

Better HPC Better Living
為更好的生活努力

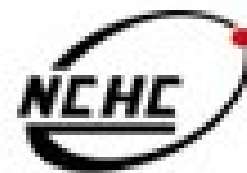
自由軟體在雲端科技發展趨勢中的角色與貢獻

Roles and contributions of Open Source in Cloud Computing Era

王耀聰 副研究員

jazz@nchc.org.tw

財團法人國家實驗研究院
國家高速網路與計算中心



國家實驗研究院
National Applied Research Laboratories

國家高速網路與計算中心
National Center for High-Performance Computing

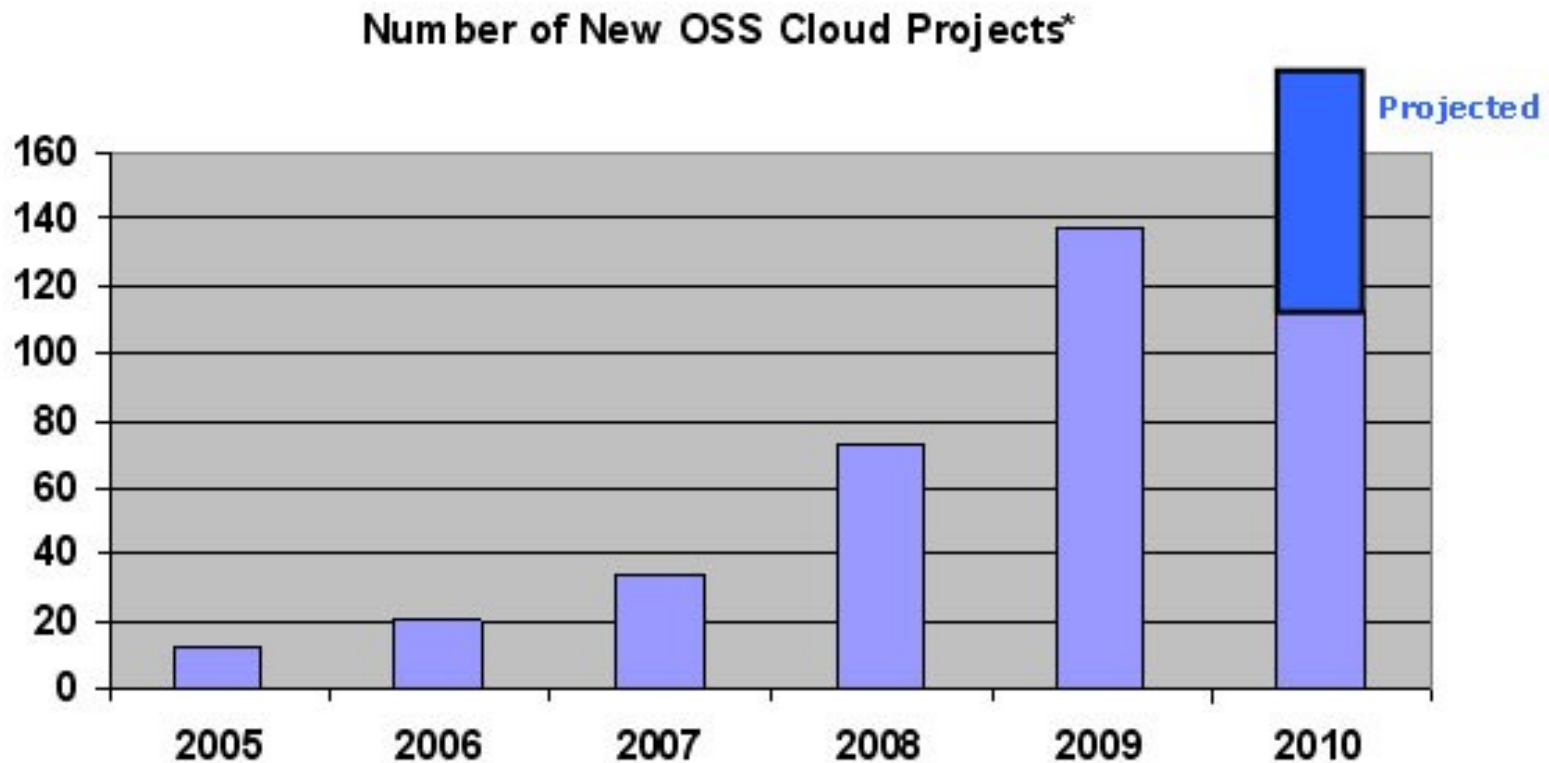
Rapid Growth of Open Source Projects for Cloud

