叢集與格網計算環境管理
以 DRBL 爲例

蕭志棋 黃國連 柯森 林芳邦

National Center for High-performance Computing, Taiwan

http://drbl.nchc.org.tw, http://drbl.sf.net

Sep 2004

Outline

- Introduction to Diskless Remote Boot in Linux (DRBL)
- Embedded system - DRBL-based mobile sensors
- Cluster computing - scalable cluster management
- Education - smart classroom
- Grid environment - DRBL-G
- Q&A

Outline

- **Introduction to Diskless Remote Boot in Linux (DRBL)**
- **Embedded system - DRBL-based mobile sensors**
- **Cluster computing - scalable cluster management**
- **Education - smart classroom**
- **Grid environment - DRBL-G**
- **Q&A**

DRBL - Diskless Remote Boot in Linux

- Developed by NCHC open source task force, based on
  - Network boot mechanism 
    - PXE (Preboot Execution Environment)
    - Etherboot
  - NFS file system
  - Other management programs
- Major Features
  - Virtually centralized management
  - Integrate distributed system, either embedded or cluster, into a virtually single system

Diskfull, diskless or systemless

- **diskfull - client nodes have dedicated disks**
- **diskless - client nodes have no disks**
- **systemless - client nodes have dedicated disks, but they don't contain a disk bootloader and they are boot from the network, disk is for swap, tmp**
- **Why diskless ?**
  - **It's easier to manage one image than many individual installations**
  - **save budget**
  - **a disk is a mechanical part that is subject to failure**
  - **lesser mechanical parts, greater reliability**

Ref: http://www.linuxworld.com/story/43718_p.htm

Diskfull, diskless or systemless

- **Using diskless, systemless or diskfull for clients ?**
  - **depends on applications**
- **DRBL provides diskless and systemless mode for clients, we also have a program called "Clonezilla" to deploy a diskfull system**
- **Clonezilla has been used in many schools to clone the M$ Windows or Linux system image**

Development and tuning in DRBL

Package Installation Live CD

NIS, SSH, NTP...

initrd, busybox pcitable

init, booting, runlevel

kernel space

DEVFS

NBD

KNFSD

TMPFS

- Live CD is available in http://drbl.nchc.org.tw

Schematic figure for DRBL

server

switch

client nodes

		pxe/etherboot	pxe/etherboot
DHCP	IP →	192.168.0.1	192.168.0.40
TFTP	kernel →	boot	boot
NFS	file system →	/, /usr, /home ...	/, /usr, /home ...
NIS	account →	user login	user login

Adopted areas

- Embedded system
  - DRBL-based mobile sensors
- Cluster computing
  - Scalable cluster management
- Education
  - Smart classroom
- Grid environment
  - DRBL-G

Hierarchical integration

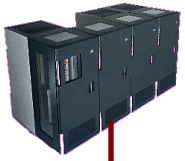
Fu Shan

SMPS

三民國小

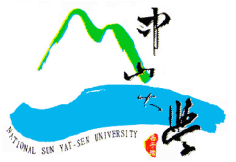
DRBL-G

Storage/Data



NCTU

NSYSU



Computers
- DRBL

THU

Data logger

(CR10X,,campbell)

Taiwan UniGrid

NCHC

Network Backbone

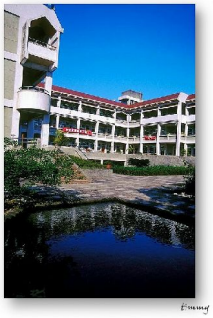
Embedded system

DFES

Control system

Domain Knowledge Center

LTES



Observation Station-DRBL

End Users

Cluster computing

Computers
- DRBL

Tele-robotics - DRBL

Outline

- Introduction to Diskless Remote Boot in Linux (DRBL)
- Embedded system - DRBL-based mobile sensors
- Cluster computing - scalable cluster management
- Education - smart classroom
- Grid environment - DRBL-G
- Q&A

