



**ClassCloud: switch your PC classroom
into Cloud Computing Testbed
for Scientific Education**

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



Powered by **DRBL**

ClassCloud: turn your PC classroom into Cloud Testbed for Education

PART 1 :

(50 %)

What is Cloud Computing?

PART 2 :

(25 %)

What is DRBL?

PART 3 :

(25 %)

How we use DRBL to deploy Cloud ?

- IaaS : Virtualization (DRBL-Xen)***
- PaaS : Data Processing (DRBL-Hadoop)***
- SaaS : Bioinformatics (DRBL-biocluster)***



Part 1 : the trend of Cloud Computing

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



Powered by DRBL



What is Cloud Computing ?
Could we have a simple definition ?

Is it about buying NEW Hardware and Software?



Is it a trap to another bubble economy ?



Cloud Computing is as simple as 5..4..3..2..1...



National Definition of Cloud Computing

5 Characteristics

4 Deployment Models

3 Service Models

Detail definition:
[http://csrc.nist.gov/
groups/SNS/cloud-
computing/cloud-
def-v15.doc](http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc)

On-demand self-service.

Broad network access

Resource pooling

Rapid elasticity

Measured Service

4 Deployment Models of Cloud Computing

Public Cloud
Public Data
Non-sensitive

Target Market
is **S.M.B.**



Dynamic Resource Provisioning between multiple clouds

Hybrid Cloud

Enterprise is key market

Sensitive Data



Private Cloud

Community Cloud
Data for Sharing
Academia

3 Service Models of Cloud Computing

IaaS

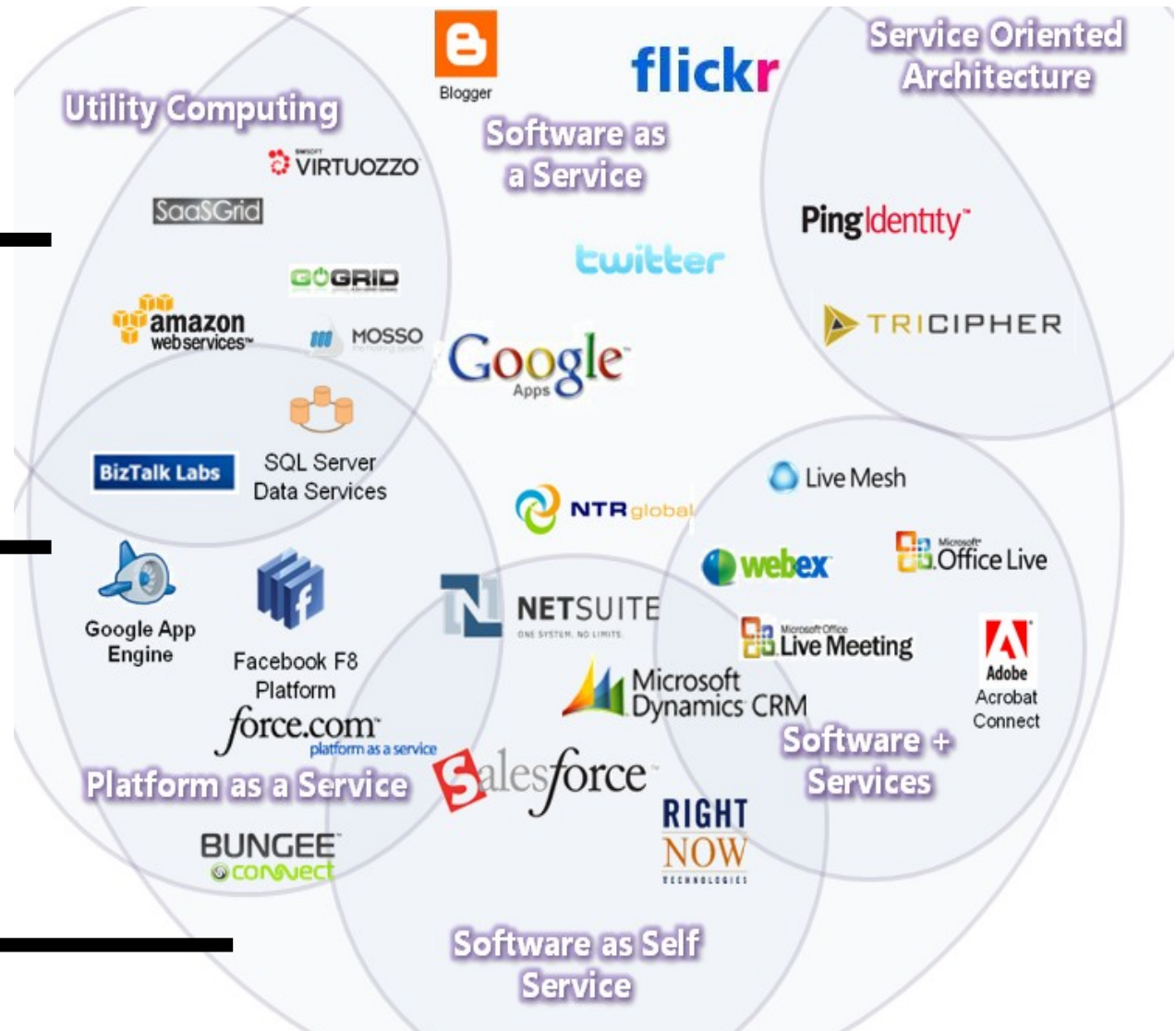
Infrastructure as a Service

PaaS

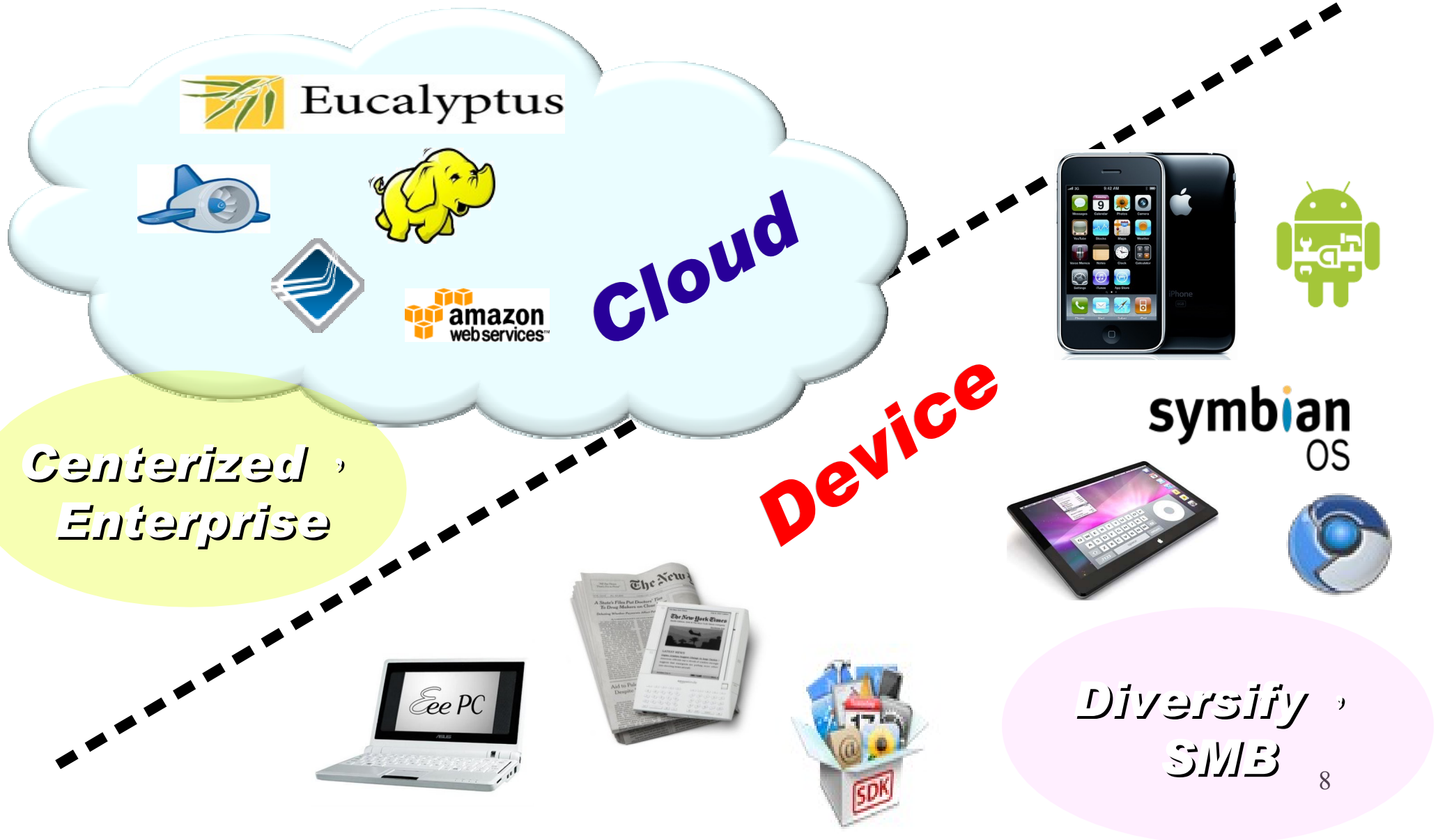
Platform as a Service

SaaS

Software as a Service



2 R&D directions : Cloud or Device



One key spirit of Cloud Computing

Anytime

Anywhere

With Any Devices

Accessing Services via Network

Cloud Computing =~ Network Computing

*Key spirit of Cloud ~
Everything as a Service !!*

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 ”*

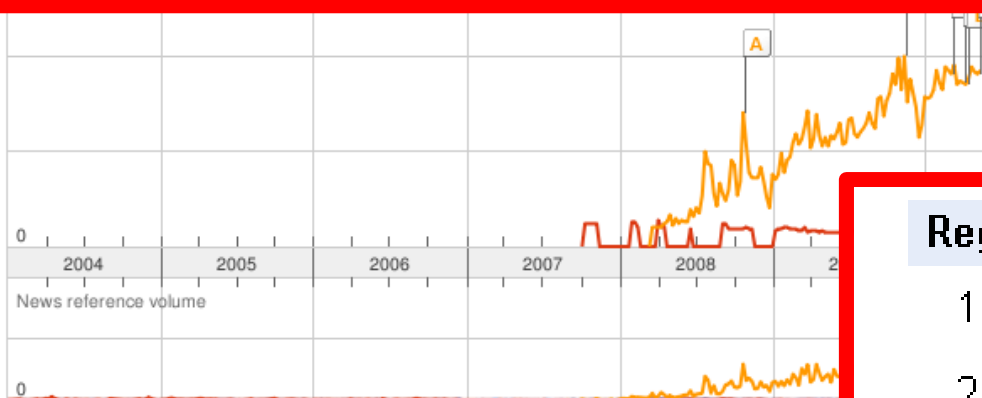
Is Cloud the trend of next 10 years ?

