雲端運算趨勢與統計學習

Trends of Cloud Computing and its relation to Statistics

Jazz Wang

Yao-Tsung Wang

jazz@nchc.org.tw

Powered by DRBL

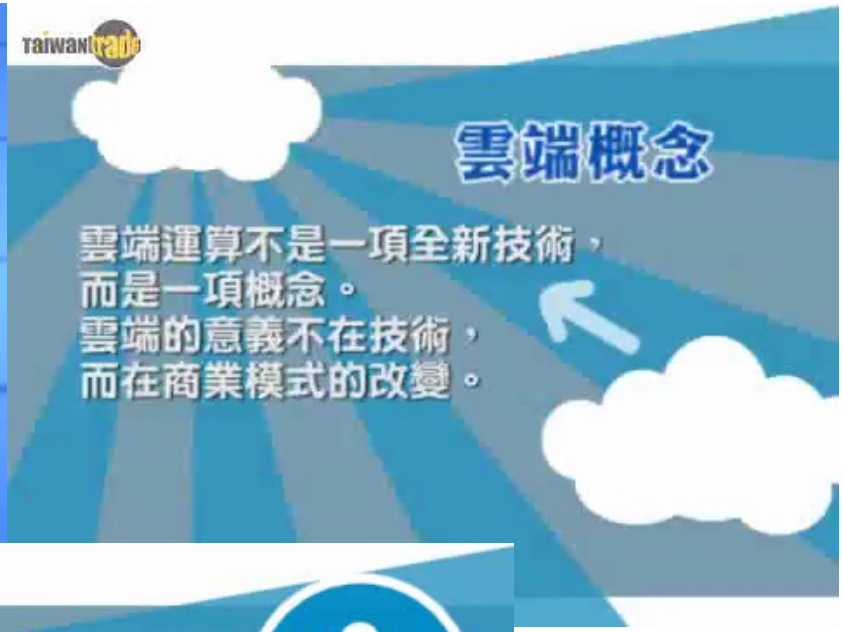
Course Information 課程資訊

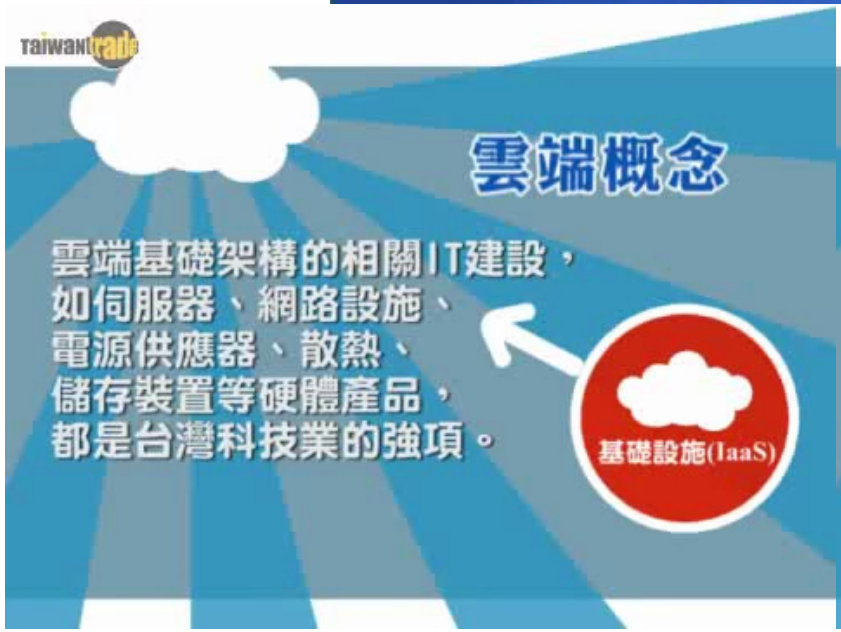
- 講師介紹：
  - 國網中心　王耀聰　副研究員 / 交大電控碩士
  - jazz@nchc.org.tw
- 所有投影片、參考資料與操作步驟均在網路上
  - 由於雲端資訊變動太快，愛護地球，請減少不必要之講義列印。
- 礙於缺乏實機操作環境，故以影片展示與單機操作爲主
  - 若有興趣實機操作，請參考國網中心雲端運算課程錄影
  - http://trac.nchc.org.tw/cloud
  - http://www.classcloud.org/media
  - http://www.screentoaster.com/user?username=jazzwang
- 若需要實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

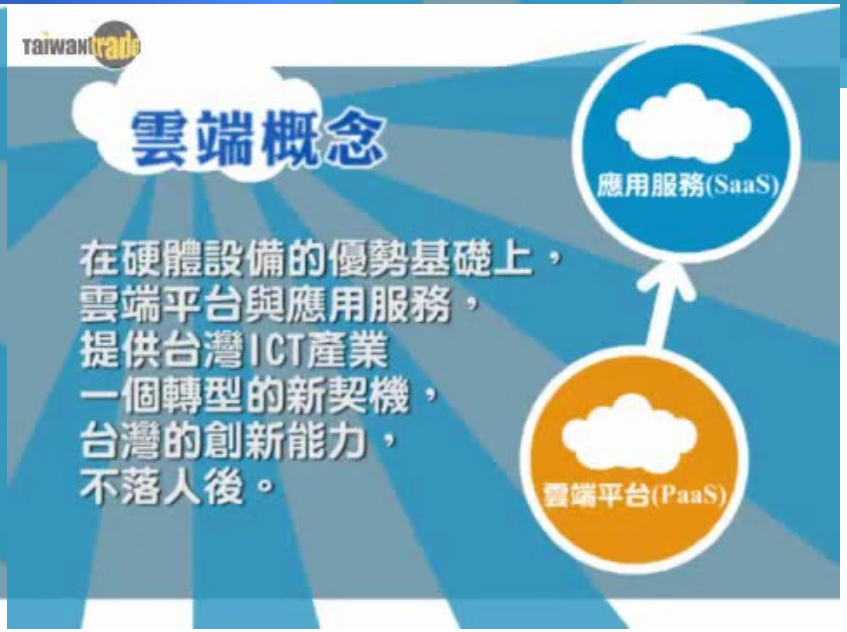
Source: http://www.2010taipeiexpo.tw/ct.asp?xItem=17186&CtNode=5952&mp=3

什麼是雲端運算啊？

What is Cloud Computing ?







http://www.youtube.com/watch?v=bJLSAcU6O3U
http://www.youtube.com/watch?v=VIMtd3nfPqc

