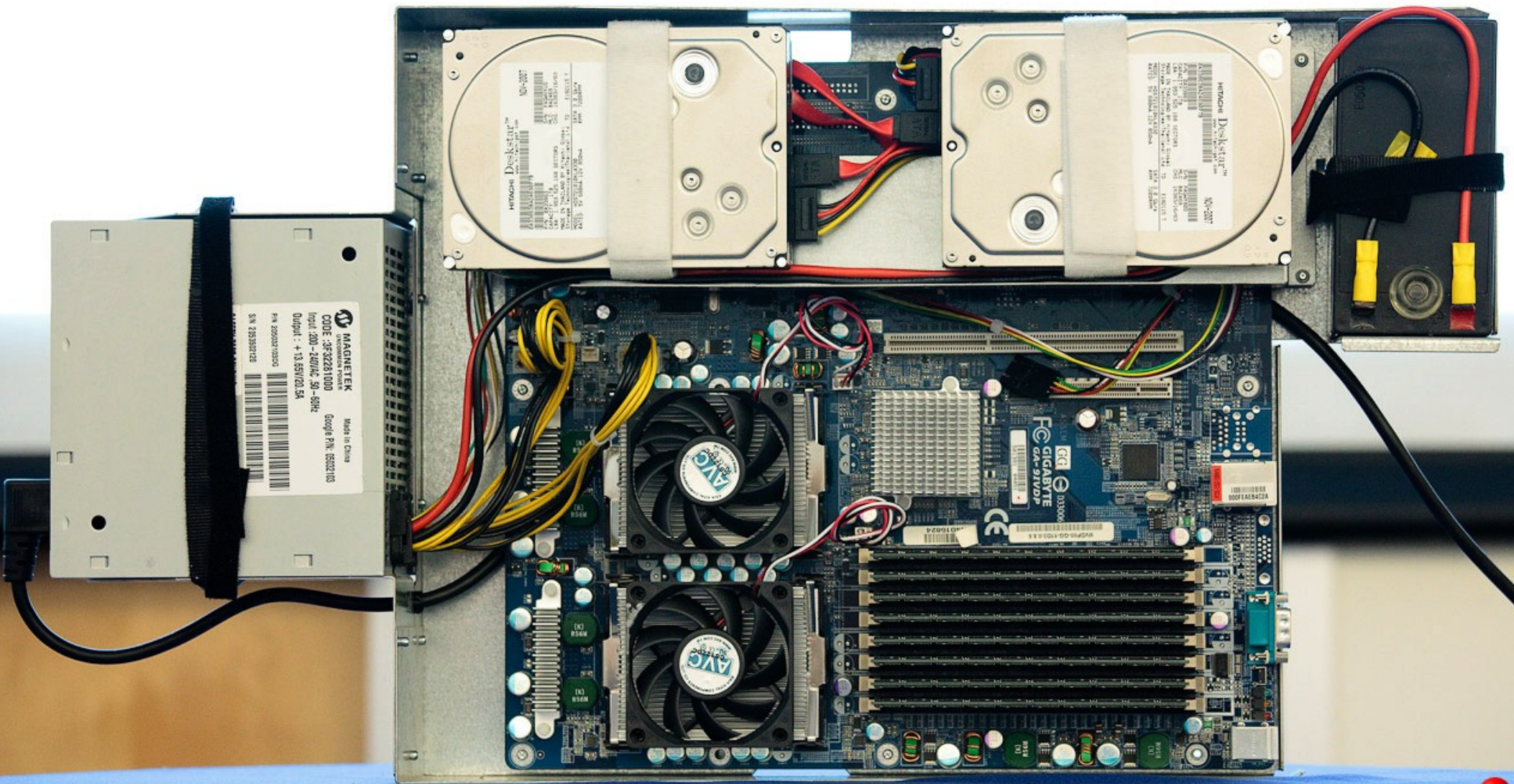


趨勢 (3) 爲了建立資料中心，主機板大量出貨！

建議 (3) 設計支援虛擬化、省電、高工作溫度的主機板

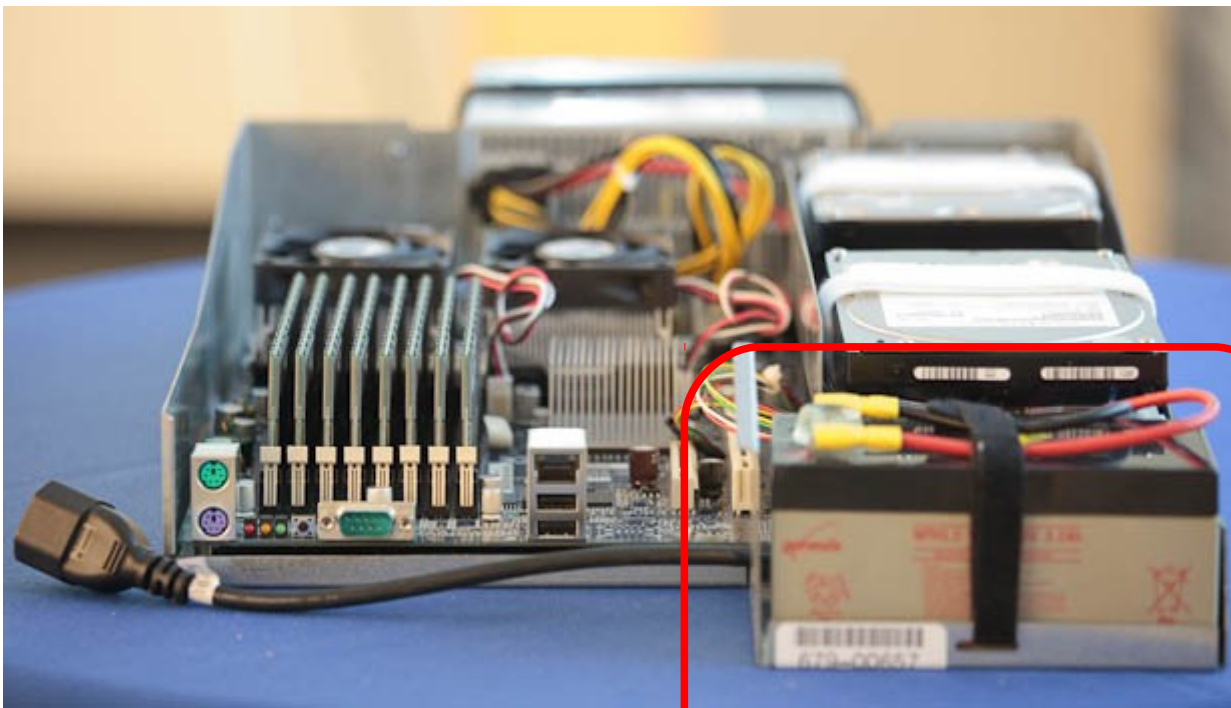
行動 (3) 評估 Open Compute Project

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，現任台達電)

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 (爲了 Big Data 運算特性)
3. 記憶體插滿 (這依舊是爲了 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)

Open Compute Project of Facebook 臉書的開放運算計畫



OPEN
Compute Project

SEARCH



f Log in

Specs & Designs

View on Github

Open

Featured Projects

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



Intel Motherboard

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



AMD Motherboard

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



Power Supply

The Open Compute Project 450W power converter, single voltage 12.5VDC, class supply used in high-efficiency applications includes...



Chassis

The Open Compute Project chassis is the other components in a server, including 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



7 人說讚。快免費註冊來查看你的朋友對什麼說讚。



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為了相關的協定，舉。以往科技公由，根據不同的了一大型的科技Google、facebook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.bnnext.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

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

光纖卡

伺服器

機上盒

屠家照護

儲存器

交換器

準系統

手機

物聯網

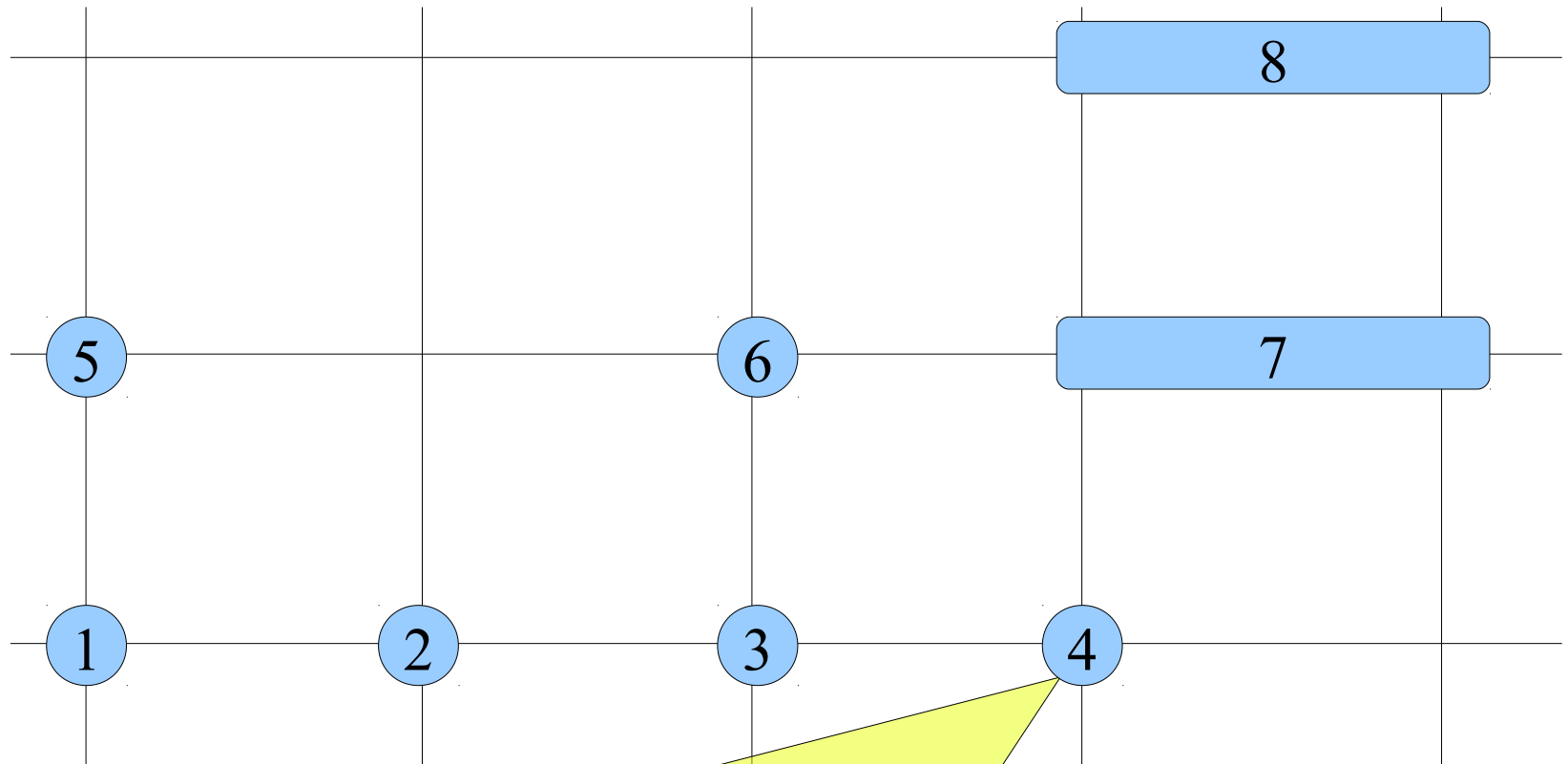
路由

主機板

Web 2.0
網頁服務

Big Data
資料分析

Virtualization
虛擬化技術



趨勢 (4) 嵌入式裝置進入多核心時代！

建議 (4) 選擇應用軟體多的嵌入式作業系統！

行動 (4) 評估導入嵌入式虛擬化技術！

Virtualization for Embedded Device ?

虛擬化只用在「雲」嗎？那「端」呢？



Motivation for Embedded Virtualization

嵌入式軟體採用虛擬化技術的三大動機

多核心
Multiple
Cores

co-existence of different OS environment
on the same platform

讓同一個硬體平台上共存不同的作業系統

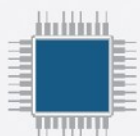
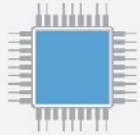
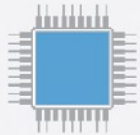
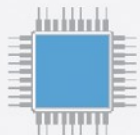
isolating critical components

from an untrusted OS environment

在不受信任的作業系統上建立獨立的管制區

an indirection level for remote control
of OS environments

為作業系統建立一個間接的遠端遙控機制



Real Time OS

General Purpose OS

參考來源：Embedded systems virtualization: Consider a Hypervisor

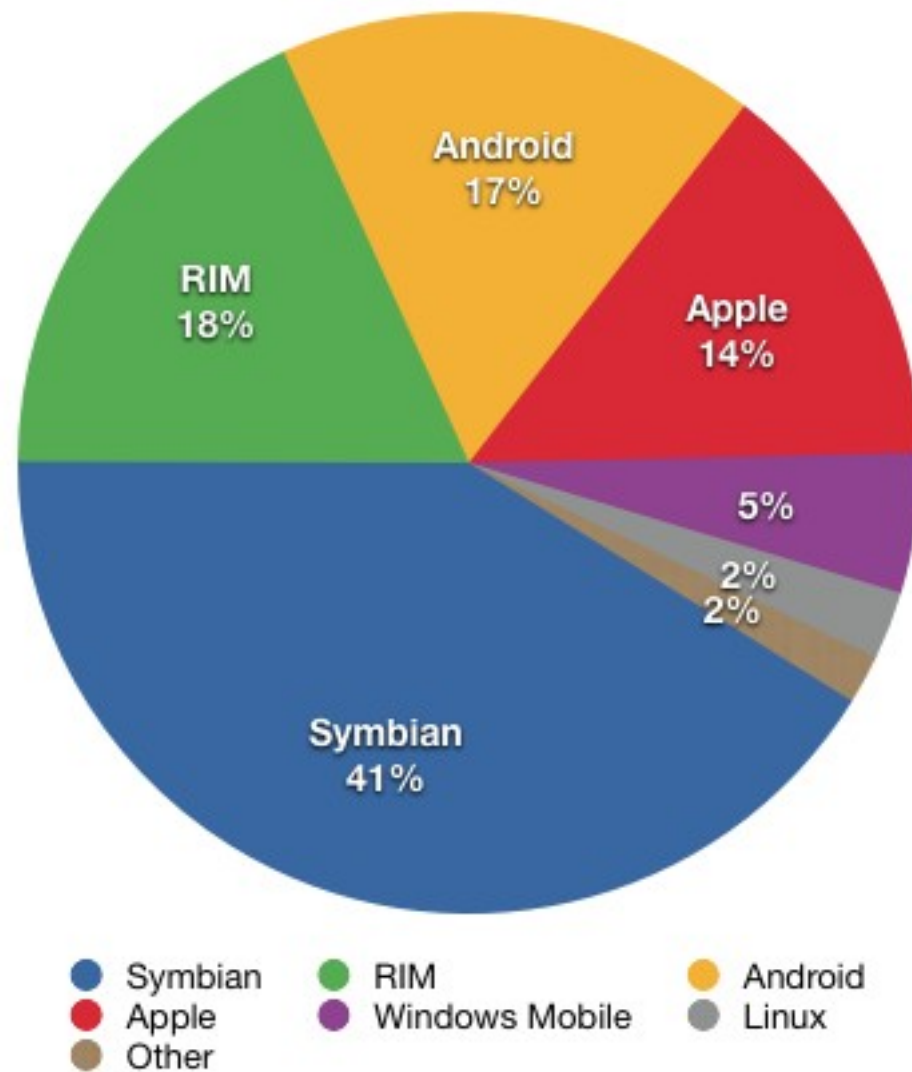
<http://www.eetimes.com/design/automotive-design/4016811/Embedded-systems-virtualization-Consider-a-Hypervisor>

參考來源：10 questions to ask when choosing a virtualization solution

<http://www.eetimes.com/design/eda-design/4006406/10-questions-to-ask-when-choosing-a-virtualization-solution>

Application : Key drivers of Multiple OS in one device

應用程式是在同一個裝置跑多作業系統的關鍵驅動力



圖片來源：http://www.iphonic.tv/iphone_screen_amaze.png

圖片來源：http://en.wikipedia.org/wiki/File:Smartphone_share_current.png

參考來源：iPhone - Playing game4iphone with a WiiMote

<http://www.youtube.com/watch?v=2AqgIWPnr1c>

參考來源：Android running on iPhone!