雲端相關的自由軟體專案於過去三年急速成長

Open Source Projects for Cloud on Rise, According to Black Duck Software Analysis

根據 Black Duck 的分析，近年來雲端相關的自由軟體專案有大幅度的成長

Open Source Projects for Cloud on Rise



* Projects specifically referencing cloud. Many OSS projects are suitable for cloud but don't specify it.

圖片來源：

<http://www.blackducksoftware.com/oss/projects/> 2

Three Roles and Contribution of Open Source for Cloud

自由軟體在雲端科技發展中的三大角色與貢獻

Role #1: The foundation of Cloud Services !

角色一：自由軟體是雲端服務的幕後技術基石

Contribution #1: Reduce IT Cost 貢獻一：節省軟體成本

Role #2: Interoperability Standard between Clouds !

角色二：跨越不同雲端服務的互通標準！

Contribution #2: Less Vendor Lock-In 貢獻二：減少技術壟斷

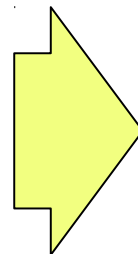
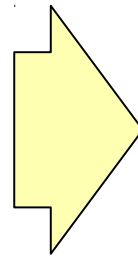
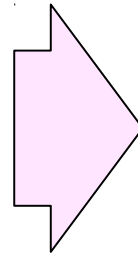
Role #3: Open Platform for Open Education !

角色三：雲端人才培育需要開放平台！

Contribution #3: Open Ladder to Cloud 貢獻三：搭建登雲天梯

Role #1: The foundation of Cloud Services !

角色一：自由軟體是雲端服務的幕後技術基石



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, Mashups, Workflows, ...

控制管理 Control

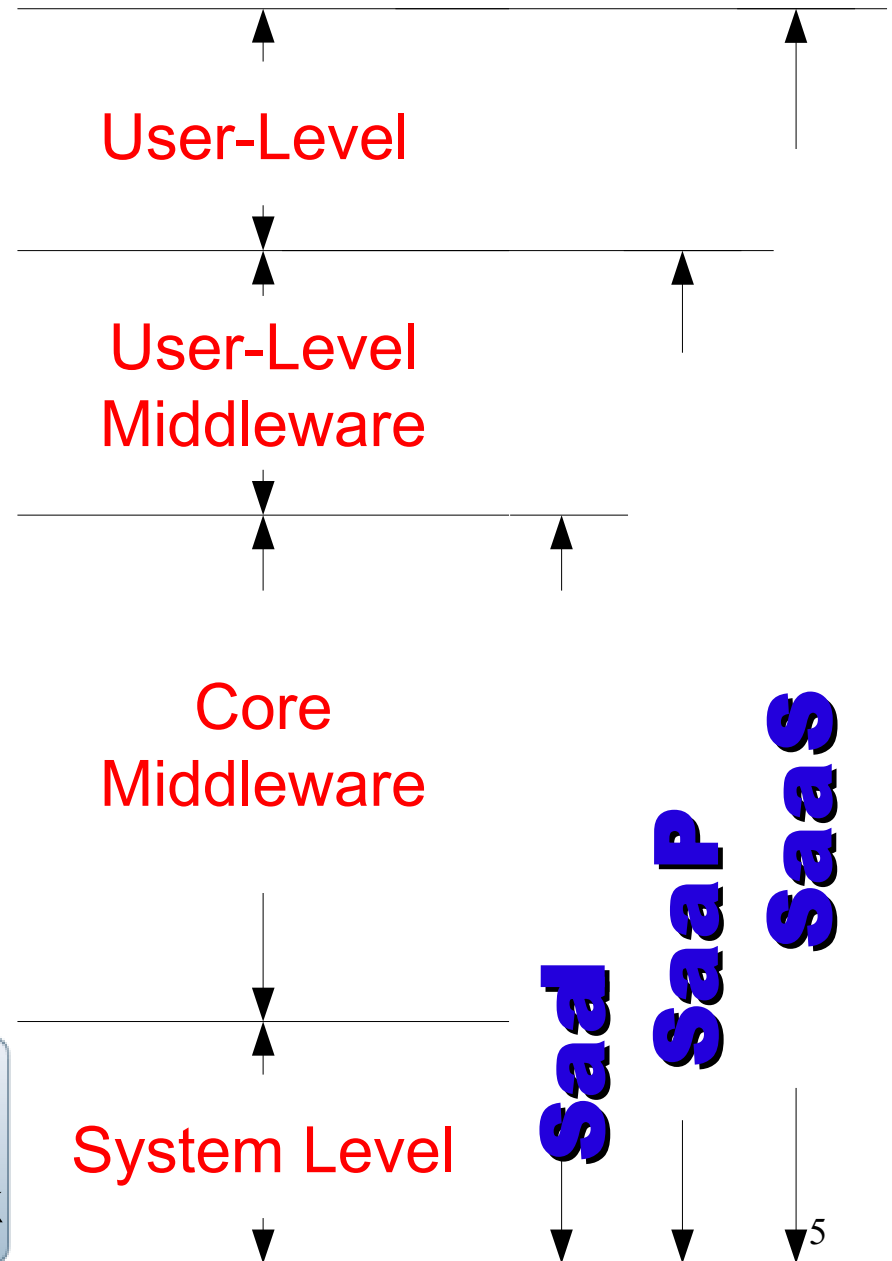
Qos Negotiation, Admission Control, Pricing, SLA Management, Metering...

