



Xen-OSCAR: OSCAR Testing with Xen

Geoffroy Vallee, Stephen L. Scott
Oak Ridge National Laboratory

Introduction

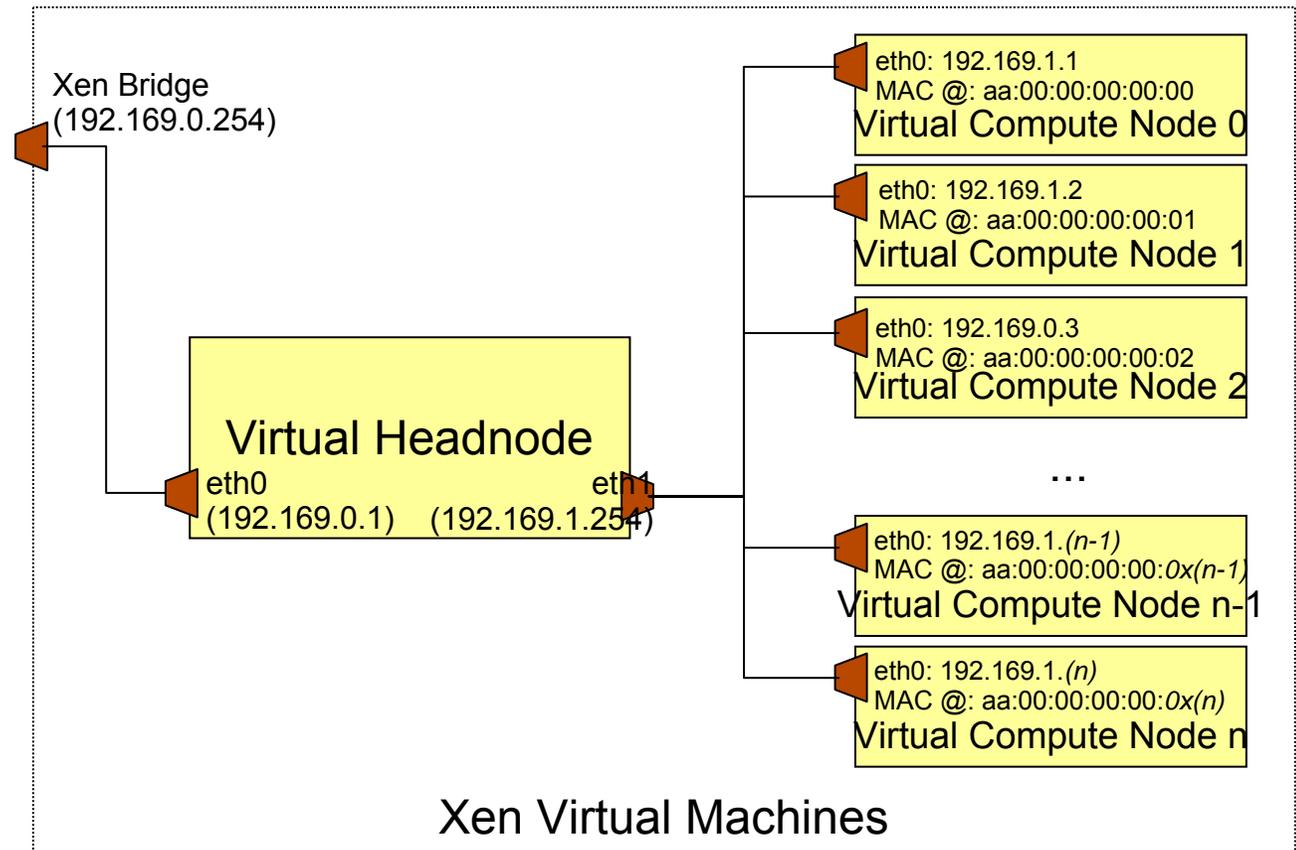
- OSCAR testing is difficult (for both OPKG and OSCAR core)
 - need to setup a “fresh” head-node every time
 - need to access a “real” cluster even during the development phase
- A possible solution: use a virtual cluster (VC) on a single physical machine
- At the same time, virtualization solutions become mature

OS Virtualization - Introduction

- Aims at executing several operating systems on a single host machine
- Virtual machines (VM) are isolated from one another
- Full virtualization vs. para-virtualization
 - full virtualization: the guest OS is not modified (not possible of the standard x86 architecture)
 - para-virtualization: the guest OS is modified to support virtualization (and to improve performances)