Embedded system - DRBL-based mobile sensors

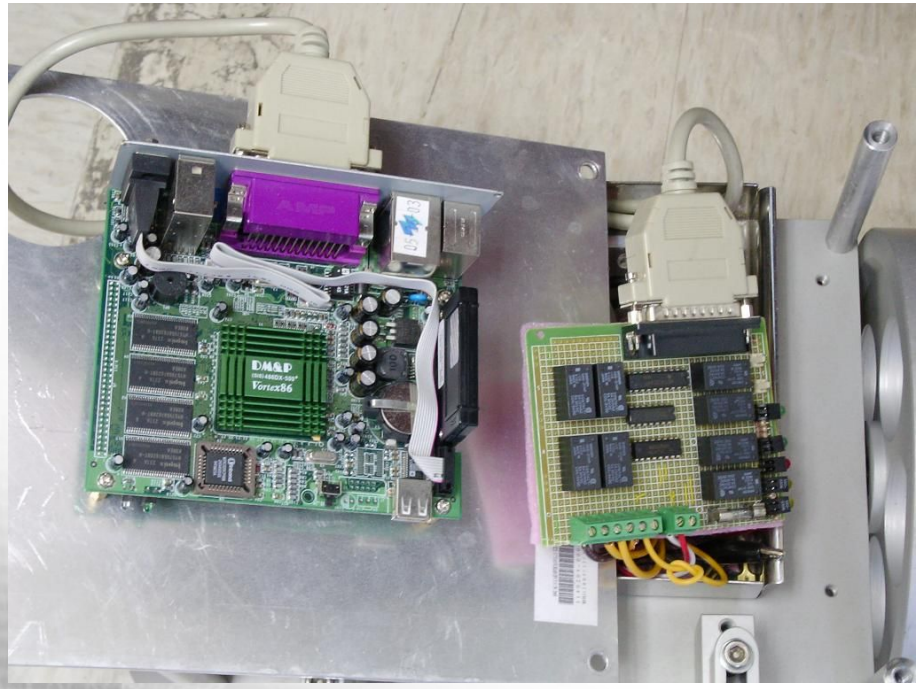
From sensors to robot

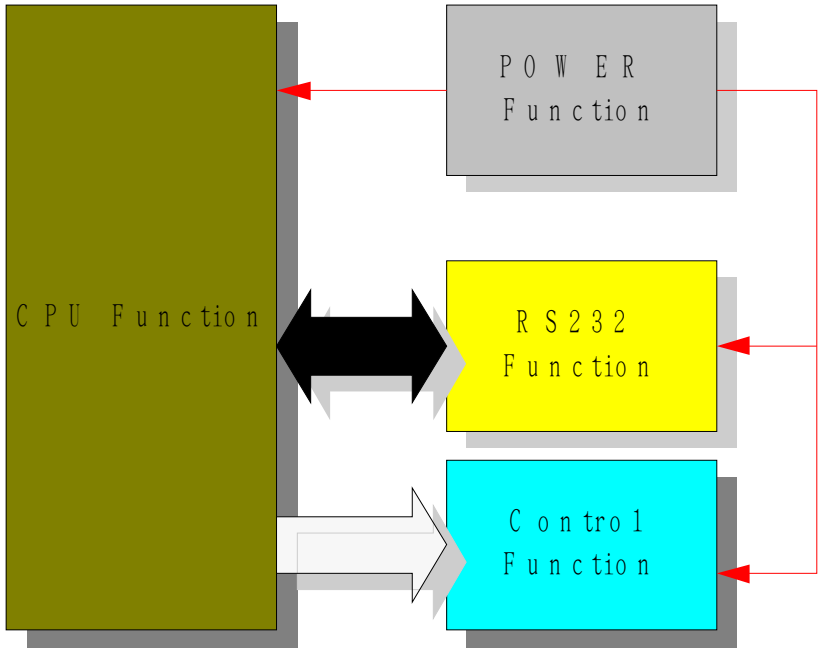
Eco-grid

Agriculture-grid

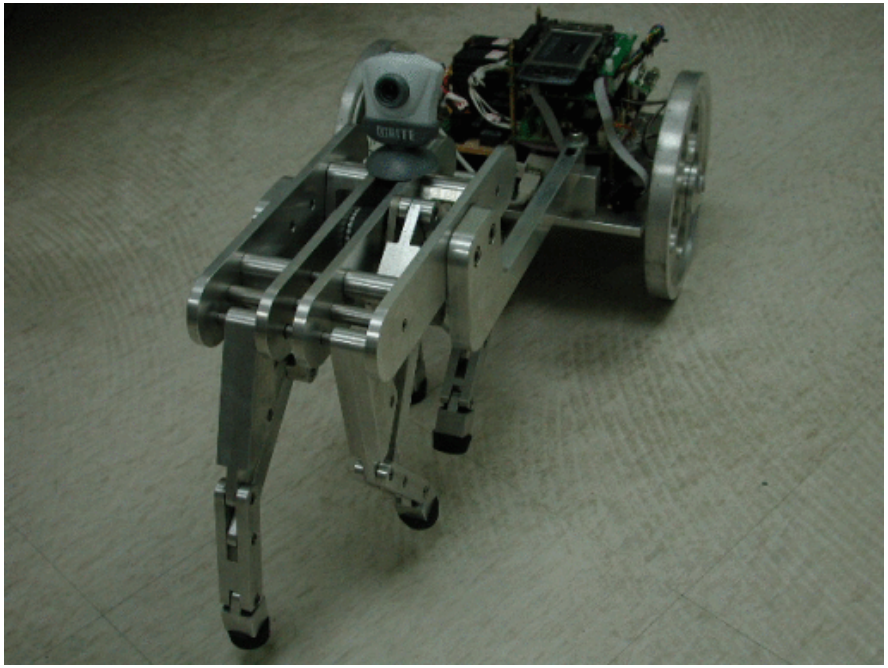
Wireless Robot Net

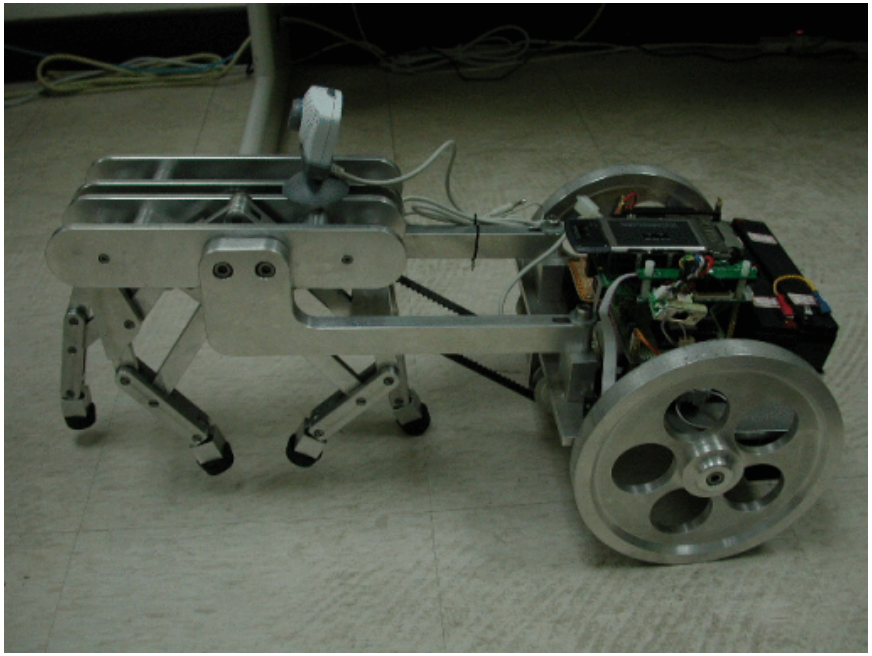
- **Basic idea: intelligent and mobile sensors using Grids.**
- **Ex. IP-driven Horse Robot**
  - **Redesign the electronic control panel: PC104 w/ RS232, IDE, Ethernet and USB**
  - **Use both IDE HD and/or DRBL system setup**
  - **RS232 & Ethernet for the IP-driven, IPv4 or IPv6, wireless robot control**
  - **USB for webcam**

