Performance Considerations

- Typical High Energy Physics (HEP) analyses showed a near native performance (3-6%) per core for Xen and KVM (with 'virtio' paravirtualization support).

Goals

- Share a computing resource between different user groups
- Provide multiple computing environments
- Isolate group partitions

Possible solutions:

- Common computing cluster, static partitioned
  - No load-balancing between user group partitions
- Common computing cluster, dynamic partitioned
  - Load-balancing possible

Current Implementation:

- Same concept two implementations: vimmagemanager at DESY
- batch at KIT

Several postdocs, PhD and diploma students are involved

The Testbed: IC1 Cluster at SCC

- Shared between nine different KIT department
- 200 compute nodes
- 2 x Intel Quadcore Xeon
- 17.5 TFlop peak performance
- 2 user group partitions:
  - HPC partition: parallel computing (MPI)
  - HEP (High Energy Physics) Partition: Serial computing, High Throughput Computing (HTC)
- Operating Systems:
  - Suse Linux Enterprise 10.0Sp2 on the hardware machines
  - Scientific Linux 5 in virtualized worker nodes
- Virtualization technique: Kernel Virtual Machine (KVM)
- Batch system: MAUI/Torque (PBS)

Scientific Linux 5.1 on the hardware machines

Operating Systems:

- Scientific Linux 5 in virtualized worker nodes
- Virtualization technique: Kernel Virtual Machine (KVM)
- Batch system: MAUI/Torque (PBS)

Dynamic Virtualization of Worker Nodes

- Wrapper script (Prologue/Epilogue) around the actual computing job inside the batch system
- Virtual machines are deployed on the hardware node disc
- Wrapper script starts VM and checks status of the VM
- Jobs are piped to the VM via ssh
- Job runs on VM
- After job execution VM is destroyed and image is deleted
- New images are prepared for following jobs

DESY Solution:

- Main goal: Bringing virtualization to GridKA_lite users
- Using the Sun Grid Engine (SGE) as back-end
- SGE support by the gLite Grid middleware
- lcg-CE (CE) as Grid front-end
- Operating Systems:
  - Scientific Linux 5.1 on the hardware machines
  - No middleware installed, runs SGE execd and shepherd
  - SL 4.7 in virtualized worker nodes
  - Has middleware installed, but no batch client

Implementation Details:

- Short prologue, epilogue and started scripts
- Basically run vimmagemanager
- Execution of job payload via ssh on VM
- Wrapper script starts VM and checks status of the VM
- Proof of principle implementation: 1 CE, 1 SGE master, 1 host, 2 VMs
- Stress test with > 100 jobs submitted simultaneously via WMS successfully
- No changes to CE, minor SGE configuration changes

Virtualization Techniques

- Typical High Energy Physics (HEP) analyses showed a near native performance (3-6%) per core for Xen and KVM (with 'virtio' paravirtualization support).

Educational usage of Virtualization

- Virtualization tutorials at GridKA School 2007, 2008 and 2009
- GridKA School computing environment partially virtualized to provide school service and hands-on machines

Conclusion

- Dynamic Virtualization: Opportunisitic use of computing resources
  - Avoids limitation of the compute environment (Operation System, architecture, worker node setup)
  - Status: Development and testing at KIT and DESY

Publications: