Course Information 課程資訊

- 講師介紹 About Me ：
  - 國網中心　王耀聰　副研究員 / 交大電控碩士
  - Yao-Tsung Wang (Jazz) / Associate Researcher, NCHC
  - Master, Electrical and Control Engineering, NCTU, Taiwan
  - jazz@nchc.org.tw
- 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。
  It changes too fast, don't waste your paper !!
- 更多資訊 All Info. will be updated :
  - http://trac.nchc.org.tw/cloud
  - http://www.classcloud.org/media - training course video archives
  - http://www.screentoaster.com/user?username=jazzwang
- 若需要實驗環境 If you need an environment of Hadoop :
  - http://hadoop.nchc.org.tw
- 相關問題討論 If you have questions about Hadoop ：
  - http://forum.hadoop.tw

運用自由軟體打造私有雲端

Build Your Own Private Cloud using Open Source

Jazz Wang

Yao-Tsung Wang

jazz@nchc.org.tw

Powered by DRBL

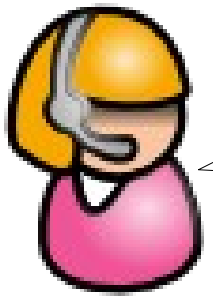
Let's have a QUICK REVIEWS about Cloud

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

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

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

私有雲端
Private Cloud

Community Cloud
社群雲端

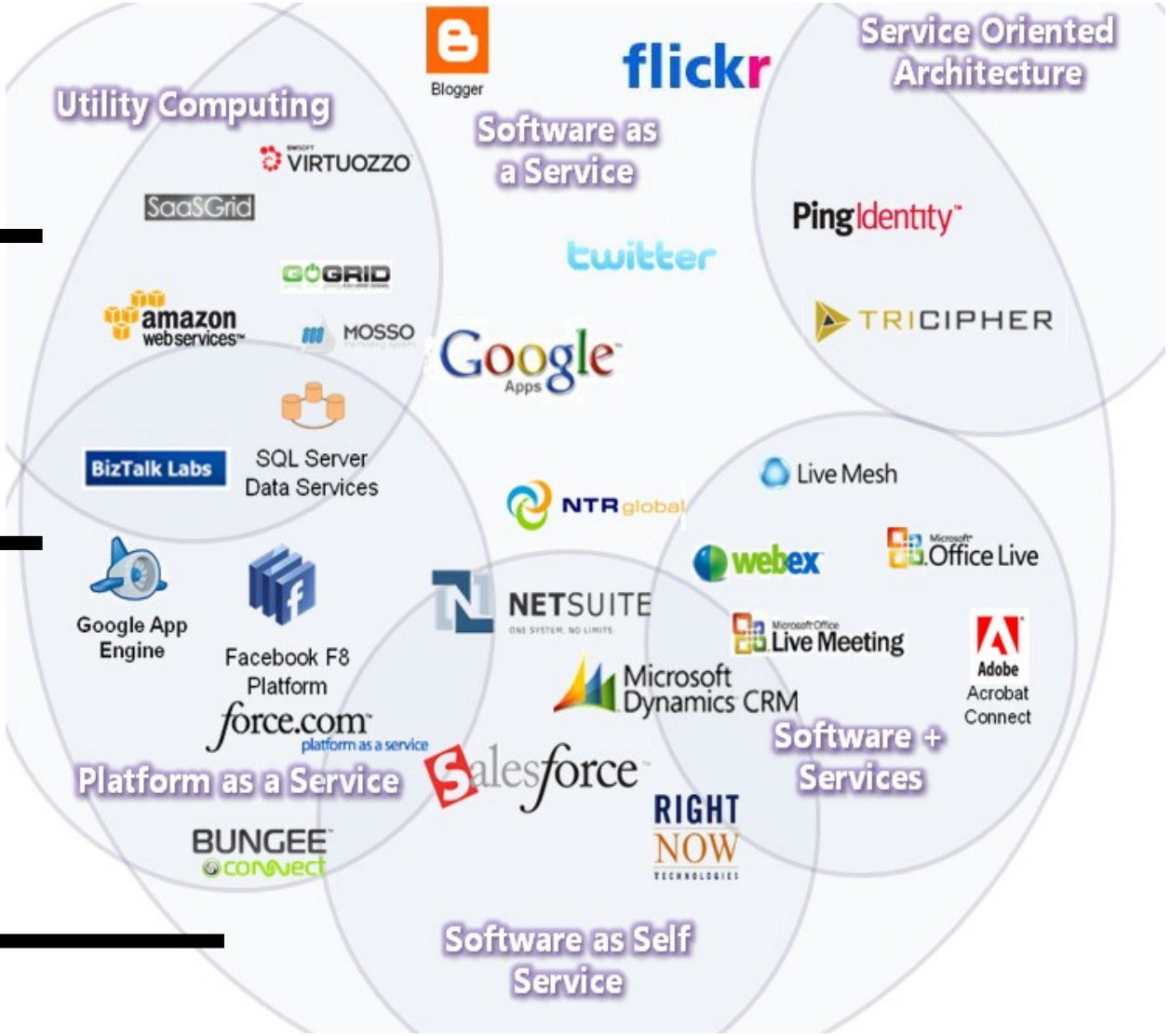
Academia **學術**為主

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」？

端

集中，大廠
Centerized，
Enterprise

symbian
OS

多元，中小廠
Diversify，
SMB

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

Key spirit of Cloud ~

形成服務才是重點！！

Everything as a Service !!

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

Everything as a Service 哈米鬼都是一種服務

- AaaS Architecture as a Service
- BaaS Business as a Service
- CaaS Computing as a Service
- DaaS Data as a Service
- DBaaS Database as a Service
- EaaS Ethernet as a Service
- FaaS Frameworks as a Service
- GaaS Globalization or Governance as a Service
- HaaS Hardware as a Service
- IMaaS Information as a Service
- **IaaS Infrastructure or Integration as a Service**
- IDaaS Identity as a Service
- LaaS Lending as a Service
- MaaS Mashups as a Service
- OaaS Organization or Operations as a Service
- **SaaS Software or Storage as a Service**
- **PaaS Platform as a Service**
- TaaS Technology or Testing as a Service
- VaaS Voice as a Service

Customer-Oriented

客戶導向，服務至上

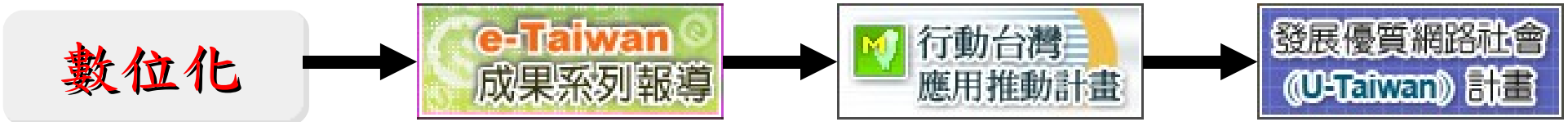
能把 AAA 做好就很強了

Authentication
Authorization
Accounting
as
a
Service

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

Evolution of Cloud Services

雲端服務只是軟體演化史的必然趨勢



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

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/

1960 PDP-1

.

.

.

1965 PDP-7

.

.

.

1969 1st Unix

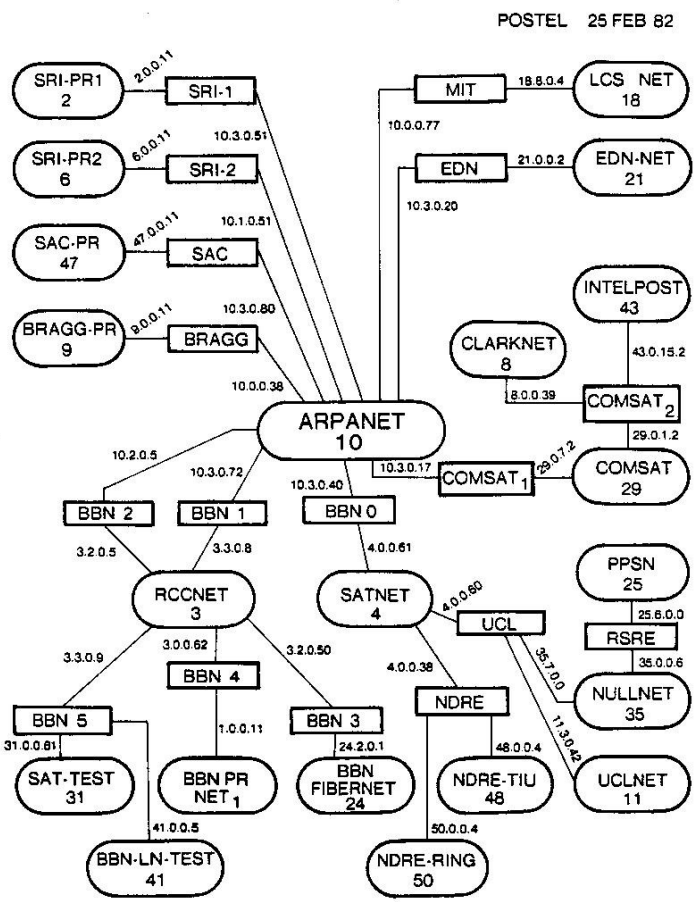
Mainframe
Super Computer

1977 Apple II

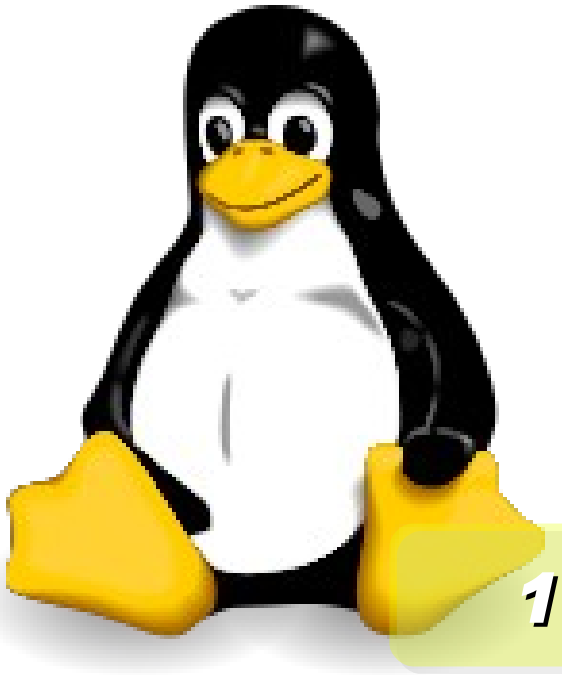
1981 IBM 1st PC 5150

Back to Year 1970s ...

1982 TCP/IP



1983 GNU



1991 Linux

Back to Year 1980s ...

