

# Course Information 課程資訊



- 講師介紹：
  - 國網中心 王耀聰 副研究員 / 交大電控碩士
  - [jazz@nchc.org.tw](mailto:jazz@nchc.org.tw)
- 所有投影片、參考資料與操作步驟均在網路上
  - 由於雲端資訊變動太快，愛護地球，請減少不必要之講義列印。
- 礙於缺乏實機操作環境，故以影片展示與單機操作為主
  - 若有興趣實機操作，請參考國網中心雲端運算課程錄影
  - <http://trac.nchc.org.tw/cloud>
  - <http://www.classcloud.org/media>
  - <http://www.screentoaster.com/user?username=jazzwang>
- 若需要實驗環境，可至國網中心雲端運算實驗叢集申請帳號
  - <http://hadoop.nchc.org.tw>
- Hadoop 相關問題討論：
  - <http://forum.hadoop.tw>



# 淺談雲端運算的新趨勢 及其對政府部門之影響

The trend of Cloud Computing and its impact to public sectors

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



Powered by DRBL

# 淺談雲端運算的定義與驅動力

Part 1 : Overview of Cloud Computing

## 雲端運算的三大關鍵技術

Part 2 : Overview of Cloud Computing Core Technologies

### 深入解析雲端虛擬化技術

Part 2-1 : Deep Dive into Virtualization Technologies

### 深入解析大量資料分析技術

Part 2-2 : Deep Dive into Data Science Technologies

## 雲端運算對政府部門之衝擊與因應對策

Part 3 : What is the Impacts and How should public sector adjust ?

## 安全：雲端、智慧電網、物聯網的共通問題

Part 4 : Security Issue of Cloud, Smart Grid and Internet of Things



# 淺談雲端運算的定義與驅動力

Part 1 : Overview of Cloud Computing

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



Powered by DRBL

# WHAT



Source: <http://www.2010taipeiexpo.tw/ct.asp?xItem=17186&CtNode=5952&mp=3>



# 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。

**雲端概念**

雲端基礎架構的相關IT建設，  
如伺服器、網路設施、  
電源供應器、散熱、  
儲存裝置等硬體產品，  
都是台灣科技業的強項。

基礎設施(IaaS)

**雲端概念**

在硬體設備的優勢基礎上，  
雲端平台與應用服務，  
提供台灣ICT產業  
一個轉型的新契機，  
台灣的創新能力，  
不落人後。

雲端平台(PaaS)

應用服務(SaaS)

<http://www.youtube.com/watch?v=bJLSAcU6O3U>

<http://www.youtube.com/watch?v=VIMtd3nfPqc>

當紅「雲端運算」 你瞭解了嗎？  
雲端產業 8分鐘就上手



什麼是雲端運算啊？可以個簡單的定義嗎？

**What is Cloud Computing ?**

雲端運算怎麼聽起來要買一些新硬體、新軟體啊？

**Is it about buying NEW Hardware and Software?**



雲端運算可能只是拿來振興經濟的幌子吧？

**Is it a trap to another bubble economy ?**

我聽你們在那裡講五四三.....

**Cloud Computing is as simple as 5..4..3..2..1...**



# 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



Microsoft

Google

**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

Target Market

is **S.M.B.**

主要客戶為  
中小企業

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

## 雲端運算的三種服務模式 (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



# 2 perspectives : Services vs Technologies

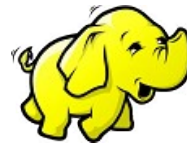
您想聽的是「雲端服務」還是「雲端技術」？

Google YouTube e W



雲端服務

Microsoft



雲端技術



Cloud computing hype spurs confusion, Gartner says

<http://www.computerworld.com/s/article/print/9115904>

淺談雲端運算 (Cloud Computing)

[http://www.cc.ntu.edu.tw/chinese/epaper/0008/20090320\\_8008.htm](http://www.cc.ntu.edu.tw/chinese/epaper/0008/20090320_8008.htm)

# 2 R&D directions : Cloud or Device

兩大研究方向：你該選「雲」還是「端」？



# 1 key spirit of Cloud Computing

用一句話說明雲端運算！服務才是王道！

Anytime 隨時

Anywhere 隨地

With Any Devices 使用任何裝置

Accessing Services 存取各種服務

Cloud Computing =~ Network Computing

雲端運算 =~ 網路運算

Key spirit of Cloud ~

形成服務才是重點！！

Everything as a Service !!

WHAT



花精靈-小球

WHEN



花精靈-小葵

# 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

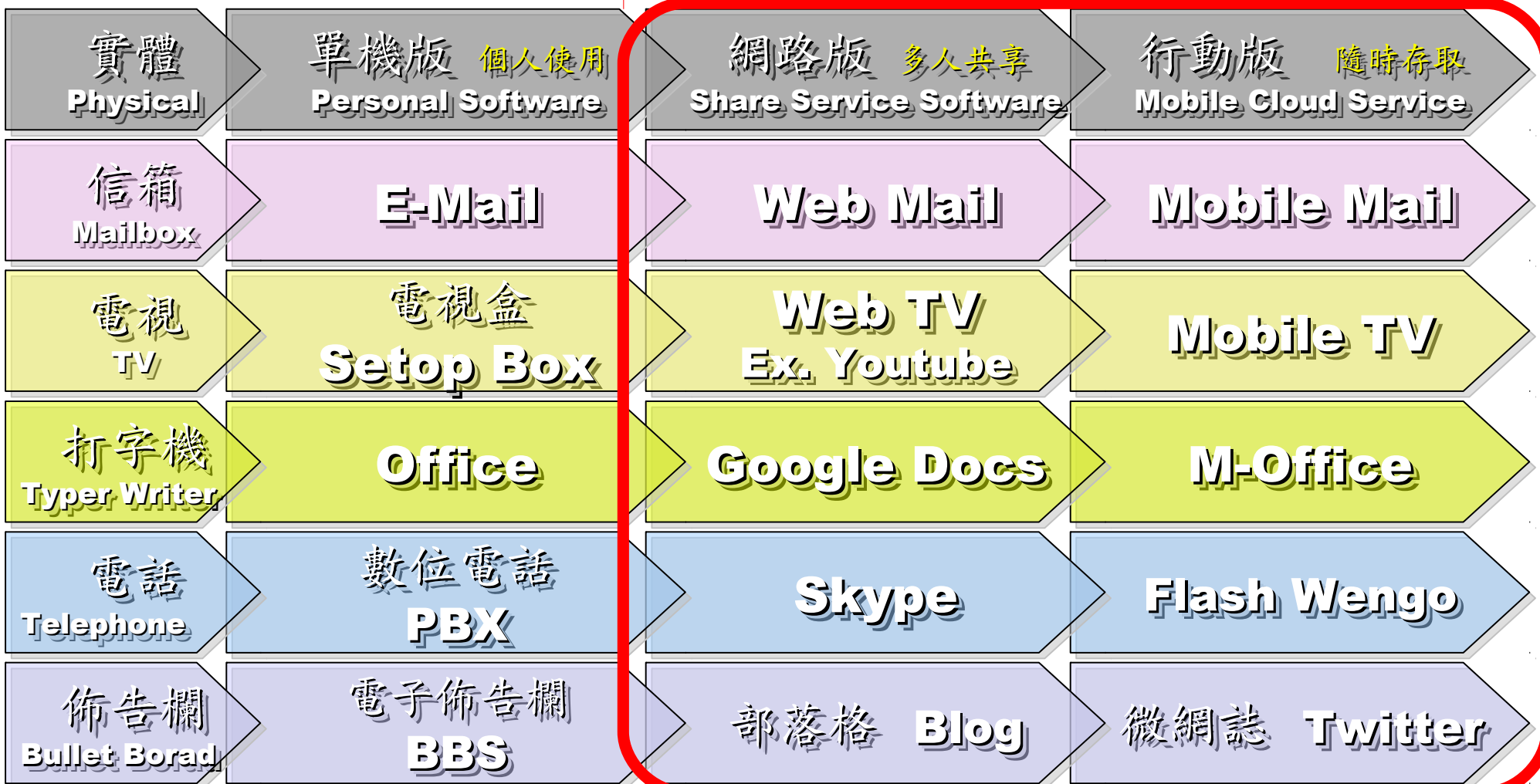
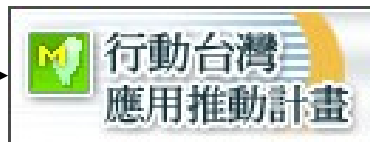
2006年8月24日

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



# Evolution of Cloud Services

雲端服務只是軟體演化史的必然趨勢

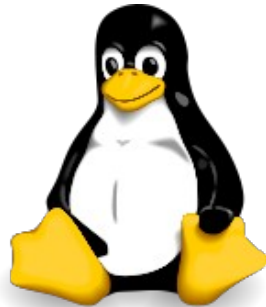




# Brief History of Computing

## 運算技術演進簡史

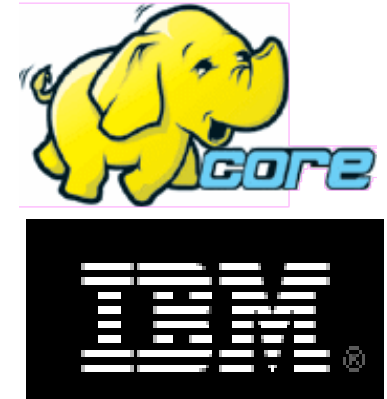
1991



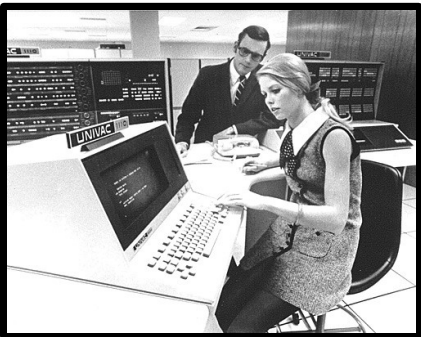
2002



2004



1960



1977



1993



2003



2006



**Mainframe**  
Super  
Computer

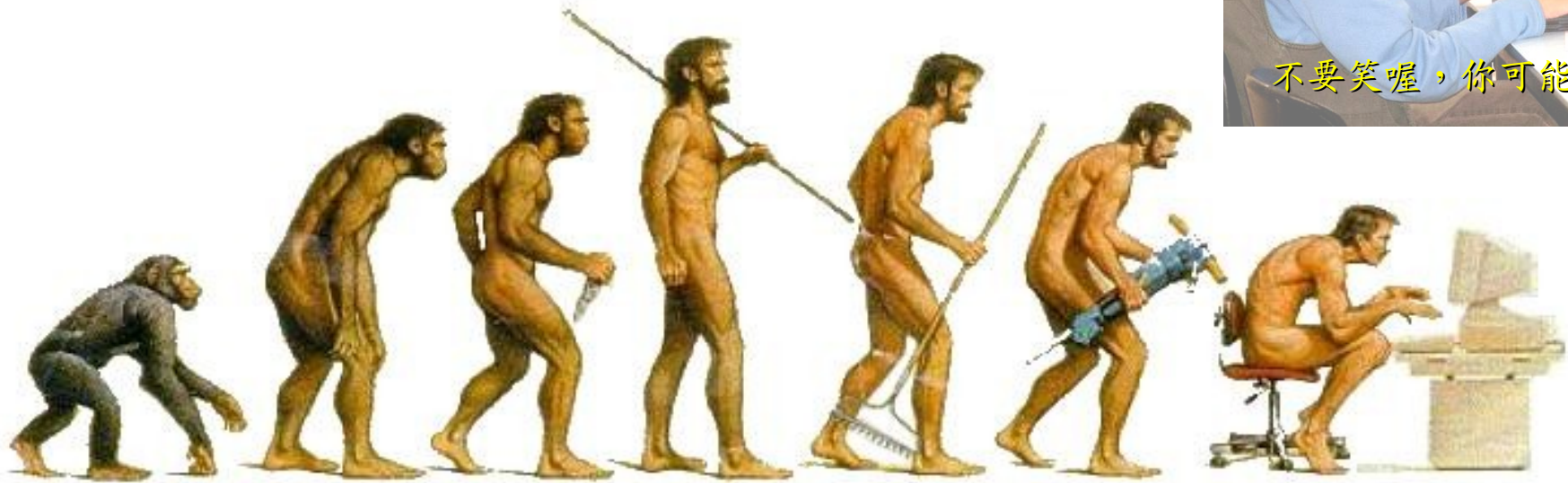
**PC / Linux**  
Cluster  
Parallel

**Internet**  
Distributed  
Computing

**Virtual Org.**  
Grid  
Computing

**Data Explode**  
Cloud  
Computing

# Evolution



(OR is it?)

**What can we learn from the past ?!**

在這漫長的演化中，我們到底學到些什麼？！

**WHAT**



花精靈-小球

**WHEN**



花精靈-小葵

**WHY**



花精靈-小鬱

Source: <http://www.2010taipeiexpo.tw/ct.asp?xItem=17186&CtNode=5952&mp=3>

**Lesson #1: One cluster can't fit all !**

**教訓一：叢集的單一設定無法滿足所有需求！**

**Answer #1: Virtual Cluster 新服務：虛擬化叢集**

**Lesson #2: Grid for Heterogeneous Enterprise !**

**教訓二：格網運算該用在異業結盟的資源共享！**

**Answer #2: Peak Usage Time 尖峰用量發生時間點**

**Lesson #3: Extra cost to move data to Grid !**

**教訓三：資料搬運的網路與時間成本！**

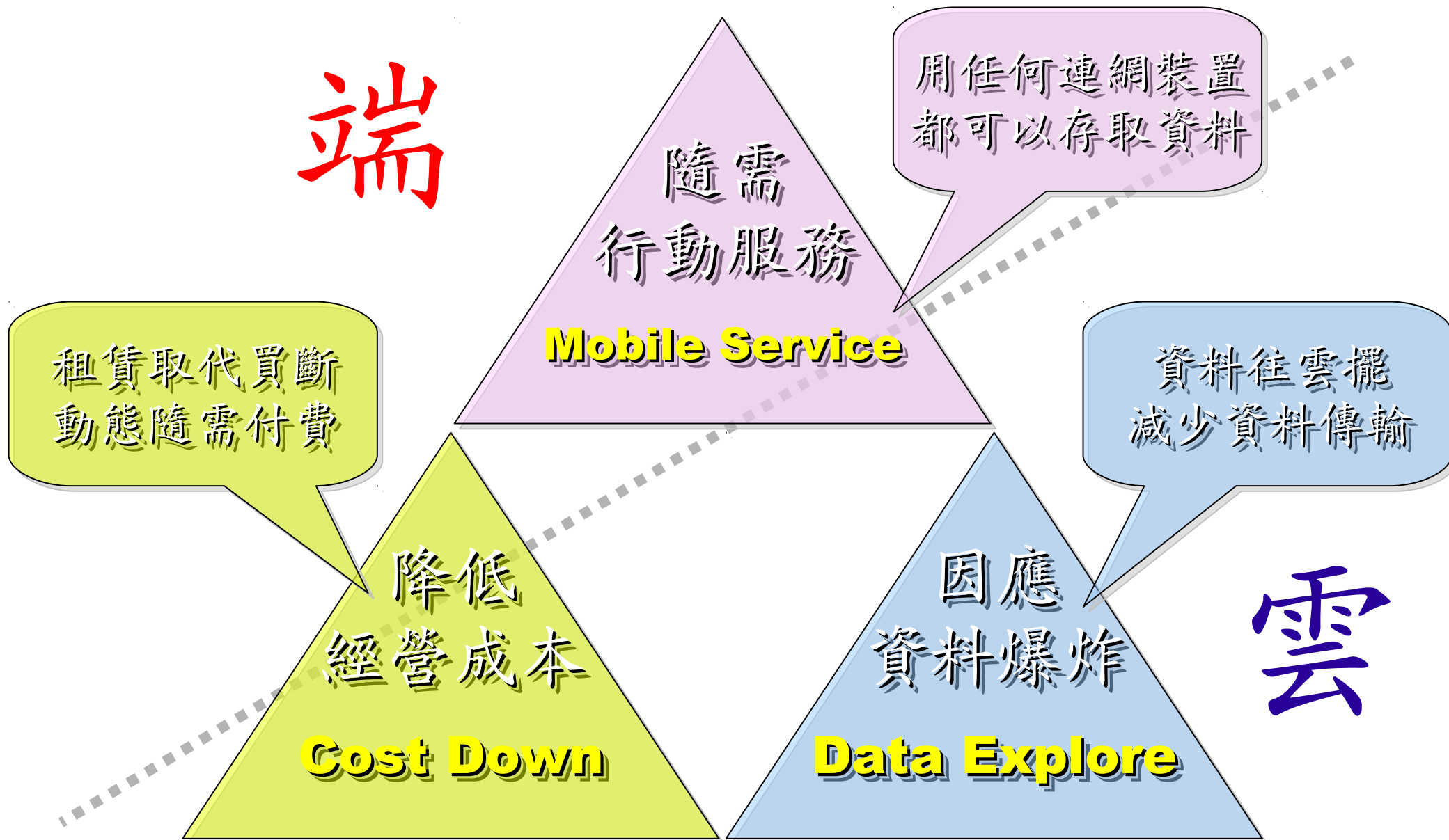
**Answer #3: Total Cost of Ownership 總擁有成本**

