



運用自由軟體打造雲端運算實驗環境

Building Experimental Cloud Infrastructure with Open Source

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



Powered by DRBL

What is Cloud Computing?

何謂雲端運算？請用一句話說明！

***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

More definition?

其他定義請參考：

NIST Notional

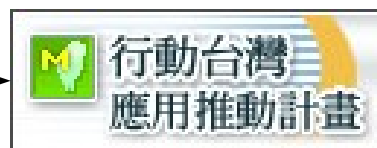
Definition of Cloud

Computing

Evolution of Cloud Services

雲端服務只是軟體演化史的必然趨勢

數位化



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！！

Types of Cloud Computing

雲端運算的三種型態

  
Public Cloud
公用雲端

Target Market
is **S.M.B.**
主要客戶為
中小企業

Hybrid Cloud

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

以**大型企業**
為主要客戶
Enterprise is
key market



私有雲端
Private Cloud

Types of Cloud Service Provider

雲端服務的市場區隔

SaaS

Software as a Service

軟體即服務

PaaS

Platform as a Service

平台即服務

IaaS

Infrastructure as a Service

架構即服務



Everything as a Service 啥米鬼都是一種服務

- 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

引用自：

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

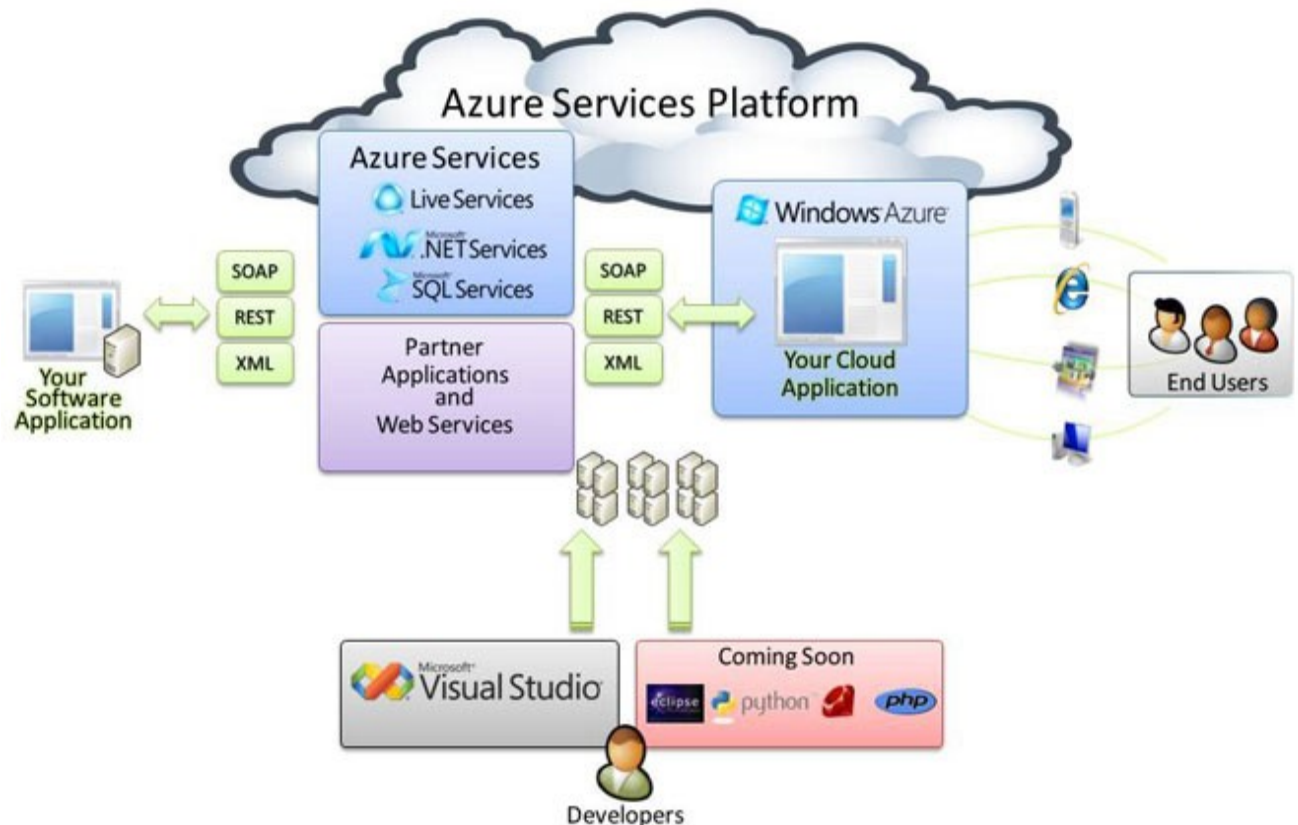


- Amazon Web Service (AWS)
- 虛擬伺服器：**Amazon EC2**
 - Small (Default) \$0.10 per hour \$0.125 per hour
