

Virtualisation for Oracle databases and application servers

Carlos Garcia Fernandez

Luigi Gallerani

Anton Topurov

Carlos Garcia Fernandez at cern.ch

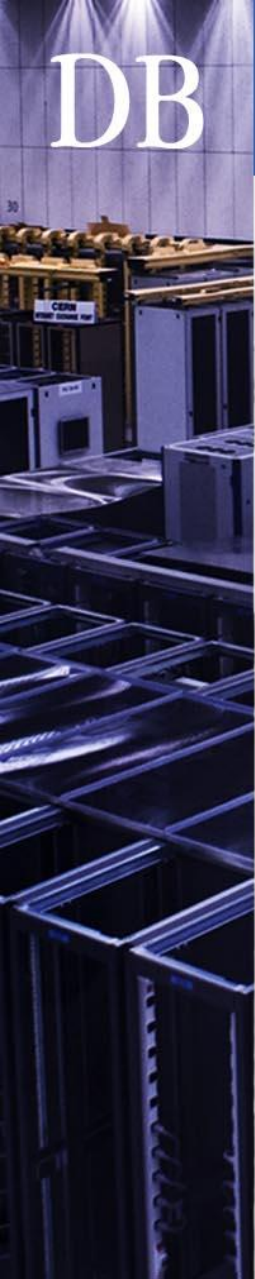
- What is virtualisation?
- Tests and Management of Oracle VM
- CERN infrastructure: CERN ELFms
- Integration steps of Oracle VM 2.1.5
- Update to version 2.2
- Guests installation
- Conclusion and future work

DB

Database Services

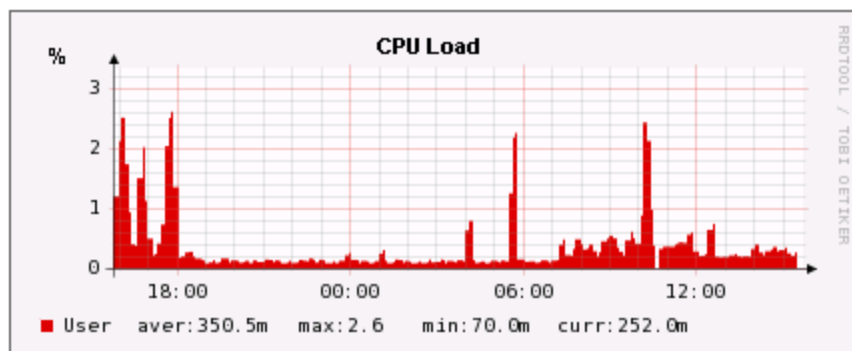
CERN IT
Department

Virtualisation



- **Virtualisation** is a term that refers to the *abstraction of computer resources*.
- **Paravirtualisation** is a virtualisation technique where the software interface to virtual machines is similar, but not identical, to that of the underlying hardware, thereby *requiring guest operating systems to be adapted*.
- **Hardware-assisted virtualisation** is a virtualisation technique that enables efficient *full virtualisation using help from hardware capabilities*, primarily from the host processors.
- **Oracle VM**: is the Oracle solution for server virtualization that supports both Oracle and non-Oracle applications. First version integrated at CERN 2.1.5

- **Growing number** of Oracle database instances and application server instances
- Need to **control** the necessary **resources** in terms of physical space, manpower, electricity and cooling.
- **Relocation** from one physical machine to another as needed
- Server **consolidation**: P2V transformation



Old machines:

- 8 GB RAM

New machines:

- 48 GB RAM



DB

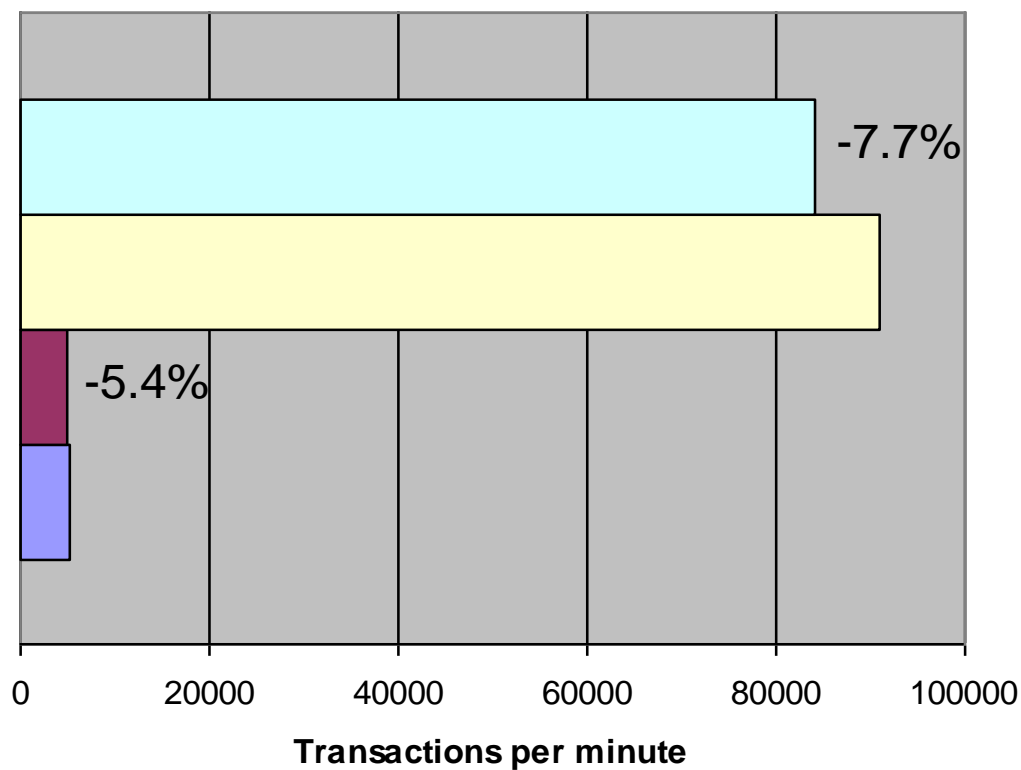
Database Services

CERN IT
Department

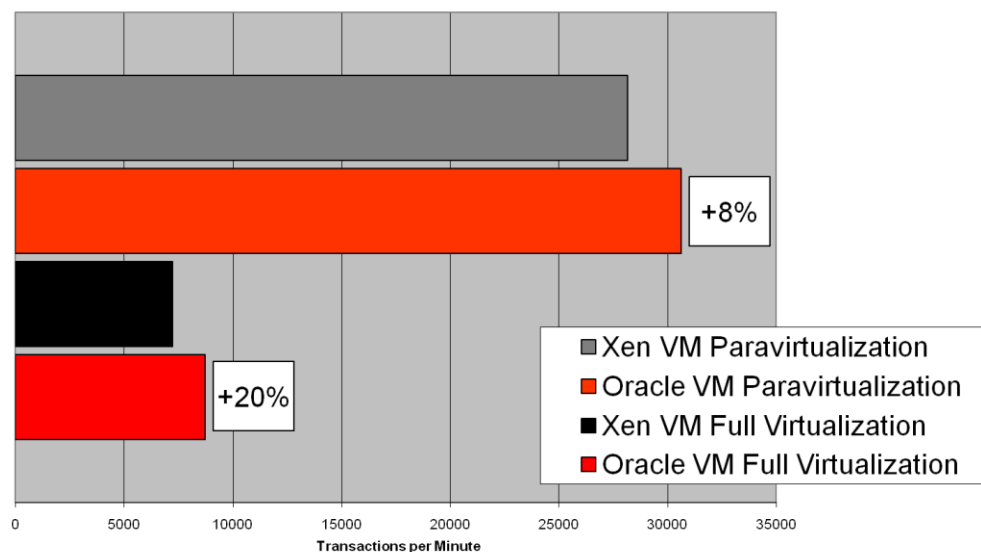
Performance and management of Oracle VM

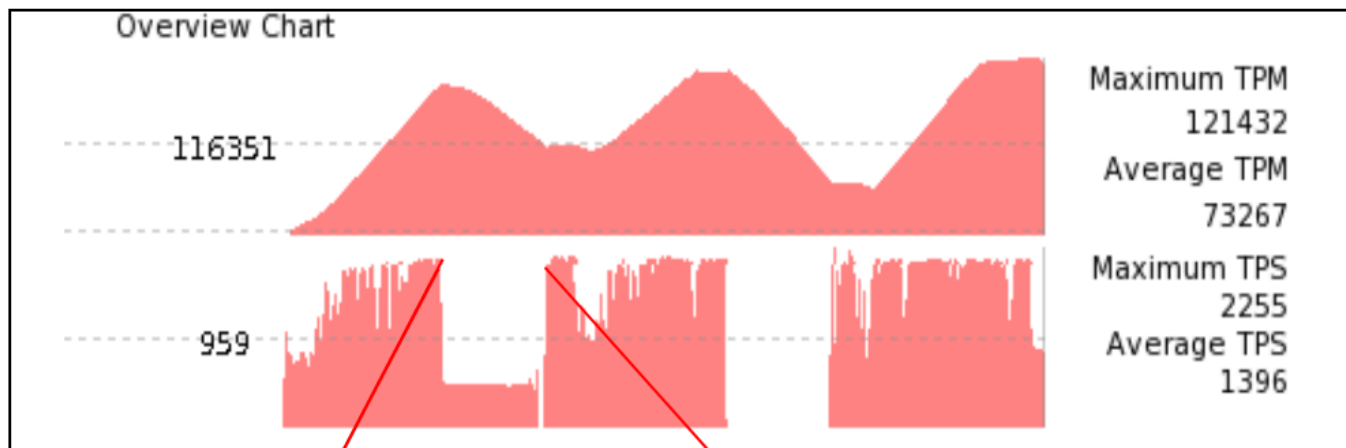
- Oracle Databases:
 - Oracle VM versus pure Xen
 - Paravirtualisation vs Hardware-Virtualisation
 - Live Migration
- Tests:
 - Performance tests using Swingbench
 - Stress tests
 - Order Entry tests
 - With and without load-balancing between the cluster nodes

Bare Metal to Oracle VM



- Performance comparisons of databases
 - Using Oracle VM
 - Using virtual machines on top of pure Xen
- Gained between 10% and 20% of performance in Oracle VM vs. pure Xen





Node 1

```
#xm list
Name      ID Mem VCPUs  State Time(s)
Domain-0  0  834  8      r----- 1773.7
virt04    8 4096  8      -b----- 517.4
```

xm migrate virt04 node2 --live

```
# xm list
Name      ID Mem VCPUs  State Time(s)
Domain-0  0  834  8      r----- 1785.7
migrating-virt04 8 4096  8      r----- 538.3
```

```
# xm list
Name      ID Mem VCPUs  State Time(s)
Domain-0  0  834  8      r----- 1851.5
```

Node 2

```
# xm list
Name      ID Mem VCPUs  State Time(s)
Domain-0  0  834  8      r----- 2410.8
```

```
# xm list
Name      ID Mem VCPUs  State Time(s)
Domain-0  0  834  8      r----- 2444.8
virt04    11 4096  0      -bp----- 0.0
```

```
# xm list
Name      ID Mem VCPUs  State Time(s)
Domain-0  0  834  8      r----- 2481.1
virt04    11 4096  8      -b----- 6.4
```

- Command Line interface
 - Used mostly all the time
 - Easily scriptable
- Oracle VM manager
 - Some incompatibilities with CERN network infrastructure
 - MAC address specified randomly with no possible modification
 - A pool couldn't be controlled by different managers
 - Need some work around to install in central DBs
 - Feedback has been sent to Oracle
- Oracle Enterprise Manager Virtualisation Pack
 - Same issues as Oracle VM manager

