Virtualization and containerization

Thomas Berreis
Outline

- Virtualization
  - Definition
  - Advantages / Features
  - Performance
- Lightweight virtualization: Container
  - Features
  - Performance
- Conclusion
## Introduction

<table>
<thead>
<tr>
<th>RANK</th>
<th>SITE</th>
<th>SYSTEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>DOE/SC/Oak Ridge National Laboratory United States</td>
<td><strong>Titan</strong> - Cray XK7, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x Cray Inc.</td>
</tr>
<tr>
<td>3</td>
<td>DOE/NNSA/LLNL United States</td>
<td><strong>Sequoia</strong> - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom IBM</td>
</tr>
<tr>
<td>4</td>
<td>RIKEN Advanced Institute for Computational Science (AICS) Japan</td>
<td>K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect Fujitsu</td>
</tr>
<tr>
<td>5</td>
<td>DOE/SC/Argonne National Laboratory United States</td>
<td><strong>Mira</strong> - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM</td>
</tr>
</tbody>
</table>

Fig: [http://www.top500.org/lists/2015/11/](http://www.top500.org/lists/2015/11/)
Introduction

- global sharing
- validation checks
- portability
- scalability

Definition

- technology to split physical environment into logical units
- possible for hardware and software
- hypervisor as abstraction layer
- less of bare-metal
Definition

Virtual Machine File System

Hypervisor

Fig: http://www.catbird.com/sites/default/files/catbird_threat-surface.jpg
Advantages

- energy-saving
- space-saving
- faster provisioning
- isolated applications
- environments for developing and testing

Fig: http://www.pressetext.com/news/photo/medium/20081112011/0
Advantages

- higher availability
- virtual disaster recovery
- easier administration
- intelligent management of resources
- enormous hardware compatibility

VMware Compatibility Guide

http://www.vmware.com/resources/compatibility/search.php
Techniques - RDMA

- RDMA (Remote Direct Memory Access)
  - Allows to exchange data in main memory without involving processor, cache or operating system
  - Minimizes overhead
  - Improves performance
Techniques - RDMA

Techniques - SR-IOV

- SR-IOV (Single Root I/O Virtualization)
  - presents single I/O device as multiple separate devices
  - each virtual device has its own
    - Configuration space, base address registers
    - Send/receive queues with own interrupts
  - Specific NIC driver needed
Techniques - DirectPath I/O

- Passthrough / VMDirectPath
  - allows direct and exclusive access to I/O devices by bypassing the virtualization layer
- Incompatibility with many virtualization features
  - Fault Tolerance
  - Snapshots
  - Live Migration
Performance

- VMware Test Configuration
  - 4x HP DL380p G8 (3.3 GHz, 128 GB RAM)
  - Hypervisor: VMware vSphere
  - ConnectX-2 QDR InfiniBand 10 Gb / RDMA
Performance - Bandwidth

**READ**

**WRITE**

Fig: https://blogs.vmware.com/cto/running-hpc-applications-vsphere-using-infiniband/
Performance - Latency

**READ**

**WRITE**

---

Fig: https://blogs.vmware.com/cto/running-hpc-applications-vsphere-using-infiniband/
Performance - Summary

- Virtualized HPC performance close to bare-metal

Performance - Summary

- Latency almost identical in future releases(?)

Celebrities

- ESX (VMware)
- Hyper-V (Microsoft)
- KVM (Red Hat)
- XEN (Citrix)
Use Case - Problem

- International work …
- Reconstruct some Chinese research findings …
- 红旗 Linux … what the heck?!
Use Case - Solution

- Strange operating system?
- Incompatible system libraries?
- **Virtualization!**
  - Use of independent well-known environments
  - Of course compatible with our software
  - Easy to provide and remove
Container
Container

- without hypervisor but also with virtualization layer
- uses system libs and kernel (limited to host ecosystem)
- non virtualized drivers
- namespaces to isolate processes
- CGroups to isolate or limit resource usage
- partly layered file systems
Container

Virtual Machines

- App A
- Bins/Libs
- Guest OS

Docker

- App A
- Bins/Libs
- Docker Engine

Host Kernel/OS

Server

Fig: http://www.spantree.net/blog/2015/04/29/10-things-to-know-about-docker.html
Performance

NASA Advanced Supercomputing Parallel Benchmarks
(Hypervisor: QEMU KVM)

Performance

- Other high performance applications

Performance

- SYSBENCH OLTP with Red Hat 7 & Docker

Celebrities

- Docker
- LXC
- OpenVZ
- Solaris Zones
- FreeBSD Jails
Use Case - Problem

- Develop an application that fits in your HPC environment …
- You have to test your application in this environment …
- You cannot virtualize this damn special OS …
- Developing on your client is pointless …
Use Case - Solution

- Multitenancy architecture
  - Work in a container within this environment
  - Independent to other applications / containers
  - Test, QA and Production possible in same environment
  - Without any overhead
Conclusion

- each kind of virtualization has pros and cons
- container not replacing virtual machines
- containerization is not only a passing fad
- requirements are crucial
- containerization quite adapted for HPC
Conclusion

Fig: http://www.spantree.net/blog/2015/04/29/10-things-to-know-about-docker.html
Conclusion – a fusion?

- basically possible

- but
  - advantages gets lost
  - bottleneck effect

Fig: https://mydigitalworkplace.files.wordpress.com/2012/05/bottleneck.gif
Conclusion

Fig: http://blogs.vmware.com/performance/files/2014/10/linpack.png
Customers

- PayPal
- Groupon
- Uber
- eBay
- Spotify
- BBC News
- yelp

- Deutsche Telekom
- Beiersdorf
- Symantec
- Adobe
- SAP
- Vodafone
- HP
Conclusion

Much more about virtualization we cannot discuss today:

- The UberCloud Experiment
- AWS | Amazon Elastic Compute Cloud
- OpenStack Open Source Cloud Computing Software
- Apache CloudStack Open Source Cloud Computing
- VMware vSphere Big Data Extensions
- Virtualized InfiniBand
- Kubernetes by Google
- ...
THANK YOU!
References

References


- [https://www.thomas-krenn.com/de/wiki/VMware_VMDirectPath_zum_Durchreichen_von_PCI_Karten](https://www.thomas-krenn.com/de/wiki/VMware_VMDirectPath_zum_Durchreichen_von_PCI_Karten) Slide 12 | Werner Fischer, 18.05.2015


References


References


- [http://www.top500.org/project/linpack/](http://www.top500.org/project/linpack/) Slide 33 | 11.2015


- [https://www.docker.com/customers](https://www.docker.com/customers) Slide 34 | 11.2015

- [https://www.mellanox.com/page/virtualization](https://www.mellanox.com/page/virtualization) | 11.2015

- [http://nowlab.cse.ohio-state.edu/static/media/publications/abstract/huangwei-ics06.pdf](http://nowlab.cse.ohio-state.edu/static/media/publications/abstract/huangwei-ics06.pdf) | Wei Huang, 06.2006