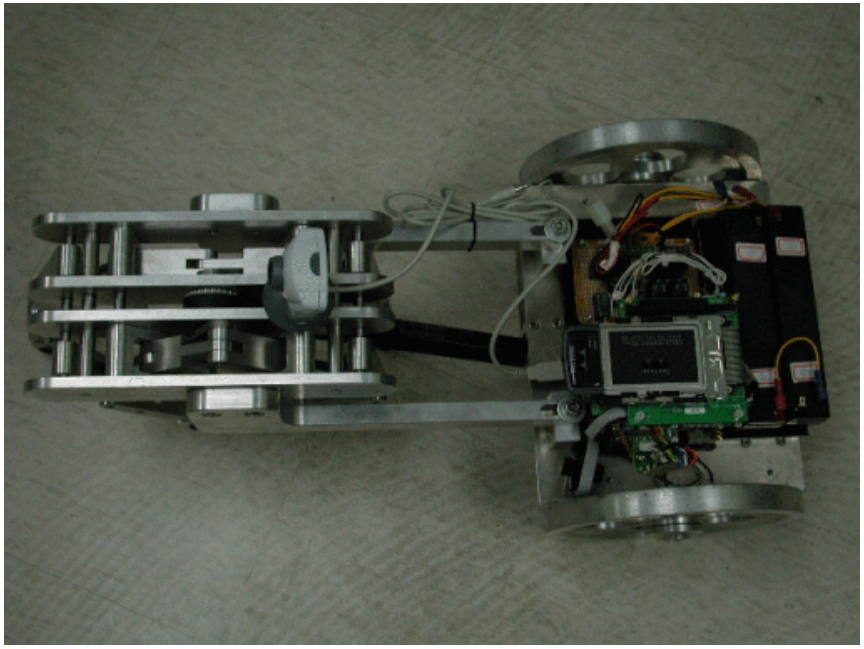


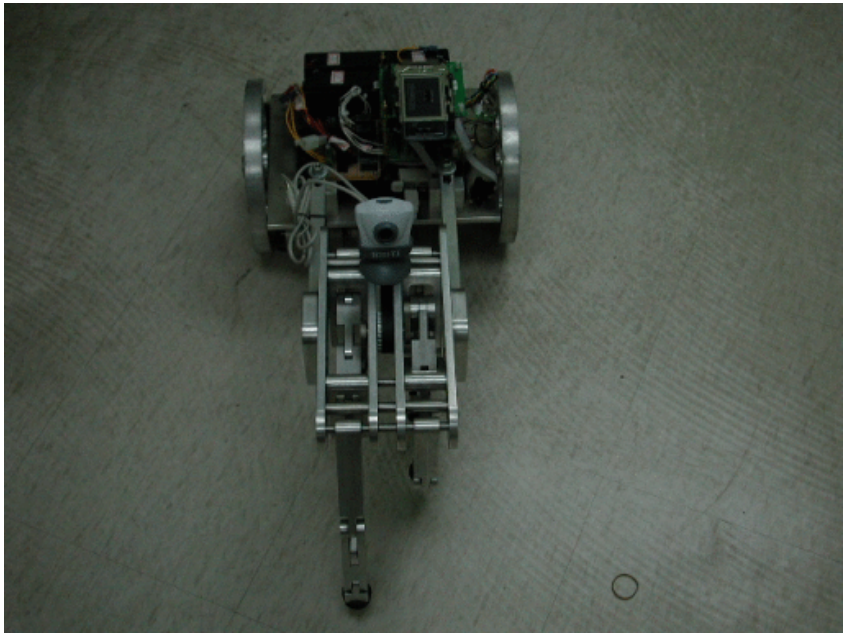


- **Vortex86 200MHz SoC**
- **128MB SDRAM**
- **Setup in DRBL, easy to deploy different systems**









*movie

Outline

- Introduction to Diskless Remote Boot in Linux (DRBL)
- Embedded system - DRBL-based mobile sensors
- Cluster computing - scalable cluster management
- Education - smart classroom
- Grid environment - DRBL-G
- Q&A

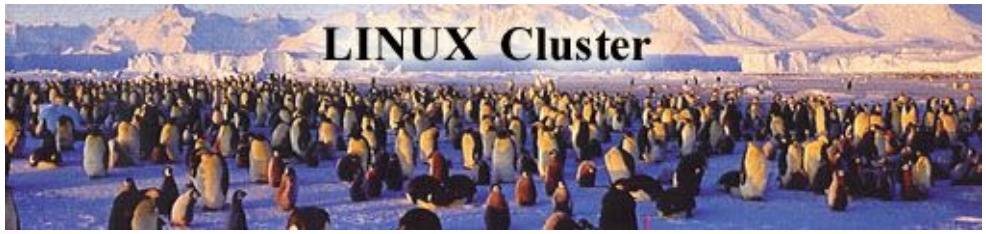
DRBL in cluster computing

- The advantages come with diskless/systemless environment – management, budget, reliability
- Manage one system image only
  - homogeneous system
- Client machines can be workstations at daytime, and become cluster computing nodes at night
  - These nodes can be very quickly integrated into a cluster without any alteration of the main OS stored on their disks
- Almost "zero time (effortless)" installation for the OS and application program in client nodes

NPACI Rocks & OSCAR

- NPACI Rocks
- An Open Source High Performance Linux Cluster Solution
- http://rocks.npaci.edu

- OSCAR – Open Source Cluster Applications Resources
- http://oscar.sourceforge.net/



DRBL, Rocks & OSCAR

	DRBL	Rocks	OSCAR
diskless/systemless	Y	N	N (*1)
diskfull	Y(*2)	Y	Y
cluster packages	Y(*3)	Y	Y
scale up	good(*4)	excellent	excellent

1. **A project called "Thin-OSCAR" provides the diskless/systemless OSCAR cluster**
2. **Using Clonezilla, DRBL can deploy diskfull nodes**
3. **Packages are same with those in Rocks and OSCAR, but some are not included**
4. **From good to excellent**
   - **Good: NFSRoot, however many-to-1 problem in large scale**
   - **Excellent: (1) ramdisk and multicast model or (2) hierarchical network topology , the scale can be excellent**