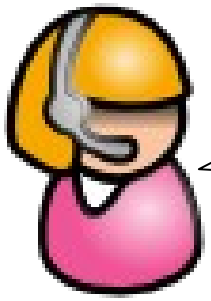
當紅「雲端運算」 你瞭解了嗎？
雲端產業 8分鐘就上手

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

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

Public Cloud
公用雲端

amazon web services™
Microsoft
Google

Target Market is S.M.B.
主要客户爲中小企業

Hybrid Cloud

以大型企業爲主要客户
Enterprise is key market

IBM

私有雲端
Private Cloud

Community Cloud
社群雲端

Academia 學術爲主

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」？

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

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

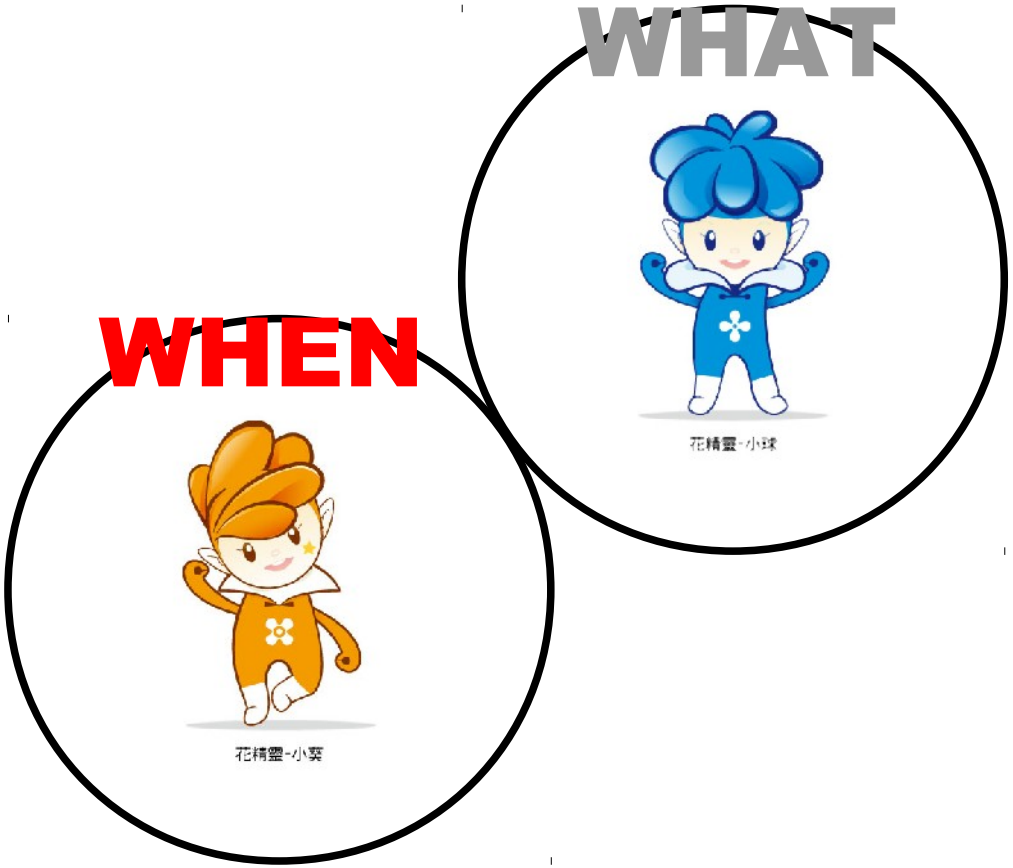
Key spirit of Cloud ~

形成服務才是重點！！

Everything as a Service !!

Cloud Computing =~ Network Computing

雲端運算 =~ 網路運算



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

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

Source: http://www.cnet.co.uk/i/c/blg/cat/software/cloudcomputing/clouds1.jpg

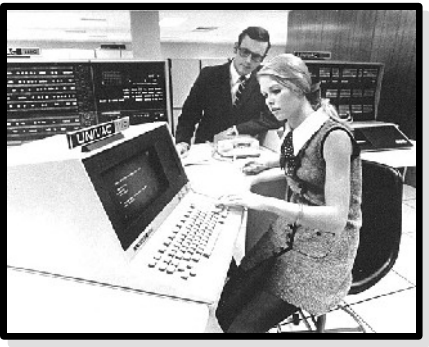
Evolution of Cloud Services
雲端服務只是軟體演化史的必然趨勢

e-Taiwan 成果系列報導 → 行動台灣 應用推動計畫 → 發展優質網路社會 (U-Taiwan) 計畫 → 智慧台灣 領航系列報導

實體 Physical	單機版 個人使用 Personal Software	網路版 多人共享 Share Service Software	行動版 隨時存取 Mobile Cloud Service
信箱 Mailbox	E-Mail	Web Mail	Mobile Mail
電視 TV	電視盒 Setop Box	Web TV Ex. Youtube	Mobile TV
打字機 Typer Writer	Office	Google Docs	M-Office
電話 Telephone	數位電話 PBX	Skype	Flash Wengo
佈告欄 Bullet Borad	電子佈告欄 BBS	部落格 Blog	微網誌 Twitter

