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 : Virtaulization (DRBL-Xen)
- PaaS : Data Processing (DRBL-Hadoop)
- SaaS : Bioinformatics (DRBL-biocluster)

Part 1 : the trend of Cloud Computing

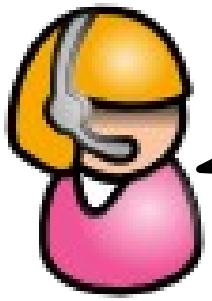
Jazz Wang
Yao-Tsung Wang
jazz@nchc.org.tw

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

On-demand self-service.

Broad network access

Resource pooling

Rapid elasticity

Measured Service

4 Deployment Models of Cloud Computing

Dynamic Resource Provisioning between multiple clouds

Public Cloud

Public Data

Non-sensitive

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

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

Cloud

Device

symbian
OS

Centerized
Enterprise

Diversify
SMB

One key spirit of Cloud Computing

Anytime

Anywhere

With Any Devices

Accessing Services via Network

Key spirit of Cloud ~

Everything as a Service !!

Cloud Computing =~ Network Computing

CIO 2010 : Virtualization, Cloud and Web 2.0

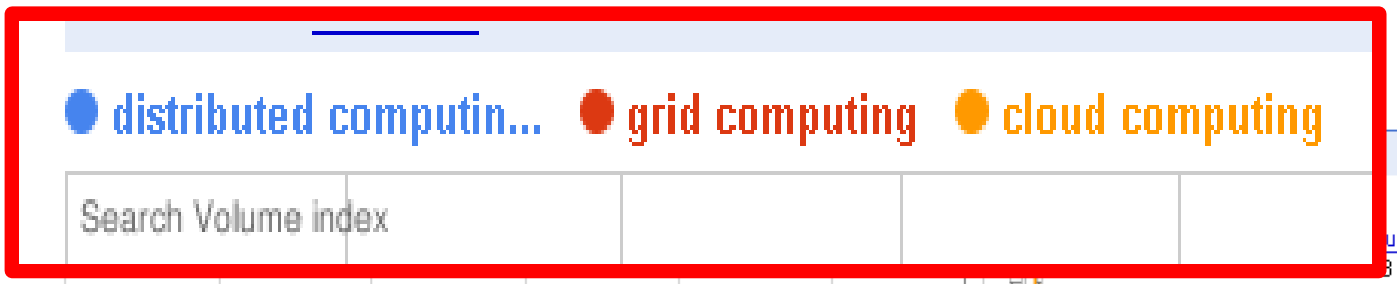
CIO strategic technologies reflect increased interest in "lighter-weight" solutions

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 ?



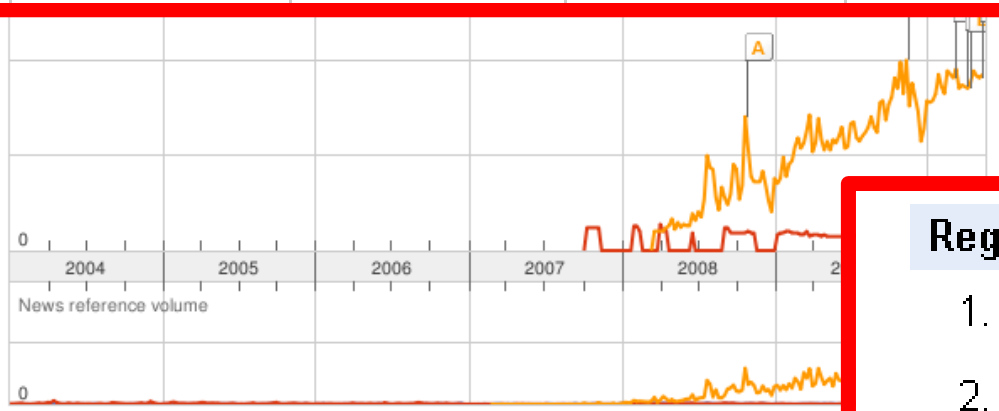
Sign in to see and export additional Tren

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



Rank by cloud computing

Regions

1. India
2. Singapore
3. South Korea
4. Hong Kong
5. Taiwan
6. Ireland

Cities

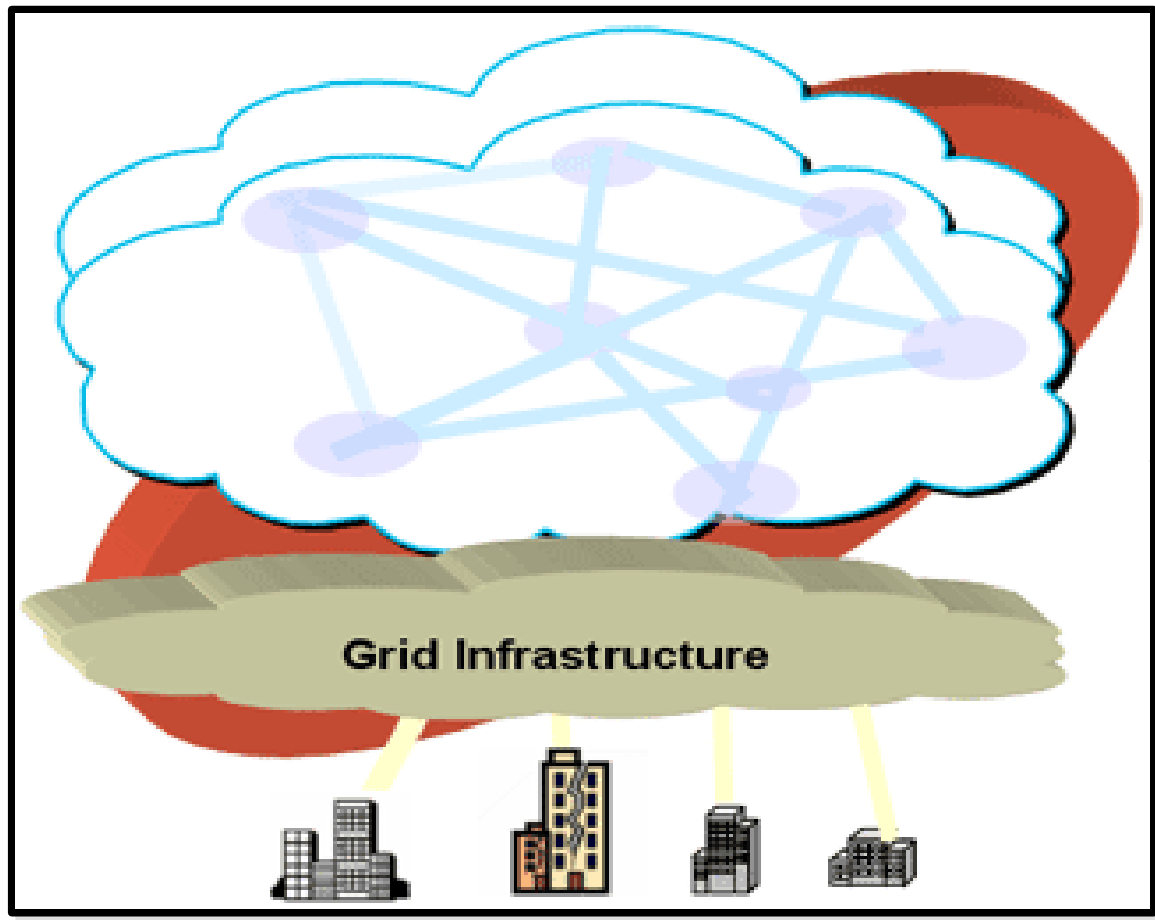
1. Bangalo
2. Mahape
3. Mumbai
4. Chennai
5. San Jos
6. Delhi, In