 - All Data Transfer \$0.10 per GB
- 儲存服務：**Amazon S3**
 - \$0.150 per GB – first 50 TB / month of storage used
 - \$0.100 per GB – all data transfer in
 - \$0.01 per 1,000 PUT, COPY, POST, or LIST requests
- 觀念：**Paying for What You Use**

- Google App Engine (GAE)
- 讓開發者可自行建立網路應用程式於Google平台中。
- 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: Python、Java



- 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



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

User-Level

User-Level
Middleware

Core
Middleware

System Level

IaaS
PaaS
SaaS

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, ...

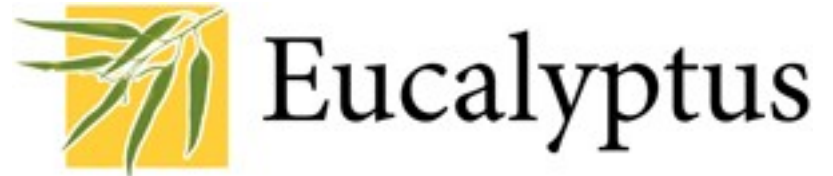
虛擬化

VM, VM management and Deployment

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

硬體設施

Infrastructure: Computer, Storage, Network

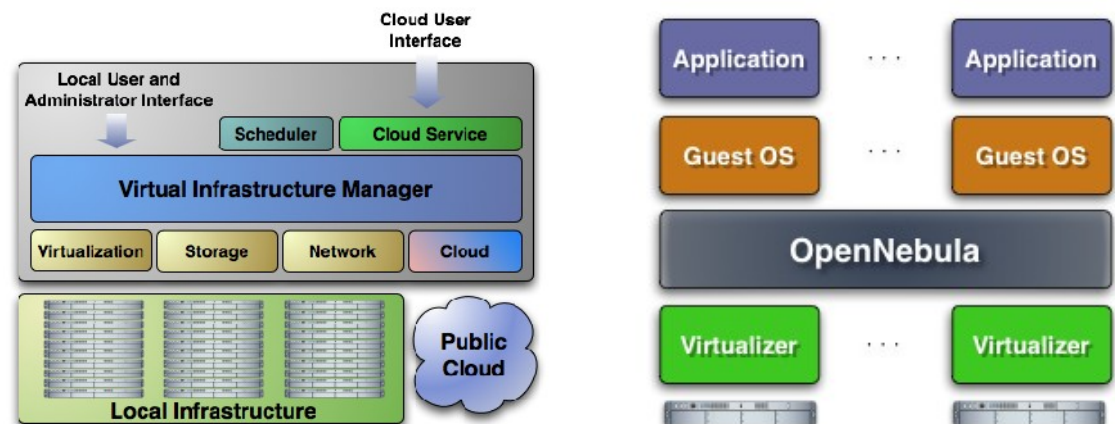


- <http://open.eucalyptus.com/>
- 原是加州大學聖塔芭芭拉分校(UCSB)的研究專案
- 目前已轉由Eucalyptus System這間公司負責維護
- 創立目的是讓使用者可以**打造自己的EC2**
- 特色是相容於 Amazon EC2 既有的用戶端介面
- 優勢是Ubuntu 9.04 已經收錄 Eucalyptus 的套件
- [Ubuntu Enterprise Cloud powered by Eucalyptus in 9.04](#)
- 目前有提供 Eucalyptus 的官方測試平台供註冊帳號
- 缺點：目前仍有部分操作需透過指令模式