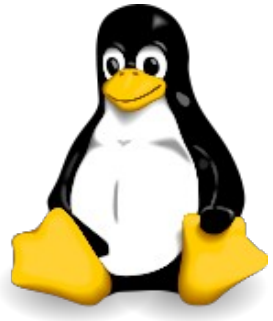
Brief History of Computing
運算技術演進簡史

1960



1977

1991



1993

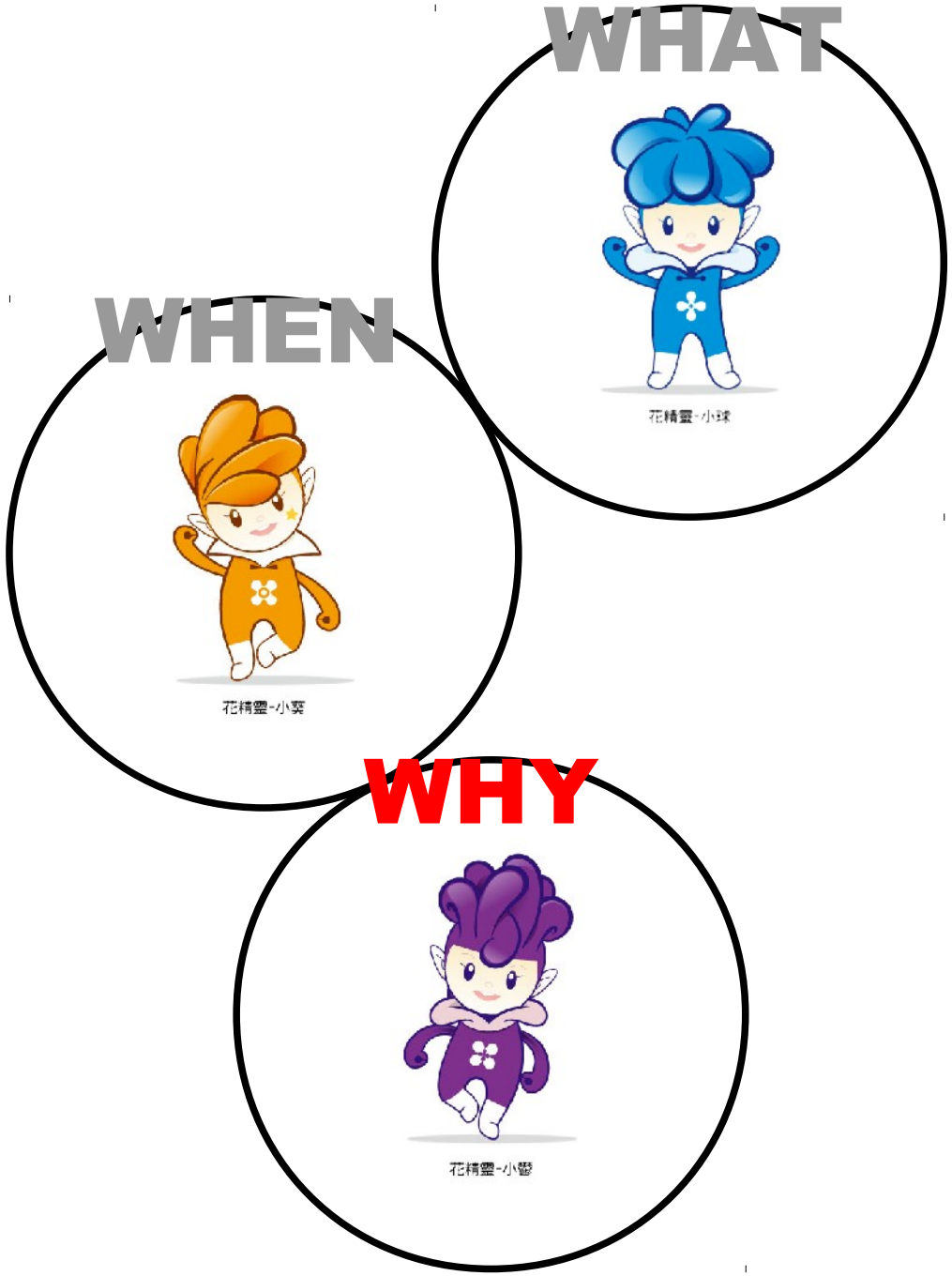
2002

2003

2004

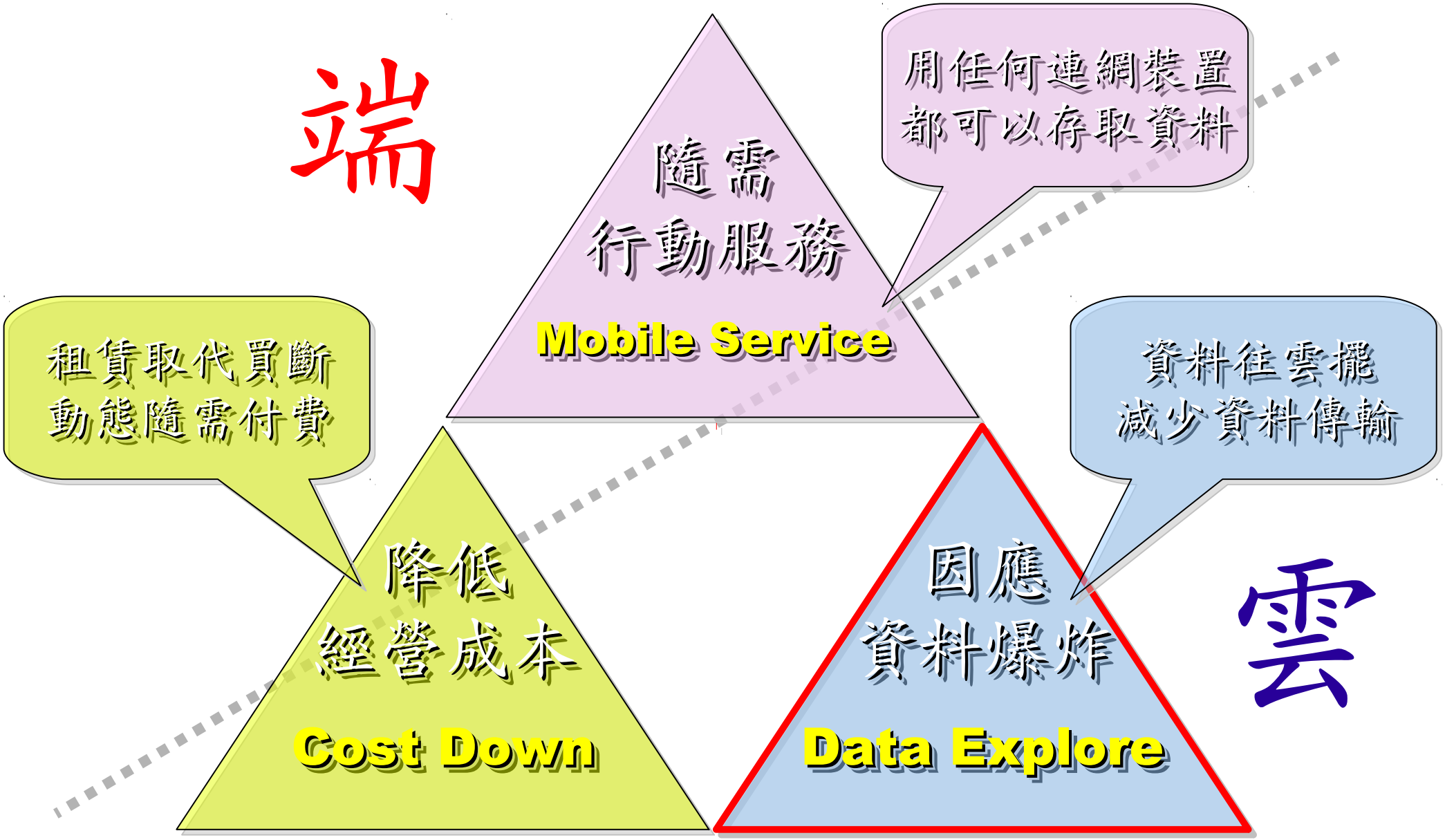
2006

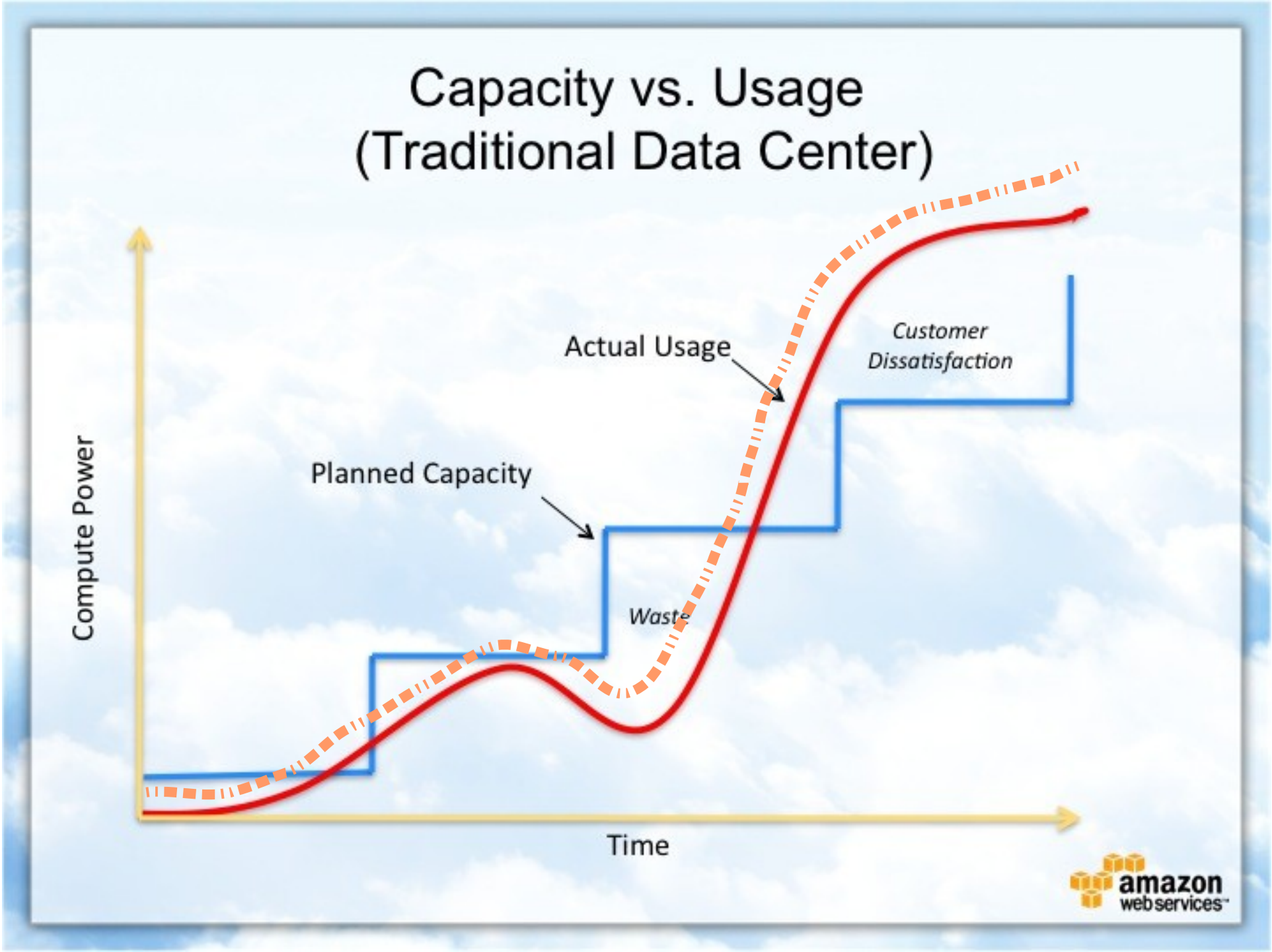
Mainframe Super Computer → **PC / Linux Cluster Parallel** → **Internet Distributed Computing** → **Virtual Org. Grid Computing** → **Data Explode Cloud Computing**



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

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





Source : http://awsmedia.s3.amazonaws.com/chart01_traditional_720x540.jpg

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

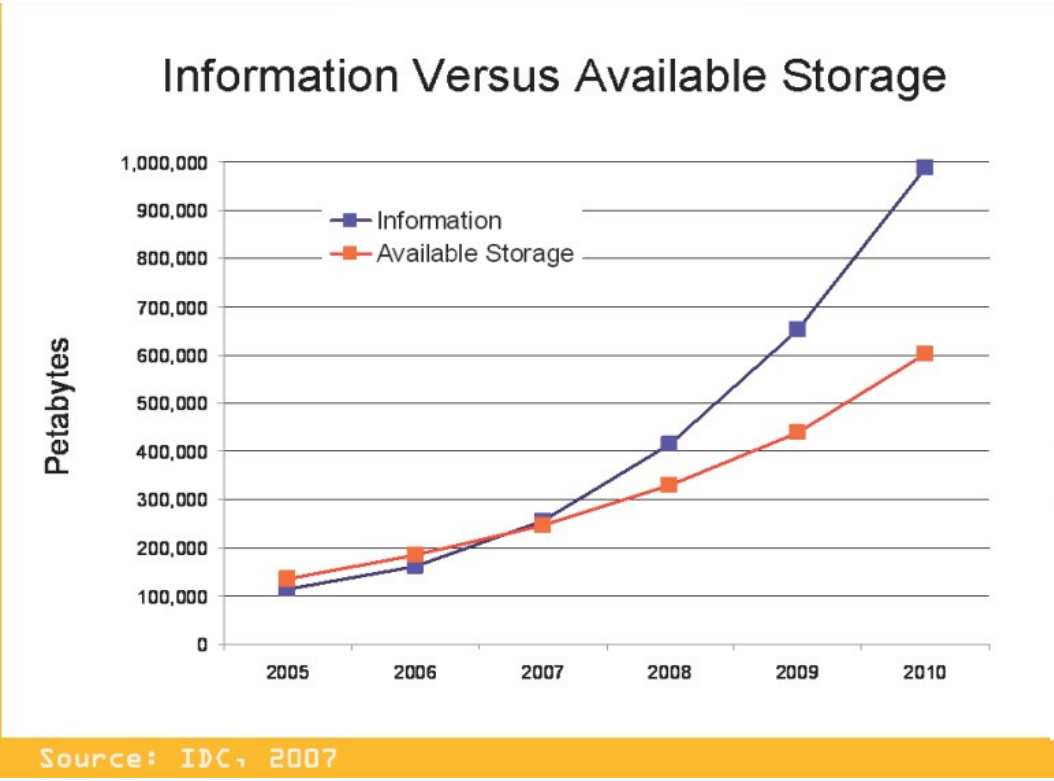
Lesson #3: Extra cost to move data to Grid !
教訓三：資料搬運的網路與時間成本！