Regions

1. India
2. Singapore
3. South Korea
4. Hong Kong
5. Taiwan
6. Ireland

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

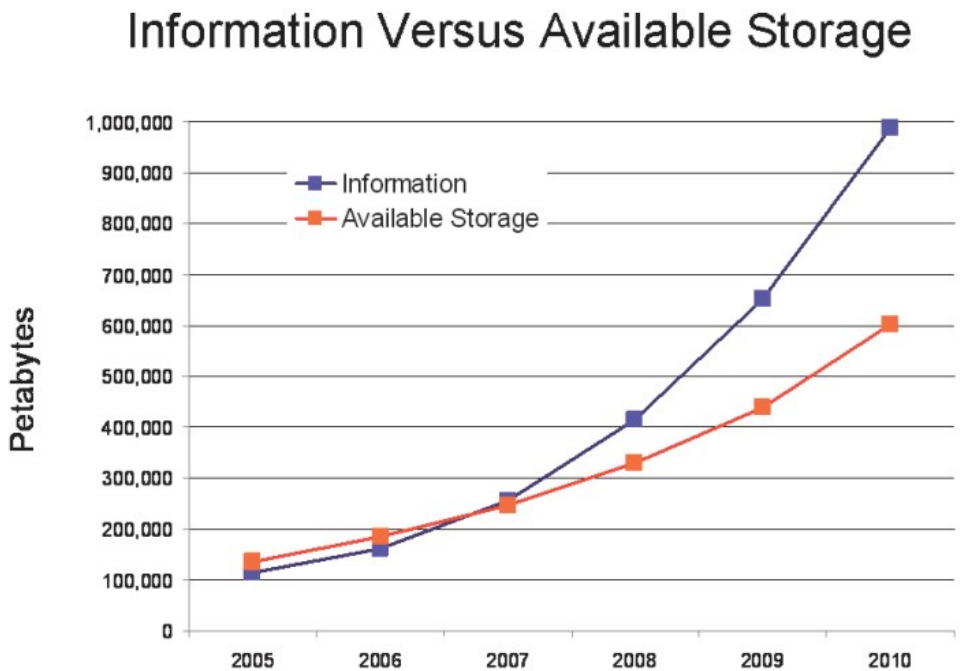
Internet Distributed Computing

Virtual Org. Grid Computing

Data Explode Cloud Computing

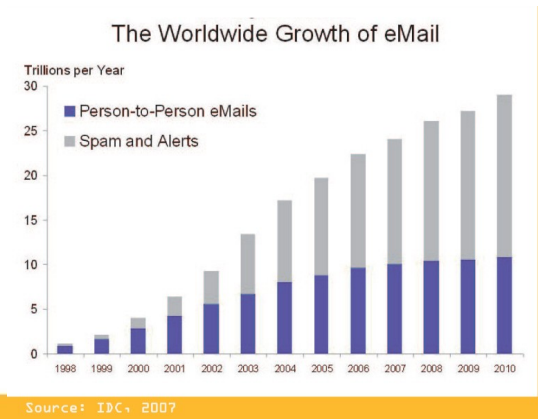
2007 Data Explore

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

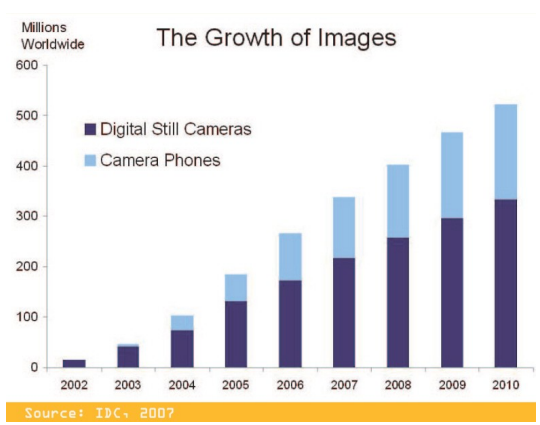


Source: http://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf

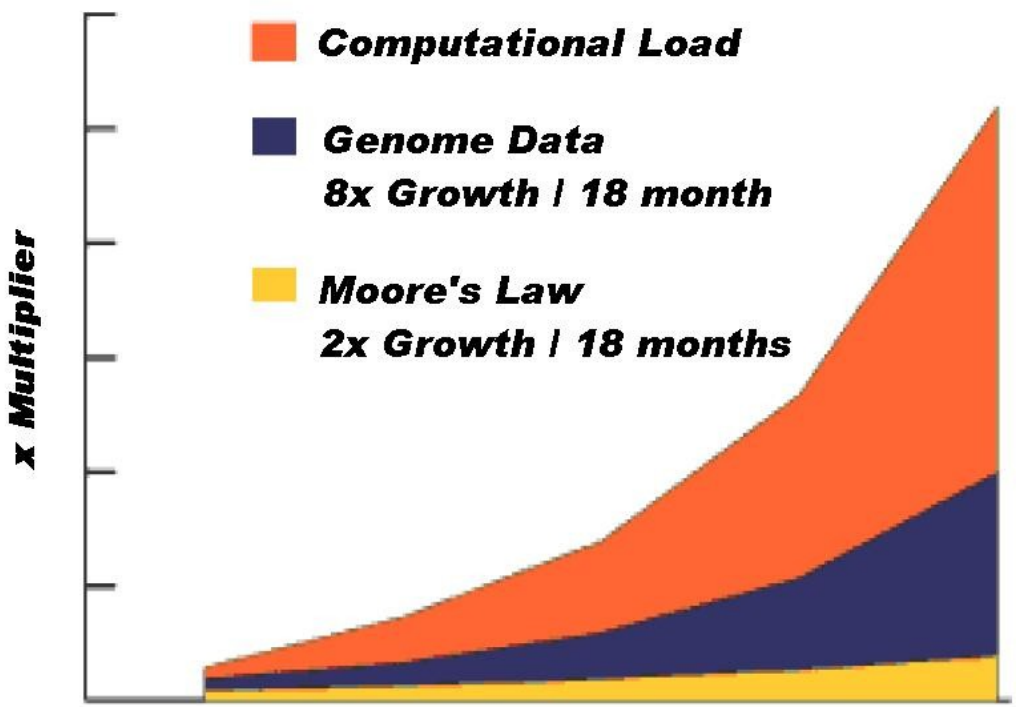
Source: IDC, 2007



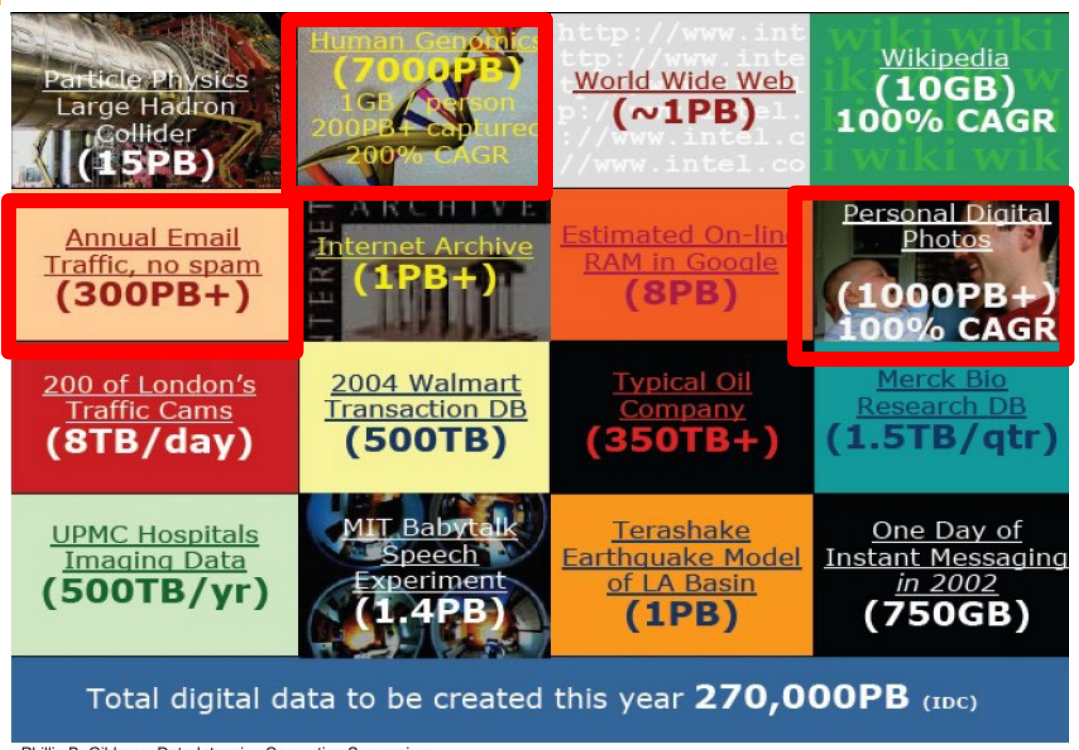
Source: IDC, 2007



Source: IDC, 2007



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

Private Cloud

Community Cloud

Data for Sharing

Academia

Reference Cloud Architecture

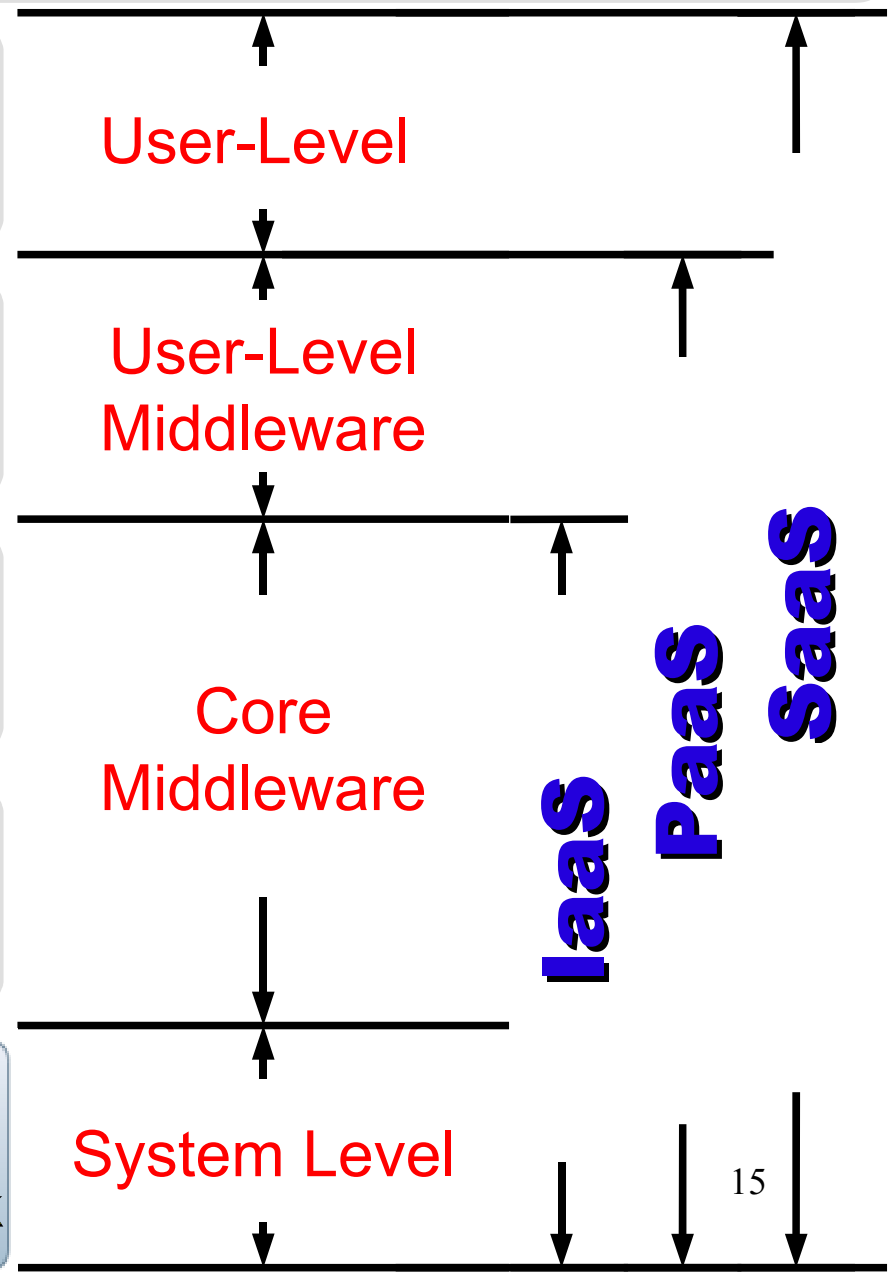
Application
Social Computing, Enterprise, ISV,…

Programming
Web 2.0, Mashups, Workflows, …

Management