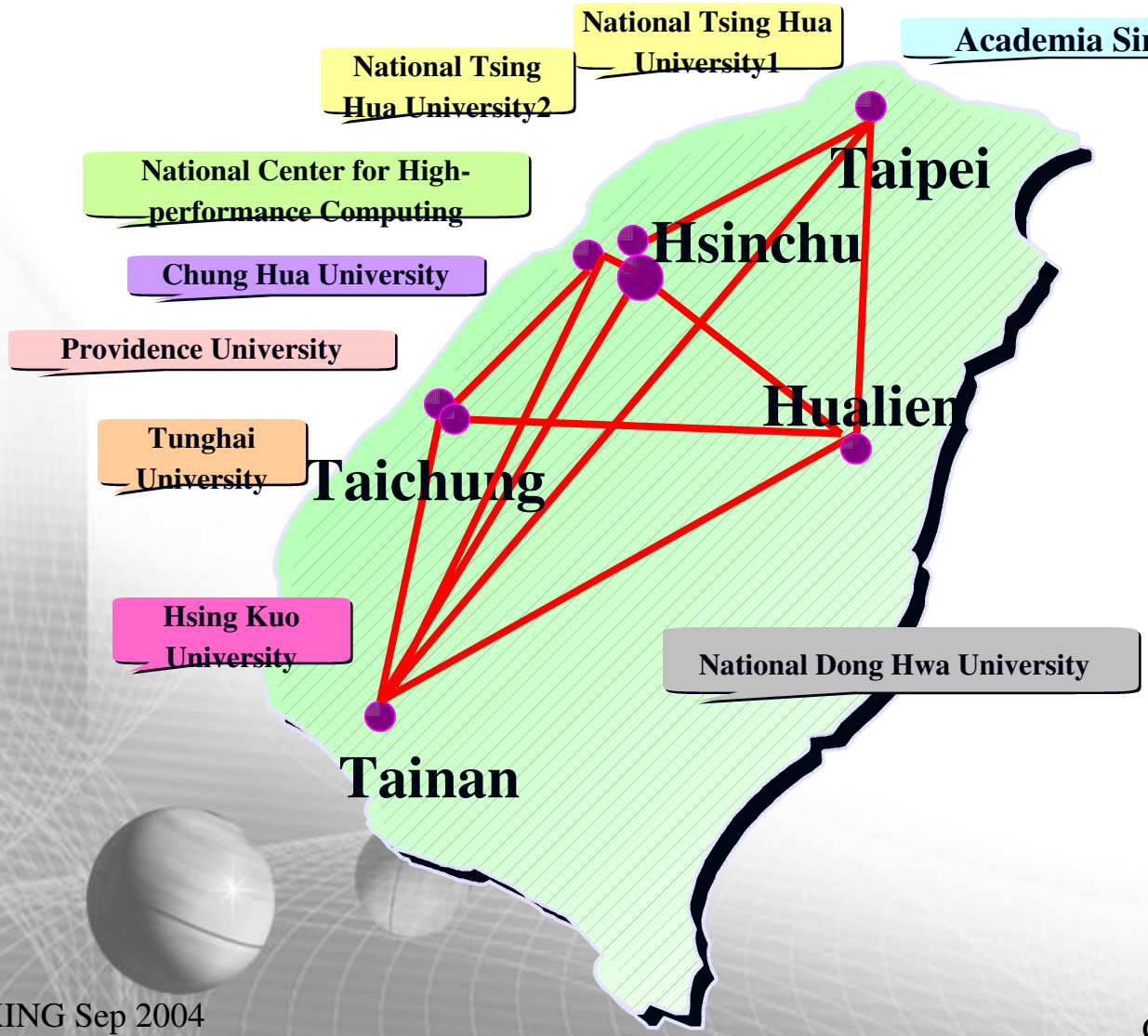
PC clusters using DRBL in NCHC

- Two PC clusters are deployed with DRBL
  - ASE cluster : 8 + 1 nodes
  - GT3 cluster : 4 + 1 nodes
- Deploying DRBL to large scale cluster computing
  - Formosa I diskfull approach by clonezilla
  - DRBL extention to Formosa I and Taiwan Unigrid

NCHC PC Cluster - Formosa I

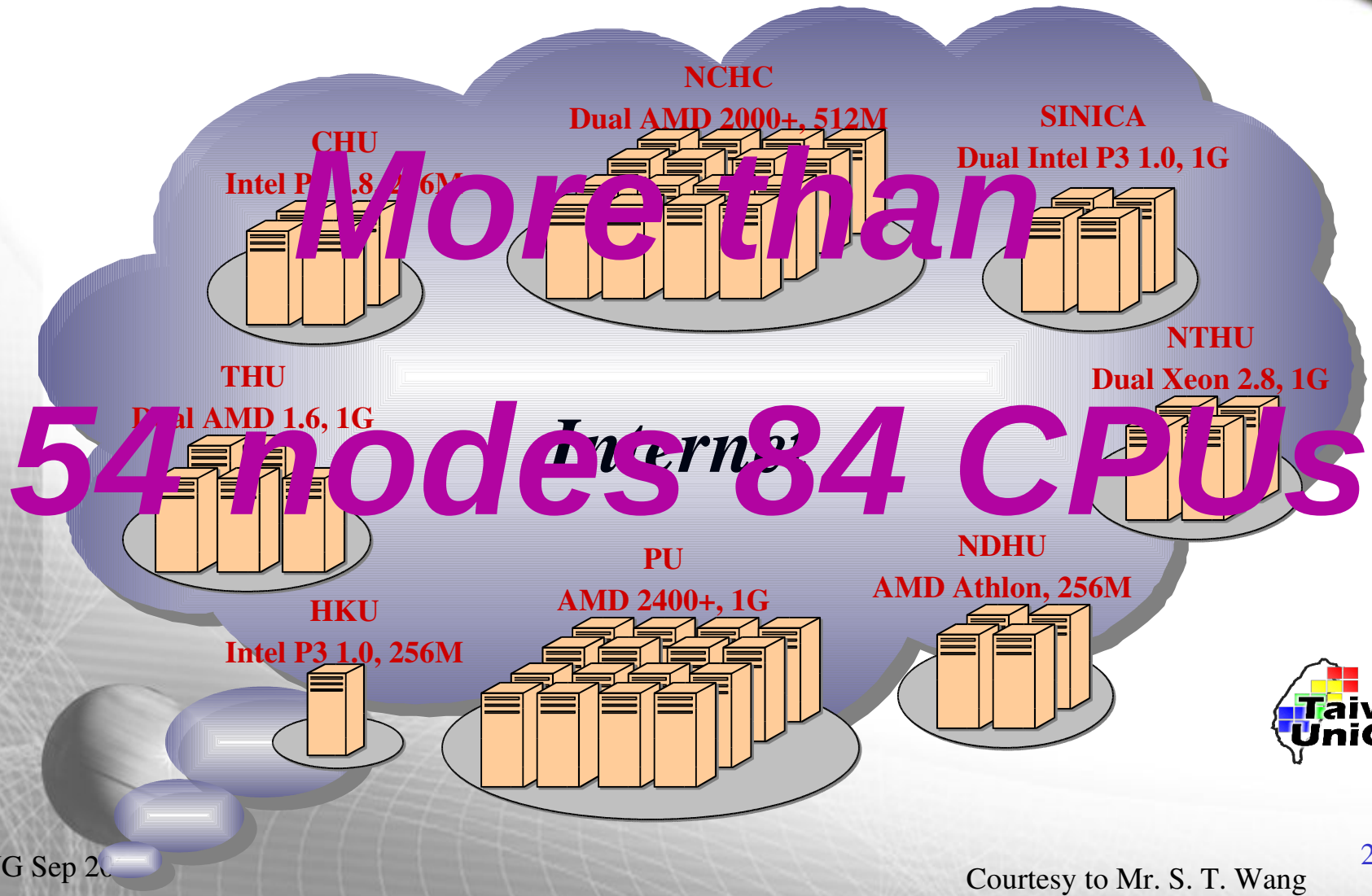
- Built in Oct 2003
- Rank 242 in Top500 on Jun/2004
- CPUs : 300 (2 CPUs X 150 nodes)
- Memory : 384GB (4 GB X 32 + 2 GB X 128)
- Gigabits swtich & NICs
- R_{peak} : 1.68 TFLOPS
- R_{max} : 1.002 TFLOPS