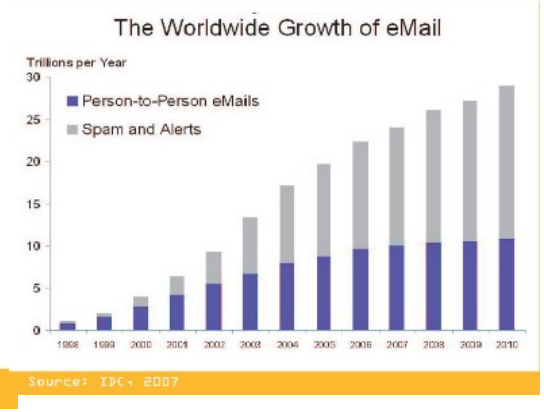
Answer #3: Total Cost of Ownership 總擁有成本

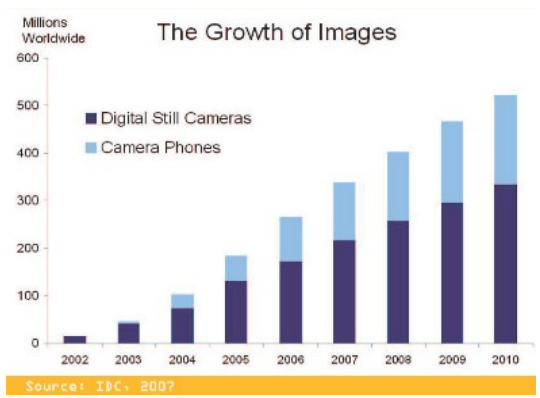
Cost Down is the Key Drive !!
降低營運成本才是企業導入雲端運算的關鍵考量！!

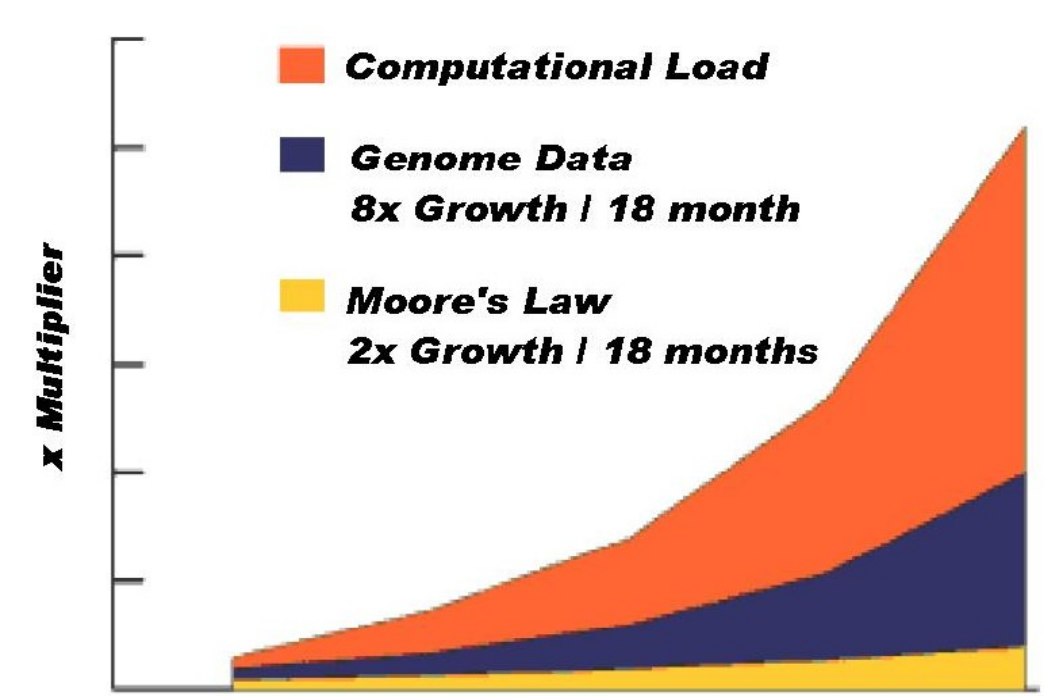
2007 Data Explore

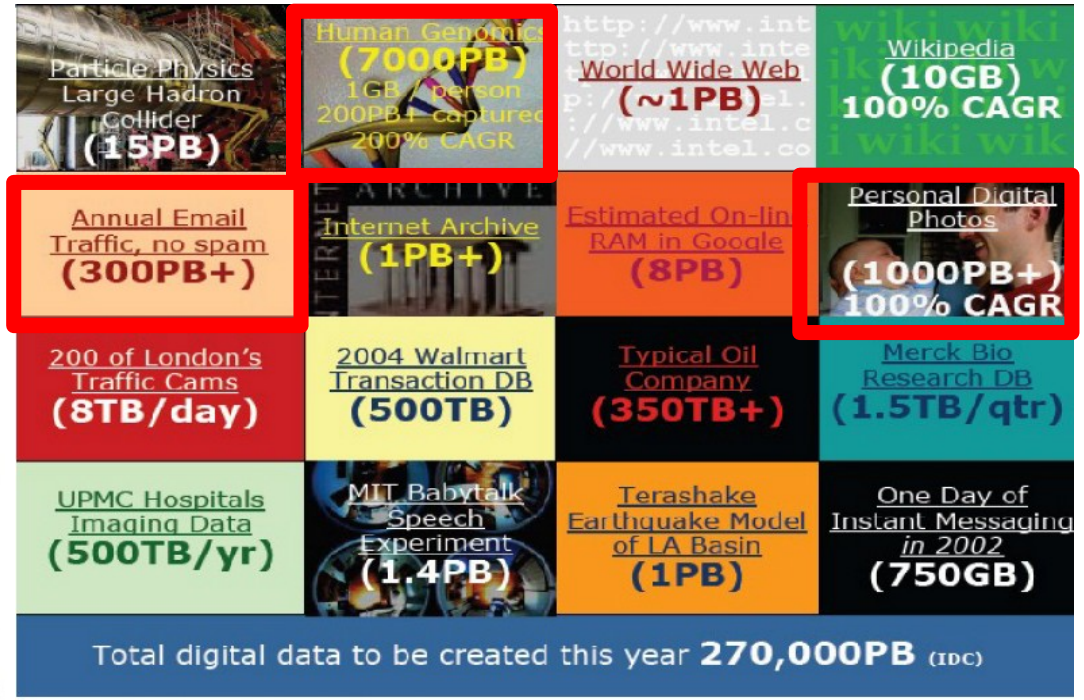
Top 1 : Human Genomics – 7000 PB / Year
Top 2 : Digital Photos – 1000 PB+/ Year
Top 3 : E-mail (no Spam) – 300 PB+ / Year







