Virtualization and containerization

Thomas Berreis

Outline

- Virtualization

  - Definition

  - Advantages / Features

  - Performance

- Lightweight virtualization: Container

  - Features

  - Performance

- Conclusion

Introduction

RANK	SITE	SYSTEM
1	National Super Computer Center in Guangzhou China	**Tianhe-2 (MilkyWay-2)** - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P NUDT
2	DOE/SC/Oak Ridge National Laboratory United States	**Titan** - Cray XK7 , Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x Cray Inc.
3	DOE/NNSA/LLNL United States	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom IBM
4	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect Fujitsu
5	DOE/SC/Argonne National Laboratory United States	**Mira** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM

Fig: http://www.top500.org/lists/2015/11/

Introduction

- global sharing
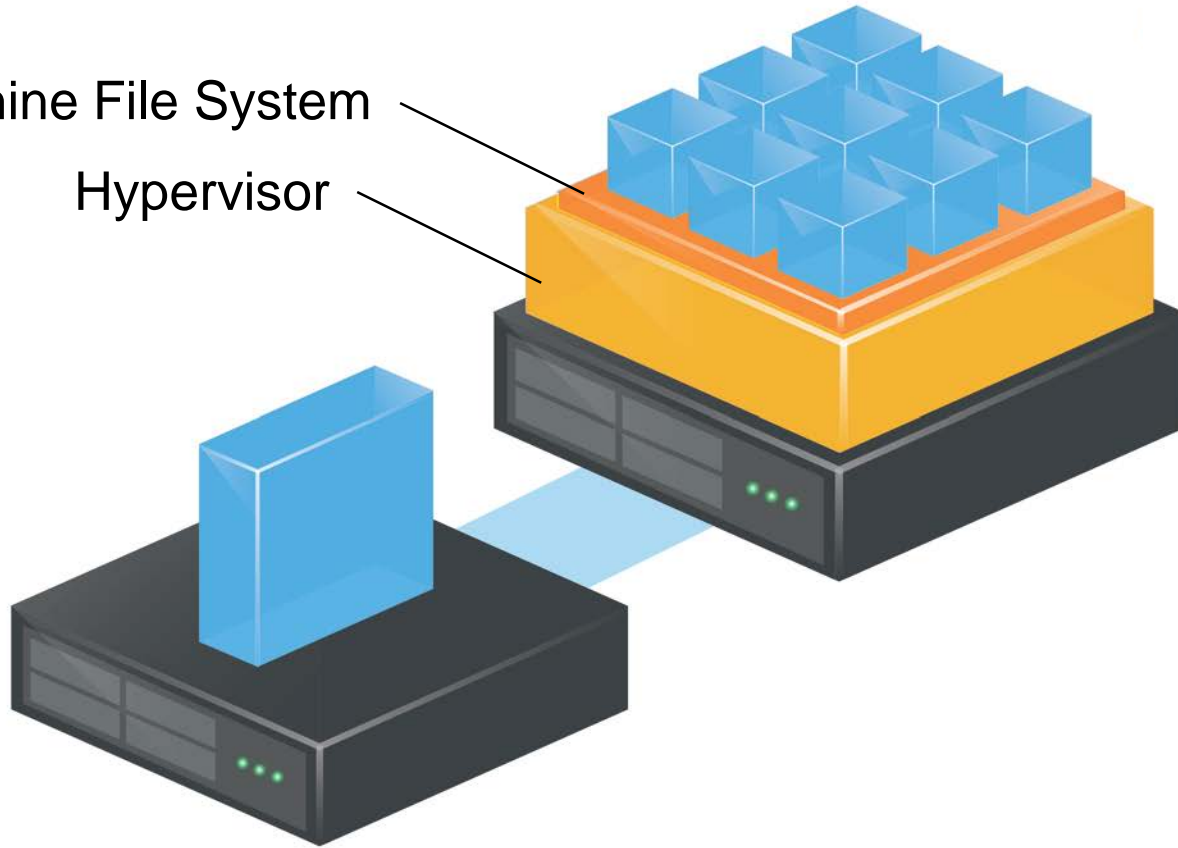
- validation checks

- 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

Advantages

- **energy-saving**

- **space-saving**

- faster provisioning

- isolated applications

- environments for developing and testing

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

PHYSICAL   **VIRTUAL**

Fig: http://blogs.vmware.com/cto/files/2014/11/ndm2014-sc14-simons.pdf
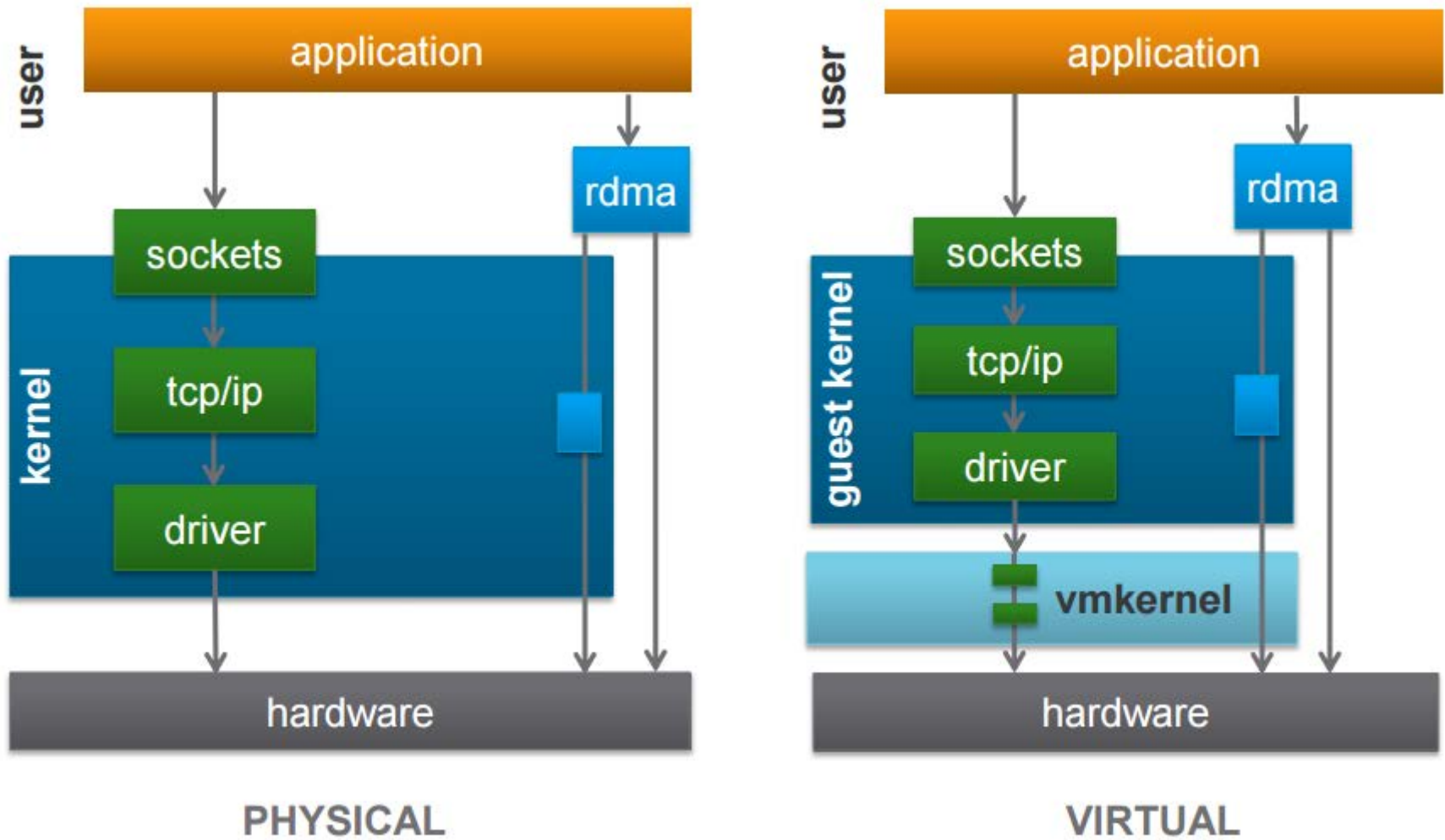
Techniques - SR-IOV

