

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**



淺談雲端運算的新趨勢

Part 1 : Overview the trend of Cloud Computing

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



Powered by **DRBL**



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

On-demand self-service.

隨需自助服務

Broad network access

隨時隨地用任何網路裝置存取

Resource pooling

多人共享資源池

Rapid elasticity

快速重新佈署靈活度

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*

Community Cloud

社群雲端

Academia 學術為主

以**大型企業**
為主要客戶

Enterprise is
key market



私有雲端

Private Cloud

3 Service Models of Cloud Computing

雲端運算的三種服務模式

IaaS

Infrastructure as a Service

架構即服務

PaaS

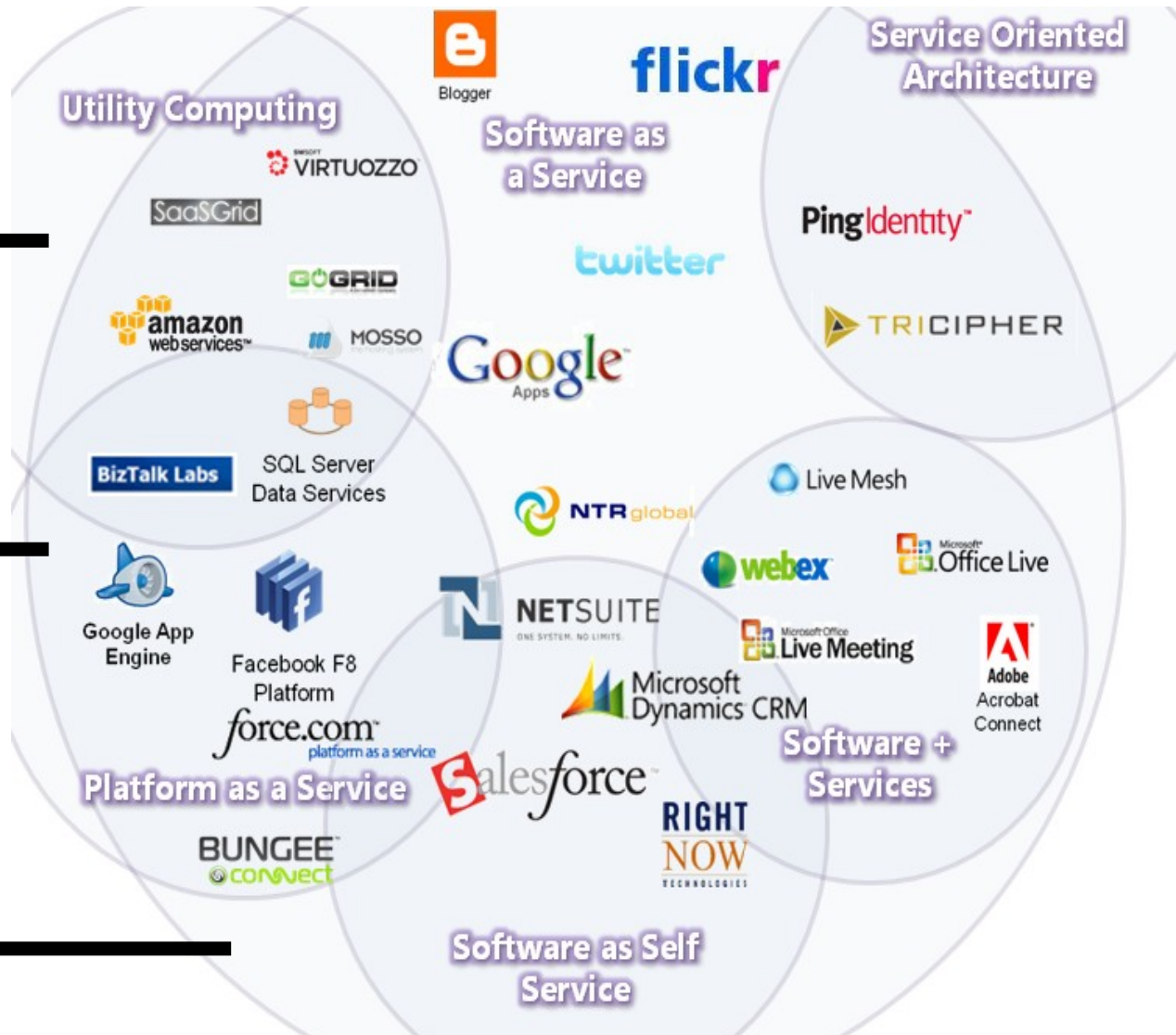
Platform as a Service

平台即服務

SaaS

Software as a Service

軟體即服務



2 R&D directions : Cloud or Device

兩大研究方向：你該選「雲」還是「端」？



One 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 ~

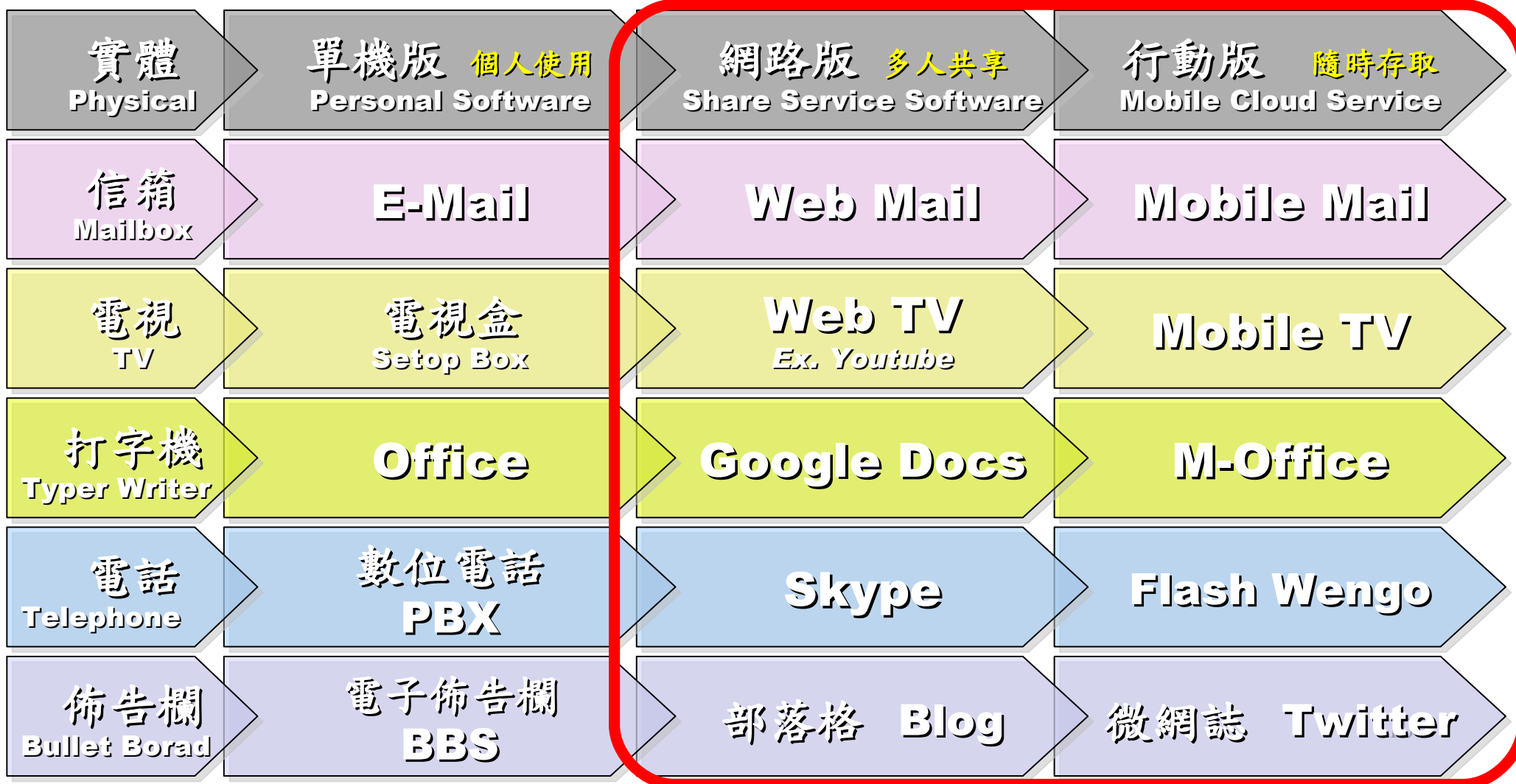
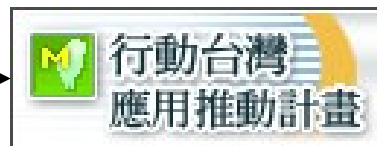
形成服務才是重點！！

Everything as a Service !!

Evolution of Cloud Services

雲端服務只是軟體演化史的必然趨勢

數位化



Rome wasn't built in a day !

羅馬不是一天造成的！



圖片來源：<http://www.mjfq.com/pic/20070822/20070822234234402.jpg>

When did the Cloud come ?!

這朵雲幾時飄過來的？！

Brief History of Computing (1/5)



Source: <http://pinedakrch.files.wordpress.com/2007/07/>

Mainframe
Super
Computer

1960 PDP-1

'

'

'

1965 PDP-7

'

'

'

1969 1st Unix

1977 Apple II

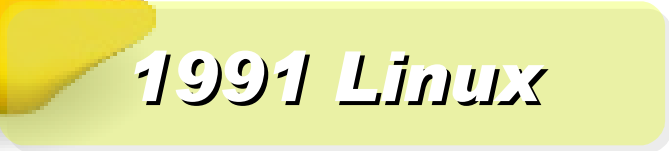


1981 IBM 1st PC 5150



Back to Year 1970s ...

1982 TCP/IP



14

Brief History of Computing (2/5)



Source: <http://www.nhc.org.tw>

Mainframe
Super
Computer

PC | Linux
Cluster
Parallel

**1990 World Wide Web
by CERN**

...

...

**1993 Web Browser
Mosaic by NCSA**



1991 CORBA

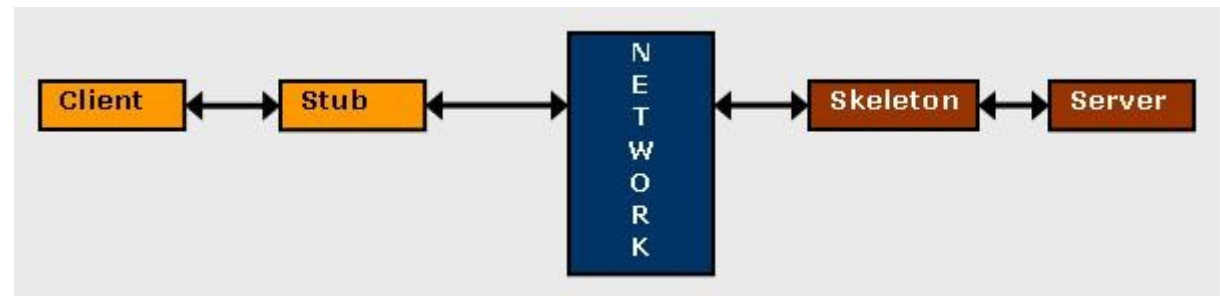
...

Java RMI

Microsoft DCOM

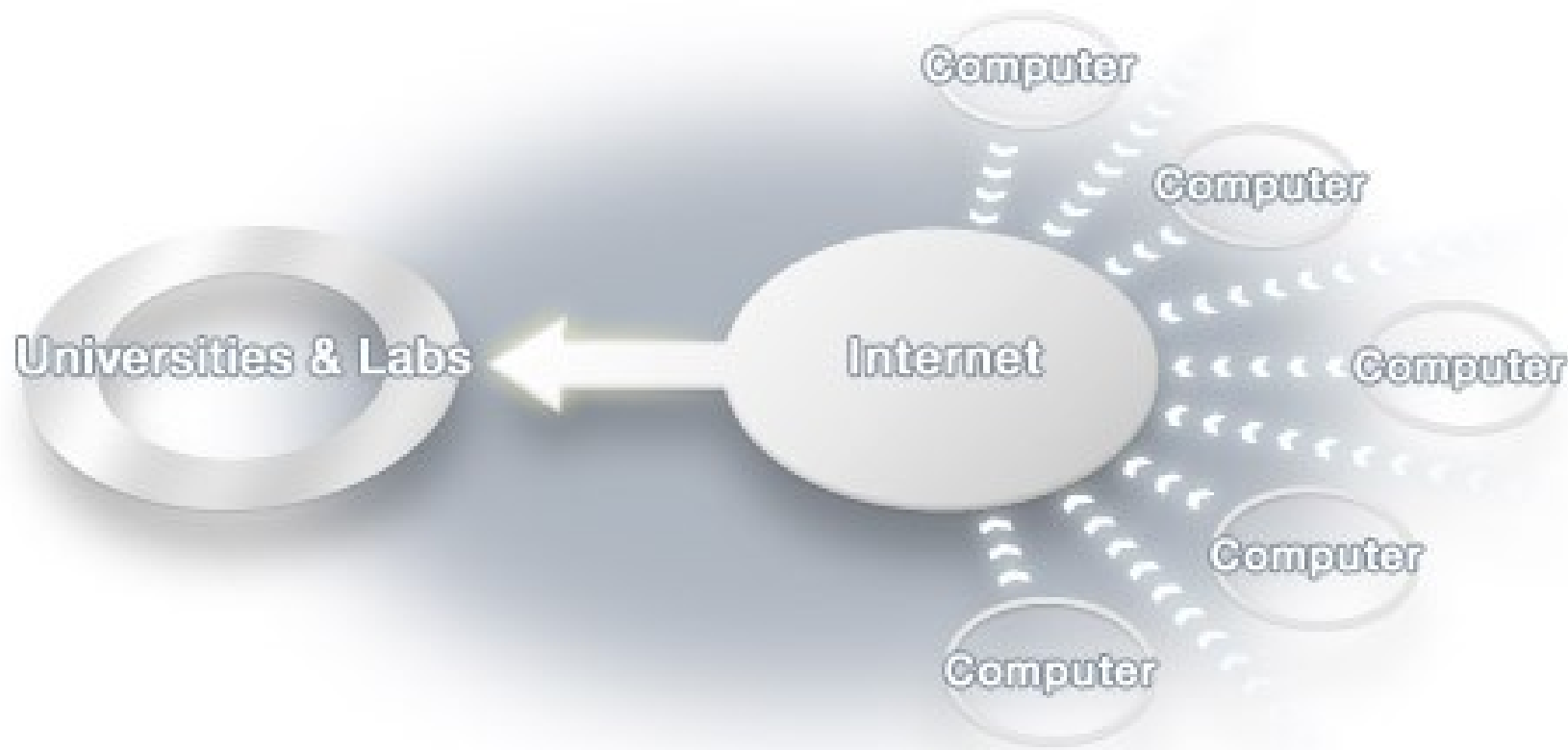
...