Qos Neqotiation, Ddmission 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 Neqotiation, Ddmission Control, Pricing, SLA Management, Metering…

OpenNebula, Enomaly, Eucalyptus , OpenQRM, ...

Virtualization
VM, VM management and Deployment

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

Part 2 : Introduction to DRBL

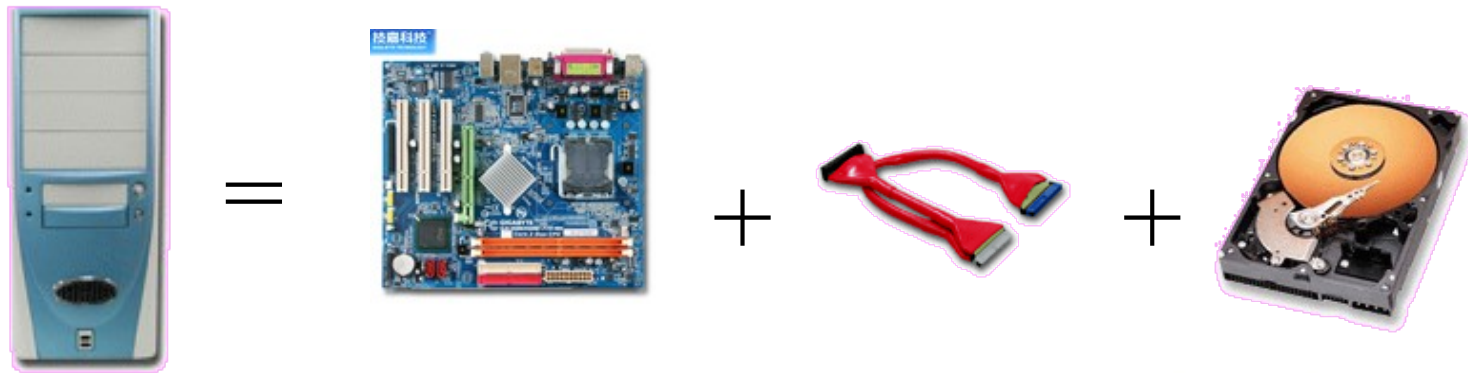
Jazz Wang
Yao-Tsung Wang
jazz@nchc.org.tw

Powered by DRBL

What is DRBL ??

- ***Diskless Remote Boot in Linux***
- Network is cheap, and our time is expansive
- 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

source: http://www.mren.com.tw

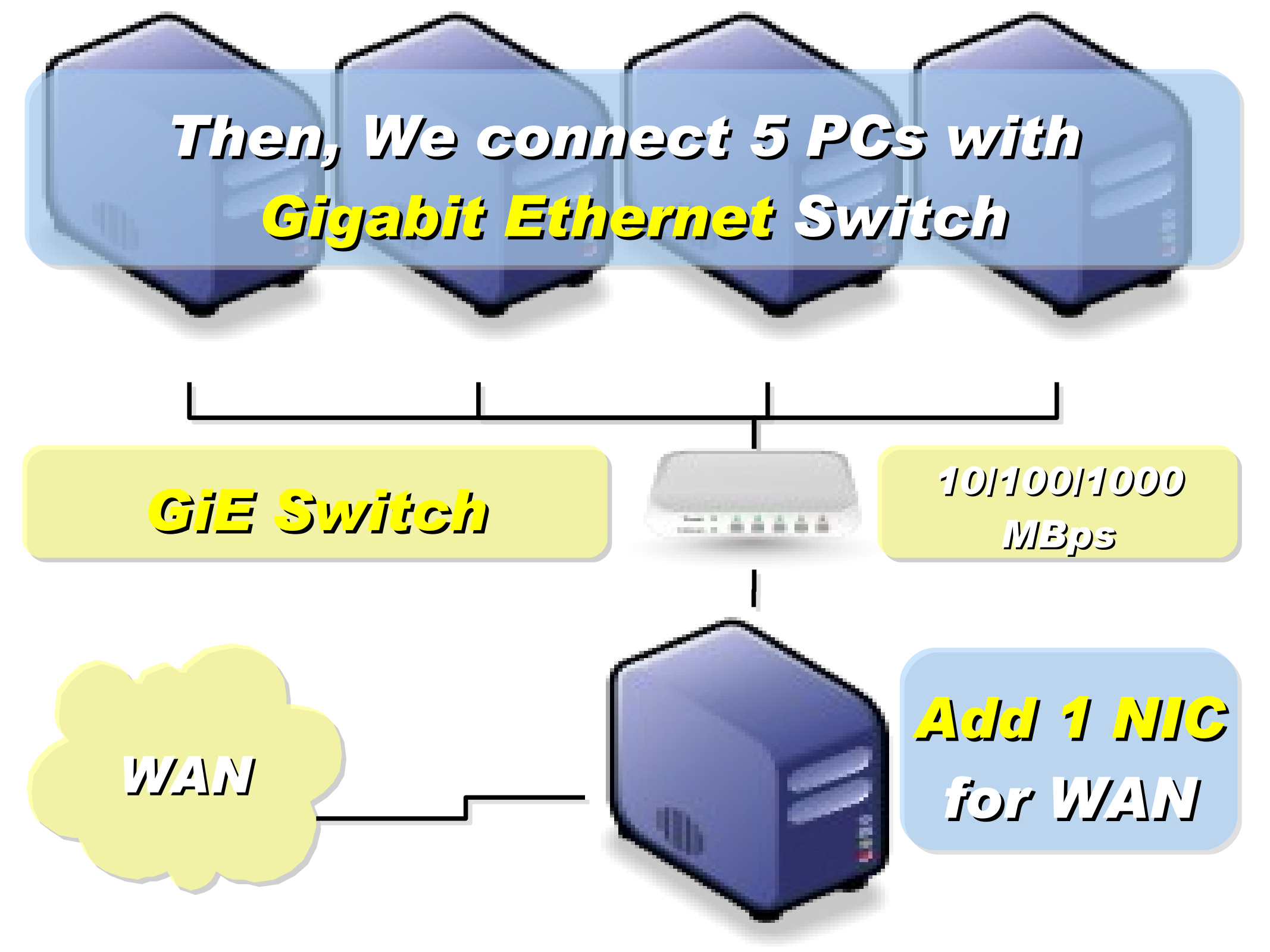
At First, We have “ 4 + 1 ” PC Cluster

It'd better be

2^n

Manage

Scheduler


Then, We connect 5 PCs with
Gigabit Ethernet Switch
GiE Switch
10/100/1000
MBps
WAN
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

Basic System Setup for Cluster

Messaging

Account Mgnt.

