ComputeMode^{TM} Grid Manager
Version 1.4.1

`http://www.computemode.org/`

ICATIS SAS
37 avenue du Granier
38240 Meylan
France

`http://www.icatis.com/`

Contents

List of Figures

List of Tables

Glossary

In the following definitions list, the letters '(CM)' indicate a meaning specific to ComputeMode Grid Manager.

'Always CM' schedule: (CM) this schedule has the peculiarity to always boot the nodes in computation mode

batch manager: also known as Queue Manager, Job Manager, Task Manager

CM administrator: (CM) ComputeMode Grid Manager administrator - related to the root Unix account on the CM server and to the admin account on the web interface Note: the cmwebadmin administrator may differ from the cmserver administrator

CM boot mode: (CM) a type of boot mode whose role is to help instantiate a CM image OS (for the client nodes)

CM: (CM) ComputeMode Grid Manager

CMGM: (CM) see CM

client node: (CM) the workstation which will be able to boot in Computation Mode (by opposition to CM server)

computation mode: (CM) the mode in which the client nodes should be to successfully the tasks submitted to the CM server

computing mode: (CM) see computation mode

DHCP: Dynamic Host Configuration Protocol

grid user: (CM) a user for which the cmserver administrator has created a Unix account and who is allowed to submit jobs through a batch manager

job manager: see batch manager

NFS: Sun's Network File System

NIS: Network Information Service (a.k.a. "yellow pages")

OS: operating system

Owner: (CM) see user

PXE: Pre-execution Environment

queue manager: see batch manager

RAW PXE boot mode: (CM) a kind of boot mode whose role is to chain other pxelinux static configuration file
Such a mode can be used for instance to make a machine boot through the network a floppy disk image (FreeDOS, Symantec's Ghost, . . .)

remote wake-up: see Wake-on-LAN

RWU: see remote wake-up

standard mode: (CM) the mode in which the client nodes would be if there were no CM server

Task Manager: see Batch Manager

User Mode: (CM) see Standard Mode

User: (CM) the owner of a given node, that is, the person who is using a workstation on a regular basis

Wake-on-LAN: a functionality which lets a user wake up a powered-off system through the network by sending a specifically crafted packet (used in the phrase: to send Wake-on-LAN packets)

WoL: see Wake-on-LAN

boot image: (CM) it is indeed a Linux boot image which is composed of a Linux kernel and a specific initrd tuned for CM - extra parameters may be used too.

boot mode: (CM) indicates how to boot a machine

cmserver, server: (CM) ComputeMode Grid Manager server

cmwebadmin-data: ComputeMode Grid Manager web resources (images, style sheets, logos)

cmwebadmin: (CM) ComputeMode Grid Manager web administration interface

cmwebadmin: ComputeMode Grid Manager web administration interface

image OS: (CM) this is the file hierarchy located in /cm/<OS_NAME> (for instance /cm/debian/...) and which is aimed at being mounted on the clients - it may be hosted on a read only server and requires specific tuning

label: (CM) it is a logical tag attached to a node - its main role is to simplify the handling of large number of hosts by binding it to a symbolic name - labels may for instance be bound to the room number or the hardware brand

local boot mode: (CM) a special reserved boot mode which basically tells the machine to boot on its local hard disk

node, host: (CM) it is a client node, that is a machine which boots a CM OS - a node uses a schedule, and may hold an unspecified number of labels (from none to any number)

off boot mode: (CM) this is a RAW PXE boot mode whose aim is to shut down the machine which has used it

processing node: (CM) see Client Node

schedule: (CM) it is a calendar for a week telling CM which OS to apply to certain days and certain hours

special labels, system labels: (CM) some label names have special meanings to CM - basically they are labels and can be applied and removed as any other label but they cannot be edited or renamed and they may have some special behavior

ssh: secured shell

Chapter 1

ComputeMode Grid Manager Overview

This section aims at describing slightly more thoroughly what CM can and cannot do.

1.1 Quick overview

A ComputeMode server can help you:

- deploy easily and non intrusively a distributed test-bed

- extend an already existing computing system through the aggregation of unused computing resources (idle workstations)

- have a unified view of all your resources

1.2 General overview

ComputeMode relies on a master-slave architecture built by using a central server as a master.

The CM server maintains the availability of registered PCs on the local network. Each PC owner, in accordance with your company policy, may choose to let his/her PC be used at nights, or during weekends or vacations: most workstations are used interactively 50 hours a week. A PC in such a state will be known to be in computation mode, and will also be a client (or processing or slave) node. The other mode is known as 'user mode'.

The two modes are:

- in user mode, in which the machine is working the way it's used to (most likely under Microsoft Windows for instance). The owner of the machine should not even notice that his or her PC is managed as a ComputeMode processing node. In particular, the computational resources of the PC will not be used while in User Mode (unless virtualization support is enabled)

- The Computation Mode (from where the ComputeMode name comes) is activated when the machine is in a time period where the PC is declared as "idle". In case the ComputeMode Server detects some computational peak, the PC can be remotely switched to Computation Mode. The switch from user mode is done through an automatic reboot of the machine and proceeds to a remote boot. The remote boot is handled by the ComputeMode Server with the "pre-execution environment" (PXE) protocol, which is natively available from the BIOS of PCs since 1999. While in Computation Mode, the machine is running under the Linux Operating System, and does not have any access to any local hard disk.

The ComputeMode administrator can easily manage the computing PCs through a web-based interface available on the ComputeMode server.

A grid user (standard Unix user) may submit computational jobs to the system through the use of a classical batch manager by logging in on the server (through ssh for instance), or computation scheduler (for large parametric computation campaigns for instance)

The batch manager will then take care of:

- reserving adequate resources for the computations

- scheduling the execution of the jobs

An Open Source (GPL) batch manager is shipping with CM and named 'OAR', yet other products such as Platform's LSF, OpenPBS, TORQUE or Sun Grid Engine are known to work with ComputeMode.

Load balancing with an already installed job manager is also something which can be accomplished: for instance, your dedicated cluster usually handle the work load but for peak periods, some extra-CPU power would be useful. . .

If an owner wants to When the ComputeMode server detects that some client node has become unavailable, the client nodes returns to the previous User Mode, so that the owner will not even notice that his or her PC has been used by ComputeMode.

The submitted Jobs can take advantage of the NFS distributed file system, which is made available from the ComputeMode Server. Each Grid User has his or her own private directory and can use it for data required by the computational Jobs, as well as to retrieve the produced output files.

A specific case may happen if a PC Owner comes back and needs his or her PC while a computational Job is being processed. This can happen, for instance, if the Owner has an urgent need and comes back at night to work on his/her PC. In such a situation, the owner always has priority over the machine. Just by using the keyboard, the Owner can abort any ongoing computational activity on his/her PC, and ComputeMode will restore the machine to its User Mode in about one minute.

1.3 CM Grid Manager is a cluster/grid building software suite ?

Yes, it is. The word 'grid' is used but you may also see the CM infrastructure as a 'cluster'. The differences between 'cluster' and 'grid' words is somewhat fuzzy (and besides, vary according to different streams of thoughts), so to help you understand what CM can do for you:

- if you have some workstations (PC) available during nights, weekends or vacations, or some spare systems,

- if you have CPU-hungry applications (MPI-like parallel codes, parametric codes, scientific simulations, large compilation projects) running under Unix
  Note: Microsoft OS [DOS, Win 3.x, and current OS] scientific applications can often be run under Linux - this require some specific work which Icatis may help you do

- if you know what the role of a batch manager/scheduler (PBS, Platform's LSF, TORQUE, Sun's Grid Engine/N1, OAR)

- CM will help you make all those entities meet easily.

If you happen to already own and use a dedicated clusters, you may also want to give a try to ComputeMode so as to offer extra power without extra costs (new machines, administration, conditioned air, . . .)

If your application is CPU hungry but you do not know whether it can gain from being parallelized or distributed, feel free to ask us for advice.

If you want to initiate students to distributed infrastructures, as well as cluster tools and technologies, or if you simply want to experiment with a cluster prior to buying a dedicated system, CM will help you reach easily those goals.

If you are working in a scientific laboratory, you may want to use CM to offer your searchers or colleagues an extra infrastructure to test-bed your computations.

The scope and possibilities offered by CM are rather extensive: if you have some doubts, or want a preliminary study or any kind of support or help, please contact us (see 13 on page 74).

and Icatis is available to help you install, design and setup your own server and network infrastructure.

1.4   ComputeMode Grid Manager is Open Source

ComputeMode is indeed a whole set of software available under Open Source licenses (most tools being available through a GNU GPL) - an acknowledgement list will be given in appendix C.1 on page 78.

It is currently based on Linux and ships a Debian GNU/Linux OS for the server. The solution to ship a compete distribution was chosen since lots of configuration have to be done to allow the proper functioning of the whole system. If you really want to use your own distribution, you may seek help from the mailing lists or from Icatis.

As such it is available under Open Source license or Open Source-friendly licenses :

- cmwebadmin: the web interface administration is available under the GNU GPL 2

- all the other codes or patches are available for free (sources are available on ComputeMode web site)

- cmwebadmin-data: (basically, images, style sheets, logos) are available under a Creative Common license (Attribution-NonCommercial-NoDerivs 2.5) which basically lets you redistribute it unmodified for free telling you got it from Icatis. cmwebadmin will work without it but it should be much esthetically pleasant with it -

you are free to do a substitution package, this license is basically a solution to enforce our trademark and logos - if you find this license too restrictive for some projects you are doing, we are open to discussion so please contact us (see 13 on page 74).

You may also purchase a professional version (basically extra functionalities plus support) as well as specific support and services from Icatis.

1.5 Current requirements

As CM is a distributed architecture it has some requirements on the server and the client nodes, as well as your network topology.

1.5.1 Server requirement

The server should be a fast machine dedicated to running ComputeMode. To give very minimal figures, the settings should be worth at least:

- Pentium class processor - at least 1 GHz (the higher, the better)

- 256MB memory

- 100Mbps Ethernet

- 4GB hard disk (SCSI or IDE and a few RAID cards are also supported - SATA is currently not tested) : note the HDD will be repartitioned and reformatted upon ComputeMode installation; hence the full contents of the hard disk will be erased.

You will have to alter these figures according to your needs as, of course, more client nodes require more power. To give a few figures : a Dual Xeon 2.4GHz with 2GB RAM and a gigabit network cards handles around 120 client nodes. If your application make lots of I/O, you may have to boost the network link and the storage device.

1.5.2 Client requirement

Processing nodes will be remote booted when required by the ComputeMode Server. This process involves running a disk-less Linux (through a network boot) on each of these PCs. Not much requirements are made onto these PCs, but of course the faster the machines, the better the computing performance will be:

Pentium-class processor

- 80MB of RAM: do note this is the minimum value to make somethings and that this value will have to be increased to cope with your application needs

- 100 Mbps Ethernet : Remote-Wake-Up (a.k.a. Wake-on-LAN) support is not mandatory but it would make a great addition

- PXE compliance: this feature makes a PC able to network boot. PXE capability is supported on most PC with an integrated LAN network card. It has been integrated by major vendors' corporate PCs since 1999 (and is required by PC98 and PC99 recommendations from Microsoft and by the Wired for Management Initiative from Intel).

- PXE should be enabled on each client node through their BIOS setup. Network boot is usually enabled by default by PC manufacturers, but you should check that each PC is properly PXE-enabled. The PXE/Network boot must be configured as the first boot device in order to fully take advantage of ComputeMode boot mechanism. Some systems also activate PXE boot when the system receives a Wake-on-LAN message.

1.5.3 Network architecture requirements

The current version of ComputeMode relies heavily on the use of PXE (also known as network boot) to setup disk-less distributions. Thus, if your network already uses PXE for other purposes, or if your workstations are too old and do not support PXE boot, this may be a show-stopper for this current version - feel free to ask Icatis for guidance in such a case as ther may be fine tuning or alternatives out of the scope of this manual.

CM also uses DHCP but it is configured so as to be transparent and not interact with your current set up.

Besides, the following constraints should be enforced:

- All ComputeMode nodes should be in the same broadcast range. For instance, having all the machines connected on the same switch, sharing the same subnet or VLAN will be sufficient. This is because ComputeMode relies on the DHCP/PXE protocol, which requires the server to be able to receive Ethernet frames broadcasted on the LAN by any of the Processing Nodes.

- PXE should *not* be in use on your local network. PXE is sometimes used by corporations that use it for deploying new Operating System images to network PCs. If PXE is already in use on your network, the PXE from the ComputeMode Server will conflict with the corporate PXE. You should check with your network administrator whether PXE is on your network or not. In such a case however, advanced configurations of the PXE environment may be provided by Icatis, adapted to your local needs, using for instance an advanced PXE bootstrap chaining mechanism Tip: To check whether PXE is in use on your premises, boot a PC and let it proceed into PXE network book (this should happen at the end of its BIOS bootstrap) - of course, the ComputeMode server has be disconnected during that test. If the PC exits without finding any PXE server, then there are chances that no PXE server is present on your local network.

1.5.4 Access nodes requirements

Some hosts of your network will be used to access the ComputeMode system. For instance, you will need to access the ComputeMode Web interface located on the ComputeMode server. You will also need to access the system to submit computational jobs and retrieve their results. Such operations may be done using any PC from your network (including processing nodes).

The ComputeMode server embeds a minimal X display with a web browser, but as it is a server, it will most likely be more comfortable to access it from your office instead of the server room !

The requirements are rather pieces of advice:

- a standard-compliant web browser supporting CSS and Javascript: if Javascript is not supported, every functionality is still available but in a less user-friendly manner. Such browsers are : Gecko-based ones (Firefox 1.x), Opera (8.+). Microsoft's IE5 (Windows 2000) and IE6 (Windows XP) are tested but less thoroughly, so they may render pages with minor display glitches - IE7 is expected to offer better results for it is expected to (at last) properly support W3C standards.

- a 1024x768 display (at least) - a smaller screen will make using cmwebadmin less user-friendly

- a ssh client: several are available at no charge and under Open Source licenses (see OpenSSH with Cygwin or PuTTY). You may use commercial ssh clients, too. If you do not intend to either submit job or administer the CM server, you probably do not need such a tool.

1.6 ComputeMode Grid Manager flavors

ComputeMode Grid Manager ships currently in two flavors available through the Internet. If you do not want to download the ComputeMode software from the Internet, please contact Icatis (see section 13 on page 74) and ask for a physical CD-ROM to be sent.

1.6.1 ComputeMode ISO image

To install a native ComputeMode server, you will first need to get a ComputeMode Server installation CD-ROM.

The ComputeMode CD-ROM is distributed in the form of an ISO file which is several hundred MB large: you must ensure that your Internet connection is able to perform such a download. Once the download is finished, burn the `*.iso` file image onto a blank CD-ROM (or CD-RW) and reboot.

The detailed server setup is explained in a later chapter.

1.6.2 ComputeMode VMware appliance

For testing purpose (as there is some kind of performance limitations), a VMware appliance may be downloaded from :

> ComputeMode `http://www.computemode.org/`

It is a fully functional ComputeMode server which lets you test the software prior to dedicating a real server to it.

Part I

Getting a ComputeMode server up and running

Chapter 2

Starting with the ComputeMode CD-ROM

This section aims at describing how to install from scratch a CM server using the available CM ISO image. Do note that such an installation will destroy any previously installed OS. If you cannot afford this, then consider testing CM through the VMware appliance (see next chapter) or by installing CM in a virtual machine (see section 2.2).

2.1   Installing a CM server on a real machine

First, make sure the network and system requirements are fulfilled.

In the following section, we will assume your CM server will have:

- have a IP of 1.2.3.42
- a netmask of 255.255.255.0
- a gateway of 1.2.3.254
- a DNS server of 1.2.3.1

This is the IP you will have to use to access the CM server from within your site.

You can then proceed to the installation of ComputeMode by simply booting on the CM CD-ROM.

The installation screens should be rather straightforward. When requested by the installer, enter the ComputeMode Server network parameters.. Once the installation is finished, go to your own machine (which will be used as an Access Node), open your Web browser and navigate to the following ComputeMode Server address, prepended by "http://". For example, open the "http://1.2.3.42/" Web page. This is the Compute-Mode Server administration page.

2.2   Installing a CM server in a virtual machine

In most virtualization software, I/O operations are privileged actions which require the virtualization software intervention, hence performances are often degraded compared to a real dedicated server.

2.2.1   VMware Workstation or VMware Server

If you already use VMware's virtualization products you may simply install the CM software inside a virtual machine. You have however to be sure that it will get access to the same network as your clients (whether these are real or virtual):

- if the client nodes are real machines, it will be easier to use the bridge mode

Note: you may have to alter the `/dev/vmnet0` permissions to allow to allow promiscuous mode (you will know when you need it as VMware will complain - for instance if you want to run a network sniffer to troubleshoot possible networking issues)

- if you want your clients to be hosted by the same VMware system as your server, then use a team, NAT or host-only networking to avoid connecting to your physical network. Please consult the manual of the VMware product you want to use.

2.2.2   qemu

qemu is a free CPU emulator mainly developed by Fabrice Bellard released under the GNU GPL. The author developed a closed-source but freely available kernel-level accelerator. It may be downloaded from the qemu website `http://fabrice.bellard.free.fr/qemu/` and will greatly enhance the CPU performances.

```
dd of=cm.img bs=1M seek=4000 count=0
qemu-img create cm.img 4G
qemu -hda cm.img -boot d -cdrom cminstaller-X.Y.Z.iso -user-net
```

2.3   Thorough Server Setup

Once you have booted on the CD-ROM, CM installer will take you through the following steps and show the following screens.
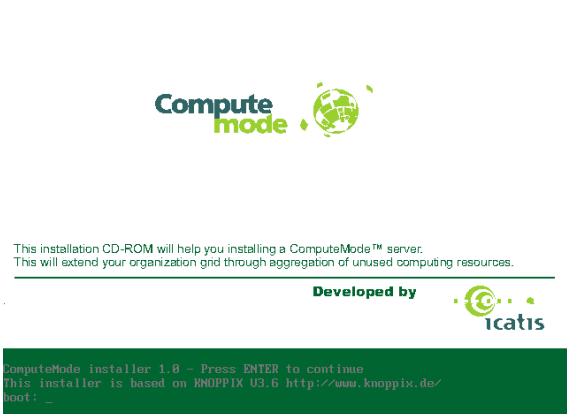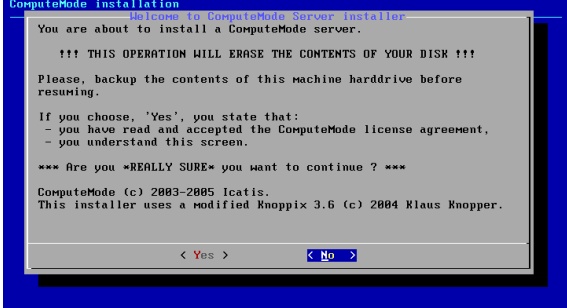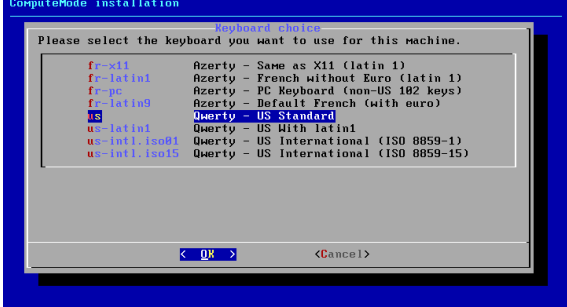
Screenshot	Comments
	This is the initial screen you will see.
	If you agree the license, choose OK.
	Choose your keyboard flavor.

Table 2.1: Installation: early screens

With regards to the table 2.1, screens are relatively obvious. The following additions may be given:

- License: basically we are distributing a complete OS system with software available under several license, by hitting yes, you acknowledge that you understand what this means.

- Keyboard: currently there are only QWERTY-US and AZERTY-FR keyboards supported - if you have any other flavor, this can be changed at a later time when the CM server has rebooted.
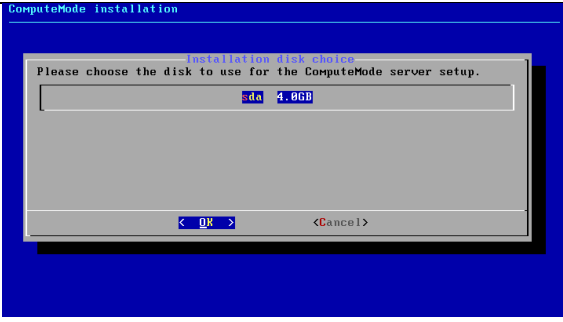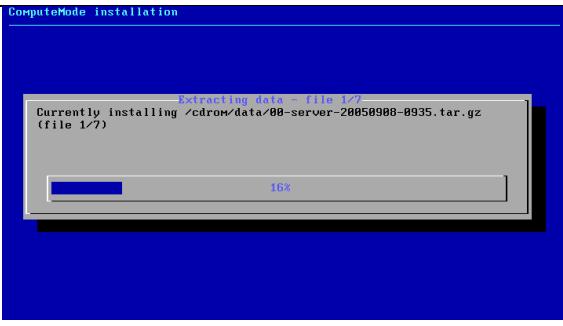
Screenshot	Comments
	Choose the hard drive you want to use.
	Get a coffee or soda.