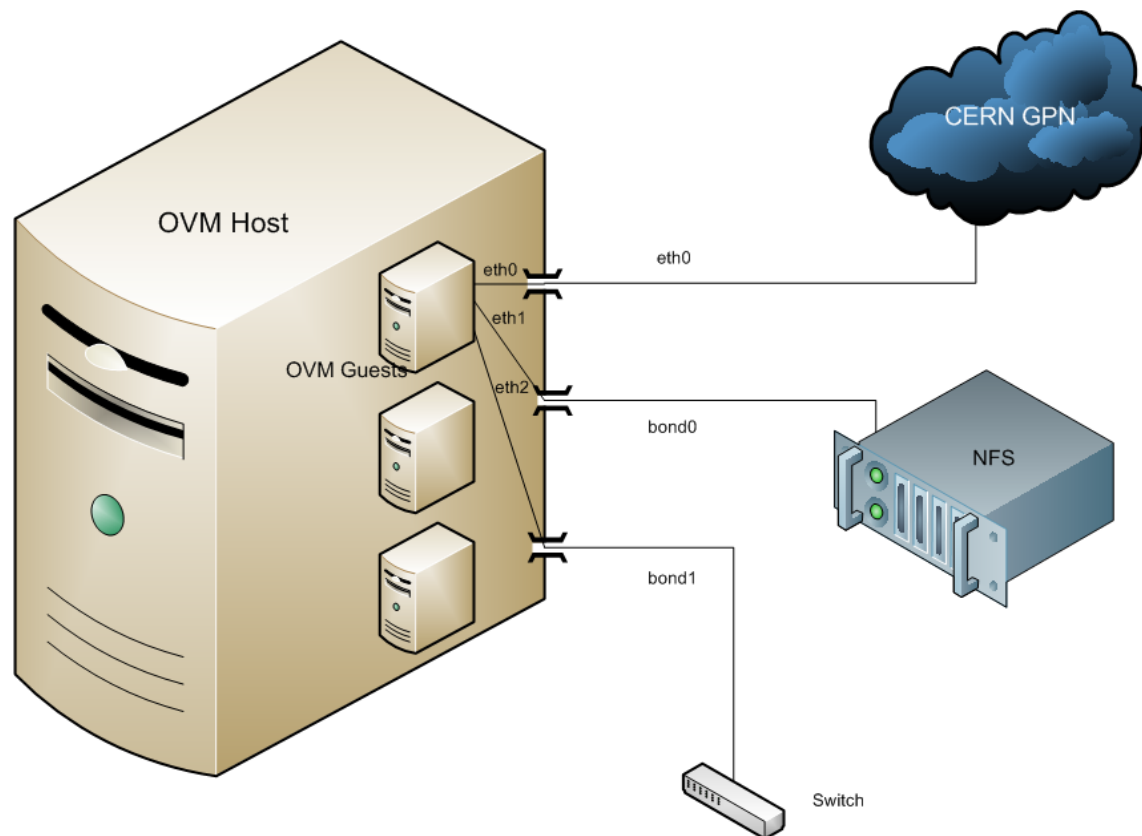
DB

Database Services

CERN IT
Department

CERN Fabric Management

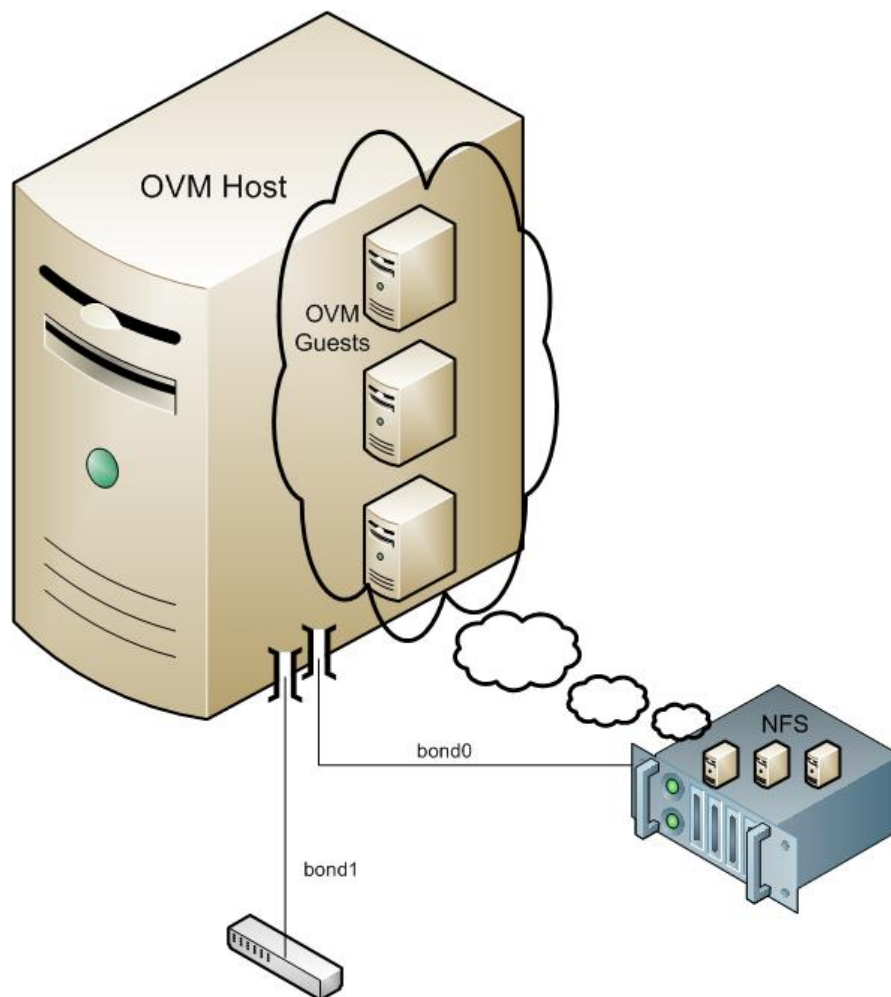




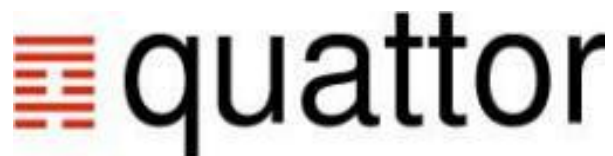
Advantages:

- Use of **same architecture** as the non-virtualised servers
- Eases the **migration** from physical environment to virtual (**P2V**)

- Images have to be placed in the same storage so we can **migrate** them
- NFS eases the increase of volumes **size**
- Use of the **same approach** as physical machines