虛擬化 Virtualization

VM, VM management and Deployment

硬體設施 Hardware

Infrastructure: Computer, Storage, Network



Building IaaS with Open Source

用自由軟體打造 IaaS 服務 (1) 虛擬化

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

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

程式語言 Programming
Web 2.0, 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

What is Virtualization ??

虛擬化技術有哪些呢??

Application Virtualization 應用程式虛擬化

Desktop Virtualization
Client Virtualization 桌面虛擬化

Presentation Virtualization 顯示虛擬化

OS-level Virtualization 作業系統虛擬化

Network Virtualization 網路虛擬化

Storage Virtualization 儲存虛擬化

資料庫虛擬化

資料虛擬化

Virtualization

Data

Data Virtualization

大家對虛擬化的
狹義認知

Open Source for Virtualization

虛擬化技術對應的自由軟體

Application Virtualization

Ex. VMWare ThinApp

Desktop
Client

建置虛擬化叢集不可只
注重作業系統虛擬化，
須同時具備底層支援的
網路與儲存虛擬化！！

Redhat SPICE

Presentation

VNC, FreeNX

OS-level Virtualization

Xen, KVM, OpenVZ

Network Virtualization

OpenFlow vSwitch

Storage Virtualization

Lessfs, SDFS

Building IaaS with Open Source

用自由軟體打造 IaaS 服務 (2) 虛擬叢集控制

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

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

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

Hadoop (MapReduce),
Sector/Sphere, AppScale

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

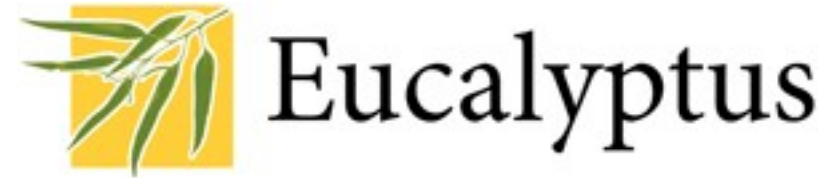
**OpenNebula, Enomaly,
Eucalyptus, OpenQRM, ...**