MPICH

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 Mgnt.

OpenPBS

File Sharing

NFS

Extra

Messaging

MPICH

Account Mgnt.

SSHD

NIS

YP

GCC

GNU Libc

Bash

Perl

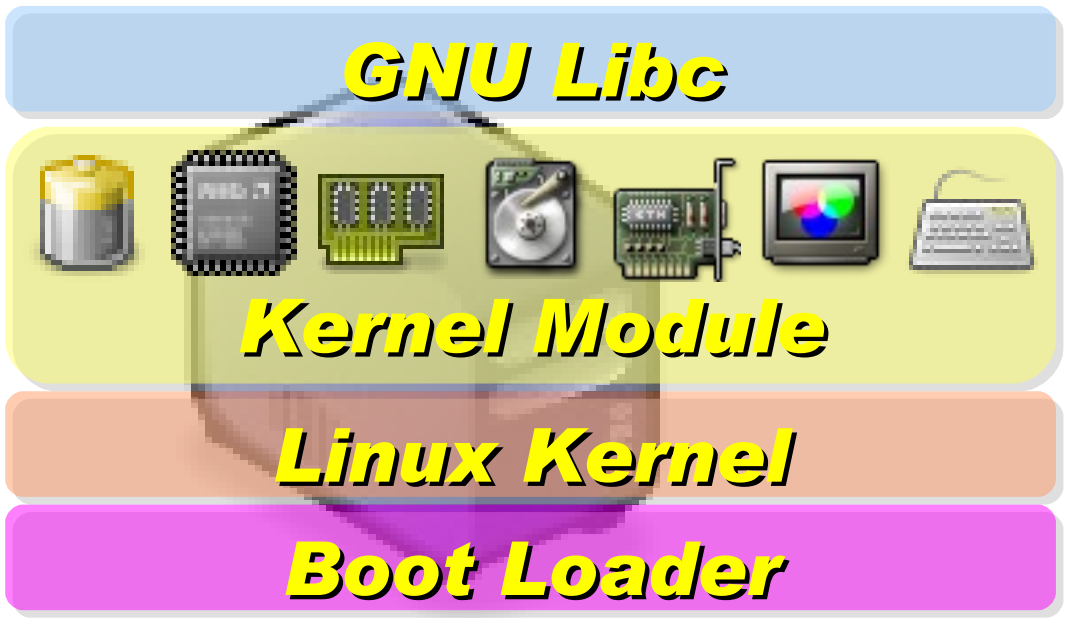
Kernel Module

Linux Kernel

Boot Loader

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

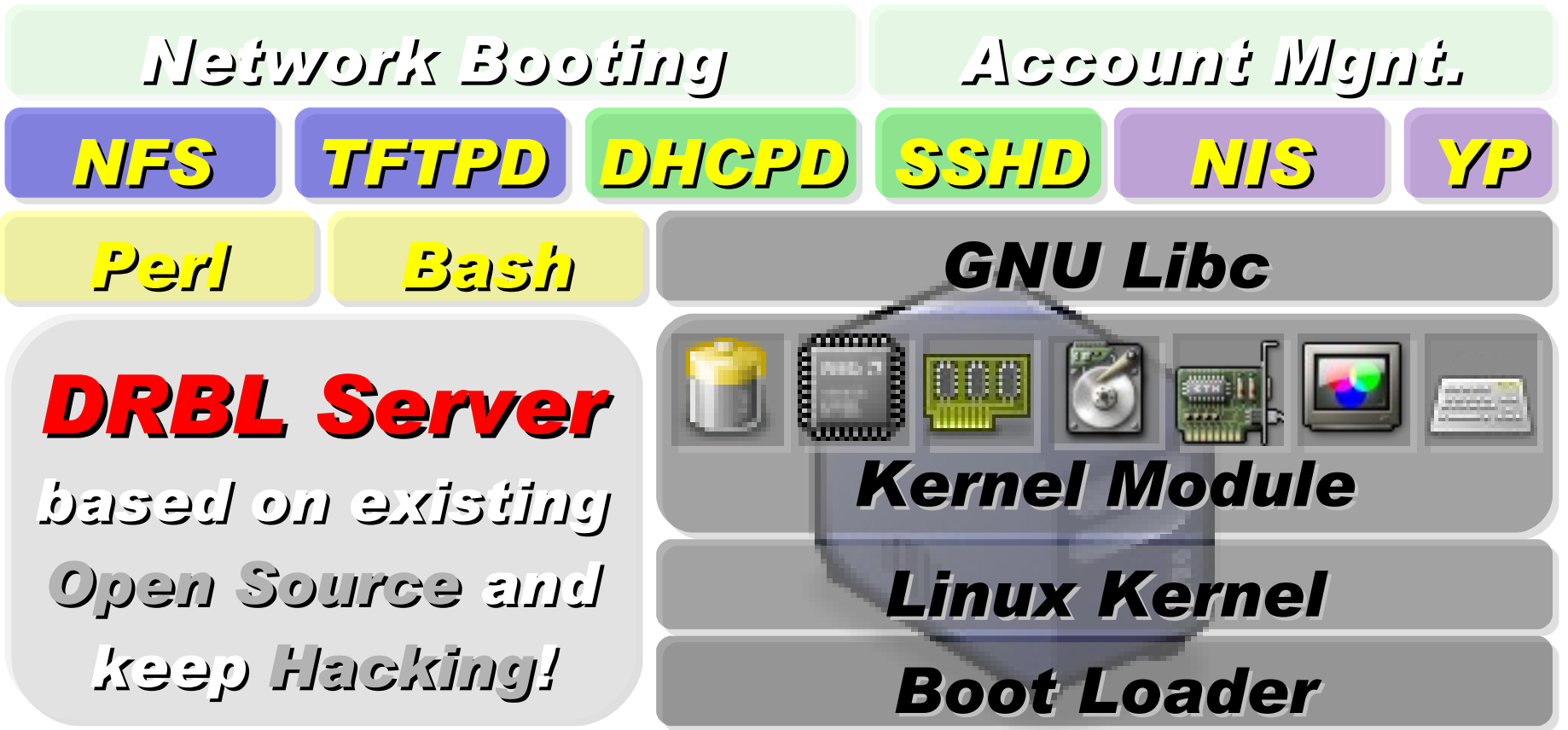
Redhat, Fedora, CentOS, Mandriva, Ubuntu, Debian, ...



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

There are lots of service needed:

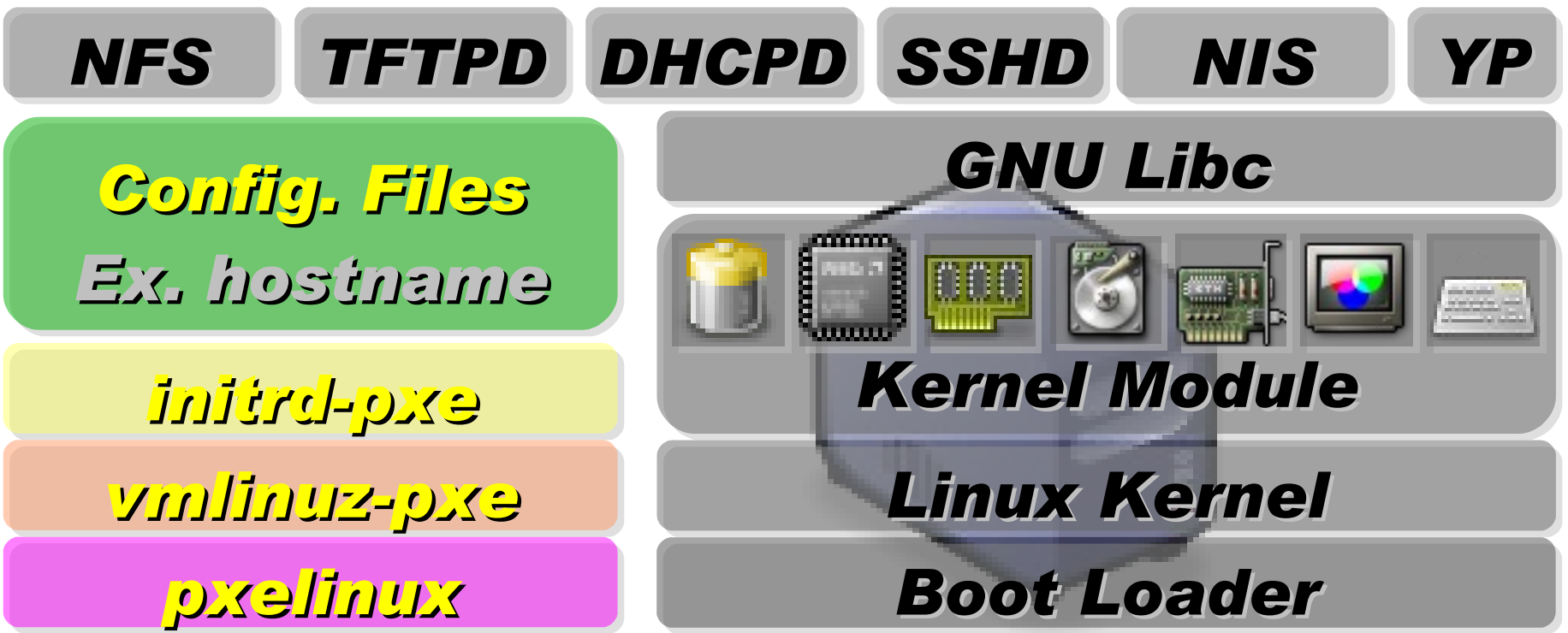
SSHD, DHCPD, TFTPD, NFS Server,
NIS Server, YP Server ...