Distributed Objects



Back to Year 1990s ...

Brief History of Computing (3/5)



Source: <http://www.scei.co.jp/folding/en/dc.html>

Mainframe
Super
Computer

PC | Linux
Cluster
Parallel

Internet
Distributed
Computing

1997 Volunteer Computing
1999 SETI@HOME



2003 Globus Toolkit 2



2002 Berkley BOINC



2004 EGEE gLite



Back to Year 2000s ...

Brief History of Computing (4/5)



Source: <http://gridcafe.web.cern.ch/gridcafe/whatisgrid/whatis.html>

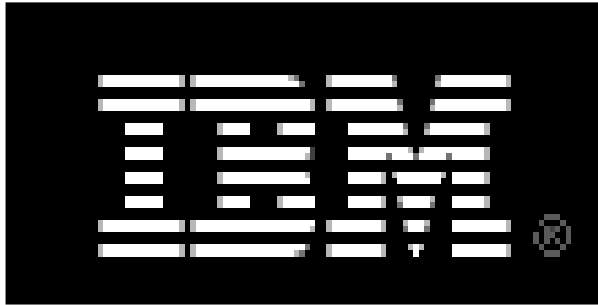
Mainframe
Super
Computer

PC | Linux
Cluster
Parallel

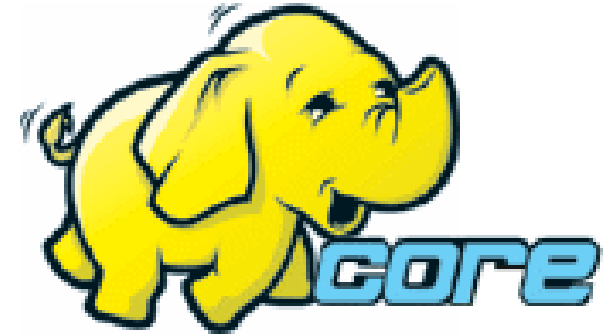
Internet
Distributed
Computing

Virtual Org.
Grid
Computing

2001 Autonomic Computing
IBM



2006 Apache Hadoop



2005 Utility Computing
Amazon EC2 / S3

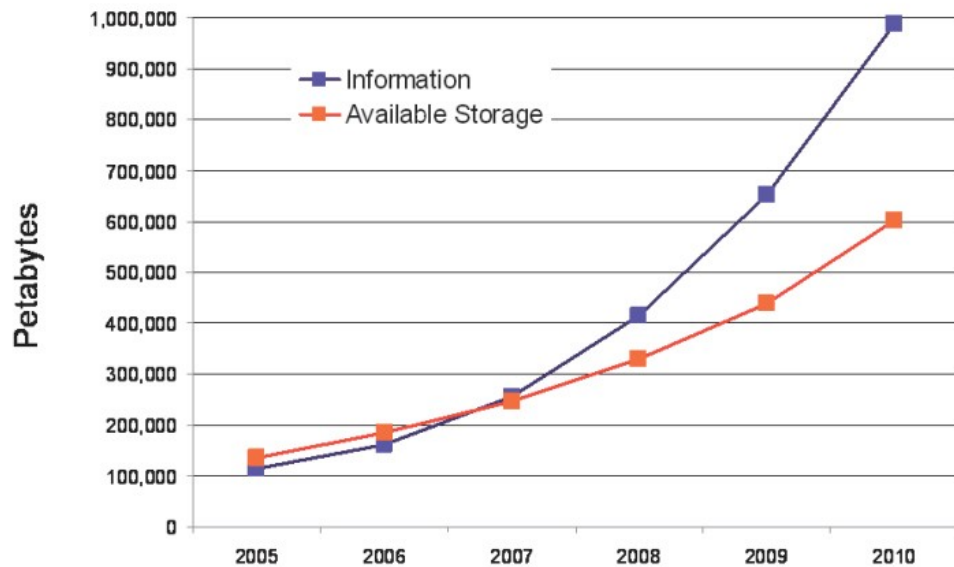


2007 Cloud Computing
Google + IBM



Back to Year 2007 ...

Information Versus Available Storage



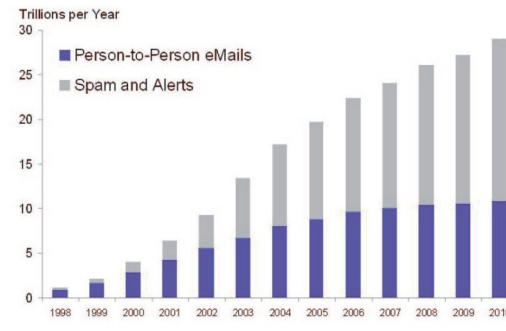
Source: <http://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf>

Source: IDC, 2007

2007 Data Explore

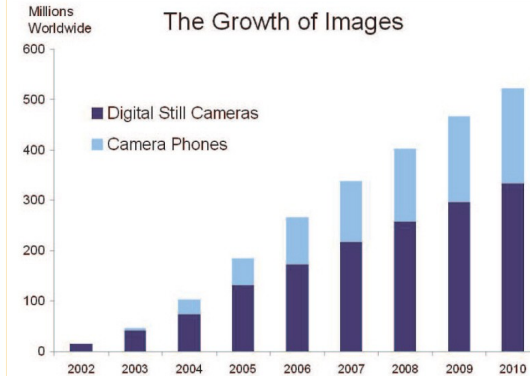
Top 1 : Human Genomics – 7000 PB / Year
Top 2 : Digital Photos – 1000 PB+/ Year
Top 3 : E-mail (no Spam) – 300 PB+ / Year

The Worldwide Growth of eMail



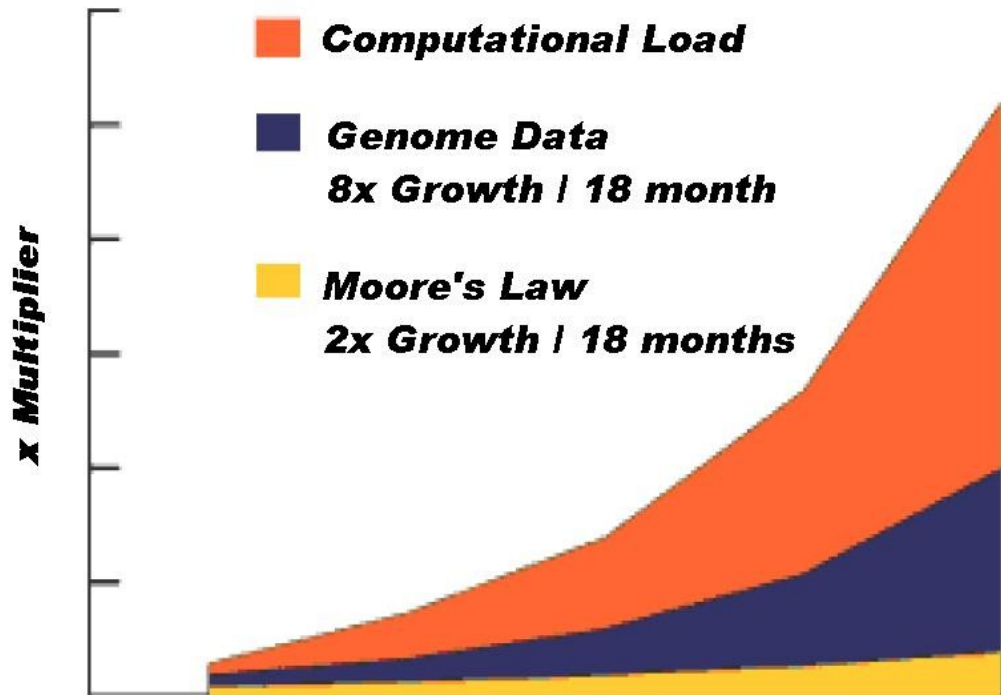
Source: IDC, 2007

The Growth of Images

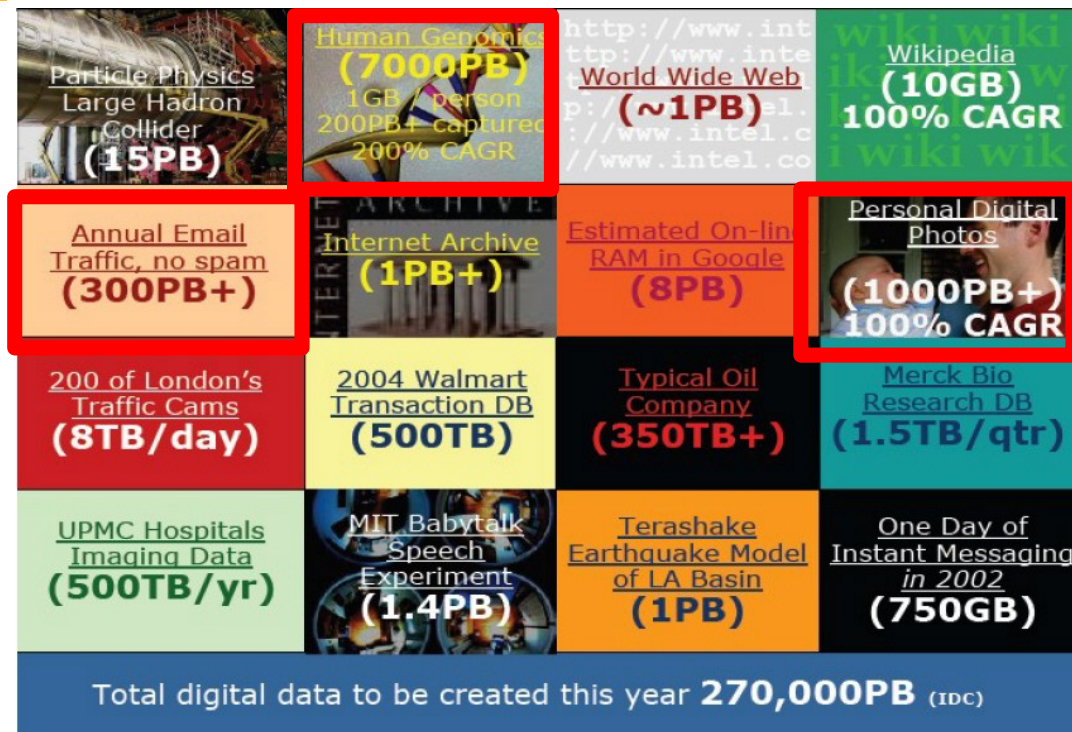


Source: IDC, 2007

Computational Load
Genome Data
 8x Growth / 18 month
Moore's Law
 2x Growth / 18 months