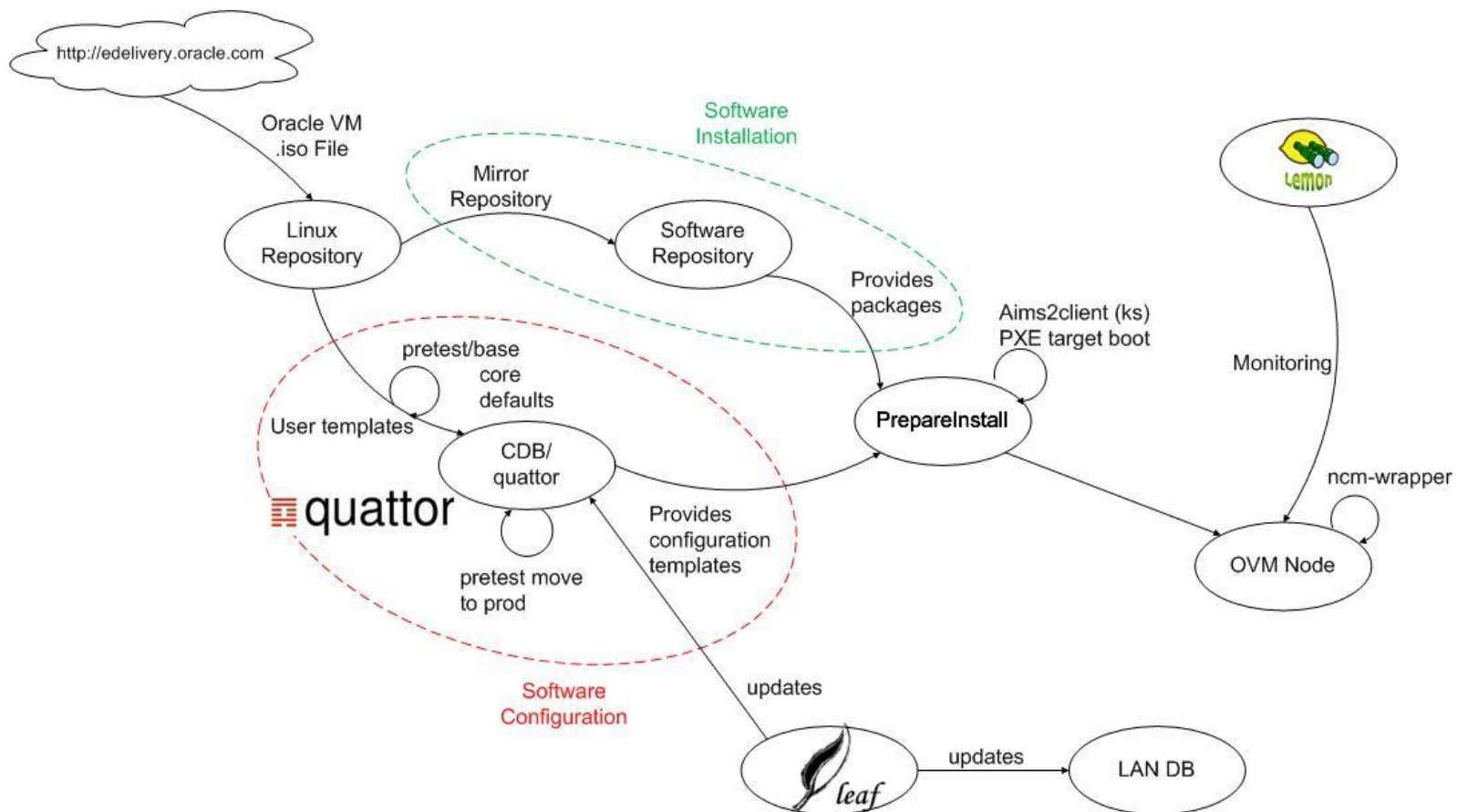


- All the systems running databases are being configured via **Quattor** including the database software installation.
- In order to reach the same level of management, we have to **use the central Linux installation service and Quattor** for OracleVM.
- All this process is done at CERN using **CERN ELFms**



- ELFms stands for Extremely Large Fabric management system
- It is divided in the following steps:
 - **Specifying configuration:**
 - Description in PAN language templates.
 - **Installing machines:**
 - Add DHCP entries and generate PXE configurations
 - Mechanism Kickstart/Anaconda.
 - **Configuring services:**
 - Done by ncm components (Node Configuration Manager)

- It's done with a perl script called **PrepareInstall**.
- It generates an **Anaconda/KickStart** file from the node information retrieved from CDB
- It prepares the **Sindex** service for download of sensitive files during installation
- It configures the **AIMS** installation service to:
 - upload the Kickstart file
 - configure and restart the dhcp server
 - select the PXE image to be used
- Once PrepareInstall has finished, you can **reboot** the machine to install **the node**.