Taiwan UniGrid Topology



Courtesy to Mr. S. T. Wang

Hardware Infrastructure

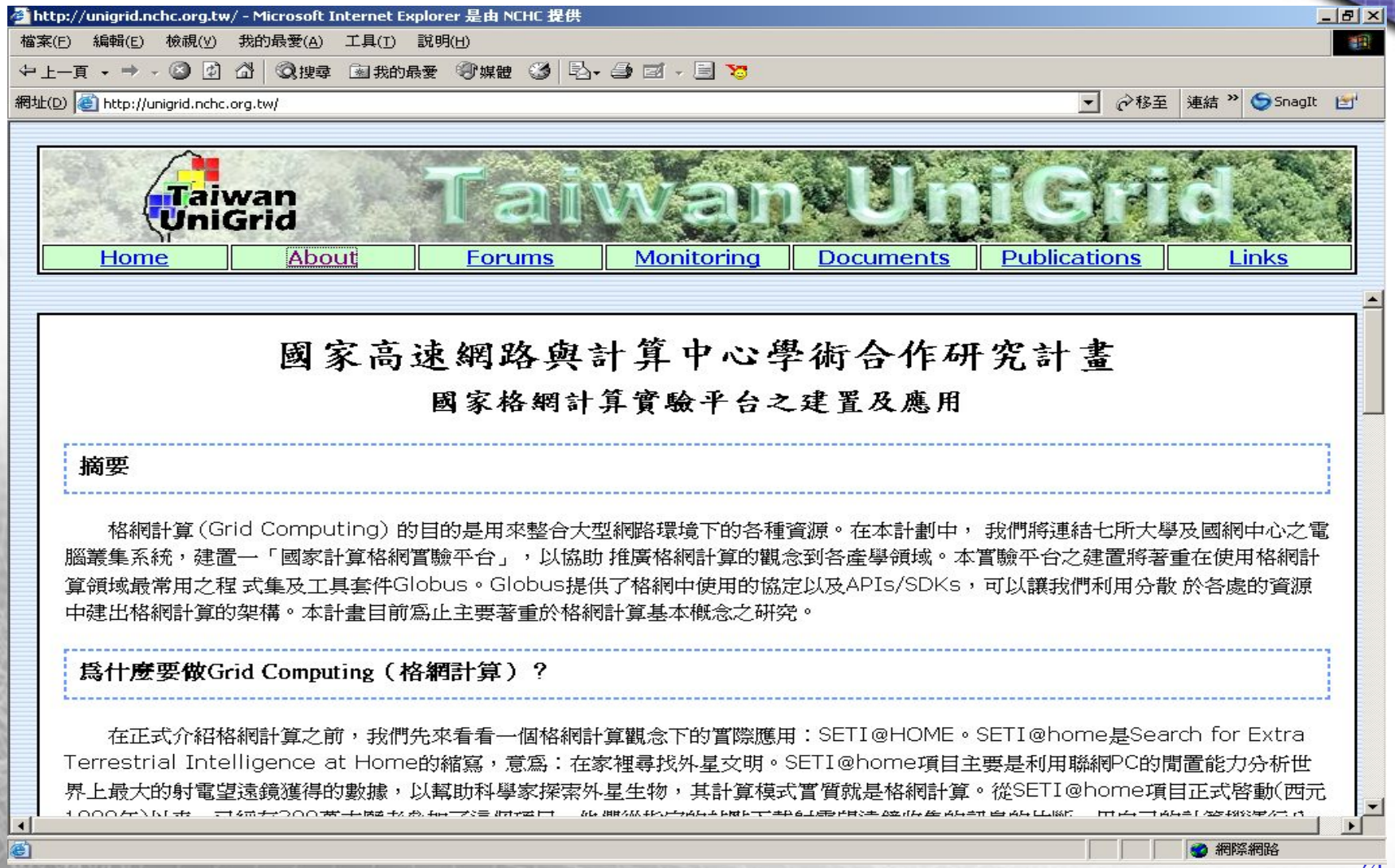


Taiwan UniGrid

Courtesy to Mr. S. T. Wang

Unigrid Project Webpage

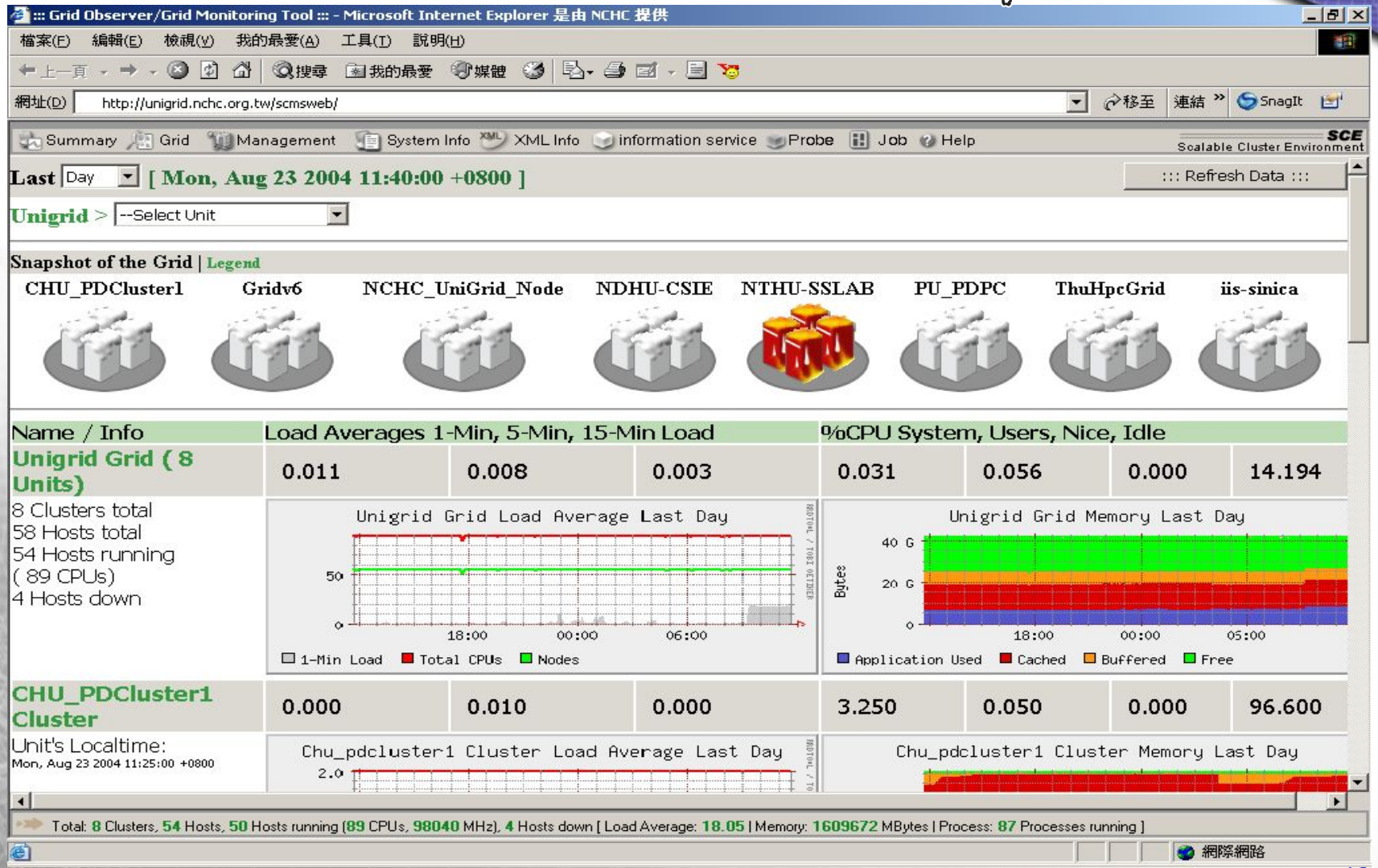
- **http://unigrid.org.nchc.tw/**



Courtesy to Mr. S. T. Wang

System Monitoring Webpage

- http://unigrid.nchc.org.tw/scmswe



Courtesy to Mr. S. T. Wang

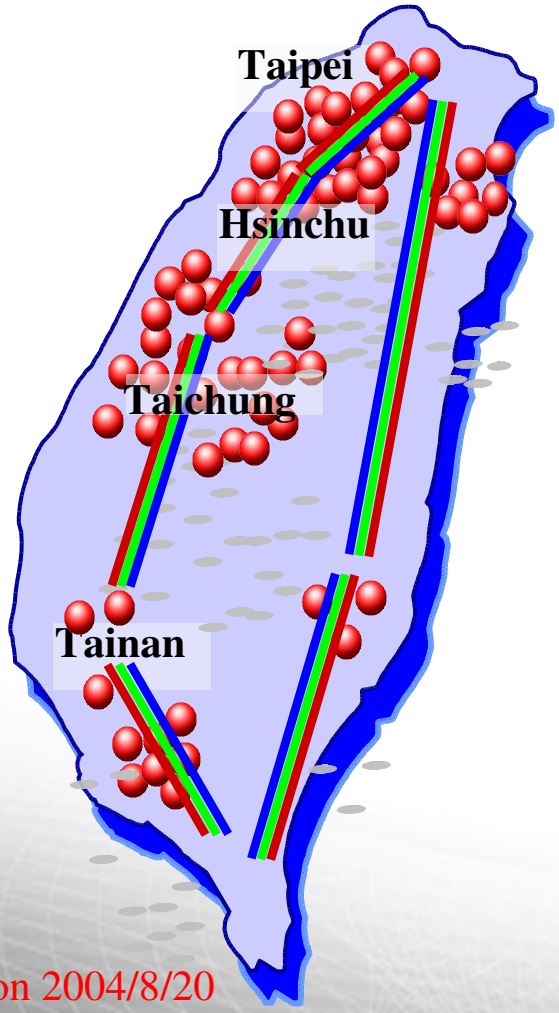
Outline

- Introduction to Diskless Remote Boot in Linux (DRBL)
- Embedded system - DRBL-based mobile sensors
- Cluster computing - scalable cluster management
- Education - smart classroom
- Grid environment - DRBL-G
- Q&A

A different kind of light-weight grid computing – computer classroom and office

> 100 sites, > 4000 PCs

- **DRBL users:**
  - **Public sector**
    - **Primary/high school 55**
    - **University/college 17**
    - **Hospitals 1**
    - **TV stations 1**
    - **Governments 3**
    - **NPO 5**
  - **Private sector**
    - **Internet and business companies 35**



*users number on 2004/8/20

Outline

- Introduction to Diskless Remote Boot in Linux (DRBL)
- Embedded system - DRBL-based mobile sensors