DRBL Server

based on existing Open Source and keep Hacking!

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



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

GNU Libc

Kernel Module

initrd-pxe

vmlinuz-pxe

Linux Kernel

pxelinux

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

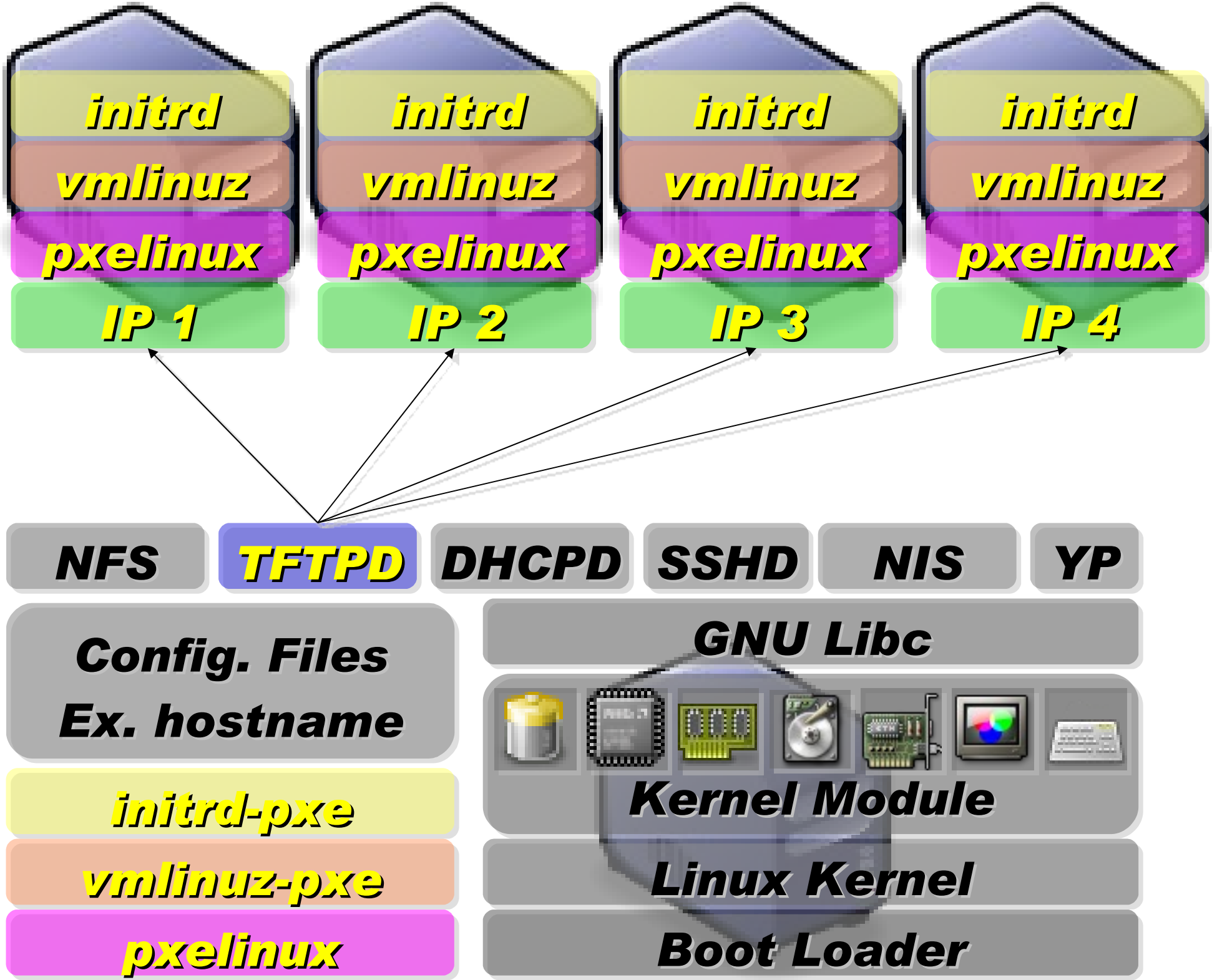
pxelinux

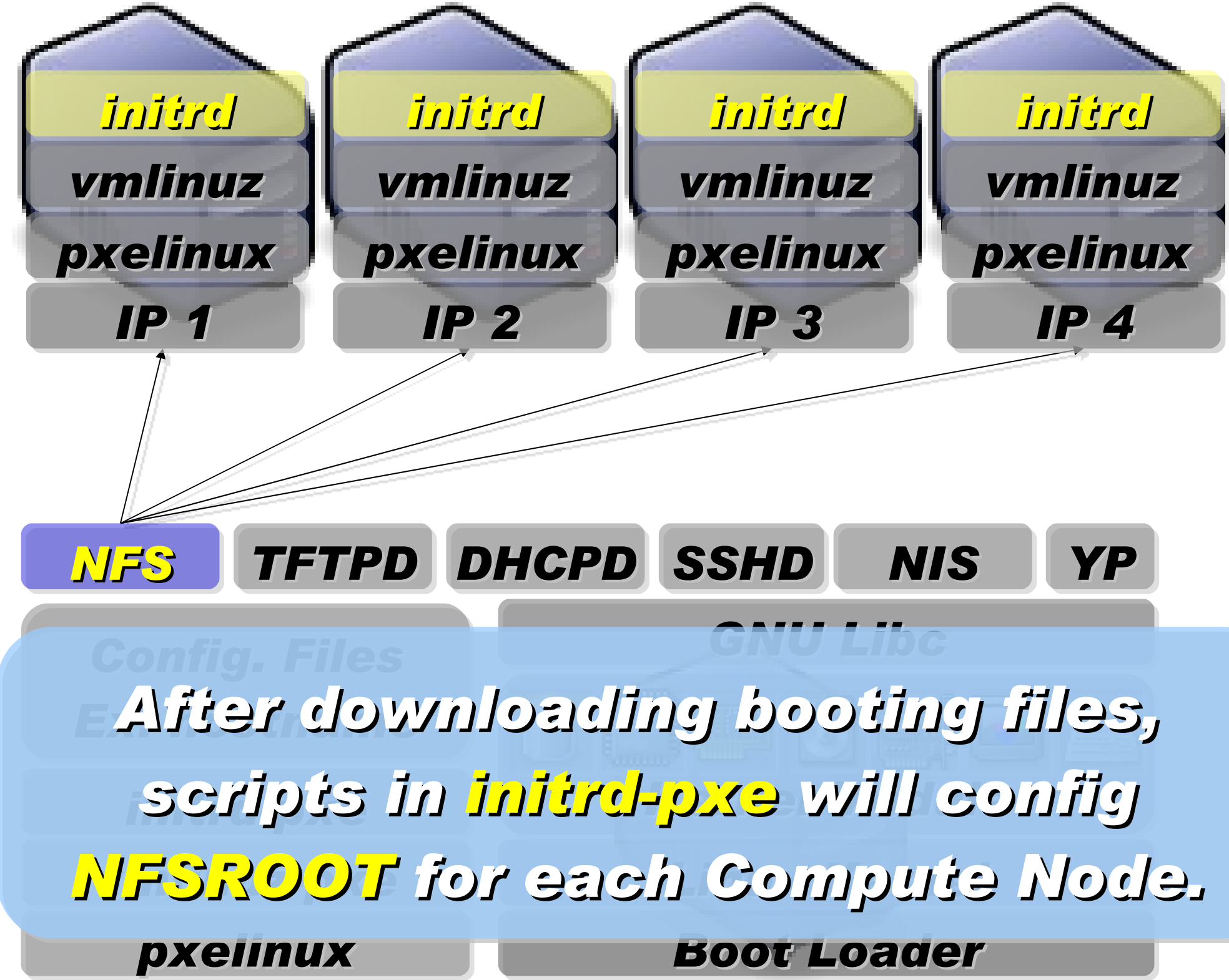
GNU Libc

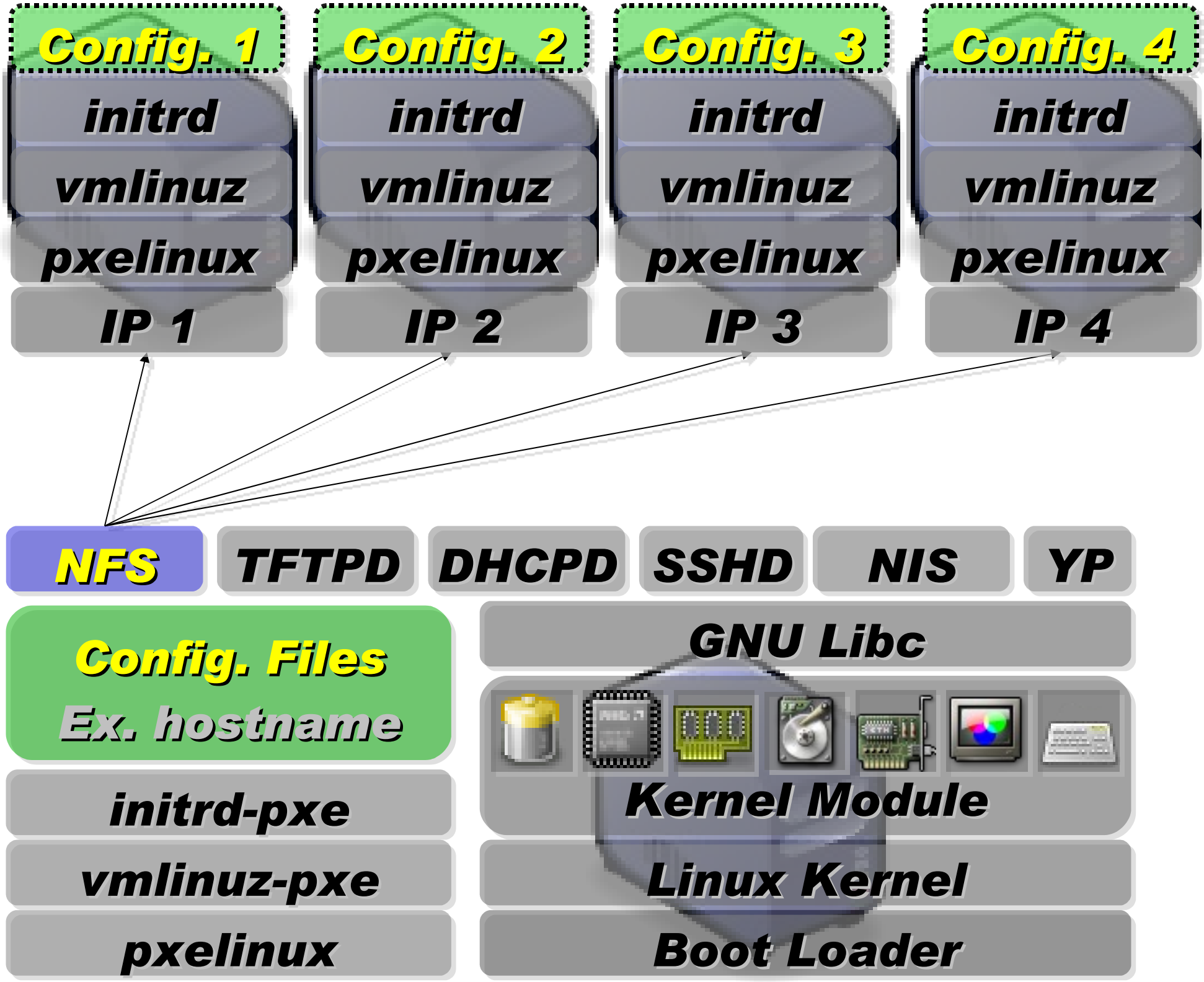
Kernel Module

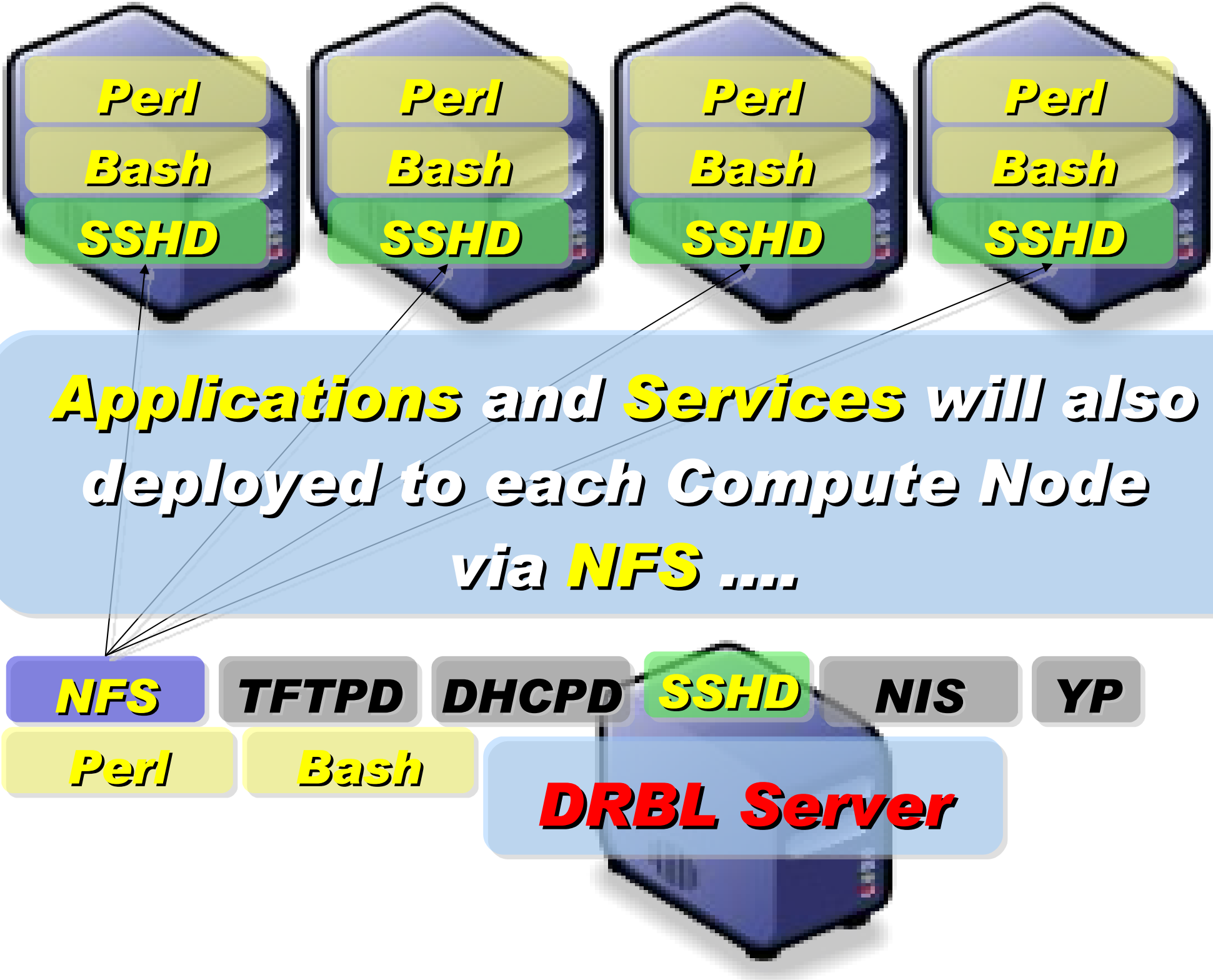
Linux Kernel

Boot Loader


initrd
vmlinuz
pxelinux
IP 1
initrd
vmlinuz
pxelinux
IP 2
initrd
vmlinuz
pxelinux
IP 3
initrd
vmlinuz
pxelinux
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


initrd
vmlinuz
pxelinux
IP 1
initrd
vmlinuz
pxelinux
IP 2
initrd
vmlinuz
pxelinux
IP 3
initrd
vmlinuz
pxelinux
IP 4
NFS
TFTPD
DHCPD
SSHD
NIS