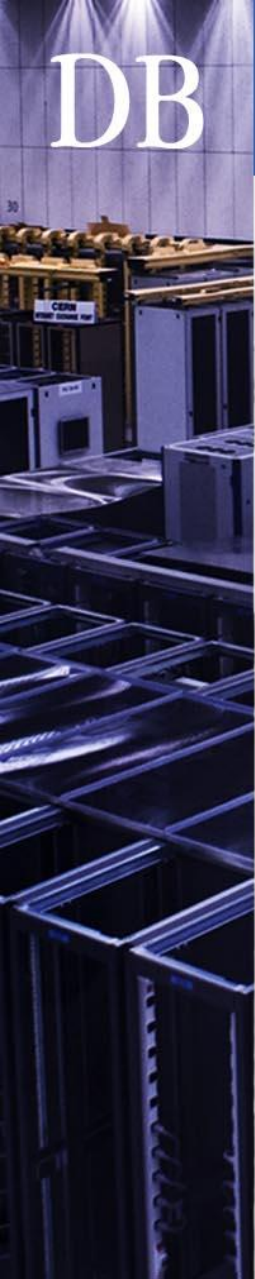


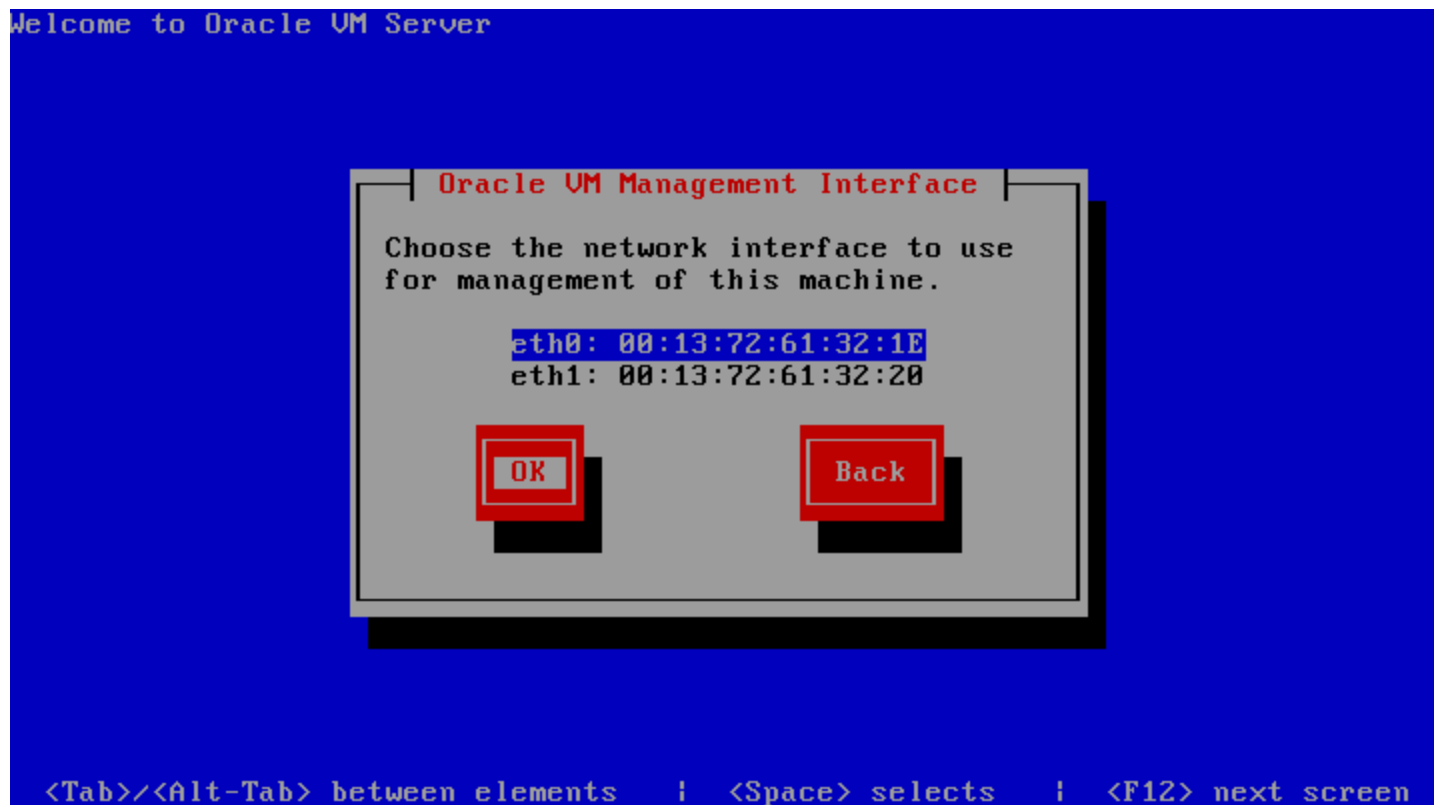
DB

Database Services

CERN IT
Department

Oracle VM integration steps





- Parameter to add to the kickstart file:
ovsmgmtif eth0



-Parameter to add to the kickstart file:
ovsagent XXXXXX

Note: XXXXXX is the password for the agent

- Modify bridge script to get the bond interfaces
 - /etc/xen/scripts/network-bridges

```
#!/bin/bash

dir=$(dirname "$0")

run_all_ethernets()
{
    for f in /sys/class/net/*; do
        netdev=$(basename $f)
        if [[ $netdev =~ "^eth[0-9]+$" ]]; then
            devnum=${netdev:3}
            $dir/network-bridge "$@" "netdev=${netdev}"
            "bridge=xenbr${devnum}"
        fi
        if [[ $netdev =~ "^bond[0-9]+$" ]]; then
            devnum=${netdev:4}
            $dir/network-bridge "$@" "netdev=${netdev}"
            "bridge=xenbo${devnum}"
        fi
    done
}

run_all_ethernets "$@"
```

- Mount the /OVS folder to store the images in a NFS
- Lost of connection problems with the NFS
 - OracleVM **Server Agent** automatically **mounts** the image folder, called "**Repository**", using information defined by the script `/opt/ovs-agent-2.3/utils/repos.py`
 - **Mount point** is `/var/ovs/mount/UUID`, where UUID is a hash unique descriptor for the NFS folder. UUID is managed by the server agent.
 - The `/OVS` folder is then automatically linked by OVM server agent to the UUID folder mounted.
 - The machine was configured to **manually mount** the `/OVS` in the NFS, but **OVS-Agent changed automatically the mount point** causing the lost of the connection.