**Cost Down is the Key Drive !!**

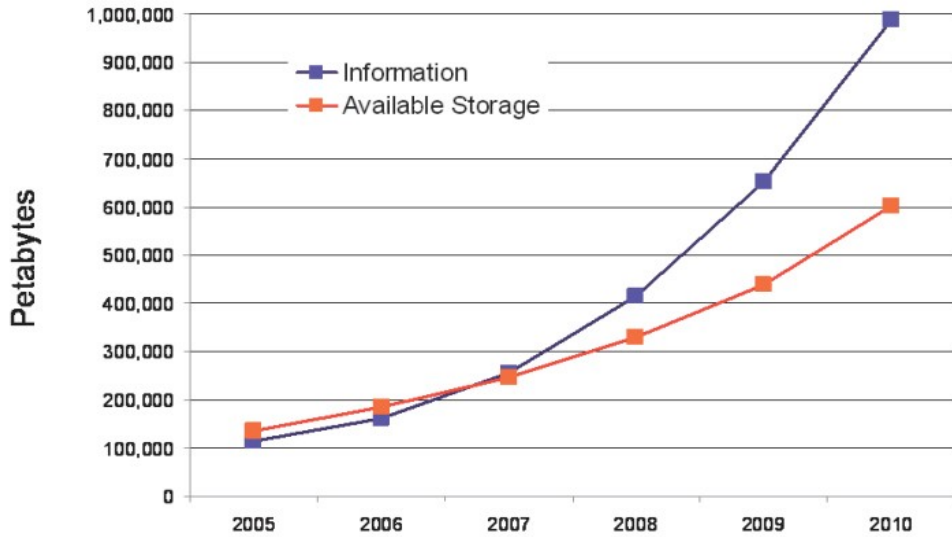
**降低營運成本才是企業導入雲端運算的關鍵考量！！**

# Key Driving Forces of Cloud Computing

## 雲端運算的關鍵驅動力



# Information Versus Available Storage



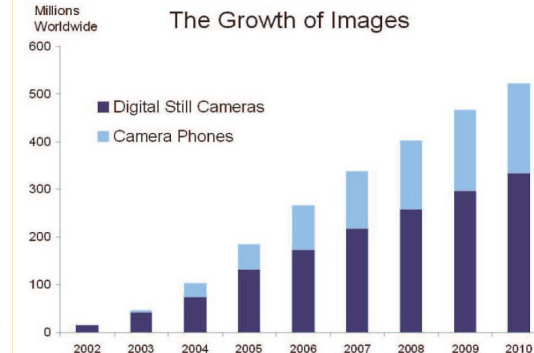
Source: IDC, 2007

# 2007 Data Explore

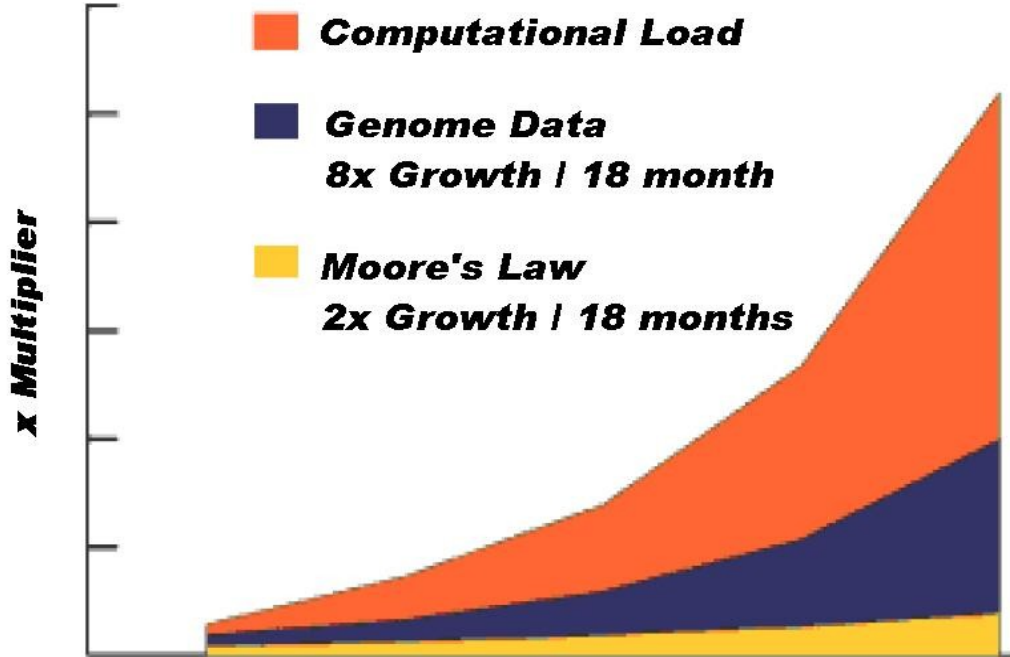
**Top 1 : Human Genomics - 7000 PB / Year**  
**Top 2 : Digital Photos - 1000 PB+ / Year**  
**Top 3 : E-mail (no Spam) - 300 PB+ / Year**



Source: IDC, 2007



Source: IDC, 2007



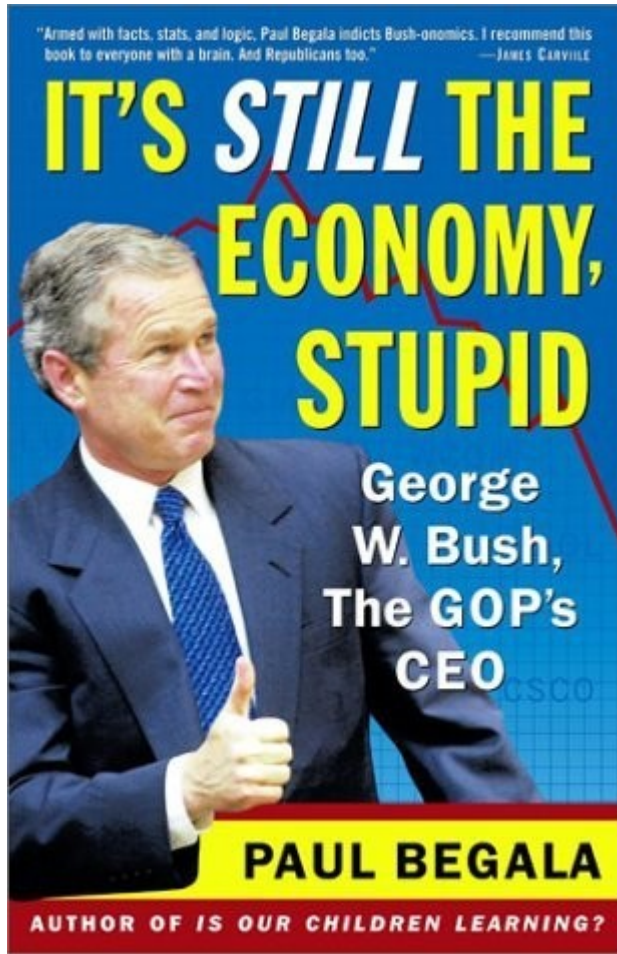
Particle Physics Large Hadron Collider (15PB)	Human Genomics (7000PB) 1GB / person 200PB+ captured 200% CAGR	World Wide Web (~1PB)	Wikipedia (10GB) 100% CAGR
Annual Email Traffic, no spam (300PB+)	Internet Archive (1PB+)	Estimated On-line RAM in Google (8PB)	Personal Digital Photos (1000PB+) 100% CAGR
200 of London's Traffic Cams (8TB/day)	2004 Walmart Transaction DB (500TB)	Typical Oil Company (350TB+)	Merck Bio Research DB (1.5TB/qtr)
UPMC Hospitals Imaging Data (500TB/yr)	MIT Babytalk Speech Experiment (1.4PB)	Terashake Earthquake Model of LA Basin (1PB)	One Day of Instant Messaging in 2002 (750GB)
Total digital data to be created this year <b>270,000PB</b> (IDC)			

Phillip B. Gibbons, Data-Intensive Computing Symposium

Source: <http://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf>

Source: [http://lib.stanford.edu/files/see\\_pasig\\_dic.pdf](http://lib.stanford.edu/files/see_pasig_dic.pdf)

# IT'S THE DATA, STUPID!



「笨蛋！重點在經濟」

( **"It's the economy, stupid"** )

卡維爾 ( **James Carville** ) 自創這句標語，  
促使柯林頓當上美國第 **42** 屆總統。

- **1992** 年

「笨蛋！重點還是在經濟」

( **"It's STILL the economy, stupid"** )

卻讓小布希嘲笑是幼稚的總統。

- **2002** 年

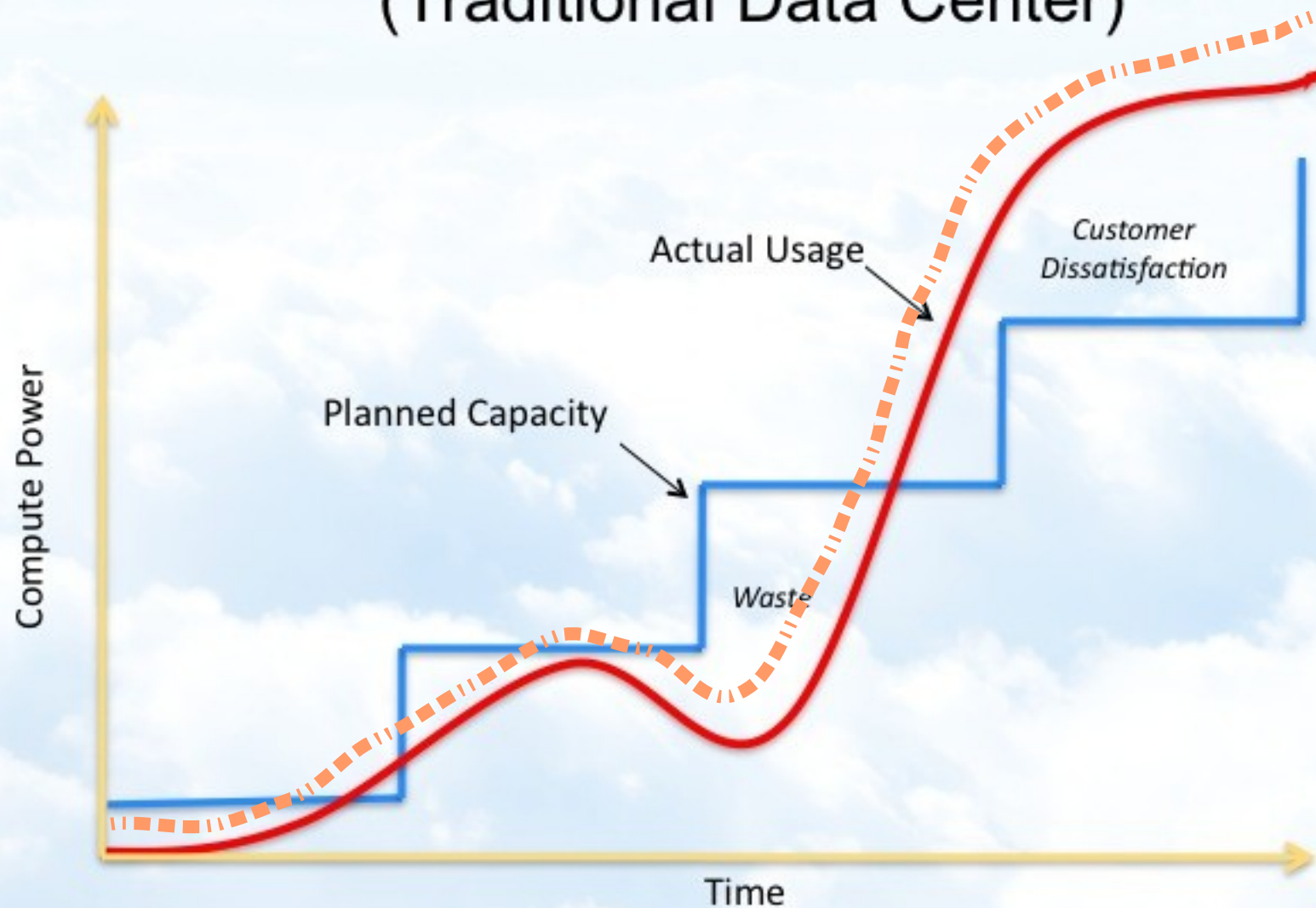
雲端時代，谷歌會說：「笨蛋！重點在資料」

( **"It's the data, stupid"** )

誰掌握了你的資料，就有機會掌握你的荷包  
想想看，電腦、手機掉了，您心疼的是甚麼呢？

- **2007** 年

# Capacity vs. Usage (Traditional Data Center)

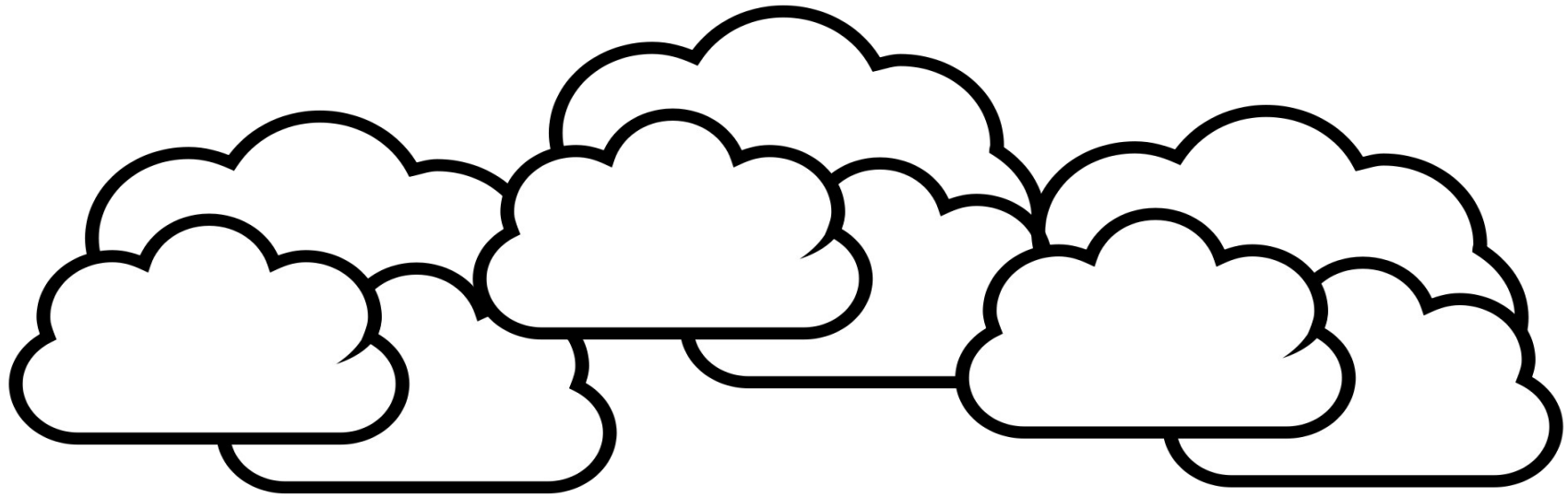


Source : [http://awsmedia.s3.amazonaws.com/chart01\\_traditional\\_720x540.jpg](http://awsmedia.s3.amazonaws.com/chart01_traditional_720x540.jpg)



# Data is the source of Wisdom !!

用雲掌握資料，加以分析，形成智能給端用



嵌入式的新思維：未來，**端**的智能來自於**雲**

**Devices share the wisdom of Cloud**



**WHAT**



花精靈-小球

**WHEN**



花精靈-小葵

**WHY**



花精靈-小鬱

**WHO**



花精靈-百兒

Source: <http://www.2010taipeiexpo.tw/ct.asp?xItem=17186&CtNode=5952&mp=3>



**Flying to the Cloud ...  
or  
Falling to the Ground ...**

Source: [http://media.photobucket.com/image/falling%20ground/preeto\\_f10/falling.jpg](http://media.photobucket.com/image/falling%20ground/preeto_f10/falling.jpg)

該使用別人打造的雲端，還是自己打造專屬雲端呢？

# How can we build Cloud Services ??

觀察雲端關鍵提供者，找尋打造雲端服務的模式

Public Cloud

公用雲端



Target Market

is **S.M.B.**

主要客戶為  
中小企業

雲端服務參考模型  
Reference Model

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



- Amazon Web Service ( AWS )
- 虛擬伺服器：**Amazon EC2**
  - Small (Default) \$0.085 per hour(L) - \$0.12 per hour(W)
  - All Data Transfer \$0.15 per GB
- 儲存服務：**Amazon S3**
  - \$0.15 per GB – first 50 TB / month of storage used
  - \$0.15 per GB – all data transfer in
  - \$0.01 per 1,000 PUT, COPY, POST, or LIST requests
- 觀念：**Paying for What You Use**

參考來源：<http://eblog.cisnet.org.tw/post/Cloud-Computing.aspx>  
<http://aws.amazon.com/ec2/pricing/>  
<http://aws.typepad.com/aws/2010/02/aws-data-transfer-prices-reduced.html>  
<http://aws.amazon.com/s3/#pricing>

# Amazon Web Services 亞馬遜的網頁服務

## E-Commerce

Fulfillment Web Service (FWS)

## Database

SimpleDB, Rational DB Service

## Compute

EC2, Elastic MapReduce (EMR)

## Network

VPC, Elastic Load Balance

## Storage

S3, Elastic Block Storage (EBS)

## Support

AWS Premium Support

## Payment / Billing

Amazon DevPay

## Messaging

Simple Queue / Notification Service

## Web Traffic

Alexa Web Info / Top Sites

## Content Delivery

Amazon CloudFront

較外顯的基本服務

較內隱的流程細節

- Google App Engine (GAE)
- 讓開發者可自行建立網路應用程式於 Google 平台之上。
- 提供：
  - 500MB of storage
  - up to 5 million page views a month
  - 10 applications per developer account
- 限制：
  - 程式設計語言只能用 Python 或 Java

計費標準：

- 連出頻寬 \$0.12 美元/GB, 連入頻寬 \$0.10 美元/GB
- CPU 時間 \$0.10 美元/時
- 儲存的資料 \$0.15 美元/GB-每月
- 電子郵件收件者 \$0.0001 美元/每個收件者

參考來源：<http://code.google.com/intl/zh-TW/appengine/>  
<http://code.google.com/intl/zh-TW/appengine/docs/billing.html>