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

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

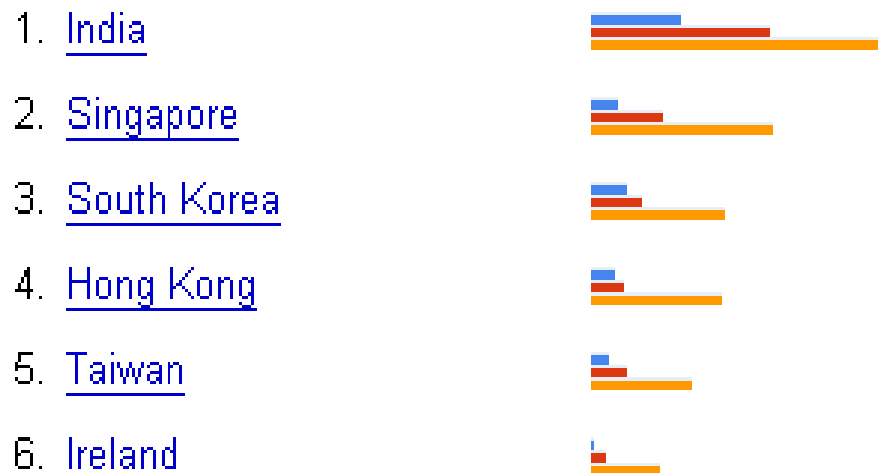
All regions All years

Search Volume index



- 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



Regions

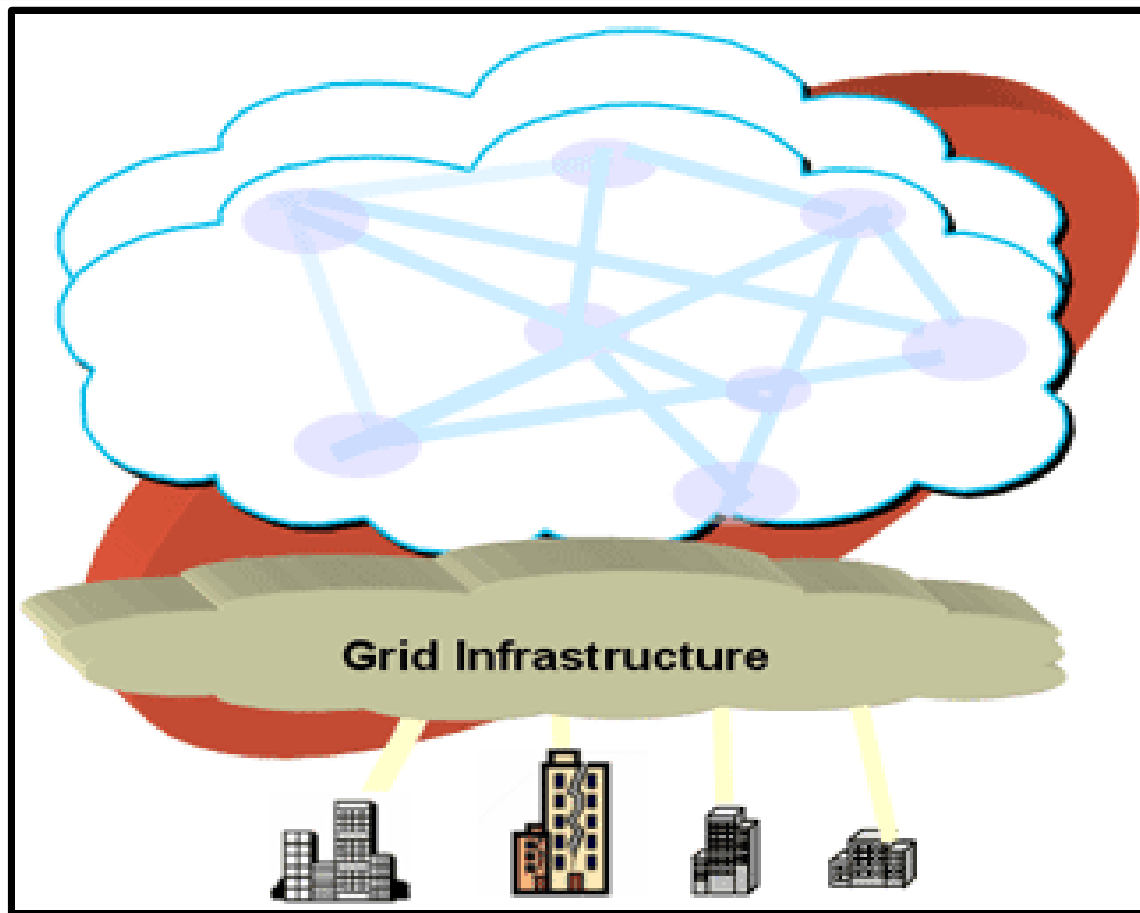


Cities

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

Is Cloud too HOT in Asia-Pacific Area ?!

Brief History of Computing



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

**Mainframe
Super
Computer**

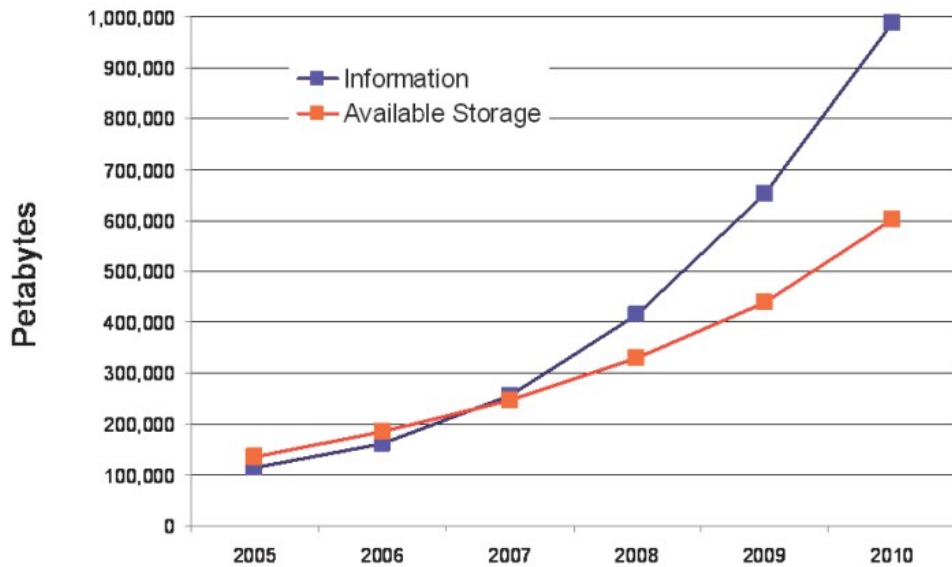
**PC | Linux
Cluster
Parallel**

**Internet
Distributed
Computing**

**Virtual Org.
Grid
Computing**

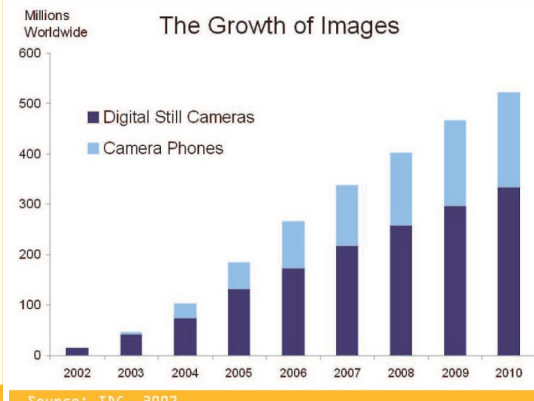
**Data Explode
Cloud
Computing**

Information Versus Available Storage



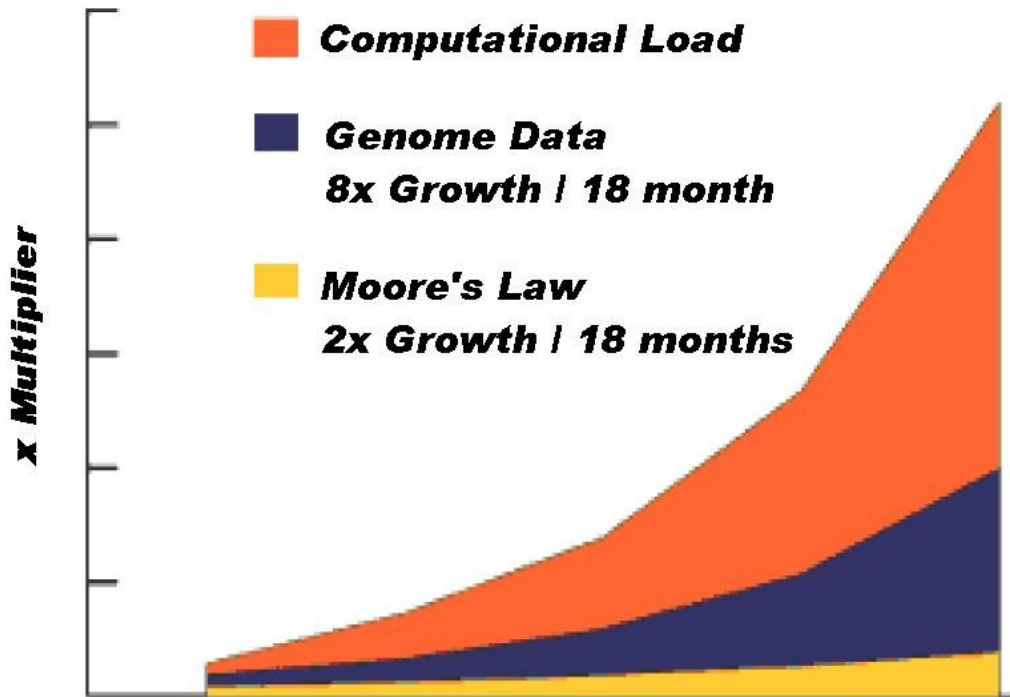
2007 Data Explore

- 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://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf>
 Source: IDC, 2007

Source: IDC, 2007



Particle Physics Large Hadron Collider (15PB)	Human Genomics (7000PB) 1GB / person 200PB+ captured 200% CAGR	World Wide Web (~1PB)	Wikipedia (10GB) 100% CAGR
Annual Email Traffic, no spam (300PB+)	Internet Archive (1PB+)	Estimated On-line RAM in Google (8PB)	Personal Digital Photos (1000PB+) 100% CAGR
200 of London's Traffic Cams (8TB/day)	2004 Walmart Transaction DB (500TB)	Typical Oil Company (350TB+)	Merck Bio Research DB (1.5TB/qtr)
UPMC Hospitals Imaging Data (500TB/yr)	MIT Babytalk Speech Experiment (1.4PB)	Terashake Earthquake Model of LA Basin (1PB)	One Day of Instant Messaging in 2002 (750GB)
Total digital data to be created this year 270,000PB (IDC)			

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

Phillip B. Gibbons, Data-Intensive Computing Symposium

How can we build our Private Cloud ??