<http://linuxoniphone.blogspot.com/2010/04/ive-been-working-on-this-quietly-in.html>

<http://www.youtube.com/watch?v=5yO2KQHkt4A>

參考來源：Android on Motorola Rokr E6

<http://www.youtube.com/watch?v=-QAWd4ljV3g>

參考來源：Dual Boot Windows Mobile and Android

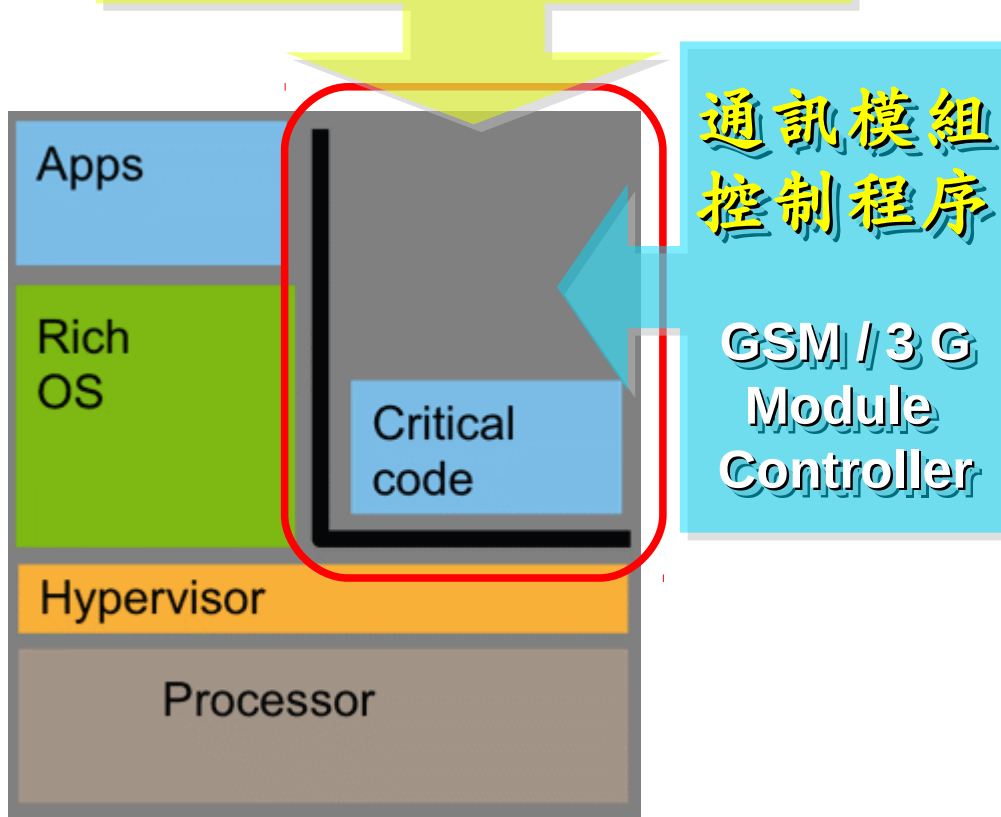
<http://www.redmondpie.com/how-to-dual-boot-windows-mobile-and-android-on-windows-phone-9140407/>

<http://www.youtube.com/watch?v=Nvj4ObHmxCI>

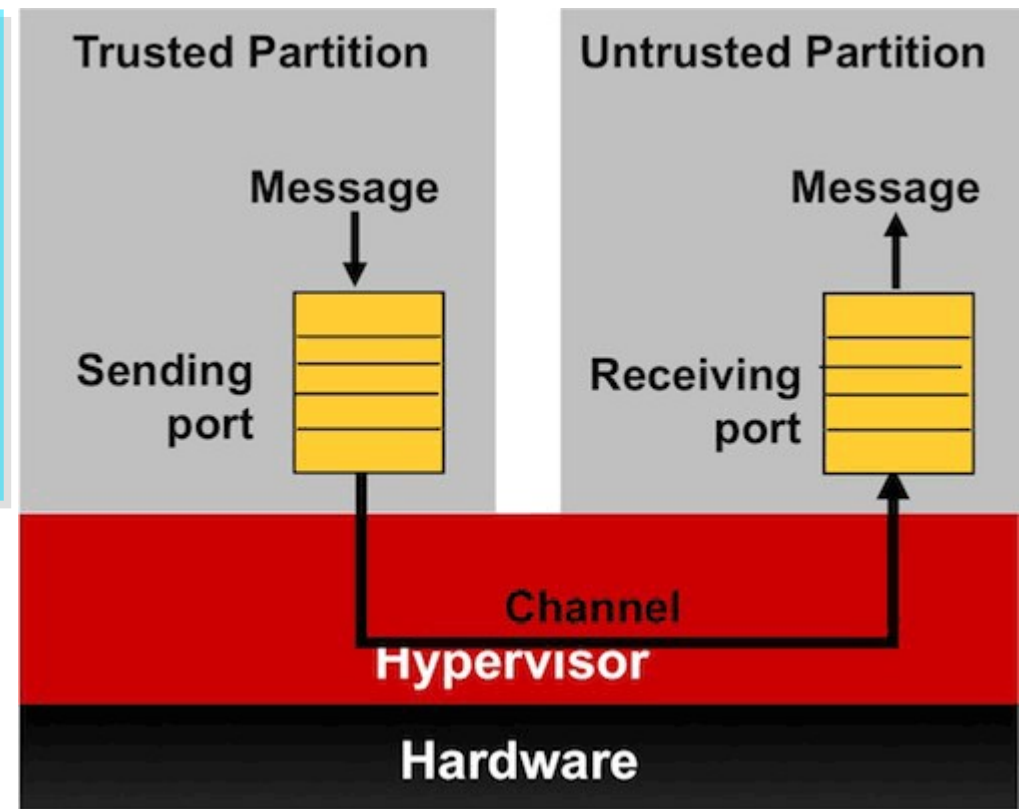
Virtualization for Embedded Security

爲了安全性，必須建立獨立的管制區

DRM 音樂授權
Digital Rights Management



區域間用訊息傳遞方式溝通
Communicate with Message Channel



參考來源：Embedded systems virtualization: Consider a Hypervisor

<http://www.eetimes.com/design/automotive-design/4016811/Embedded-systems-virtualization-Consider-a-Hypervisor>

參考來源：Securing Smart Grid Devices - Using Virtualization to Protect the Grid

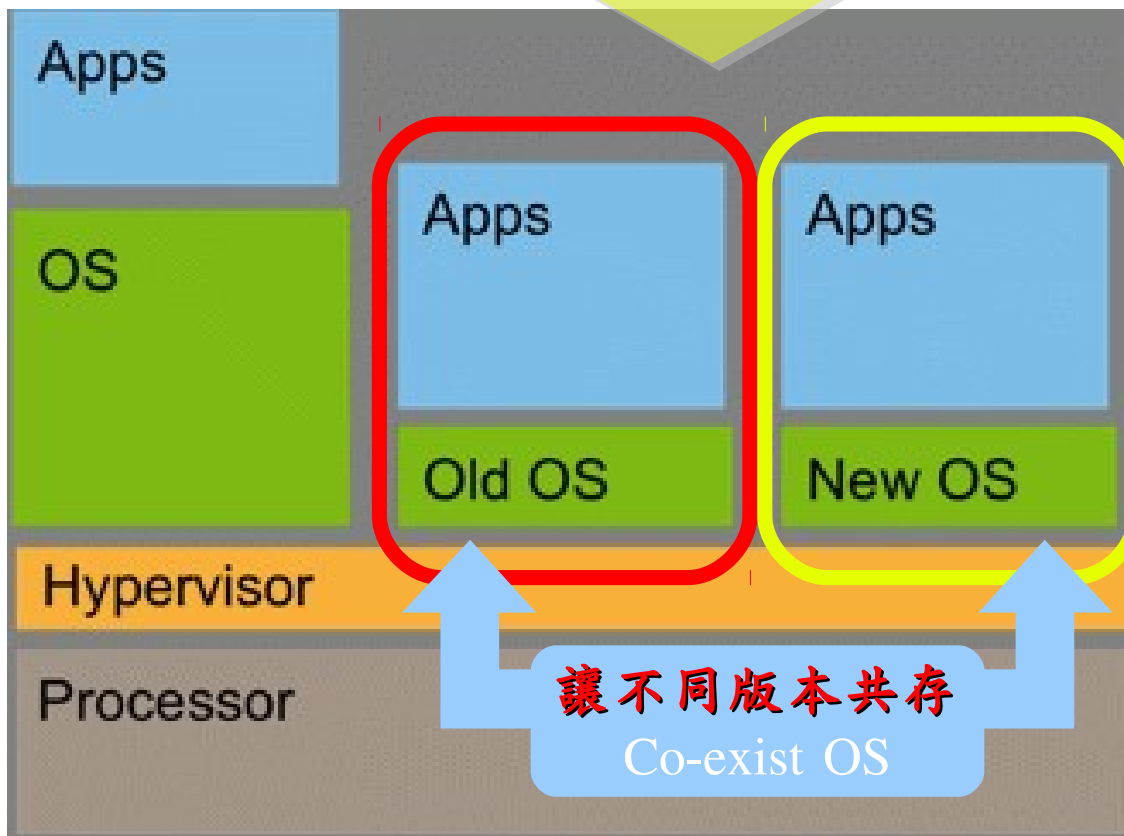
<http://embeddedinnovator.com/securing-smart-grid-devices>

Upgrade and Data Security : Key drivers of remote control

有助於作業系統升級並保護手機被偷時的資料安全

FOTA 韌體升級管制
firmware-over-the-air

遠距資料清除
Remote control to wipe data



輸入 *#06# 查 IMEI

Open Source for Embedded Virtualization

嵌入式虛擬化技術的自由軟體

Embedded XEN
on ARM platforms
<http://embeddedxen.sf.net>

sourceforge FIND AND DEVELOP OPEN SOURCE SOFTWARE

[Find Software](#) [Develop](#) [Create Project](#) [Blog](#) [Site Support](#) [About](#)

[SourceForge.net](#) > [Find Software](#) > Embedded XEN on ARM platforms

 **Embedded XEN on ARM platforms** Alpha by rossierd

[Summary](#) [Files](#) [Support](#) [Develop](#)

Requirement	KVM	Xen	Embedded Hypervisor
Embedded Processor Support	no	yes	yes
Real-time	no	no	yes
Fast communication	no	no	yes
Footprint	>10MB	16MB	<64kB
Code size	100s kLoC	100s kLoC	10 kLoC

image with bedded

EDIT

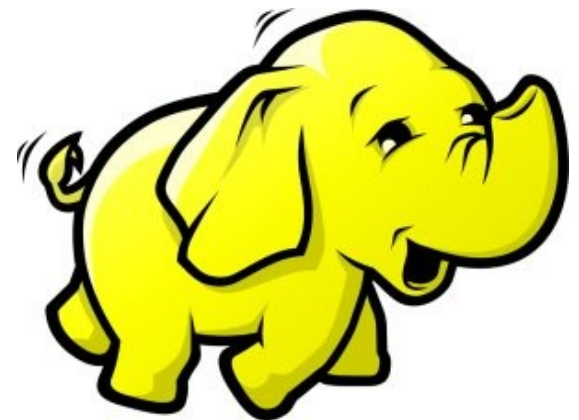
[v project details](#)



平台即服務：資料分析技術

Introduction to Big Data and related Technologies

Jazz Wang
Yao-Tsung Wang
jazz@nchc.org.tw



WHAT



What is Big Data ?

何謂海量資料

趨勢

Trends

定義

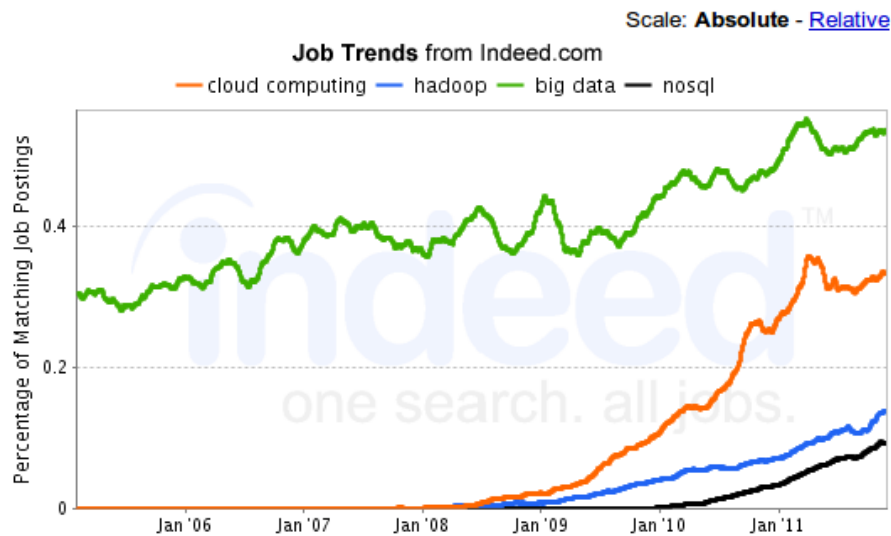
Definitions

挑戰：管理維度

The Six Dimensions

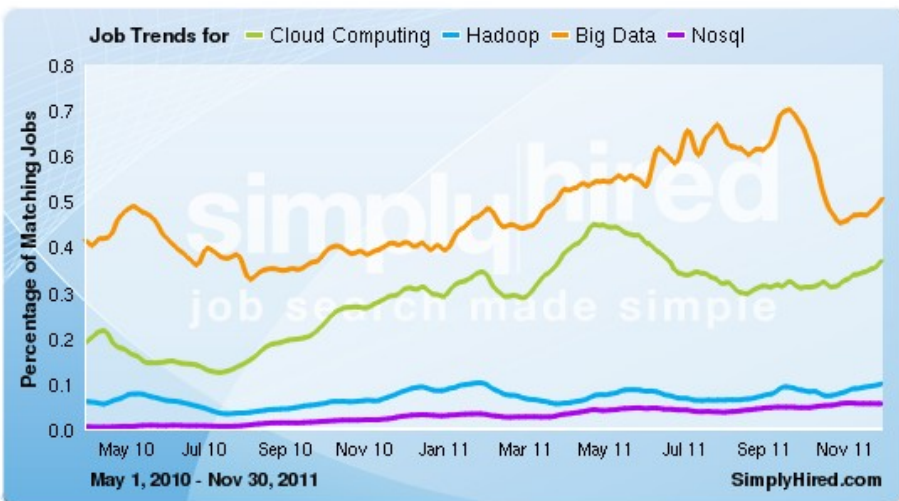
Trends of Market Needs 市場需求趨勢

cloud computing, hadoop, big data, nosql Job Trends



Indeed.com searches millions of jobs from thousands of job sites. This job trends graph shows the percentage of jobs we find that contain your search terms.