Table 2.2: Installation: hard disk setup

The table 2.2 illustrates some of the sample screens you will go through when you have to choose an installation disk.

Note: Your harddisk will be wiped when you start installation - all data on the chosen disk will be lost! This cannot be stressed enough.
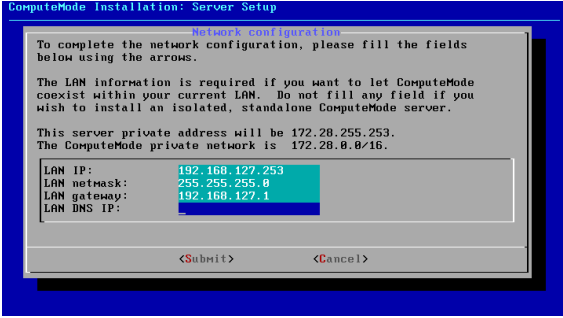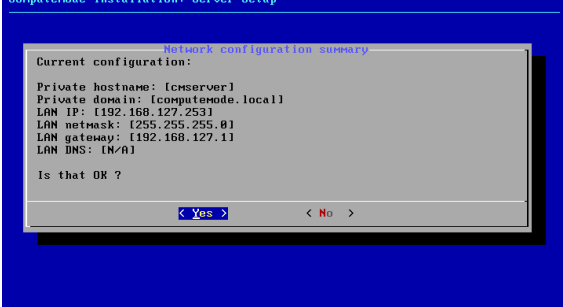
Screenshot	Comments
	Fill the network form :
	• If you specify an IP address and a netmask, in the same subnet as the one used on your LAN, you will be able to access cmwebadmin from your workstation.
	• If you specify an IP address as well as a netmask, which is in the same subnet as the one used on your LAN, you will be able to access cmwebadmin from your workstation.
	• If a gateway is specified, the CM server will be able to access a machine located beyond your router (outside your LAN, or the Internet - which may be convenient to fetch system updates).
	• If a DNS is specified, the CM server will be aware of the naming of the other machines within your intranet.
	If the network configuration is OK, hit 'Yes'

Table 2.3: Installation: network configuration

As CM is mainly aimed at being a headless system (through a remote administration facility), network configuration is the part which will require your attention. Explanations are given in table 2.3 - if you are not sure about what to do, feel free to contact us (see section 13 on page 74).

Note: The private network 172.28.0.0/16 is reserved for CM use - if you are already using this subnetwork, please contact us for an updated version.
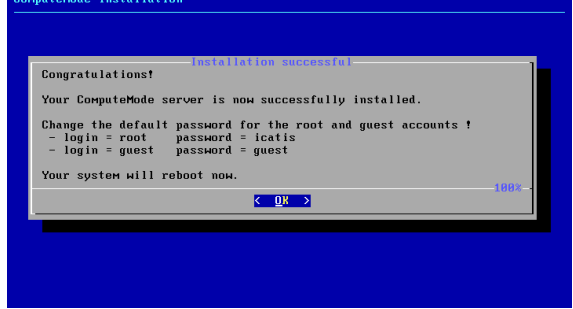
Screenshot	Comments
	Hit 'Yes' to reboot.

Table 2.4: Installation: last screen before reboot

If everything went smoothly, then your system should now tell you it will reboot (see table 2.4) and take you to the graphical login as shown on table 2.5.

If not, there might have been some problem during the installation, feel free to drop a message (see 13 on page 74)..

Screenshot	Comments
GNU GRUB version 0.95 (638K lower / 161728K upper memory) **Debian GNU/Linux, kernel 2.4.27-2-686** Debian GNU/Linux, kernel 2.4.27-2-686 (recovery mode) Use the ↑ and ↓ keys to select which entry is highlighted. Press enter to boot the selected OS, 'e' to edit the commands before booting, or 'c' for a command-line. Compute mode • Developed by icatis	GRUB boot loader
ComputeMode Server Debian GNU/Linux 3.1 cmserver tty1 cmserver login: _	Console login
Login Authentication **Welcome to** **cmserver** Login name: | debian Start WM — NoChange Options — Login Help Start Over Go!	Graphical login

Table 2.5: Boot loading screen shots

You may now head directly to chapter 4 on page 29.

Chapter 3

Starting the ComputeMode Virtual Appliance

This chapter describes how to use the VMware appliance of the ComputeMode Grid Manager software.

3.1 VMware Player

The appliance version of CM is built to work out-of-the-box with VMware Player. You can get this tool at no charge from the VMware Player download page `http://www.vmware.com/download/pl`

3.2 Recommended usage

Since VMware adds some overhead on I/O operations (disk, network), and since CM is mainly doing I/O, it should mainly be used for evaluation purposes:

- software evaluation : you do not have to get a specific setup

- quick all-in-one cluster setup: just start it, you can get a small cluster running within minutes

3.3 Appliance Walk-through

3.3.1 Requirements

You need to know: your netmask, an available IP address, your DNS servers, if PXE is used in your service.

3.3.2   Starting the engine

The appliance has been designed to be simple to start and provides a common web page as shown on figure 3.1.
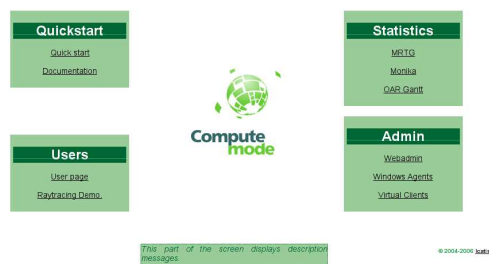
Figure 3.1: Welcome screen

The main part of CM is in the 'Admin' box, especially the 'Webadmin' link. You may now proceed to the chapter 7 on page 43 if you want to have detailed explanations about administration or simply proceed to the next chapter to resume your ComputeMode walk-through.

Part II

Launching Computations

Chapter 4

Starting Computing

Now that your CM server is running - everything from now on will apply whether your CM server may be real or virtual.

4.1   Logging into cmwebadmin

Once you have logged in your Unix account (or the guest account), you will be shown with a login screen (see figure 4.1). You may also reach cmwebadmin through a web browser located on your desktop machine. cmwebadmin is accessible through HTTP or HTTPS.

Figure 4.1: cmwebadmin login screen

The default account and password are given in section A on page 76. It is a safe practice to change this password now.

Once logged you are taken to the nodes management page (see figure 4.2).

You may choose which labels to show by selecting the drop-down menu at the top. Upon selection, the page will be automatically reloaded (if you have Javascript) to only show the nodes holding the selected label.

The nodes are listed and some of their properties are show in columns. Most column contents are obvious yet a few deserve explanations:

- Schedule column: it is the name of schedule which applies to the node

- WoL column: it stands for Wake-on-LAN and it indicates whether a Wake-on-LAN (a.k.a. Remote Wake Up) message should be sent to a node which has to boot a non-Local boot mode

- Notes column: it is left for the user to add comments

- Schedule boot mode column: this column indicates the boot mode the node should be in if it booted now ('now' being the time of the web page generation) - this column is not editable (read/only) and is automatically set by CM

- Labels column: indicates the labels tied to a given node
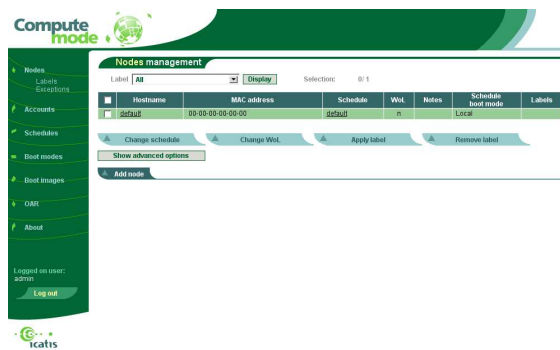
Figure 4.2: Nodes management page

Clicking on the 'Show/Hide advanced options' button, will toggle the display of several buttons.

Note: the buttons shown by your CM version may differ from the one shown in figure 4.3 as the display is highly configurable and dependent on which modules are available and enabled.
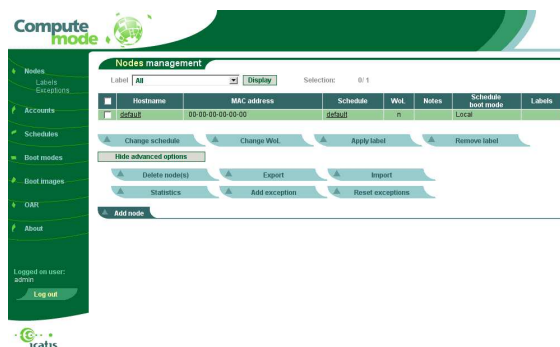
Figure 4.3: Nodes management page - with advanced options shown

4.2 Getting some computation nodes registered

Client (or computation) nodes may be added automatically, or by hand. The following sections is to explain the two options.

4.2.1 Automatic discovery

First, you have to check that this feature is enabled in the BIOS settings of the machines you want to use. A sample BIOS screen is shown in figure 4.4.
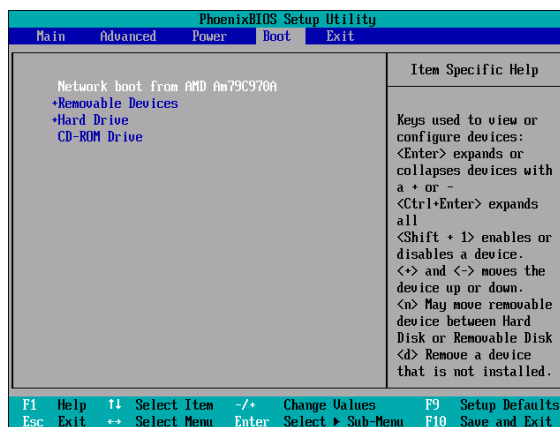
Figure 4.4: Enabling network boot in the BIOS

You can now check whether the node is listed in the nodes management screen. There should not been any node if you never booted it.
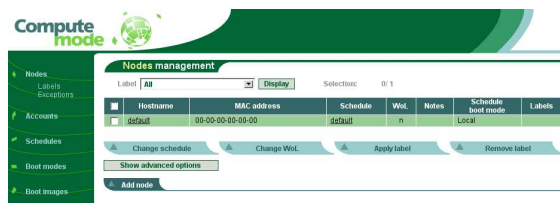
Figure 4.5: Nodes management before client node first boot

Once you start the client machine: you should see a screen similar to the one in figure 4.6. During this step, you machine tries to find a PXE boot server within your LAN. This request is caught by cmserver which then registers this node as an unconfigured node and adds an automatic name of U_xx-xx-xx-xx-xx-xx (where xx-xx-xx-xx-xx-xx is your client node MAC address). If you choose to reload the nodes management page (as shown on figure 4.7), you will see a new node was added with the label 'Unconfigured'.