Virtual Cluster Architecture

- A single physical machine
- One virtual head-node and several compute nodes
- All network NICs are virtual
- Each VM has its own FS



Xen - Introduction

- Full virtualization w/ hardware support (VT/Pacifica)
- Para-virtualization solution for standard x86 architecture
 - modification of the kernel (“Xen architecture”)
 - new INTEL and AMD processors have now an hardware support of virtualization
- No complete boot sequence (the Hypervisor “start” the VM)
- Open source

OSCAR - Overview

- Suite to install and manage a Beowulf cluster
- Based on the imaging approach
 - images for compute nodes are created on the head-node
 - an image is used to install a compute node
 - compute nodes network boot
 - based on the MAC address a specific image is copied on the compute node
 - once installed, compute nodes reboot and are ready
 - OSCAR is currently based on SIS

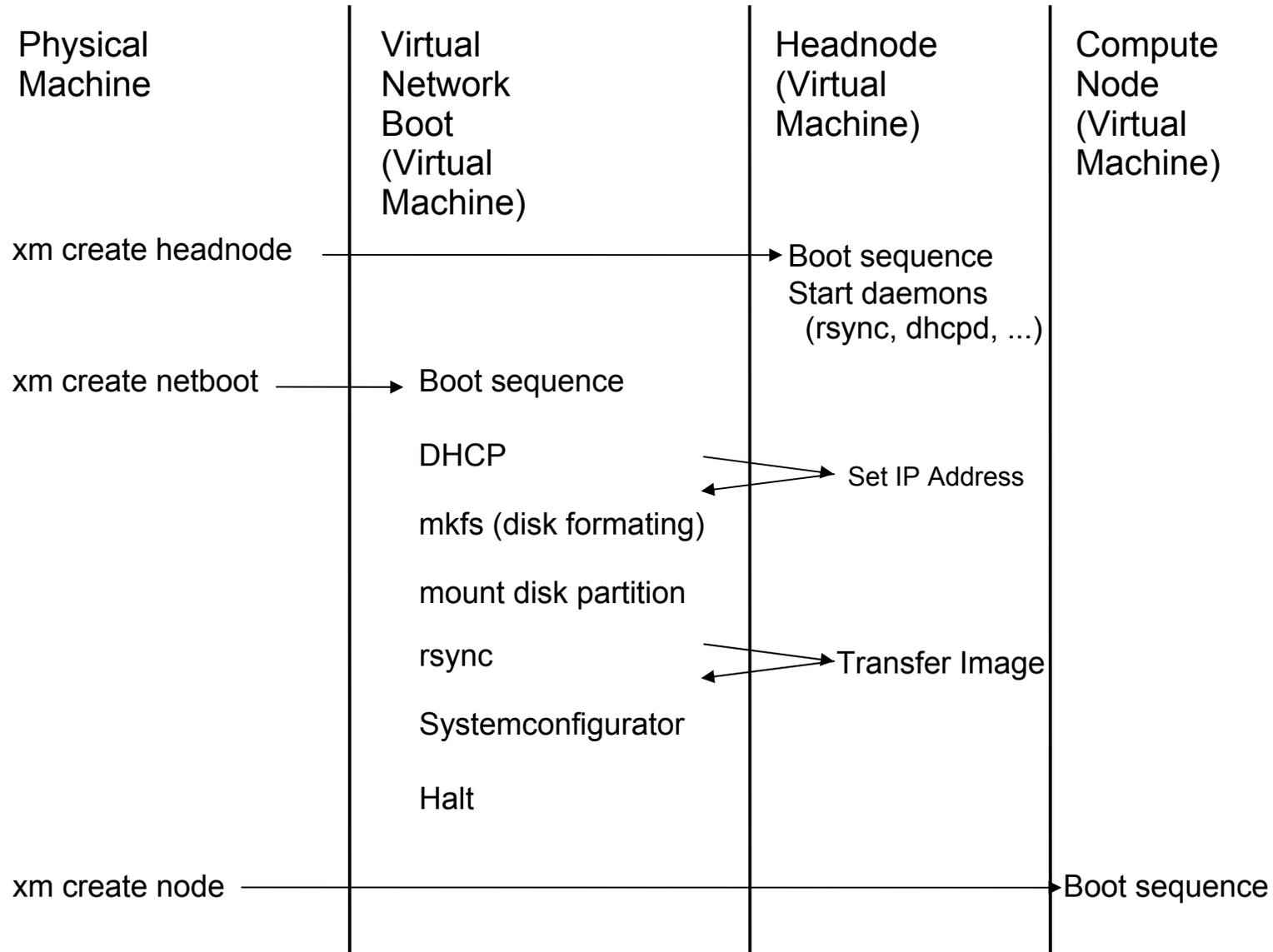
Xen-OSCAR - Issues

- How do you specify the virtual cluster node?
 - Users may want a specific Linux distribution on the head-node
 - Users may want to specify the number of compute node
- How do you setup Xen?
 - Xen needs configuration files
- How do you create Xen VMs?
- How do you simulate a network boot?
 - No complete sequence boot

SIS and a Virtual Cluster

- SIS installation process