Public Cloud
Public Data
Non-sensitive



Target Market
is **S.M.B.**

Hybrid
Cloud

Enterprise is
key market

Sensitive Data



Community Cloud
Data for Sharing
Academia

Private Cloud

Reference Cloud Architecture

Application

Social Computing, Enterprise, ISV, ...

Programming

Web 2.0, Mashups, Workflows, ...

Management

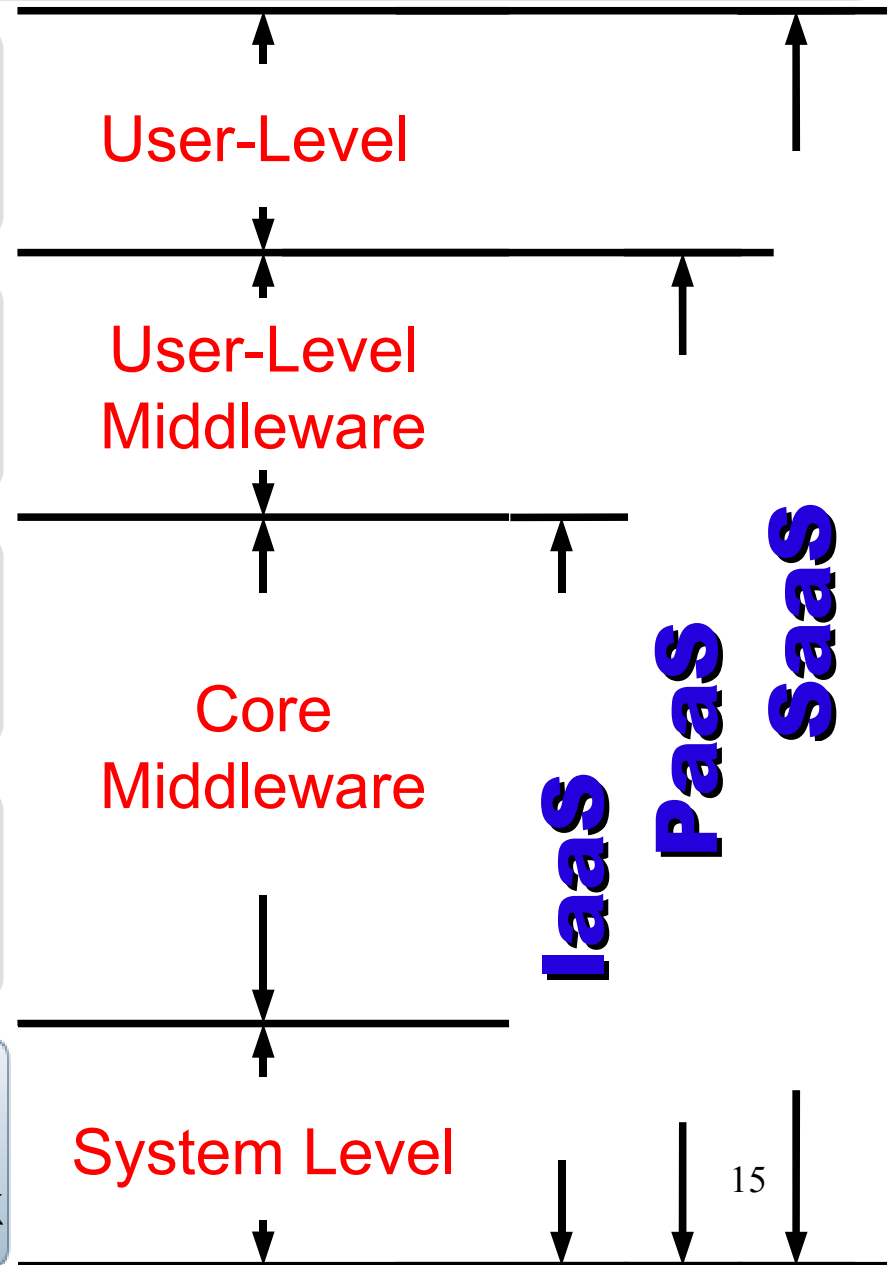
Qos Negotiation, Admission Control, Pricing, SLA Management, Metering...

Virtualization

VM, VM management and Deployment

Physical Hardware

Infrastructure: Computer, Storage, Network



Open Source for Private Cloud

Application

Social Computing, Enterprise, ISV, ...

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

Programming

Web 2.0, Mashups, Workflows, ...

Hadoop (MapReduce),
Sector/Sphere, AppScale

Management

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

OpenNebula, Enomaly,
Eucalyptus, OpenQRM, ...

Virtualization

VM, VM management and Deployment

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

Physical Hardware

Infrastructure: Computer, Storage, Network



Part 2 : Introduction to DRBL

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



Powered by **DRBL**

What is DRBL ??

- **Diskless Remote Boot in Linux**
- Network is cheap, and our time is expensive
- In simple words, DRBL is
 - Replace IDE/SATA cable with network cable
 - 40+ student PCs connected to one DRBL server



**Diskfull
PC**



=



+



+



**Diskless
PC**



Server



At First, We have "4 + 1" PC Cluster

It'd better be
2ⁿ



Manage
Scheduler

*Then, We connect 5 PCs with
Gigabit Ethernet Switch*



GiE Switch



*10/100/1000
Mbps*



***Add 1 NIC
for WAN***



Compute Nodes

4 Compute Nodes will communicate via LAN Switch. Only Manage Node have Internet Access for Security!



WAN



Manage Node



Compute Nodes

Basic System Setup for Cluster

Messaging

MPICH

Account Mgmt.

SSHD

NIS

YP

GCC

GNU Libc

Bash

Perl



Kernel Module

Linux Kernel

Boot Loader

On **Manage Node**,
We need to install **Scheduler** and
Network File System for sharing
Files with **Compute Node**

Job Mgmt.

OpenPBS

File Sharing

NFS

Extra

Messaging

MPICH

GCC

Bash

Perl

Account Mgmt.

SSHD

NIS

YP

GNU Libc



Kernel Module

Linux Kernel

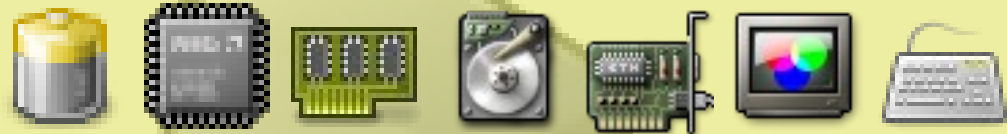
Boot Loader

1st, We install Base System of **GNU/Linux on **Management Node**. You**

can choose:

**Redhat, Fedora, CentOS, Mandriva,
Ubuntu, Debian, ...**

GNU Libc



Kernel Module

Linux Kernel

Boot Loader

2nd, We install **DRBL package and
configure it as **DRBL Server**.**

**There are lots of service needed:
SSH, DHCP, TFTP, NFS Server,
NIS Server, YP Server ...**

Network Booting

Account Mgmt.