If you feel like it, you can verify the MAC address you may have seen in figure 4.6 is the same as the one you will see in figure 4.7.

Note: if several machines happen to boot at the same time, there may be issues to know which one is which unless you know the association MAC address - machine name.

Figure 4.6: Client node first boot: PXE request

Figure 4.7: Nodes management after client node first boot:a new node is listed

4.2.2 Manual addition

If you know your machine MAC address, you may simply add the corresponding entry into cmwebadmin: just click the 'Add node' button. You will have to fill the form - see the next section, or chapter 7 on page 43, for further information about every line.

4.3 Editing the newly-added node properties

If you click on the 'U_xx-xx-xx-xx-xx-xx' name in the nodes management, you will be taken to the node edition properties screen. You should chose the 'Always CM' schedule so as to observe something when you reboot the client node. When you are satisfied with the form, click 'Update' button (see figure 4.8).

Figure 4.8: Node properties edition

If everything went smoothly, you will be taken back to the nodes management screen with a message shown at the top of the page as shown on figure 4.9.
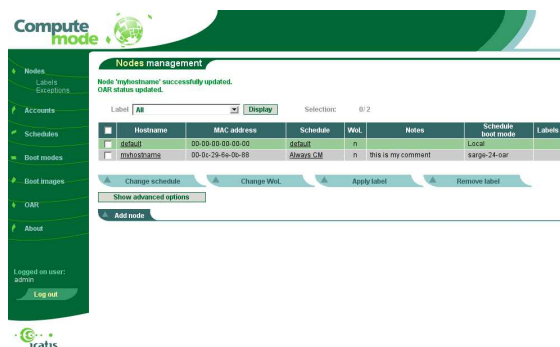
Figure 4.9: Node management after successful node properties edition

Now, if you reboot the machine for which you have just modified the configuration, screens similar to the ones in table 4.1 will be shown.

4.4   Getting some work done: cluster computing demonstration

4.4.1   Principle of the demonstration

A simple cluster demonstration is provided so that you may test your setup. This application is made to be simple and visual: basically a 3D scene is split in stripes, each stripe is a task submitted to the job manager. The job manager dispatches the tasks on every available computation node.

Meanwhile, the ComputeMode server will merge and update the display of the computed stripes till every task has been executed.

Screen shot	Comments
	Boot underway
	Boot over

Table 4.1: CM client boot screens

4.4.2 How to start it ?

You can check how many nodes you have by selecting the OAR menu option and choosing Monika. A page similar to the one below will be shown.

ComputeMode OAR nodes

Summary:

OAR node status	Free	Busy	Total
Nodes	4	0	4

Reservations:

cmserver Free pc1 Free pc2 Free pc3 Free

Job details:

No job currently in queues

- **Monika** - ©**ID-Imag** -

Figure 4.10: OAR page listing the ComputeMode nodes

Then you have to select the 'POV demonstration item' in CM demonstration page (`http://.../~pov/`)

Figure 4.11: POV scenes listing

A few sample scenes are given - choose one and click its name - the rendering process will now start. If you launch Monika again (in another window) you will see that several tasks have been queued (see the figure below).

Figure 4.12: OAR page listing the ComputeMode nodes executing jobs

After a few seconds, the first results will begin to appear:

Figure 4.13: POV scene rendering

The nodes scheduling may be seen through the Gantt drawings (as an option in the OAR sub-menu):
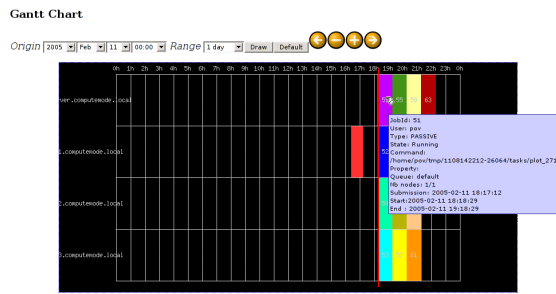
Figure 4.14: DrawOARGantt showing submitted jobs

Eventually the full picture is displayed.

Povray rendering of scenes/chess2.pov

* Povray scene to render : scenes/chess2.pov
* Povray job ID = 11081 42212-26064 (results will be at
http://192.168.7.253/~pov/tmp/chess2-11081 42212-26064.html
* Setting up tasks environment...Ok

* Launching rendering process - Please be patient, this may take some time to initialize...
!!! WARNING: Please to not try to reload this page manually as it will actually queue a new rendering process !!!

```
POVRAYDEMO: submitting tasks... ###################################
Waiting for results... [##                          ] (02/32)  +000s = 000s [pc3:1]
Waiting for results... [######                      ] (06/32)  +015s = 015s [pc3:2] [pc1:1] [pc2:1] [cmserver:1]
Waiting for results... [##########                  ] (10/32)  +014s = 029s [pc3:3] [pc1:2] [pc2:2] [cmserver:2]
Waiting for results... [#############               ] (13/32)  +015s = 044s [pc3:4] [pc1:3] [pc2:3] [cmserver:2]
Waiting for results... [################            ] (16/32)  +016s = 060s [pc3:4] [pc1:4] [pc2:4] [cmserver:3]
Waiting for results... [##################          ] (19/32)  +017s = 077s [pc3:5] [pc1:5] [pc2:5] [cmserver:3]
Waiting for results... [#######################     ] (23/32)  +018s = 095s [pc3:6] [pc1:6] [pc2:6] [cmserver:4]
Waiting for results... [##########################  ] (27/32)  +017s = 112s [pc3:8] [pc1:7] [pc2:7] [cmserver:4]
Waiting for results... [############################] (30/32)  +015s = 127s [pc3:8] [pc1:8] [pc2:8] [cmserver:5]
Waiting for results... [############################] (32/32)  +008s = 135s [pc3:9] [pc1:8] [pc2:8] [cmserver:6]
Cleaning environment
```

Figure 4.15: Complete 3D scene rendered

Part III

Grid Manager for users

Chapter 5

As a Grid User, How Do I. . .

This section list is a bit built like a FAQ section.

5.1 . . . log in on the CM server ?

Your CM administrator has to create a Unix account.

Once done, you may log in by using a ssh client.

5.2 . . . submit jobs to the batch manager ?

The batch scheduler installed by default is OAR and provides the de facto standards command line tools: *sub/*del/*stat.

Please consider reading the complete OAR manual at: `http://oar.imag.fr/` if you have advanced needs. For sake of completeness, only the most common usage will be shown below.

5.2.1 oarnodes (monitoring)

oarnodes provides information about the nodes registered in ComputeMode: basically it will let you know which nodes may join the computing cluster. It will also list their properties if they have already booted and registered in OAR previously (CPU, RAM, etc.)

Example:

```
oarnodes
```

or

```
oarnodes -l
```

5.2.2   oarstat (monitoring)

oarstat displays the current jobs status and its behavior may be altered with the following options:

- -f: prints further information about each job
- -a: prints more detail and try to use a table format

More concise high level interfaces are also available: "Monika" and "DrawOARGantt". These analysis tools are available as web pages in cmwebadmin. These pages are accessible from cmwebadmin welcome page (see section 3.1 on page 27).

Monika displays a snapshot of current OAR status regarding to nodes occupation and a list of running or pending jobs.

As for DrawOARGantt, it displays a Gantt diagram of the past nodes reservations.

5.2.3   oarsub (submit)

oarsub lets a task be submitted to the job manager.

If you want to have an interactive shell, type:

```
oarsub -I
```

You may specify the number of nodes you want by adding the -l option followed by a number:

```
oarsub -I -l nodes=4
```

The contents of the nodes attributed for your job will be in the $OARNODES environment variable.

You may also submit scripts (files) or in-line scripts:

```
oarsub -l nodes=1 ~/myscript.sh
```

or (for in-line submission):

```
oarsub -l nodes=1 '~/mybin param1 param2'
```

Note: when submitting scripts files, you have to make sure the script is executable (`chmod +x ~/myscript.sh`)

5.2.4   oardel (delete a job)

To remove a job from the OAR queue, you have to get its job ID (use oarstat for this purpose) and then use the following command:

```
oardel <job_ID>
```

If the job is currently running on a client node, it will be killed.

Chapter 6

As an owner, how do I. . .

This chapter aims at giving quick recipes to solve common issues for an administrator.

6.1 . . . let my system be a part of the CM grid ?

Everything is on your administrator and your site policy :

- you may have to switch down your computer at night or when you leave

- you may have a special screen saver installed by your IT staff

- you may have a ComputeMode agent installed

By default if you boot, during a computation period and, if your machine was registered in the CM server by the administrator, then you can see the CM boot process happen.

6.2 . . . get out of the CM grid ?

To achieve this, you simply have to hit 'alt-ctl-del' when the system is in computation mode and a wait a few seconds so that the servers acknowledges it. Once done, you will never see CM again unless you tell your system administrator that it is OK to use your machine as a computation node again.

Part IV

Grid Manager for administrators

Chapter 7

ComputeMode web administration interface

ComputeMode Web Administration interface (cmwebadmin for short) is the central place where most of CM behavior can be tuned.

cmwebadmin is built as follows:

- a menu on the left of the screen which will take you to management pages (for nodes, users, etc.)

- on any management page, you have a list of items and buttons at the bottom of the page to affect the selected objects

7.1 The 'Nodes' menu item
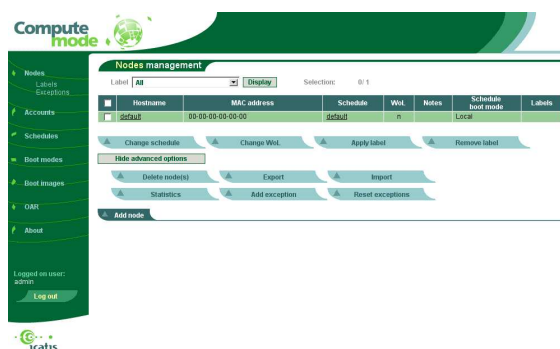
Figure 7.1: Nodes management page - with advanced options shown

7.1.1 Editing or adding a node

To reach the node edition page, simply click its name or its MAC address in the node management page.

To add a new node, click the 'Add' button in the management page.