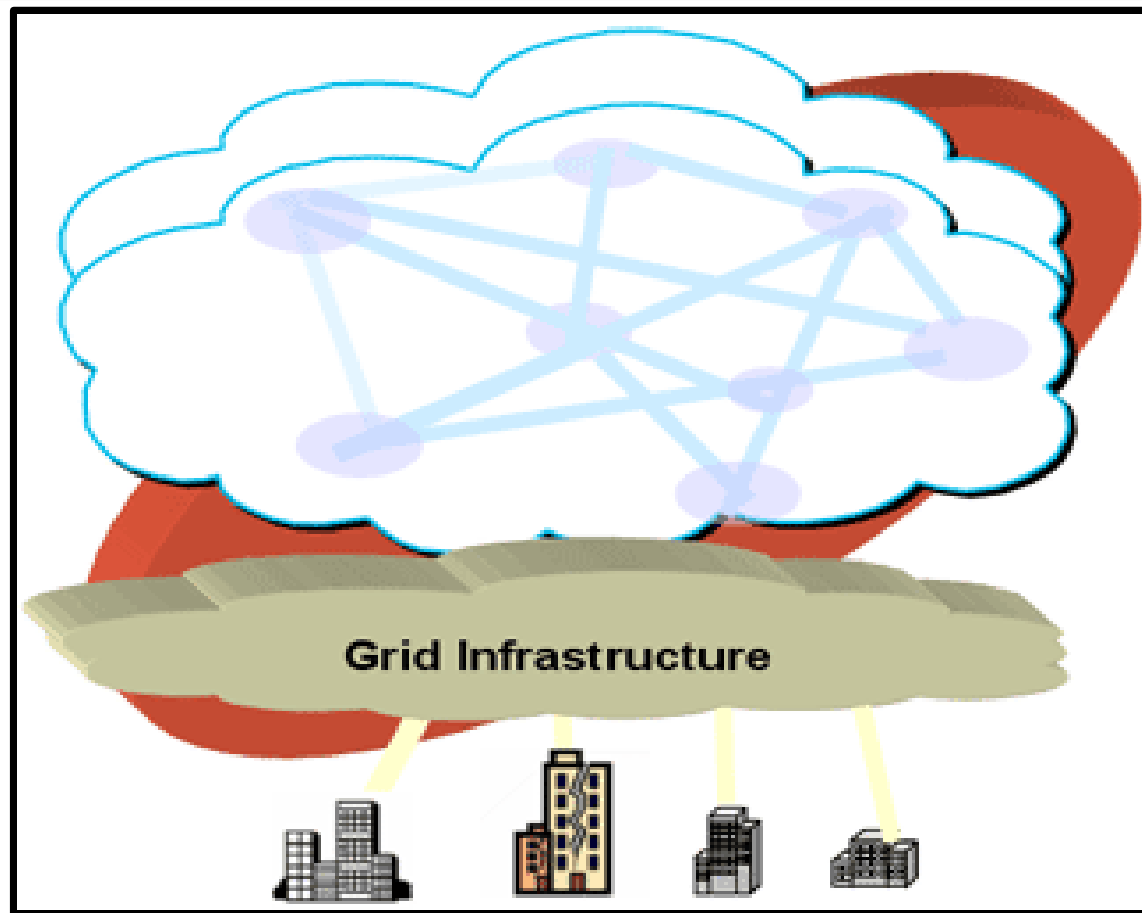


Source: http://lib.stanford.edu/files/see_pasig_dic.pdf



Phillip B. Gibbons, Data-Intensive Computing Symposium

Brief History of Computing (5/5)



Source: <http://mmdays.com/2008/02/14/cloud-computing/>

mainframe
super
computer

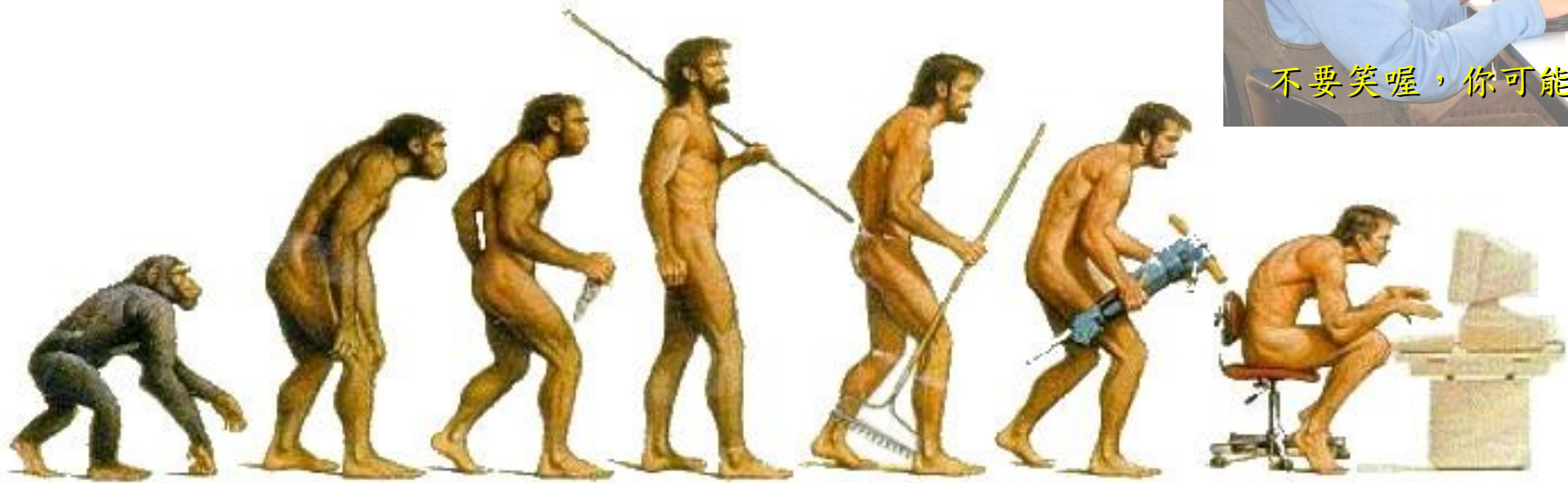
PC / Linux
Cluster
Parallel

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 ?!

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

This is why Cloud Computing matters ?!

這就是為什麼雲端運算變得熱門？！

What are the trend of next 10 years ?

什麼是下個十年的熱門技能？

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

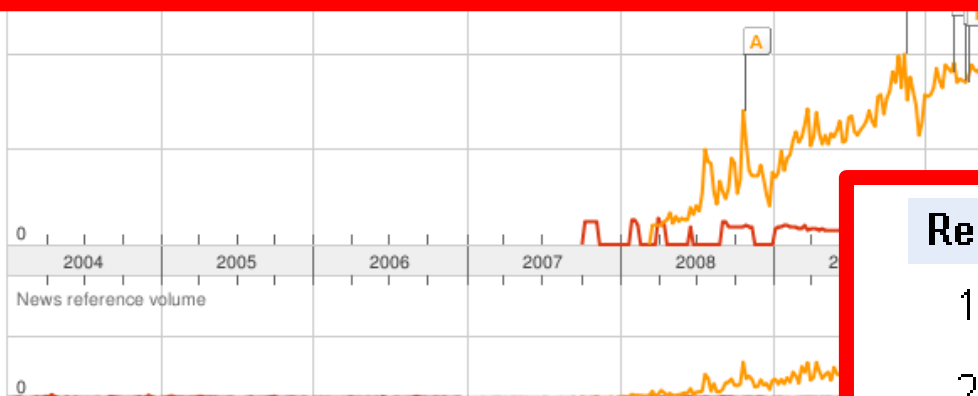
[Sign in](#) to see and export additional Tren

Search Volume Index

All regions

All years

"cloud" computing space



B Microsoft's cloud computing system is growing up

Philadelphia Inquirer - Nov 17 2009

C Google looks to be 'cloud-computing' rainmaker for other online business services

Winnipeg Free Press - Mar 10 2010

Regions

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

Regions

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

Cities

1. Bangalor
2. Mahape
3. Mumbai
4. Chennai
5. San Jos
6. Delhi, In

似乎亞洲國家特別熱愛雲端?! *Too Hot in Asia ?!*

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?!

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 !!

為了讓這些技能與產業接軌，亟需教育訓練²⁸!!



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

Source: 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？

Let's Talk about Public Cloud

讓我們先來談談公用雲端服務

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

Academia **學術**為主

IBM®

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**

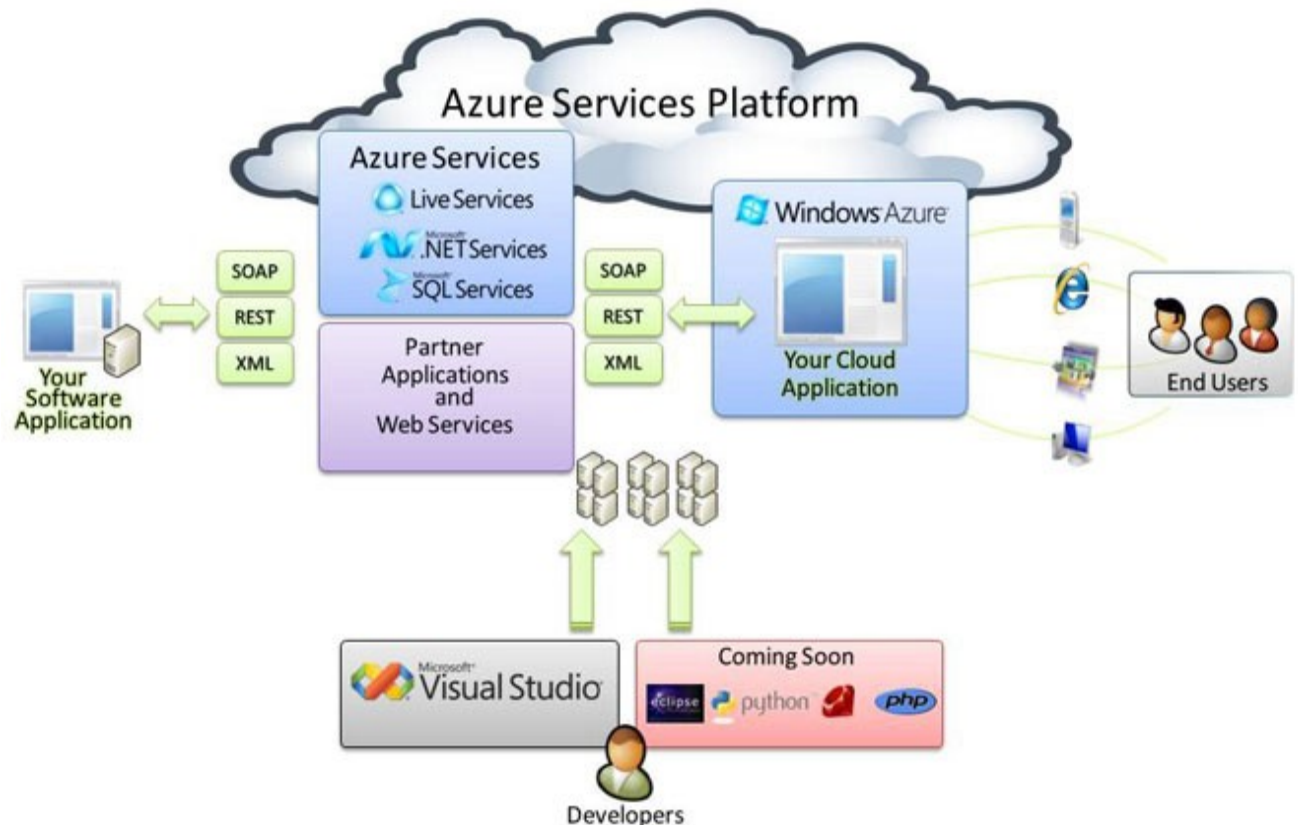
參考來源：
<http://eblog.cisanet.org.tw/post/Cloud-Computing.aspx>
<http://aws.amazon.com/ec2/pricing/>
<http://aws.typepad.com/aws/2010/02/aws-data-transfer-prices-reduced.html>
<http://aws.amazon.com/s3/#pricing>

- 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

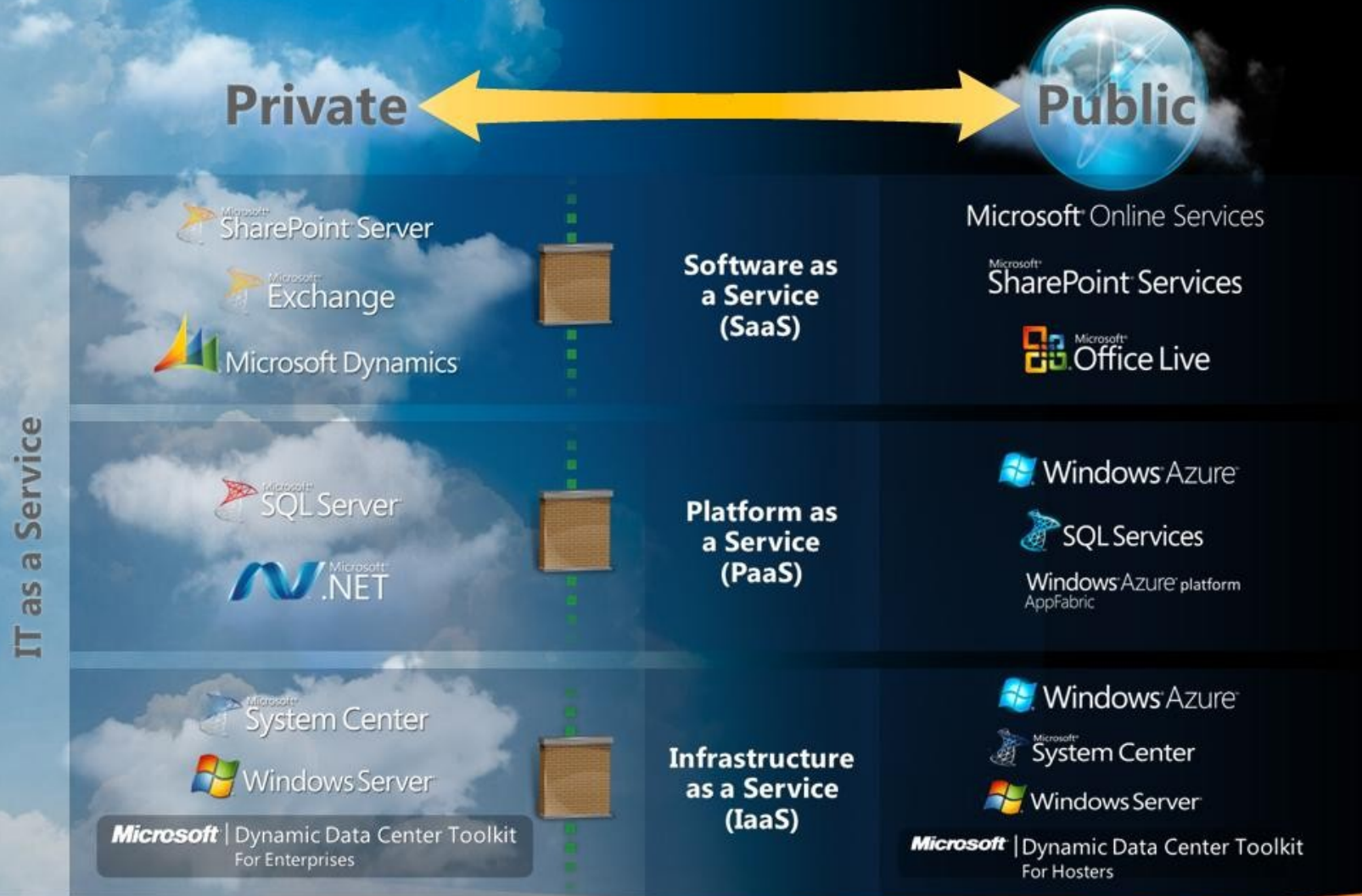


Public Cloud #3: *Microsoft* 微軟

- 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



Microsoft Cloud Computing 全貌



Dallas
→ **DaaS**

Azure
AppFabric
→ **PaaS**
(類似 **GAE**)

SQL Azure
→ **PaaS**
(雲端 **SQL**)

Window
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			

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*

Community Cloud

社群雲端

Academia 學術為主

以大型企業
為主要客戶
Enterprise is
key market



私有雲端

Private Cloud

Reference Cloud Architecture

雲端運算的參考架構

應用

Social Computing, Enterprise, ISV,...

程式語言

Web 2.0 介面, Mashups, Workflows, ...

控制

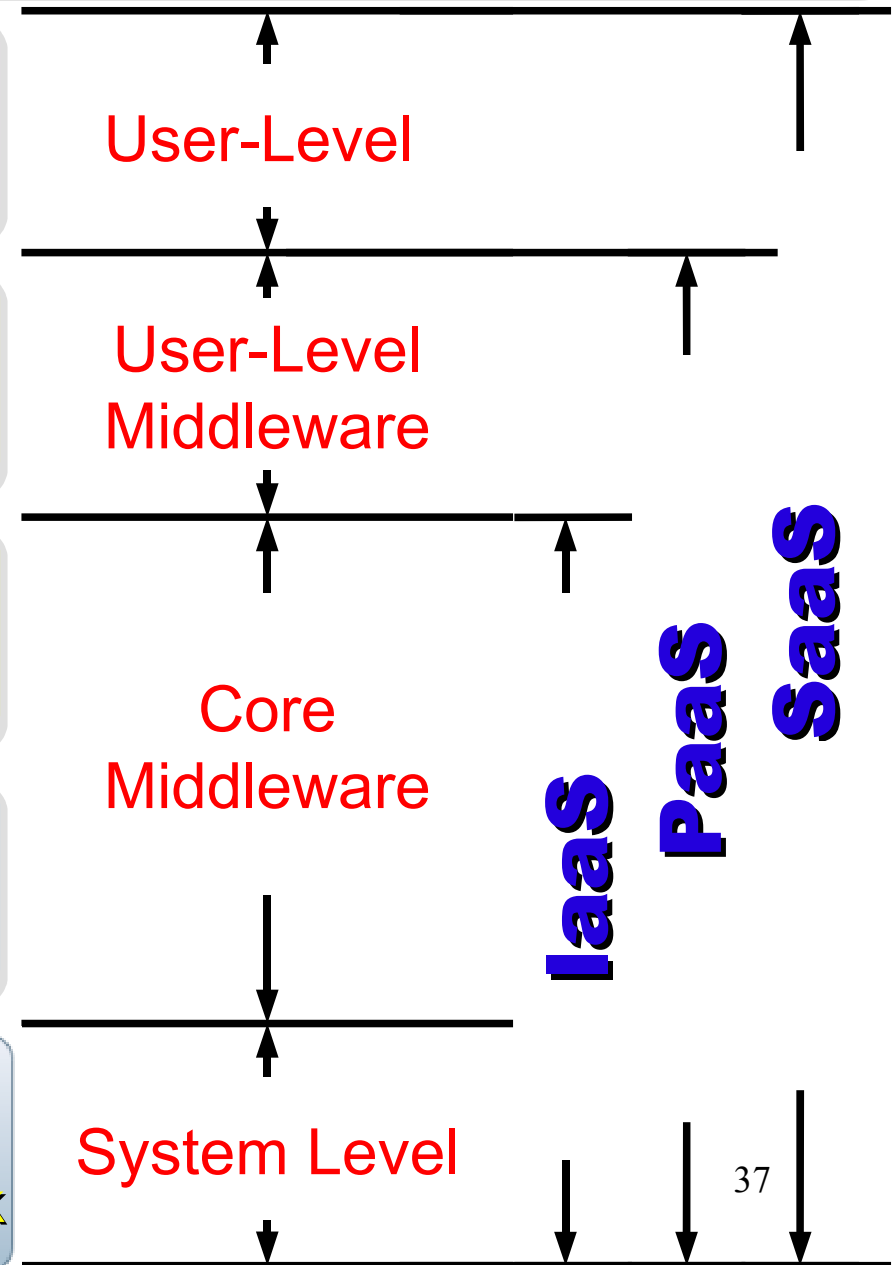
Qos Negotiation, Admission Control, Pricing, SLA Management, Metering...

虛擬化

VM, VM management and Deployment

硬體設施

Infrastructure: Computer, Storage, Network



Open Source for Private Cloud

建構私有雲端運算架構的自由軟體

應用

Social Computing, Enterprise, ISV,...

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

程式語言

Web 2.0 介面, Mashups, Workflows, ...

Hadoop (MapReduce),
Sector/Sphere, AppScale

控制

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

OpenNebula, Enomaly,
Eucalyptus, OpenQRM, ...

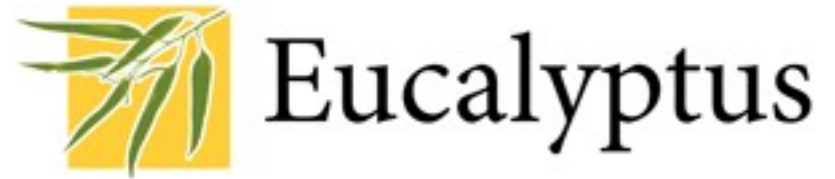
虛擬化

VM, VM management and Deployment

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

硬體設施

Infrastructure: Computer, Storage,
Network

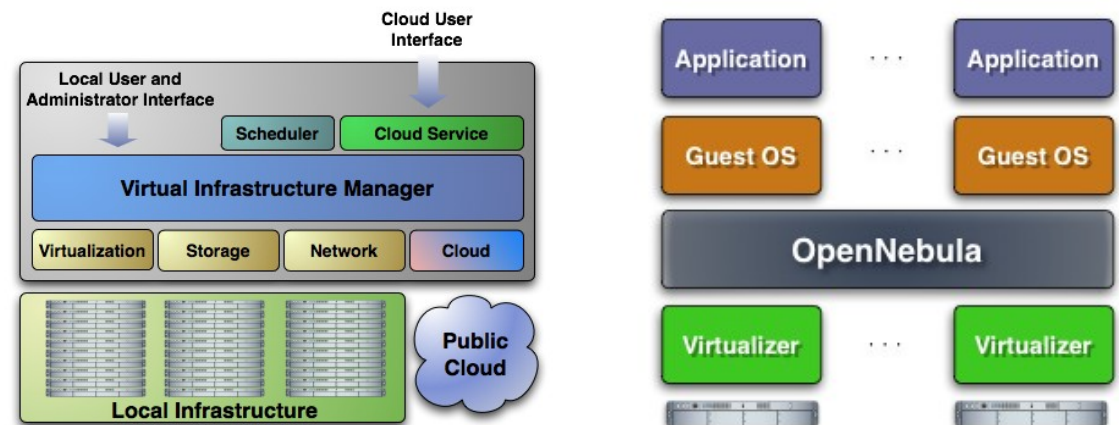


- <http://open.eucalyptus.com/>
- 原是加州大學聖塔芭芭拉分校(UCSB)的研究專案
- 目前已轉由Eucalyptus System這間公司負責維護
- 創立目的是讓使用者可以**打造自己的EC2**
- 特色是相容於 Amazon EC2 既有的用戶端介面
- 優勢是Ubuntu 9.04 已經收錄 Eucalyptus 的套件
- [Ubuntu Enterprise Cloud powered by Eucalyptus in 9.04](#)
- 目前有提供 Eucalyptus 的官方測試平台供註冊帳號
- 缺點：目前仍有部分操作需透過指令模式

- <http://www.opennebula.org>
- 由歐洲研究學會(European Union FP7)贊助
- 將實體叢集轉換成具管理彈性的虛擬基礎設備
- 可管理**虛擬叢集**的**狀態、排程、遷徙(migration)**
- 優勢是Ubuntu 9.04 已經收錄 OpenNebula 的套件
- 缺點：需下指令來進行虛擬機器的遷徙(migration)。



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



- <http://hadoop.apache.org>
- Hadoop 是 Apache Top Level 開發專案
- 目前主要由 Yahoo! 資助、開發與運用
- 創始者是 Doug Cutting，參考 Google Filesystem，以 Java 開發，提供 HDFS 與 MapReduce API。
- 2006 年使用在 Yahoo 內部服務中
- 已佈署於上千個節點。
- 處理 Petabyte 等級資料量。
- Facebook、Last.fm、Joost ... 等著名網路服務均有採用 Hadoop。



- <http://sector.sourceforge.net/>
- 由美國資料探勘中心 (National Center for Data Mining) 研發的自由軟體專案。
- 採用 C/C++ 語言撰寫，因此效能較 Hadoop 更好。
- 提供「類似」Google File System 與 MapReduce 的機制
- 基於[UDT高效率網路協定](#)來加速資料傳輸效率
- [Open Cloud Consortium](#)的 [Open Cloud Testbed](#)，有提供測試環境，並開發了[MalStone效能評比軟體](#)。



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2007年繼格網運算之後的新趨勢!!

Cloud Computing become new trend since year 2007 !!

WHY

資料集中、虛擬化、異業資源共享

Data-intensive, Virtualization, Heterogeneous

HOW

採用自由軟體也能打造私有雲端

Hadoop, Sectore/Sphere, Eucalyptus, and more



雲端運算對政府部門之衝擊與因應對策

What is the Impacts and How should public sectore adjust ?

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



Powered by **DRBL**

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: Share Public Data 共享公用資料

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

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

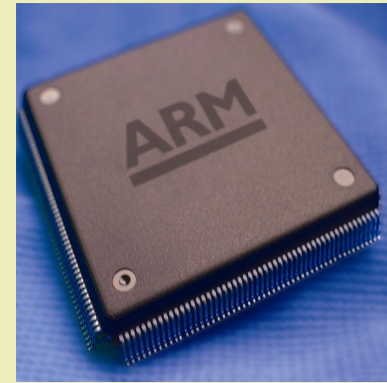
Source: Gartner Executive Programs : “ Leading in Times of Transition: The 2010 CIO Agenda ”⁴⁶

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

Database Virtualization

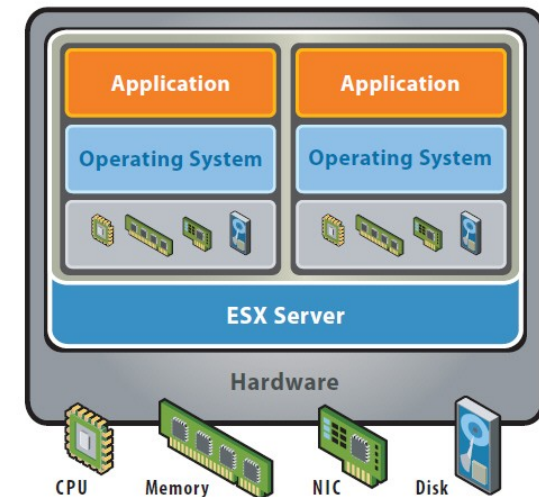
庫資料

Data Virtualization

Adjust #1: Virtualization

公部門導入雲端參考策略 (1) **Virtualization**

- 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[™]



How to Evaluate the need of Virtualization ?

如何評估對虛擬化的需求??

- 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 ?