Brief History of Computing (2/5)

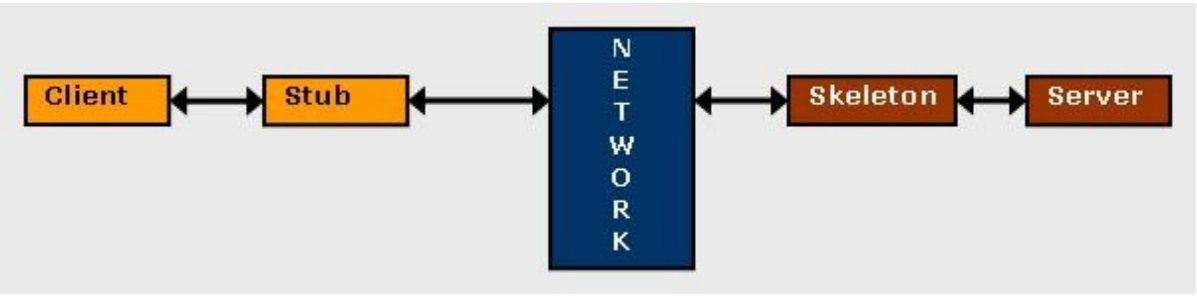
Source: http://www.nchc.org.tw

Mainframe
Super Computer

PC / Linux
Cluster Parallel

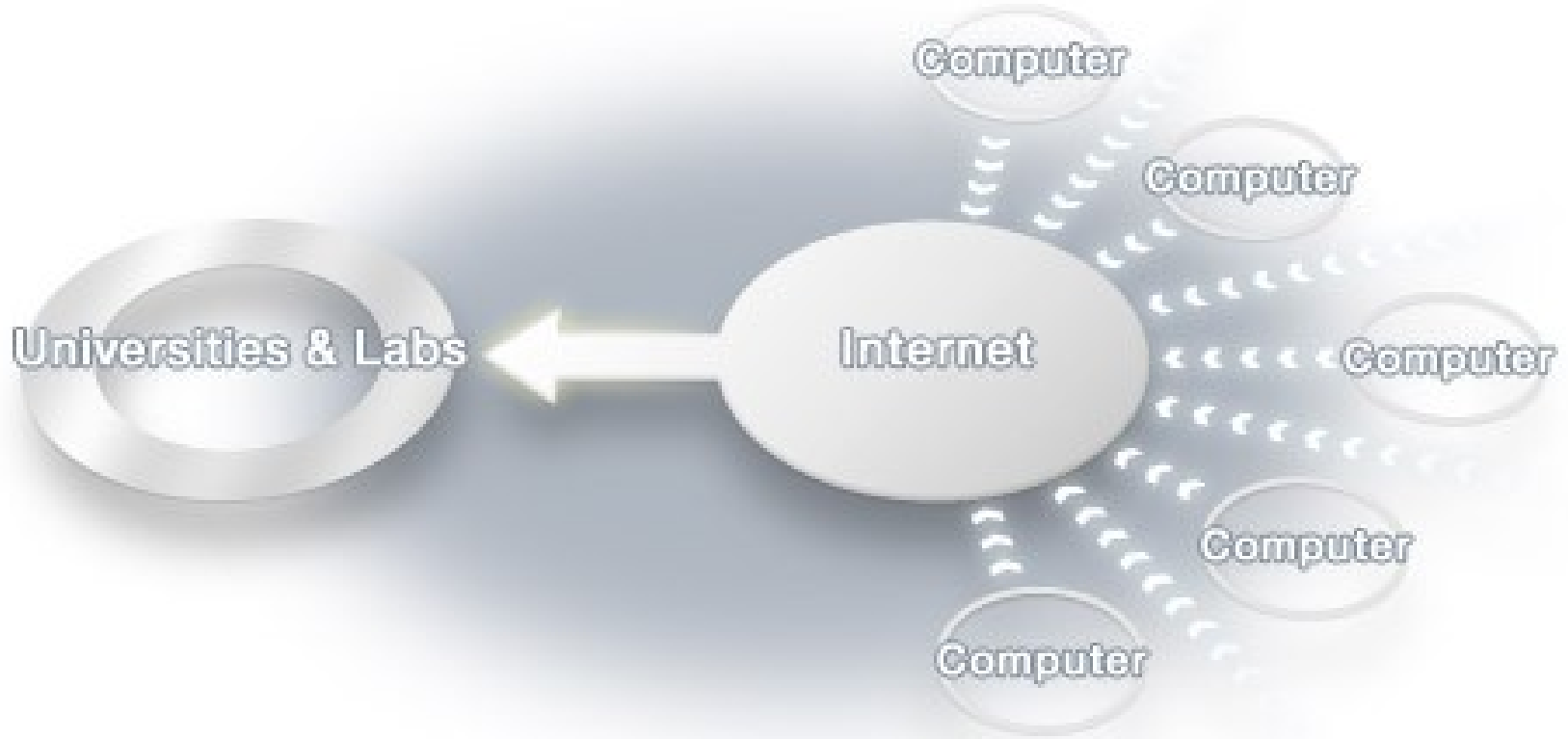
1990 World Wide Web
by CERN
…
…
1993 Web Browser
Mosaic by NCSA