 - the compute node boots through the network (PXE, using a CDROM or a floppy)
 - the compute node contact the head-node
 - based on the MAC address, an image is copied via *rsync*
- One solution: use a VM just like a CDROM during the installation phase
 - need a specific Xen image
 - need to specialize this image to match features

VC – Installation Process



Cluster Installation: real vs. virtual

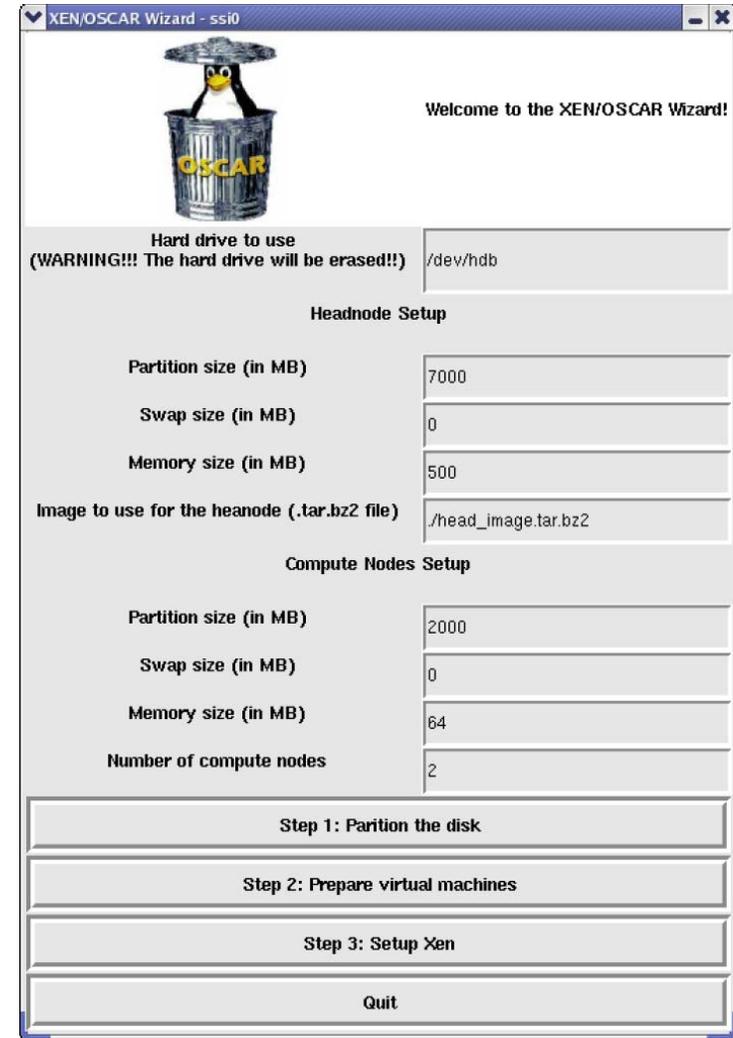
- Physical installation vs. virtual installation

	Real Cluster	Virtual Cluster
<i>Boot of the compute node</i>	PXE, floppy, CDRROM	Virtual Boot (specific VM)
<i>Installation method</i>	SIS – SystemInstaller scripts (rsync)	Specific scripts (based on SIS scripts)
<i>Post-configuration</i>	SIS – SystemConfigurator	SIS – SystemConfigurator
<i>Reboot</i>	Physical reboot	Stop of the VM for the virtual boot and start of the installed VM

- Not possible to use the standard SIS
- Not possible to test SIS w/ a VC

Xen-OSCAR

- Four steps
 - cluster specification
 - virtual FS initialization
 - Xen setup
 - installation of the virtual cluster
- GUI federates these steps



Step 1: Virtual Cluster Description

- Automatically install Xen
- Virtual cluster description
 - headnode: FS, memory size, image to use (distro + RPMs)
 - compute nodes: FS, memory size
- Xen-OSCAR provides images for headnodes

XEN/OSCAR Wizard - ssi0

Welcome to the XEN/OSCAR Wizard!