- <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[Ma1Stone效能評比軟體](#)。

軟體研發 (1) : 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



root / **drbl-hadoop-0.1**

root / **hadoop-register**

Name ▲
↑ ../
📁 drbl-hadoop
📁 drbl-hadoop-mount-disk

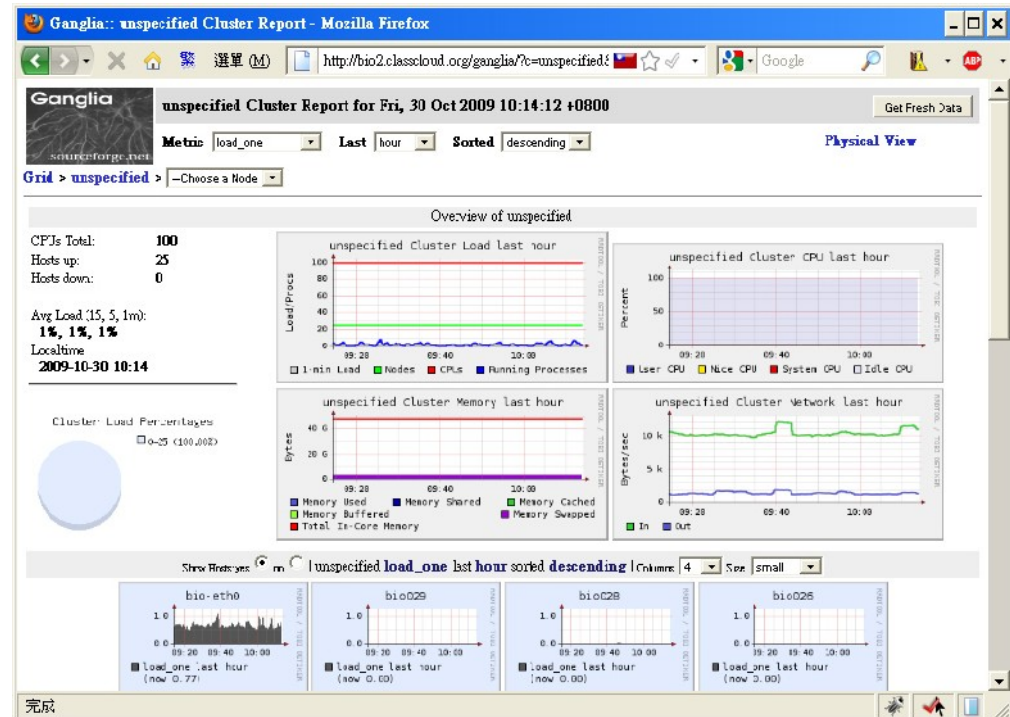
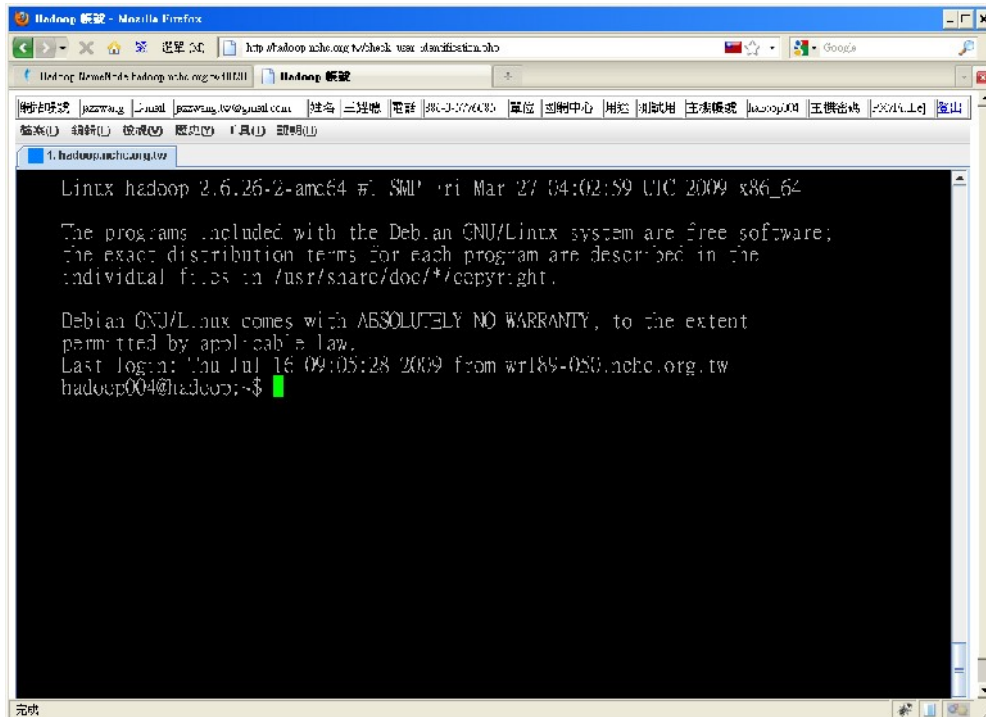
Name ▲	Size	Rev	Age	Last
↑ ../				
▶ 📁 etc		103	4 weeks	wa
📄 adduser.php	1.3 kB	85	6 weeks	wa