Find [Cloud Computing jobs](#), [Hadoop jobs](#), [Big Data jobs](#), [Nosql jobs](#)



美國軟體就業市場分析，根據 indeed 與 simply hired 兩間公司的趨勢觀察，都得到一樣的結果：

Big Data > Cloud Computing > Hadoop > NoSQL

To

CIO technologies	Ranking of technologies CIOs selected as one of their top 3 priorities in 2012			
Ranking	2012	2011	2010	2009
Analytics and business intelligence	1	5	5	1
Mobile technologies	2	3	6	12
Cloud computing (SaaS, IaaS, PaaS)	3	1	2	16
Collaboration technologies (workflow)	4	8	11	5
Virtualization	5	2	1	3
Legacy modernization	6	7	15	4
IT management	7	4	10	*
Customer relationship management	8	18	*	*
ERP applications	9	13	14	2
Security	10	12	9	8
Social media/Web 2.0	11	10	3	15

Gartner CIO Agenda 2012 前三名：
 [1] Business Intelligence (Big Data)
 [2] Mobile technology
 [3] Cloud Computing

How BIG? 讓我們先來認識一下容量單位

Bit (b)	1 or 0
Byte (B)	8 bits
Kilobyte (KB)	1,000 bytes
Megabyte (MB)	1,000 KB
Gigabyte (GB)	1,000 MB
Terabyte (TB)	1,000, GB
Petabyte (PB)	1,000 TB
Exabyte (EB)	1,000 PB
Zettabyte (ZB)	1,000 EB

Data Explosion!! 始於 2007 的「資料大爆炸」時代

Information Versus Available Storage

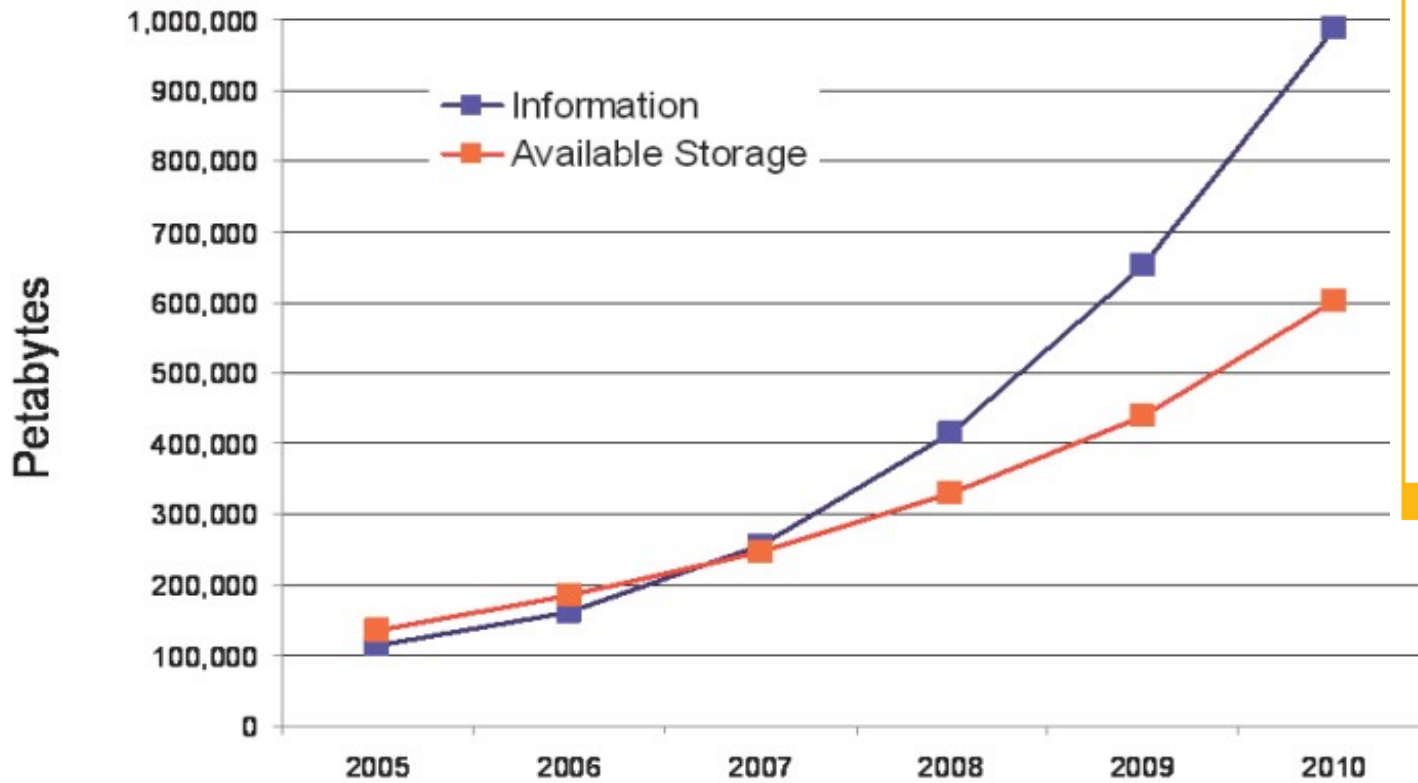
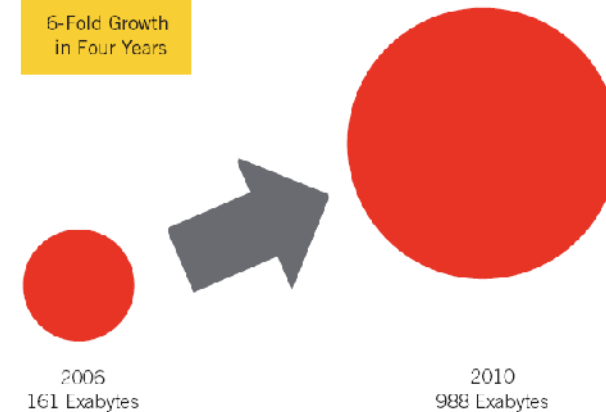


Figure 1

Information Created, Captured and Replicated

6-Fold Growth
in Four Years



Source: IDC, 2007

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

Source: IDC, 2007

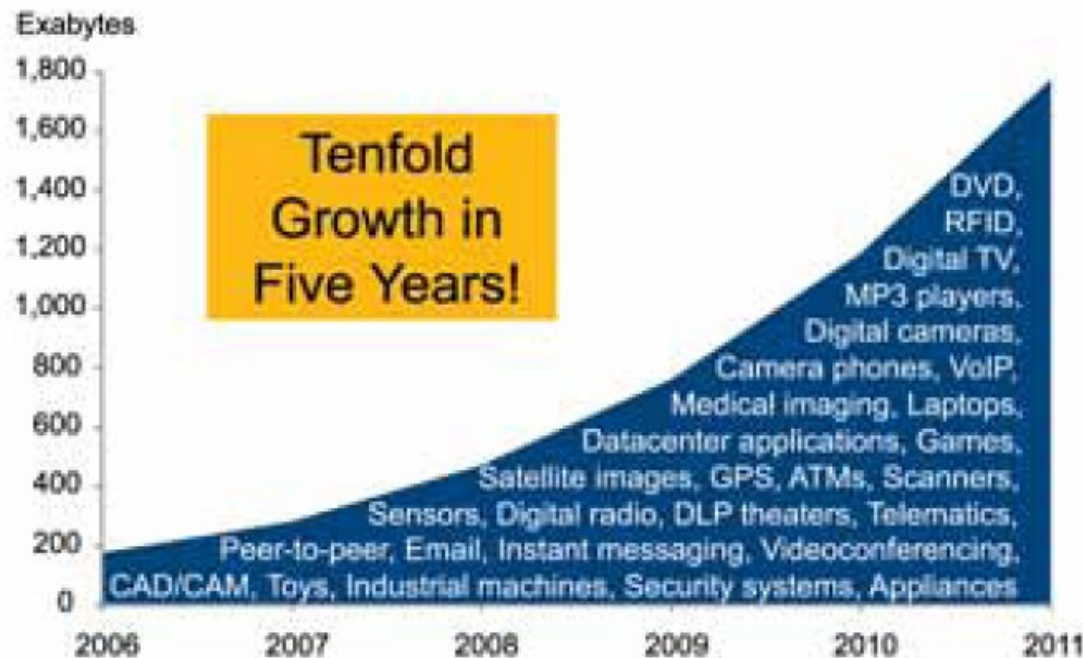
出處：The Expanding Digital Universe,
A Forecast of Worldwide Information Growth Through 2010,
March 2007, An IDC White Paper - sponsored by EMC
<http://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf>

2006 161 EB
2010 988 EB (預測)

Data Explosion!! 始於 2007 的「資料大爆炸」時代

Figure 1

Digital Information Created, Captured, Replicated Worldwide



Source: IDC, 2008

2009 年，IDC 預估
2011 年會成長**十倍**！
(相較 2006 年)

2006	161	EB
2007	281	EB
2010	988	EB (預測)
2011	1773	EB (預測)

出處：[The Diverse and Exploding Digital Universe, An Updated Forecast of Worldwide Information Growth Through 2011](http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf)
March 2008, An IDC White Paper - [sponsored by EMC](http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf)
<http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf>

Data expanded 1.6x each year !! 每年約略 1.6 倍



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	1200	EB	(1.2 ZB)
2011	1773	EB	(預測)
2011	1800	EB	(1.8 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>

What is Big Data?! 何謂『海量資料』？

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]

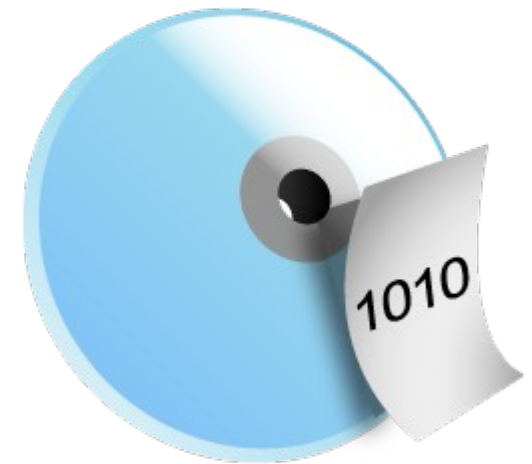
出處：http://en.wikipedia.org/wiki/Big_data



多個檔案，容量 100TB



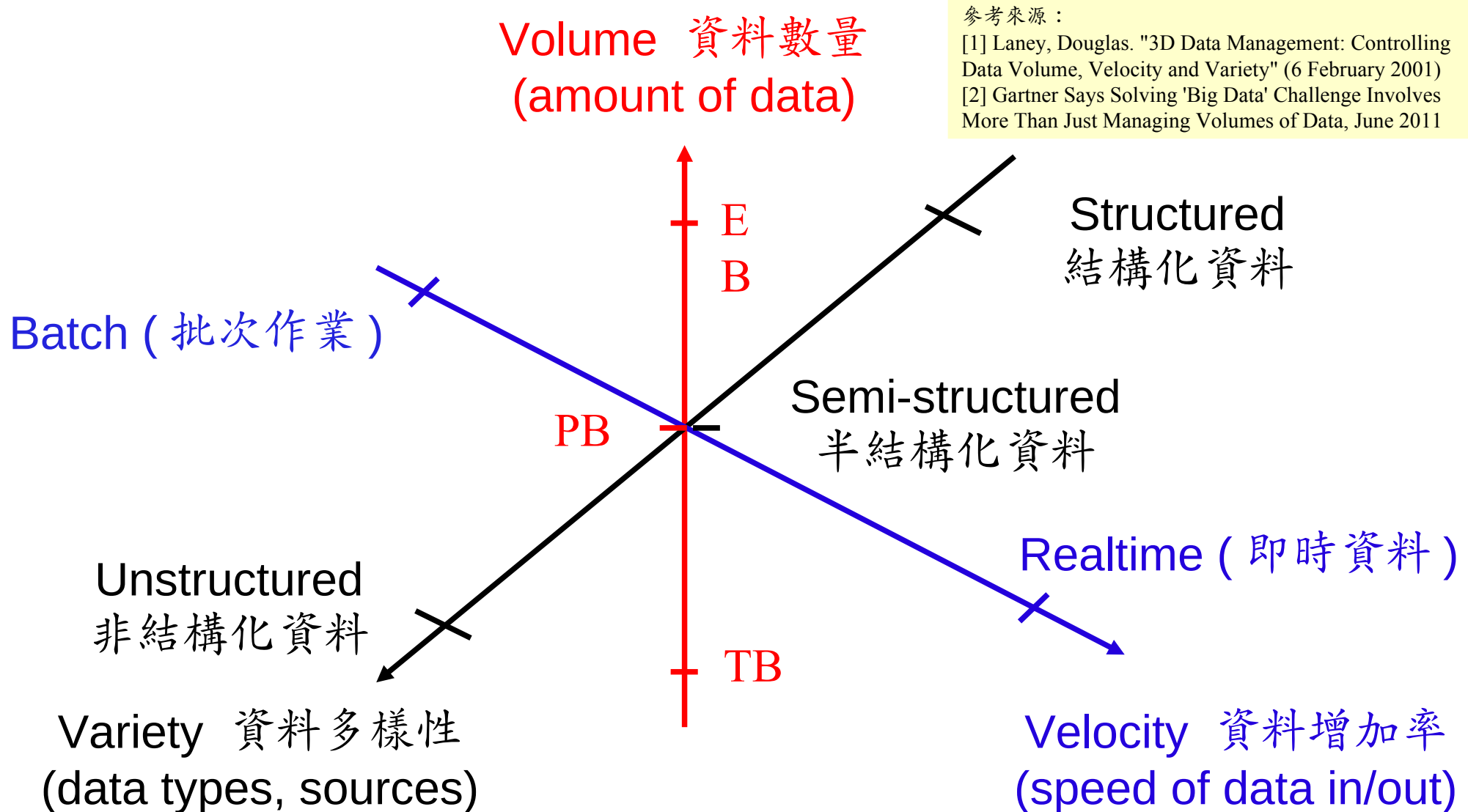
一個資料庫，容量 100TB



一個檔案，容量 100TB

Gartner Big Data Model ? 海量資料的模型 ?

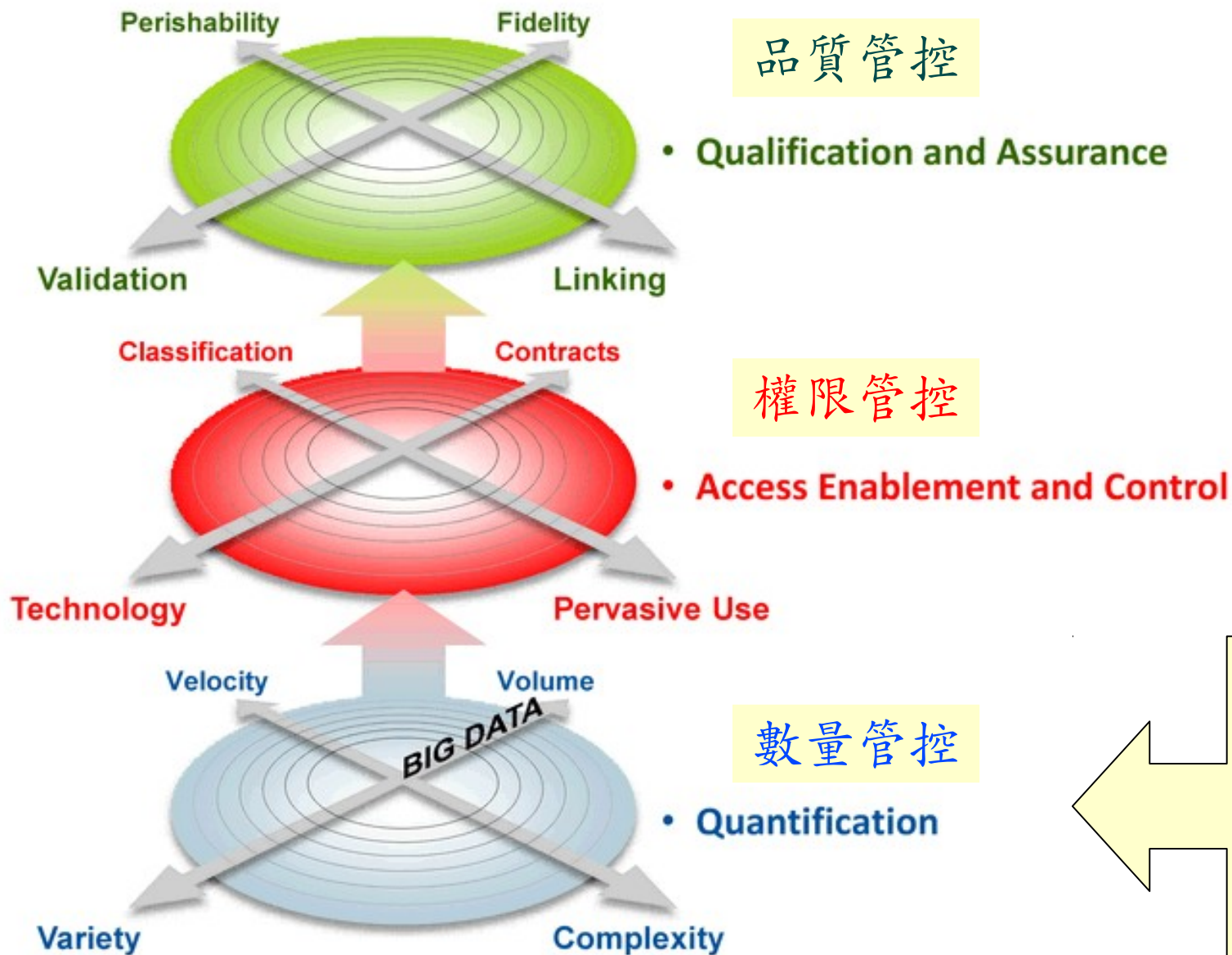
海量資料的挑戰在於如何管理「數量」、「增加率」與「多樣性」



Six Dimensions of Big Data? 六個維度？



12D of Information Management? 12 個維度？



Big Data
只是終極
資訊管理
的開端！

Source: Gartner (March 2011), 'Big Data' Is Only the Beginning of Extreme Information Management, 7 April 2011, <http://www.gartner.com/id=1622715>

What is Big Data ?