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

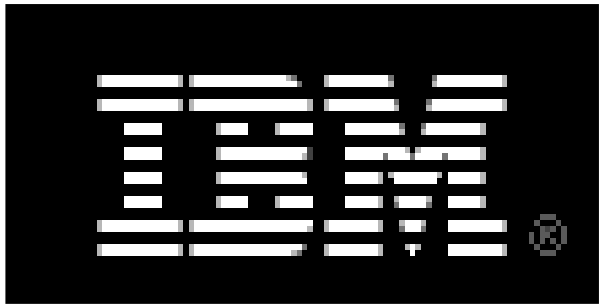
Mainframe
Super Computer

PC / Linux
Cluster Parallel

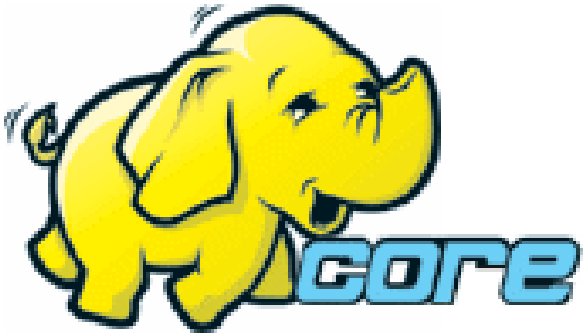
Internet
Distributed Computing

Virtual Org.
Grid Computing

2001 Autonomic Computing
IBM



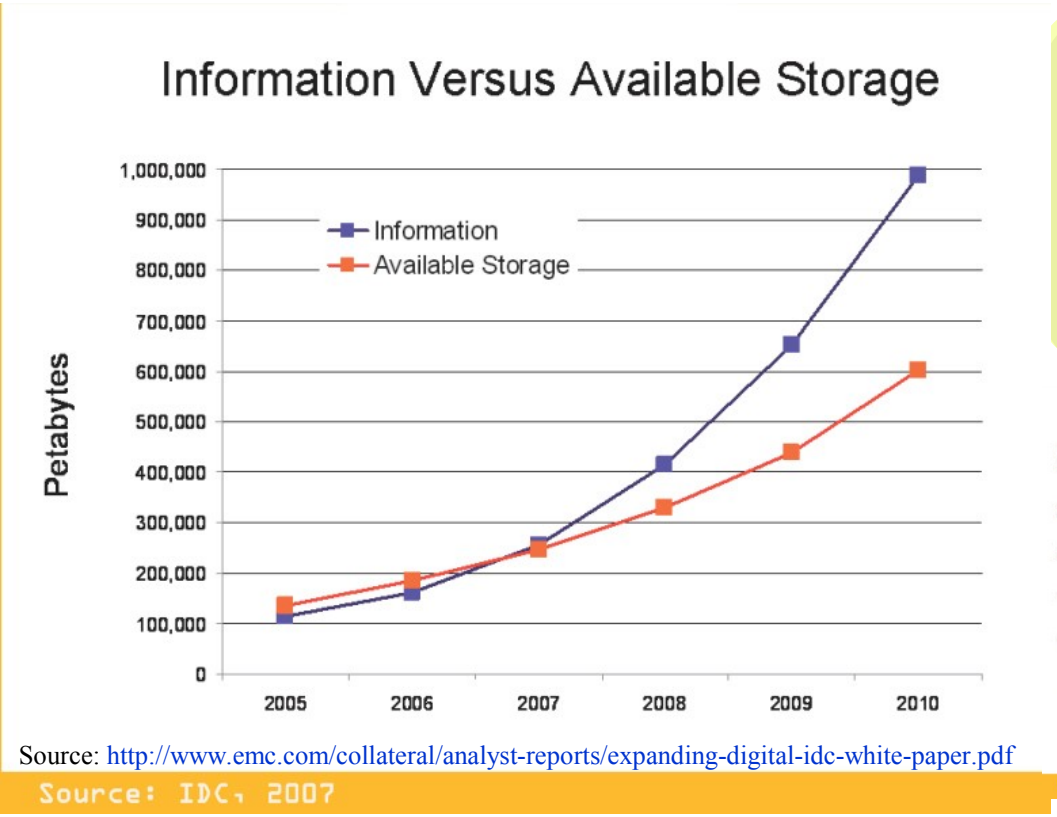
2006 Apache Hadoop



2005 Utility Computing
Amazon EC2 / S3

2007 Cloud Computing
Google + IBM

Back to Year 2007 ...