YP
Config. Files
GNU Libc
pxelinux
Boot Loader
After downloading booting files, scripts in initrd-pxe will config NFSROOT for each Compute Node.


Config. 1
Config. 2
Config. 3
Config. 4
initrd
vmlinuz
pxelinux
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


Perl
Bash
SSHD
Perl
Bash
SSHD
Perl
Bash
SSHD
Perl
Bash
SSHD
Applications and Services will also deployed to each Compute Node via NFS ....
NFS
TFTPD
DHCPD
SSHD
NIS
YP
Perl
Bash
DRBL Server


SSHD
SSHD
SSHD
SSHD
With the help of NIS and YP,
You can login each Compute Node
with the Same ID / PASSWORD
stored in DRBL Server!
SSH Client
NFS
TFTPD
DHCPD
SSHD
NIS
YP
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 Neqotiation, Ddmission Control, Pricing, SLA Management, Metering…

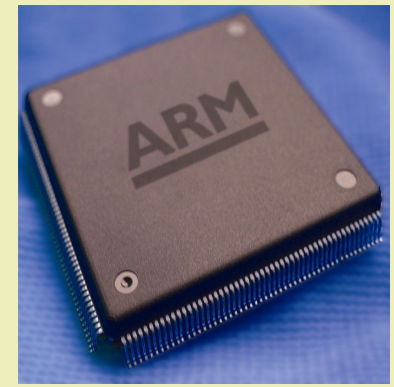
OpenNebula, Enomaly, Eucalyptus , OpenQRM, ...