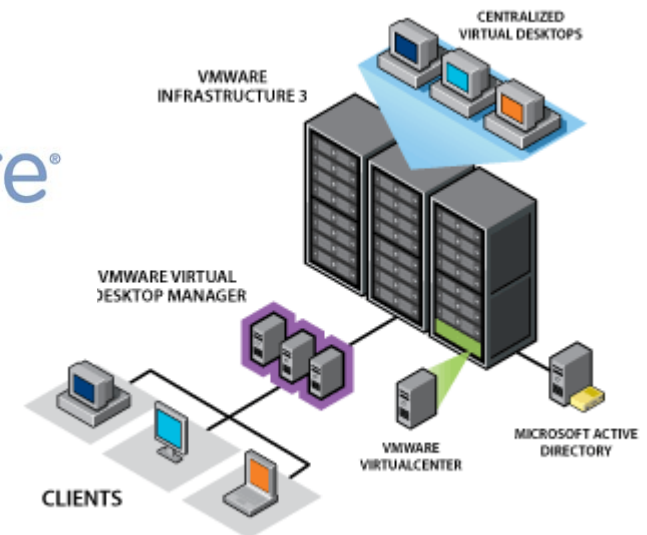
下次改採購精簡型電腦與桌面虛擬化??

- 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)


Windows Server® 2008
Hyper-V™

 **CITRIX®**

 **vmware®**



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.nchc.org.tw/cloud/attachment/wiki/jazz/09-10-14/TSMC/08-08-17_DRBL_Green_Computing_v2.pdf

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）

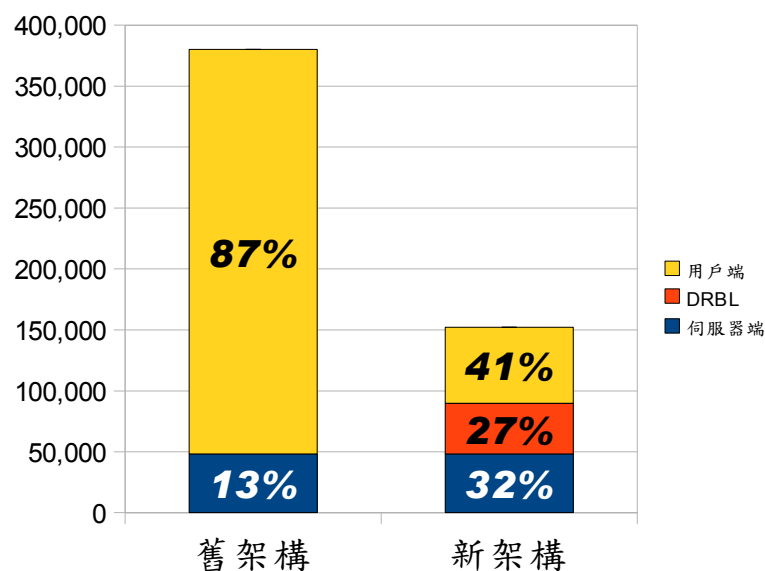
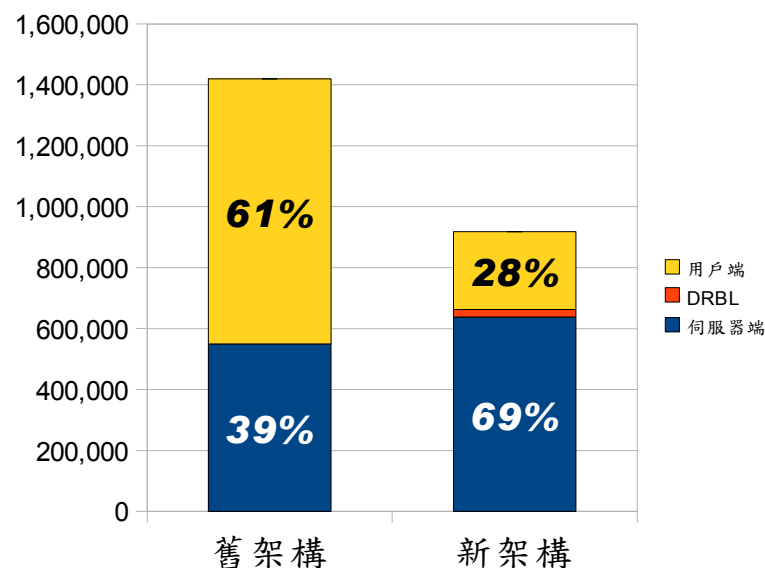
- 降低軟體授權開支

- 購買 **Windows** 作業系統與 **Office** 授權只為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	

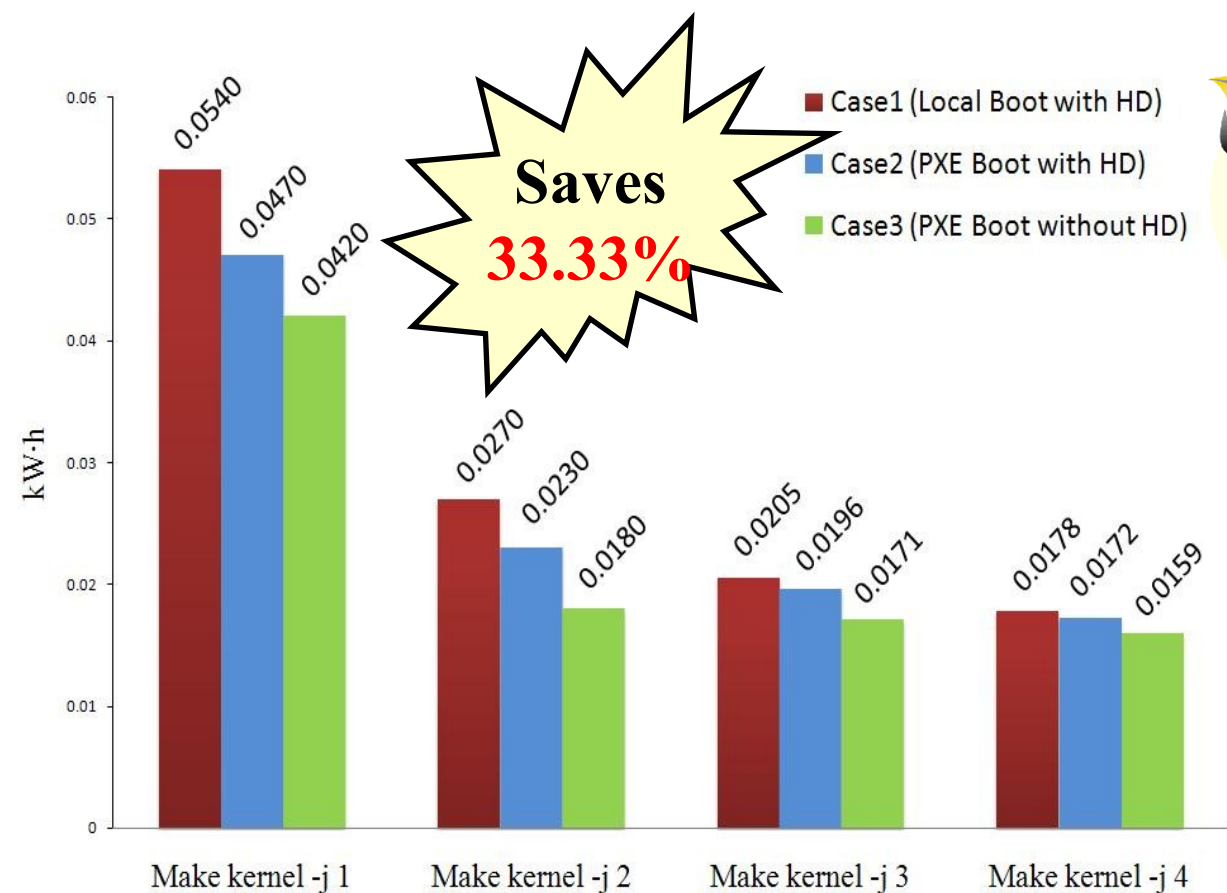
節省 **50 萬**
總**建置**成本

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，
企鵝龍 (DRBL) 無碟架構不僅省電最高達 33%，效能還比較好



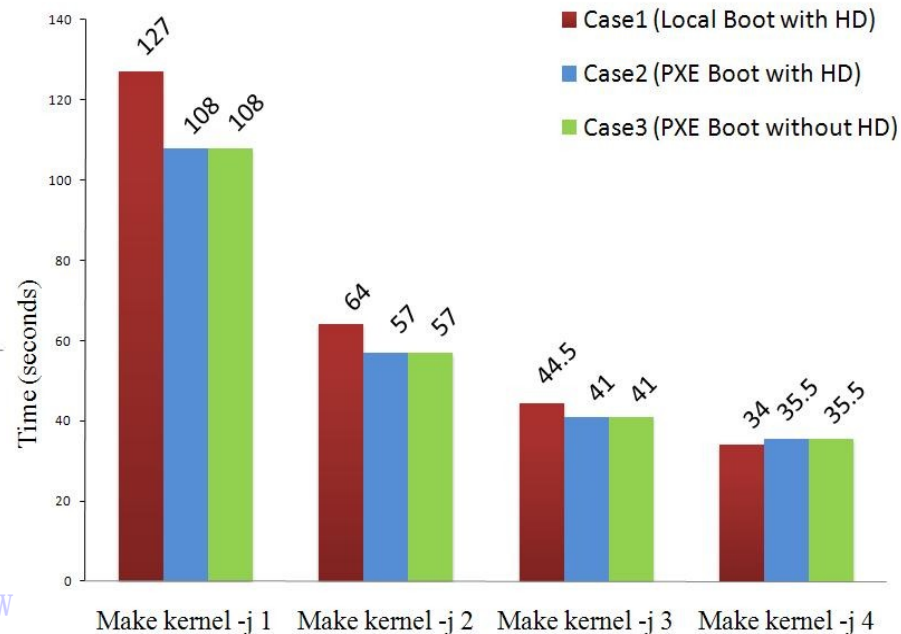
企鵝龍 **DRBL**
(Diskless Remote Boot in Linux)

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: Share Public Data

公部門導入雲端參考策略(2) 共用公開資料

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?!

Adjust #2: Share Public Data

公部門導入雲端參考策略(2) 共用公開資料

- 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 #3: Single AAA Service

公部門導入雲端參考策略(3) 單一身分認證

Everything as a Service 啥米鬼都是一種服務

- AaaS
- BaaS
- CaaS
- DaaS
- DBaaS
- EaaS
- FaaS
- GaaS
- HaaS
- IaaS

Architecture as a Service
Business as a Service
Computing as a Service
Data as a Service
Database as a Service
Ethernet as a Service
Frameworks as a Service
Globalization as a Service

Customer-Oriented

Authentication, Authorization, Accounting as a Service

- OaaS
 - **SaaS**
 - **PaaS**
 - TaaS
 - VaaS
- Software as a Service
Lending as a Service
Mashups as a Service
Organization or Operations as a Service
Software or Storage as a Service
Platform as a Service
Technology or Testing as a Service
Voice as a Service

引用自：

https://www.ibm.com/developerworks/mydeveloperworks/blogs/sbose/entry/gathering_clouds_of_xaas

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 網頁標準

Open Platform

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

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 ?

Concurrent 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 !!



私有雲端技術之產業應用

Enterprise Applications of Private Cloud

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



Powered by **DRBL**

Cyberinfrastructure of TSMC

台積電的資訊架構 @ Year 2000 ?

後端資訊系統 (internal operation)

FPS 預測規劃系統

Forecast Planning System

TOM 全方位訂單管理系統

Total Order Management (資訊流)

MES 製造執行管理系統

Manufacturing Execution System (物流)

CRP

VMI

JIT

SAP ERP 企業資源規劃

Enterprise Resource Planning (金流)