虛擬化 Virtualization
VM, VM management and Deployment

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

硬體設施 Hardware
Infrastructure: Computer, Storage,
Network

Eucalyptus



- <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

關於 Eucalyptus 的更多資訊，請參考

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

OpenNebula

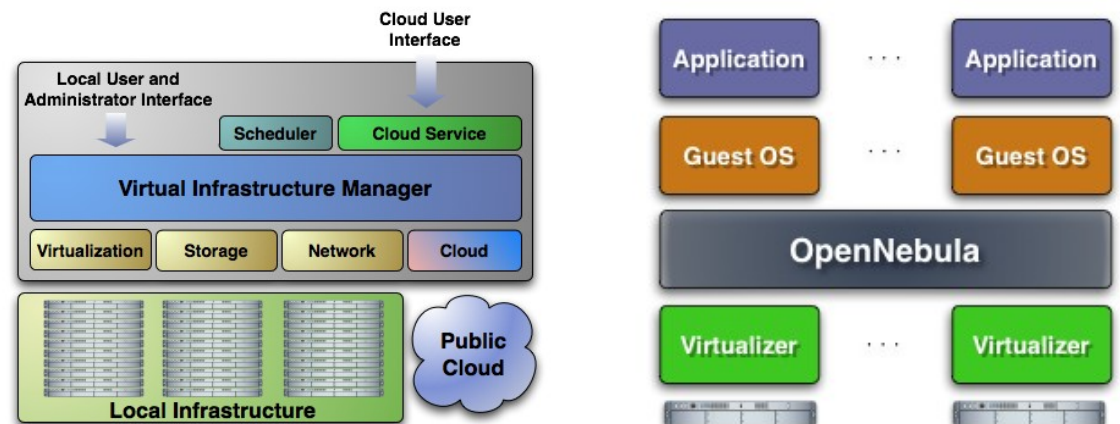
- <http://www.opennebula.org>
- 由歐洲研究學會 (European Union FP7) 贊助
- 將實體叢集轉換成具管理彈性的虛擬基礎設備
- 可管理**虛擬叢集**的狀態、排程、遷徙 (migration)
- 優勢是Ubuntu 9.04 已經收錄 OpenNebula 的套件
- 缺點：需下指令來進行虛擬機器的遷徙 (migration) 。

OpenNebula.org



關於 OpenNebula 的更多資訊，請參考

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



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, 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, ...

硬體設施 Hardware
Infrastructure: Computer, Storage,
Network

Hadoop

- 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。



Sector / Sphere

- <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[Ma1Stone效能評比軟體](#)。



National Center for Data Mining
University of Illinois at Chicago



Open Data Group

<http://www.opendatagroup.com/>

Building SaaS with Open Source

用自由軟體打造 SaaS 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, 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

快速佈署內網搜尋引擎的軟體工具 Crawlzilla

- 雲端運算在大型資料分析、資料探勘方面的示範應用
- 可協助企業或校園建置內網的網站、檔案搜尋
- 支援叢集式安裝、中文分詞，並提供網頁管理介面
- 更多資訊詳見專案網站：<http://code.google.com/p/crawlzilla>

Web UI (Crawlzilla Website + Search Engine)

JSP + Servlet +
JavaBean

Nutch

Lucene

Crawlzilla System Management

Tomcat

Hadoop

PC1 (Master)

PC2

PC3

PC4

CrawlZilla 網頁管理介面



HOME

Crawl

資料庫管理

系統狀態

使用者設定

登出系統

管理介面功能介紹

* 抓取網頁設定：透過網頁介面建立搜尋引擎索引

* 資料庫管理：查詢、刪除已爬取的索引資料庫

* 系統狀態：瀏覽目前系統狀態

* 管理員設定：包含個人化與語言設定

* 修改密碼

搜尋引擎快速連結

CrawlZilla 搜尋引擎範例

3@books

3@businessweekly

3@yahoo

系統功能

修改管理員密碼

相關資源

CrawlZilla@GoogleCode



Role #2: Interoperability Standard between Clouds !

角色二：跨越不同雲端服務的互通標準！

IBM、思科（Cisco Systems）等 20 多家科技公司於 2009 年 3 月 30 日發佈「開放雲端宣言」（Open Cloud Manifesto），希望為雲端運算建立標準，確保業者間的相容性，讓客戶更容易轉換平台。

「客戶使用雲端運算的經驗愈多，就愈明白相容性有多重要。他們不想被綁死在單一業者。」
— IBM 開放標準部門副總裁諾斯沃西（Karla Norsworthy）



圖片來源：<http://opencloudmanifesto.org/>

2009年四月28日，Open Grid Forum宣佈成立Open Cloud Computing Interface (OCCI)工作小組，以協助制定雲端服務之間的互通國際標準。