Gmail / 電子郵件



Contact / 通訊錄



Postini / 通訊安全



Calendar / 行事曆



Talk / 即時通



Group / 網上論壇



Doc / 文件



Video / 影音



Sites / 協作平台



Mobile / 行動使用Apps



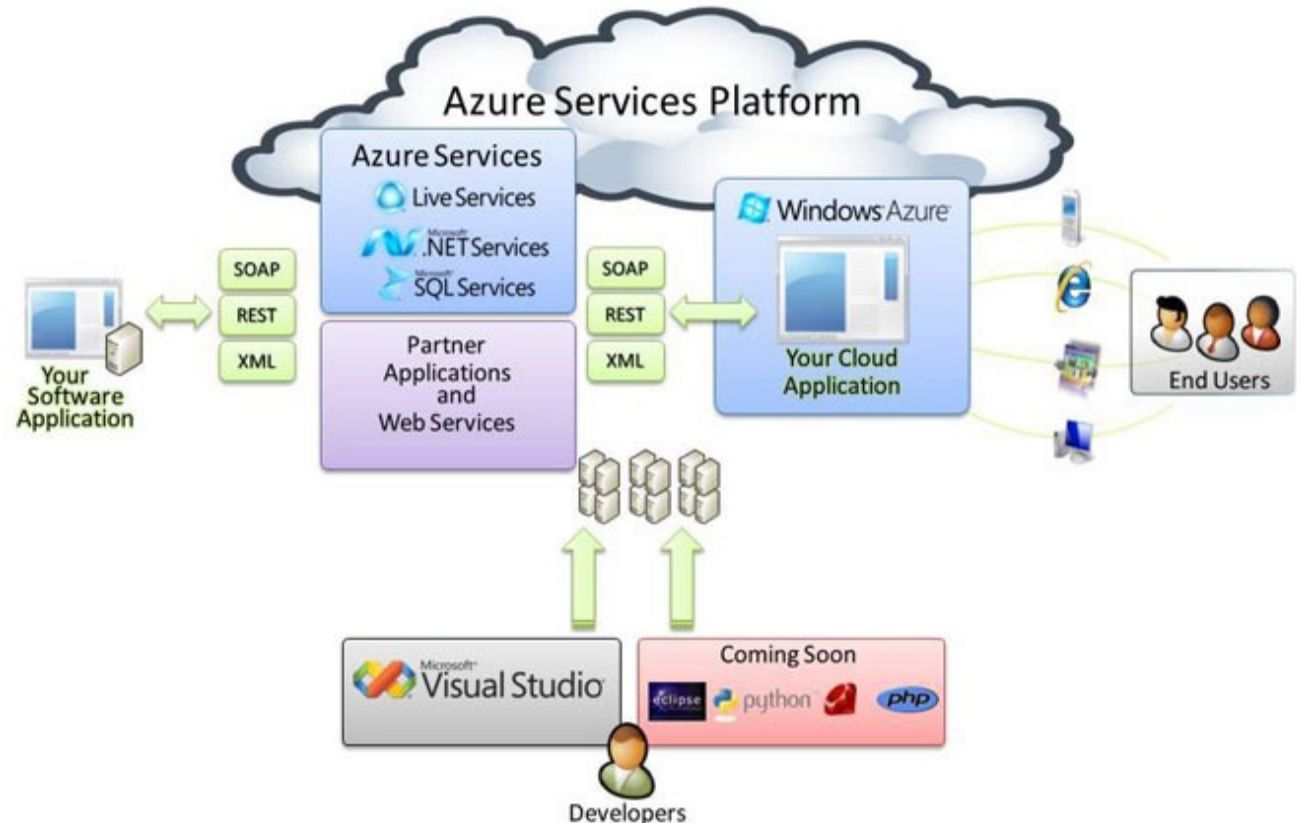
Admin / 管理後台

- **Google Apps**
- **Google Apps for Government**
- **Google Apps for ISPs**
- **Google Apps for Business**
- **Google Apps for Non-profits**

如果無法掌握雲端技術，至少該學會使用雲端服務！



- Microsoft Azure 是一套雲端服務作業系統。
- 作為 Azure 服務平台的開發、服務代管及服務管理環境。
- 服務種類：
  - .Net services
  - SQL services
  - Live services



XaaS Key Player :

Microsoft 微軟

# Microsoft Cloud Computing 全貌

Private

Public

Microsoft SharePoint Server  
Microsoft Exchange  
Microsoft Dynamics

Software as a Service (SaaS)

Microsoft Online Services

Microsoft SharePoint Services

Microsoft Office Live

Microsoft SQL Server  
Microsoft .NET

Platform as a Service (PaaS)

Windows Azure

SQL Services

Windows Azure platform  
AppFabric

Microsoft System Center  
Windows Server

Infrastructure as a Service (IaaS)

Windows Azure

Microsoft System Center

Windows Server

Microsoft Dynamic Data Center Toolkit For Enterprises

Microsoft Dynamic Data Center Toolkit For Hosters

IT as a Service

Dallas  
→ DaaS

Azure AppFabric  
→ PaaS  
(類似 GAE)

SQL Azure  
→ PaaS  
(雲端 SQL)

Windows Azure  
→ PaaS  
(類似 EC2)

Hyper-V  
→ IaaS  
(虛擬化)

# Public Cloud Comparison:

# 公用雲端的比較

	On-Premises Apps	Small-to-Medium Web Apps	Large Web Apps	Parallel Processing Apps	Web Apps with Back-end Processing	Store Blob Data
GoGrid, Flexiscale, Others	X	X				
Amazon Web Services	X	X	X	X	X	X
Windows Azure 2009 July CTP		X	X	X	X	X
Google AppEngine			X			
Salesforce.com Force Platform			X			

25

**WHAT**



花精靈-小球

**WHEN**



花精靈-小葵

**HOW**



花精靈-蘭兒

**WHY**



花精靈-小鬱

**WHO**



花精靈-百兒

# How can we build our Private Cloud ??

那我們如何打造私有雲端呢 ??

Public Cloud

公用雲端



Microsoft

Google

Target Market

is **S.M.B.**

主要客戶為

中小企業

Hybrid  
Cloud

以大型企業  
為主要客戶

**Enterprise** is  
key market

Community Cloud

社群雲端



私有雲端

Private Cloud

Academia 學術為主

# Reference Cloud Architecture

## 雲端運算的參考架構

### 應用軟體 Application

Social Computing, Enterprise, ISV, ...

### 程式語言 Programming

Web 2.0 介面, Mashups, Workflows, ...

### 控制管理 Control

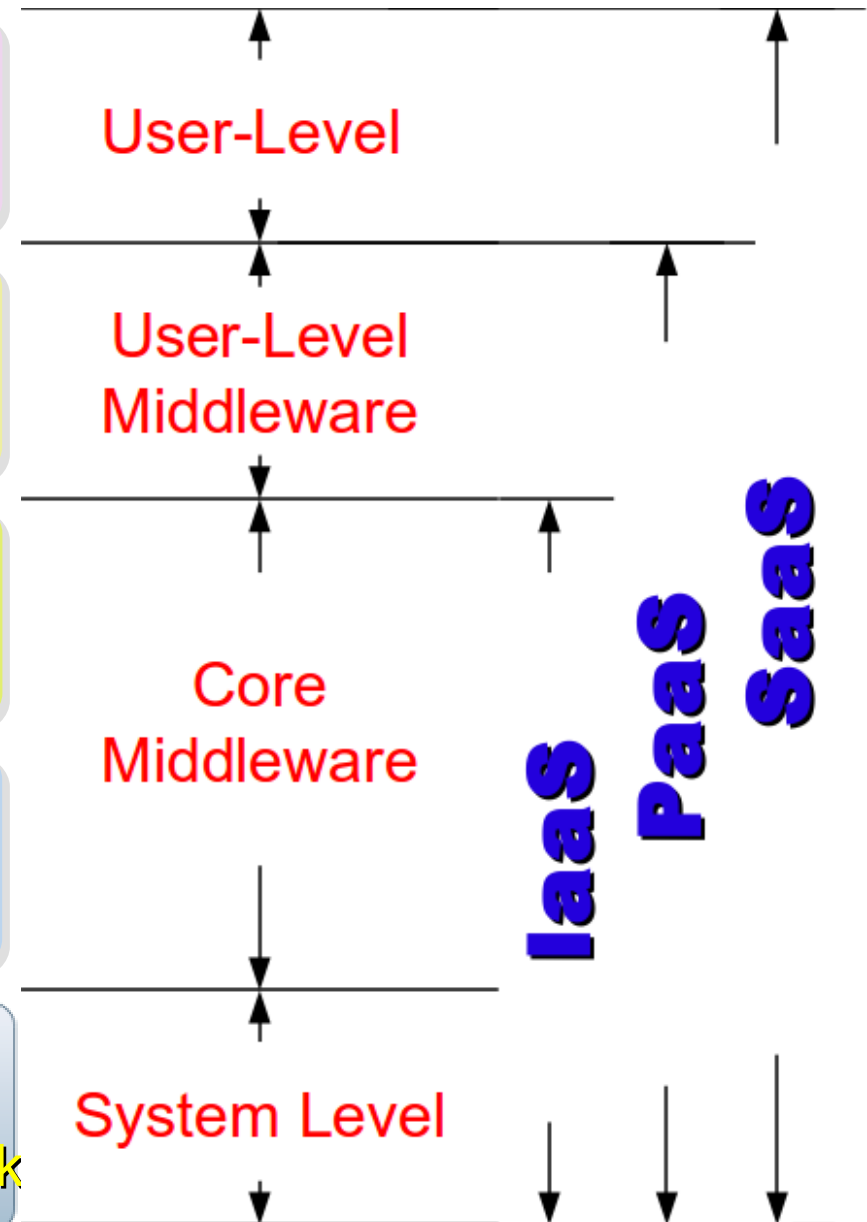
Qos Negotiation, Admission Control, Pricing, SLA Management, Metering...

### 虛擬化 Virtualization

VM, VM management and Deployment

### 硬體設施 Hardware

Infrastructure: Computer, Storage, Network



# Open Source to build Private Cloud

## 建構私有雲端的自由軟體

### 應用軟體 Application

Social Computing, Enterprise, ISV, ...

**eyeOS, Nutch, ICAS, X-RIME, ...**

### 程式語言 Programming

Web 2.0 介面, Mashups, Workflows, ...

**Hadoop (MapReduce), Sector/Sphere, AppScale**

### 控制管理 Control

Qos Negotiation, Admission Control, Pricing, SLA Management, Metering...

**OpenNebula, Enomaly, Eucalyptus, OpenQRM, ...**

### 虛擬化 Virtualization

VM, VM management and Deployment

**Xen, KVM, VirtualBox, QEMU, OpenVZ, ...**

### 硬體設施 Hardware

Infrastructure: Computer, Storage, Network



# 雲端運算的三大關鍵技術

Part 2 : Overview of Cloud Computing Core Technologies

**Jazz Wang**

**Yao-Tsung Wang**

**jazz@nchc.org.tw**



Powered by **DRBL**



# What are the trend of next 10 years ?

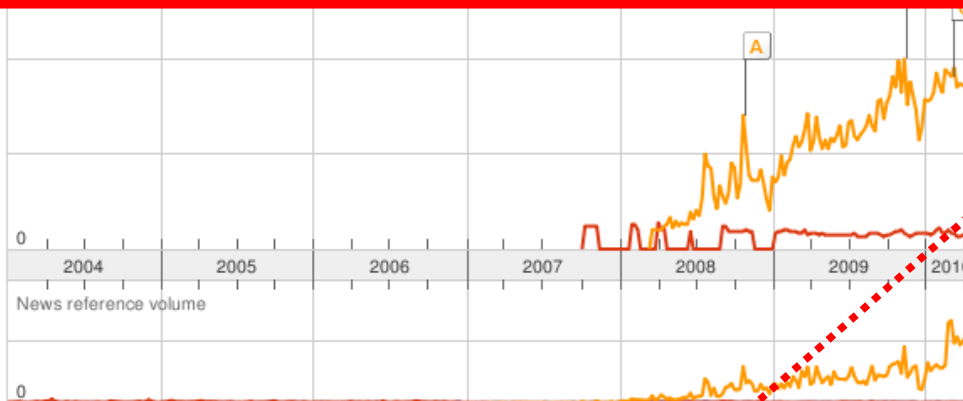
## 什麼是下個十年的熱門工作技能？

● distributed computin... ● grid computing ● cloud computing

[Sign in](#) to see and export additional Tren

All regions All years

Search Volume index



Rank by cloud computing

### Regions

1. [India](#)
2. [Singapore](#)
3. [South Korea](#)
4. [Hong Kong](#)
5. [Taiwan](#)
6. [Ireland](#)

### Cities

1. Bangalore, India
2. Mahape, India
3. Mumbai, India
4. Chennai, India
5. San Jose, CA, USA
6. Delhi, India

### Regions

1. [India](#)
2. [Singapore](#)
3. [South Korea](#)
4. [Hong Kong](#)
5. [Taiwan](#)
6. [Ireland](#)

似乎亞洲國家特別熱愛雲端?! Too Hot in Asia ?!

# Are the trends telling the truth ?

你確定沒有被圖表晃點嗎？

Searches [Websites](#) All regions

cloud computing does not have enough search volume for ranking

Scale is based on the average worldwide traffic of **iphone** in all years. [Learn more](#)

cloud computing 0 **iphone** 1.00 **ipad** 0.10 **android** 0.10 **chrome** 0.25



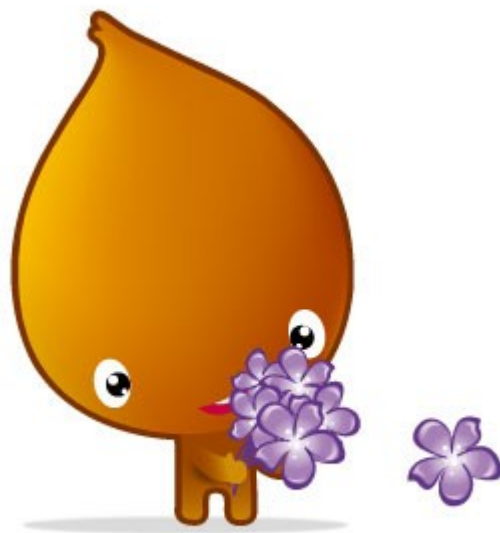
- A** [iPhone 3G Success!!!](#)  
Dallas Morning News - Jul
  - B** [iPod...iPhone...now, iPad](#)  
Economic Times - Jan 27
  - C** [Apple unveils the "iPad"](#)  
AFP - Jan 28 2010
  - D** [iPad Gag Apps Missing: N](#)  
PC World - Apr 5 2010
  - E** [Apple unveils iPhone 4 and OS4](#)  
Myjoyonline.com - Jun 7 2010
  - F** ['iPhone 4 major leap on iP](#)  
Times of India - Jun 24 2010
- [More news results »](#)

雲端運算

Regions

Cities

Languages



種子-芽比

**Let's Take a look at the Trends of Cloud**  
**透過觀察雲端的趨勢，來歸納雲端關鍵技術**

# Trend #1: Data are moving to the Cloud

趨勢一：資料開始回歸集中管理

Access data anywhere anytime 為了隨時存取

Reduce the risk of data lost 降低資料遺失風險

Reduce data transfer cost 減少資料傳輸成本

Enhance team collaboration 促進團隊協同合作

**How to store huge data ?!**

如何儲存大量資料呢 ?!

**Trend #2: Web become default Platform!**

**趨勢二：網頁變成預設開發平台**

**Open Standard** 網頁是開放標準

**Open Implementation** 實作不受壟斷

**Cross Platform** 瀏覽器成爲跨平台載具

**Web Application** 網頁程式設計成爲顯學

**Browser difference become entry barrier ?!**

**瀏覽器的差異造成新的技術門檻 ?!**

**Trend #3: HPC become a new industry**

**趨勢三：高速計算已悄悄變成新興產業**

**Parallel Computing** 平行運算的技能

**Distributed Computing** 分散運算的技能

**Multi-Core Programming** 多核心程式設計

**Processing Big Data** 處理大資料的技能

**Education and Training are needed !!**

**爲了讓這些技能與產業接軌，亟需教育訓練 !!**

## 善用雲端架構 打造企業人才庫

對於雲端的運用，多半仍停留在創造新商機的層次，然而善用雲端運算，可以替組織創造更多業務、行銷和人才培訓的機會。

作者：麥肯錫 出處：天下雜誌

過去五年，麥肯錫觀察重要科技發展，其中雲端、大量資訊 (big data)、智慧裝置 (smart assets) 三項，以超乎想像的速度發展。這三大技術，帶來五大趨勢，可被應用在企業營運及組織運作。先分別來看這三項技術：

第一、雲端運算。「雲端」在台灣已被一般民眾熟知。但我認為大家多半仍停留在雲端運算如何能創造新商機，卻很少好好思索，該怎麼運用雲端運算來替組織創造更多機會。特別是服務提供者，譬如電信業者、有線電視業者等，都應更有效應用雲端運算，為業務帶來更多機會。

第二、大量資訊。目前，絕大多數台灣企業，分析大量龐雜資料，仍使用類似微軟工具如 excel 等來整理。事實上，大量資訊經過快速運算分析，能更省時、省費用、有效的進行行銷活動。

第三、智慧裝置。如何善用監控器、智慧電表這類智慧裝置，來更優化公司營運。

參考來源：善用雲端架構 打造企業人才庫，作者：麥肯錫，出處：天下雜誌 455 期 (2010/09)  
<http://www.cw.com.tw/article/print.jsp?id=41776>

### 雲端運算

### 大量資訊分析

### 智慧裝置

# New Data Science : Social Network + Realtime Search

## 當「社交網路」遇上「即時搜尋」 = 即時市場行銷分析

### 創意行銷 / 臉書行銷 每天400萬顧客在線上

【經濟日報/潘俊琳】

2010.10.11 02:20 am

社交網站臉書Facebook的興起，重新定義了網路行銷的概念，大量的人潮讓業者彷彿看到滾滾錢潮，但臉書「開放平台」的模式，讓習慣有規則可循的行銷業者，必須開始學習全新的社群行銷，試著擁抱這項利器並串連消費者。

根據美國comScore的統計，美國網友8月分共花了1,140萬分鐘在臉書上，首次超越停留在Google旗下網站的時間，而臉書全球已經有超過5億的使用者，其中有35%的人每天登入。

#### 快速分享 即時知道顧客反應

聖洋科技執行長邱繼弘表示，台灣臉書每個月約有700萬的累計使用人次，以60%每天上臉書的人口來算，就有420 萬人天天上線。

邱繼弘指出，臉書最大的行銷價值在於「開放平台」，只要符合它的基本規範，任何人、任何公司都可以在上面「免費」發揮自己的行銷創意。過去想要利用網路行銷，企業必須自己架站，林林總總的後台建設非常繁瑣，有多少人會來也是個問號？