NFS

TFTP

DHCP

SSH

NIS

YP

Perl

Bash

GNU Libc

DRBL Server

*based on existing
Open Source and
keep Hacking!*



Kernel Module

Linux Kernel

Boot Loader

After running “**drblsrv -i**” & “**drblpush -i**”, there will be **pxelinux**, **vmlinux-pex**, **initrd-pxe** in TFTPROOT, and different **configuration files** for each Compute Node in NFSROOT

NFS

TFTPD

DHCPD

SSHD

NIS

YP

Config. Files

Ex. hostname

initrd-pxe

vmlinux-pxe

pxelinux

GNU Libc



Kernel Module

Linux Kernel

Boot Loader

3nd, We enable *PXE* function in *BIOS* configuration.

BIOS PXE

BIOS PXE

BIOS PXE

BIOS PXE

NFS

TFTPD

DHCPD

SSHD

NIS

YP

Config. Files

Ex. hostname

initrd-pxe

vmlinuz-pxe

pxelinux

GNU Libc



Kernel Module

Linux Kernel

Boot Loader

**While Booting, *PXE* will query
IP address from *DHCPD*.**

BIOS PXE

BIOS PXE

BIOS PXE

BIOS PXE

NFS

TFTPD

DHCPD

SSHD

NIS

YP

Config. Files
Ex. hostname

initrd-pxe

vmlinuz-pxe

pxelinux

GNU Libc



Kernel Module

Linux Kernel

Boot Loader

**While Booting, *PXE* will query
IP address from *DHCPD*.**

IP 1

IP 2

IP 3

IP 4

NFS

TFTPD

DHCPD

SSHD

NIS

YP

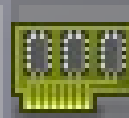
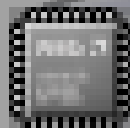
Config. Files
Ex. hostname

initrd-pxe

vmlinuz-pxe

pxelinux

GNU Libc



Kernel Module

Linux Kernel

Boot Loader

After PXE get its IP address, it will download booting files from **TFTPD.**

IP 1

IP 2

IP 3

IP 4

NFS

TFTPD

DHCPD

SSHD

NIS

YP

Config. Files

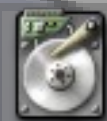
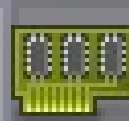
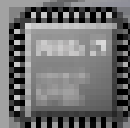
Ex. hostname

initrd-pxe

vmlinuz-pxe

pxelinux

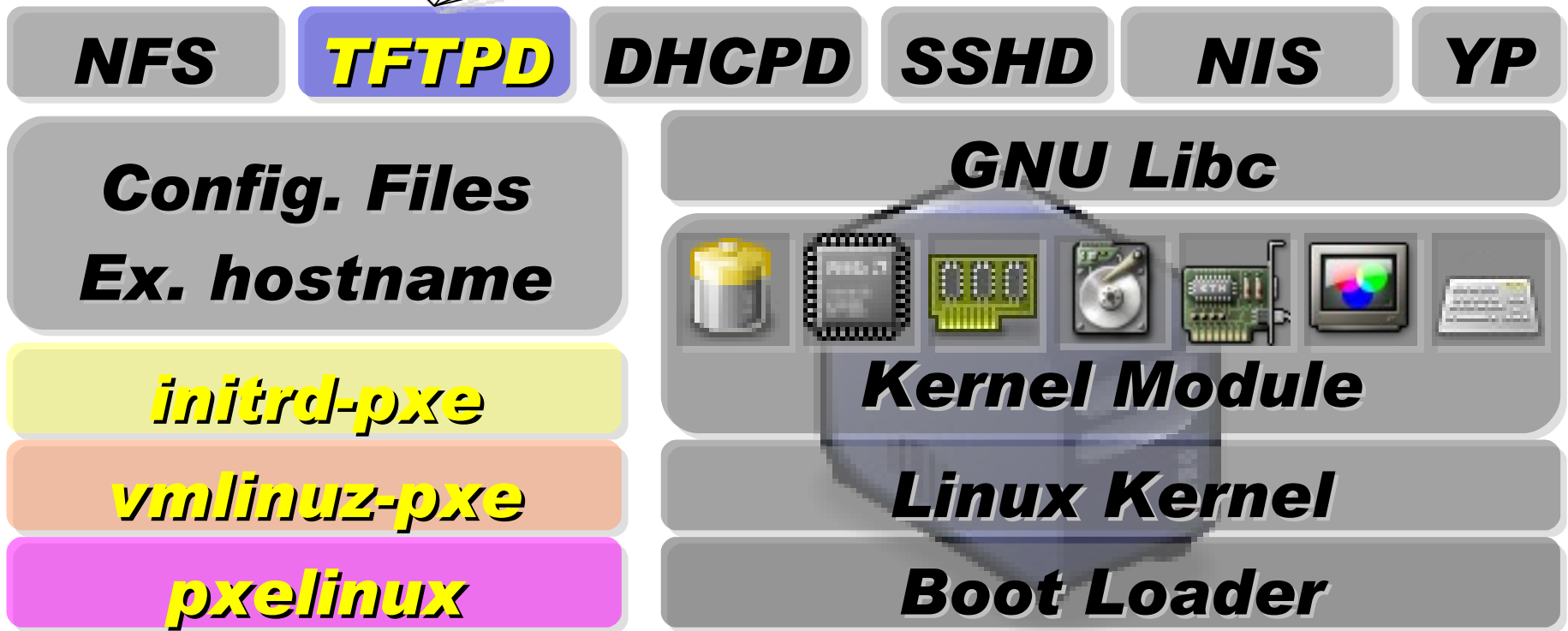
GNU Libc

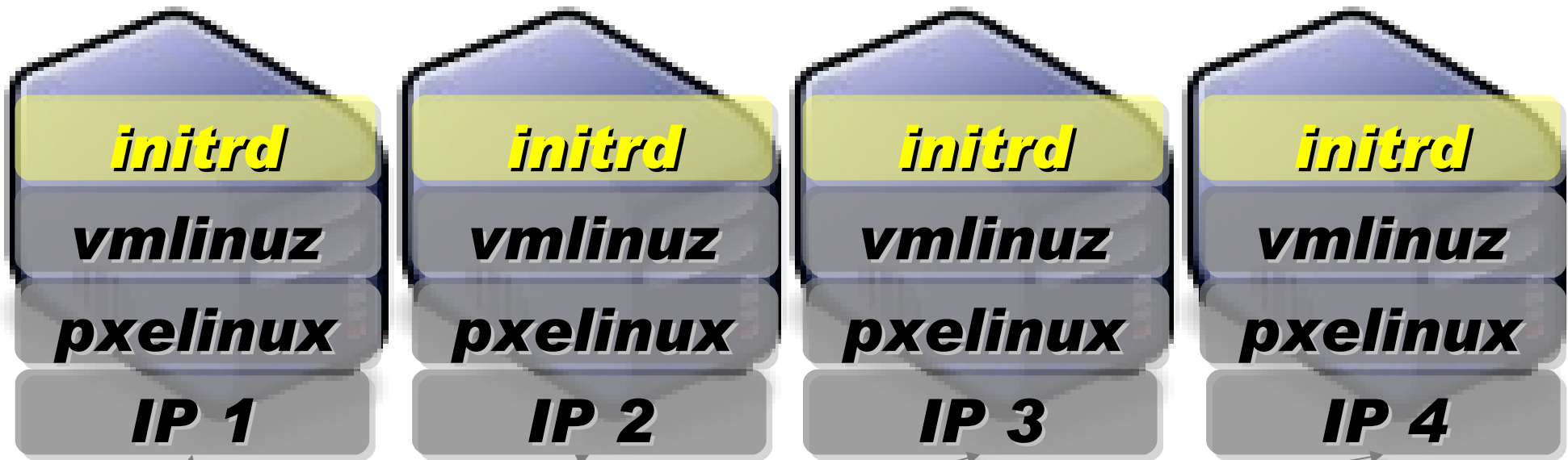


Kernel Module

Linux Kernel

Boot Loader





NFS **TFTPD** **DHCPD** **SSHD** **NIS** **YP**

Config. Files *GNU Libc*

After downloading booting files, scripts in **initrd-pxe will config **NFSROOT** for each Compute Node.**