- 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

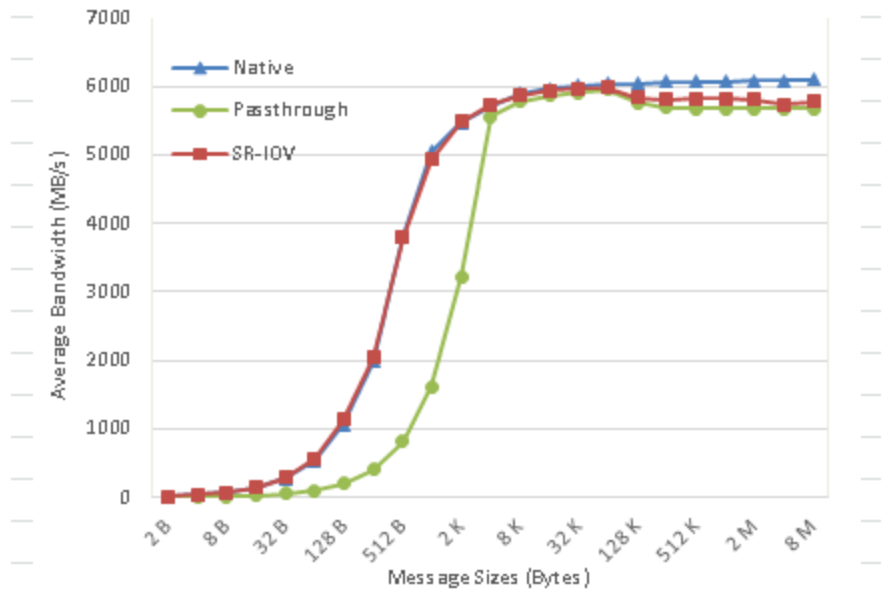   - Snapshots

    - Live Migration

Performance

- VMware Test Configuration

  - 4x HP DL380p G8 (3.3 GHz, 128 GB RAM)