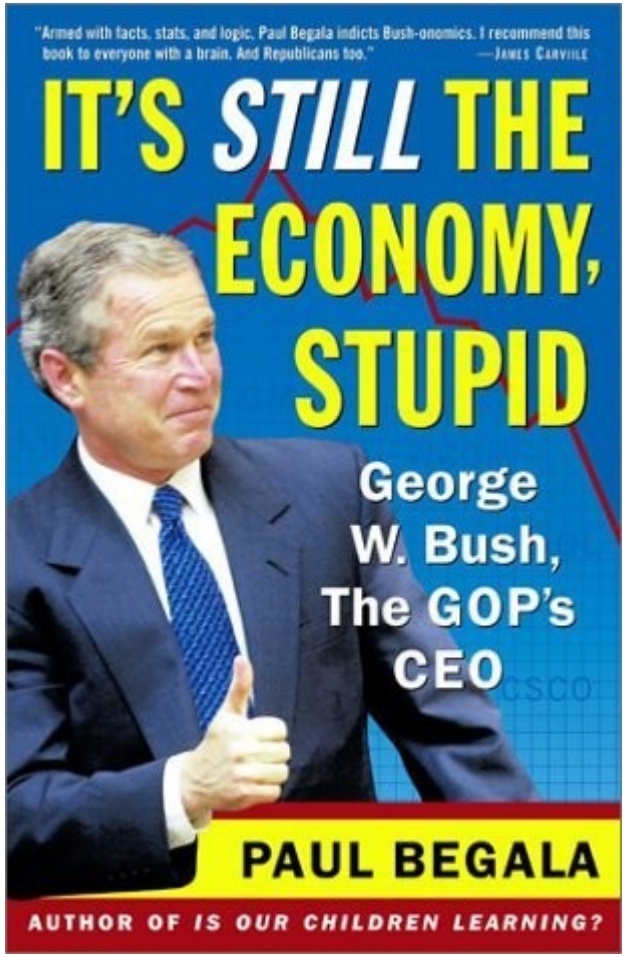




Phillip B. Gibbons, Data-Intensive Computing Symposium

Source: http://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf
Source: http://lib.stanford.edu/files/see_pasig_dic.pdf

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」

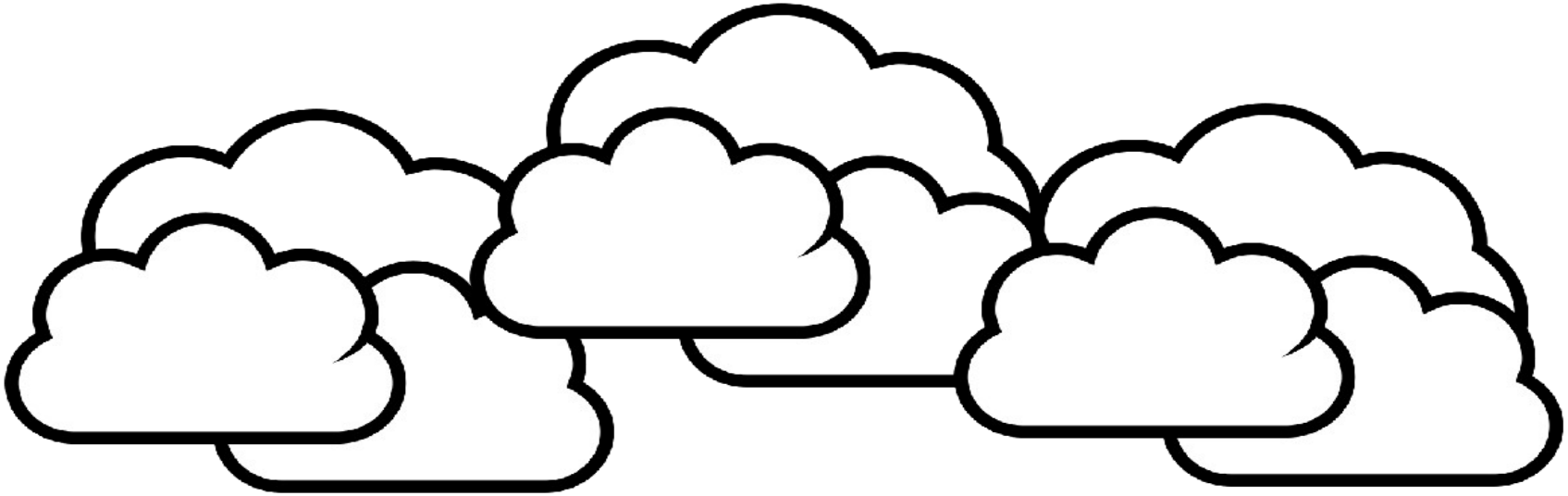
（**"It's the data, stupid"**）

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？

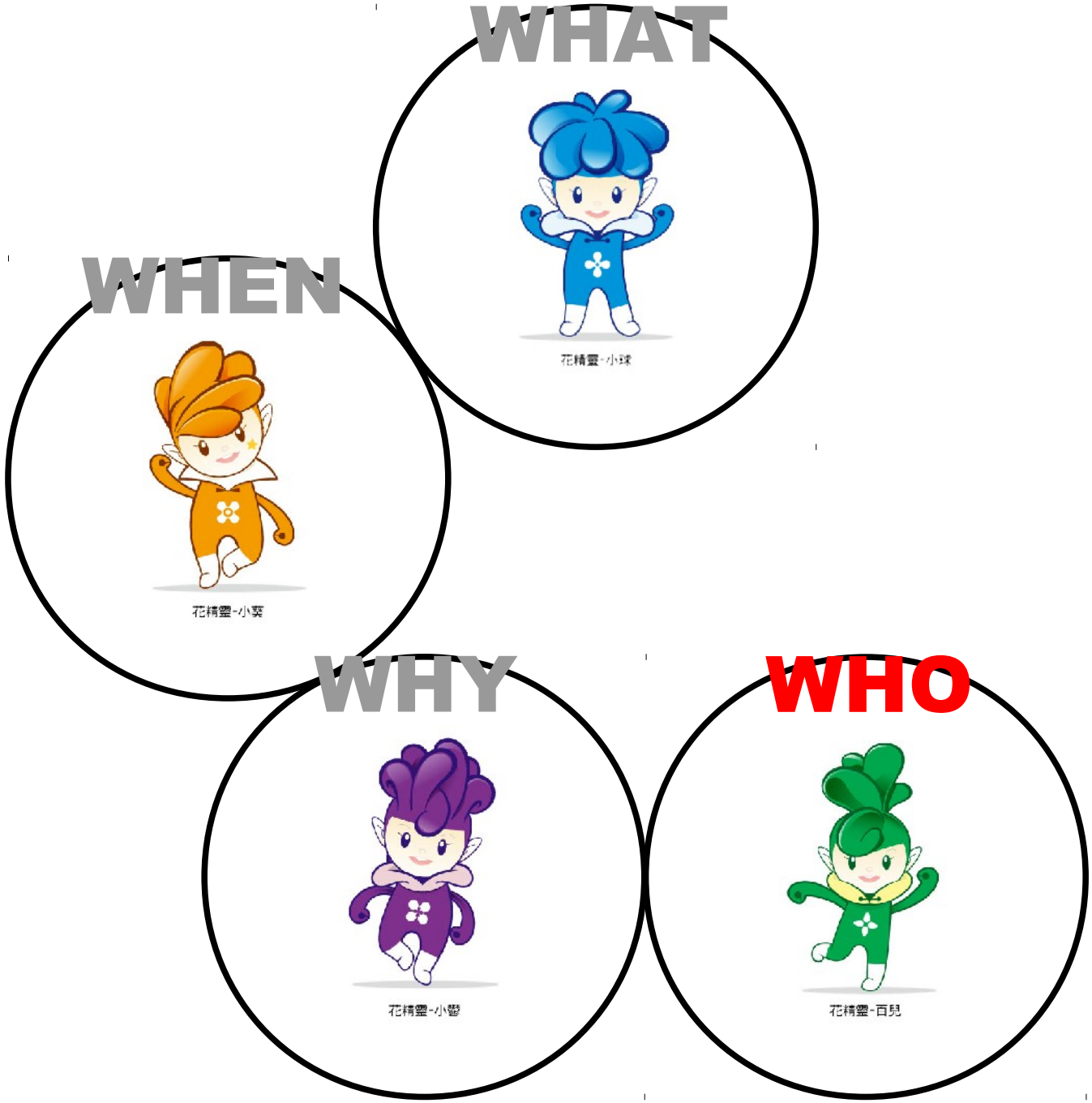
- 2007 年

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



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

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

amazon web services™
Microsoft
Google

雲端服務參考模型
Reference Model

Target Market is **S.M.B.**
主要客户爲
中小企業

Hybrid Cloud

以大型企業爲主要客户
Enterprise is key market

IBM

私有雲端
Private Cloud

Community Cloud
社群雲端

Academia 學術爲主

IaaS Key Player: Amazon 亞馬遜網路書店

- 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.cisanet.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