但臉書幫企業解決了後台建設以及人潮，不論是企業或個人，只要成立自己的「粉絲專頁」，然後發揮行銷創意，回收可能比自己架站還更豐碩。因為臉書玩家只要在粉絲專頁按「讚」，就成為「粉絲團」的一員，往後企業發布在粉絲專頁的訊息，所有粉絲團成員都會收到，如果粉絲團的成員覺得某個行銷訊息不錯，只要按「分享」這個訊息就會出現在粉絲個人的臉書上，他所有的朋友就會看到這則行銷訊息，這是目前最高明的病毒式行銷。

社交網路

即時搜尋

評價排行榜



參考來源：創意行銷 / 臉書行銷 每天 400 萬顧客在線上

【經濟日報 / 潘俊琳】

<http://udn.com/NEWS/FINANCE/FIN11/5901891.shtml>



## 2011 年 10 大策略科技

科技	影響
雲端運算	大型企業將會在 2012 年成立動態採購小組，專門負責雲端運算相關的決定以及管理。
媒體平板以及行動應用	2010 年將會有 12 億人使用具備上網能力的手機。隨著行動上網裝置以及應用程式日趨普及，與地點(location)、動作(motion)相關的應用軟體，可望進一步推動裝置的銷售。
社交溝通以及協作 (collaboration)	多數的公司在 2016 年已經把社交科技整合至多數的企業應用中，整合的範圍包含內部社交 CRM、溝通及協作以及外部社交網站。
影片	2013 年每位工作者看到的內容中，將有 25% 都是照片、影音。
次世代分析	隨著電腦、行動裝置運算能力、連結能力更強，影響企業如何決策，SAS 是長期領導廠商，IBM 以及甲骨文(Oracle) 事後起之秀。
社交分析	衡量人、主題以及想法的關係，範圍不限於社交網路，IBM 預計在 2011 年成為該領域的主要廠商之一。
情境感知運算 (context-aware computing)	較人工智慧更為寬廣，預計在 2013 年時 Fortune 500 大企業中超過半數會有相關採用方案。
儲存等級記憶體 (storage class memory)	快閃記憶體在消費性裝置、娛樂設備中的使用更多。
無所不在的運算 (ubiquitous computing)	儘管 Gartner 已經提及這個概念許多年，但隨著手機、射頻晶片更為普及，越多的物件可以連上網路。
架構化(fabric-based) 的基礎建設以及電腦	運算能力模組化，系統可以透過不同的模組來建構，可望提升效能。

資料來源：DIGITIMES 整理，2010/10

製表：雷佳宜、李盈瑩

雲端運算

平板行動應用

社交溝通協作

多媒體內容

次世代分析

社交分析

情境感知運算

儲存等級記憶體

無所不在的運算

模組化基礎建設

Source : <http://www.gartner.com/it/page.jsp?id=1454221>

Source : [http://www.digitimes.com.tw/tw/dt/n/shwnws.asp?CnId=4&cat=400&cat1=20&id=0000205798\\_CUZ63ZS3LCRY7E7UBK6V8](http://www.digitimes.com.tw/tw/dt/n/shwnws.asp?CnId=4&cat=400&cat1=20&id=0000205798_CUZ63ZS3LCRY7E7UBK6V8)

# 端

平板行動應用

社交溝通協作

多媒體內容

次世代分析

社交分析

情境感知運算

儲存等級記憶體

無所不在的運算

模組化基礎建設

雲端運算

**SaaS :**  
**Web 2.0**

**PaaS :**  
**Big Data**

**IaaS :**  
**Virtualization**

社交網路

評價排行榜

即時搜尋

智慧裝置

大量資訊分析

雲端運算

# 雲

# Two Type of Cloud Architecture ?

雲端架構的兩大陣營？

**SaaS :**  
**Web 2.0**

**IaaS :**  
**Virtualization**



想盡辦法誘你用計算跟網路  
**Computing Intensive**

**SaaS :**  
**Web 2.0**

**PaaS :**  
**Big Data**



想盡辦法誘你提供資料作分析  
**Data Intensive**



# 深入解析雲端虛擬化技術

Part 2-1 : Deep Dive into Virtualization Technologies

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



Powered by DRBL

# 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* ”

# How hot is Cloud and Virtualization ?

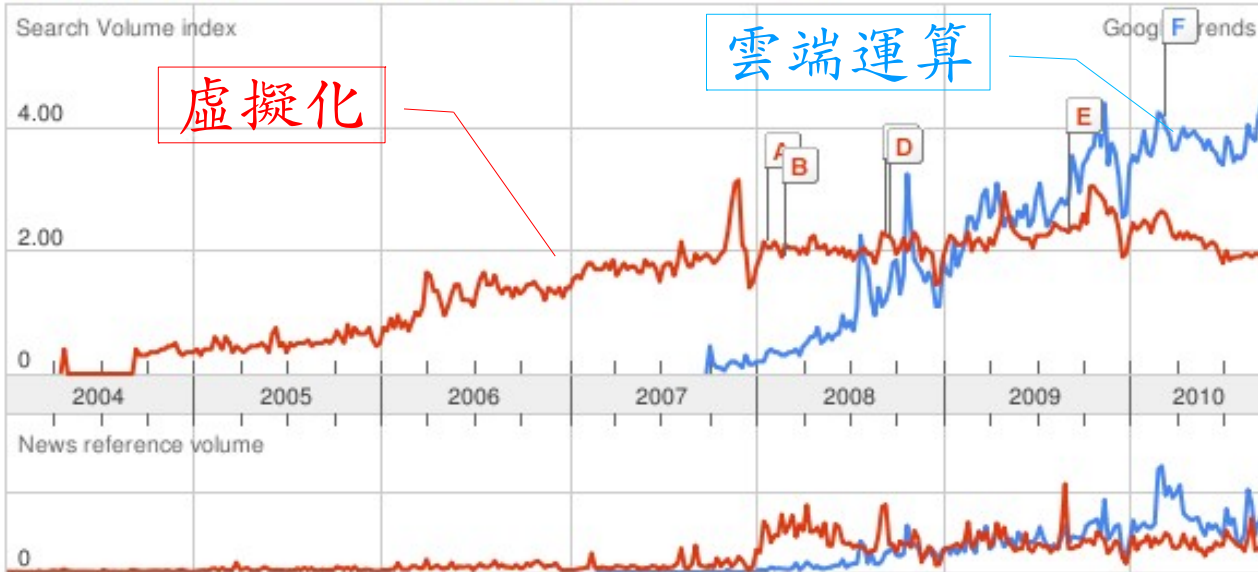
## 雲端運算與虛擬化有多熱門呢？

Searches [Websites](#)

All regions

Scale is based on the average worldwide traffic of [cloud computing](#) in all years. [Learn more](#)

cloud computing — 1.00 virtualization — 1.45



- A** [Microsoft joins 'virtualization' push](#)  
Globe and Mail - Jan 22 2008
- B** [Microsoft enters virtualization with Windows](#)  
Canada.com - Feb 27 2008
- C** [Microsoft Launches Virtualization Blitzkrieg](#)  
CRN - Sep 8 2008
- D** [CA to Deliver Integrated Virtualization Manag Program](#)  
WELT ONLINE - Sep 15 2008
- E** [Wyse and VMware Users and Channel Part](#)  
Reuters - Aug 31 2009
- F** [Google looks to be 'cloud-computing' rainma](#)  
Winnipeg Free Press - Mar 10 2010

[More news results »](#)

Rank by

### Regions

1. <a href="#">India</a>	<div style="width: 100%;"><div style="width: 100%; height: 10px; background-color: blue;"></div><div style="width: 80%; height: 10px; background-color: red;"></div></div>
2. <a href="#">Singapore</a>	<div style="width: 100%;"><div style="width: 100%; height: 10px; background-color: blue;"></div><div style="width: 80%; height: 10px; background-color: red;"></div></div>
3. <a href="#">Hong Kong</a>	<div style="width: 100%;"><div style="width: 100%; height: 10px; background-color: blue;"></div><div style="width: 80%; height: 10px; background-color: red;"></div></div>
4. <a href="#">South Korea</a>	<div style="width: 100%;"><div style="width: 100%; height: 10px; background-color: blue;"></div><div style="width: 80%; height: 10px; background-color: red;"></div></div>
5. <a href="#">Taiwan</a>	<div style="width: 100%;"><div style="width: 100%; height: 10px; background-color: blue;"></div><div style="width: 80%; height: 10px; background-color: red;"></div></div>
6. <a href="#">Malaysia</a>	<div style="width: 100%;"><div style="width: 100%; height: 10px; background-color: blue;"></div><div style="width: 80%; height: 10px; background-color: red;"></div></div>
7. <a href="#">United States</a>	<div style="width: 100%;"><div style="width: 100%; height: 10px; background-color: blue;"></div><div style="width: 80%; height: 10px; background-color: red;"></div></div>

### Cities

1. Bangalore, India	<div style="width: 100%;"><div style="width: 100%; height: 10px; background-color: blue;"></div><div style="width: 80%; height: 10px; background-color: red;"></div></div>
2. Mahape, India	<div style="width: 100%;"><div style="width: 100%; height: 10px; background-color: blue;"></div><div style="width: 80%; height: 10px; background-color: red;"></div></div>
3. Mumbai, India	<div style="width: 100%;"><div style="width: 100%; height: 10px; background-color: blue;"></div><div style="width: 80%; height: 10px; background-color: red;"></div></div>
4. Chennai, India	<div style="width: 100%;"><div style="width: 100%; height: 10px; background-color: blue;"></div><div style="width: 80%; height: 10px; background-color: red;"></div></div>
5. Delhi, India	<div style="width: 100%;"><div style="width: 100%; height: 10px; background-color: blue;"></div><div style="width: 80%; height: 10px; background-color: red;"></div></div>
6. San Jose, CA, USA	<div style="width: 100%;"><div style="width: 100%; height: 10px; background-color: blue;"></div><div style="width: 80%; height: 10px; background-color: red;"></div></div>
7. Singapore, Singapore	<div style="width: 100%;"><div style="width: 100%; height: 10px; background-color: blue;"></div><div style="width: 80%; height: 10px; background-color: red;"></div></div>

### Language

1. Korean	<div style="width: 100%;"><div style="width: 100%; height: 10px; background-color: blue;"></div><div style="width: 80%; height: 10px; background-color: red;"></div></div>
2. English	<div style="width: 100%;"><div style="width: 100%; height: 10px; background-color: blue;"></div><div style="width: 80%; height: 10px; background-color: red;"></div></div>
3. Dutch	<div style="width: 100%;"><div style="width: 100%; height: 10px; background-color: blue;"></div><div style="width: 80%; height: 10px; background-color: red;"></div></div>
4. German	<div style="width: 100%;"><div style="width: 100%; height: 10px; background-color: blue;"></div><div style="width: 80%; height: 10px; background-color: red;"></div></div>
5. Chinese	<div style="width: 100%;"><div style="width: 100%; height: 10px; background-color: blue;"></div><div style="width: 80%; height: 10px; background-color: red;"></div></div>
6. Portuguese	<div style="width: 100%;"><div style="width: 100%; height: 10px; background-color: blue;"></div><div style="width: 80%; height: 10px; background-color: red;"></div></div>
7. Thai	<div style="width: 100%;"><div style="width: 100%; height: 10px; background-color: blue;"></div><div style="width: 80%; height: 10px; background-color: red;"></div></div>

# 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

Database Virtualization

資料虛擬化

Data Virtualization

# Open Source for Virtualization

## 虛擬化技術對應的自由軟體

Application Virtualization  
應用程式虛擬化

Ex. VMWare ThinApp

Desktop Virtualization  
桌面虛擬化

Redhat SPICE

Presentation Virtualization  
顯示虛擬化

VNC, FreeNX

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, 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

OpenNebula, Eucalyptus

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, SDFS, ZFS

# Building IaaS with Open Source

## 用自由軟體打造 IaaS 服務

應用軟體 Application  
Social Computing, Enterprise, ISV, ...

eyeOS, Nutch, ICAS,  
X-RIME, ...

程式語言 Programming  
Web 2.0 介面, Mashups, Workflows, ...

Hadoop (MapReduce),  
Sector/Sphere, AppScale

控制管理 Control  
Qos Negotiation, Admission Control,  
Pricing, SLA Management, Metering...

OpenNebula, Enomaly,  
Eucalyptus, OpenQRM, ...

虛擬化 Virtualization  
VM, VM management and Deployment

Xen, KVM, VirtualBox,  
QEMU, OpenVZ, ...

硬體設施 Hardware  
Infrastructure: Computer, Storage,  
Network

- 原是加州大學聖塔芭芭拉分校 (UCSB) 的研究專案
- **It was a research project of UCSB, USA**
- 目前已轉由 Eucalyptus System 這間公司負責維護
- **Now Eucalyptus System provide technical supports.**
- 創立目的是讓使用者可以**打造自己的 EC2**
- **It designed to help user to build their own Amazon EC2**
- 特色是相容於 Amazon EC2 既有的用戶端介面
- **Its feature is compatible with existing EC2 client.**
- 優勢是 Ubuntu 9.04 已經收錄 Eucalyptus 的套件
- **Ubuntu Enterprise Cloud powered by Eucalyptus in 9.04**
- 目前有提供 Eucalyptus 的官方測試平台供註冊帳號
- **You can register trail account at <http://open.eucalyptus.com/>**
- 缺點：目前仍有部分操作需透過指令模式
- **Cons : you might need to type commands in some case**



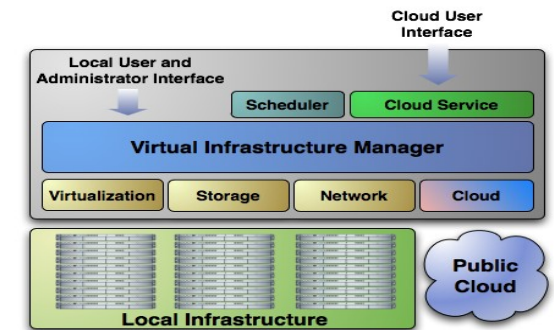
Eucalyptus

關於 Eucalyptus 的更多資訊，請參考

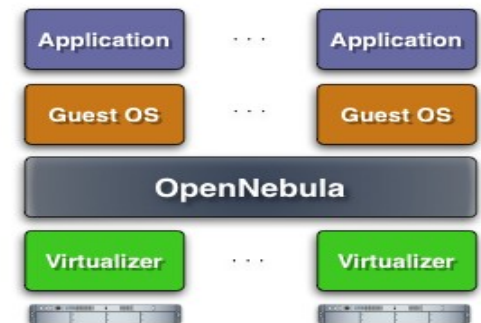
<http://trac.nchc.org.tw/grid/wiki/Eucalyptus> 60

- <http://www.opennebula.org>
- 由歐洲研究學會 (European Union FP7 ) 贊助
- **Sponsor by European Union FP7**
- 將實體叢集轉換成具管理彈性的虛擬基礎設備
- Turn Physical Cluster into Virtual Cluster
- 可管理**虛擬叢集**的狀態、排程、遷徙 (migration)
- **manage status, scheduling and migration of virtual cluster**
- [Ubuntu 9.04 provide package of opennebula](#)
- 缺點：需下指令來進行虛擬機器的遷徙 (migration)。
- **Cons** : You need to type commands to check or migration

OpenNebula.org



關於 OpenNebula 的更多資訊，  
請參考 <http://trac.nchc.org.tw/grid/wiki/OpenNEbula>



# 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

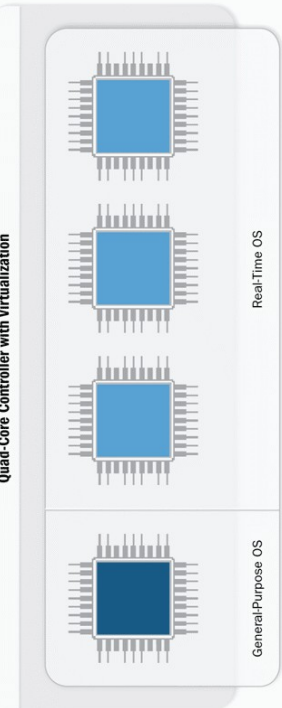
co-existence of different OS environment  
on the same platform

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



參考來源：**Embedded systems virtualization: Consider a Hypervisor**

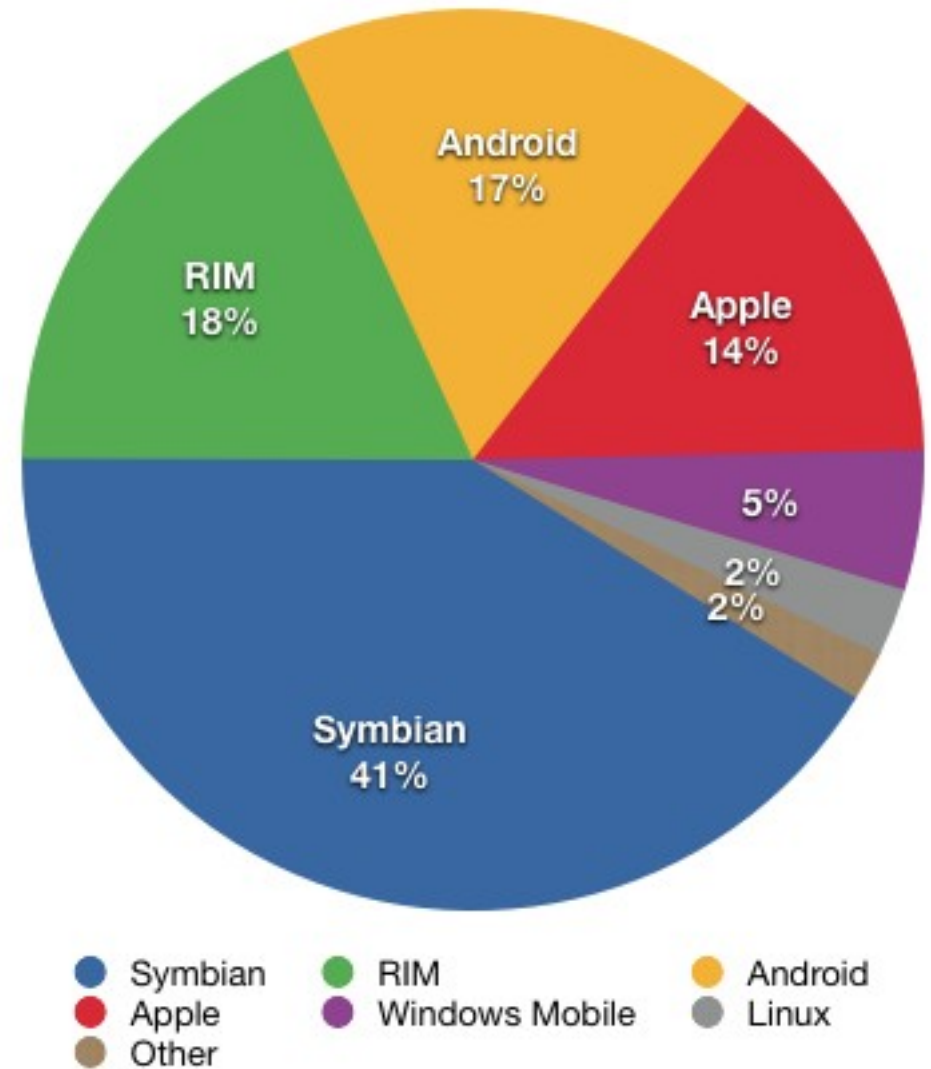
<http://www.eetimes.com/design/automotive-design/4016811/Embedded-systems-virtualization-Consider-a-Hypervisor>