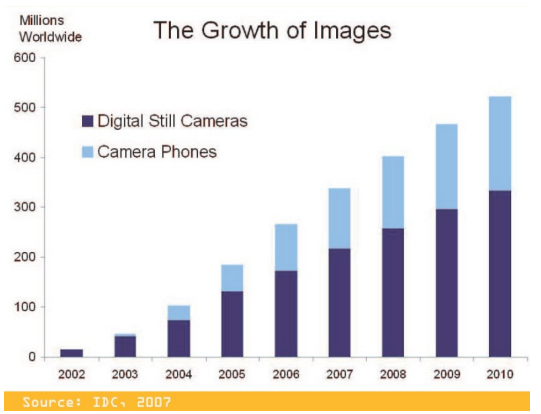


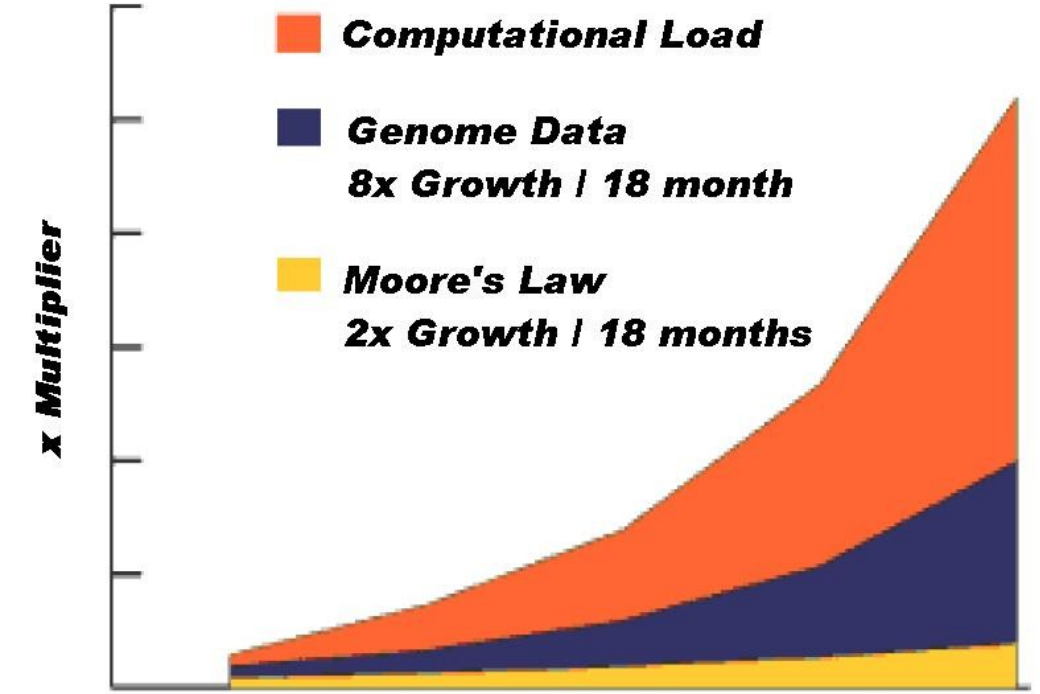
Source: http://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf

