Virtual machines

ALICE

Costin.Grigoras@cern.ch

Experience and use cases

- Services at CERN
- Worker nodes at sites
  - CNAF
  - GSI
- Site services (VoBoxes)
  - GSI
  - Subatech
  - FZK

Services at CERN

- 3 hosts, 32 cores, 64GB RAM, 28 fast disks
  - Ubuntu & VirtualBox
- 25 guests
  - Build servers:
    (SLC5 32 & 64bit, Ubuntu) * (AliEn, AliROOT)
  - Experimental build servers for development and testing
  - 12 other service machines
- Vanilla operating systems everywhere

(Software distribution)

- SLC5 libc-compatible binaries
  - Packaged with all dependencies
- Largest set to install: 433MB
  - 43MB : AliEn + system dependencies
  - 270MB : AliROOT
  - 120MB : ROOT
- Delivered to the WNs either as
  - Shared file system (NFS) - unpacked
  - Torrent - tarballs
- Same binaries available to both jobs and users
  - Very easy to switch from local to PROOF to Grid

Services at CERN

- Excellent solution for compacting rack space (power, investment)
  - Good environment for building, testing, prototyping
- Applies only for non-demanding services
  - Friendly enough to coexist with the rest of the guests on the same host
  - Some hiccups
    - Unreliable network throughput
    - Time flow issues with SLC default kernels under VB

Services at CERN

- Large number of concurrent VMs require good quality hardware
  - Replacement of the entire storage plane in two high-end servers to be able to support the current number of guests
  - Would be there also for processes on the physical host

Worker nodes at sites - CNAF
Francesco Noferini

- Virtual machines used for WNs and services
  - Not for storage, stability issues with GPFS
- KVM : 1 VM / core, up to 16 VMs per machine
- Happy with the network performance
  - *However ping between the two VoBoxes:*

    rtt min/avg/max/mdev = 9.166/11.290/15.837/1.360 ms
- Policies define the OS flavor for each experiment (role) and the appropriate VM images are started
  - Though recycling them when possible

Worker nodes at sites - issues

- SpecINT conversion factor cannot be determined

```
$ cat /proc/cpuinfo
cpu family      : 6
model           : 6
model name      : QEMU Virtual CPU version 0.9.1
stepping        : 3
cpu Mhz         : 2500.154
cache size      : 32 KB
```

Pentium II
Mendocino(300 to 500MHz)

- Correct accounting not possible
- Even if it were published, could it be trusted? (CPU time in particular)

Worker nodes at sites – issues

- Fixed number of machines with fixed resources => inflexible limits on the jobs
  - Jobs which could otherwise succeed are killed for overshooting their resources (even by a small fraction)
  - Not possible to dynamically adjust the VM resources with regards to the available resources on the host

Site services – GSI experience

Mykhaylo Zynovyev, Victor Penso

- AliEn and gLite services are running stable on VMs since 2006
- Tools: Xen/KVM, OpenNebula, Chef, Lustre, Torque
- Tests with ALICE analysis trains running on VMs have shown acceptable performance
  - Trains are CPU bound
  - Data is accessed from Lustre
  - Performance overhead is within 10%

Site services – GSI plans

- Provide users with infrastructure to submit and manage private virtual analysis clusters on demand
  - To be used with PoD (http://pod.gsi.de/) for interactive processing
  - To be used with job scheduler for batch processing
- Make use of OCCI (Open Cloud Computing Interface) API for VM management tools

Site services - Subatech

Jean-Michel Barbet

- 2 VMWare servers mounting the same SAN storage
  - Allowing hot-move of the running machines, high availability even when upgrading hosts
- 10 guests (LCG & CREAM CEs, BDII, DPM head, Quattor, PBS, AliEn VoBox)
  - Disk servers and MySQL – only on physical hosts due to poor performance on VMs

Site services - Subatech

- Very easy to clone and test new environment; snapshots as backup before every significant change
- “Pick and choose” policy on what to run on VMs and physical hosts

Site services - FZK

Artem Trunov

- Experimenting with the xrootd redirector
  - KVM, 2 hosts, 2 guests, shared GPFS filesystem for the VM images
  - Light service, so no problems expected

Bottom line

- We like a lot the virtual machines
  - Vanilla virtualization is an excellent tool to build diverse (OS) build and test systems
  - Careful selection of hardware to avoid overloads
  - The technology is used where applicable – not a 'universal solution to all problems'
- Site services and WNs – transparency is a must
  - We'd rather not know that something is running on a VM

Bottom line

- Site services and WNs
  - Multitude of adopted virtualization platforms (with their positive and negative sides)
  - Mastering storage from a VM is still an open issue, especially data servers
  - Generally accepted for services that are not I/O demanding, also not for Dbs
  - The adoption of virtualization technology is very uneven and does not depend on the site size (T1 / T2)