- Cluster computing - scalable cluster management
- Education - smart classroom
- Grid environment - DRBL-G
- Q&A

DRBL-G

Implemented by

Chien-Lin Huang, Gary Wu, Julian Yu-Chung Chen, Weicheng Huang

NCHC

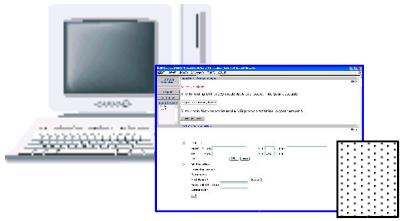
DRBL-G

- Use DRBL to setup grid-enabled clusters
  - Two testing PC clusters are grid-enabled with DRBL
    - ASE cluster : 8 + 1 nodes
    - GT3 cluster : 4 + 1 nodes
- DRBL-G environment
  - Use globus + shell script + DRBL to integrate the infrastructure
  - Web portal

DRBL-G

- A web portal is designed to
  - Describe the workflow
  - Monitor the job execution
  - Generate the script based on the workflow description
- Use DRBL to provide a pool of grid-enabled machines
  - DRBL Server
    - Load sharing between DRBL clients
    - Provide a 'sandbox' environment for DRBL clients
      - no sensitive data
      - only /tmp have read-write permission
  - DRBL Clients
    - A pool of grid-enabled machines
    - Take care of grid proxy generation and the script execution

DRBL-G



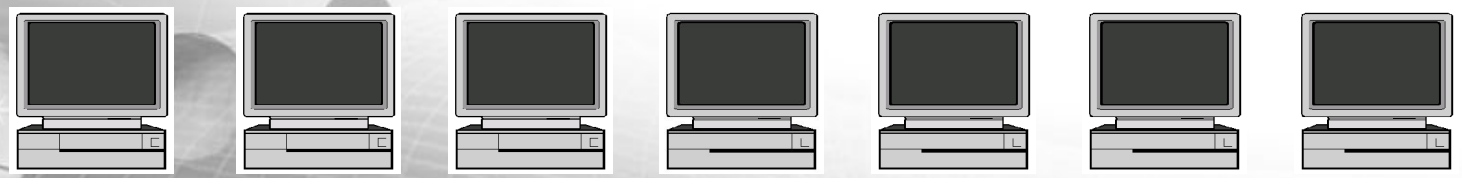
Describe the workflow and generate the script