  - Hypervisor: VMware vSphere

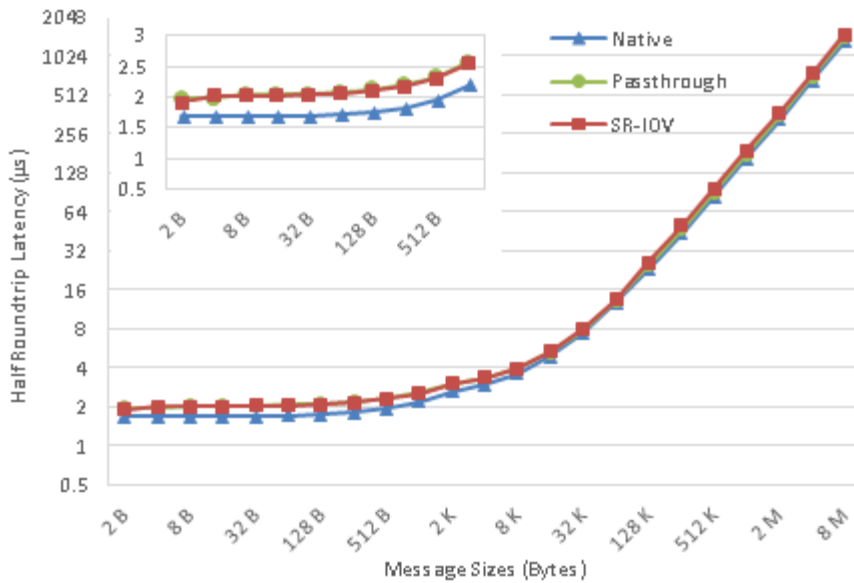 - ConnectX-2 QDR InfiniBand 10 Gb / RDMA
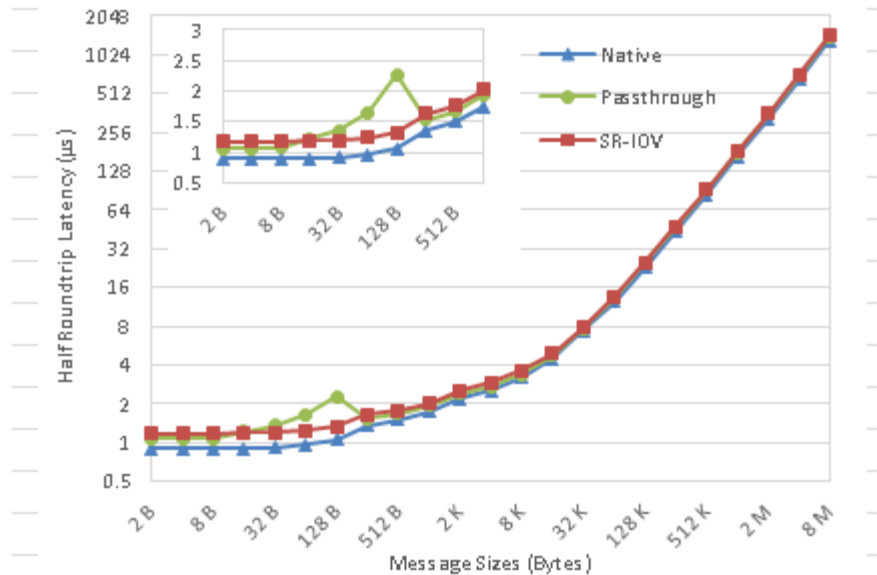
Performance - Bandwidth