何謂海量資料

Why should we care?

為何需要關切

資料 Data

知識 Knowledge

智慧 Wisdom

WHY



花精靈-小葵

Why we call it “ SMART ” !!
智慧打哪兒來？！

Smart Phone
智慧手機

Smart Car
智慧車輛

Smart Grid
智慧電網

SMART
哪裡長
智慧了？

Smart City
智慧城市

Smart Home
智慧家庭

Smart Meter
智慧電錶

資料 Data

知識 Knowledge

智慧 Wisdom

Can Machine understand You? 讓機器更懂你?

iPhone

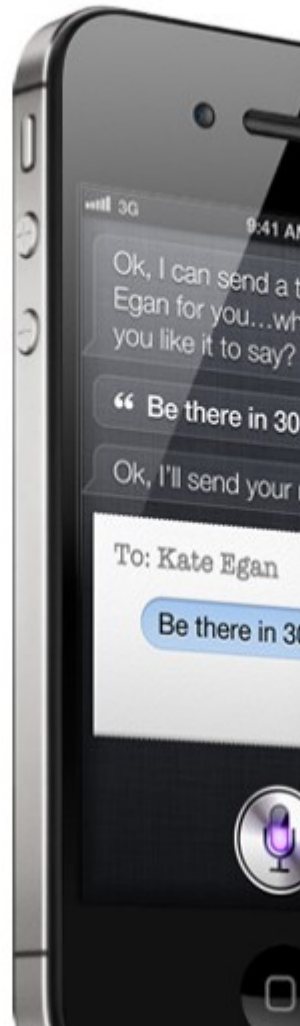
Features Built-in Apps



Siri. Beta

Your wish is
its command.

Siri on iPhone 4S lets you use your voice to send messages, schedule meetings, place phone calls, and more. Ask Siri to do things just by talking the way you talk. Siri understands what you say, knows what you mean, and even talks back. Siri is so easy to use and does so much, you'll keep finding more and more ways to use it.



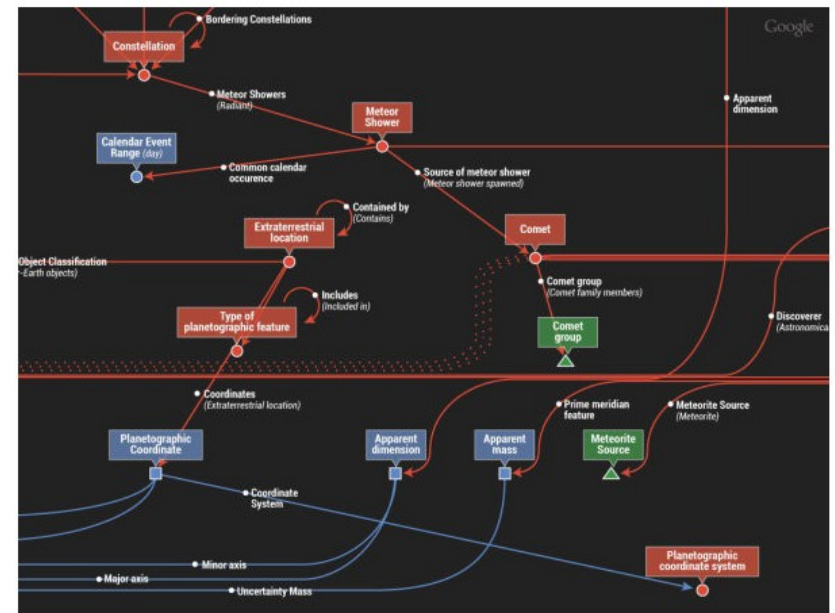
Google將發展「人工智慧」 永久改變搜尋引擎

2012年02月15日 00:11

點評：超級阿斯拉，衝啊！（阿斯拉：好的，準人！）

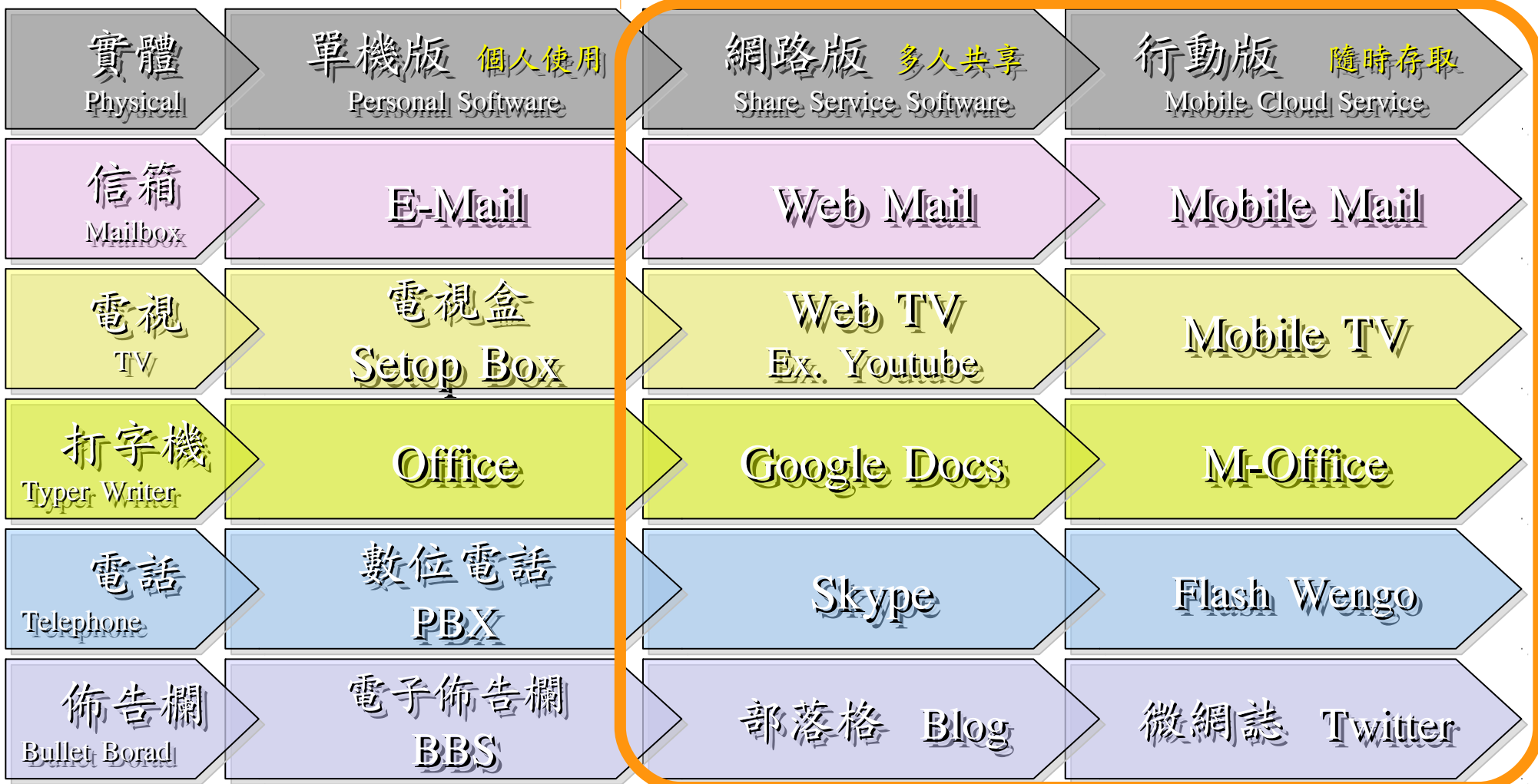
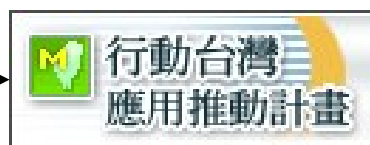
記者黃郁棋 / 綜合報導

每個人都再猜，下一波網路革命是什麼？每個人都再猜，未來的世界會如何運作？Google的資深副總Amit Singhai透露了一點訊息。「Google正努力從『單字』層面進展到『意義』層面，未來搜尋引擎提供的不只是關鍵字搜尋，搜尋引擎甚至會『明白』你到底要什麼。」



▲ Google未來將會朝「人工智慧」前進。(圖 / 取自mashable.com)

Evolution of Software / Service 軟體演化勢必走向『智能化』



How to handle it ?

三大因應策略

儲存虛擬化

Dedup.

資料安全

Security

智慧服務

SaaS

HOW



花精靈-圖兒

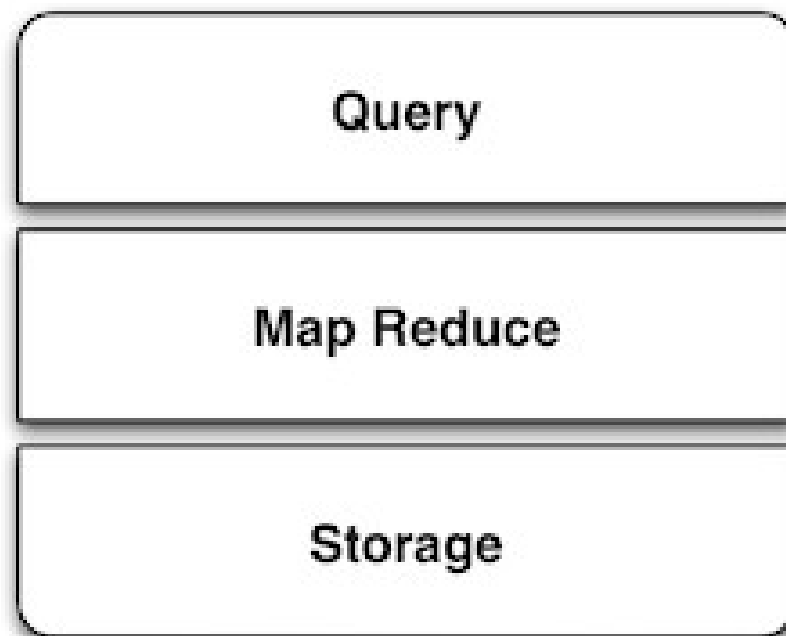
The SMAQ stack for big data

海量資料處理的資訊架構

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)

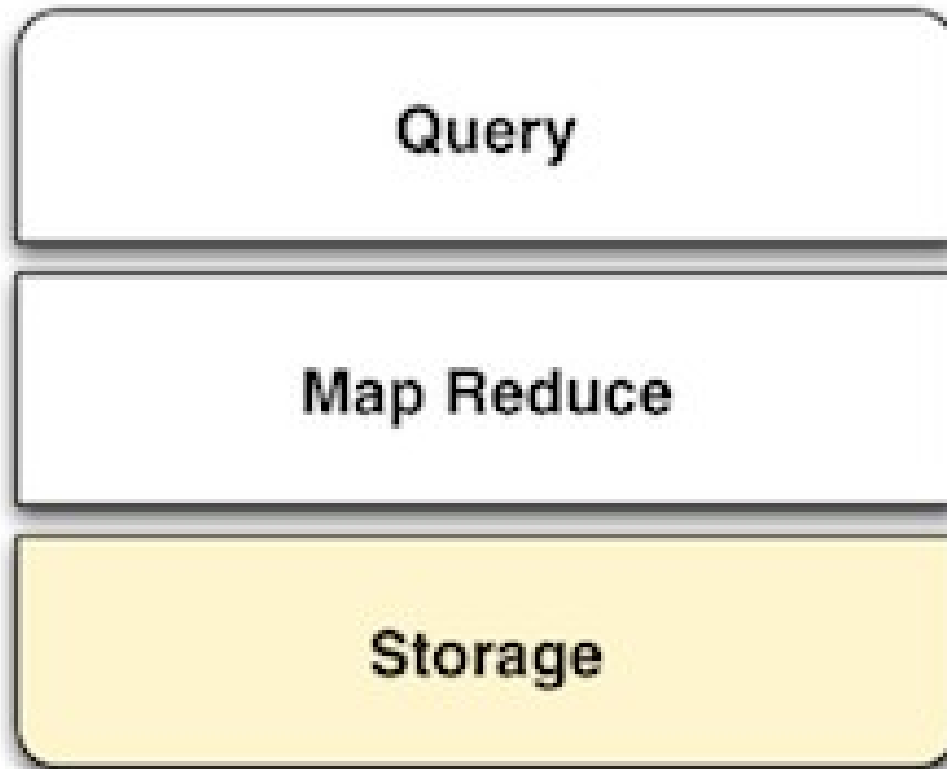


參考來源：The SMAQ stack for big data，Edd Dumbill，22 September 2010，
<http://radar.oreilly.com/2010/09/the-smaq-stack-for-big-data.html>

