CERN Virtual Infrastructure

Status update

Jan Van Eldik, Tim Bell
CERN – IT/OIS
3rd November 2010

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

Agenda

- What is the CERN Virtual Infrastructure?
  - CVI in numbers
- Service overview
  - SCVMM – Virtual Machine Manager
  - Integration with Cern environment
  - Templates
  - Hypervisors
  - User communities, Hostgroups
  - Backup
- Linux VMs
  - Integration Components, performance, issues

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

What is CVI?

- The CERN Virtual Infrastructure provides the underlying support for custom virtual machines in the CERN computer centre
- These VMs have a long-term lifetime of months/years
- Simple user kiosk for requesting a virtual machine in hours rather than days/weeks with physical hardware
- Contrasts with the batch virtualisation
  - Based on submitted workload and integrated with batch system
  - Lifetime is short so live migration less critical
  - Scaling requirements are much larger than with CVI

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

CVI in numbers

CVI service has grown spectacularly in 2010:

- Number of Virtual Machines: 680
  - Was 300 on January 1st
- Windows VMs / Linux VMs: 430 / 250
  - Was 260/40
- Number of hypervisors: 170
  - Was 40
- Many user communities benefit from CVI
  - Beams developers: 200 VMs
  - Quattorized Physics Services: 130 VMs

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

SCVMM

- Reminder: CVI is based on Microsoft's System Center Virtual Machine Manager
- Enterprise class centralized management
  - comparable to vSphere
- Rich feature set:
  - Allows grouping of hypervisors, with delegation of administrative privileges
  - VM migration, High availability
  - Checkpointing
  - PowerShell Snap-In for administration / scripting