軟體研發 (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-bioclust** – 彙整安裝多人共用生物資訊叢集的批次檔

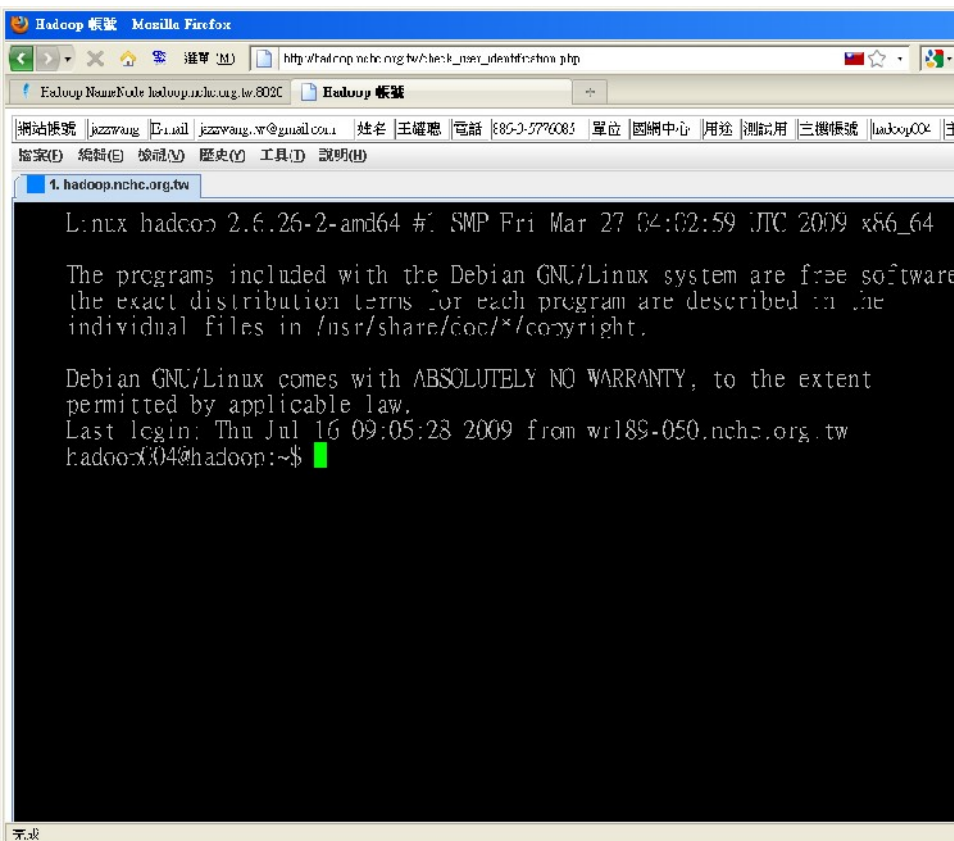
svn co <http://trac.nchc.org.tw/pub/grid/drbl-bioclust>

- 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**、**BLAST**、**SIM4**、**Clustal**、**PipMaker**、**Phylip**、**Velvet**、**Bowtie**、**SOAP**
- 成果：<http://bio2.classcloud.org>



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



```
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 09:05:23 2009 from wr185-050.nchc.org.tw
hadoop:~$
```



hadoop Hadoop Map/Reduce Administration - Mozilla Firefox

http://hadoop.nchc.org.tw:50030/jobtrac

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
none

完成

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-

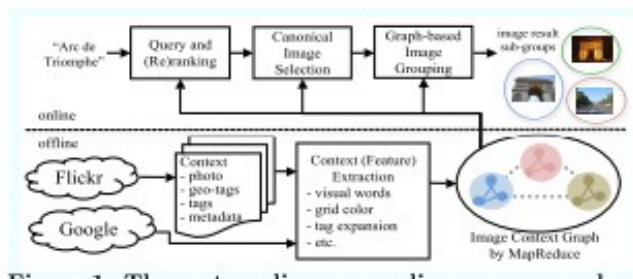
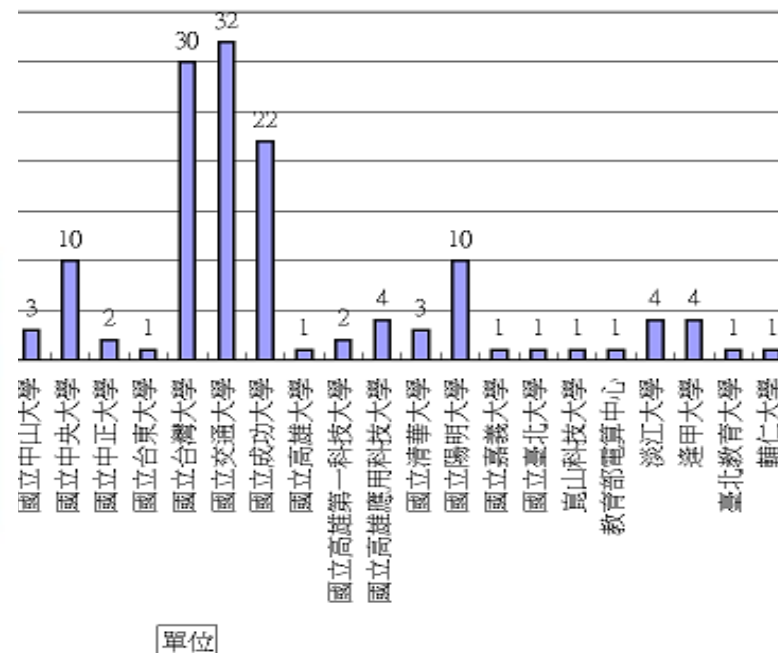