Monitor the job execution

Move/spawn the script to/on a special host

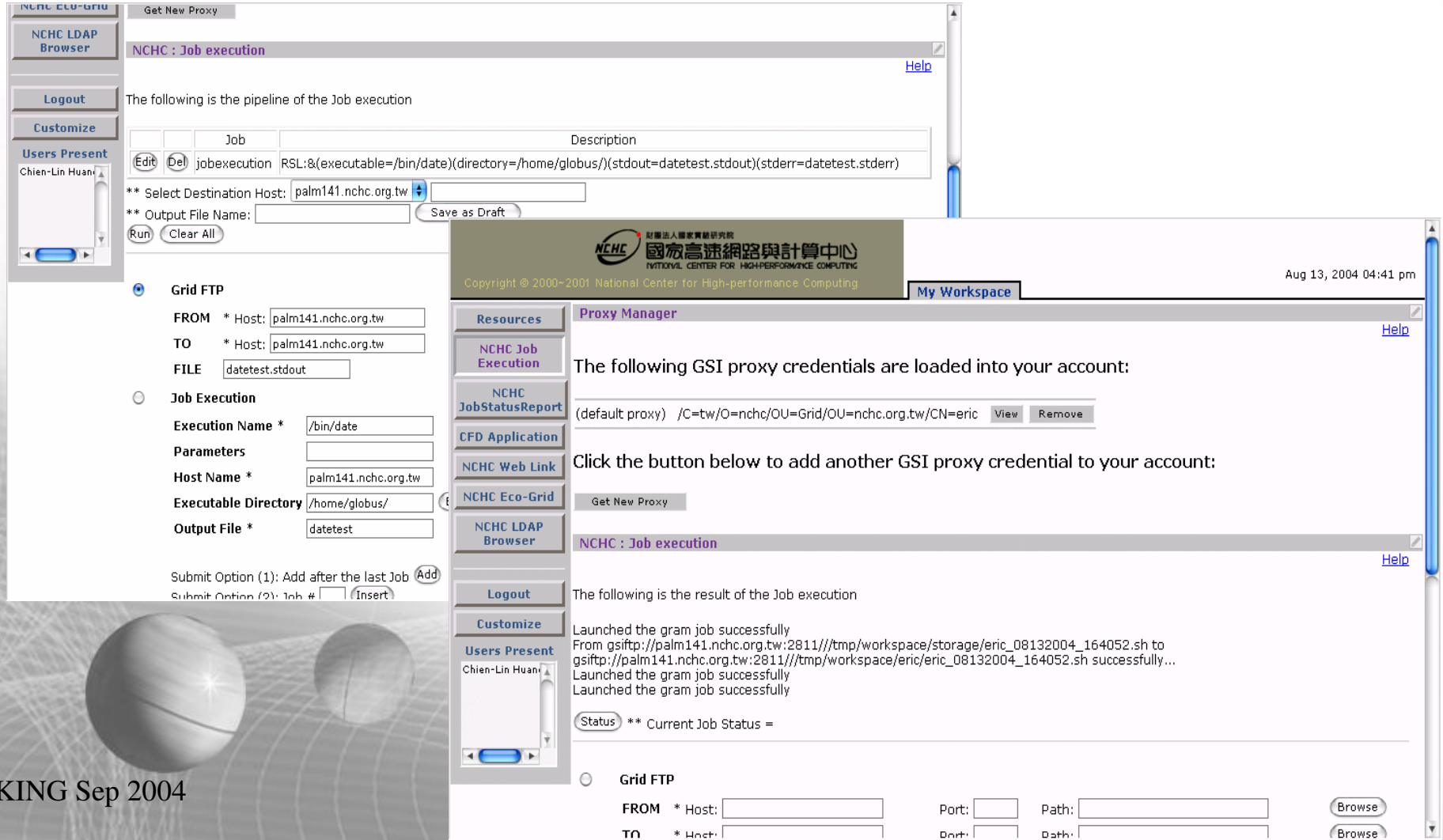
A special host that takes care of grid-proxy generated and executes the script