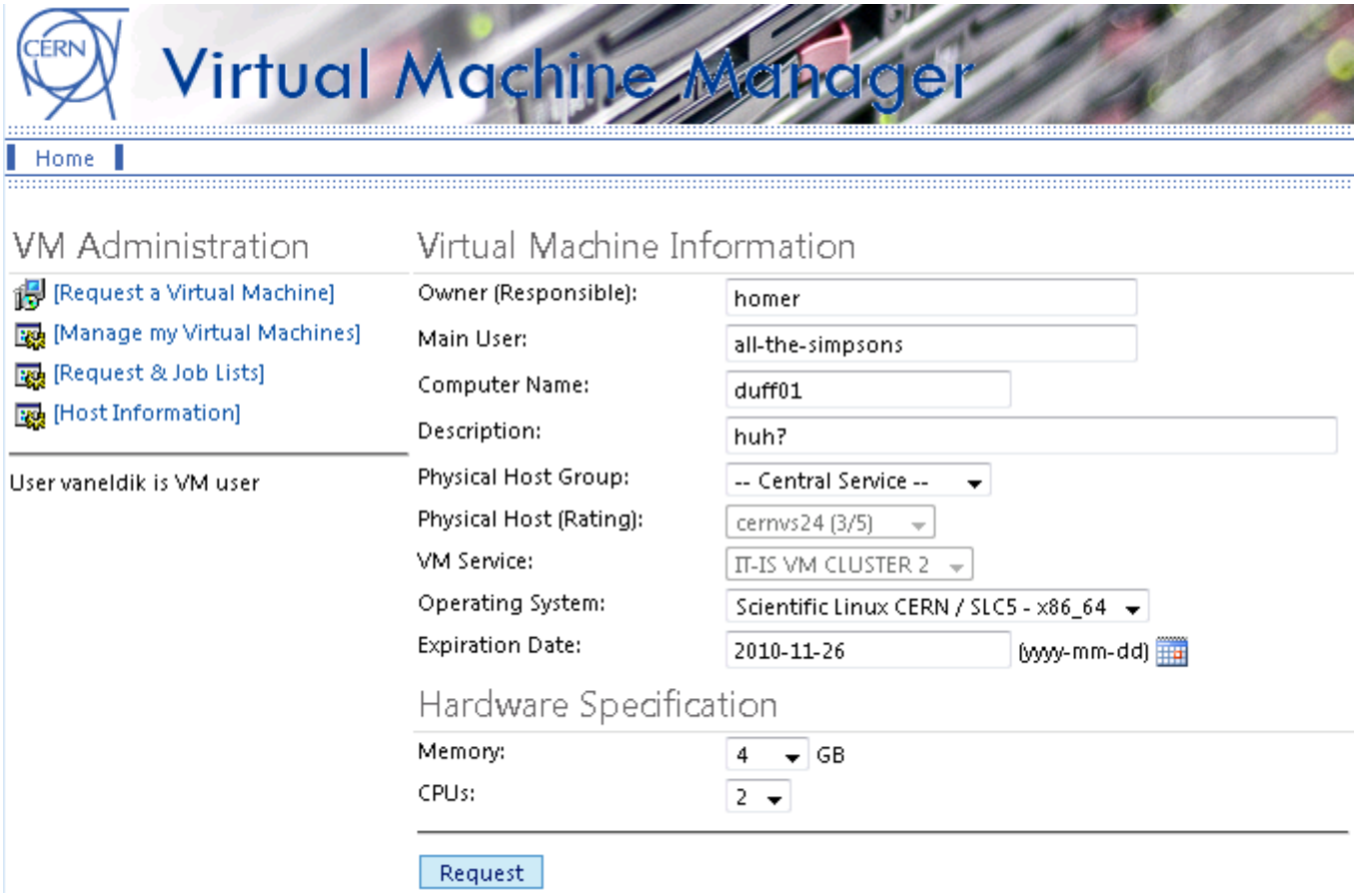
CERN addons - Interfaces

Web and SOAP interfaces

- Browser and OS agnostic
- VM creation, deletion, migration, ...
- Integrates SCVMM with Network Database (and hence with DNS) and Active Directory
- SOAP methods called by Web interface
  - and heavily used by Linux clients
- Interfaces recently re-implemented
  - addresses scalability issues
  - improves stability and performance
  - refresh of templates

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

CVI web interface



CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

Templates

VMs instantiated based on Templates:

- Windows
  - Win7 (x64, x86), WinXP, Server 2008 R2 x64, Server 2008 x86
  - Vista disabled (=hidden from view for standard users)
  - Note: SCVMM allows for automatic VHD updates, using Windows update service
- Linux
  - SLC5 (x86_64, i386)
  - SLC4 disabled

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

Templates - II

- PXE installations of Windows and Linux
- Work-in-progress: support for ‘customer-supplied' images
  - Required by several large customer groups
    - BE/CO, ETICS, EN department
  - Customers will be able to create templates from their customized VMs
  - These templates can then be used to instantiate more VMs
  - We can restrict availability of such templates to customer’s dedicated hypervisors

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

Hostgroups

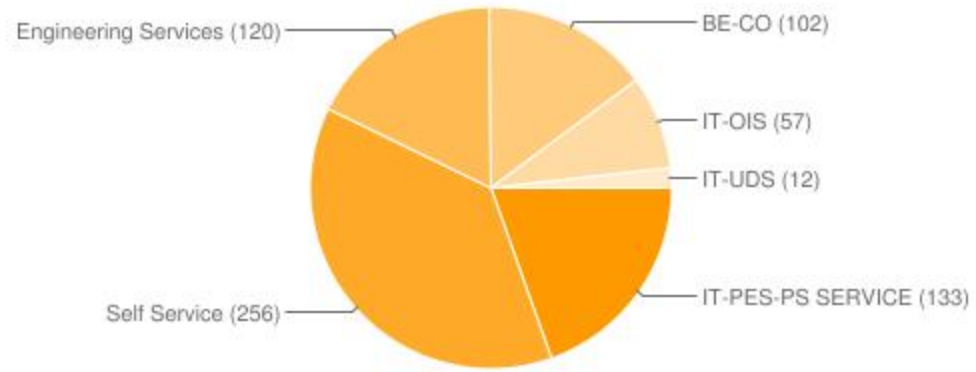
170 Hypervisors in CVI are grouped in 6 top-level ‘hostgroups’:

- 5 dedicated hostgroups
  - For large, well-defined communities:
    - Physics Services, Engineering, Beams development, Operating Systems Support, Conferencing
  - Admin privileges delegated
    - To migrate VMs, modify virtual hardware, etc
- Self-Service hostgroup
  - Shared ‘public’ resource
  - Many short-lived test/development VMs