Source: IDC, 2007

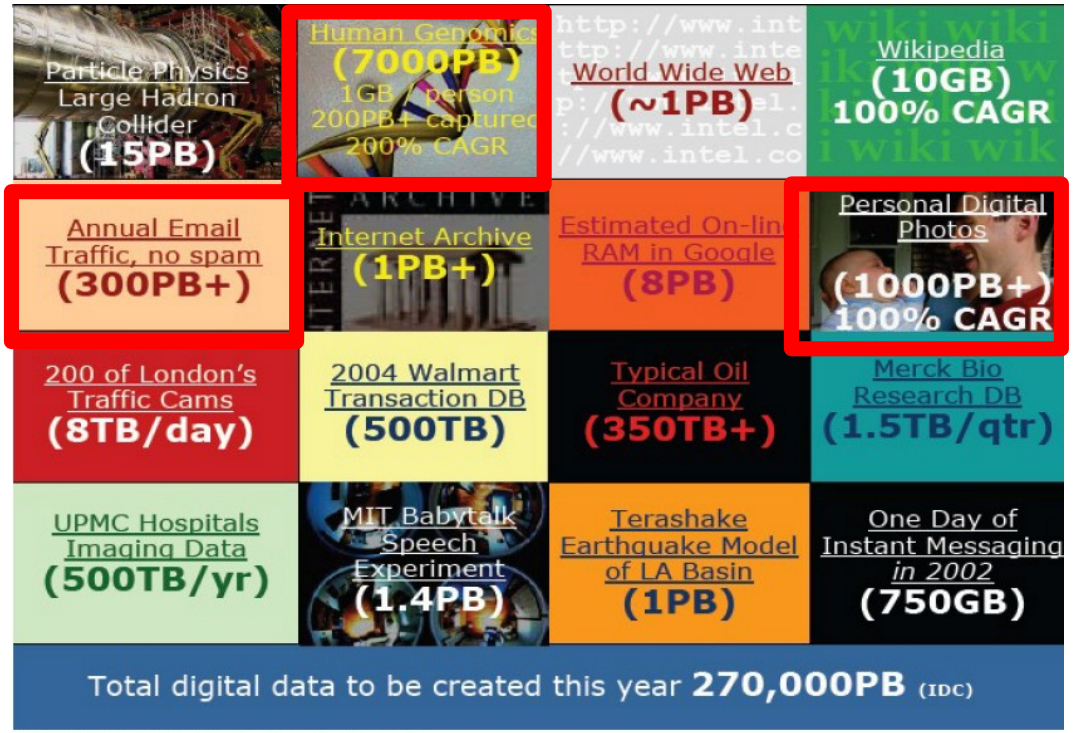
2007 Data Explore

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



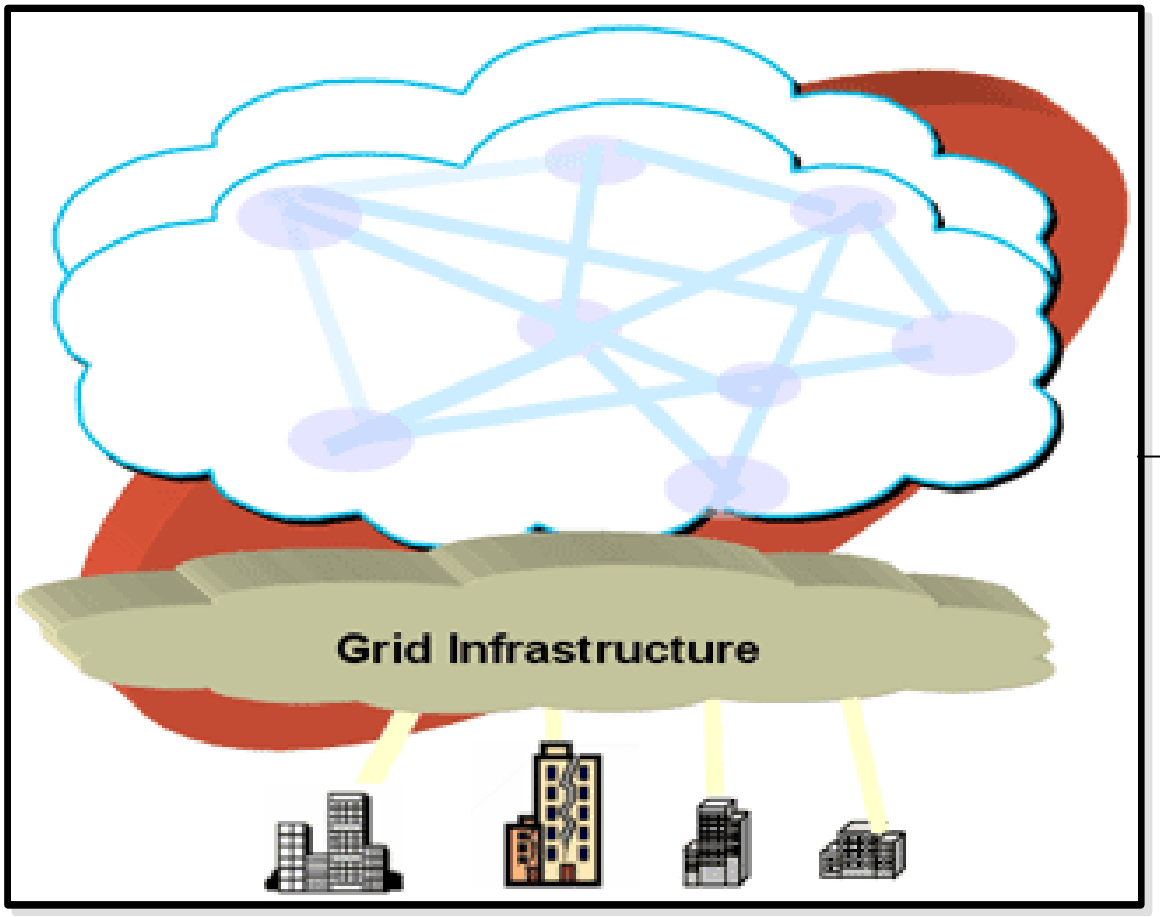


Source: http://lib.stanford.edu/files/see_pasig_dic.pdf



Phillip B. Gibbons, Data-Intensive Computing Symposium

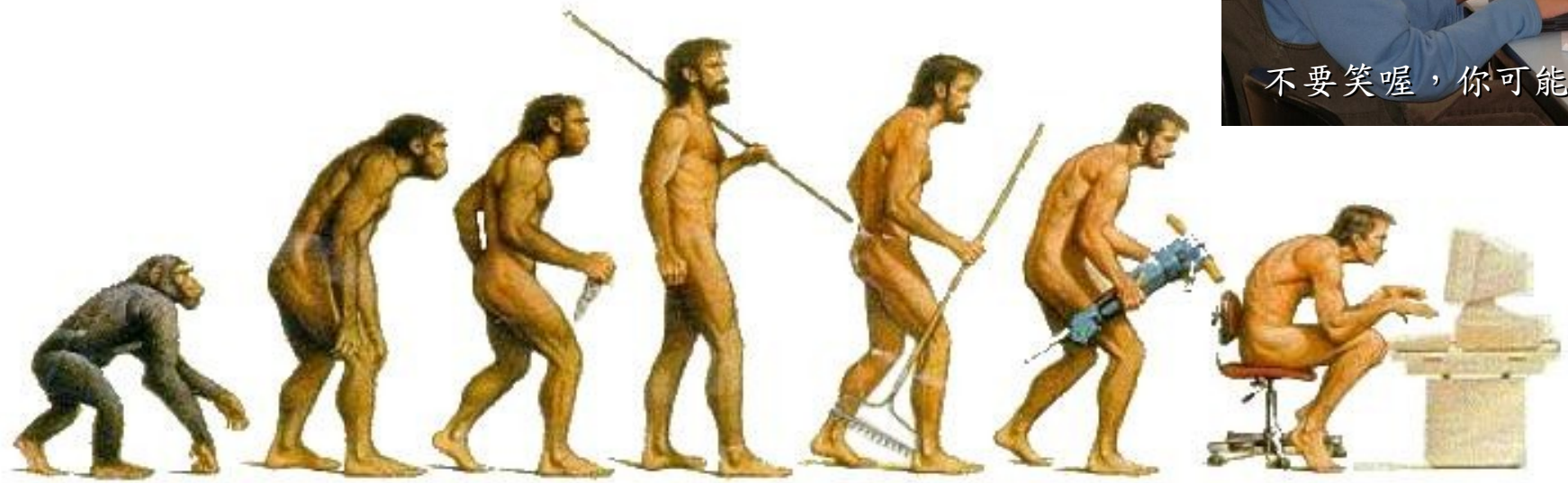
Brief History of Computing (5|5)



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

nframe per puter	**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 ?!