ABOUT OGF RESOURCE CENTER EVENTS DOCUMENTS AREAS/GROUPS
MEMBERS **NEWS** STANDARDS CONTACT US SITE MAP GRIDFORGE

OGF ANNOUNCES OPEN CLOUD COMPUTING INTERFACE (OCCI) WORKING GROUP

OGF Officially Launches Working Group To Create An API For Cloud Infrastructure Services

CHICAGO, 28 April 2009 The Open Grid Forum (OGF) has officially launched the Open Cloud Computing Interface Working Group (OCCI-WG). Its aim is the rapid development of a clean, open API for cloud infrastructure delivered on-demand. Infrastructure as a Service (IaaS) is one of three primary segments of the emerging cloud industry in which compute, storage and network resources are provided as a service. The API will be based on existing service provider functionality, and will cover management of the entire life-cycle of virtual machines and other similar workloads (including virtual private servers and physical dedicated servers). It will as much as possible complement and build on existing work such as the Open Virtualisation Format (OVF), as well as serving as an integration point for others standards efforts including DMTF and SNIA.

Craig Lee, OGF's President, stated: "OCCI is an extremely important development that will define a much needed, open standard API in the cloud space. This is, however, just the first step of many. We are committed to working with all stakeholders and organizations to move the ball forward. As always, OGF strives to facilitate productive technical environments for work across the spectrum of distributed computing technologies."

Existing specifications for cloud infrastructure services (including the Amazon EC2 API and Sun Cloud APIs) are provided by single vendors and OCCI will be the first multi-vendor initiative to deliver a royalty-free, open standard API. This will improve cloud computing interoperability and increase competition in the space, facilitating the emergence of a marketplace of services, which will in turn increase customer choice and decrease the risk of vendor lock-in. OCCI will also be important enabler for the creation of hybrid cloud architectures that bridge multiple data centers and cloud services, through adapters and tooling that support the open API. This and related open standards will lower the cost of migration to and from public clouds, delivering automated management of peak capacity, minimizing outages and reducing costs.

OCCI will rely heavily on existing Internet standards and work already done by cloud providers, so as it is easy to both implement and consume. This will benefit software vendors and service consumers alike. For example software vendors who wish to deliver their applications as a service

圖片來源：<http://www.ogf.org/News/news.php?id=132>

數位時代 Beta 3.0 - Topics and Links新聞精選

全球廿五間科技公司將攜手合作，計畫促進更健全、更開放的雲端科技。

這項稱作OpenStack的合作計畫在今天公布，以IT代管服務公司Rackspace和美國太空總署NASA為首，將以Open Source（開程式碼）的方式，把雲端平台的技術分享到世界每一個角落。

有了OpenStack以後，任何一個想要提供雲端服務的單位，只要準備好伺服器 and 網路，就可以下載OpenStack的開放軟體，架設自己的雲端平台。目前這個計畫的網站提供了兩套軟體：OpenStack Compute用來設定和管理大量的伺服器，而OpenStack Storage則提供一套雲端儲存系統，只要使用多台一般市面上出售的伺服器，就可以儲存上千兆個位元組的資料。

Rackspace和NASA在這項計畫中，都有卓著的貢獻。Rackspace開放了自家公司產品「雲端檔案」和「雲端伺服器」的程式碼，這兩項產品是該公司雲端設備的基礎。NASA也將貢獻出他們用以架設「Nebula雲端平台」的技術。

NASA Nebula是目前全世界最強大的雲端平台之一。上週Microsoft Research所公布的最新火星影像，就是在NASA的Nebula平台上，可用Microsoft Research的影像，目前也存放在Nebula

藉由開程式碼的方式，增加不同雲端科技產品之間的互通性。

以Nebula平台的

參與這項計畫的公司還有Intel、Dell、Citrix等，而其他知名公司如Microsoft也在陸續洽談中。這個計畫的目標，是希望為雲端科技提供一個良好的開發環境，同時藉由開程式碼的方式，增加不同雲端科技產品之間的互通性。這樣一來，雲端科技一旦有了新的突破，不同的產品都能夠同時受惠。

Role #3: Open Platform for Open Education !

角色三：雲端人才培育需要開放平台！

<http://www.captivating-content.com/Hidden%20Treasure.jpg>

第一波雲端運算
的成長主要帶動了
台灣硬體產業的
經濟復甦 ...