PaaS / SaaS Key Player : Google 公歌

- 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

PaaS / SaaS Key Player : Google 公歌

Gmail / 電子郵件

Contac / 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

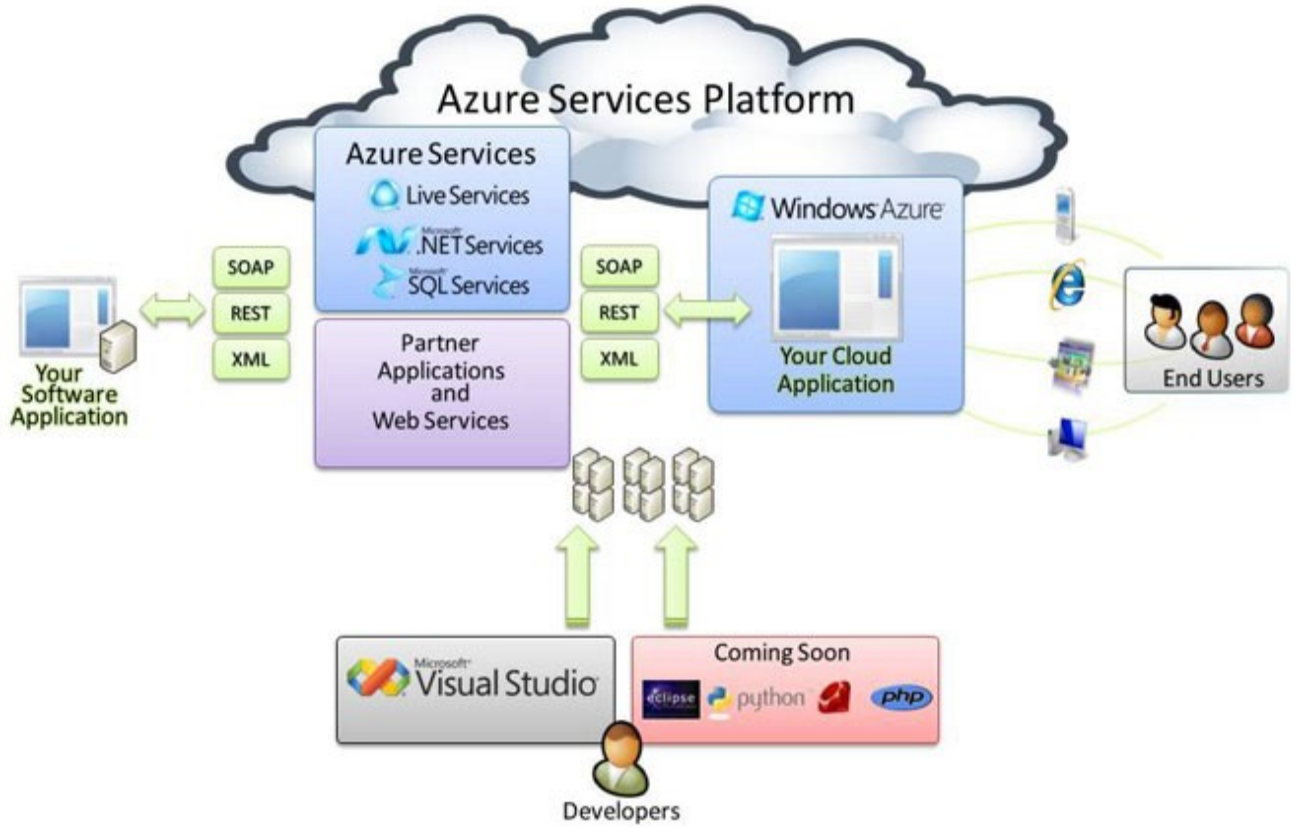
Admin / 管理後台

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

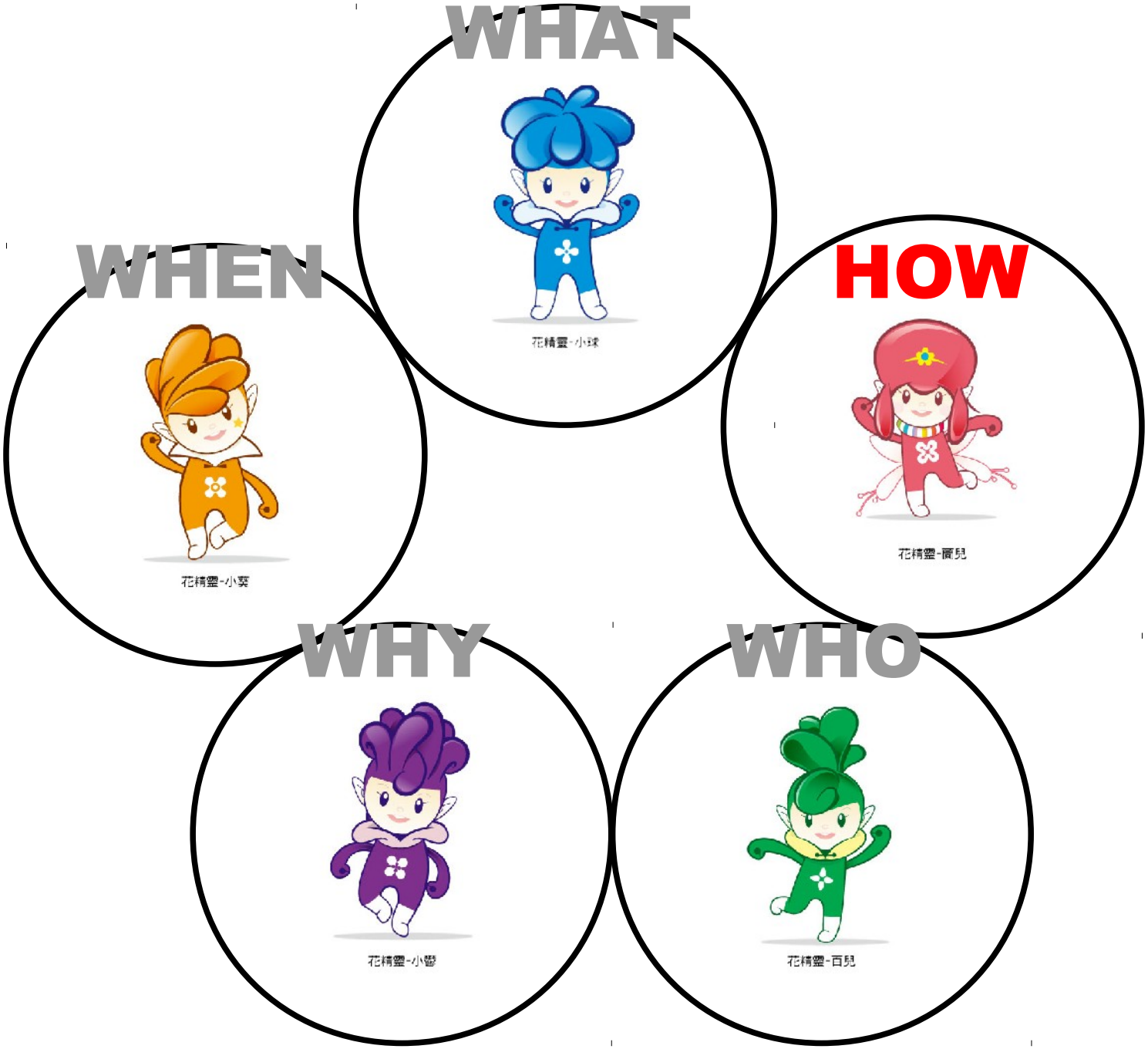
如果無法掌握雲端技術，至少該學會使用雲端服務！

XaaS Key Player : Microsoft 微軟

- Microsoft Azure 是一套雲端服務作業系統。
- 作爲 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