PIDB
Product Information Data-Base
產品資訊資料庫

前端資訊系統 eFoundry

TSMC-Direct

TSMC-Online 1.0 / 2.0

TSMC-YES

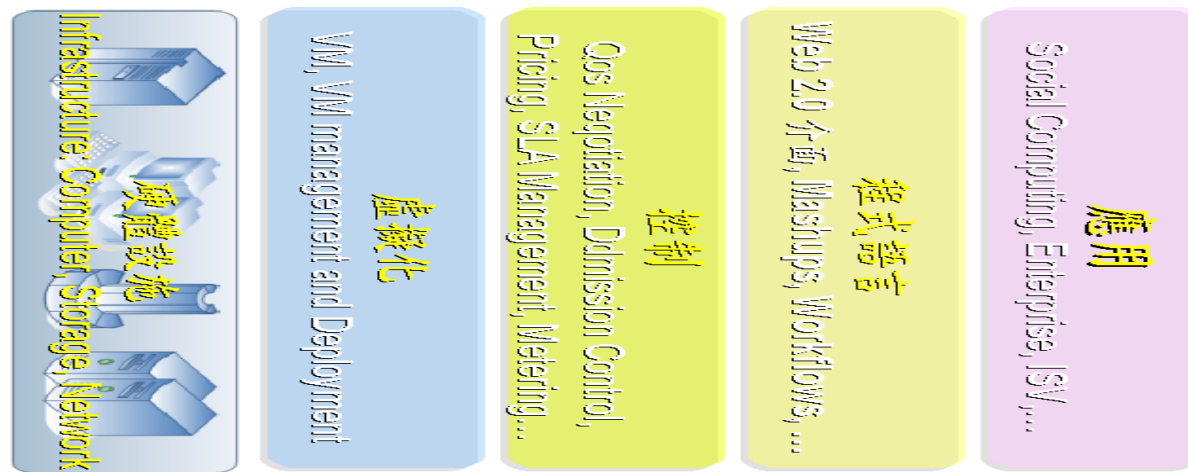
Internet Layout Viewer

Design Sphere Access

Logistics
商業流共享

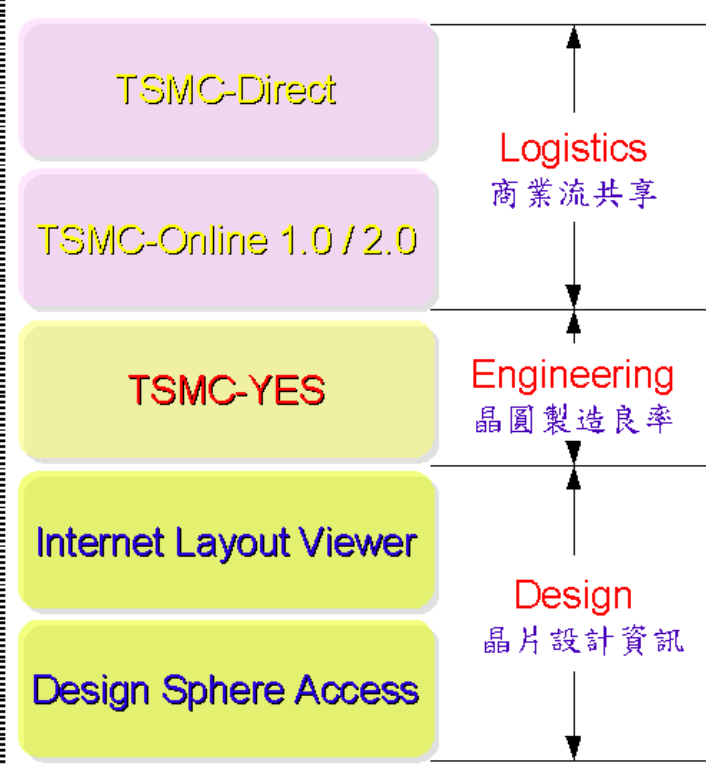
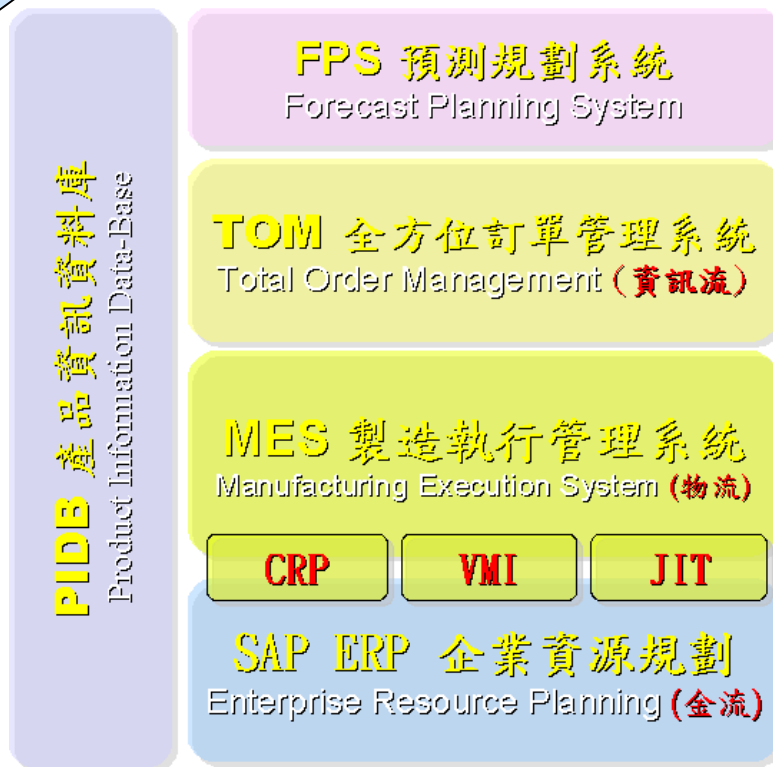
Engineering
晶圓製造良率

Design
晶片設計資訊



後端資訊系統 (internal operation)

前端資訊系統 eFoundry



Possible Cloud Service for TSMC (1)

台積電雲端導入參考策略 (1) **Grid WebOS**

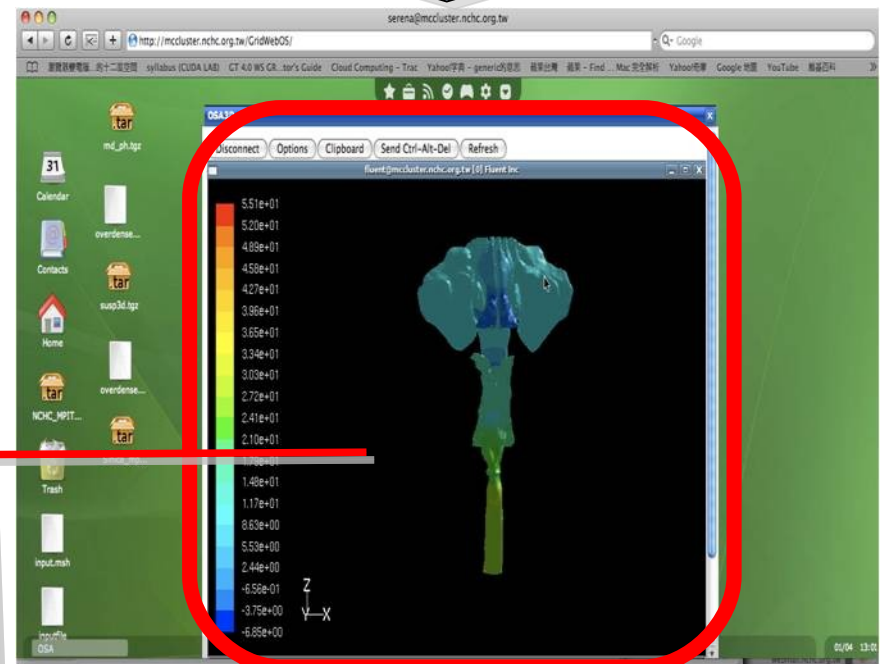
關於 Grid WebOS 的更多資訊，請參考
<https://service.nchc.org.tw/crs/GridComputing/GridComputing.php>

- 導入前：
 - 每間小公司都需要採購專用軟體
 - 商業軟體**授權金額高**
- 導入評量指標：
 - **總軟體使用率** (時間、人次)
- 導入需求技術：
 - **權限控管、授權分配 (排程) 機制**
 - **使用率統計** → **成本攤提比例**
- 預期效益：
 - **上下游供應鏈共同分攤軟體授權**
 - **協同驗證**設計與製程可行性

Internet Layout Viewer

Design Sphere Access

?

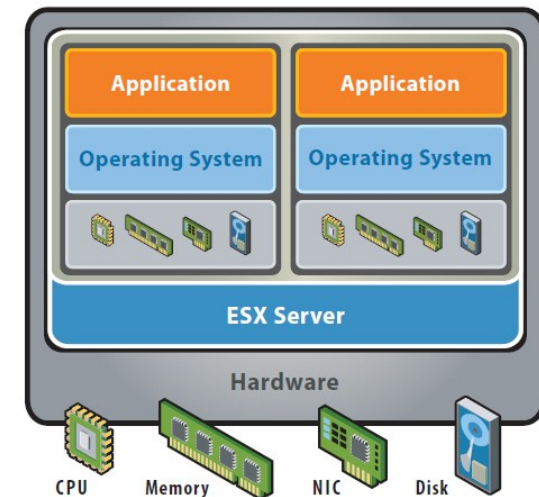


Ex. 讓協力廠商遠端執行 **Cadence**

Possible Cloud Service for TSMC (2)

台積電雲端導入參考策略 (2) Virtualization

- 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™



Possible Cloud Service for TSMC (3)

台積電雲端導入參考策略 (3) Skills for 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





工商服務時間 * @_@ *

About NCHC Cloud Computing Research Group ...

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



Powered by **DRBL**

NCHC Cloud Computing Research Group

團隊小檔案：國網中心雲端運算研究小組

- 主要研究雲端運算的基礎架構組成元件
- 團隊成員：6 名
 - 王耀聰 – drbl-xen / drbl-hadoop (~6 Years) 架構
 - 陳威宇 – Hadoop / NutchEz / ICAS (~3 Years) 應用
 - 郭文傑 – Xen / OpenNebula / Eucalyptus (~3 Years) 元件
 - 涂哲源 – Xen GPU / OpenMP / VirtualGL (~3 Years) 元件
 - 鄭宗碩 – Google App Engine (~2 Years) 新技術
 - 鄧偉華 – AMQP / OpenID (~2 Years) 新技術
- 定位：
 - 研發快速佈建軟體，提供實驗平台服務，開辦訓練課程育才
- 獨特性：
 - 基於企鵝龍 (DRBL)，可快速佈署雲端運算的叢集環境

站在巨人的肩膀－國網中心自由軟體開發

多元化資訊教學的新選擇！

以個人叢集電腦 (PC Cluster) 經驗發展 DRBL&Clonezilla



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

是自由！不是免費…

分送、修改、存取、使用軟體的自由。免費是附加價值。

何謂企鵝龍 DRBL ??

- **Diskless Remote Boot in Linux**
- 網路是便宜的，人的時間才是昂貴的。
- 企鵝龍簡單來說就是.....
 - 用網路線取代硬碟排線
 - 所有學生的電腦都透過網路连接到一台伺服器主機



**Diskfull
PC**



=



+



+



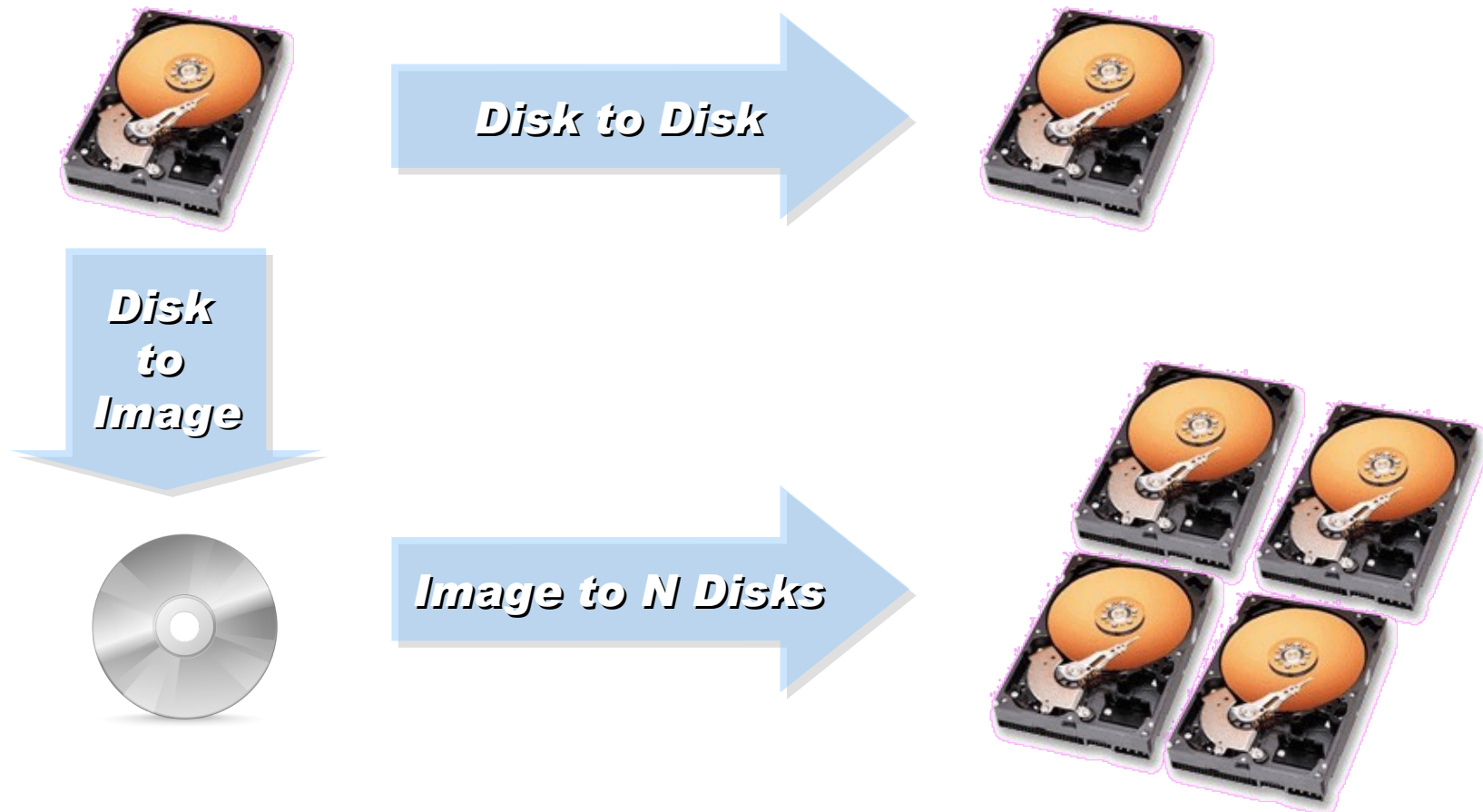
**Diskless
PC**



Server

何謂再生龍 Clonezilla ??