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

The SMAQ stack for big data

海量資料處理的資訊架構



用來儲存分散、沒有關聯
的非結構化資料

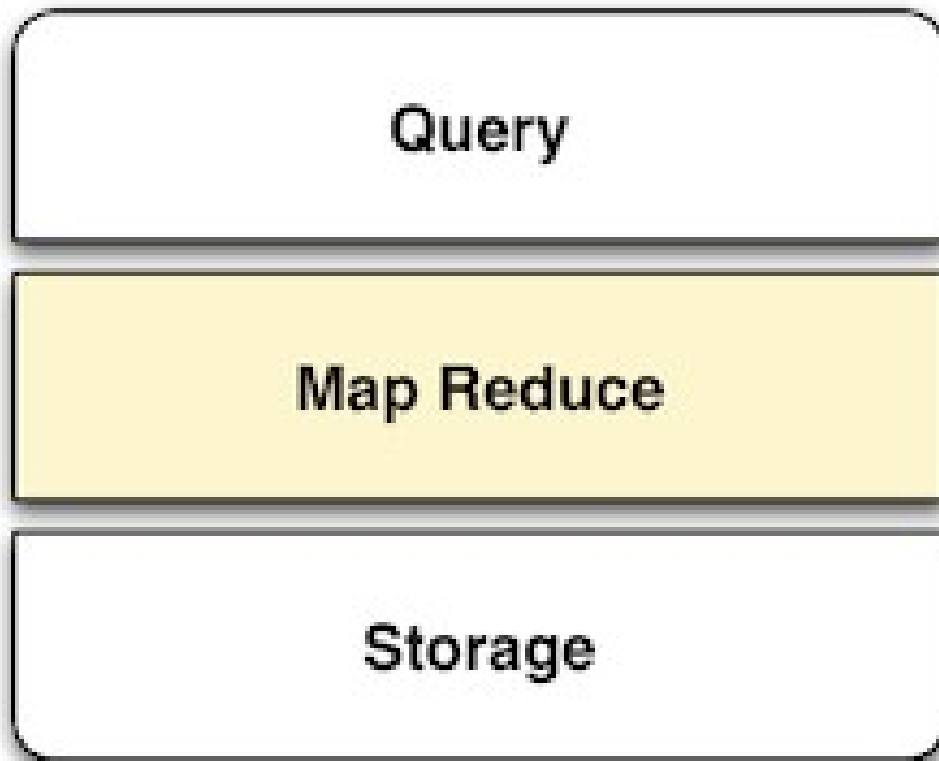
Key features

- Distributed
- Non-relational or unstructured

The SMAQ stack for big data

海量資料處理的資訊架構

運用批次處理的方式，將
運算工作平均分散到許多
的伺服器做運算。

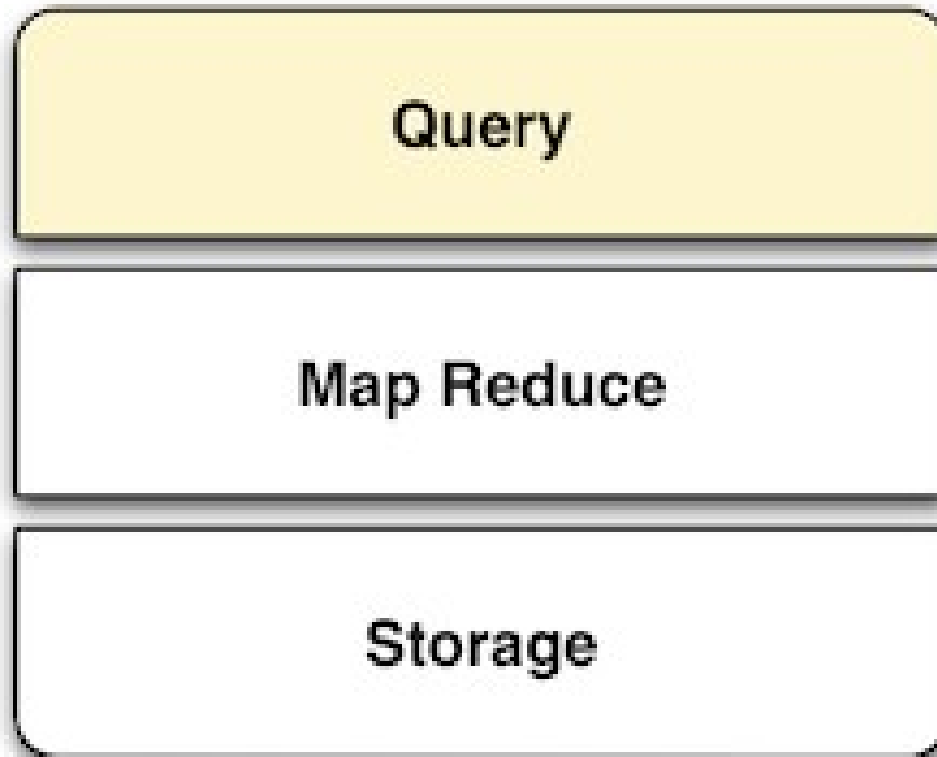


Key features

- Distributes computation over many servers
- Batch processing model

The SMAQ stack for big data

海量資料處理的資訊架構



Key features

- Efficient way of defining computation
- Platform for user friendly analytical systems

將算完的結構化資料儲存到可供查詢的資料庫系統

Three Core Technologies of Google

Google 的三大關鍵技術

- Google 在一些會議分享他們的三大關鍵技術
- Google shared their design of web-search engine
 - SOSP 2003 :
 - “The Google File System”
 - <http://labs.google.com/papers/gfs.html>
 - OSDI 2004 :
 - “MapReduce : Simplified Data Processing on Large Cluster”
 - <http://labs.google.com/papers/mapreduce.html>
 - OSDI 2006 :
 - “Bigtable: A Distributed Storage System for Structured Data”
 - <http://labs.google.com/papers/bigtable-osdi06.pdf>



Open Source Mapping of Google Core Technologies

Google 三大關鍵技術對應的自由軟體

BigTable

A huge key-value datastore

HBase, Hypertable
Cassandra,

MapReduce

To parallel process data

Hadoop MapReduce API
Sphere MapReduce API, ...

Google File System

To store petabytes of data

Hadoop Distributed File System (HDFS)
Sector Distributed File System

更多不同語言的 MapReduce API 實作：

<http://trac.nchc.org.tw/grid/intertrac/wiki%3Ajazz/09-04-14%23MapReduce>

其他值得觀察的分散式檔案系統：

- IBM GPFS - <http://www-03.ibm.com/systems/software/gpfs/>
- Lustre - <http://www.lustre.org/>
- Ceph - <http://ceph.newdream.net/>

光纖卡

伺服器

機上盒

居家照護

儲存器

交換器

準系統

手機

物聯網

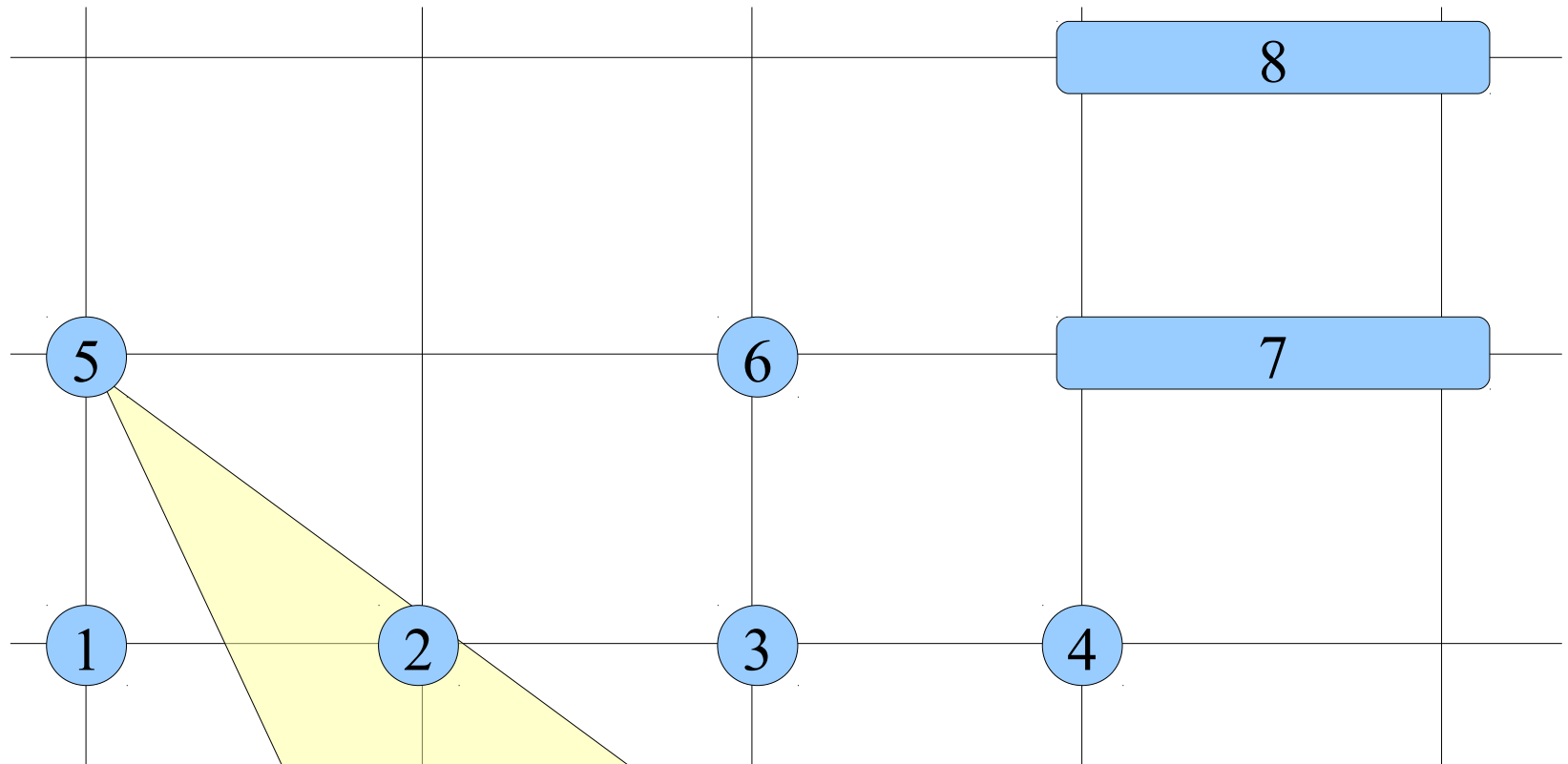
路由

主機板

Web 2.0
網頁服務

Big Data
資料分析

Virtualization
虛擬化技術



趨勢 (5) 建構資料分析架構，需要龐大的儲存設施！
 建議 (5) 思考如何從硬體架構增加 I/O
 行動 (5) 從 RAID、PCIe SSD 介面著手

電子工程專輯

欲打印此文章，從您的瀏覽器菜單中選擇“文件”後再選“打印”。

PCIe將改變固態硬碟市場競爭局面

上網時間:2012年04月11日

有一群廠商打算推出採用 PCI Express (PCIe)匯流排的固態硬碟(SSD)，號稱可改善目前採用 Serial ATA (SATA)或是Serial-Attached SCSI (SAS)介面的固態硬碟性能；新一波產品將延續SATA與SAS兩種協議分庭抗禮的局面，並讓市場競爭重點加速轉向新軟體功能。

去年，包括戴爾(Dell)、英特爾(Intel)、美光(Micron)、甲骨文(Oracle)與Stec等共80家廠商，組織了一個產業團體著手定義 NVMe (Non-Volatile Memory Express)介面標準，採用該標準的首批固態硬碟預計在今年稍晚上市。