pxelinux **Boot Loader**

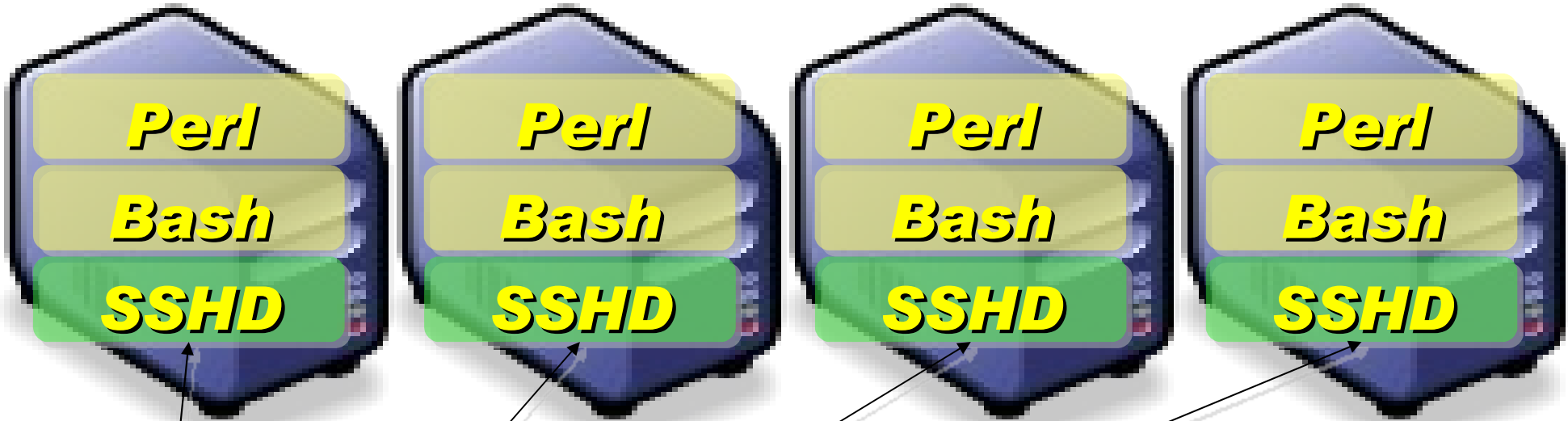


NFS **TFTPD** **DHCPD** **SSHD** **NIS** **YP**

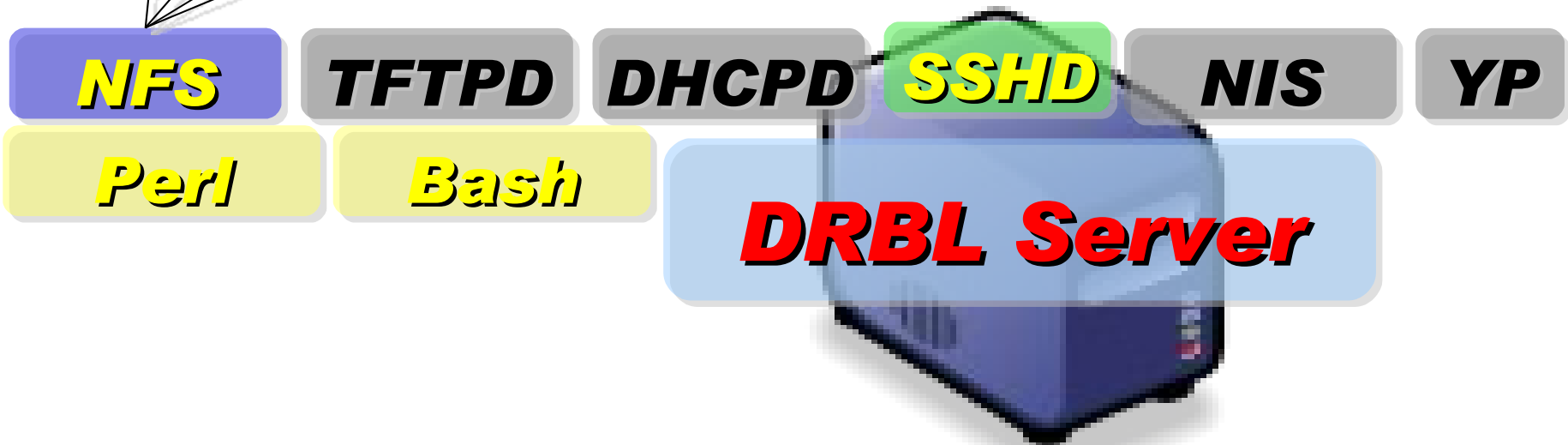
Config. Files
Ex. hostname

initrd-pxe
 vmlinuz-pxe
 pxelinux





**Applications and Services will also
deployed to each Compute Node
via **NFS****





*With the help of **NIS** and **YP**,
You can login each Compute Node
with the **Same ID / PASSWORD**
stored in **DRBL Server!***

SSH Client



DRBL Server





Part 3 : How we use DRBL to deploy Cloud Testbed ?

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



Powered by **DRBL**

Building IaaS using DRBL-Xen

Application

Social Computing, Enterprise, ISV, ...

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

Programming

Web 2.0, Mashups, Workflows, ...

Hadoop (MapReduce),
Sector/Sphere, AppScale

Management

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

OpenNebula, Enomaly,
Eucalyptus, OpenQRM, ...

Virtualization

VM, VM management and Deployment

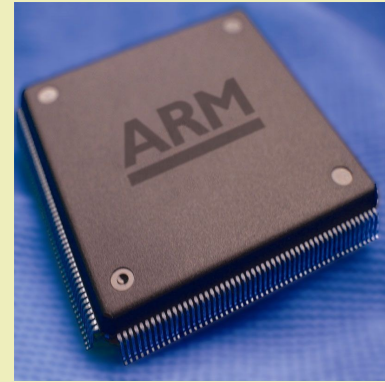
Xen, KVM, VirtualBox,
QEMU, OpenVZ, ...

Physical Hardware

Infrastructure: Computer, Storage, Network

Virtualization ?? Emulator ??

Virtual Hardware / OS



mame4iphone

Latest Version:

MAME 0.138 15 May 10

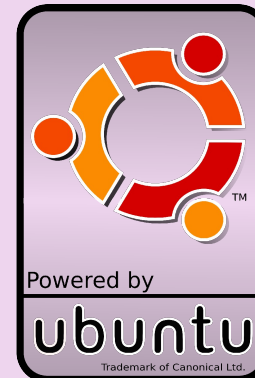
— DOWNLOAD NOW —

[Download source updates to MAME 0.138](#)



Mac4Lin

QEMU



Physical Hardware / OS

What is Virtualization ??