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](http://www.iphonic.tv/iphone_screen_amaze.png)

圖片來源：[http://en.wikipedia.org/wiki/File:Smartphone\\_share\\_current.png](http://en.wikipedia.org/wiki/File:Smartphone_share_current.png)

參考來源：**iPhone - Playing mame4iphone 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/>

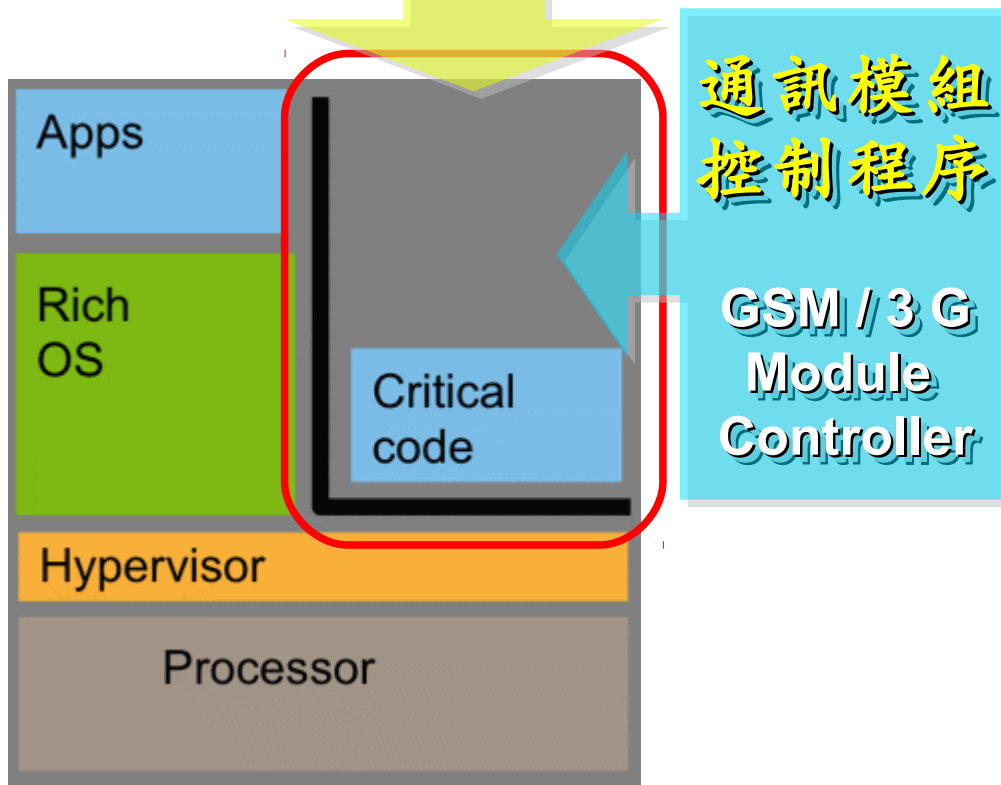
<http://www.youtube.com/watch?v=Nvj4ObHmxCI>



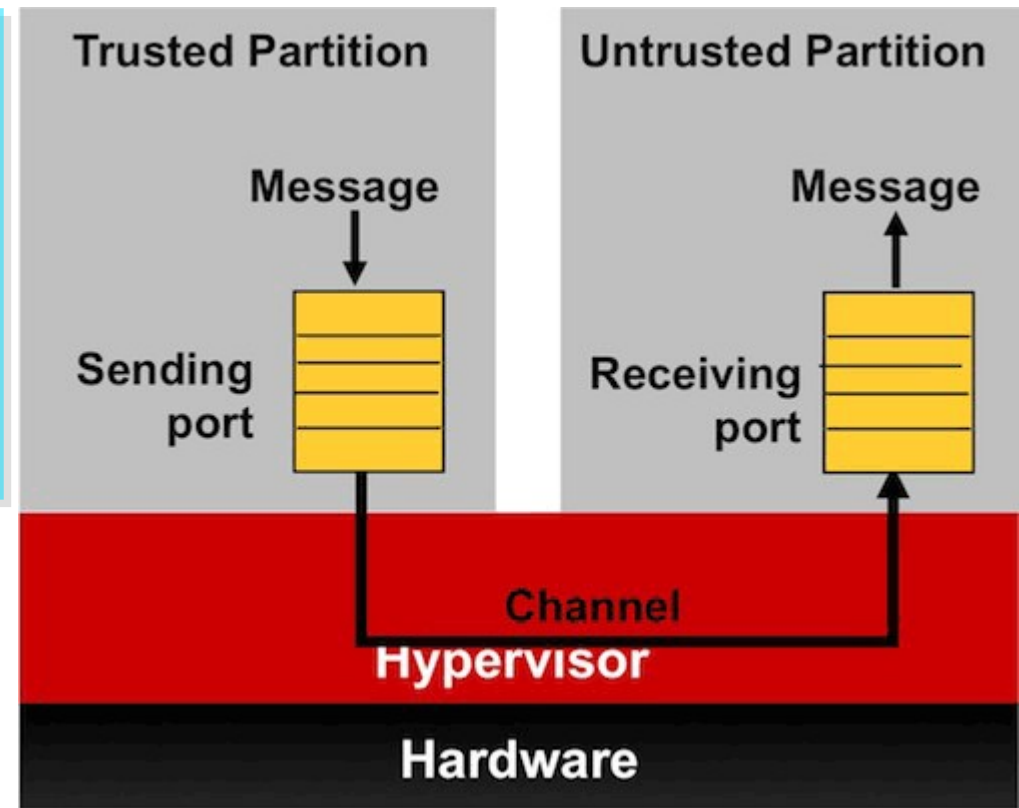
# Virtualization for Embedded Security

## 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**

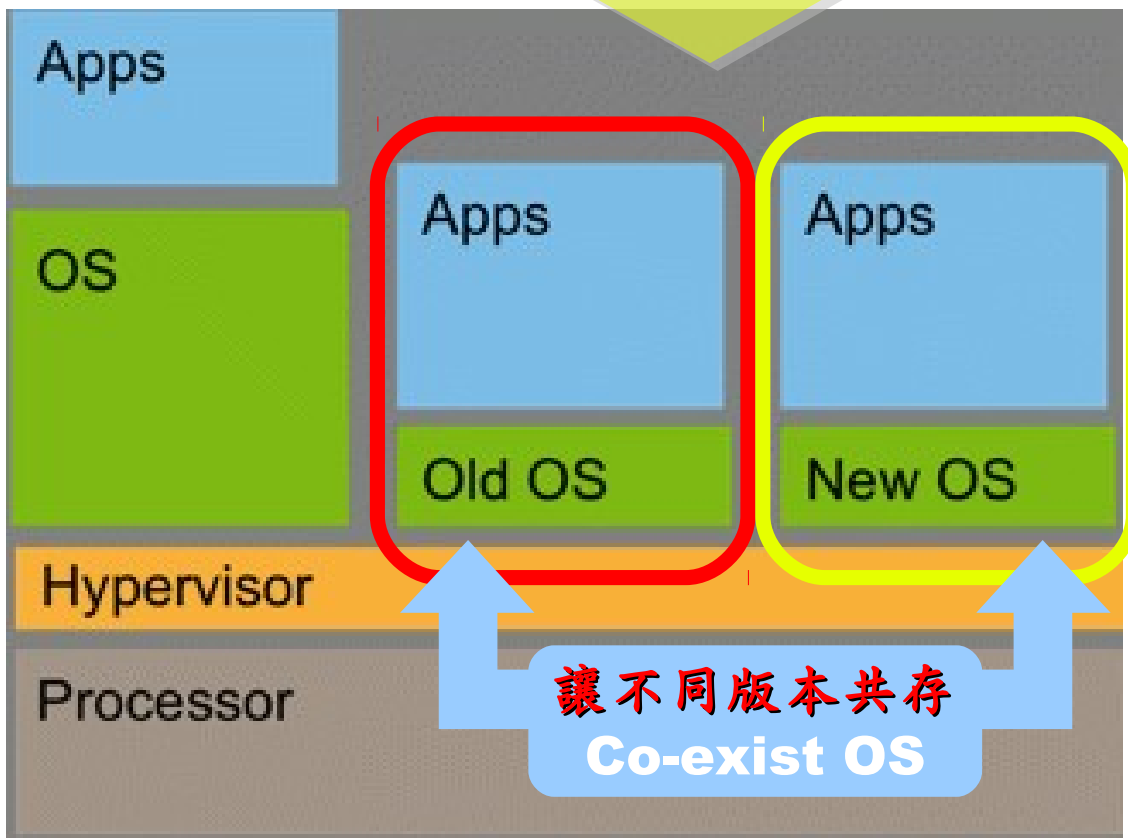
<http://embeddedinnovator.com/securing-smart-grid-devices>

# 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



**\*#06#**

參考來源：**Embedded systems virtualization: Consider a Hypervisor**

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

# 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>

### **KVM** for ARM

<http://www.ncl.cs.columbia.edu>

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](#)

EmbeddedXEN is an academic project and consists of a single multi-kernel ARM-target binary image with bedded

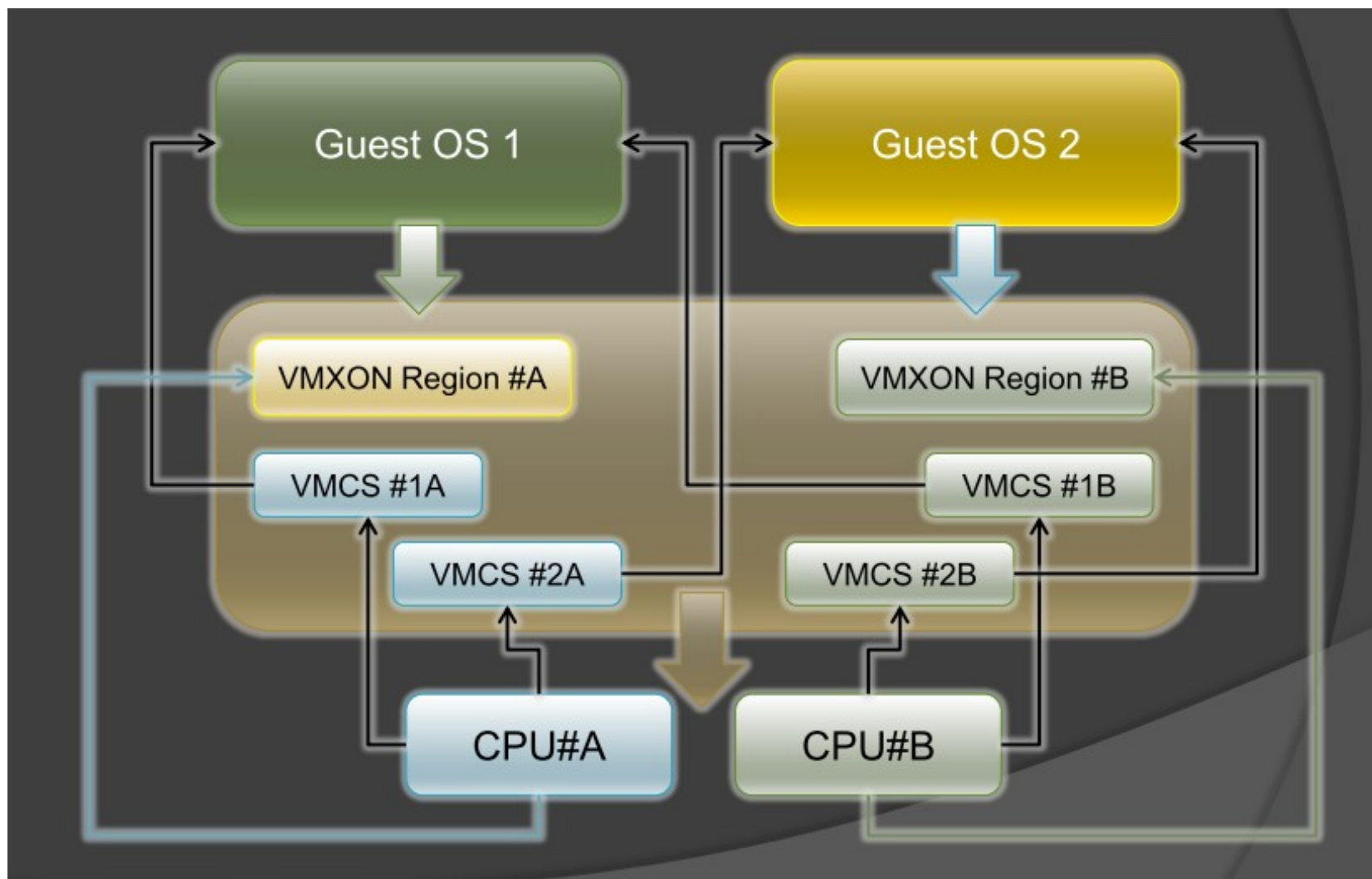
EDIT

[view project details](#)

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

# 虛擬化衍生的新興資安問題

透過虛擬機器，竊取鍵盤輸入、植入後門 .....



圖片來源： Hacks in Taiwan Conference 2010

[http://www.hitcon.org/hit2010/download/6\\_New%20Battlefield%20For%20Malware%20Game.pdf](http://www.hitcon.org/hit2010/download/6_New%20Battlefield%20For%20Malware%20Game.pdf)

王大寶 & PK / Hypervisor - New Battlefield For Malware Game 虛擬機 - 惡意程式攻防的新戰場



# 深入解析雲端大量資料分析技術

Part 2-2 : Deep Dive into Data Science Technologies

**Jazz Wang**

**Yao-Tsung Wang**

**jazz@nchc.org.tw**



Powered by DRBL

# Building PaaS with Open Source

## 用自由軟體打造 PaaS 雲端服務

應用軟體 Application  
Social Computing, Enterprise, ISV, ...

eyeOS, Nutch, ICAS,  
X-RIME, ...

程式語言 Programming  
Web 2.0 介面, Mashups, Workflows, ...

Hadoop (MapReduce),  
Sector/Sphere, AppScale

控制管理 Control  
Qos Negotiation, Admission Control,  
Pricing, SLA Management, Metering...

OpenNebula, Enomaly,  
Eucalyptus, OpenQRM, ...

虛擬化 Virtualization  
VM, VM management and Deployment

Xen, KVM, VirtualBox,  
QEMU, OpenVZ, ...

硬體設施 Hardware  
Infrastructure: Computer, Storage,  
Network

# 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/>



# Hadoop

- <http://hadoop.apache.org>
- Hadoop 是 Apache Top Level 開發專案
- **Hadoop is Apache Top Level Project**
- 目前主要由 Yahoo! 資助、開發與運用
- **Major sponsor is Yahoo!**
- 創始者是 Doug Cutting，參考 Google Filesystem
- **Developed by Doug Cutting, Reference from Google Filesystem**
- 以 Java 開發，提供 HDFS 與 MapReduce API。
- **Written by Java, it provides HDFS and MapReduce API**
- 2006 年使用在 Yahoo 內部服務中
- **Used in Yahoo since year 2006**
- 已佈署於上千個節點。
- **It had been deploy to 4000+ nodes in Yahoo**
- 處理 Petabyte 等級資料量。
- **Design to process dataset in Petabyte**



**Facebook, Last.fm,  
Joost, Twitter  
are also powered  
by Hadoop**

# Sector / Sphere

- <http://sector.sourceforge.net/>
- 由美國資料探勘中心研發的自由軟體專案。
- **Developed by National Center for Data Mining, USA**
- 採用 C/C++ 語言撰寫，因此效能較 Hadoop 更好。
- **Written by C/C++, so performance is better than Hadoop**
- 提供「類似」Google File System 與 MapReduce 的機制
- **Provide file system similar to Google File System and MapReduce API**
- 基於 [UDT 高效率網路協定](#) 來加速資料傳輸效率
- **Based on UDT which enhance the network performance**
- [Open Cloud Testbed](#) 有提供測試環境，並開發 [Ma1Stone 效能評比軟體](#)
- **Open Cloud Consortium provide Open Cloud Testbed and develop Ma1Stone toolkit for benchmark**



National Center for Data Mining  
University of Illinois at Chicago



Open Data Group  
<http://www.opendatagroup.com/>

# What is Hadoop ?

用一句話解釋 **Hadoop** 是什麼??

*Hadoop is a **software platform** that lets one easily write and run applications that **process vast amounts of data.***

*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。

# Features of Hadoop ...

## Hadoop 這套軟體的特色是 ...

- **海量 Vast Amounts of Data**
  - 擁有儲存與處理大量資料的能力
  - Capability to **STORE** and **PROCESS** vast amounts of data.
- **經濟 Cost Efficiency**
  - 可以用在由一般 PC 所架設的叢集環境內
  - Based on large clusters built of **commodity hardware**.
- **效率 Parallel Performance**
  - 透過分散式檔案系統的幫助，以致得到快速的回應
  - With the help of HDFS, Hadoop **have better performance**.
- **可靠 Robustness**
  - 當某節點發生錯誤，能即時自動取得備份資料及佈署運算資源
  - Robustness to add and remove computing and storage resource without shutdown entire system.