<http://www.eettaiwan.com/articleLogin.do?artId=8800664853>

光纖卡

伺服器

機上盒

屠家照護

儲存器

交換器

準系統

手機

物聯網

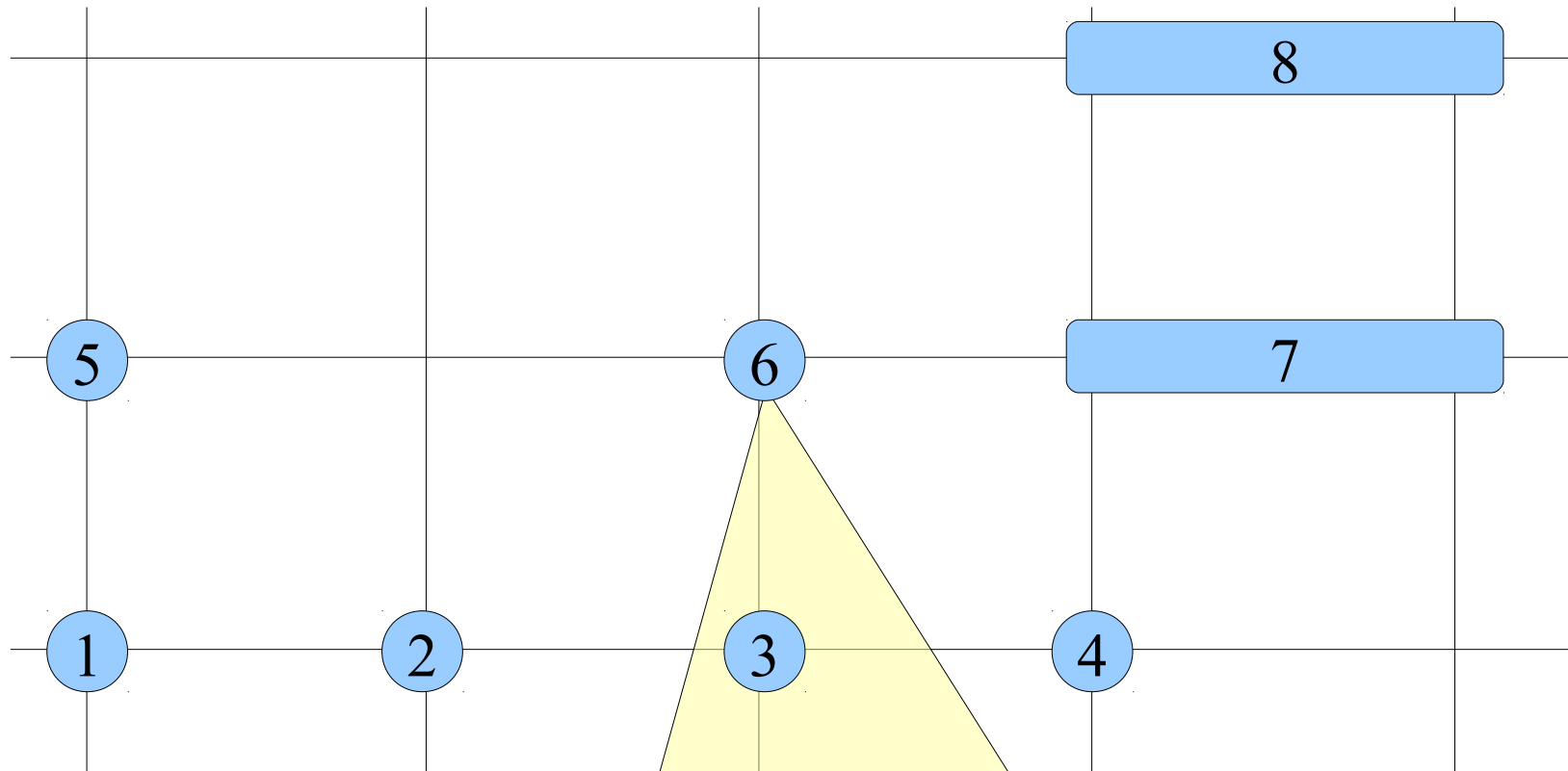
路由

主機板

Web 2.0
網頁服務

Big Data
資料分析

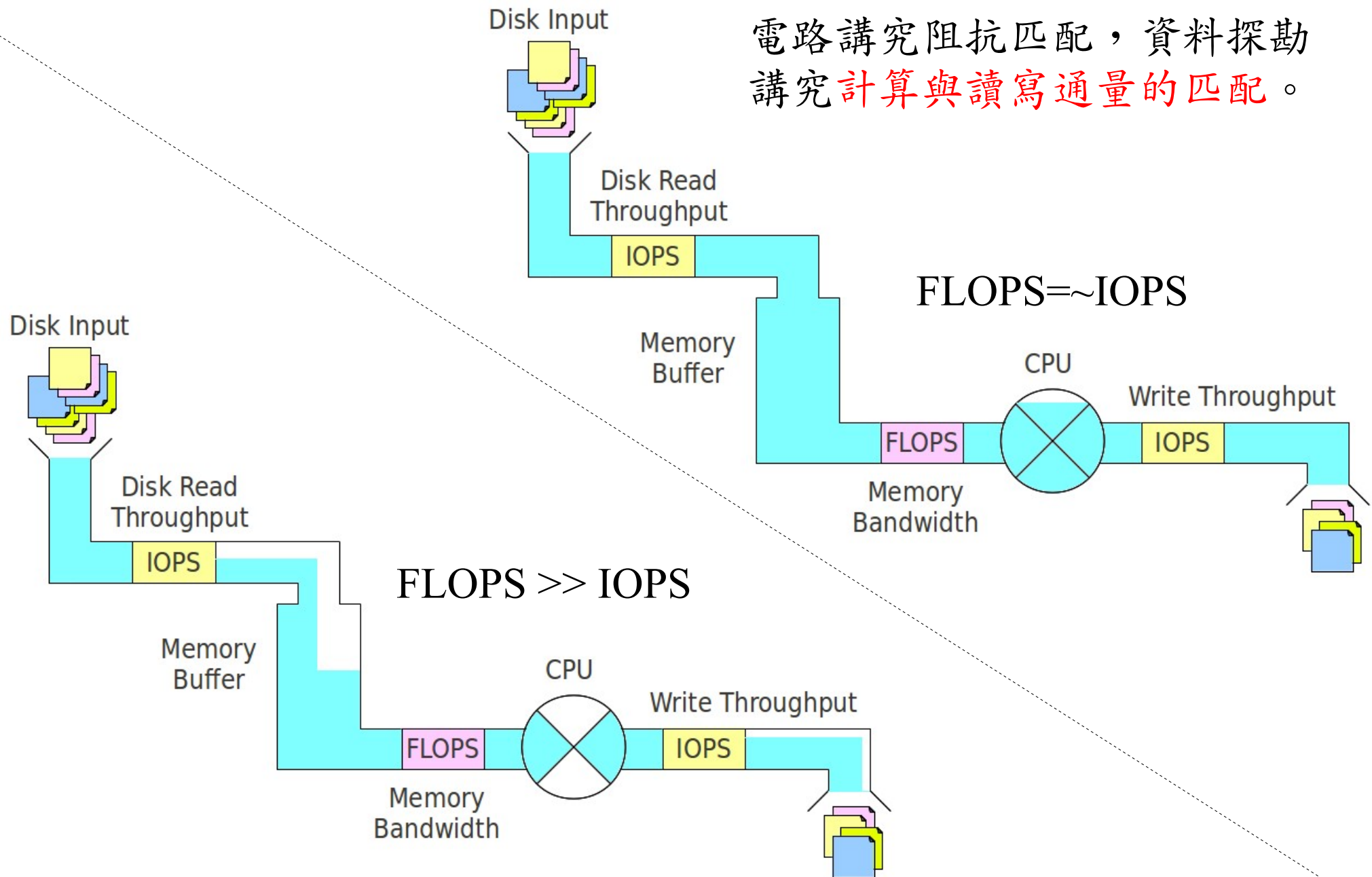
Virtualization
虛擬化技術



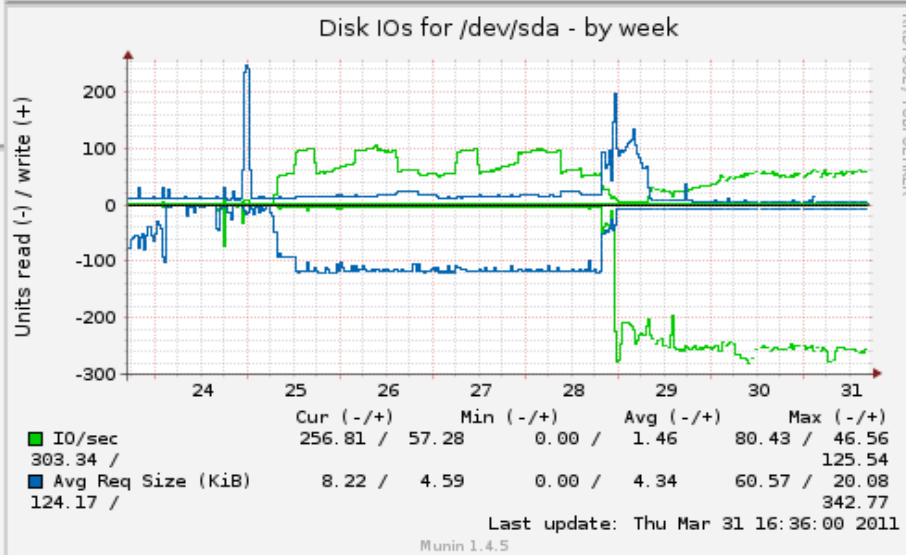
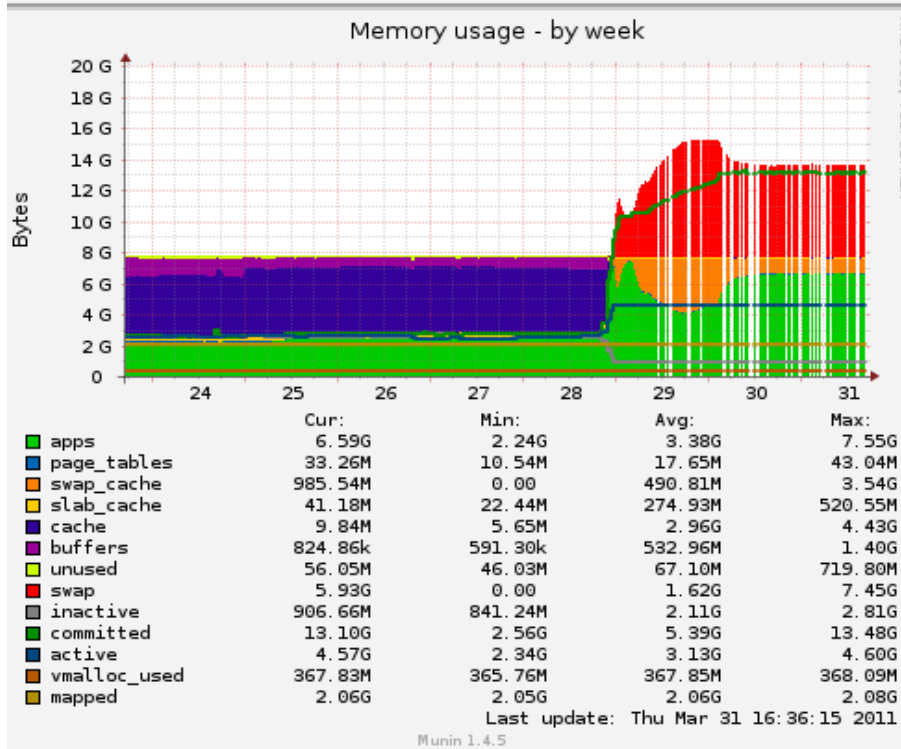
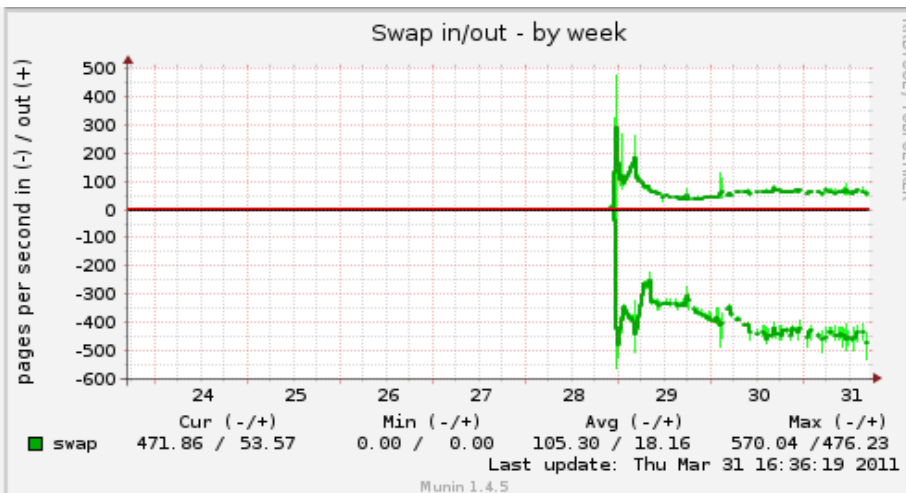
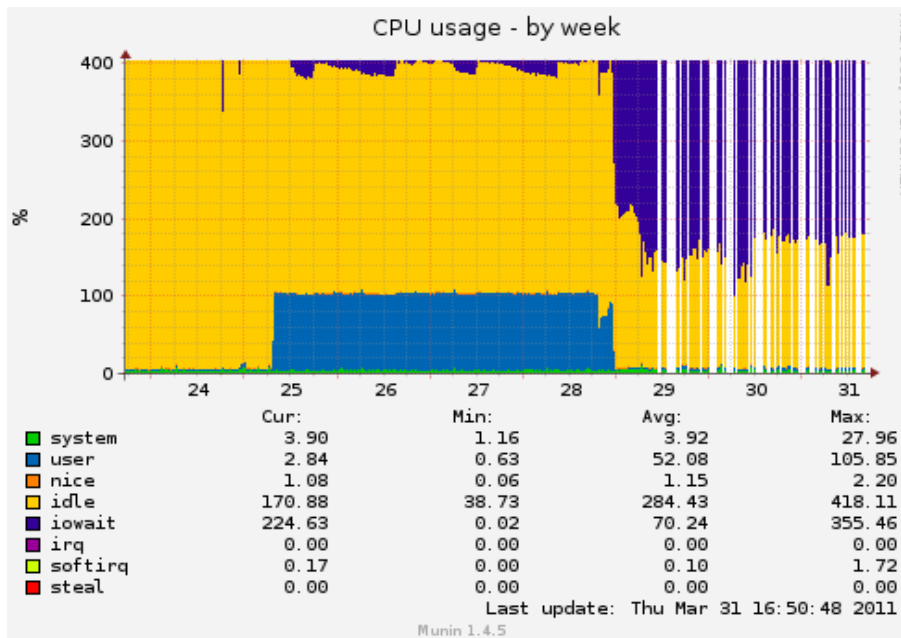
趨勢 (6) 建構資料分析架構，需要龐大的儲存設施！
 建議 (6) 思考硬體架構該如何匹配，才能增加 I/O
 行動 (6) 從南北橋晶片下手

I/O Impedance ? 資料通量達成匹配？

電路講究阻抗匹配，資料探勘講究計算與讀寫通量的匹配。



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)



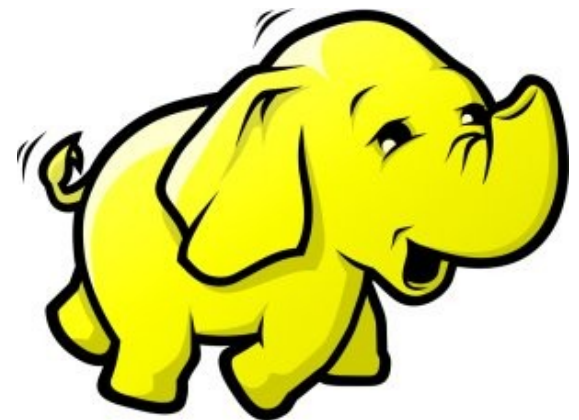
軟體即服務：網頁服務技術

Web 2.0, Distributed Databases and NoSQL

Jazz Wang

Yao-Tsung Wang

jazz@nchc.org.tw



New IT Architecture toward Cloud Computing !!

用雲掌握資料，加以分析，形成智能給端用



雲

資料中心
提供服務

雲端設計新思維：端的智能來自於雲的服務

Devices share the wisdom of Cloud

端



各類裝置
存取服務

4. 下載最佳路徑

1. 上傳行進軌跡

6. 上傳最佳路徑執行狀態
(判斷是否需要修正)