The fields in the form have the following uses:

Figure 7.2: Adding a node

- MAC address: the machine Ethernet address

- name: the hostname the machine will be known by

- owner: currently, simply used for presentation (in doubt, choose 'admin')

- note: basically available for the user - a few special expressions may be used to alter a node behavior (for instance the `ip=x.y.z.t` option, see below)

- Wake-on-LAN: indicates whether the node should be woken up when it should be in a computation time slot

- node schedule menu: indicates which schedule to use

A few read-only fields are shown when you edit a node (instead of adding it), namely OAR properties and the list of labels having been applied.

When satisfied with the node entries, click the 'Add' (or 'Update') button.

Special field: Comment

Some options can be given to alter the behavior of CM by means of the comments field:

- `ip=x.y.z.t` (where x.y.z.t are the components of an IP address): add such a comment and you will allow the owner of a client node to edit its availability (if the user pages are accessible) from the IP specified

- other options may follow

7.1.2 Change schedule

Select the nodes for which you want to alter the schedule in the management page, and click the 'Change schedule' button.

A summary page indicating which the nodes chosen are, and asking you for the new schedule is shown such as in figure 7.3.
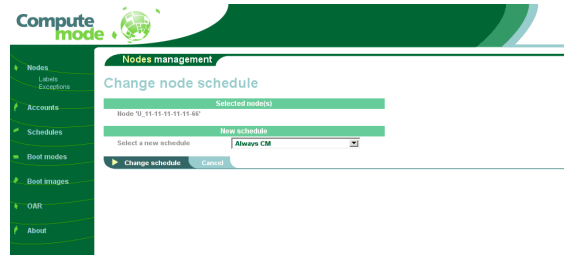
Figure 7.3: Changing a node schedule

Click the 'Change schedule' button to confirm.

7.1.3 Change WoL

The principle is similar to the one of the 'Change schedule' button (see section 7.1.2): select your nodes and use the 'Change WoL' button instead.

7.1.4 Apply Label

The principle is similar to the one of the 'Change schedule' button (see section 7.1.2): select your nodes and use the 'Apply Label' button instead.

Note: the label to apply must be created before by using the 'Nodes/Labels' menu (see section 7.2.4 on page 49)

7.1.5 Remove Label

The principle is similar to the one of the 'Change schedule' button (see section 7.1.2): select your nodes and use the 'Remove Label' button instead.

Note: the nodes will not be deleted - the same thing applies for labels not tied to any node

7.1.6 Delete Nodes (advanced options)

Select the nodes to delete, click the 'Delete' button. A summary confirmation screen is shown. Click 'Delete' to confirm.

7.1.7 Export Nodes (advanced options)

Select the nodes you want to export, then click the 'Export' button.
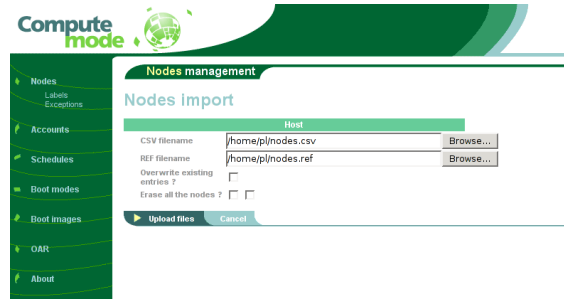
Figure 7.5: Importing nodes

Figure 7.4: Exporting nodes

Once this is done, you have to download both REF and CSV files. The CSV file may be imported in your spreadsheet. Both files are text files which may be open in any text editor.

7.1.8 Import Nodes (advanced options)

Click the 'Import' button - no node has to be selected. You will be taken to a screen similar to the one shown in figure 7.5.

First, you will need to have both CSV and REF file - this is mandatory.

Then you have to options:

- either you choose to overwrite existing entries (this is enabled by checking the box)

- if yes, you may also choose to completely erase *all the nodes* (this is enabled by checking both check-boxes) in the 'Erase all nodes' option

The REF format describe the columns of the CSV format. The file format will be given at a later time.

7.1.9   Statistics (advanced options)

Select the nodes of which you want to have statistics, then click the 'Statistics' button.
You will then be taken to a page similar to the one shown in figure 7.6.
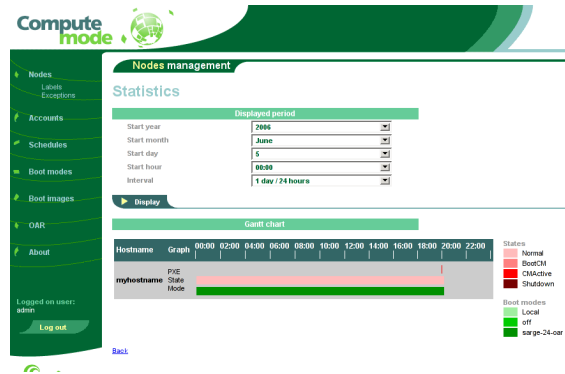
Figure 7.6: Node statistics

You may change the date and time period for which you want statistics to be drawn.

Note: if you select lots of nodes and a long time period, over a slow network link, the
page rendering may either be slow, or even timeout.

7.1.10   Add Exception (advanced options)

Select the nodes to which you want to add exceptions, then click the button 'Add ex-
ception'.

	Standard screen
	Click on an edit box to have a calendar shown.
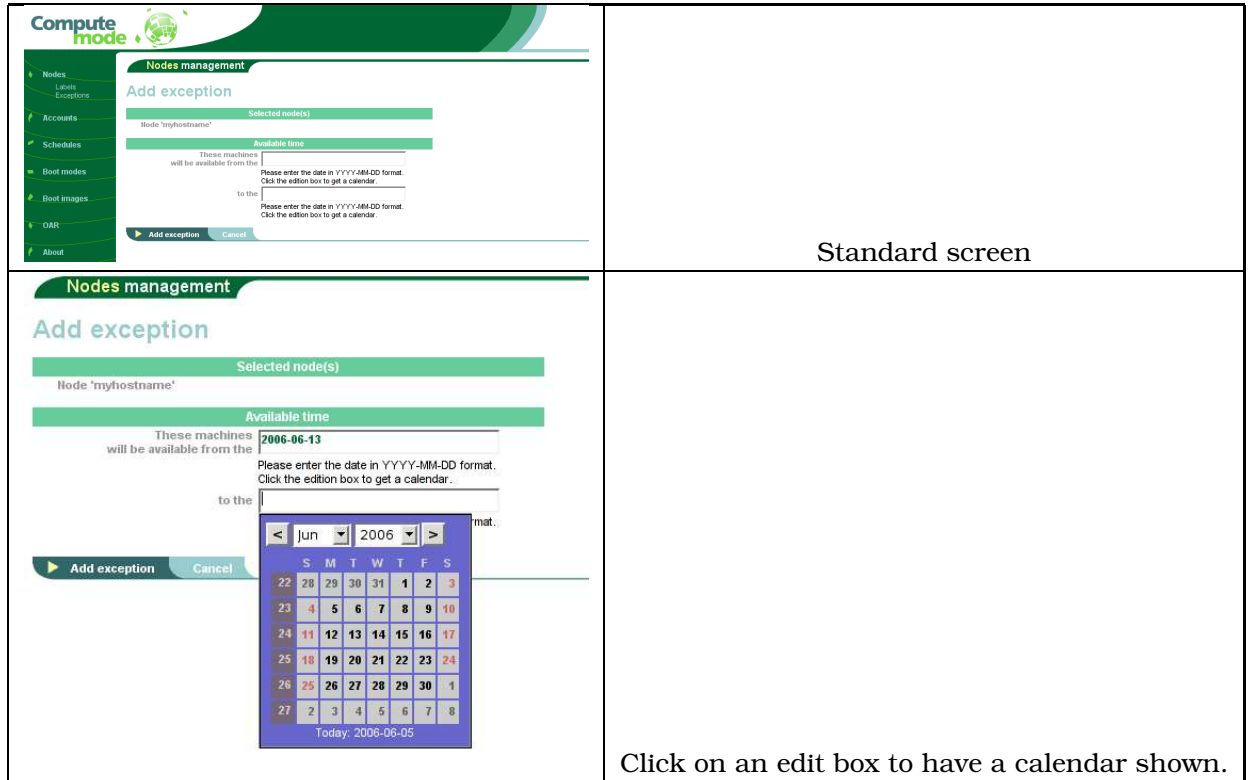
Figure 7.7: Adding an exception to a set of nodes

7.2 The 'Nodes / Labels' sub-menu item
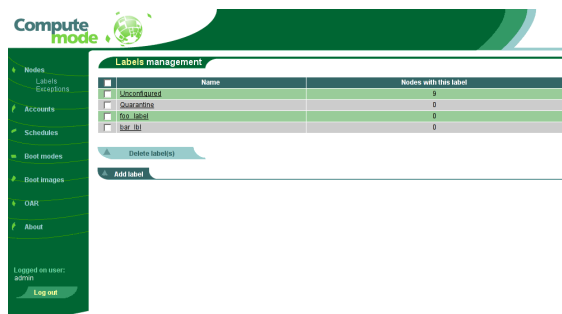
Figure 7.8: Nodes / Labels management page

The column named 'Nodes with this label' indicates the number of nodes which currently have this label applied.

There are two special labels which will be described below.

7.2.1 Special label 'Unconfigured'

This label is added to automatically newly-discovered nodes.

When you edit the node properties and click 'Update' this label will be automatically removed.

7.2.2 Special label 'Quarantine'

This label is automatically added to the nodes which were in Computing modes and whose owner hit the ctl-alt-del to get their machine back.

To disable automatic 'Quarantine' also called within CM node inhibition, edit CM configuration file: config.php (which should be located in /var/www/cm/).

In this file, find the line reading:

```
$GLOBALS['ALLOW_INHIBIT'] = true;
```

Replace it by :

```
$GLOBALS['ALLOW_INHIBIT'] = false;
```

7.2.3 Label edition

You may rename a label by clicking its name in the labels management page (see figure 7.8). The nodes with the old label will maintain the new one.

7.2.4 Add label

Simply click the 'Add' button in the management page (see figure 7.8). A form will be shown asking you to enter a label name (basically only alpha-numerical characters) - click 'Add' to create the label.

7.2.5 Delete labels

Select the labels to erase in the management page (see figure 7.8) the click the 'Delete' button. A confirmation screen will be shown.

Note: the nodes will not be erased, only the label deleted will be removed

7.3 The 'Nodes / Exceptions' sub-menu item

Exceptions can only be listed and deleted in this page (see figure 7.9). They cannot currently be edited. They are "positive" exceptions in the sense they indicate when machines are free to join the computing mode.

Figure 7.9: Nodes / Exceptions management page

7.3.1   Delete exceptions

Select the exceptions you want to remove then click

This is an advanced feature. Further information will be added later.

7.4   The 'Accounts' menu item

This is where you go if you want to change the administrator information. Please note that the accounts shown here are only related to cmwebadmin and are not tied at all with the Unix accounts.
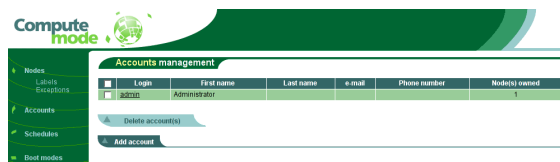
Figure 7.10: Accounts management page

7.4.1   Account edition

Figure 7.11: Accounts edition

The main field in this form are obviously the login field and the password.
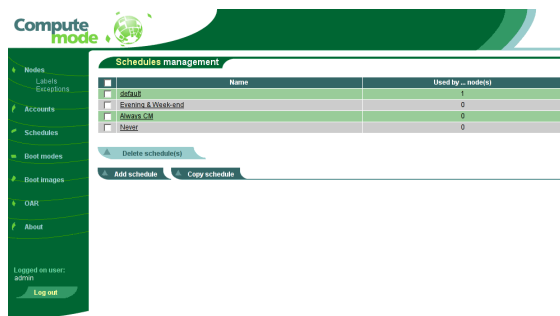
Note: if the password field is left empty, the previous password is kept.

7.4.2   Add/Delete account

Currently, there is no use adding extra CM accounts - the only limitation is there must be at least one admin account.

7.5   The 'Schedules' menu item

This page list the existing schedule (see figure 7.12) as well as the number of nodes which are using it. Please note that some schedule are reserved (though you can edit them to fit your needs): default is seen as a template for new schedules, 'Always' and 'Never' should correspond respectively to always-computing and always-local.

Figure 7.12: Schedules management page

7.5.1   Schedule edition

Schedule edition consists in editing setting which boot mode to apply based on the day of the week and the time of the day.

You can select some parts by clicking then dragging the mouse to select a rectangular portion of the week (this requires Javascript).

Figure 7.13: Schedule edition

The differences between the standard and professional versions lies in the number of supported boot mode per schedule.

7.5.2 Add schedule

Clicking the 'Add' button in the nodes management page (as shown in figure 7.12) will create a new schedule as a copy of the one named 'default'.

7.5.3 Copy schedule

Select the schedule you want to copy by ticking its check-box, then click the 'Copy' button. A new schedule named 'Copy of . . .' will be created.

7.5.4 Delete schedule

Select the schedules to delete in the management page (as shown in figure 7.12), then click the 'Delete' button. You will be shown with a confirmation page to commit the

deletion.

7.6 The 'Boot modes' menu item

This menu (see figure 7.14) lists the boot modes available as well as their types (see below).

Figure 7.14: boot modes management page

7.6.1 Local boot mode edition

This is a special boot mode (not configurable) which tells a machine to boot locally. It cannot be edited or added since, there is, indeed, only one way to boot a local hard disk (think singleton design pattern).
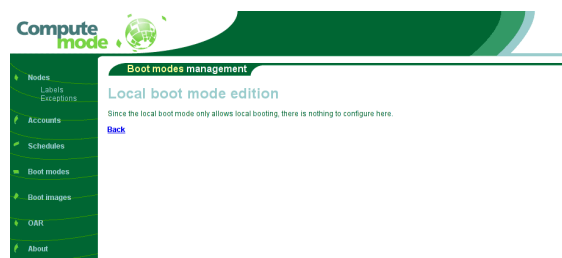
Figure 7.15: Local boot mode

7.6.2 Raw PXE boot mode edition

This is a special boot mode which is only available in Icatis extensions. For instance, it may help to start network floppy disks or provide other similar features and is aimed at professional users.

Regarding the professional version, you have currently to put the file on the server by separate means (scp) - an optional upload file configuration may be shipped later.

	free version: not available
	professional version

Figure 7.16: RAW PXE boot mode edition

7.6.3 ComputeMode boot modes edition

This mode is editable but some parameters may be slightly tricky - feel free to ask for help.

If you are altering an existing ComputeMode boot mode or creating a new one from scratch, you have to know that a few options are mandatory namely MASTER and NFSROOT.
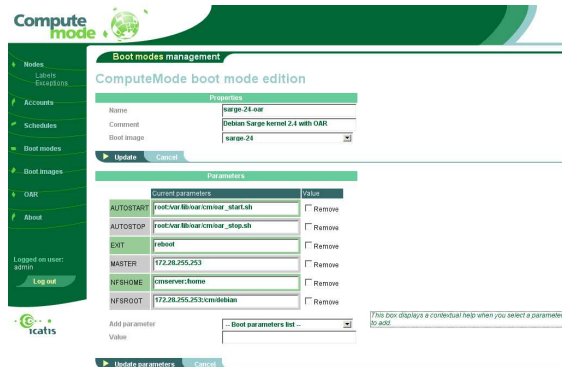
Figure 7.17: ComputeMode boot mode edition

7.7 The 'Boot images' menu item

This menu describes which Linux boot images may be booted and with which parameters it should run (see figure 7.18).
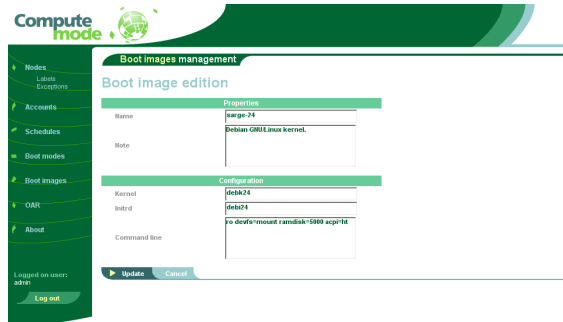
Figure 7.19: Boot image edition
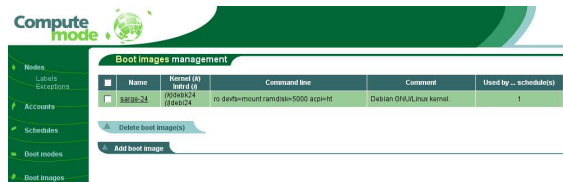
Figure 7.18: boot images management page

7.7.1 Adding or editing a boot image

You will be shown with a screen similar to the one in figure 7.19.

A few informations may be given regarding specific fields:

- name and note: should be meaningful strings for humans

- kernel: indicates the name of the kernel file in /var/lib/tftpboot/PXEclient/

- initrd: must be the name of the initrd file in /var/lib/tftpboot/PXEclient/

- command line: must be options given on the kernel command line

7.7.2 Delete boot images

Select the boot image to delete in the management page (see figure 7.18) then click the 'Delete' button. You will be shown with a summary and a confirmation button.

7.8 The 'OAR' menu item

Some statistics regarding OAR batch system may be shown in this page.
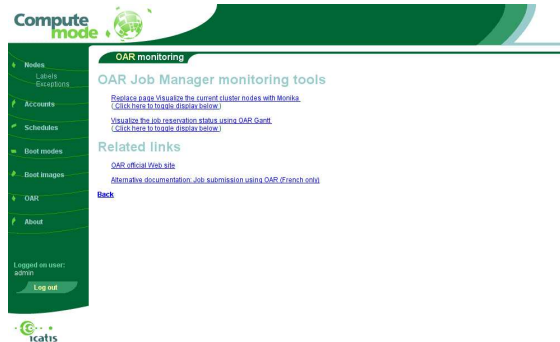
Figure 7.20: OAR statistics page

7.9 The 'About' menu item

This page holds nothing special but a list of the components used but for the sake of completeness, here is the mandatory screen shot.
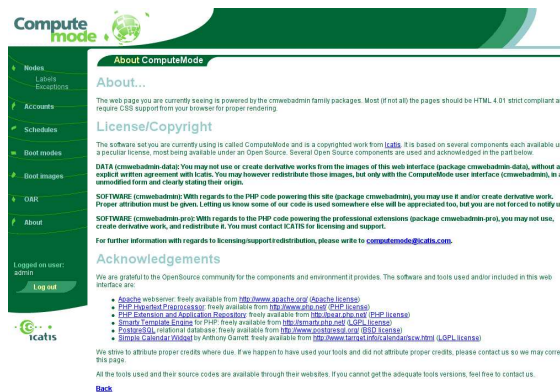
Figure 7.21: cmwebadmin about page

Chapter 8

Client nodes OS execution

Several solutions are available to add nodes to your CM grid.

8.1 Native node by rebooting

The simplest way to add client nodes is probably to boot the nodes and enable PXE: the CM server will then detect the PXE request and add the new machine under the name U_XX-XX-XX-XX-XX-XX where XX-XX-XX-XX-XX-XX is indeed replaced by the MAC address. The newly detected machine will have the default label 'Unconfigured' applied.

8.2 Virtualized nodes

Another means to test CM is to start a virtual machine (such as VMware Player) and enable PXE so that it may boot on the CM server.
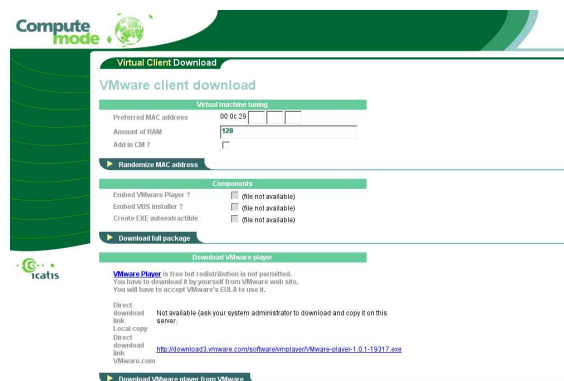
Figure 8.1: CM virtual machine for VMware Player download page

CM provides a page to download and register easily such a configuration package for VMware Player. Some files may not be redistributed for licensing issues - but you can do it by yourself - so you will have to do a bit of work if you want to enable most options available here. If you are interested in doing this, then please read section B on page 77.

8.3 Screen saver running virtualized nodes

The idea there is basically to install in Windows a new screen saver component. When the machine has been idle for a while, instead of starting fancy graphics and animations screen saving, it will simply darkens the display and launch a virtual machine which will network boot through PXE on the CM server.

The screen saver is currently only available for the professional version. Please contact Icatis for further information (see section 13 on page 74).

Chapter 9

As a CM administrator, how do I. . .

This chapter aims at giving quick recipes to solve common issues for an administrator.

9.1 . . . change CM administrator's password ?

1. Click on the 'Users' menu item.

2. Click on the account name you want to edit.

3. Enter a new password (it will not be shown).

4. Click the 'Update' button.

Note: The new password is enabled at once.

9.2 . . . reset cmwebadmin administrator's password ?

This should never happen since, as everybody knows it, nobody has ever forgotten a password :-)

This procedure is slightly annoying but as it's a last resort solution.