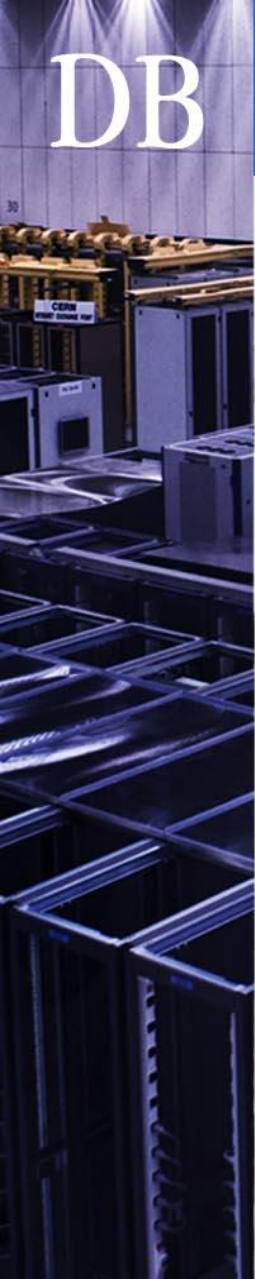
- **Provide** to the linux team the **packages** of this new version
- Create a **repository** for the new version
- Update the **default version** for the packages
 - Generated automatically in pretest running some scripts
 - Verify that we have all the packages we want and in the proper version
- Test the ***pretest*** installation
 - If working move it to *prod*

DB

Database Services

CERN IT
Department

Guest installation steps



- We wanted to have the most **transparent** VMs for the **users**
- We want to **avoid** having the **guest-host link** in quattor configuration, to ease live migration.
- We need to make some small changes in quattor templates:
 - Adapted the **cluster** templates for the hda disks
 - Adapted **RHES5** and **SLC5** as guest OS
- Selected “on-the-fly” installation vs. “golden images”
 - Better for quattor management “bare metal” images
 - Better to follow life cycle, patch installation

- Configuration file for xen:

```
name = 'virt06'  
builder = 'hvm'  
memory = 4096  
disk = [ 'file:/OVS/virt06/disk.img,hda,w' ]  
vif = [ 'type=ioemu,mac=00:16:3E:76:A6:AB,bridge=xenbr0', 'type=ioemu,bridge=xenbo0' ,  
        'type=ioemu,bridge=xenbo1' ]  
vfb = [ 'type=vnc' ]  
kernel = '/usr/lib/xen/boot/hvmloader'  
device_model = '/usr/lib/xen/bin/qemu-dm'  
root = '/dev/hda ro'  
vnc = 1  
vncunused = 1  
vnclisten = '127.0.0.1'  
apic = 1  
acpi = 1  
pae = 1  
#Boot parameter, n (network) for first PrepareInstall, cn (C drive+network) for next shutdown  
boot = 'cn'  
vcpus = 8  
serial = 'pty'  
on_reboot = 'restart'  
on_crash = 'restart'
```

- SOAP script to **add machine in the network database**

- Run **LEAFAddHost** to the virtual machine

```
LEAFAddHost --new_host=dbvrtd001 --cluster=webapps --  
  serial_number=1234 --rack=ek01 --hardware=ovm_00_00 -  
  os=rhes5 -arch=x86_64 --mac1=00:00:00:00:00:00 -  
  resource=des
```

- **PrepareInstall** the VM
 - Reboot it with “**Boot from network**” option
- Installation finishes
 - Reboot it with “**Boot from disk**” option
- **20 min** host added and **installed** from scratch

- Oracle VM and WLS-VE are great **technologies** we are keen to **exploit**
- **Long and hard work to integrate** Oracle VM in large scale environments
- We will replace **DEV** and **TEST** application servers for VMs by June
- Develop some scripted mechanism for operations with VMs (reinstall, start, stop, move)
- More news and experiences for **next HEPiX**