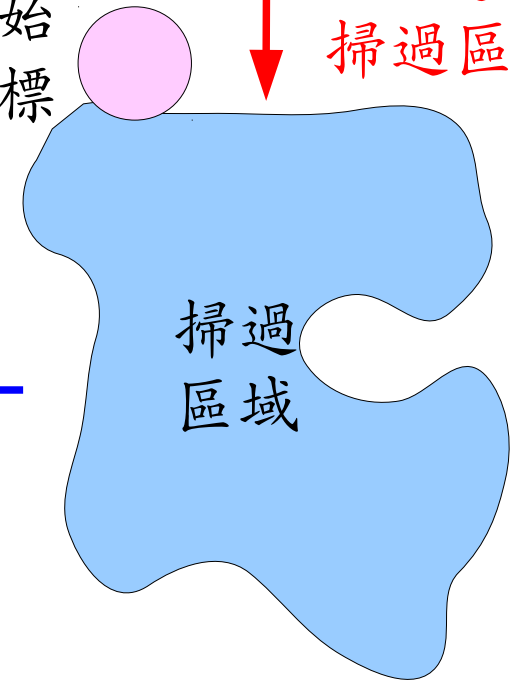
3. 運算
最佳路徑

5. 執行
最佳路徑



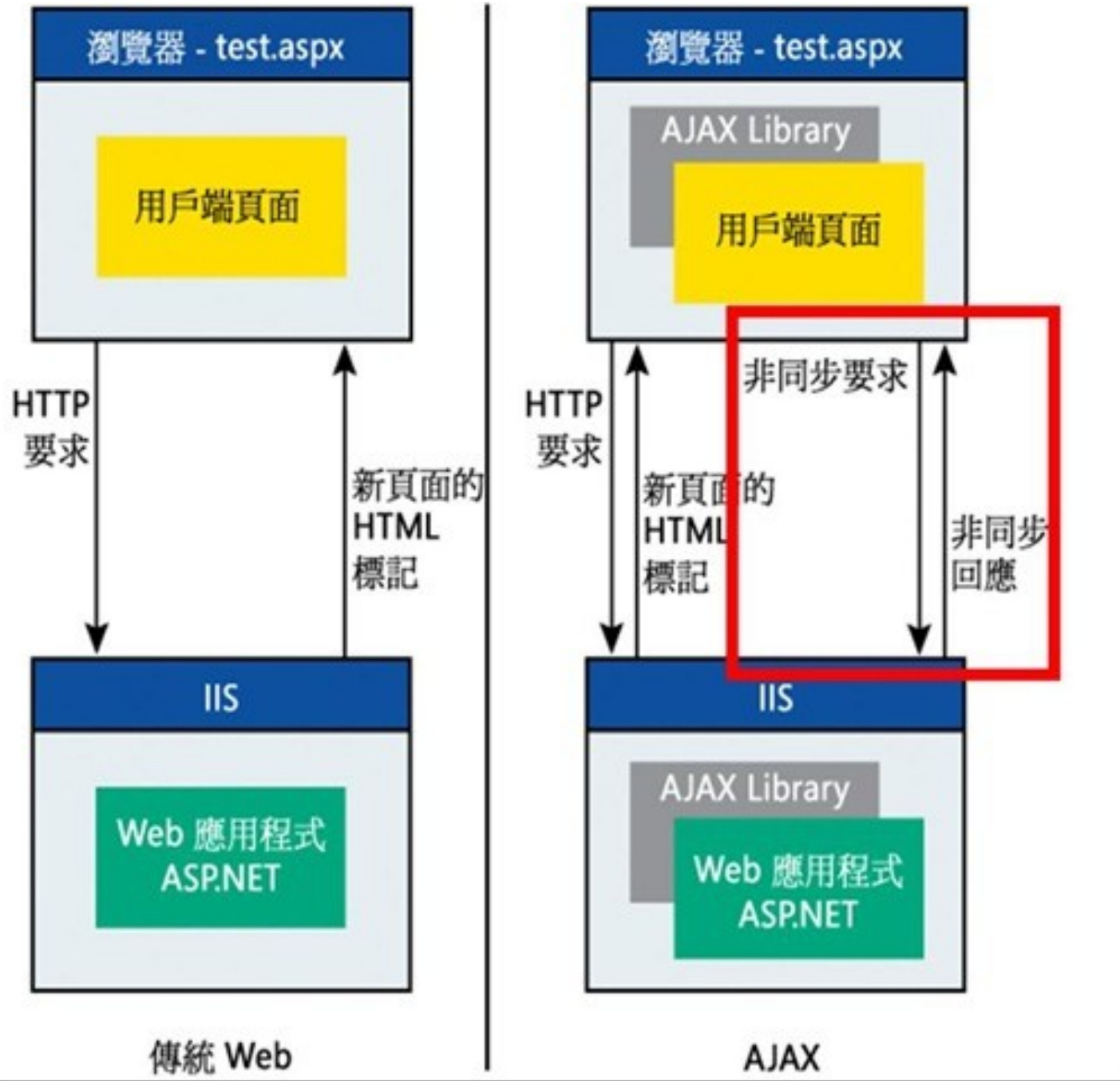
起始
座標

2. 運算
掃過區域

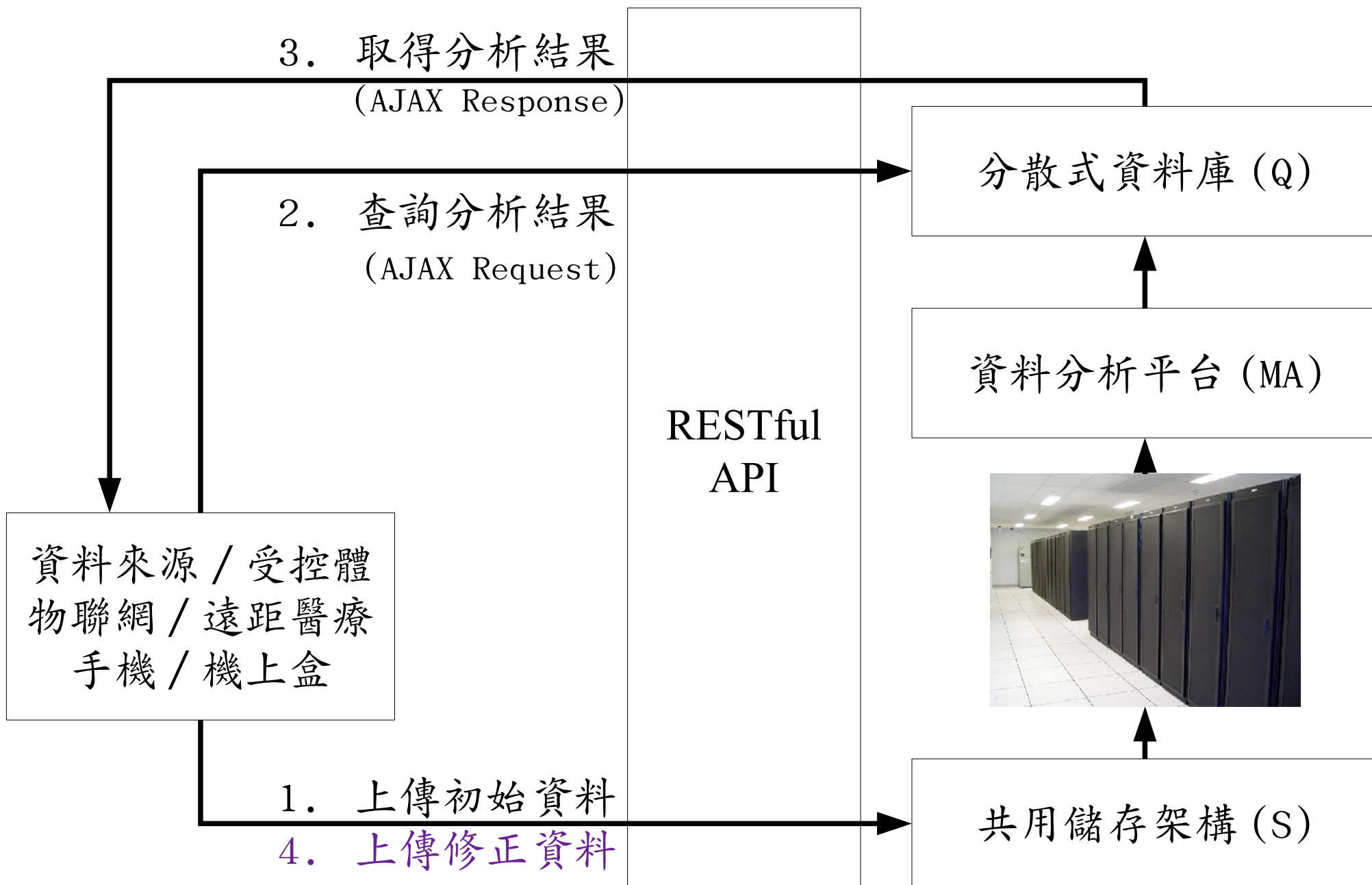


曾經有人問道：我是學電機的，那電機與控制跟雲端有甚麼關聯性？

Web 1.0 vs Web 2.0 非同步回應的資訊架構

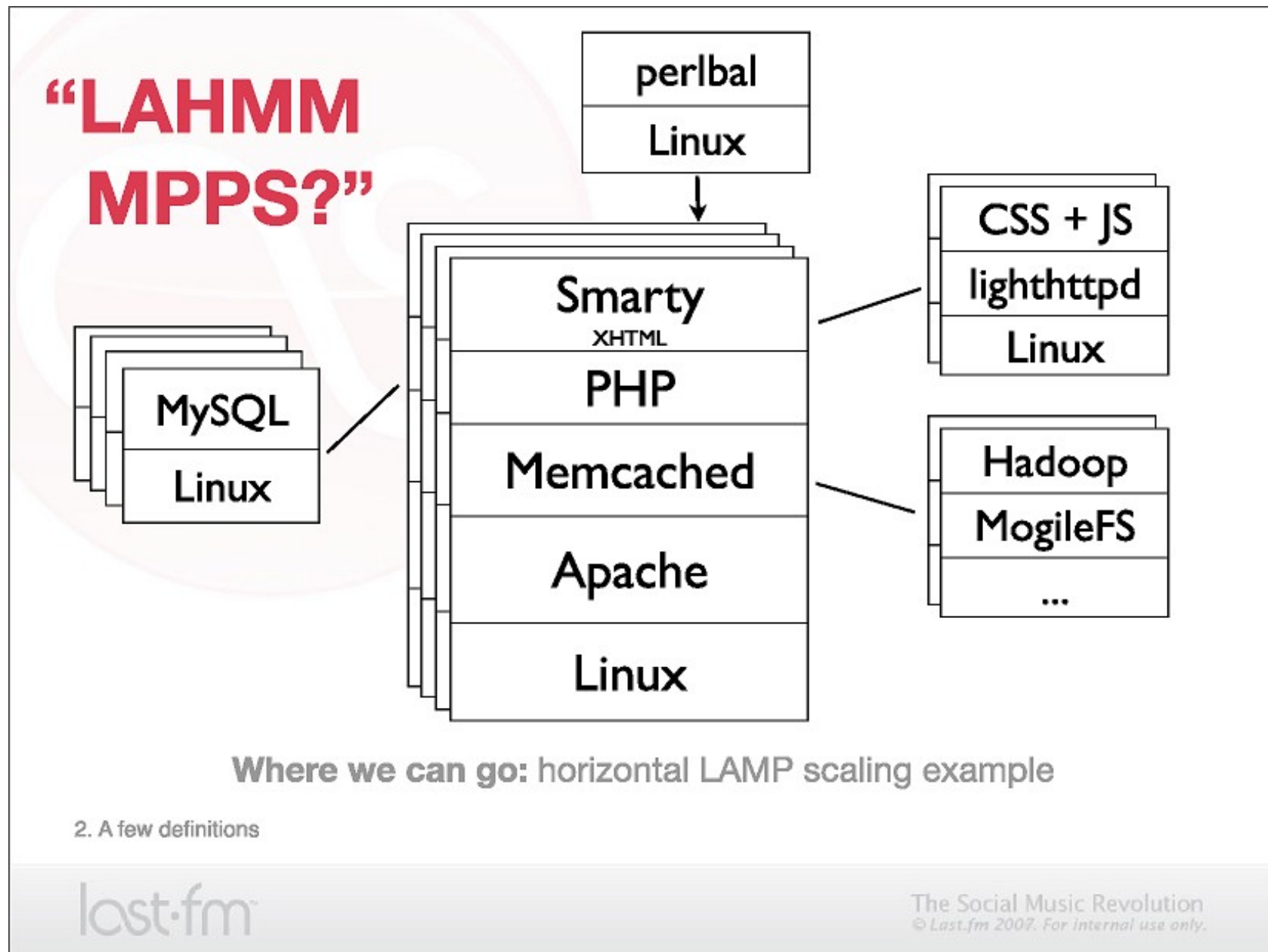


Common Architecture 通用架構



How to scale up web service in the past ?

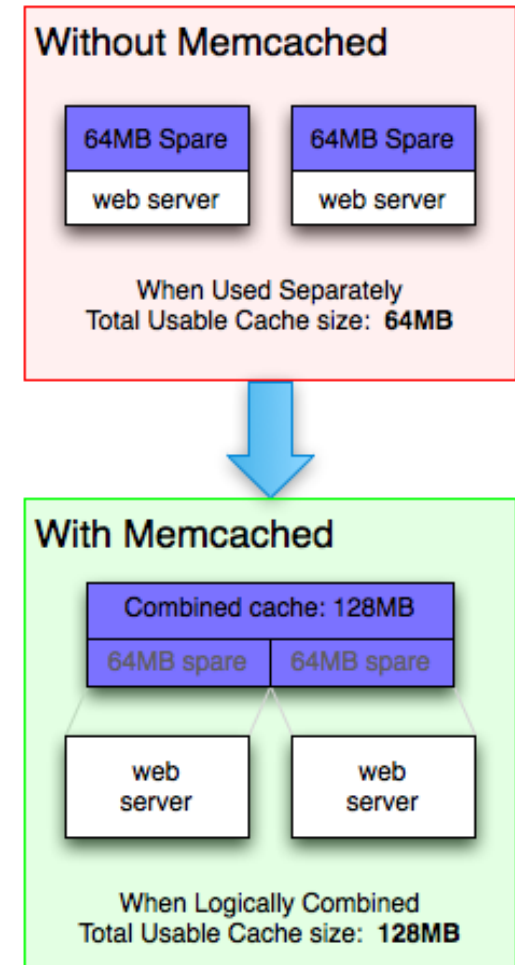
大型網頁服務的參考架構



Tools used by large scale websites

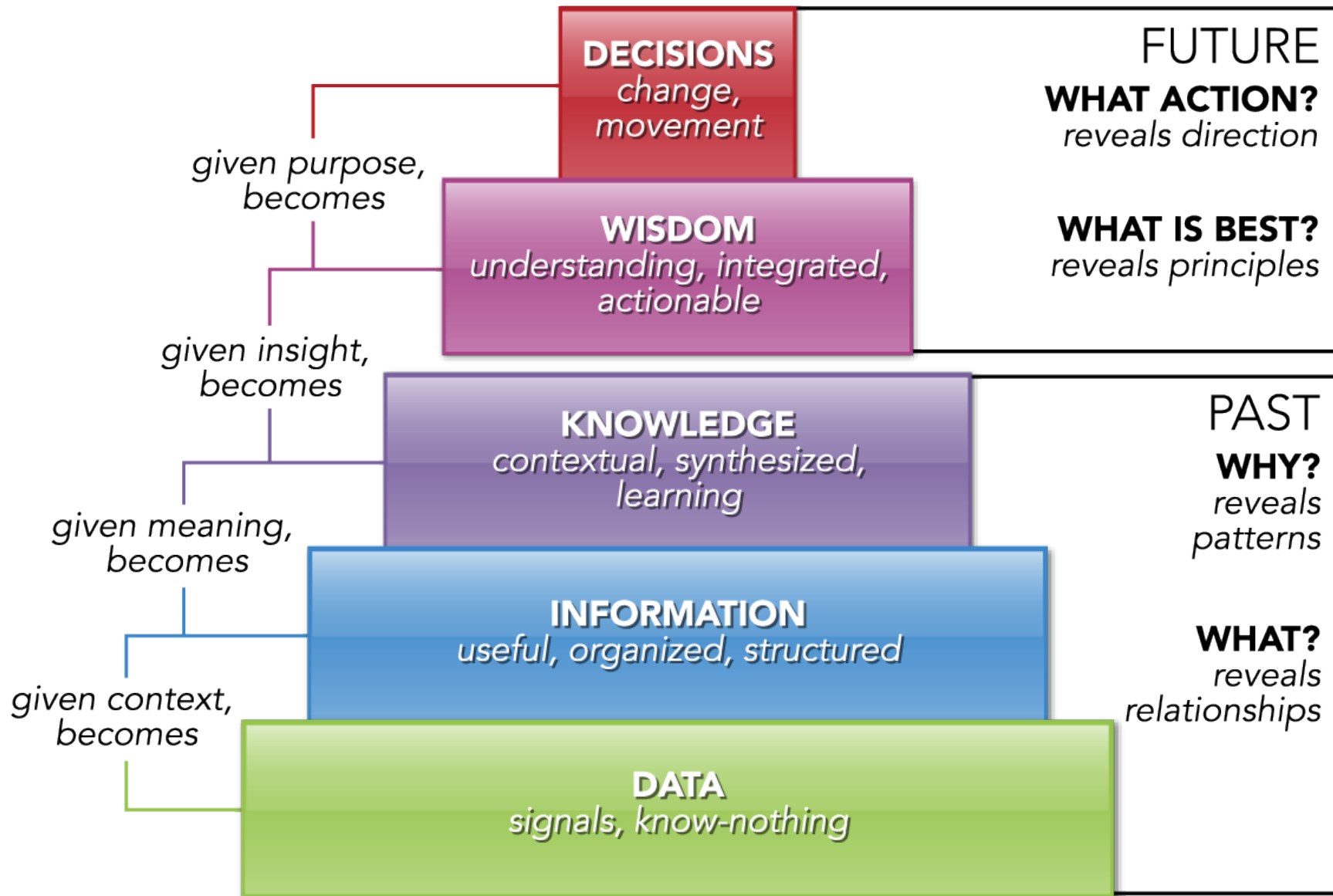
大型網頁服務常用的軟體組件

- Perlbal - <http://www.danga.com/perlbal/>
 - ◆ 多個網頁伺服器的負載平衡
 - ◆ Load balancer
- MogileFS - <http://www.danga.com/mogilefs/>
 - ◆ 分散式檔案系統
 - ◆ Distributed File System for small files
 - ◆ 有公司認為 MogileFS 比起 Hadoop 適合拿來處理小檔案
- memcached - <http://memcached.org/>
 - ◆ 共享記憶體 ??
 - ◆ Share Memory
 - ◆ 把資料庫或經常讀取的部分，用記憶體快取 (Cache) 方式
- Moxi - <http://code.google.com/p/moxi/>
 - ◆ Memcache 的 PROXY
- More Resource:
 - ◆ <http://code.google.com/p/memcached/wiki/HowToLearnMoreScalability>
 - ◆ <http://www.slideshare.net/techdude/scalable-web-architectures-common-patterns-and-approaches>



Data, Information, Knowledge, Wisdom

知識管理模型：資料、資訊、知識與智慧



Business Intelligence 商業智慧

Data Mining

資料探勘

Data Warehouse

資料倉儲

Data Integration 資料整合

若想要達成商業智慧的目標，請先做資料整合、資料倉儲與探勘平台

ERP
金流

CRM
人事

MES
倉管物流

KMS
資訊流

TOM
資訊流

Logs / Files
系統日誌

Compute 計算設施

Network 網路設施

Storage 儲存設施

虛擬化
Virtualization

Data Integration ? 怎麼做資料整合?