Application Virtualization **Ex. VMWare ThinApp**

Desktop Virtualization
Client Virtualization **Ex. XenDesktop**

Presentation Virtualization **Ex. VNC, M\$ RDP**

OS-level Virtualization **Ex. Xen, KVM**

Network Virtualization **Ex. OpenFlow**

Storage Virtualization **Ex. NetApp**

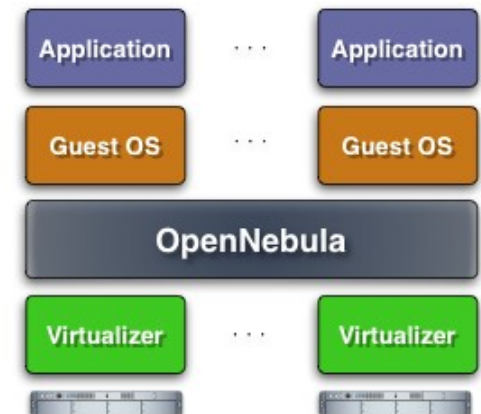
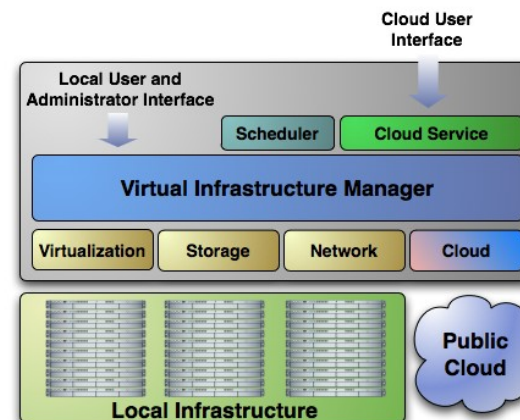
Database Virtualization

Data Virtualization



- <http://open.eucalyptus.com/>
- It was a research project of UCSB, USA
- Now Eucalyptus System provide technical supports.
- It designed to help user **to build their own Amazon EC2**
- Its feature is compatible with existing EC2 client.
- [Ubuntu Enterprise Cloud powered by Eucalyptus in 9.04](#)
- You can register trail account at <http://open.eucalyptus.com/>
- Cons : you might need to type commands in some case

- <http://www.opennebula.org>
- Sponsor by European Union FP7
- Turn Physical Cluster into Virtual Cluster
- manage **status, scheduling and migration of virtual cluster**
- [Ubuntu 9.04 provide package of opennebula](#)
- Cons : You need to type commands to check or migration



Building IaaS using DRBL-Xen

- DRBL-Xen is still need more work to intergrate into DRBL
- Manual procedure could be found at
 - http://trac.nchc.org.tw/grid/wiki/jazz/DRBL_Xen

trac
Integrated SCM & Project Management

logged in as jazz | [Logout\(o\)](#) | [Preferences](#) | [Help/Guide](#) | [About Trac](#)

[Wiki](#) | [Timeline](#) | [Roadmap](#) | [Browse Source](#) | [View Tickets](#) | [New Ticket](#) | [Search](#) | [Admin](#)

[Start Page](#) | [Index](#) | [History](#) | [Last Change](#)

Massive Deployment of Xen enabled Virtual Cluster using Diskless Remote Boot Linux

Test Environment

- VMWare Server 1.0.3
- VMWare Image debian 4.0r1 as DRBL Server

[Massive Deployment of Xen enabled Virtual Cluster using Diskless Remote ...](#)

- [Test Environment](#)
- [Pre-installation](#)
- [Install and Configure DRBL](#)
- [Configure PXELinux for Xen Network Booting](#)
- [Enable Xend in default rcX.d](#)
- [Know Issue of DRBL \(\) with Xen](#)
- [Create Xen DomU configure files](#)
- [Xen DomU PXE Booting](#)
 - Case 1: HVM supported
 - Case 2: only Para-Virtualization supported
- [Reference](#)

DRBL_Xen

DRBL_Xen

State: Powered off

Guest OS: Other Linux

Configuration file: C:\Virtual Machines\DRBL_Xen\drbl_xen.vmx

Version: Current virtual machine for VMware Server 1.0.3

Commands

- [Start this virtual machine](#)
- [Edit virtual machine settings](#)

Devices

- Memory 256 MB
- Hard Disk (SCSI 0:0)
- DRBL-Xen

Building PaaS using DRBL-Hadoop

Application

Social Computing, Enterprise, ISV,...

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

Programming

Web 2.0, Mashups, Workflows, ...

Hadoop (MapReduce),
Sector/Sphere, AppScale

Management

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

OpenNebula, Enomaly,
Eucalyptus, OpenQRM, ...

Virtualization

VM, VM management and Deployment

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

Physical Hardware

Infrastructure: Computer, Storage, Network

- <http://hadoop.apache.org>
- Hadoop is Apache Top Level Project
- Major sponsor is Yahoo!
- Developed by Doug Cutting
- Written by Java, it provides HDFS and MapReduce API
- Used in Yahoo since year 2006
- It had been deploy to 4000+ nodes in Yahoo
- Design to process dataset in Petabyte
- Facebook 、 Last.fm 、 Joost are also powered by Hadoop

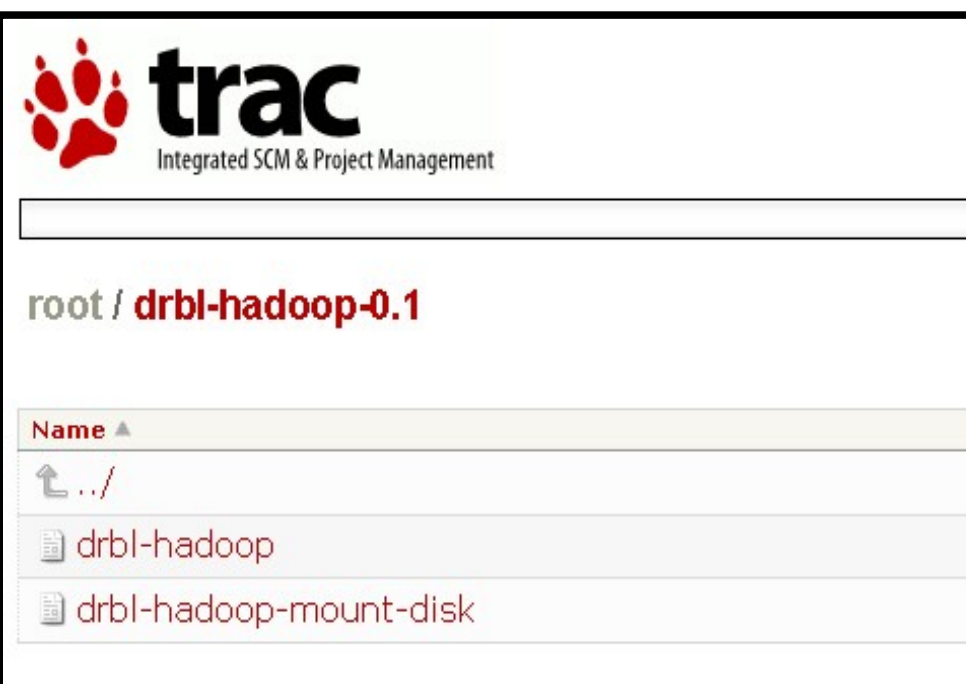


- <http://sector.sourceforge.net/>
- Developed by National Center for Data Mining, USA
- Written by C/C++, so performance is better than Hadoop
- Provide file system **similar to** Google File System and MapReduce API
- Based on [UDT](#) which enhance the network performance
- [Open Cloud Consortium](#) provide [Open Cloud Testbed](#) and develop [MalStone toolkit for benchmark](#)