Source: http://cyberpingui.free.fr/humour/evolution-white.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！

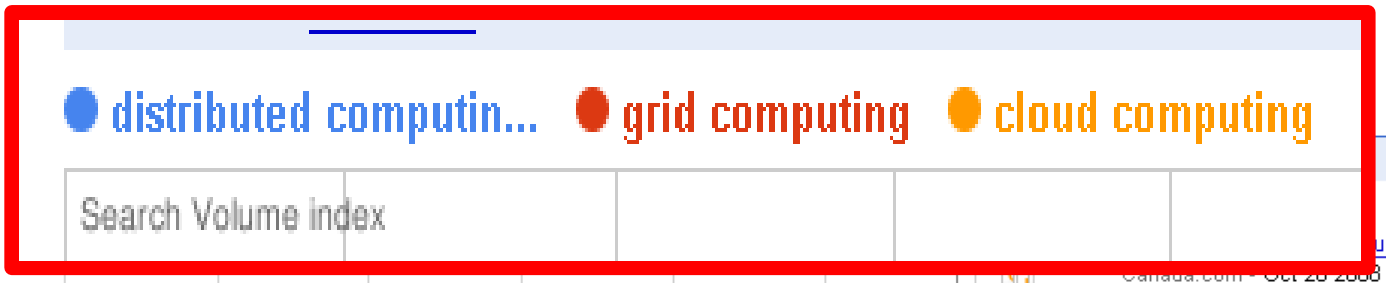
Answer #3: Total Cost of Ownership 總擁有成本

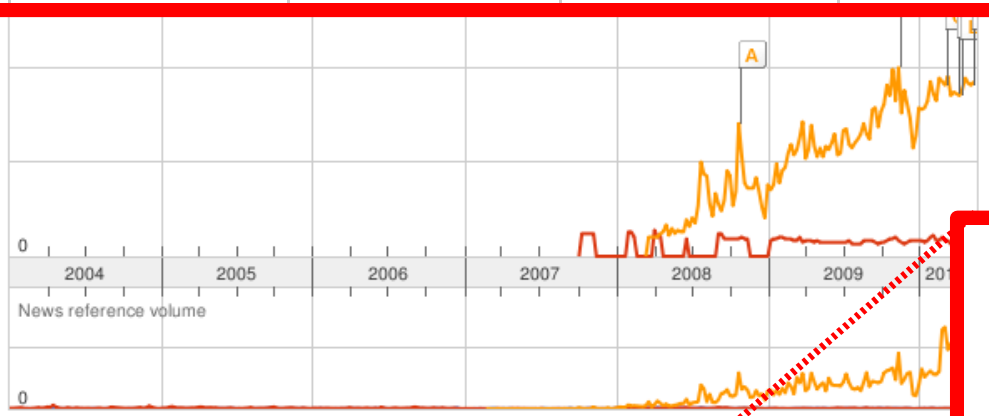
This is why Cloud Computing matters ?!

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？





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

Sign in to see and export additional Tren

All regions | All years

ud" 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

似乎亞洲國家特別熱愛雲端?! **Too Hot in Asia ?!**

CIO 2010 : Virtualization, Cloud and Web 2.0

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 "*

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！！

Source: http://media.photobucket.com/image/falling%20ground/preeto_f10/falling

該使用別人打造的雲端，還是自己打造專屬雲端呢？

Let's SKIP Public Cloud
公用雲端服務，講過了就跳過啦！！

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

私有雲端
Private Cloud

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

amazon web services™
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

IBM

私有雲端
Private Cloud

Community Cloud
社群雲端

Academia 學術為主

Reference Cloud Architecture
雲端運算的參考架構

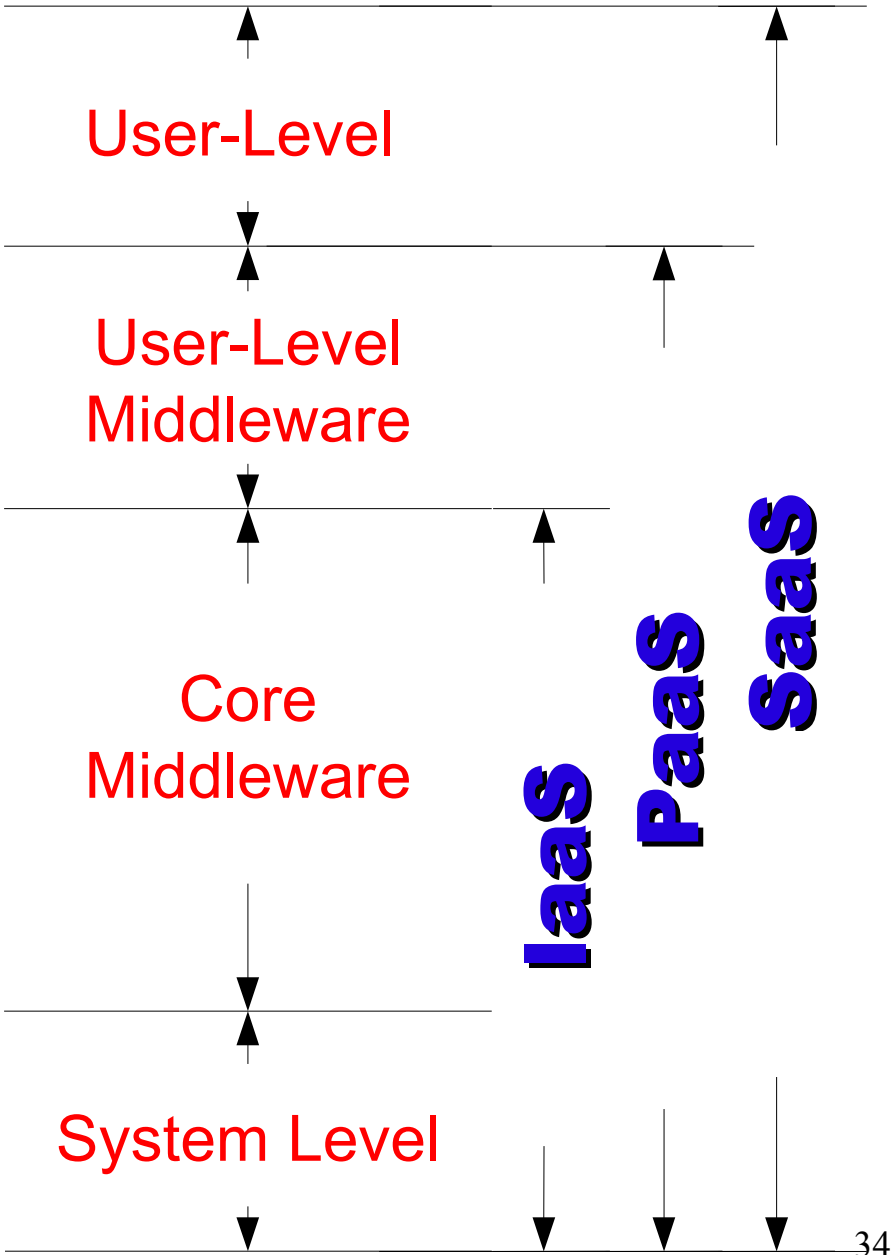
應用軟體 Application
Social Computing, Enterprise, ISV,…