Virtualization
VM, VM management and Deployment

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

Virtualization ?? Emulator ??

Virtual Hardware / OS



mame4iphone

Mac4Lin

QEMU

Microsoft

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

Source: http://en.wikipedia.org/wiki/Virtualization

Open Cloud #1: Eucalyptus

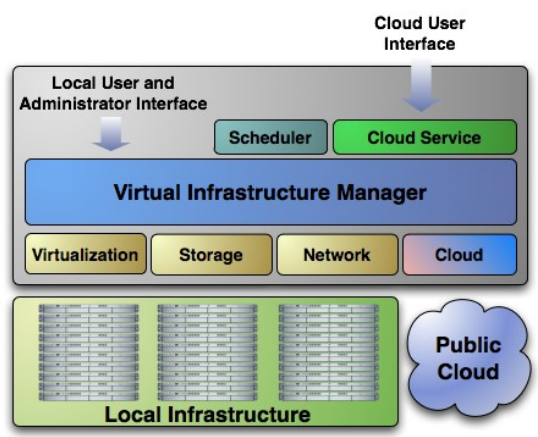
- http://open.eucalyptus.com/
- It was a research project of UCSB, USA
- Now Eucalyptus System provide technical supports.
- It designed to help user to build their own Amazon EC2
- Its feature is compatible with existing EC2 client.
- Ubuntu Enterprise Cloud powered by Eucalyptus in 9.04
- You can register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

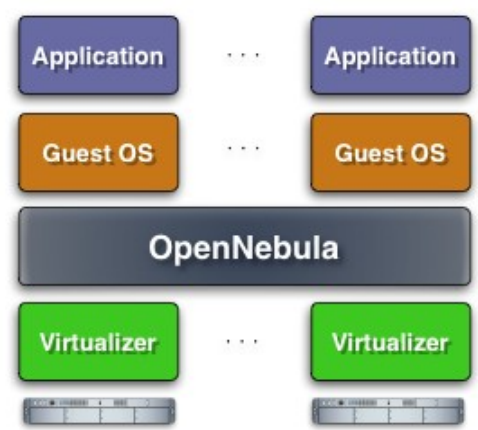
Open Cloud #2: OpenNebula

- 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

SEVENTH FRAMEWORK PROGRAMME

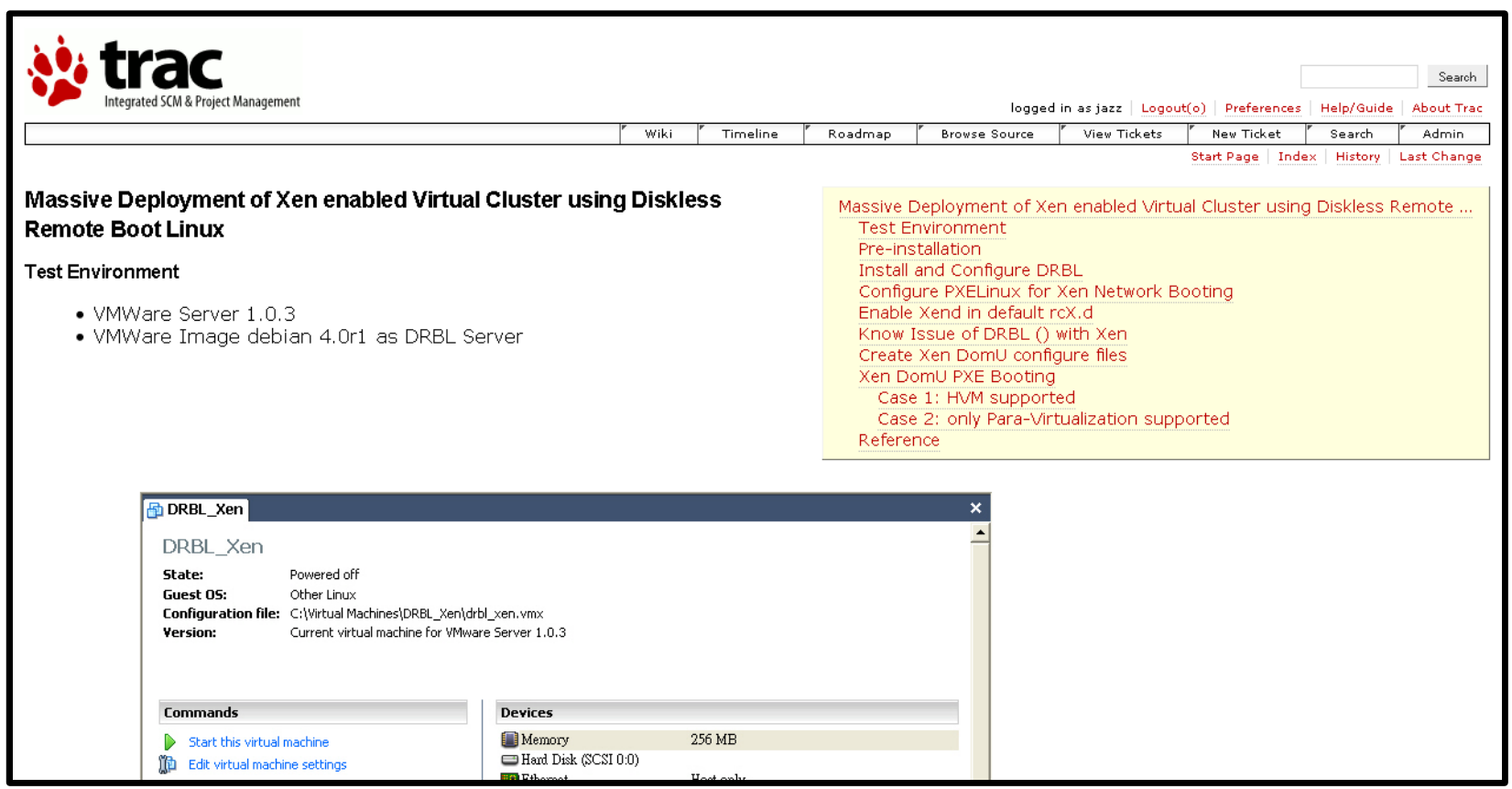
RESERVOIR





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



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 Neqotiation, Ddmission Control, Pricing, SLA Management, Metering…

OpenNebula, Enomaly, Eucalyptus , OpenQRM, ...