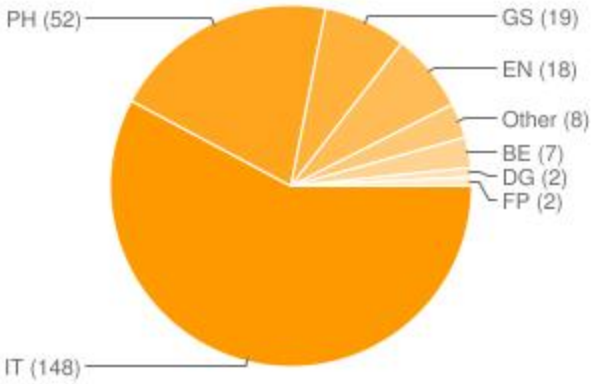
CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

Distribution of hostgroups

680 Virtual Machines divided over 6 hostgroups:



256 VMs in Self-service hostgroup, per department



CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

Hypervisor backup of VMs

- for disaster recovery only
  - or: if user missed expiration warnings, and his VM gets deleted
- TSM client 6.2.1.1 running on Hypervisors
  - backup of full VHDs and VM configuration
- very few restores (only 2 in 2010)
- works fine, but is very expensive...
  - [WIP] implement ‘opt-in’ policies to reduce the number of VMs to be backed up
  - [WIP] reduce the number of copies in TSM
- Note: Physics Services excluded from backup (service manager choice)

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

Network configurations

- hypervisors on primary IP services
  - can be private network
- VMs on 'clusters' of secondary IP services
  - Campus Network, LCG, private, ...
  - add secondary IP services to a clusters when necessary
    - Number of IP addresses for VMs no longer bottleneck
  - IP address of VM won't change when VM is migrated to a hypervisor within the same primary IP service!

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

Hypervisors

Two different hardware types currently in use

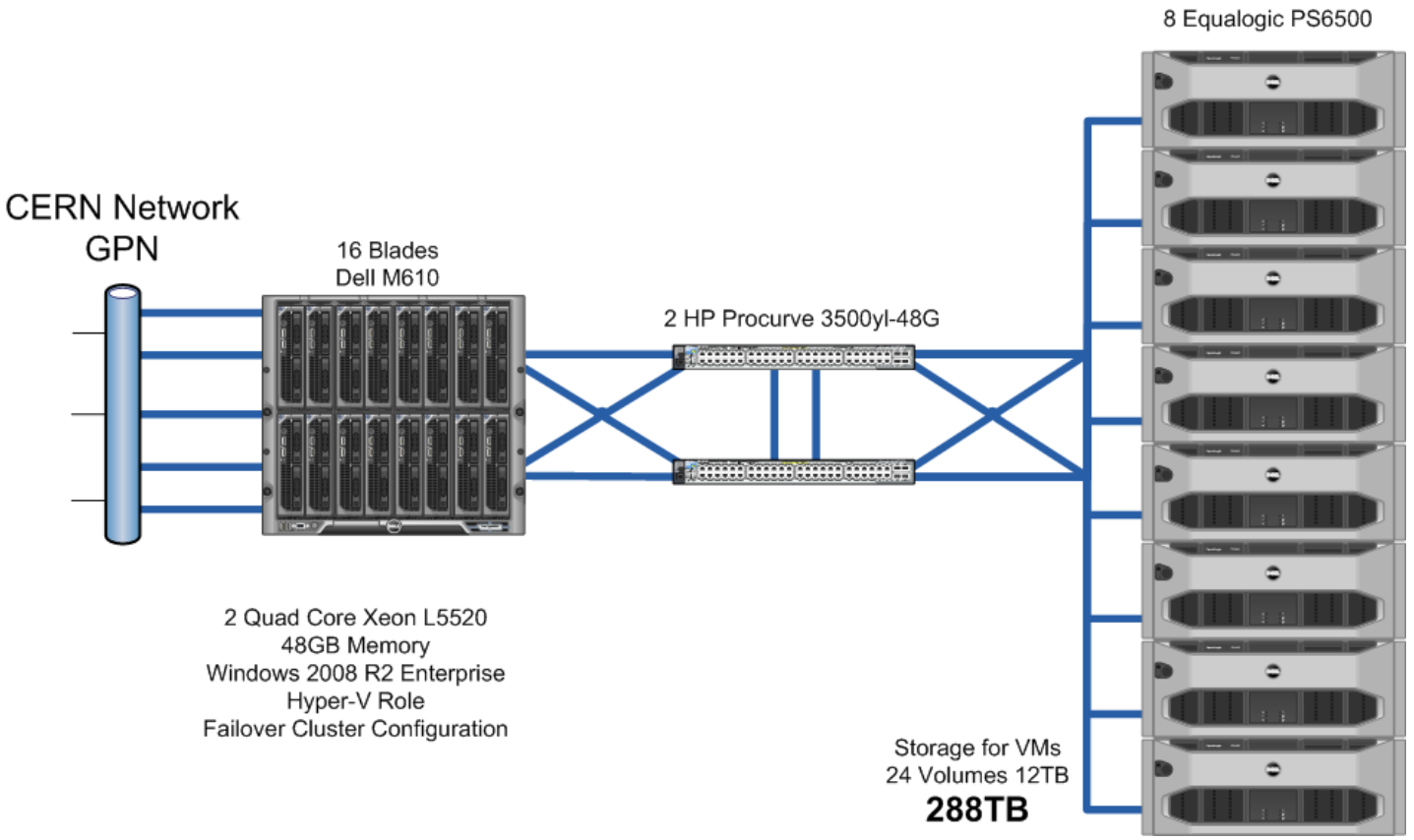
- Direct Attached Storage
  - ~90 Pyramid diskservers:
    - 24 GB memory
    - 2 quadcore E5410 CPUs
    - 1 TB of RAID-10 diskspack
  - Allows for ‘Quick migration’ of VMs
    - session of VM unaffected
    - network interruption of ~1 minute
  - Machines are easy to operate, but not well suited for High Availability solutions