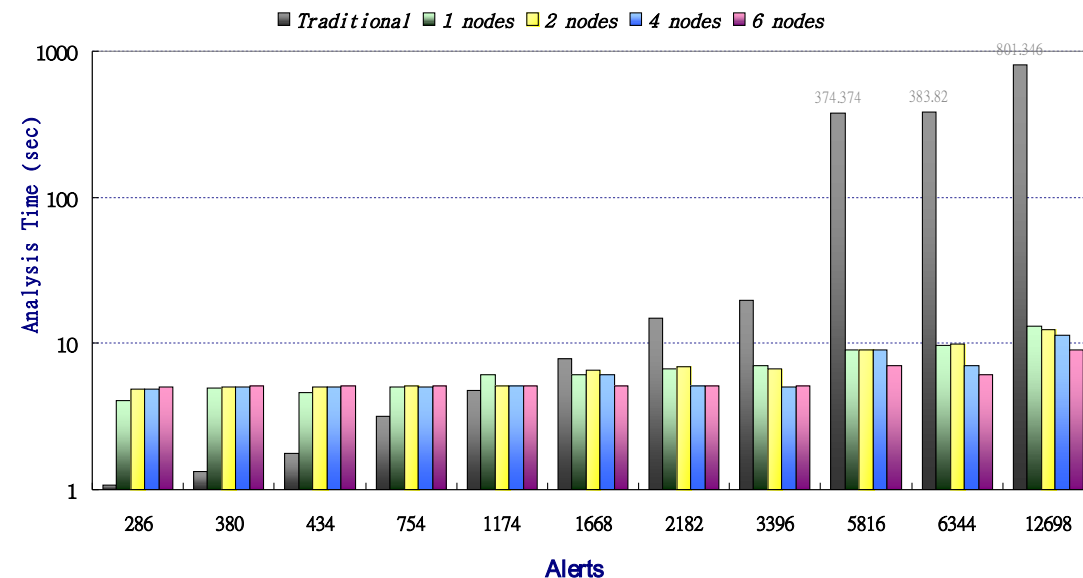
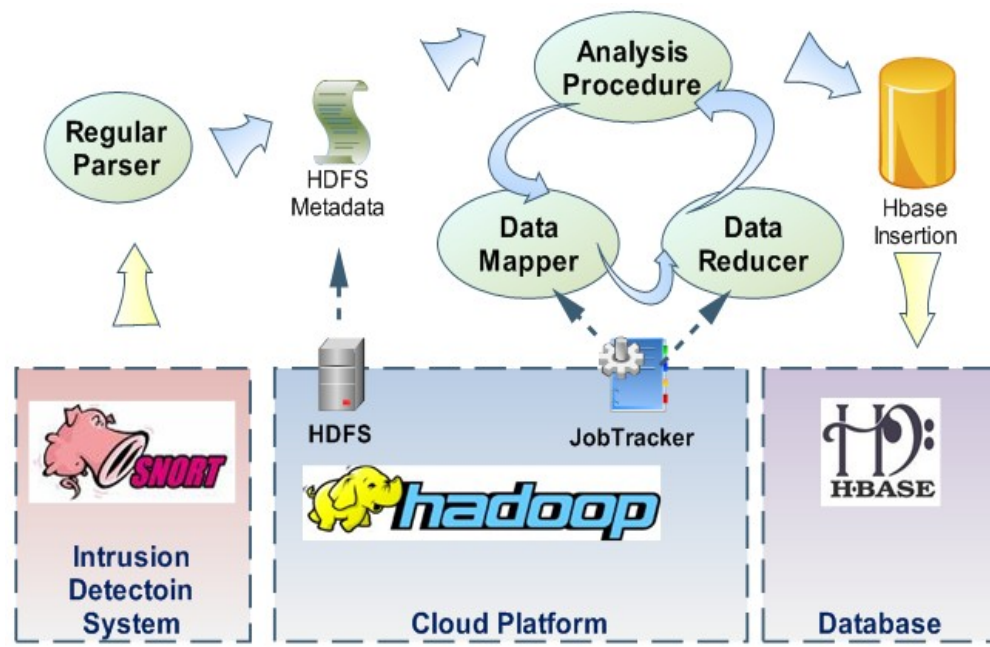
- **Clone** (複製) + **zilla** = **Clonezilla** (再生龍)
- 裸機備分還原工具
- **Norton Ghost** 的自由軟體版替代方案



軟體研發 (1) : 雲端入侵偵測分析系統

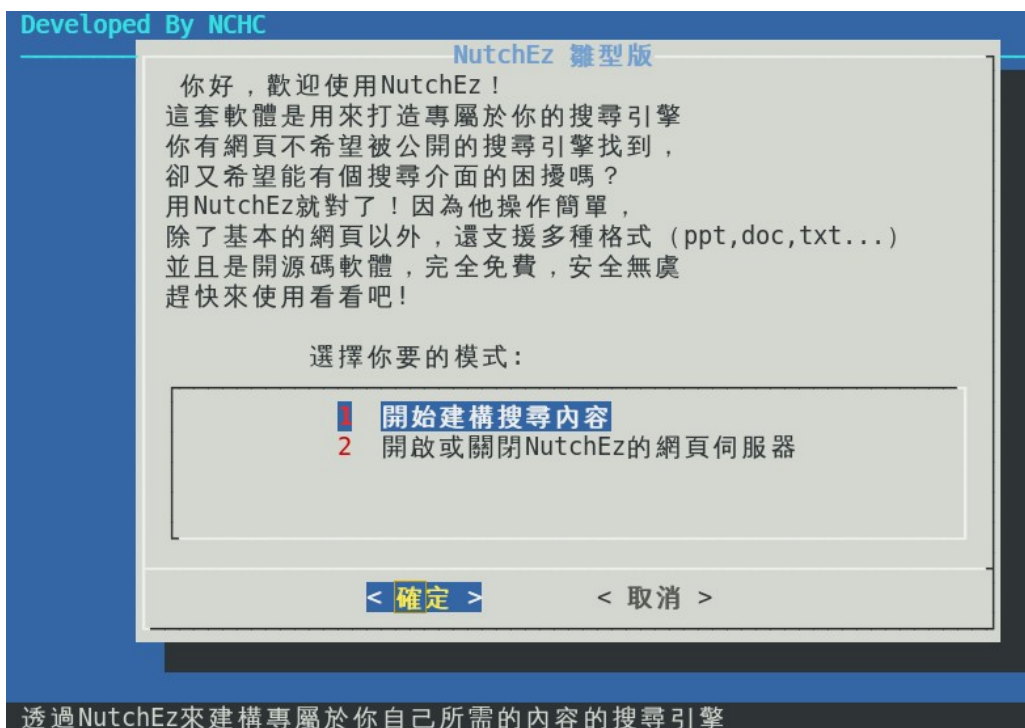
(IDS-log Cloud Analysis System , ICAS)

- 持續開發中，待整理套件
- 結合 **Hadoop** 與 **HBase** 來處理 **SNORT** 產生的網路入侵報告。
- 雲端運算處理資料格式相似且資料量大的情況下，能展現其效益，並提供高容錯率、低獨占系統資源、多工作同時執行等能力
- **Key-Value** 資料庫寫入慢，讀取效能相對快，但缺乏其他語言支援。
關聯式資料庫對小量資料的讀寫的效率較好，且支援的語言也較多。



軟體研發 (2) : 簡易架設個人搜尋引擎 (NutchEz)

- 已釋出中文版套件 <http://trac.nchc.org.tw/cloud/wiki/NutchEz>
- 合適用來建立屬於組織內部的網頁搜尋引擎
- 核心為強大的 **Nutch**，建於 **Hadoop** 上，貢獻：簡化安裝步驟
- 效能數據：搜尋 **699 doc, 322 pdf, 9 ppt, 13 odt**. 費時 **11 分**
- 系統負載：**CPU Quad 4 2.4G (19%)/ 4GB RAM (20%)**




軟體研發 (3) : 用企鵝龍佈署 Hadoop 雲端實驗環境

- 持續開發中，待整理套件
- **drbl-hadoop** – 掛載本機硬碟給 **HDFS** 用

svn co <http://trac.nchc.org.tw/pub/grid/drbl-hadoop>


- **hadoop-register** – 註冊網站與 **ssh applet**

svn co <http://trac.nchc.org.tw/pub/cloud/hadoop-register>

**trac**
Integrated SCM & Project Management

root / **drbl-hadoop-0.1**

Name ▲
⬆ ../
📄 drbl-hadoop
📄 drbl-hadoop-mount-disk

**trac**
Integrated SCM & Project Management

root / **hadoop-register**

Name ▲	Size	Rev	Age	Last
⬆ ../				
▸ 📁 etc		103	4 weeks	wa
📄 adduser.php	1.3 kB	85	6 weeks	wa

實驗服務：hadoop.nchc.org.tw 多人雲端實驗叢集

- **DRBL Server - 1 台 (hadoop)**，加大 **/home** 與 **/tftpboot** 空間。
- **DRBL Client - 19 台 (hadoop101~hadoop119)**
- 使用 **Cloudera** 的 **Debian** 套件，並針對多人環境進行讀寫權限加強。
- 使用 **drbl-hadoop** 的設定跟 **init.d script** 來協助部署
- 使用 **hadoop-register** 來提供使用者註冊與 **ssh applet** 介面

The image shows two side-by-side screenshots. The left screenshot is a terminal window titled 'Hadoop 帳號 - Mozilla Firefox' showing a Linux prompt and system information. The right screenshot is a web browser window titled 'hadoop Hadoop Map/Reduce Administration - Mozilla Firefox' showing the Hadoop administration interface.

Terminal Window (Left):

```
Linux hadoop 2.6.26-2-amd64 #1 SMP Fri Mar 27 04:02:59 UTC 2009 x86_64

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

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Thu Jul 16 09:05:28 2009 from wr189-050.nchc.org.tw
hadoop004@hadoop:~$
```

Web Browser Window (Right):

hadoop Hadoop Map/Reduce Administration

State: RUNNING
Started: Sun Jul 19 22:48:19 EDT 2009
Version: 0.18.3-4cloudera0.3.0, r
Compiled: Fri May 29 23:29:49 UTC 2009 by root
Identifier: 200907192248

Cluster Summary

Maps	Reduces	Total Submissions	Nodes	Map Task Capacity	Reduce Task
0	0	711	19	38	38

Running Jobs

Running Jobs

完成

人才培育：雲端運算基礎課程（一～三）開放課程

- 雲端運算基礎課程（一）：Hadoop 簡介、安裝與實作
- 雲端運算基礎課程（二）：Xen 虛擬化叢集建置、管理與應用
- 雲端運算基礎課程（三）：Google App Engine 體驗課程
- 最新課程訊息與課程錄影詳見 <http://trac.nchc.org.tw/cloud/>

教育訓練網

最新消息 | 會員專區 | 常見問題 FAQ | 住宿資訊 | 交通導引 | 電子報 | 電子書

課程總覽 ▾ 最近六個月的課程 ▾ 課程介紹

課程編號：NE-2009-TH06

課程名稱：雲端運算基礎課程(一)：Hadoop簡介、安裝與範例實作

課程領域：電腦及網路

相關領域：無

上課方式：實體教室

上課地點：竹科 B 教室 **竹**

上課時間：2009/11/24 (二) ~ 2009/11/25 (三) 09:30 ~ 16:30

上課總天數：2 天，共計 12 個小時

報名截止(含)：2009/11/22 (日) 17:00

費截止(含)：2009/11/23 (一) 05:00

費截止(含)：2009/11/23 (一) 17:00

提供午餐：是

招生人數：8 ~ 20 人

講師：國家高速網路與計算中心 王耀聰 先生

國家高速網路與計算中心 陳威宇 先生

教育訓練網

最新消息 | 會員專區 | 常見問題 FAQ | 住宿資訊 | 交通導引 | 電子報 | 電子書

課程總覽 ▾ 課程介紹

課程編號：NE-2009-CH05

課程名稱：雲端運算基礎課程(二)：Xen 虛擬化叢集建置、管理與應用

課程領域：電腦及網路

相關領域：無

上課方式：實體教室

上課地點：台中 電腦教室 A **中**

上課時間：2009/10/27 (二) ~ 2009/10/28 (三) 09:30 ~ 16:30

上課總天數：2 天，共計 12 個小時

報名截止(含)：2009/10/25 (日) 17:00

費截止(含)：2009/10/26 (一) 05:00

費截止(含)：2009/10/26 (一) 17:00

提供午餐：是

招生人數：8 ~ 20 人

講師：國家高速網路與計算中心 徐哲源 先生

國家高速網路與計算中心 郭文傑 先生

教育訓練網

最新消息 | 會員專區 | 常見問題 FAQ | 住宿資訊 | 交通導引 | 電子報 | 電子書

課程總覽 ▾ 課程搜尋結果 ▾ 課程介紹

課程編號：NE-2009-CH06

課程名稱：雲端運算基礎課程(三)：Google App Engine 體驗課程

課程領域：電腦及網路

相關領域：無

上課方式：實體教室

上課地點：台中 電腦教室 A **中**

上課時間：2009/11/30 (一) 09:30 ~ 16:30

上課總天數：1 天，共計 6 個小時

報名截止(含)：2009/11/27 (五) 17:00

費截止(含)：2009/11/27 (五) 17:00

費截止(含)：2009/11/27 (五) 17:00

提供午餐：是

招生人數：8 ~ 20 人

講師：國家高速網路與計算中心 鄭宗碩 先生

報名費用：一般人士 1000 元

學生 500 元

對學界的幫助 (1): 實驗叢集間接促成研究成果

- 促成台大資工系資訊網路與多媒體研究所發表論文至 **ACM Multimedia 2009**
- 自 **2009** 年四月至 **2009** 年九月，雲端實驗叢集共註冊 **238** 人，服務 **37** 個學術單位 (計 **154** 人)，**5** 個研究單位 (計 **21** 人)、**19** 間業界公司 (計 **20** 人)、**2** 所醫院 (計 **3** 人) 及不願提供單位的一般民眾計 **30** 人。累計於五個月內執行 **3341** 個 **Job**。
- 註冊人數排行前五大依序為**交通大學**、**台灣大學**、**成功大學**、**中央大學**與**陽明大學**