Hard drive to use (WARNING!!! The hard drive will be erased!!) /dev/hdb

Headnode Setup

Partition size (in MB) 7000

Swap size (in MB) 0

Memory size (in MB) 500

Image to use for the headnode (.tar.bz2 file) ./head_image.tar.bz2

Compute Nodes Setup

Partition size (in MB) 2000

Swap size (in MB) 0

Memory size (in MB) 64

Number of compute nodes 2

Step 1: Partition the disk

Step 2: Prepare virtual machines

Step 3: Setup Xen

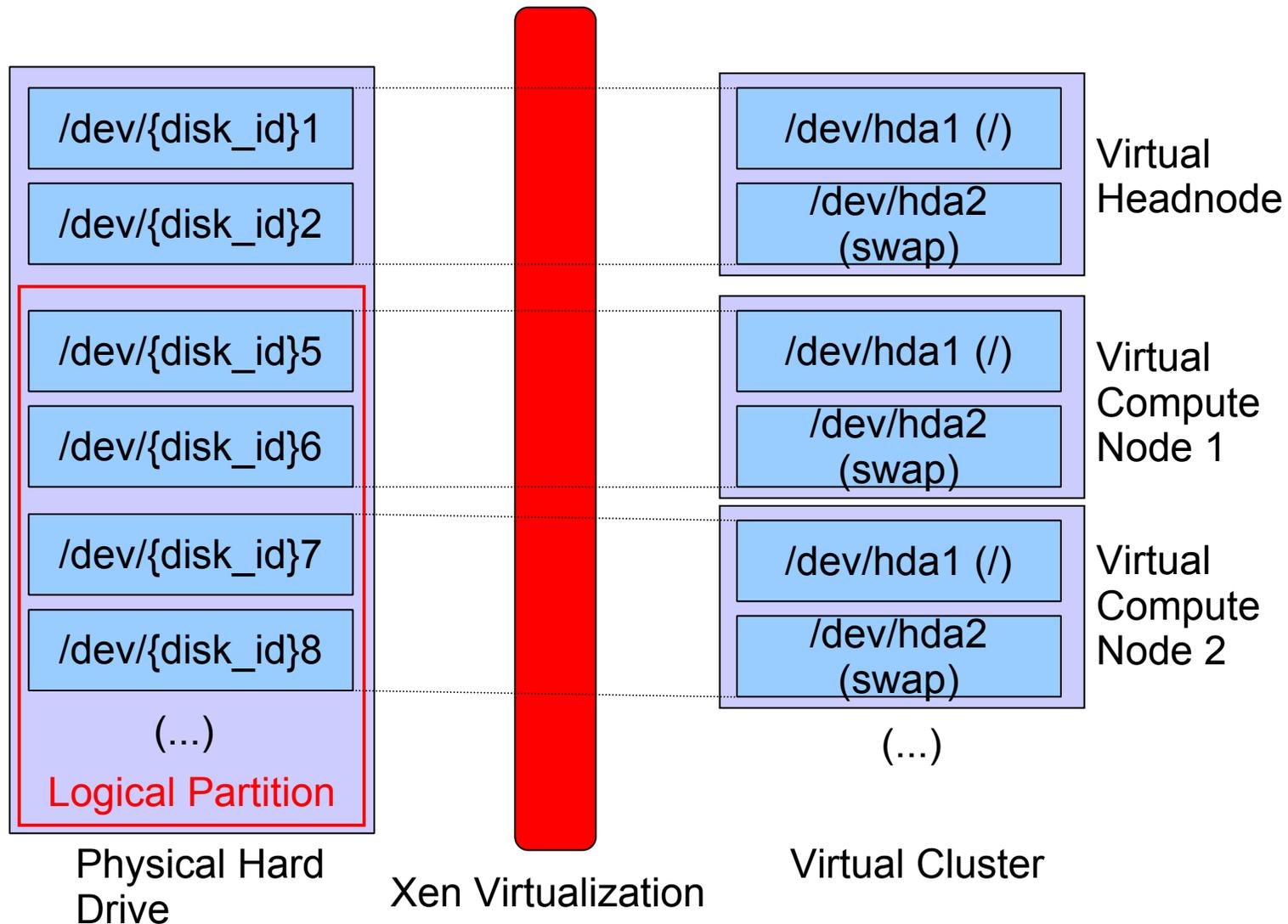
Quit

Step 2: Virtual FS Initialization

- Create file systems (one partition + swap for each VM)
- Install the system of the headnode
 - tar.bz2 image
 - mount locally the partition where the system has to be installed