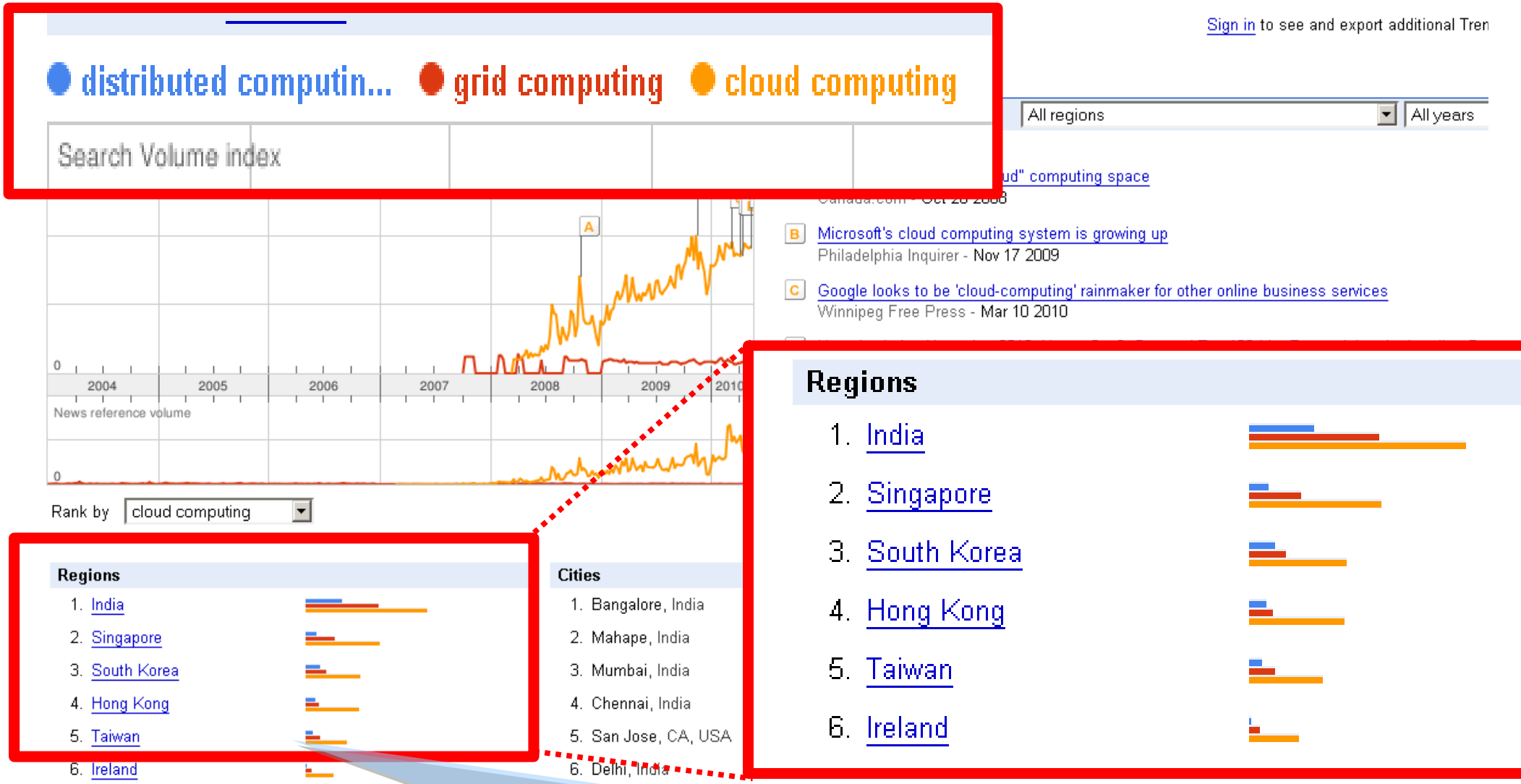


參考來源：http://tech.cipper.com/index.php/archives/332



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

What are the trend of next 10 years ?
什麼是下個十年的熱門工作技能？



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？

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。

雲端運算

大量資訊分析

智慧裝置

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

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 年成爲該領域的主要廠商之一。
情境感知運算(context-aware computing)	較人工智慧更爲寬廣，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、射頻晶片更爲普及，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