Figure 1: The system diagram: online query reranking, canonical image selection, and image result grouping are based on the image context graph, constructed offline by Hadoop MapReduce over rich context features in Flickr photos.



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	教育訓練網	教育訓練網
<p>消息 會員專區 常見問題 FAQ 住宿資訊 交通導引 電子報</p> <p>課程總覽 最近六個月的課程 課程介紹</p>	<p>消息 會員專區 常見問題 FAQ 住宿資訊 交通導引 電子報</p> <p>課程總覽 課程介紹</p>	<p>消息 會員專區 常見問題 FAQ 住宿資訊 交通導引 電子報 聯絡我</p> <p>課程總覽 課程搜尋結果 課程介紹</p>
<p>課程編號： NE-2009-TH06</p> <p>課程名稱： 雲端運算基礎課程(一)：Hadoop簡介、安裝與範例實作</p> <p>課程領域： 電腦及網路</p> <p>相關領域： 無</p> <p>上課方式： 實體教室</p> <p>上課地點： 竹科 日教室 竹</p> <p>上課時間： 2009/11/24 (二) ~ 2009/11/25 (三) 09:30 ~ 16:30</p> <p>上課總天數： 2天，共計 12 個小時</p> <p>報名截止(含)： 2009/11/22 (日) 17:00</p> <p>費截止(含)： 2009/11/23 (一) 05:00</p> <p>費截止(含)： 2009/11/23 (一) 17:00</p> <p>提供午餐： 是</p> <p>招生人數： 8 ~ 20 人</p> <p>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</p>	<p>課程編號： NE-2009-CH05</p> <p>課程名稱： 雲端運算基礎課程(二)：Xen虛擬化叢集建置、管理與</p> <p>課程領域： 電腦及網路</p> <p>相關領域： 無</p> <p>上課方式： 實體教室</p> <p>上課地點： 台中 電腦教室 A 中</p> <p>上課時間： 2009/10/27 (二) ~ 2009/10/28 (三) 09:30 ~ 16:30</p> <p>上課總天數： 2天，共計 12 個小時</p> <p>報名截止(含)： 2009/10/25 (日) 17:00</p> <p>費截止(含)： 2009/10/26 (一) 05:00</p> <p>費截止(含)： 2009/10/26 (一) 17:00</p> <p>提供午餐： 是</p> <p>招生人數： 8 ~ 20 人</p> <p>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</p>	<p>課程編號： NE-2009-CH06</p> <p>課程名稱： 雲端運算基礎課程(三)：Google App Engine體驗課程</p> <p>課程領域： 電腦及網路</p> <p>相關領域： 無</p> <p>上課方式： 實體教室</p> <p>上課地點： 台中 電腦教室 A 中</p> <p>上課時間： 2009/11/30 (一) 09:30 ~ 16:30</p> <p>上課總天數： 1天，共計 6 個小時</p> <p>報名截止(含)： 2009/11/27 (五) 17:00</p> <p>費截止(含)： 2009/11/27 (五) 17:00</p> <p>費截止(含)： 2009/11/27 (五) 17:00</p> <p>提供午餐： 是</p> <p>招生人數： 8 ~ 20 人</p> <p>講師： 國家高速網路與計算中心 鄭宗碩 先生</p> <p>報名費用： 一般人士 1000 元 學生 500 元</p>

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	

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 快速佈屬	實作九	當企鵝龍遇上小飛象	當企鵝龍遇上小飛象

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



Questions?

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

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



Powered by DRBL