建再多的資料中心
也是要有 **人** 管

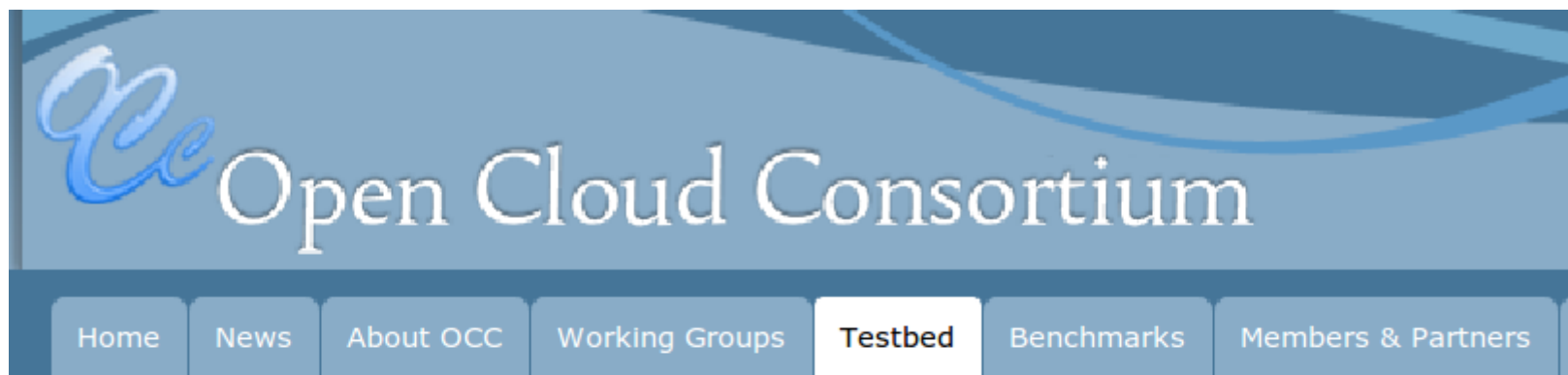


第二波雲端運算
的成長將決定於
人才培育速度能否
快速銜接產業需求

<http://www.sysadminaday.com/images/people/136-3697.JPG>

美國結合產學資源，成立 Open Cloud Consortium，
並建立 Open Cloud Testbed，作為人才培育與研究的開放平台。

會員包括：Aerospace、Booz Allen Hamilton、Cisco、Infoblox、MIT Lincoln Labs、NASA Ames Research Center、Northwestern University、Open Data Group、Raytheon、Sector Project、SIOS、University of Illinois at Chicago、Yahoo、Calit2、Johns Hopkins University、National Lambda Rail、University of Chicago



Testbed

The Open Cloud Consortium (OCC) operates a wide area testbed for cloud computing called the Open Cloud Testbed. The testbed is being rolled out in phases. We are currently in Phase 2 of the roll out.

As far as we know, the Open Cloud Testbed is the only wide area cloud that utilizes wide area 10 Gb/s networks. If you know of any other such clouds, please let us know.

國網中心 2009 年開放實驗叢集間接促成研究成果

- 國網中心於 **2009** 年四月啓用 **Hadoop** 實驗叢集，目前共 **21** 台 **PC** 服役中。
- 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 **2010** 年八月，雲端實驗叢集共註冊 **937** 人，服務 **37** 個學術單位、**5** 個研究單位、**19** 間業界公司、**2** 所醫院。累計於 **14** 個月執行 **16,424** 個 **Job**。