Canonical Image Selection and Efficient Image Graph Construction for Large-Scale Flickr Photos

Liang-Chi Hsieh, Kuan-Ting Chen, Chien-Hsing Chiang, Yi-Hsuan Yang, Guan-Long Wu
Chun-Sung Ferng, Hsiu-Wen Hsueh, Charng-Rung Tsai, Winston H. Hsu
National Taiwan University, Taipei, Taiwan

viirya@gmail.com, ktchen@cmlab.csie.ntu.edu.tw, {pacifistboy, affige}@gmail.com, {b95109, b95108, b95057, b95093, winston}@csie.ntu.edu.tw

ABSTRACT

Efficient image search clustering is prominent for image search engines for exponentially growing photo collections. In this work, we propose an image search clustering approach which selects multiple canonical images from image search results and constructs image clusters in real time on an image sub-graph for the search results. The efficiency is achieved with the help of offline-computed image context graphs by distributed computing methods. Extending our prior works, we demonstrate the results of the proposed canonical image selection and preliminary outcomes of large-scale image graph construction in this proposal. We experiment in Flickr550 dataset, containing 540,321 Flickr photos.

Categories and Subject Descriptors: H.3.5 [Informa-

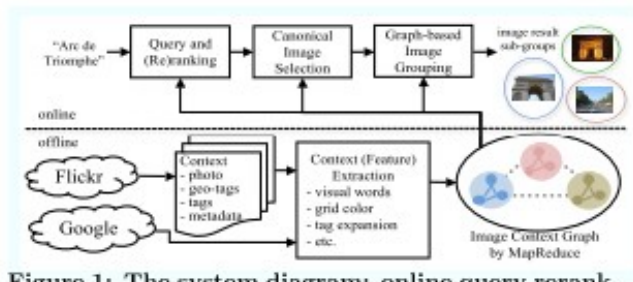
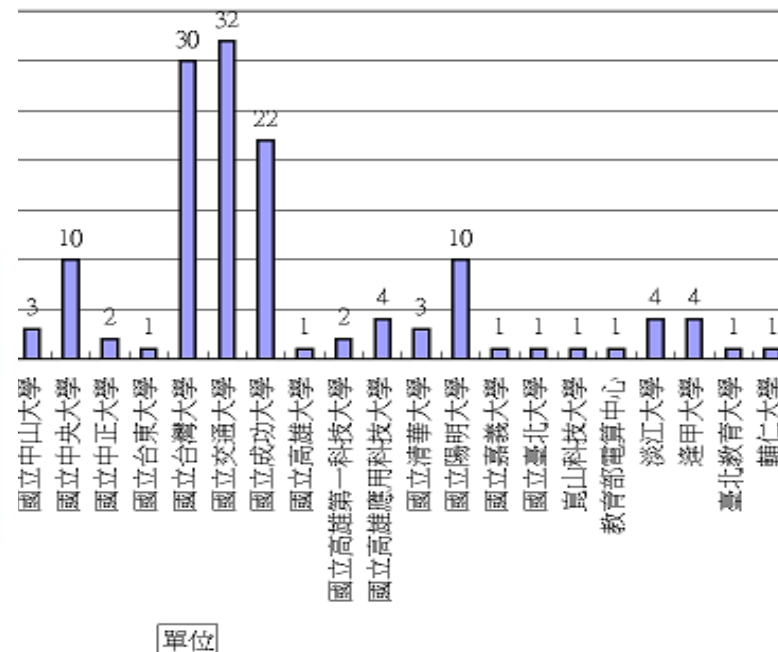


Figure 1: The system diagram: online query reranking, canonical image selection, and image result grouping are based on the image context graph, constructed offline by Hadoop MapReduce over rich context features in Flickr photos.

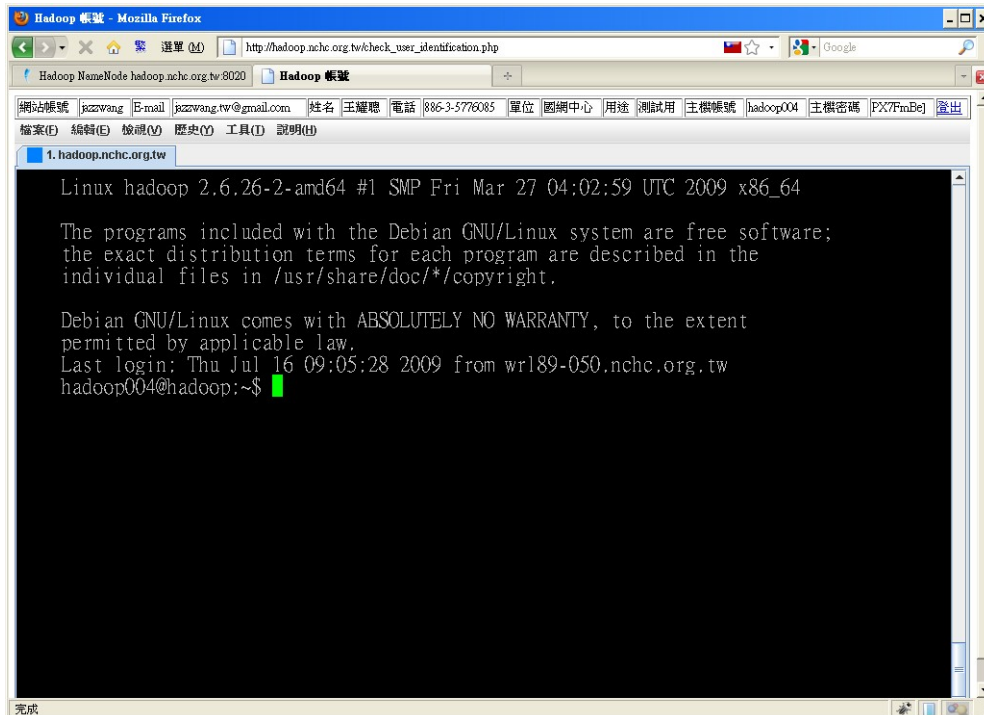


對學界的幫助 (2): 發展用企鵝龍佈署生物叢集的工具

- 持續整理中，待整理套件
- **drbl-biocluster** – 彙整安裝多人共用生物資訊叢集的批次檔

svn co <http://trac.nchc.org.tw/pub/grid/drbl-biocluster>

- 簡化安裝與測試生物資訊叢集常用軟體的程序：**DRBL**、**MPICH2**、**R**、**Rmpi**、**BioConductor**、**Ganglia**、**Nagios**、**AutoFACT**、**BLAST**、**SIM4**、**Clustal**、**PipMaker**、**Phylip**、**Eland**、**Velvet**、**Bowtie**、**SOAP**
- 成果：<http://bio2.classcloud.org>



```
Linux hadoop 2.6.26-2-amd64 #1 SMP Fri Mar 27 04:02:59 UTC 2009 x86_64

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

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Thu Jul 16 09:05:28 2009 from wr189-050.nchc.org.tw
hadoop004@hadoop:~$
```



對學界的幫助 (3): 更多開放教材－生物叢集、GAE...

- 陽明生資所 **97** 年度暑期學分班 格網及平行運算 (實驗課程) <http://trac.nchc.org.tw/course/>
- 陽明生資所 **98** 年度暑期學分班 格網及平行運算 (實驗課程) <http://bio.classcloud.org>
- 雲端運算基礎課程 (一) **Hadoop** 簡介、安裝與範例實作 <http://www.classcloud.org/media/>
- 「**Ruby on Rails** 初學」電子書 by 鄭立竺 <http://nchcrails.blogspot.com>
- Google App Engine** 電子書 by 鄭宗碩 <http://nchc-gae.blogspot.com/>
- More to come**

陽明生資所98年度暑期學分班 格網及平行運算(實驗課程) - Mozilla Firefox

<http://bio.classcloud.org/>

回課程大綱 | [實作一](#) | [實作二](#) | [實作三](#) | [實作四](#) | [實作五](#) | [實作六](#) | [實作七](#) | [實作八](#) | [實作九](#) | [實作十](#) | [實作十一](#) | [實作十二](#) | [作業](#)

陽明生資所98年度暑期學分班 格網及平行運算(實驗課程)

課程資訊

- 上課時間: 2009/7/4(六), 7/5(日), 7/11(六) 9:10~17:30 3 天, 共計 18 個小時
- 上課地點: 台北市北投區立農街二段155號 國立陽明大學 <=> [地圖](#) 圖資大樓 <=> [校園\(P3\)](#) > R401 教室
- 講 師: 王耀聰、鄧偉華
- 報名網頁課程資訊
- 國網中心部份課程網站 - <http://bio.classcloud.org> - 近期修改頁面

課程大綱

2009-07-04 (六)

- 投影片雙張一頁黑白列印版(1)

上午時段	課程內容	主講	投影片	實作步驟
09:10~09:30	課程大綱說明	王耀聰	Part-00	
09:30~10:30	第一次 Linux OS 安裝就上手 - 以 Ubuntu 9.04 安裝為例	鄧偉華	Part-01	
10:30~10:40	休息			
10:40~11:20	基本 Linux 操作 - 基礎指令	鄧偉華	Part-02	實作一
11:20~12:00	基本 Linux 操作 - 編輯器使用	鄧偉華	Part-03	實作二
下午時段	課程內容	主講	投影片	實作步驟
13:30~14:10	進階 Linux 操作(一) - SSH 遠端登入	王耀聰	Part-04	實作三
14:10~15:00	基本 Linux 程式設計 - Bash Shell Script 簡介	王耀聰	Part-05	

完成

Index of /media - Mozilla Firefox

<http://www.classcloud.org/media/>

雲端運算基礎課程 (Hadoop簡介、安裝與範例實作)

投影片	實作步驟	課程錄影(桌面+錄音,HTML+SWF檔案)	課程錄音檔(MP3檔案)
介紹課程		介紹課程	介紹課程
雲端運算簡介		雲端運算的新趨勢	雲端運算的新趨勢
Hadoop 簡介	實作一	Hadoop 簡介	Hadoop 簡介
Hadoop 架構概述		Hadoop 架構概述	Hadoop 架構概述
Hadoop Distributed File System 簡介	實作二	HDFS 簡介	HDFS 簡介
Map Reduce 介紹	實作三	Map Reduce 介紹	Map Reduce 介紹
Map Reduce 程式設計	實作四	Map Reduce 程式設計	Map Reduce 程式設計
進階 hadoop 程式開發(eclipse)	實作五	(1) Eclipse 安裝 (2) MapReduce Plugin 安裝設定 (3) Map Reduce 程式設計實例操作	(1) Eclipse 安裝 (2) MapReduce Plugin 安裝設定 (3) Map Reduce 程式設計實例操作
Hadoop 應用實例: 搜尋引擎 Nutch 簡介	實作六	Nutch 簡介與 NutchEs 展示	Nutch 簡介與 NutchEs 展示
Hadoop 叢集安裝設定解析		Hadoop 叢集設定解析	Hadoop 叢集設定解析
	實作七	實作七: Hadoop 叢集安裝操作	實作七: Hadoop 叢集安裝操作
	實作八	實作八: Hadoop 叢集進階操作	實作八: Hadoop 叢集進階操作
DRBL-Hadoop 快速佈局	實作九	當企鵝遇上小飛象	當企鵝遇上小飛象

Name Last modified Size Description

完成

Attribution-Noncommercial-Share Alike 3.0 Taiwan



姓名標示-非商業性-相同方式分享 3.0 台灣

您可自由：



分享 — 重製、散布及傳輸本著作



重混 — 修改本著作

惟需遵照下列條件：



姓名標示 — 您必須按照著作人或授權人所指定的方式，表彰其姓名（但不得以任何方式暗示其為您或您使用本著作的方式背書）。



非商業性 — 您不得為商業目的而使用本著作。



相同方式分享 — 若您變更、變形或修改本著作，您僅得依本授權條款或與本授權條款類似者來散布該衍生作品。

<http://creativecommons.org/licenses/by-nc-sa/3.0/tw/>

These slides could be distributed by Creative Commons License.



Questions?

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

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



Powered by **DRBL**