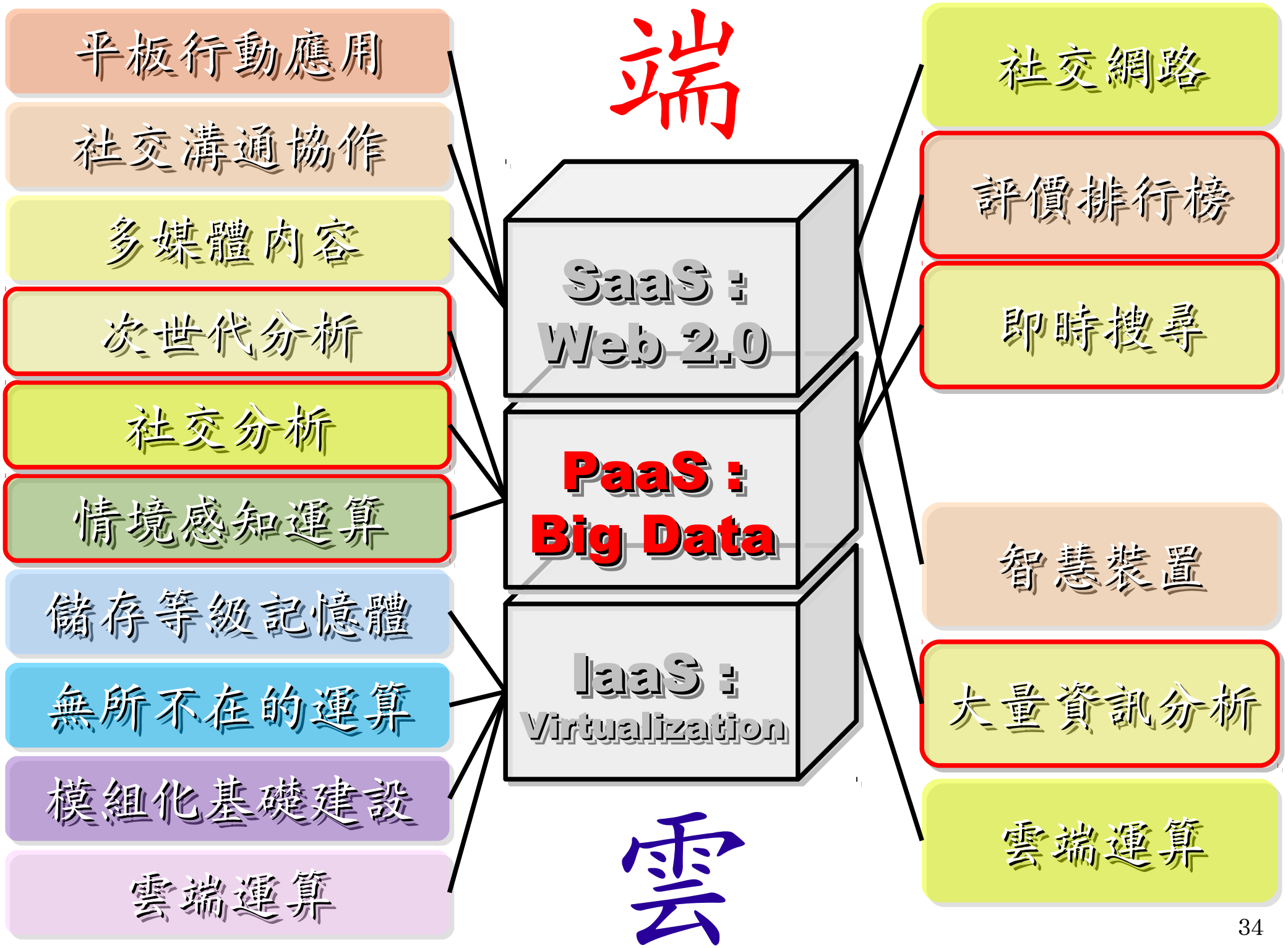
儲存等級記憶體

無所不在的運算

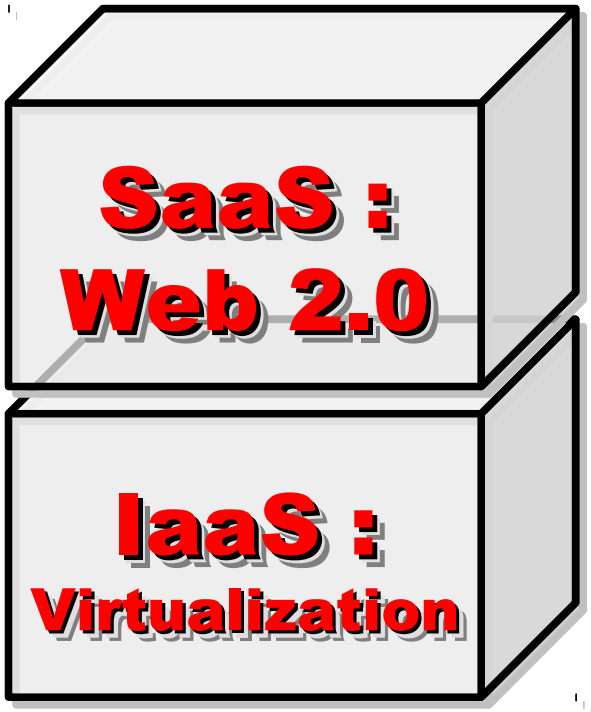
模組化基礎建設

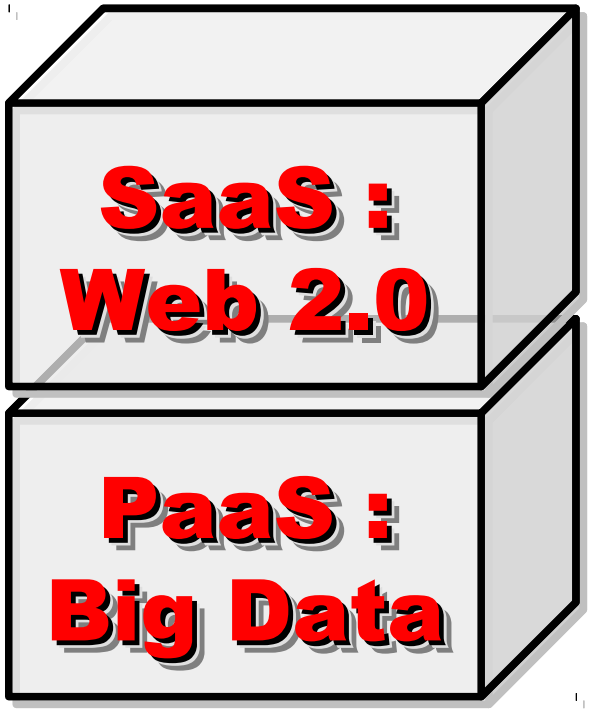
Source ： http://www.gartner.com/it/page.jsp?id=1454221

Source ： http://www.digitimes.com.tw/tw/dt/n/shwnws.asp?Cnlid=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？





amazon web services™

Google™

想盡辦法誘你用計算跟網路

Computing Intensive

想盡辦法誘你提供資料作分析

Data Intensive

Reference Cloud Architecture
雲端運算的參考架構

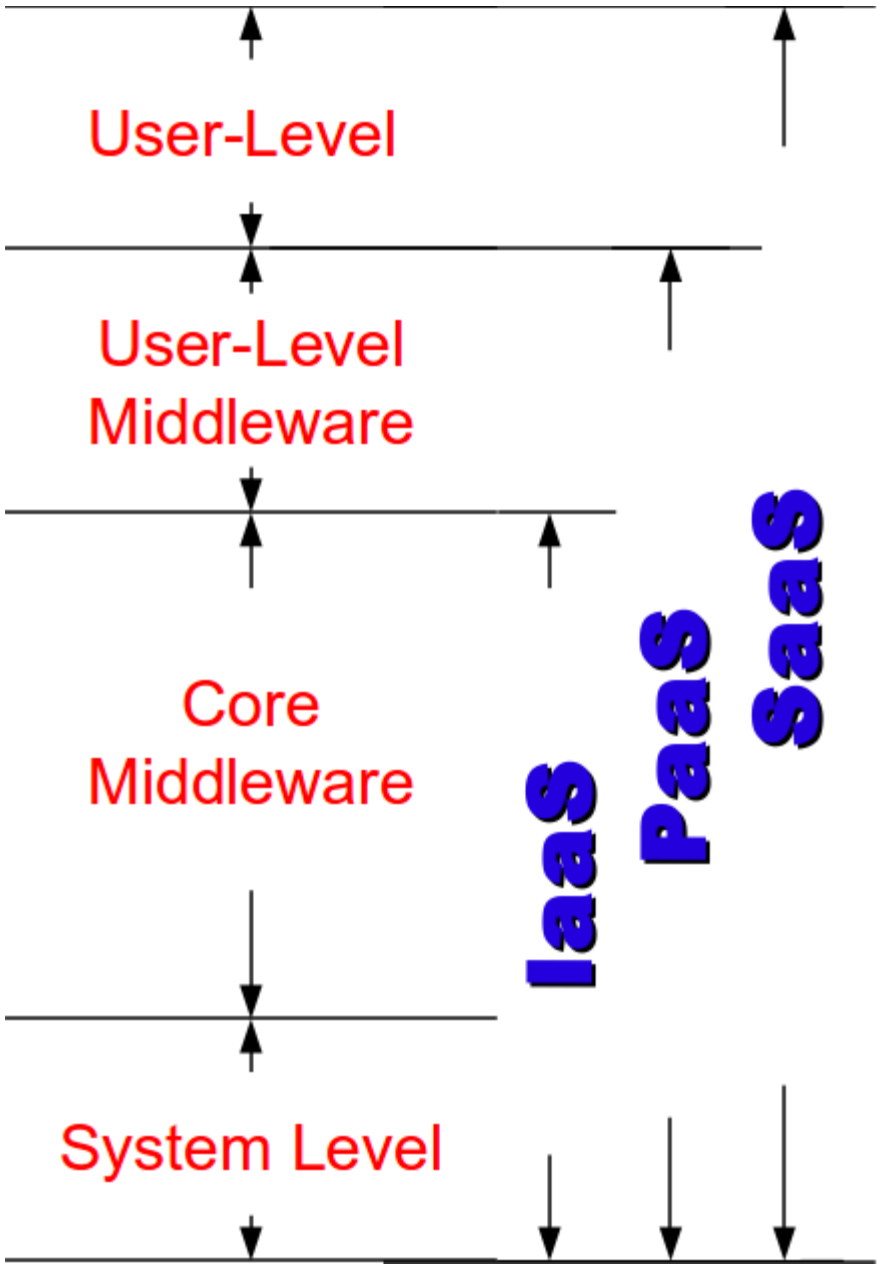
應用軟體 Application
Social Computing, Enterprise, ISV,...

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

控制管理 Control
Qos Neqotiation, Ddmission 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 Cloud Service
建構雲端服務的自由軟體

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 Neqotiation, Ddmission 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

雲端大量資料分析與統計學習

Data Science Technologies and Statistics

Jazz Wang

Yao-Tsung Wang

jazz@nchc.org.tw

Powered by DRBL

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樣股份有限公司（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

中華電信用 Hadoop 技術分析通話明細

中華電信 hicloud 雲端運算伺服器 郵件、資料庫、防火牆… 輕鬆解決企業IT資源需求
iThome週週為IT人打氣！ 雲端伺服器首選，半年免費 企業選平板？選最相容的！

中華電信用Hadoop技術分析通話明細

文/室雅薔 2011-06-12

+1 0 讚 62 人說讚。快免費註冊來查看你的朋友對什麼說讚。 + 我要收藏

中華電信利用自行開發的Hadoop大資料運算平臺，找出非結構化資料中的結構性，精簡資料後再置於資料倉儲運算，節省儲存空間

面對資料快速成長以及非結構性資料的增加，中華電信資訊處第四科科長楊秀一表示，中華電信近來利用Hadoop雲端運算技術自行開發了一個專門用來分析非結構化資料的巨量資料（Big Data）運算平臺，嘗試在資料進到資料倉儲系統之前，先進行資料的分析與處理以減少資料倉儲的資料量。

近年來行動語音市場趨於飽和，為了掌握用戶特性進行客製化行銷，一份資料要進行分析，就會被多次複製，因此即使用戶增加趨緩，但中華電信擁有的資料量仍快速暴增。

研討會訊息
· Websense TRITON電子郵件資料安全解決方案研討會
· 2011 JavaTWO專業技術大會
+更多研討會

▼ ADVERTISEMENT ▼
Microsoft

參考來源： http://www.ithome.com.tw/itadm/article.php?c=68023

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 : Simplifed 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/

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 Neqotiation, Ddmission 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

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 MalStone 效能評比軟體
- **Open Cloud Consortium provide Open Cloud Testbed and develop MalStone toolkit for benchmark**

Sector-Sphere

National Center for Data Mining
University of Illinois at Chicago
UIC

open data
Open Data Group
http://www.opendatagroup.com/

What we learn today ?

WHAT	隨時隨地用任何裝置存取各種服務!! Accessing services with any device anytime anywhere!!
WHO	亞馬遜、谷歌、微軟等! 什麼都可以是服務 ~ Amazon, Google, Microsoft and more! Everything as a Service!
WHEN	雲端運算是2006年繼格網運算之後的新趨勢!! Cloud Computing become new trend since year 2007 !!
WHY	資料爆炸、節省成本、行動應用 Data-intensive, Cost-Efficency, Mobile Applications
HOW	採用自由軟體也能打造私有雲端 Hadoop, Sectore/Sphere, Eucalyptus, and more ....

Questions?

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

Jazz Wang

Yao-Tsung Wang

jazz@nchc.org.tw

Powered by DRBL