# Founder of Hadoop – Doug Cutting

**Hadoop** 這套軟體的創辦人 **Doug Cutting**

Doug Cutting Talks About The Founding Of Hadoop

clouderahadoop

9 部影片

編輯訂閱項目



Doug Cutting Talks About The Founding Of Hadoop

<http://www.youtube.com/watch?v=qxC4urJOchs>

# History of Hadoop ... 2002~2004

## Hadoop 這套軟體的歷史源起 ... 2002~2004



- Lucene

- <http://lucene.apache.org/>
- 用Java 設計的高效能文件索引引擎API
- a high-performance, full-featured **text search engine library** written entirely in **Java**.
- 索引文件中的每一字，讓搜尋的效率比傳統逐字比較還要高的多
- Lucene create an **inverse index** of every word in different documents. It enhance performance of text searching.

# History of Hadoop ... 2002~2004

## Hadoop 這套軟體的歷史源起 ... 2002~2004

- Nutch



- <http://nutch.apache.org/>
- Nutch 是基於開放原始碼所開發的網站搜尋引擎
- Nutch is open source **web-search** software.
- 利用Lucene 函式庫開發
- It builds on **Lucene and Solr**, adding web-specifics, such as a **crawler**, a **link-graph database**, parsers for HTML and other document formats, etc.



# History of Hadoop ... 2004 ~ Now

## **Hadoop** 這套軟體的歷史源起 ... 2004 ~ Now

- Nutch 後來遇到儲存大量網站資料的瓶頸，剛好看到 Google 在一些會議分享他們的三大關鍵技術 ...
- Added DFS & MapReduce implement to Nutch
- According to **user feedback** on the mail list of Nutch ....
- Hadoop became separated project **since Nutch 0.8**
- Nutch DFS → Hadoop Distributed File System (HDFS)
- **Yahoo** hire Dong Cutting to build a team of web search engine at **year 2006**.
  - Only **14 team members** (engineers, clusters, users, etc.)
- Dong Cutting joined Cloudera at year 2009.

**YAHOO!**

 cloudera





# Who Use Hadoop ??

有哪些公司在用 **Hadoop** 這套軟體 ??

- **Yahoo** is the key contributor currently.
- **IBM** and **Google** teach Hadoop in universities ...
- [http://www.google.com/intl/en/press/pressrel/20071008\\_ibm\\_univ.html](http://www.google.com/intl/en/press/pressrel/20071008_ibm_univ.html)
- **The New York Times** used **100 Amazon EC2 instances** and a Hadoop application to process **4TB of raw image TIFF data** (stored in S3) into **11 million finished PDFs** in the space of **24 hours** at a computation cost of about **\$240** (not including bandwidth)
  - from <http://en.wikipedia.org/wiki/Hadoop>
- <http://wiki.apache.org/hadoop/AmazonEC2>
- <http://wiki.apache.org/hadoop/PoweredBy>
  - A9.com
  - ADSDAQ by Contextweb
  - EHarmony
  - Facebook
  - Fox Interactive Media
  - IBM
  - ImageShack
  - ISI
  - Joost
  - Last.fm
  - Powerset
  - The New York Times
  - Rackspace
  - Veoh
  - Metaweb

# Performance improvement of Hadoop

## Hadoop 過去幾年的效能改進 (from Yahoo)

年份	日期	節點數	耗時 ( 小時 )
2006	四月	188	47.9
2006	五月	500	42
2006	十一月	20	1.8
2006	十一月	100	3.3
2006	十一月	500	5.2
2006	十一月	900	7.8
2007	七月	20	1.2
2007	七月	100	1.3
2007	七月	500	2
2007	七月	900	2.5

Sort benchmark, every nodes with terabytes data.

# Hadoop in production run ....

## 商業運轉中的 **Hadoop** 應用 ....

- **February 19, 2008**
- **Yahoo! Launches World's Largest Hadoop Production Application**
- <http://developer.yahoo.net/blogs/hadoop/2008/02/yahoo-worlds-largest-production-hadoop.html>

Number of links between pages in the index	roughly 1 trillion links
Size of output	over 300 TB, compressed!
Number of cores used to run single Map-Reduce job	over 10,000
Raw disk used in the production cluster	over 5 Petabytes

# Hadoop in production run ....

## 商業運轉中的 **Hadoop** 應用 ....

- **September 30, 2008**
- **Scaling Hadoop to 4000 nodes at Yahoo!**
- [http://developer.yahoo.net/blogs/hadoop/2008/09/scaling\\_hadoop\\_to\\_4000\\_nodes\\_a.html](http://developer.yahoo.net/blogs/hadoop/2008/09/scaling_hadoop_to_4000_nodes_a.html)

<b>Total Nodes</b>	<b>4000</b>
<b>Total cores</b>	<b>30000</b>
<b>Data</b>	<b>16PB</b>

	<b>500-node cluster</b>		<b>4000-node cluster</b>	
	<b>write</b>	<b>read</b>	<b>write</b>	<b>read</b>
<b>number of files</b>	990	990	14,000	14,000
<b>file size (MB)</b>	320	320	360	360
<b>total MB processes</b>	316,800	316,800	5,040,000	5,040,000
<b>tasks per node</b>	2	2	4	4
<b>avg. throughput (MB/s)</b>	<b>5.8</b>	<b>18</b>	<b>40</b>	<b>66</b>



# 雲端運算對政府部門之衝擊與因應對策

Part 3 : What is the Impacts and How should public sectore adjust ?

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

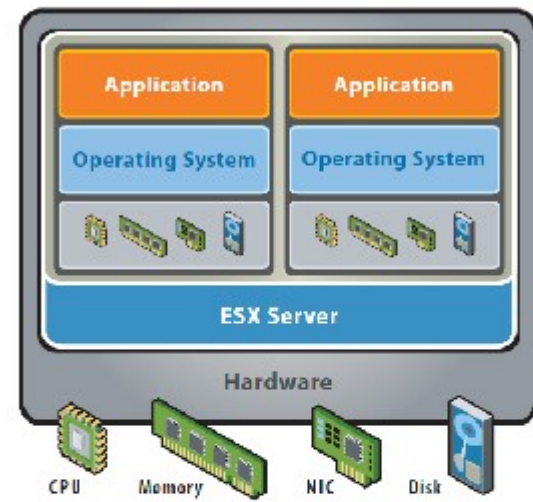


Powered by DRBL

# Adjust #1: Virtualization

## 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<sup>®</sup> 2008  
Hyper-V<sup>™</sup>



# Hidden Cost : Storage Virtualization

## 潛在成本：儲存虛擬化

- 資料整合為跨單位整合的第一步 !!
- 導入評量指標：
  - 資料所有權 / 散佈權 / 重覆比例
  - 資料成長量 = ?? PB/Year
  - 現有資料副本個數 / 災害復原能力
- 商業硬體方案：EMC、NetApp
- 商業軟體方案：IBM GPFS
- 自由軟體方案：
  - Lustre、ZFS、GlusterFS...
- 預期效益：
  - 減少資料重覆 (Data Deduplication)
  - 提供異地副本備份、災害復原機制



## Adjust #2: Web 2.0

# 導入雲端第二波主打：Web 2.0 !!

- Web 2.0 的精神在『使用者參與』
- 導入評量指標：
  - 行銷與市場調查的成本
  - 產品銷售通路的成本
  - 產品協同開發的需求
- 雲端服務解決方案：
  - **Salesforce.com** CRM
  - **Google Apps** for Business
  - **iPhone Apps, Android Apps ....**
- 預期效益：
  - 減少行銷與市場調查成本
  - 增加銷售管道與協同合作機制
  - 隨需行動服務新商機



App Store



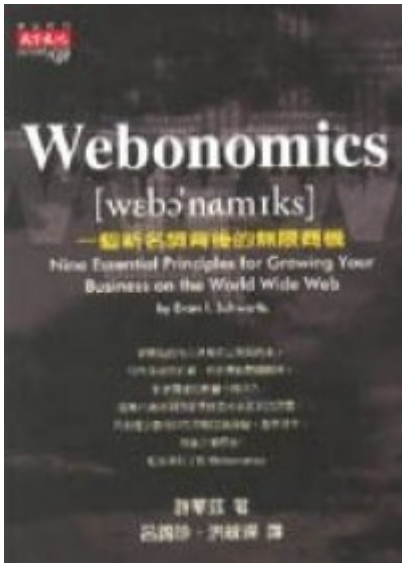
App Store

**NOKIA**





# Webonomics, The Attention Economy, Cloudonomics 網頁經濟、注意力經濟、雲端經濟：重點都在吸引眼球注意力



Mr.6團隊免費拜訪貴公司回答facebook行銷問題

歡迎學校與企業或社團邀請Mr.6團隊代表演講

Mr.6團隊徵求行銷助理中

作者：本名劉威麟，美國史丹佛電機、管理雙碩士，14歲移民加拿大，而後移居美國矽谷，大學開設兩個網站後，在互聯網經驗達13年，已出版12本書，返台後任職於美商創投，受聘領導Voofox網站公司任執行長，發起Sweet100網路正面價值推廣活動等，繼續耕耘最愛的網路產業，邀請Mr.6至企業進行網站顧問案或公開演講請來信至 [contact.mr6@gmail.com](mailto:contact.mr6@gmail.com)

[<< 觀看其他最新文章](#)

## 創意「舌刷」Orabrush如何靠YouTube行銷影片一年大賣3000萬元、113個國家

Like 229 people like this.

by Mr. 6 on October 11th, 2010, 目前有 9 則留言, [f 分享到塗爾潘](#)



參考來源：創意「舌刷」Orabrush如何靠YouTube行銷影片一年大賣3000萬元、113個國家

<http://mr6.cc/?p=5228>

## Adjust #3: Big Data

# 導入雲端第三波主打：Big Data！！

處理海量資料是雲端運算的精華

➢ 用前處理與資料量換取查詢時間

關鍵技術：

➢ 分散儲存檔案 → 在地平行運算

➢ 分散式資料庫 → Key-Value

商業解決方案：

➢ **Google** App Engine

➢ **GFS / MapReduce / BigTable**

➢ **Amazon** S3 / SimpleDB

➢ **Cloudera** Hadoop Distribution

預期效益：

➢ 降低檔案遺失風險

➢ 降低資料傳輸成本



# Big Data Analysis : Social Computing & Business Intelligence

## 「社交運算」與「商業智慧」均仰賴大量資料分析

**DIGITIMES** 網站內容的著作權為大椽股份有限公司 (DIGITIMES Inc.) 所有, 或其他授權DIGITIMES使用的內容提供者所有。

使用者下載或拷貝網站的內容或服務僅限於供個人、非商業用途之使用, 但不得以任何形式傳輸、重製、散布或提供予公眾。使用人利用時必須遵守著作權法的所有相關規定, 不可變更、發行、播送、轉賣、重製、改作、散布、表演、展示或利用DIGITIMES所屬網站上局部或全部內容及服務以賺取利益。

### 提升商業分析效果 資料倉儲業提倡資料社交化

2010/10/27 - DIGITIMES 馬培治 / 台北

社交運算(social computing)隨著包括Facebook在內的各式社交網絡服務持續發燒, 也成為企業資訊系統發展的重點之一, 繼IBM、微軟(Microsoft)與甲骨文(Oracle)等大廠提倡在應用軟體功能上支援社交功能之後, 資料倉儲(Data Warehouse)業者Teradata則提倡企業資料分析, 應納入包括社交資訊在內的多元因子, 讓不同資料源間的資料「社交化」(socialization of data), 以增強商業分析效果, 提高掌握用戶行為並輔助商業決策。

Teradata業務發展暨行銷執行副總裁Darryl D. McDonald於25日在自家全球合作夥伴與使用者大會上表示, 除了傳統企業營運資料, 各種可用來擷取資訊的資料源, 如RFID、智慧型裝置、社交網路, 乃至各種感應器, 將會對現今的企業分析帶來龐大的衝擊, 他建議企業可以開始著手思考, 如何將這些新興資料源的資料與傳統商業智慧分析的資料進行整合, 以期從更豐富的資料中, 找出過去商業分析方法看不到的隱性資訊。

McDonald表示, Facebook目前已經擁有超過5億個註冊用戶, 而推特(Tweeter)每天也有超過8,500萬條訊息產生, 若企業能夠將自身的用戶資訊或營運資料與這些龐大的資訊源進行有意義的分析, 將能夠激發在商業分析領域的創新應用。

他以參加Teradata全球合作夥伴暨使用者大會的3,000多名與會者為例進行分析, 發現這些與會者代表的公司總計具有9兆美元的資本額, 以及合計達230萬個線上社交網路服務的人際連結數, McDonald說, 這些資訊代表龐大的商機, 以及可供未來利用在業務推廣、行銷等目的使用。

參考來源：提升商業分析效果 資料倉儲業提倡資料社交化 (2010/10/27)

<http://goo.gl/2GoMo>

**Impact #1: IT Budget CUT !**  
**衝擊一：資訊基礎建設預算刪減！**

**Adjust #1: Virtualization 評估導入虛擬化技術**

**Impact #2: More Collaboration and Intergration !**  
**衝擊二：加速跨單位資訊整合與協同合作！**

**Adjust #2: Open Gov. Data 開放政府公用資料**

**Impact #3: Need Single Service Gateway !**  
**衝擊三：單一政府服務存取窗口！**

**Adjust #3: Single AAA mechanism 統一身分認證**

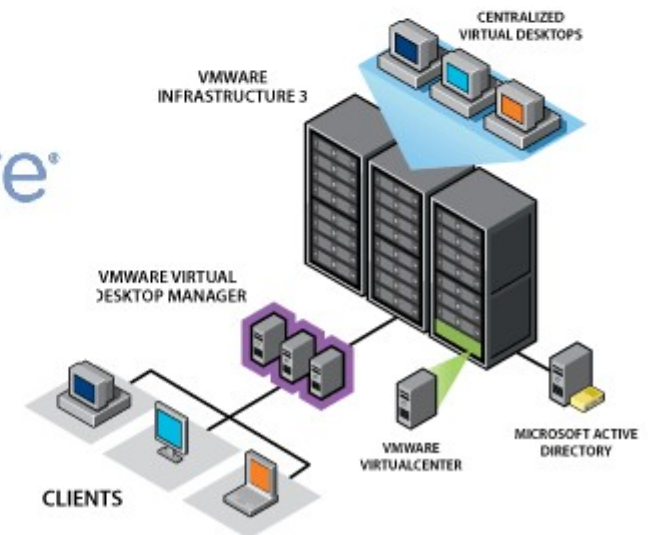
## Adjust #1: Virtualization

# 公部門導入雲端參考策略 (1) 評估導入虛擬化

- 關鍵驅動因素
  - 預算刪減？節約能源？提高現有硬體使用率？
- 導入評估工具
  - Microsoft Assessment and Planning (MAP) Toolkit for Hyper-V
  - Hyper-V 評估工具：收集 Server 使用情況，產生評估報告等
  - Open Source: MRTG, Ganglia, Nagios, ....
  - 蒐集 CPU 用量、記憶體用量、尖峰負載發生時間、批次排程時間
- 隱含額外成本
  - 若想獲得虛擬化帶來的好處（災害復原與動態負載平衡），需要額外建置共享的儲存設施。

# Thin Client and VDI for next IT procurement ? 未來改採購精簡型電腦與桌面虛擬化??

- 桌面虛擬化的需求越來越高??
- 導入評量指標：
  - 總電源開銷、空調、管理人力
  - 既存共用儲存設施
  - 資料機敏性高，不容外洩
- 商業解決方案：
  - **VMWare** Viewer 4
  - **Microsoft** Hyper-V
  - **Citrix** XenDesktop
- 預期效益：
  - 降低電力損耗 (需評估)
  - 減少資料外洩 (搭配資安政策)



# Another alternative of Virtualization

虛擬化番外篇：改用 **Terminal** + 無碟架構

用綠色自由軟體

**GREEN OSS**

打造綠色工廠

**GREEN FACTORY**

關於企鵝龍 (DRBL) 在金屬加工業的節能案例分享，請參考

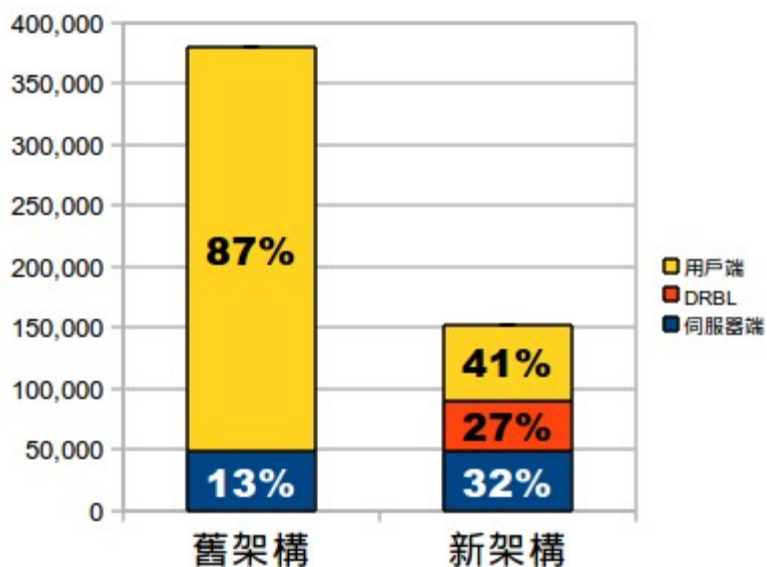
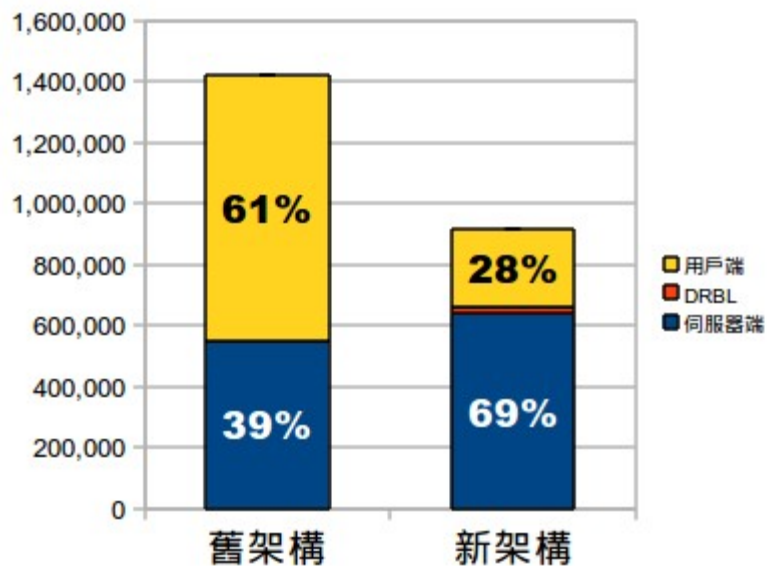
[http://trac.nhc.org.tw/cloud/attachment/wiki/jazz/09-10-14/TSMC/08-08-17\\_DRBL\\_Green\\_Computing\\_v2.pdf](http://trac.nhc.org.tw/cloud/attachment/wiki/jazz/09-10-14/TSMC/08-08-17_DRBL_Green_Computing_v2.pdf) 95