Building PaaS using DRBL-Hadoop

- Used in <http://hadoop.nchc.org.tw>
- drbl-hadoop – mount local disk for HDFS and MapReduce
svn co <http://trac.nchc.org.tw/pub/grid/drbl-hadoop>
- hadoop-register – web interface with ssh applet
svn co <http://trac.nchc.org.tw/pub/cloud/hadoop-register>



The screenshot shows the Trac web interface for the project 'drbl-hadoop-0.1'. The header includes the Trac logo and the text 'trac Integrated SCM & Project Management'. Below the header is a search bar. The main content area shows the breadcrumb 'root / drbl-hadoop-0.1'. A table lists the files in the directory:

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



The screenshot shows the Trac web interface for the project 'hadoop-register'. The header includes the Trac logo and the text 'trac Integrated SCM & Project Management'. Below the header is a search bar. The main content area shows the breadcrumb 'root / hadoop-register'. A table lists the files and directories in the directory:

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

Demo : hadoop.nchc.org.tw for multi-users

- **DRBL Server x 1 (hadoop)**
- **DRBL Client x 19 (hadoop101~hadoop119)**
- **Based on Cloudera Debian package **and enhance security setting and permission for multi-users.****

The image shows two overlapping windows from a Mozilla Firefox browser. The left window displays a terminal session with the following output:

```
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:~$
```

The right window displays the 'hadoop Hadoop Map/Reduce Administration' web interface. It shows the following status information:

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

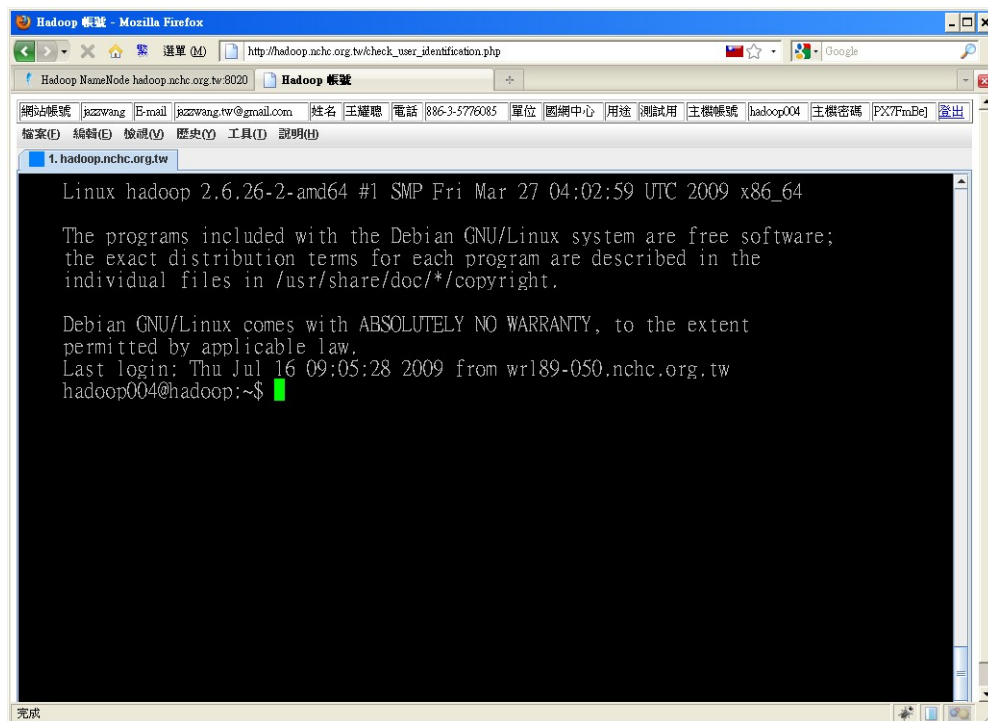
Running Jobs
none

Building SaaS using DRBL-biocluster

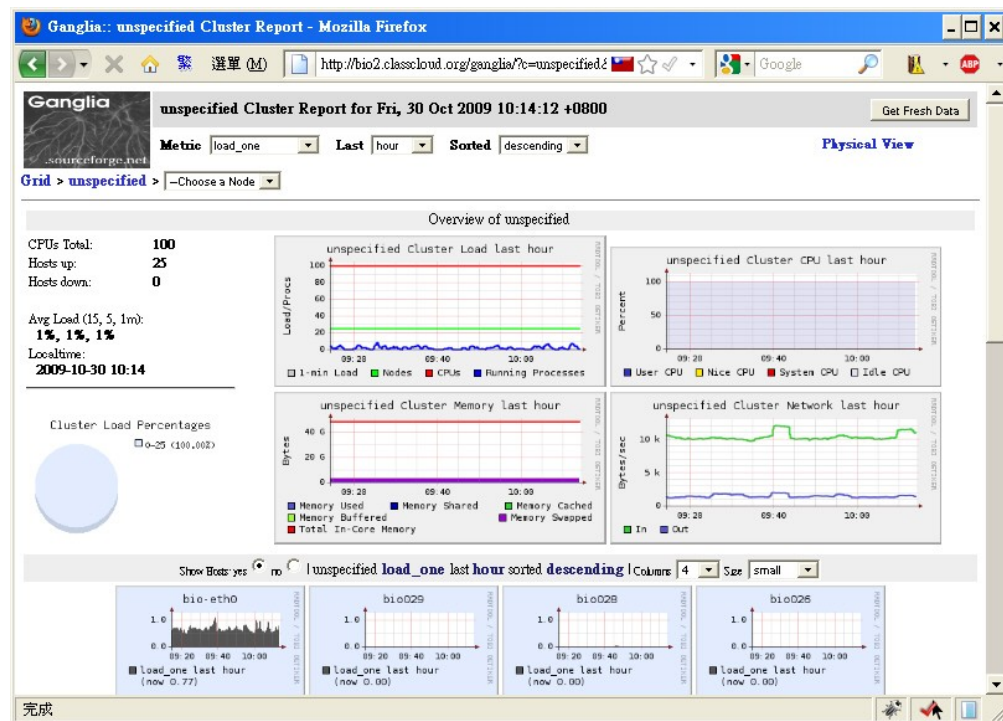
- **Need more time to package related software.**
- **drbl-biocluster – batch script of Debian to install bioinformatics related softwares**

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

- **Including DRBL、MPICH2、R、Rmpi、BioConductor、Ganglia、Nagios、AutoFACT、BLAST、SIM4、Clustal、PipMaker、Phylip、Eland、Velvet、Bowtie、SOAP**



```
Hadoop 帳號 - Mozilla Firefox
http://hadoop.nchc.org.tw/check_user_identification.php
Hadoop NameNode hadoop.nchc.org.tw:8020
網站帳號 [kczwang] E-mail [kczwang.tw@gmail.com] 姓名 [王耀聰] 電話 [886-3-5776085] 單位 [國網中心] 用途 [測試用] 主機帳號 [hadoop04] 主機密碼 [P%7FmBe]
1. hadoop.nchc.org.tw
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:~$
```



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