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

控制管理 Control
Qos 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 for Private Cloud

建構私有雲端運算架構的自由軟體

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, Ddmission Control, Pricing, SLA Management, Metering…

OpenNebula, Enomaly, Eucalyptus , OpenQRM, ...

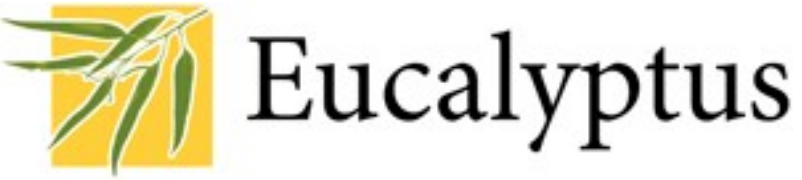
虛擬化 Virtualization
VM, VM management and Deployment

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

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, ...

Open Cloud #1: Eucalyptus



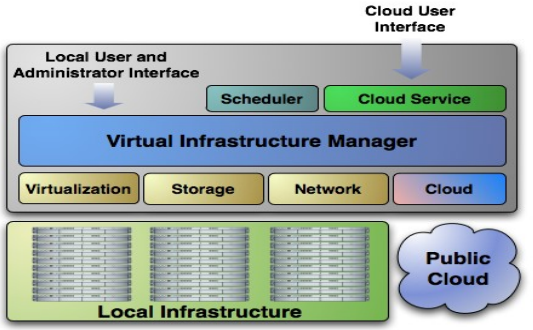
- 原是加州大學聖塔芭芭拉分校 (UCSB) 的研究專案
- **It was a research project of UCSB, USA**
- 目前已轉由 Eucalyptus System 這間公司負責維護
- **Now Eucalyptus System provide technical supports.**
- 創立目的是讓使用者可以**打造自己的 EC2**
- **It designed to help user to build their own Amazon EC2**
- 特色是相容於 Amazon EC2 既有的用戶端介面
- **Its feature is 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**

關於 Eucalyptus 的更多資訊，請參考
http://trac.nchc.org.tw/grid/wiki/Eucalyptus

Open Cloud #2: OpenNebula

- 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

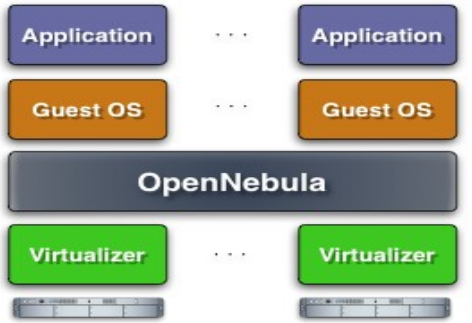
OpenNebula.org



SEVENTH FRAMEWORK PROGRAMME

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

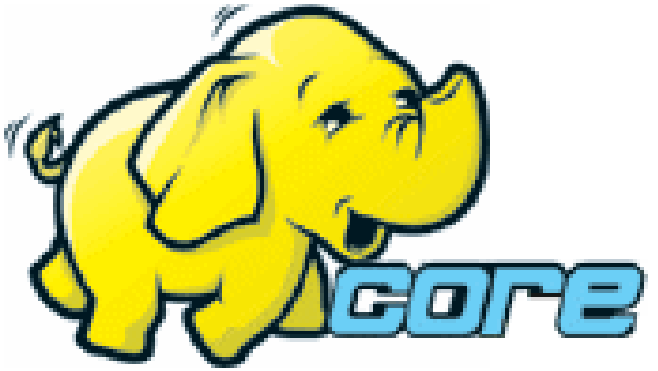
RESERVOIR



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, ...

Open Cloud #3: 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** are also powered by Hadoop

Open Cloud #4: 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**

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 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, Sector/Sphere, Eucalyptus, and more ....

Questions?

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

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

Attribution-Noncommercial-Share Alike 3.0 Taiwan

creative commons

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爲您或您使用本著作的方式背書）。

非商業性 — 您不得爲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.