# 修改理念 = 減少不必要的開支與能源消耗

- 降低硬體建置開支
  - 不用 **200W** 的桌上型電腦 (有風扇→易故障)
  - 改用 **20W** 的精簡型電腦 (無風扇→耐油氣)
  - **不裝硬碟** (少一個零件也可減少用電)
- 降低軟體授權開支
  - 購買 **Windows** 作業系統與 **Office** 授權只爲了相當簡單的報工作業並不划算
  - 改用 **Terminal 遠端桌面連線** 模式
- 降低 **MIS** 系統管理人員負擔
  - 故障時直接拿備品精簡電腦替換→**免安裝設定**



# 新舊架構成本比較 (總建置成本 / 年度營運成本)



舊架構	
伺服器端	550,000
一般用戶端	870,000
新架構	
伺服器端	638,000
DRBL	25,000
一般用戶端	255,000
合計金額	
舊架構	1,420,000
新架構	918,000
節省金額	
	502,000

舊架構	
伺服器端	48,198
一般用戶端	331,920
新架構	
伺服器端	48,198
DRBL	41,628
一般用戶端	62,442
合計金額	
舊架構	380,118
新架構	152,268
節省金額	
	227,850

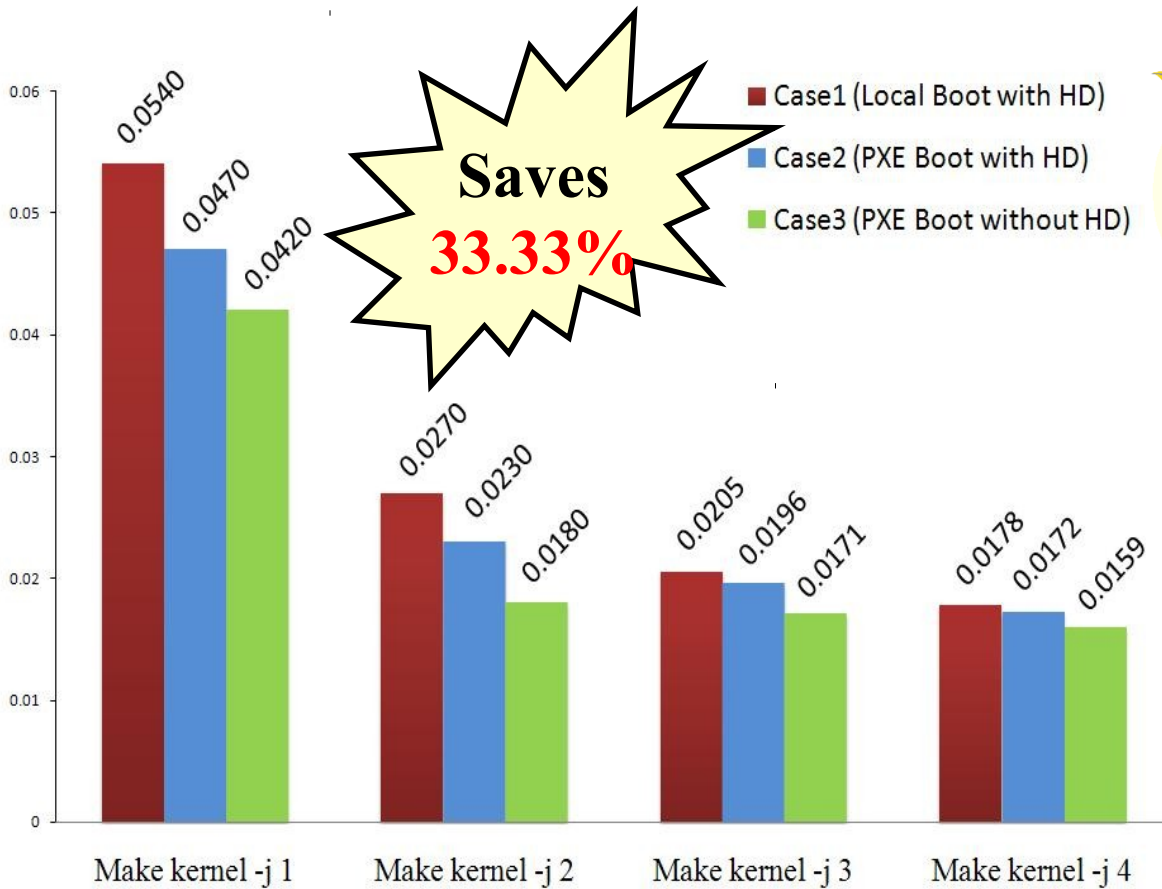
節省 **50 萬**  
總建置成本

省 **17 萬** 硬體  
省 **33 萬** 軟體

節省 **22 萬**  
年度營運成本

省 **3 萬** 電費 (60%)  
省 **19 萬** 工資 (57%)

若無很高的檔案讀寫需求，多善用 RAM Disk，無碟架構不僅省電最高達 33%，效能還比較好



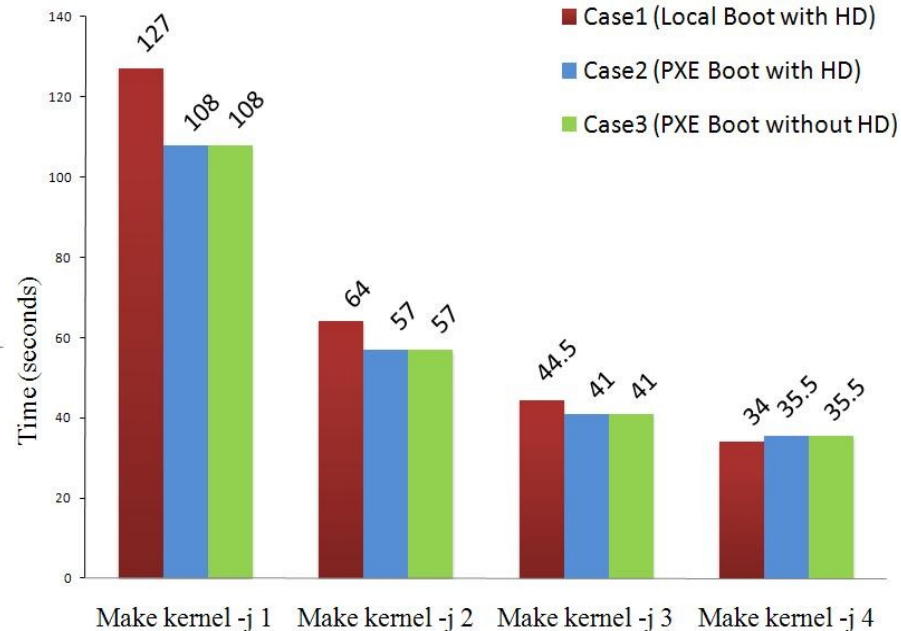
企鵝龍 **DRBL**  
(Diskless Remote Boot in Linux)

適合將整個電腦教室轉換成純自由軟體環境



再生龍 **Clonezilla**

適用完整系統備份、裸機還原或災難復原



關於企鵝龍與再生龍的更多資訊，請參考

<http://drbl.nchc.org.tw> , <http://clonezilla.nchc.org.tw>

關於 E2CC 的更多資訊，請參考

<http://trac.nchc.org.tw/grid/raw-attachment/wiki/deliverable09/1003anav.pdf>

# Adjust #2: Open Government Data

## 公部門導入雲端參考策略 (2) 開放政府公用資料

開放政府公用資料  
已成國際趨勢：

美國 - <http://data.gov>

英國 - <http://data.gov.uk>

An Official Web Site of the United States Government

Sunday, May 22, 2011 | Text: A+ A- A | Share

**DATA.GOV**  
EMPOWERING PEOPLE

HOME DATA APPS COMMUNITY METRICS OPEN DATA SITES GALLERY WHAT'S NEW

**ENVIRONMENTAL COMPLIANCE AND ENFORCEMENT DATA**

**Earthquake and Tsunami Datasets and Information**

- Worldwide M1+ Earthquakes, Past 7 Days
- RadNet Map Interface for Near-Real-Time Radiation Monitoring Data
- Search other related datasets
- World Earthquake Interactive Map Demo

SEARCH OUR CATALOGS

Search our catalogs.. SEARCH

HM Government

Log In or sign up

data.gov.uk<sup>BETA</sup>  
Opening up government

What are you looking for? Search

Home Data Apps Ideas Forum Wiki Blogs Linked Data Resources About

Data unlocking service

Cycle Journey

Emissions data - Defra data

Renewable Energy Map

**Cycle Journey**  
A free app that helps you plan your cycle journey

Over 6,900 datasets to view

**Inside Government Data**

Who's who in Government and where does the money go? Follow these links to find the data that opens it all up.

- Government spend over £25,000, by department
- Who does what in Whitehall - and how much are they paid?
- Hospitality, gifts and expenses

Share this

Facts, figures, apps and more

Find data of interest Apps Tags

**OPEN GOVERNMENT**

## Adjust #3: Single AAA Service

# 公部門導入雲端參考策略 (3) 單一身分認證

- 跨單位痛苦的第二步：身分認證 !!
- 目前已經在作：
  - 我的 e 政府 – 政府 e 化服務索引入口
  - 自然人憑證 – 實體身分認證媒介
- 導入評量指標：
  - 對外服務種類與個數
  - 民眾可能需要隨時取得的資訊
- 自由軟體方案：
  - OpenID
- 預期效益：
  - 減少資料重覆 (Data Deduplication)
  - 提供異地副本備份、災害復原機制



**Advice #1: There are other browsers !!**  
**個人建言 (1) 還有很多 IE 以外的瀏覽器 !!**

**Trend #2: Web become default Platform!**

**趨勢二：網頁變成預設開發平台**

**Open Standard** 網頁是標準

**Open Platform**

**想想您的手機是哪一種瀏覽器呢?? Opera, Safari or Chrome ?**  
工作不受壟斷

**Platform** 瀏覽器成為跨平台載具

**Web Application** 網頁程式設計成為顯學

**Browser difference become entry barrier ?!**

**瀏覽器的差異造成新的技術門檻 ?!**

**Advice #2: Ask for support of Multicore !!**  
**個人建言 (2) 軟體採購請要求多核心支援 !!**

**Trend #3: HPC become a new industry**  
趨勢三：高速計算已悄悄變成新興產業

**Parallel Computing** 平行運算

**Distributed**

分散運算的技能

有多少現有資訊系統都還沒支援多核心 !! ERP ? CRM ? 電子公文 ?

**Multi-core Programming** 多核心程式設計

**Processing Big Data** 處理大資料的技能

**Education and Training are needed !!**

為了讓這些技能與產業接軌，亟需教育訓練 !!



# 安全：雲端、智慧電網、物聯網的共通問題

Part 4 : Security Issue of Cloud, Smart Grid and Internet of Things

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



Powered by DRBL

# Scope of Cloud Security

## 雲端資安的四大範疇

用雲端  
處理資安

**Dealing Security  
issues using Cloud**

**Data Security  
In the Cloud**

雲內部  
的資安管制

**Security Issues  
Inside the Cloud**

雲端資料  
安全性

端本身  
的資安威脅

**Security Threats  
to Internet of Things**



以前你只有電腦需要防毒，現在 .....