Virtualization
VM, VM management and Deployment

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

Open Cloud #3: Hadoop

- 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

Open Cloud #4: Sector / Sphere

- 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

National Center for Data Mining
University of Illinois at Chicago

UIC

open data

Open Data Group
http://www.opendatagroup.com/

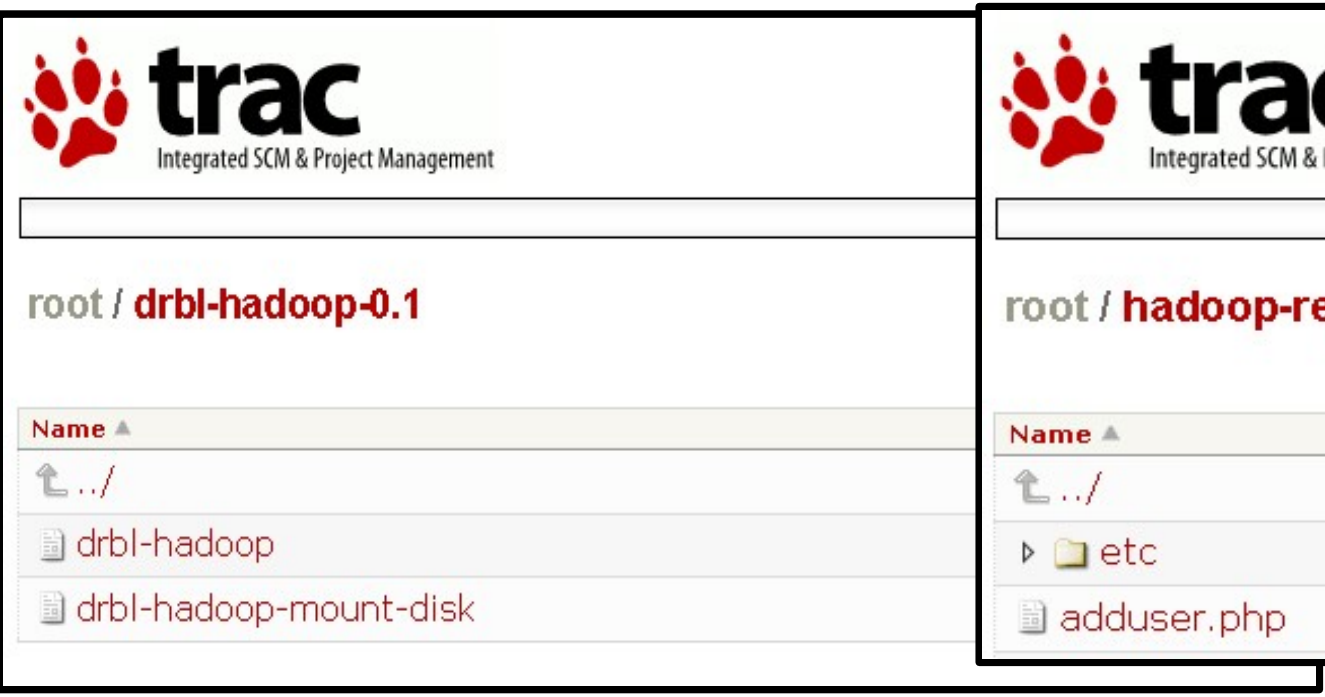
Building PaaS using DRBL-Hadoop

- **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
```

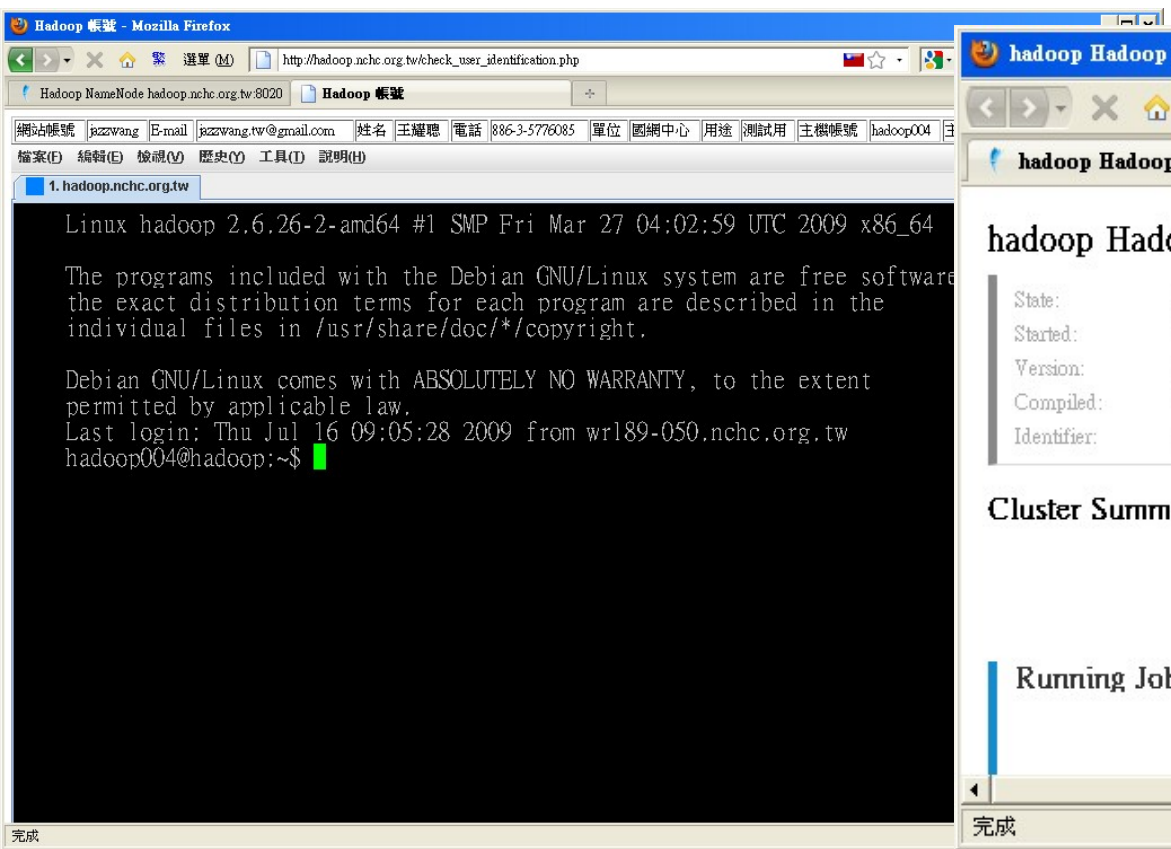
- **hadoop-register – web interface with ssh applet**

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



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.**

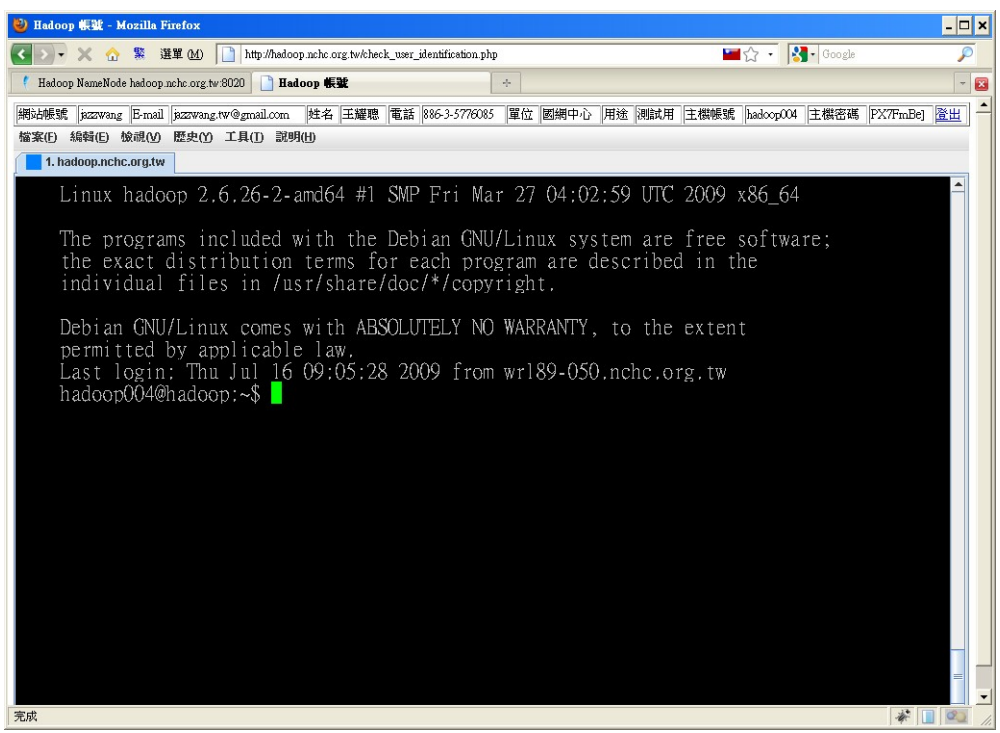


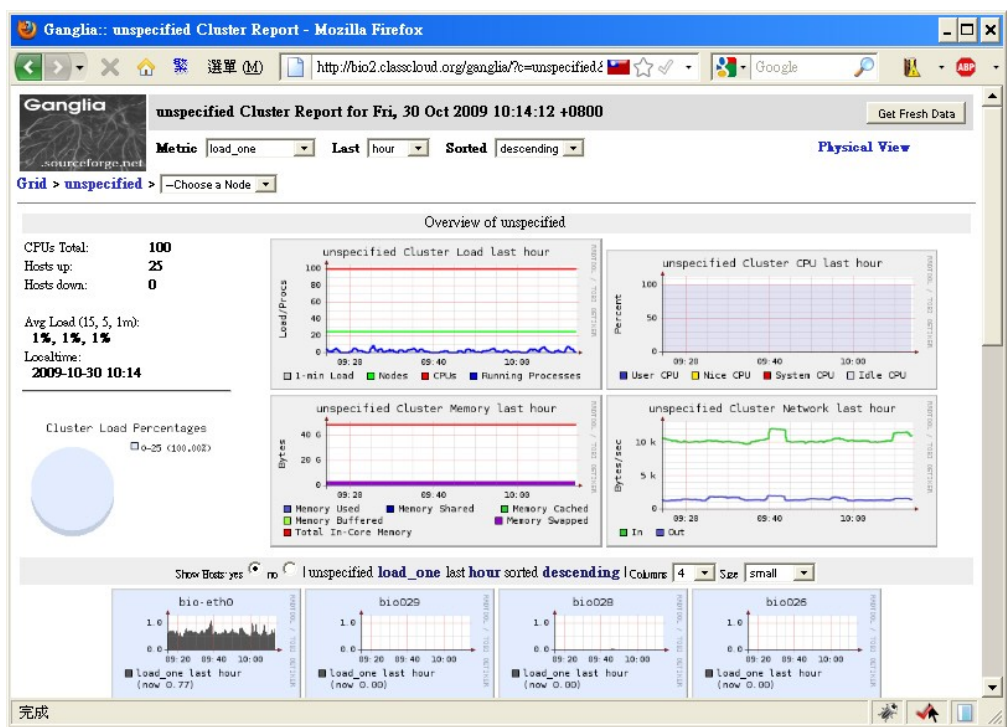
Building SaaS using DRBL-biocluster

- **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、BioCondoctor、Ganglia、Nagios、AutoFACT、BLAST、SIM4、Clustal、PipMaker、Phylip、Eland、Velvet、Bowtie、SOAP**





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.

Questions?

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

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

Powered by DRBL