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-

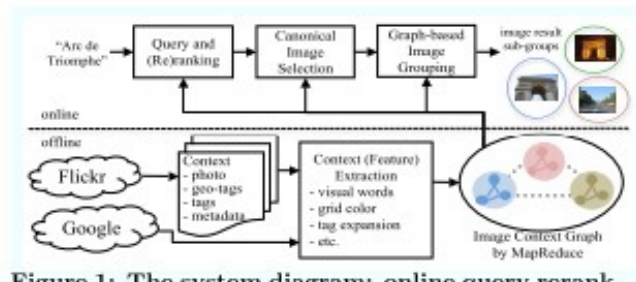
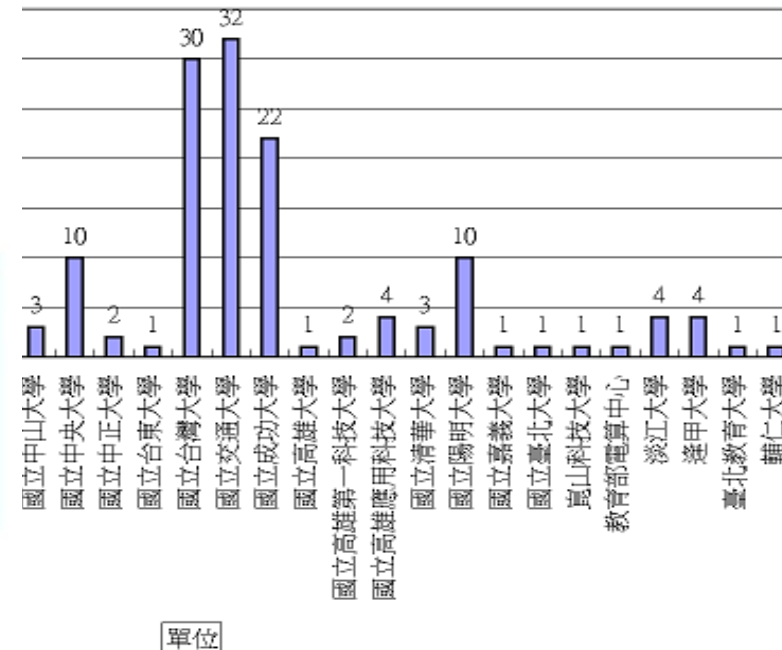


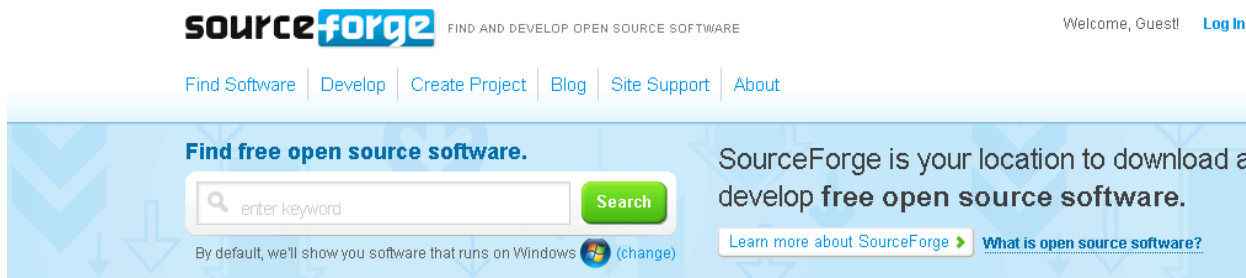
Figure 1: The system diagram: online query reranking, canonical image selection, and image result grouping are based on the image context graph, constructed offline by Hadoop MapReduce over rich context features in Flickr photos.



Conclusion: Open Source as a Strategy !!

未來自由軟體將成爲吸引注意力的一種策略

當免費、開放才能吸引使用者或開發者的注意力～
自由軟體就會變成一種行銷策略，將不得不開放！！



The screenshot shows the SourceForge homepage. At the top left is the SourceForge logo with the tagline "FIND AND DEVELOP OPEN SOURCE SOFTWARE". To the right, it says "Welcome, Guest!" and "Log In". Below the logo are navigation links: "Find Software", "Develop", "Create Project", "Blog", "Site Support", and "About". A main banner area contains the text "Find free open source software." and "SourceForge is your location to download and develop free open source software." There is a search bar with the placeholder "enter keyword" and a "Search" button. Below the search bar, it says "By default, we'll show you software that runs on Windows" with a Windows logo and a "(change)" link. To the right of the search bar are two links: "Learn more about SourceForge" and "What is open source software?".

Browse by category

Clustering
Communications

What's Hot for Windows?



1

XBMC Media Center

XBMC Media Center is a free cross-platform media ...

95% recommend

Project Hosting on Google code

Explore hosted projects:

Python	JavaScript	Django	Web	CMS	Google
Java	AJAX	Rails	Plugin	Student	Android
CPlusPlus	MySQL	.Net	Game	Windows	AppEngine