端

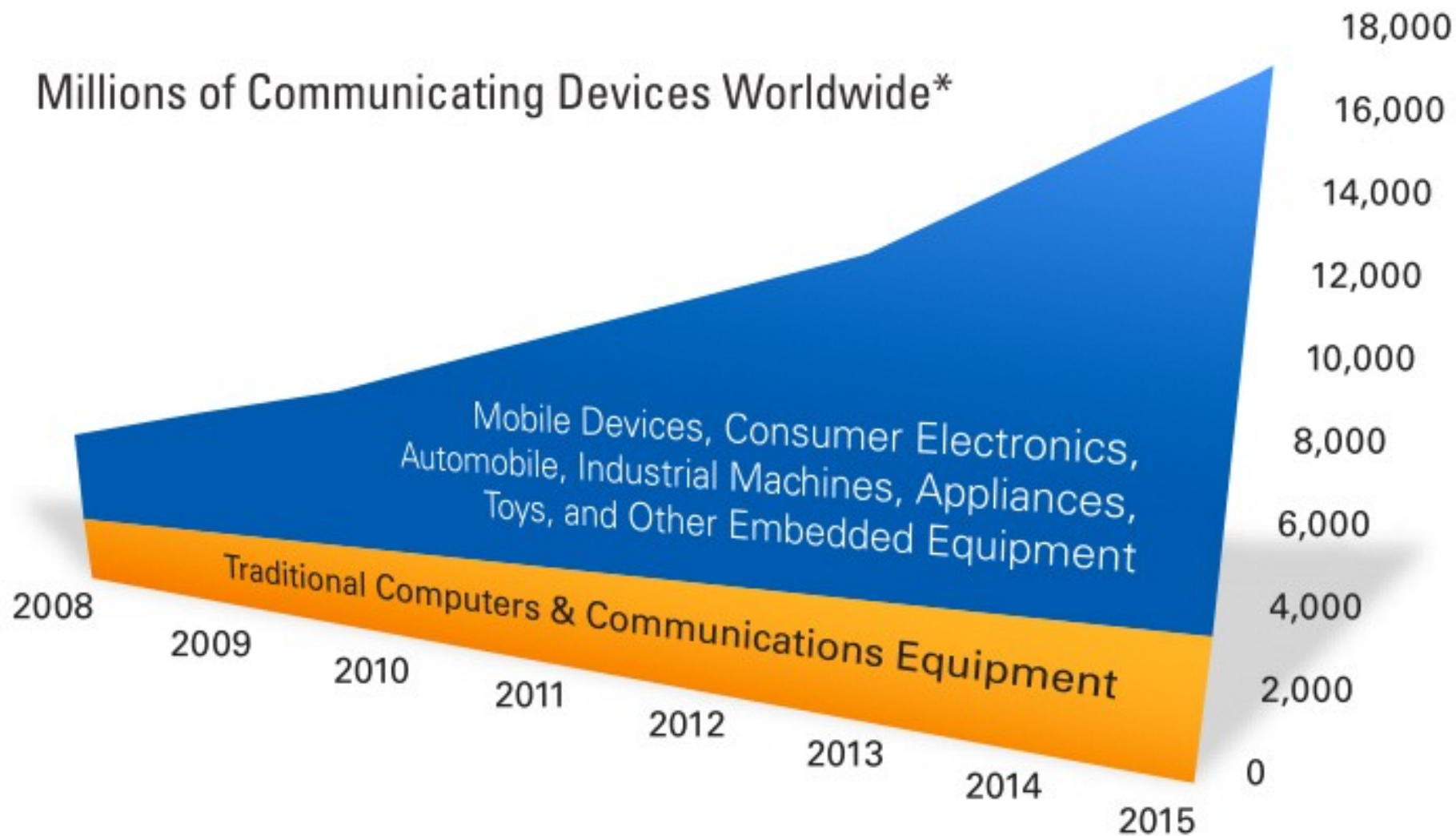
symbian OS



多元，中小廠

Diversify,  
SMB

# 全球連網裝置急速成長中



Source: IDC Device Base Model, 2009

\*Excludes voice- and SMS-only phones

# 物聯網的時代來臨

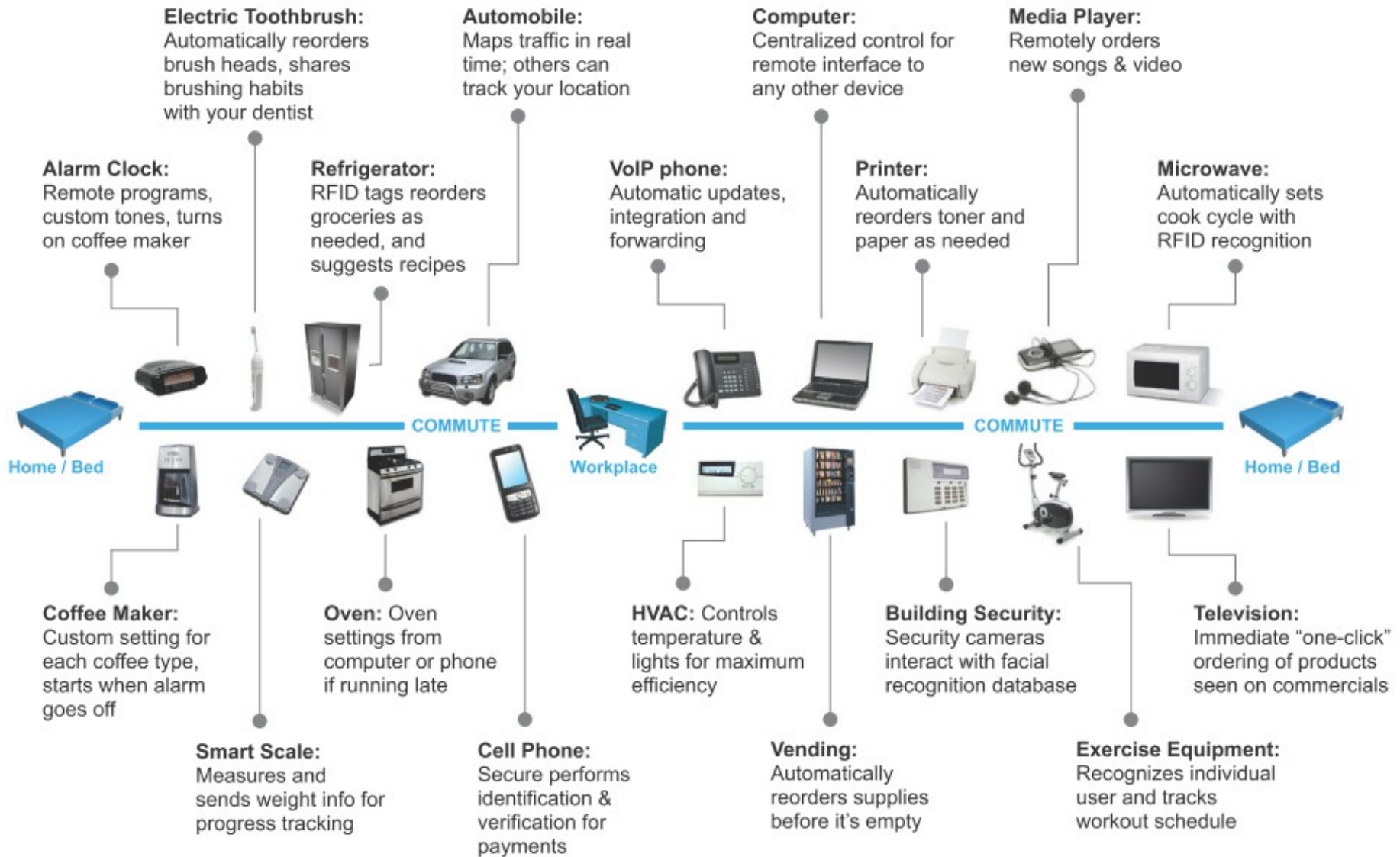
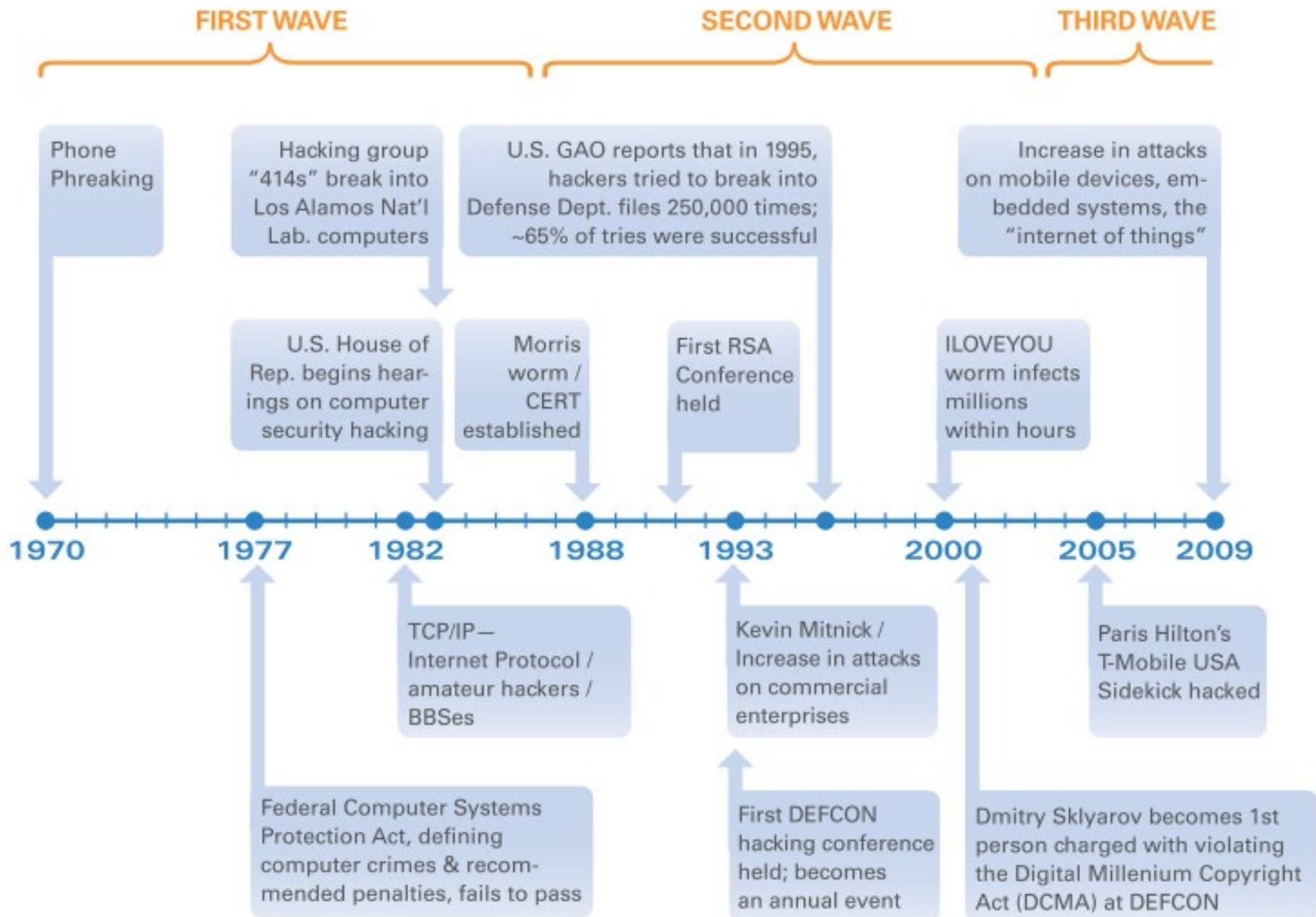


Figure 3. The Internet of Things

# 第三波網路入侵對象將鎖定在『物聯網』



# 針對行動裝置的各種資安問題與經驗

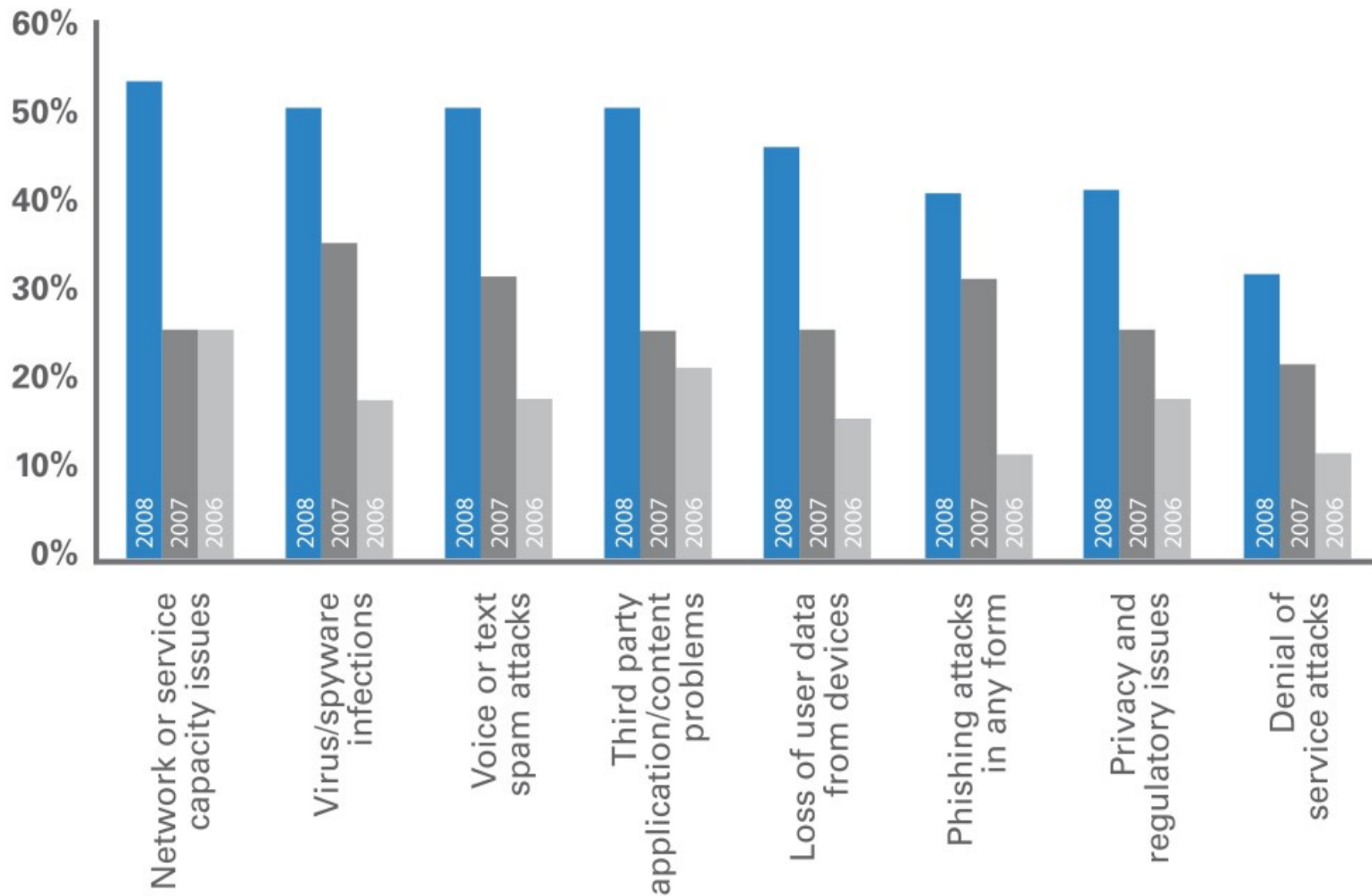
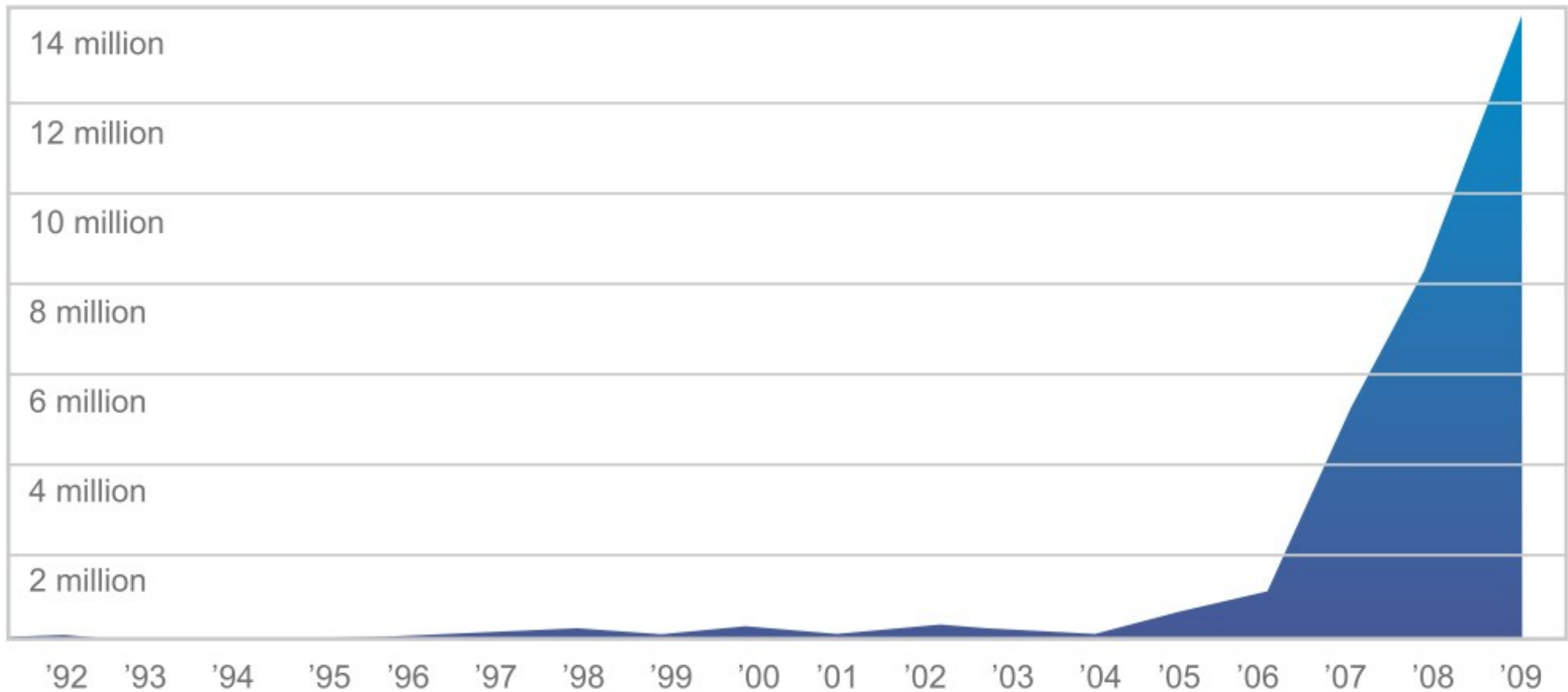


Figure 6. The increase in security issues experienced by mobile device users from 2006 to 2008; % of respondents. McAfee *Mobile Security Report 2009*

# 網路惡意程式 (Malware) 逐年激增

Malware detected by year



Over 3,000 new "species" of PC malware are released onto the Internet every hour. Now that malware is setting its sights on Device platforms.  
Source: AV LABS

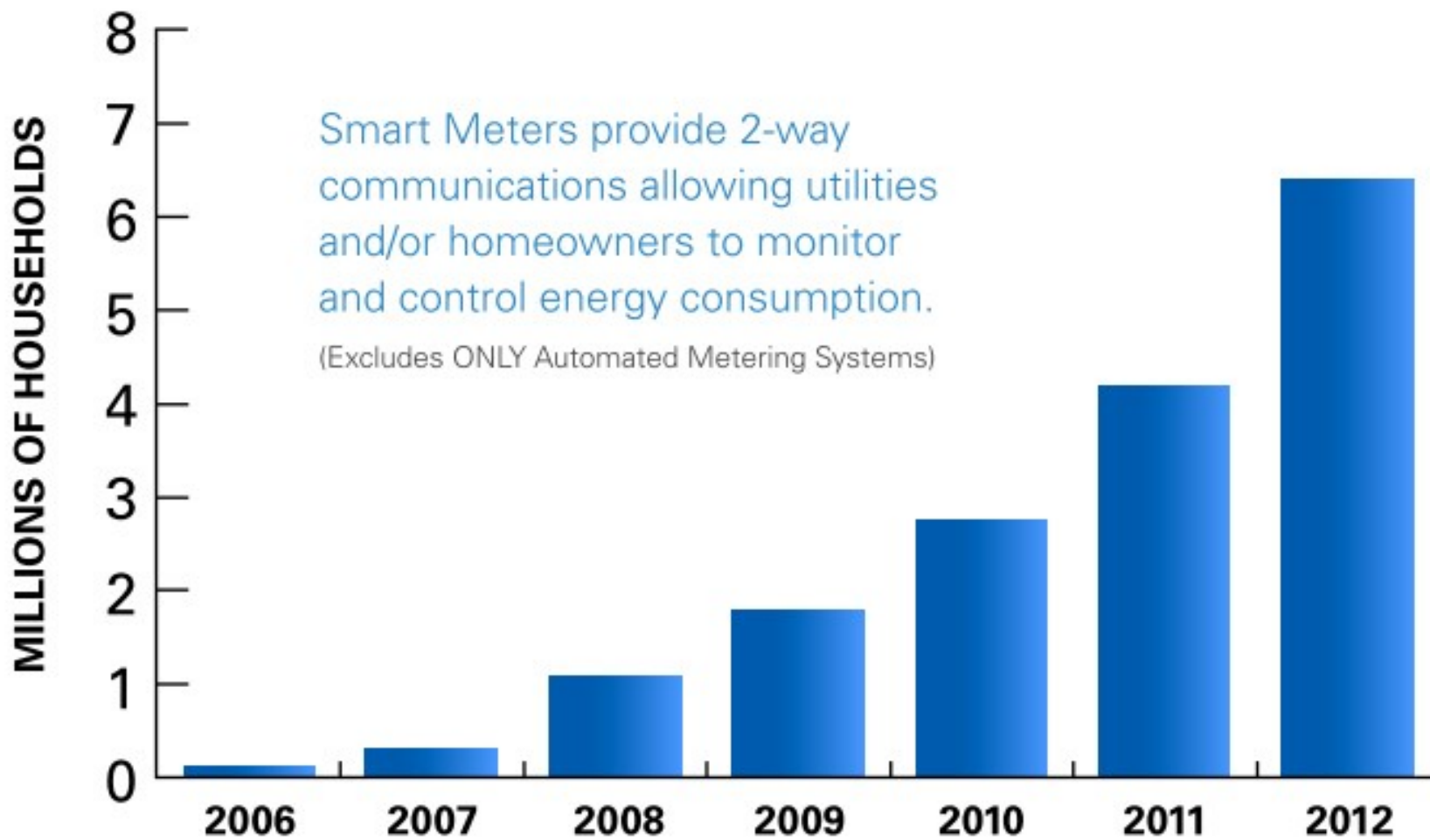
圖片來源：

U.S. Unprepared for Internet Device Flood: Unaddressed Security Problems & Talent Drought Threaten Long-Term Commercial, Government Interests

By: Kurt Stammberger, CISSP, Adrian Turner and Mat Small, Mocana With: Rich Nass, Sarah Friar, Goldman Sachs

# 如果你家的智慧電錶被入侵會怎樣？

## U.S. Households with Smart Meters



© Copyright 2009 - Parks Associates

圖片來源：

U.S. Unprepared for Internet Device Flood: Unaddressed Security Problems & Talent Drought Threaten Long-Term Commercial, Government Interests

By: Kurt Stammberger, CISSP, Adrian Turner and Mat Small, Mocana With: Rich Nass, Sarah Friar, Goldman Sachs



## Questions?

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

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



Powered by DRBL