1. Log in as 'root' on your cmserver (ssh, console).

2. Become the `postgres` user:
   `su - postgres`

3. Launch the SQL command line client:
   `psql -U cmu CMDB`

4. Enter the password CMDB password (see section A on page 76)

5. Type:
   `UPDATE users SET password = '42A0ASfCJSzOg' where login='admin';`

6. Type:
   ```
   \q
   ```

7. Type twice (one to exit the `postgres` user shell, one to exit the root shell):
   ```
   exit
   ```

Note: The new password is '`icatis`' (without the quotes).

9.3 ... register a node ?

See next section.

9.4 ... add a node ?

1. Click on the 'Nodes' menu item.

2. Click on the 'Add node' tab.

3. Fill the form.

4. Click the 'Add' button.

9.5 ... unregister a node ?

See next section.

9.6 ... make a node disappear ?

1. Click on the 'Nodes' menu item.

2. Click on the node(s) you want to delete.

3. Click the tab 'Delete node(s)'

Note 1: if you deleted a node, it may appear again in the future if it boots through PXE and auto-detection is enabled.

Note 2: if you do not want to see detected nodes, then choose to show only the nodes configured in the default screen.

9.7 ... change the default label used in the nodes management page ?

Edit `config.php` and replace the line reading:

```
$GLOBALS['NODE_MANAGEMENT_DEFAULT_LABEL_ID'] = 0;
```

by

```
$GLOBALS['NODE_MANAGEMENT_DEFAULT_LABEL_ID'] = <XXX>;
```

where <XXX> is among:

- 0: for all the nodes
- -1: for all the configured nodes
- 1: for the nodes in 'Quarantine'
- 2: for the nodes 'Unconfigured'

Note 1: getting a label ID different outside of -1 .. 2 is currently not supported

Note 2: for the record, you may try getting it by looking at the hyper-links generated by the label management pages (for instance `http://.../cm/main.php?...&labelid=42` - note well that you use this on your own and this is not supported)

9.8 . . . cope with an owner who no longer wants to be part of ComputeMode ?

See next section.

9.9 . . . black list a machine ?

See next section.

9.10 . . . put a machine in 'Quarantine' ?

1. Click on the 'Nodes' menu item.
2. Select the node you want to blacklist.
3. Click on the tab 'Apply label'
4. Choose the 'Quarantine' label.
5. Click the 'Apply' button.

CM will now act as if it were disabled for this host that is:

- if this nodes boots through with PXE, it will tell it to boot in local mode
- if a CM agents, asks what to do, it will never give an order (reboot or other) and will simply acknowledge the request

9.11 . . . remove machines from 'Quarantine' ?

See next section.

9.12 . . . remove machines from 'Unconfigured' ?

1. Click on the 'Nodes' menu item.
2. Select the node on which you want to get out of 'Quarantine'
3. Click 'Remove label'
4. Select the label named 'Quarantine'
5. Click 'Remove label'

Note: Though they have special behaviors, these are labels so they can be removed as any other label.

9.13 . . . disable auto 'Quarantine' ?

See next section.

9.14 . . . disable node automatic inhibition ?

To disable automatic 'Quarantine' also called 'node inhibition', edit the CM configuration file `config.php` (which should be located in `/var/www/cm/`).

In this file, find the line reading:

```
$GLOBALS['ALLOW_INHIBIT'] = true;
```

Replace it by :

```
$GLOBALS['ALLOW_INHIBIT'] = false;
```

9.15 . . . enable/disable node automatic detection ?

Disabling this feature requires your editing the cmwebadmin configuration file. In `config.php`, edit the line reading:

```
$GLOBALS['MODULE_NODE_AUTODETECTION'] = true;
```

and replace it by :

```
$GLOBALS['MODULE_NODE_AUTODETECTION'] = false;
```

9.16 ... enable/disable owner/user's pages ?

Disabling this feature requires your editing the cmwebadmin configuration file. In `config.php`, edit the line reading:

```
$GLOBALS['USER_ACCESS'] = true;
```

and replace it by :

```
$GLOBALS['USER_ACCESS'] = false;
```

9.17 ... force reboots at specific times ?

See next section.

9.18 ... use an agent ?

Such a functionality requires a bit more work and the use and installation of a agent. The work of the agent is basically to ask the server what he should do next - that is reboot the server into computation mode or continue as if nothing.

There are currently 2 agent versions : a Windows and a Linux/Unix one.

9.18.1 Windows agent
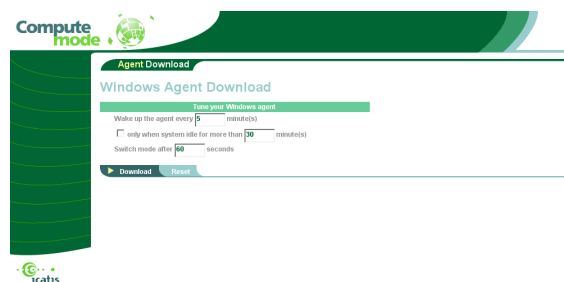
Figure 9.1: Windows Agent download page

A customized agent may be downloaded directly from a page hosted on cmserver. The figure 9.1 illustrates the customizations which could be done, namely:

- at which interval the agent has to be woken up

- should it be woken up only when the system has been idle for a certain delay

- after how much time should the mode switch should be forced

9.18.2 Linux agent

The Linux agent page is currently not shipped with the CM version. This will be fixed in a next revision. Feel free to contact us if you need one (see section 13).

9.19 . . . handle specific bank holidays ?

CM does not support calendar exceptions such as bank holidays.

A workaround consists in editing the calendar prior to event, then restoring it after.

Let's assume the standard 'evening and week end schedule' and that next Monday happens to be a bank holiday. The trick is to edit the schedule on the Friday to enable CM during all the Monday, then on Tuesday, restore the previous Monday schedule.

9.20 . . . upgrade the operating system ?

9.20.1 Server

The server is a classical Debian OS so you may use Debian standard tools (apt-get, aptitude, dpkg, . . .)

Note: some packages are more sensitive than others to upgrade (kernel, web, dhcp, nfs, etc.) that's why some are put on 'hold' in dpkg.

9.20.2 Client image OS

To update a client OS (let us say the Debian image), you have first to do a : `chroot in /cm/<image_name>/orig/`

A sample update script is given in `/usr/local/sbin/updateall.sh` but you should only apply it if you feel very confident with messing with the system - wreaking havoc by updating is quite easy: if disk space is not an issue, you should really backup the `/cm/<image_name>` folder.

As for the server, some packages are in 'hold' state in dpkg and should be updated with care (such packages are related to: kernel, perl, sudo, etc.)

9.21 . . . login as root on the client nodes ?

For security reasons, logging in as root is disabled on the client nodes on the console. Yet, this can be reactivated by editing the client image OS.

Logging in as root is allowed through ssh from the root account on the CM server.

9.22 . . . add a grid user account ?

See next section.

9.23 . . . add an Unix account ?

1. Log in as root

2. Execute:
   `adduser <the_login>`

3. Fill the requested information

4. If you are using NIS - this is enabled by default set up - you have to rebuild the NIS maps, which may be obtained by typing:
   `make -C /var/yp`

5. If you want to let this user submit jobs through OAR, you have to add him/her to the oar group, type:
   adduser <the_login> oar

9.24 . . . use an extra NFS server for home, system or data ?

This can be done by configuring the ComputeMode boot mode parameters.

Thorough explanations will be given in a future version of this manual - meanwhile, feel free to contact us (see section 13 on page 74).

9.25 . . . change the server public IP ?

This is currently done the Debian way:

- fire up a text editor and open, as root, `/etc/network/interfaces`

- go to the `eth0:0` stanza

- replace the current line with '`address x.y.z.t`' by the new address you want to use

- if necessary, edit the 'netmask' and 'gateway' lines

For further information, you may want to read `interfaces(5)` man page.

Chapter 10

ComputeMode Grid Manager architecture

This chapter will try to give view a general overview of what's happening behind the scenes.

10.1 What happens during the PXE boot ?

When the client boots, it broadcasts a DHCP request with a flag telling it wants some PXE.

The CM server sees this request and replies giving a temporary IP address and telling the client to fetch a pxelinux file from a peculiar TFTP server.

The client's PXE ROM then contact the TFTP server, fetches the file then executes it. The PXE part run by the ROM is now over.

10.2 What role does PXELinux play ?

PXElinux then tries to contact the TFTP server it knows and requests a configuration file whose name is based on the MAC address of the network card used.

That is when the CGI patch applied to the TFTP servers plays its role: if a specific pattern is matched, then, a CM backend web page is accessed and dynamically creates the pxelinux configuration file based on what the CM server currently knows (time, day, load, labels, etc).

pxelinux then receives this configuration file and acts according to its contents, which may basically of the two following flavors:

1. local boot or,

2. chain / boot something else (Linux kernel, image, floppy image, other boot loader, etc.)

In the case it is a CM OS image, several options are passed by means of kernel command line options - this will be explained in the next section.

The IP address which was obtained during the PXE negotiation described above is now released.

10.3   What role does a CM image has to play ?

Now that the Linux kernel is booting with its attached initrd, the following events take place:

- try to find and load a valid network card driver: all the previous network communications happened through the UNDI network stack, which stands for Universal Network Device Interface and is basically a low-level, low-performance interface available for early network communication - it may not be used by modern OS

- try to get a DHCP address from the CM server

- set up the disk-less Linux distribution by using a mix of NFS mounts and RAM disks

- start the distribution processes and services

10.4   How does the `/cm/<distrib_name>` file hierarchy works ?

Basically this folder contains an almost complete OS image which has been adapted for CM. 'Almost' as to be complete an image has to contain also: kernel and initrd (which both go into `/var/lib/tftpboot/PXEClient/`) and the adequates boot image and boot mode in cmwebadmin.

ComputeMode provides a centralized distribution system for client nodes. As a result, deploying (adding or upgrading) an application on the nodes is easily achieved by modifying the ComputeMode Server's network boot system repository.

Note: Altering these folders should only be performed by experienced Linux administrators.

The file system for client nodes is, currently, shared by means of the NFS protocol. For simplicity most images are located in `/cm/` - this can be changed if you update the NFS exports list as well as the boot image configuration.

ComputeMode uses currently as default a Debian-based distribution, so the directory is logically: `/cm/debian`. In this folder, several subdirectories with specific names will be found. Let's review each one of these:

- `orig/`: contains some kind of golden system image, unaltered - it was built by copying all the files of a fresh Linux installation and has been unaltered for network boot. This folder should be managed by using the distribution preferred method (dpkg/apt-get for Debian-based distribution, rpm/urpmi/yum for RPM-based ones, . . .)

- `patch/`: contains the data (configuration files, replacement and extra binaries) required to transform the standard base Linux system into a ComputeMode network boot system. As the 'patch' name may suggest it, a mechanism similar to the one for source files is used

- `rules`: is a configuration file describing how the 'patch' elements are to be used over the 'orig' at the startup of each client node. It describes the copy and link operations to be applied using a few keywords such as: `copypatch`, `copyorig`, `linkpatch`, `linkorig` and `skip`.
  This is an advanced functionality and will be discussed in a later revision - feel free to drop by the mailing list if needed.

- `utils/`: contains CM-specific scripts required to perform the network boot specific operations. The `buildroot.sh` script parses the 'rules' configuration file - other commands will perform specific operations for client nodes (start, registration, stop)

As a result, deploying an application to a given Boot Mode is achieved by adding necessary files to the 'orig' directory (possibly using the distribution's packaging facility) using a `"chroot"` command to have a consistent system view. Then, you have to check that the installed files are not overridden by the 'patch/' files. Such conflicts can be solved manually using the 'rules' file, but for most well-written applications (i.e. not needing a read/write access to the system installation when running), such do-it-yourself should not be necessary.

Note: upgrading some system parts (especially libraries) while they are used by client nodes may cause problems (crashes)

According to the rules files, and how your directories are exported, client nodes will be able to use the newly installed software at their next boot.

10.5 How do exceptions work ?

This is an advanced feature and will be explained in a later revision of this document - feel free to contact us (see section 13 on page 74).

10.6 Which processes run on the CM server ?

Please see section 11.2 on the next page.

Chapter 11

Security walk-through

This chapter aims at summarizing the processes running, and the security implications and mitigations of running a CM within your site.

11.1 Security concerns

ComputeMode is not designed to be run in a hostile environment in the sense where, if you want really want to mess with the nodes within your network almost any workstation can steal the identity of another (in more or less difficult ways).

The basic idea which is the one often seen within enterprises is that:

- security is paranoid at the network borders : nothing unsafe should get in or get out from the intranet, which most likely implies using firewalls and authenticated web (and other protocols) proxies

- within the LAN, hosts can safely communicate with one another

- employees are not evil on purpose

On the contrary, if you have hostile people whose sole role is to annoy others by sabotaging their work, it is guessed you should rethink your hiring process !

11.2 Services running on the CM server

To know which ports are open on you CM server, you can execute, as root:

```
netstat -lnp
```

Several services depends on portmap and hence do not have fixed port numbers: to find out which ports are used, you may type on the server:

```
rpcinfo -p localhost
```

11.2.1   Services required by CM server

Several services and servers are running on a standard CM server - some may be added or removed but the following ones are currently needed for proper functioning.

Service	Daemon name	Ports & protocol	Comment
DNS	bind	TCP/53 + UDP/53	DNS
		1 dynamically allocated UDP port	
		TCP/953 (localhost only)	control channel
ssh	sshd	TCP/22	
web/http	httpd	TCP/80	
web/https	httpd	TCP/443	
web	httpd	TCP/943 (localhost only)	control channel
portmapper	portmap	TCP/111=sunrpc and UDP/111=sunrpc	
NIS	rpc.yppasswd	(through portmap)	
	rpc.ypxfrd	(through portmap)	
	ypbind	(through portmap)	
mail	exim	TCP/25 on localhost	
syslog	syslogd	UDP/514	
NFS server	(none - kernel)	TCP/2049 and UDP/2049	
	rpc.mountd	through portmap	
	rpc.statd	through portmap	
NFS mounts	(none - kernel)	UDP system ports (< 1024)	
PXE proxy	pxe	UDP/4011	
TFTP server	in.tftpd	UDP/69	through xinetd
DHCP server	dhcpd	UDP/67 + ICMP	
PostgreSQL	postmaster	TCP/5432 (may be bound to localhost)	can be disabled

Table 11.1: Services running (and required by) the CM server

11.2.2   Services required by CM "friend" tools

ComputeMode Grid Manager which ships a few "friend" tools.

Currently there is only the OAR scheduler which offers a free task scheduling systems. This software is written in Perl and uses `mysql`, `ssh` and `sudo`. The client nodes are contacted when needed through `ssh` so no extra service is running there. Please check the table 11.2 for further information.

Service name	Daemon seen	Protocol/Port	Comment
OAR	perl	TCP/6666	
	mysql	TCP/3306	

Table 11.2: Services required by tools shipped with CM: OAR

11.3   Services running on client nodes

Client nodes may only be accessed from the CM server and they should all belong to the dedicated CM sub-network (by default: 172.28.0.0/16). All the services specified in table 11.3 are only accessible from the private B network used.

Service name	Daemon name	Protocol/Port	Comment
sunrpc	portmap	UDP/111 + TCP/111	tcp_wrappers
NIS	ypbind	through portmap	tcp_wrappers
NFS/status	rpc.statd	through portmap	tcp_wrappers
ssh	sshd	TCP/22	tcp_wrappers
syslog	syslogd	UDP/514	no filtering
NFS mounts	(none - kernel)	UDP system ports (< 1024) usually around UDP/800	kernel filtering
DHCP client	dhclient3	UDP/68	kernel filtering

Table 11.3: Services running on a client node

Most have their accesses disabled thanks to `inetd` and `tcp_wrappers`.

Note: Logging in as root is disabled on the client nodes on the console. Yet, this can be reactivated by editing the client image OS.
Root login is only possible from the CM server, through ssh - this can be disabled too by editing the client image OS.

11.4   Owner accesses

Owners may only access the CM server through a user page on the server. The scope of actions which can be done there is quite limited. They can:

- tell their machine is available by adding exceptions

- see what happened to their machines for the last few days

Authentication is currently voluntarily weak (IP-based) to simplify the task of users who want to let their workstation join the grid.

Note: if the CM administrator wants to disable this mode, the `config.php` file has to be edited (see section 9.16 on page 63).

Part V

Future Works & Contact

Chapter 12

Future Works

ComputeMode Grid Manager is being developed and used in several cluster and grid computing projects. The requirements and feedbacks given by users and administrators, as well as constraints (security, network architecture, . . .) help draw the directions and shape of ComputeMode Grid Manager.

Feedbacks will also help us choose which technologies and features to add to ComputeMode.

Of course, if you want to support this product, please consider purchasing the professional version or service for ComputeMode Grid Manager from Icatis SAS (see chapter 13 on the next page).

Chapter 13

Support & Contacts

13.1 Free support

Free support is provided on a best-effort basis by e-mail through the ComputeMode user mailing list.

Check the ComputeMode web site at:

```
http://www.icatis.com/
```

to find out how to subscribe.

The list preferred language is English but French is also supported.

13.2 Commercial support

Commercial grade support for ComputeMode Grid Manager is also available from Icatis SAS.

Please check our website at:

```
http://www.icatis.com/
```

Icatis SAS is a French company founded in 2004 as a start-up from a French Computer Science Laboratory (ID-IMAG `http://www-id.imag.fr/`). Our staff has previously worked for several years in the fields of High Performance Computing, cluster programming and administration and intensive data movement through several projects such as Ka-Tools, Mandrake CLIC, TopTools, I-Cluster project, i-cluster mainstream cluster (ranked in TOP500 cluster in 2001), NFSp, E-Toile, IGGI, . . .

Our core skills lies in distributed architectures (clusters, grids, network of workstations) and their administrations (Linux, Windows).

Our firm is able to offer:

- consulting and expertise in grid, cluster, application conversion, network architecture. . .

- CM support: installation, exploitation, specific developments. . .

- partnership in R&D: if you have interesting scientific applications, special needs, innovative uses of grid or cluster computing, . . .

Part VI

Appendices

Appendix A

Default passwords

Using default passwords is "evil" so please, change them once your setup is running smoothly. Note that the spelling may be slightly altered according to the keyboard you are using - on a QWERTY-US keyboard.

Component	Login	Password	Comment
UNIX account	`root`	`icatis`	change this a.s.a.p. with `passwd`
UNIX account	`guest`	`guest`	disable it when you are done: `passwd -l guest`
cmwebadmin	`admin`	`icatis`	change it (see section 7.4.1 on page 50)
PostgreSQL / CMDB	`cmu`	`cmupasswd`	change it in PG and in `config.php`
MySQL	`root`	`icatis`	change it in MySQL, in `config.php` and in `oar.conf`
OAR	`oar`	`oar`	change it in MySQL, in `config.php` and in `oar.conf`

Table A.1: CM server - default passwords

This appendix will be completed in a next version of this documentation: if you need help to change those passwords feel free to contact us (see section 13 on page 74).

Appendix B

Enabling extra-options in Virtual Client download page

This chapter tells how to enable most options in the virtual client page. You will have to perform several operations on your server, some may not be quite obvious.

This appendix will be written in a next version of this documentation: if you need help to configure this page, feel free to contact us (see section 13 on page 74).

Appendix C

Acknowledgements

C.1 Open Source Software

ComputeMode Grid Manager relies on several Open Source components, namely:

- the RDBMS used by cmwebadmin is PostgreSQL and is released under a BSD license
  `http://www.postgresql.org/`

- the web server used is Apache and is released under the Apache license 2.0
  `http://www.apache.org/`

- the interface is built above PHP4 + PEAR + SMARTY which are all released under the PHP license 3.0

- the Simple Calendar Widget is by Anthony Garrett and is released under the LGPL 2.1
  `http://www.garrett.nildram.co.uk/calendar/scw.htm`

- TFTPD-HPA is released under the GPL 2 license - this tool has been altered to support a CGI like mode (patch available on `http://www.computemode.org/`) and is known in CM as `cmtftpd-hpa`

- pxelinux which is released under a GPL 2 license
  `http://syslinux.zytor.com/`

- bind DNS server software (ISC) which is released under a BSD license
  `http://www.isc.org/sw/bind/`

- DHCP server software (ISC) which is released under a BSD-like license
  `http://www.isc.org/sw/dhcp/`

- OAR software (ID-IMAG) is released under a GPL license
  `http://oar.imag.fr/`

If some software you developed is being used and we forgot to mention it in this document, please tell us and we will fix this page in later revisions of this document.

C.2   Third-part software

This manual mentions several third-part software and trademarks. Here is the alphabetized list:

- Debian is a registered trademark of Software in the Public Interest, Inc.
  `http://www.debian.org/`

- Ghost is a product by Symantec Corporation
  `http://www.symantec.com/`

- Linux is a registered trademark by Linus Torvalds
  `http://www.kernel.org/`

- VMware Player, VMware Workstation are products of VMware, Inc.
  `http://www.vmware.com/`

FIXME:An Open Source (GPL) batch manager is shipping with CM and named 'OAR', yet other products such as Platform's LSF, OpenPBS, TORQUE or Sun Grid Engine are known to work with ComputeMode.