iSCSI

- Shared Storage
- 5 * (16 Dell Blades + Equallogic storage arrays)
- Cluster Shared Volume allows for ‘Live migration’ of VMs:
  - automatic when host is put in maintenance mode
  - allows transparent reboots of hypervisors:
    - session of VM unaffected
    - network interruption of ~few second
  - Easier to administer
    - transparent firmware upgrades, hotfixes, etc

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

Physics Services



- Five of these setups are being deployed
- ~100 Virtual Machines

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

Linux VMs – Integration Components

- Aka paravirtualized drivers
- Official release by Microsoft in Summer 2010
  - Under GPL
- ICs enhance VM performance
  - improved disk IO (details on next slide)
  - allows for synthetic gigE vNIC
    - Improved network IO performance
- ICs provide additional functionality
  - shutting down VM through VMM console interface
  - paravirtualized SCSI bus

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

Improved Disk I/O

Operating System	Arch	Direct Attached Storage		iSCSI Storage		Notes
		Write [MB/s]	Read [MB/s]	Write [MB/s]	Read [MB/s]	
SLC5	i386	70	60	15	15	
SLC5	x86_64	60	50	15	14	
SLC5 + ICs	i386	120	160	108	25	
SLC5 + ICs	x86_64	160	80	110	25	
SLC5 + ICs	i386	180	266	105	86	read_ahead_kb=1024
SLC5				110	70	Bare metal
Win2008		300	450			Bare metal

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

Integration Components

- ICs included in "staging" area of upstream kernel
- They are not included in RHEL5
  - Nor in RHEL6
- CERN Linux team have packaged them
  - kernel independent, to ease upgrades
  - RPM for SLC5 (x86_64 and i386)
    - Should work on RHEL5, CentOS5, SL5 as well
  - yum install {kmod-,}mshvic

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

Experience with ICs

- Performance of Linux VMs on Hyper-V:
  - Users are happy ☺
  - Few usecases where 'small' disk IO is very slow (mysql db table scan, file system traversal). To be understood.
- A few open issues
  - Clock drift
    - workarounds exist, but broken recent 64bit RHEL kernels. Solution expected Real Soon Now
  - Console access
    - requires Internet Explorer + Active X
    - mouse requires an additional Integration Component

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

Summary

- CVI usage has grown dramatically in 2010
  - And will continue to grow in 2011
- New user communities are joining, with new requirements
  - New interfaces being deployed
  - Customer-provided templates!
- Linux VMs on Hyper-V work well
  - Integration Components are available

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it