The script uses globus toolkit to execute jobs on all globus enabled machines



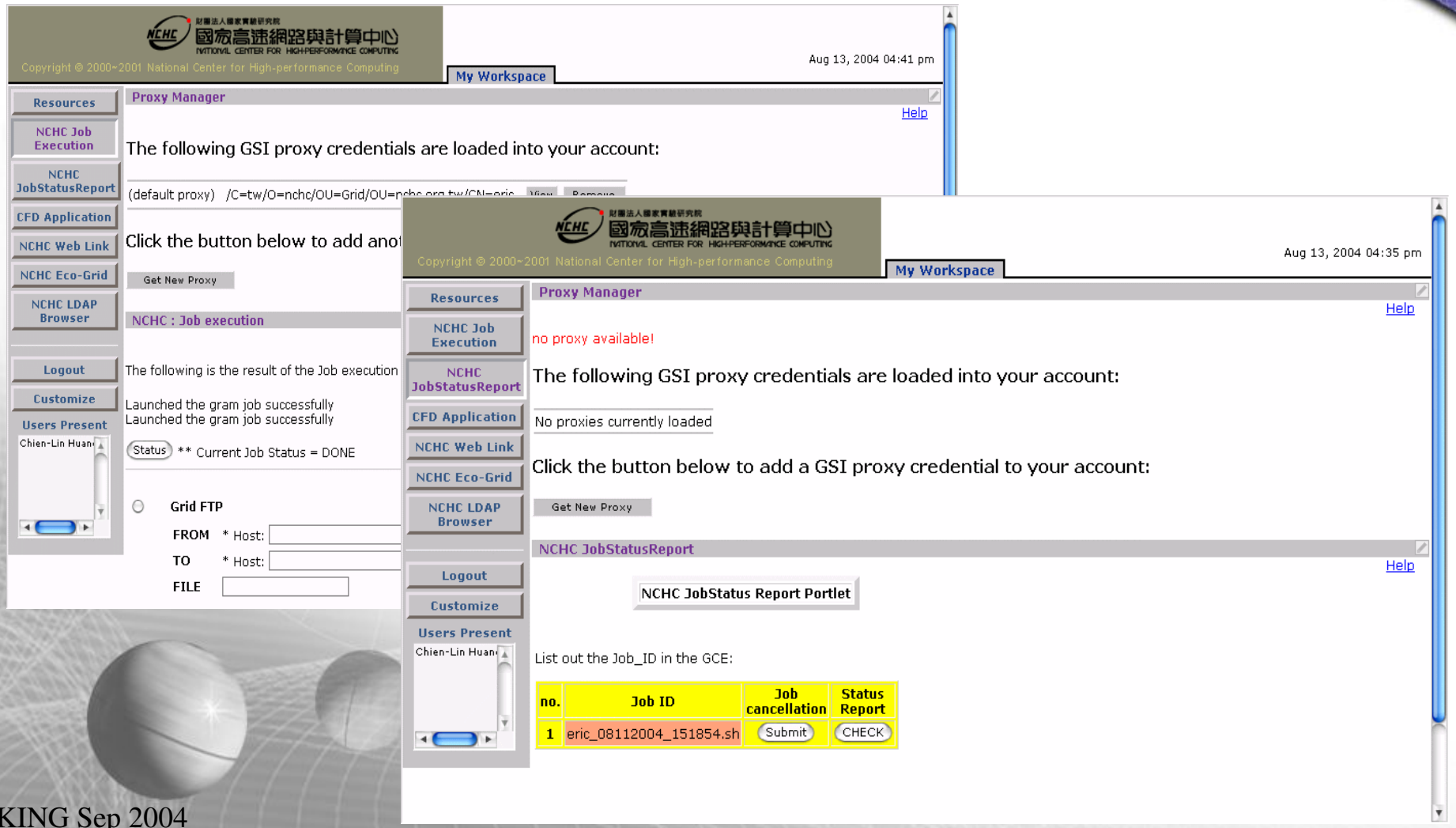
DRBL-G

- Use the Web Portal to describe the work flow

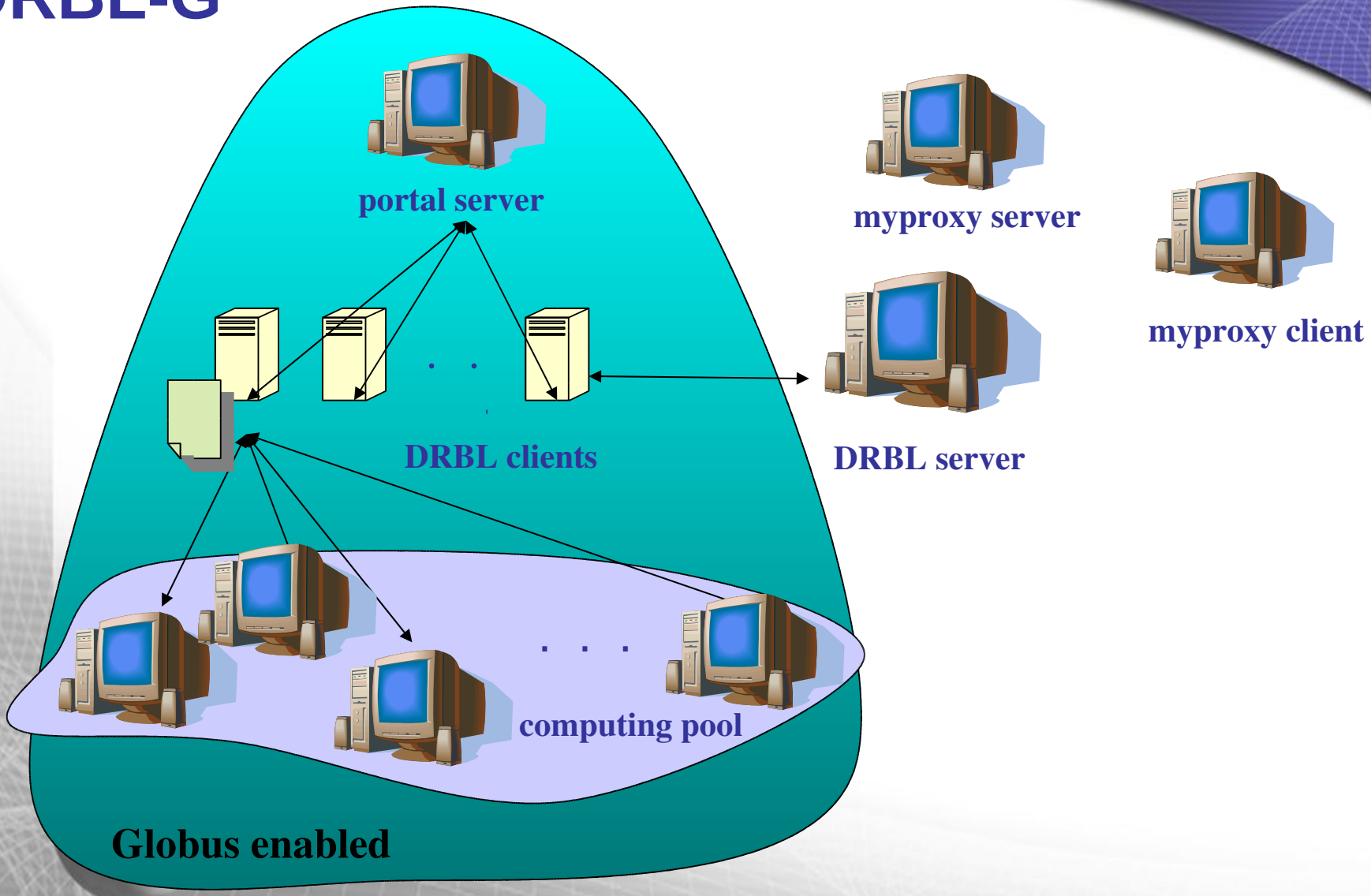


DRBL-G

- Use the Web Portal to monitor the job execution



DRBL-G


portal server
myproxy server
myproxy client
DRBL clients
DRBL server
computing pool
Globus enabled

Reference

- DRBL Project,
  - http://drbl.sf.net; http://drbl.nchc.org.tw
- EtherBoot Project,
  - http://www.etherboot.org
- Preboot Execution Environment,
  - ftp://download.intel.com/labs/manage/wfm/download/pxespec.pdf
- ECOGRID, Grid For Long Term Ecological Research
  - http://ecogrid.nchc.org.tw
- NPACI Rocks Cluster Distribution,
  - http://rocks.npaci.edu
- OSCAR : Open Source Cluster Application Resources,
  - http://oscar.openclustergroup.org
- thin-OSCAR : systemless clients support for OSCAR
  - http://thin-oscar.ccs.usherbrooke.ca/