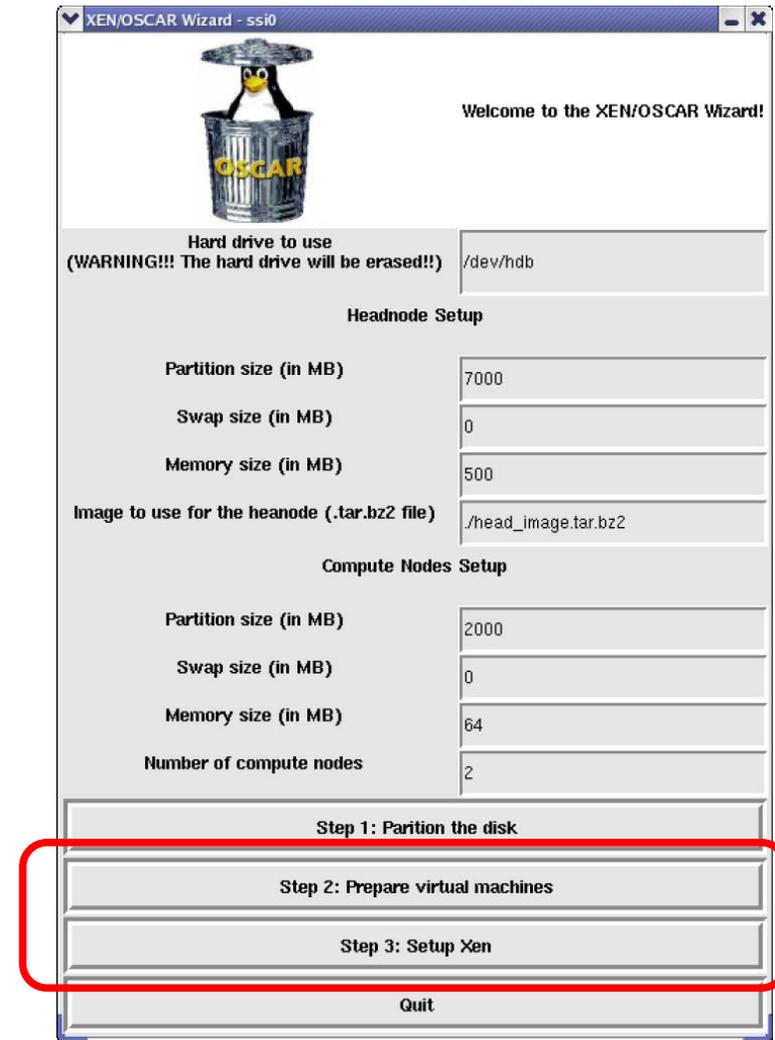


Step 2: Virtual FS Initialization (2)



Step 3: Xen Setup

- Creation of configuration files for Xen (headnode and compute nodes)
- Creation of images for the virtual boot of compute nodes
- Boot of the virtual headnode
- User may log in the virtual headnode and launch OSCAR



Step 4: Compute Nodes Installation

- Virtual boot of compute nodes (using the specific Xen image)
- Thanks to OSCAR, install compute nodes through the virtual network
- When installed, the virtual compute nodes will automatically shutdown (as during standard OSCAR installation)
- Compute nodes are ready, users may restart them

Conclusion

- Xen-OSCAR eases the installation of a virtual cluster
 - automatic Xen configuration
 - automatic creation of the headnode (thanks to headnode images)
- Ease OSCAR core and OPKG testing
- Limitations
 - do not provide a standard SIS installation process
 - do not allow to “simulate” another hardware architecture

Future Work

- Headnode images
 - support more Linux distributions
- Xen
 - integrate Xen-3 (currently based on Xen-2.0.6)
 - support more Linux distributions (currently Xen-OSCAR provides RPMs for FC4 only)
- Extend the GUI
 - virtual machine monitoring
 - Virtual node FS: use virtual block devices in addition of real partitions