Source : http://en.wikipedia.org/wiki/Data_integration

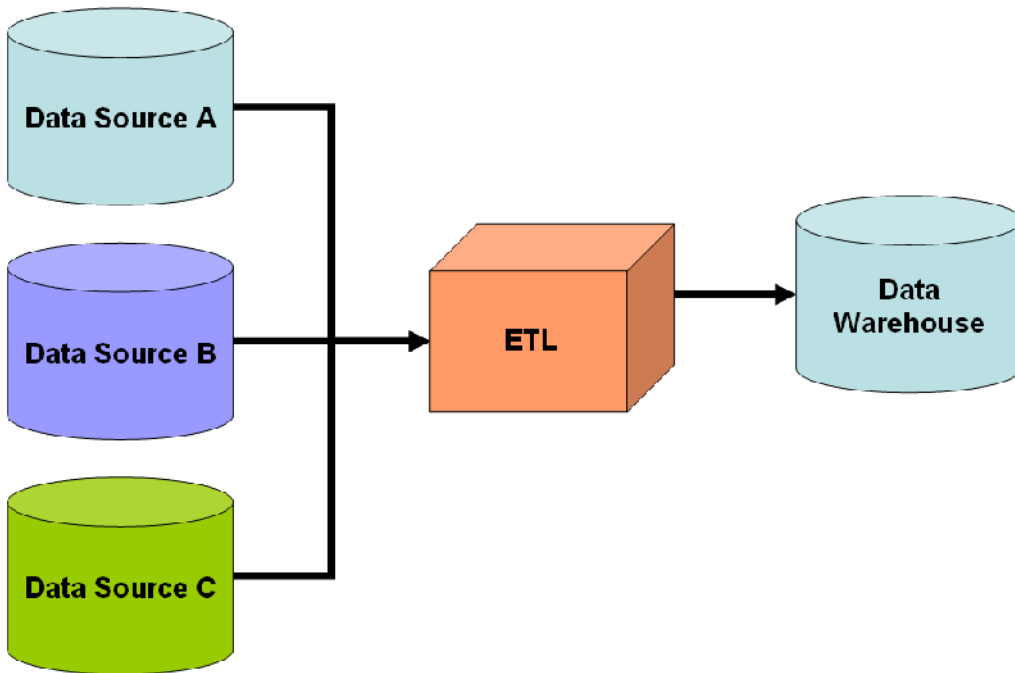


Figure 1: Simple schematic for a **data warehouse**. The ETL process extracts information from the source databases, transforms it and then loads it into the data warehouse.

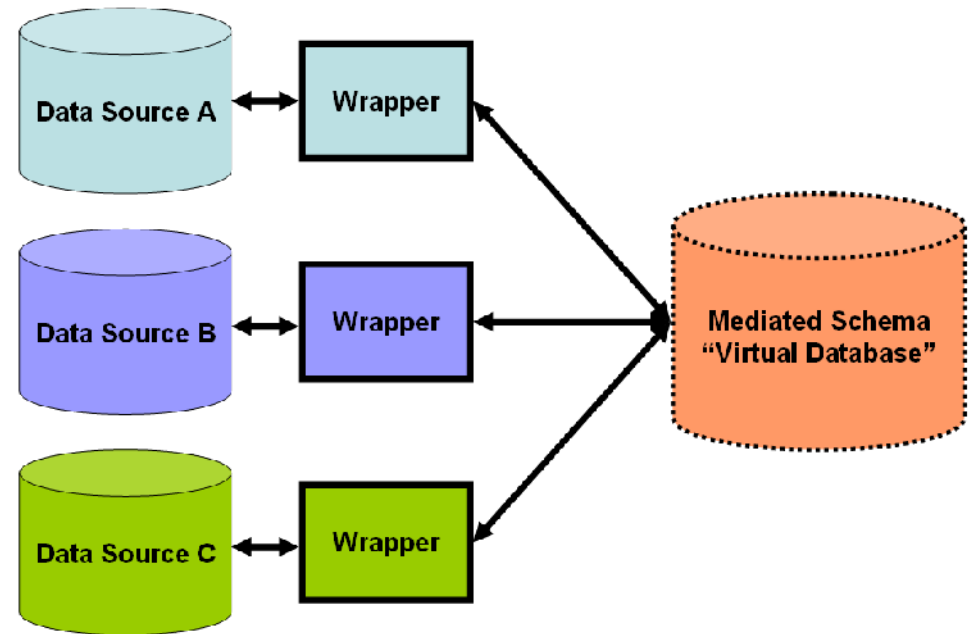


Figure 2: Simple schematic for a data-integration solution. A system designer constructs a mediated schema against which users can run queries. The **virtual database** interfaces with the source databases via **wrapper** code if required.

The way toward Business Intelligence 通往商業智慧的漫長道路

Business Intelligence

商業智慧



Data Mining

資料探勘



Data Warehouse

資料倉儲



Data Integration

資料整合



OS-level Virtualization

作業系統虛擬化



Network Virtualization

網路虛擬化



Storage Virtualization

儲存虛擬化



Data Scientist !! 成功關鍵在於資料科學家 !!

Data scientist: The hot new gig in tech

By Michal Lev-Ram, writer September 6, 2011: 5:00 AM ET

Companies that want to make sense of all their bits and bytes are hiring so-called data scientists - if they can find any.

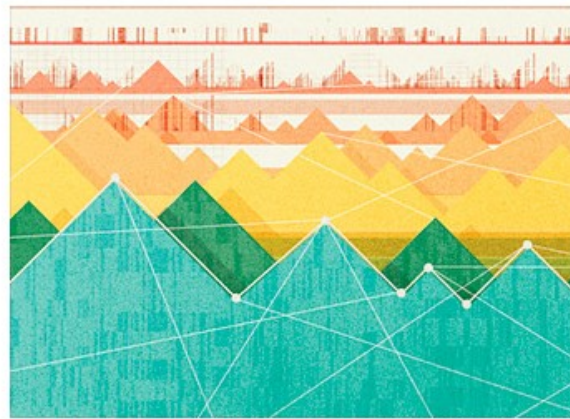


ILLUSTRATION: GAVIN POTENZI

FORTUNE -- The unemployment rate in the U.S. continues to be abysmal (9.1% in July), but the tech world has spawned a new kind of highly skilled, nerdy-cool job that companies are scrambling to fill: data scientist.

會「統計」的人照過來！

財星雜誌 (FORTUNE) 等均報導今年最熱門的職缺是「資料科學家」！

What is data science?

Data science can be broken down into four essential parts.

Mining data



Collecting and formatting the information

Statistics



Information analysis

Interpret



Representation or visualization in the form of presentations, infographics, graphs or charts

Leverage



Implications of the data, application of the data, interaction using the data and predictions formed from studying it

光纖卡

伺服器

機上盒

屠家照護

儲存器

交換器

準系統

手機

物聯網

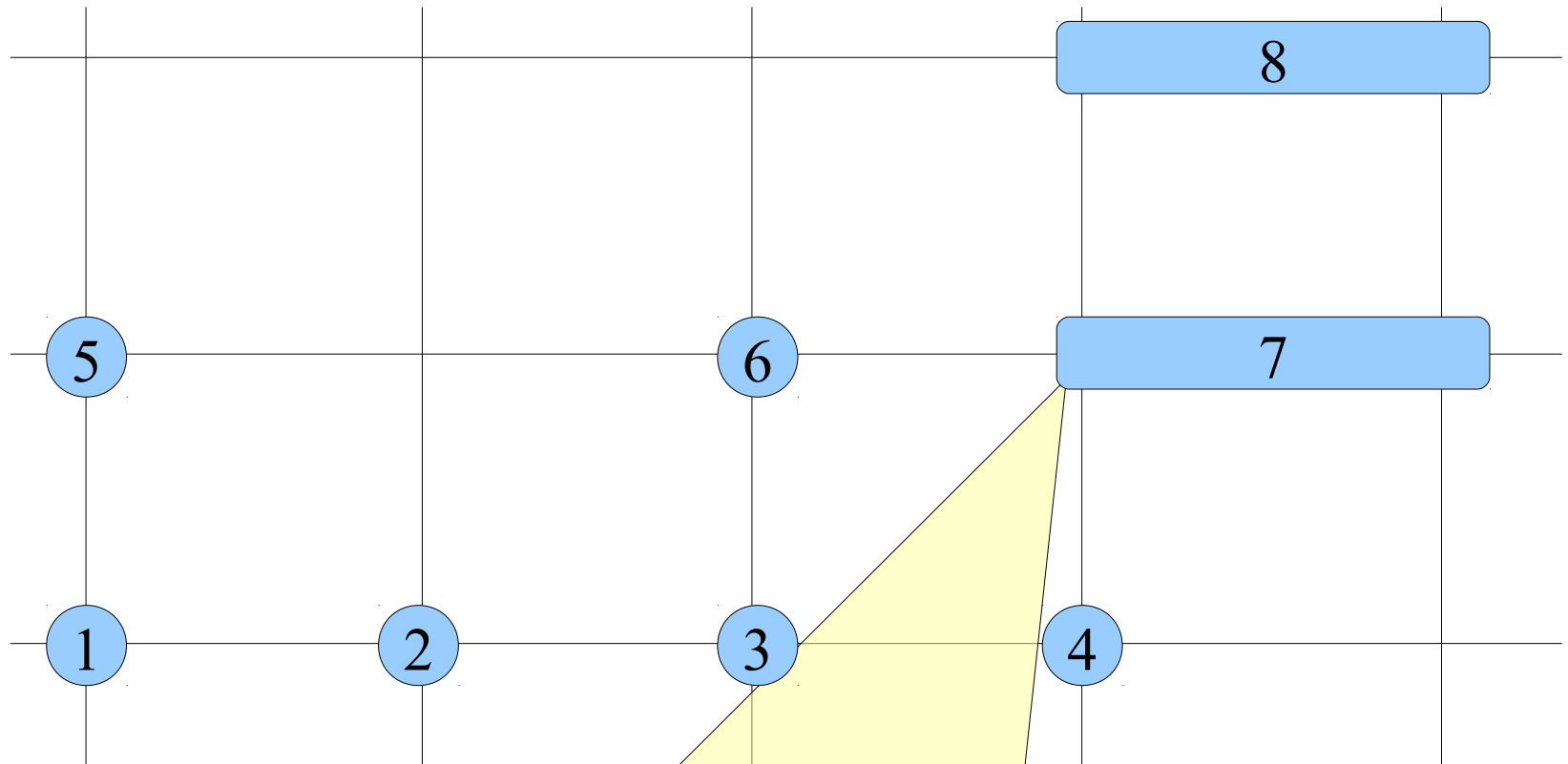
路由

主機板

Web 2.0
網頁服務

Big Data
資料分析

Virtualization
虛擬化技術



趨勢 (7) 未來人們將透過各種裝置存取整理過的資料！
 建議 (7) 跳脫單機思維，走向跨不同裝置的共用介面
 行動 (7) 從資料分析結果，建立巨型共用資料庫

光纖卡

伺服器

機上盒

屠家照護

儲存器

交換器

準系統

手機

物聯網

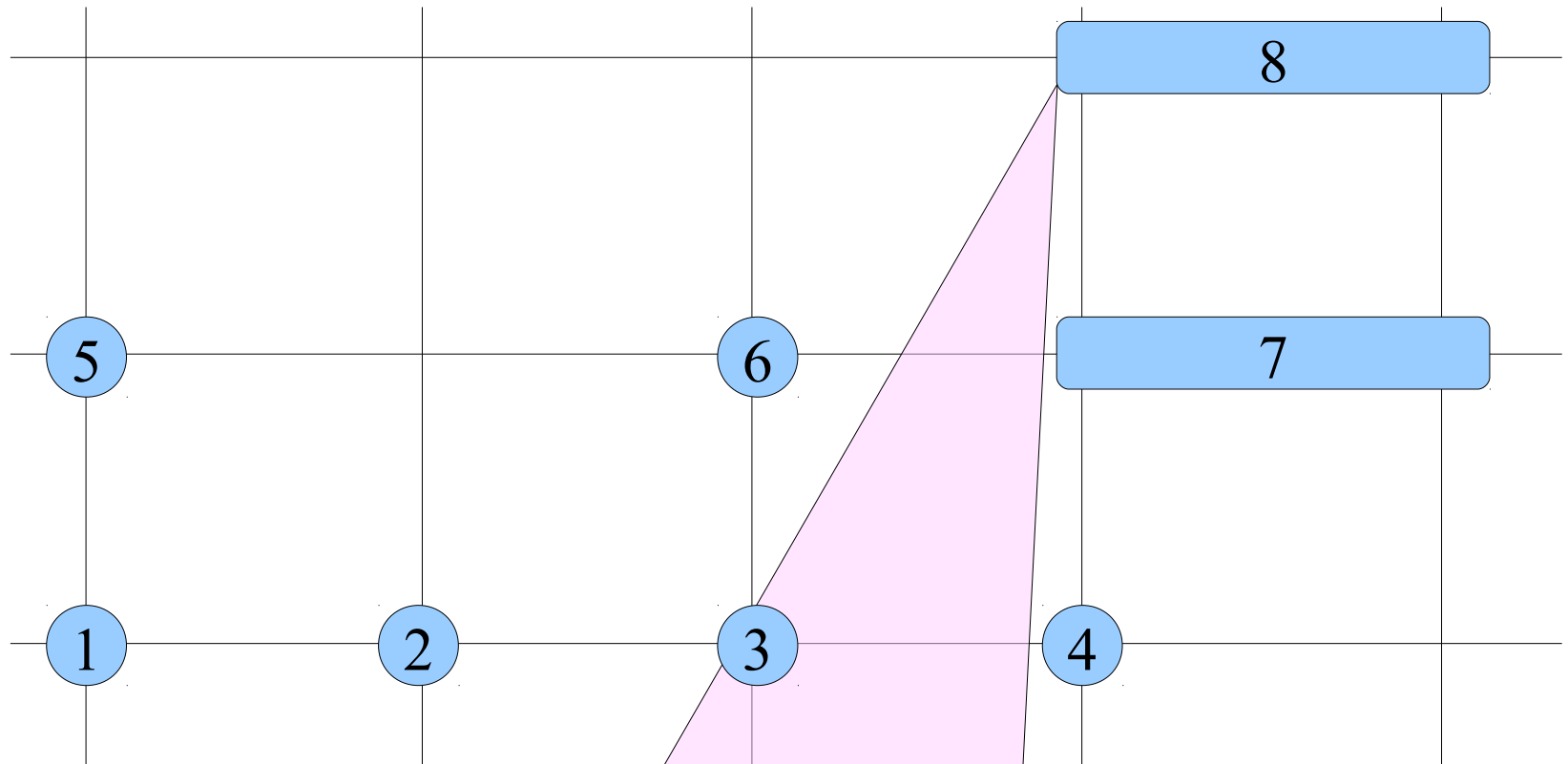
路由

主機板

Web 2.0
網頁服務

Big Data
資料分析

Virtualization
虛擬化技術



趨勢 (8) 未來人們將透過口語、手勢來跟裝置互動！
 建議 (8) 跳脫機械思維，發揮「以人為本」的精神！
 行動 (8) 用 AJAX 搭配 Apps，發揮資料整理後的效益

走向雲端運算的道路是一場「整體戰」！ 要靠各產品線各部門一同打拚！

建議 (1) 進行對儲存虛擬化支援性的測試及策略聯盟
行動 (1) 從 RAID 到分散式檔案系統的效能測試

建議 (2) 進行對網路虛擬化支援性的測試及策略聯盟
行動 (2) 導入 OpenFlow / OpenVSwitch 標準

建議 (3) 設計支援虛擬化、省電、高工作溫度的主機板
行動 (3) 評估 Open Compute Project

建議 (4) 選擇應用軟體多的嵌入式作業系統！
行動 (4) 評估導入嵌入式虛擬化技術！

建議 (5) 思考如何從硬體架構增加 I/O
行動 (5) 從 RAID、PCIe SSD 介面著手

建議 (6) 思考硬體架構該如何匹配，才能增加 I/O
行動 (6) 從南北橋晶片下手

建議 (7) 跳脫單機思維，走向跨不同裝置的共用介面
行動 (7) 從資料分析結果，建立巨型共用資料庫

建議 (8) 跳脫機械思維，發揮「以人為本」的精神！
行動 (8) 用 AJAX 搭配 Apps，發揮資料整理後的效益



Questions?

Slides - <http://trac.nchc.org.tw/cloud>

Jazz Wang
Yao-Tsung Wang
jazz@nchc.org.tw