READ
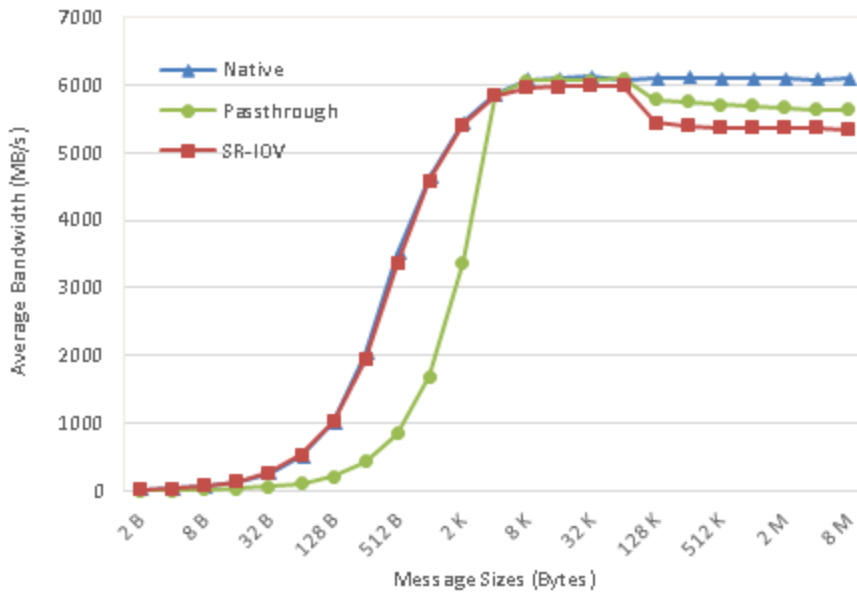
WRITE
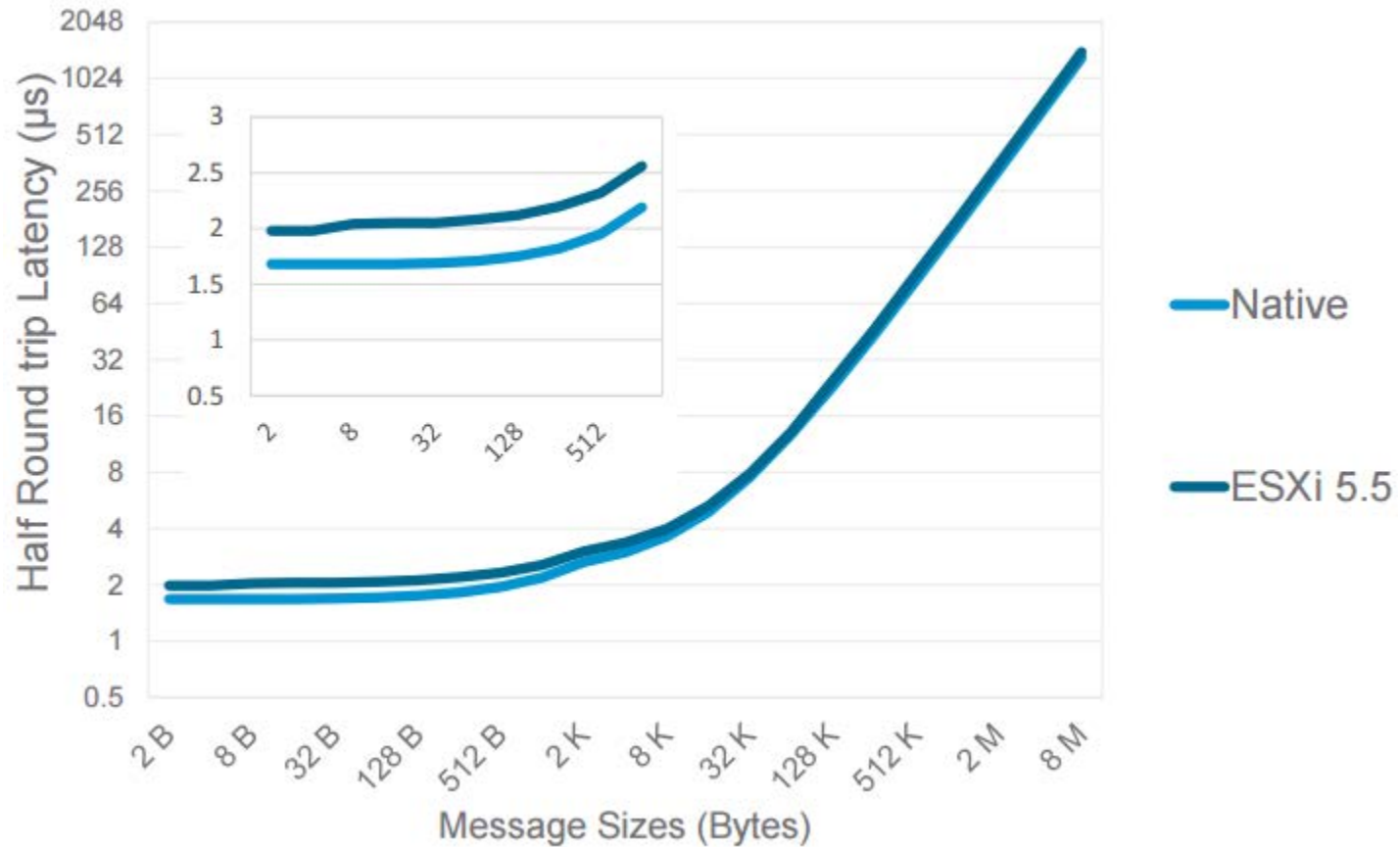
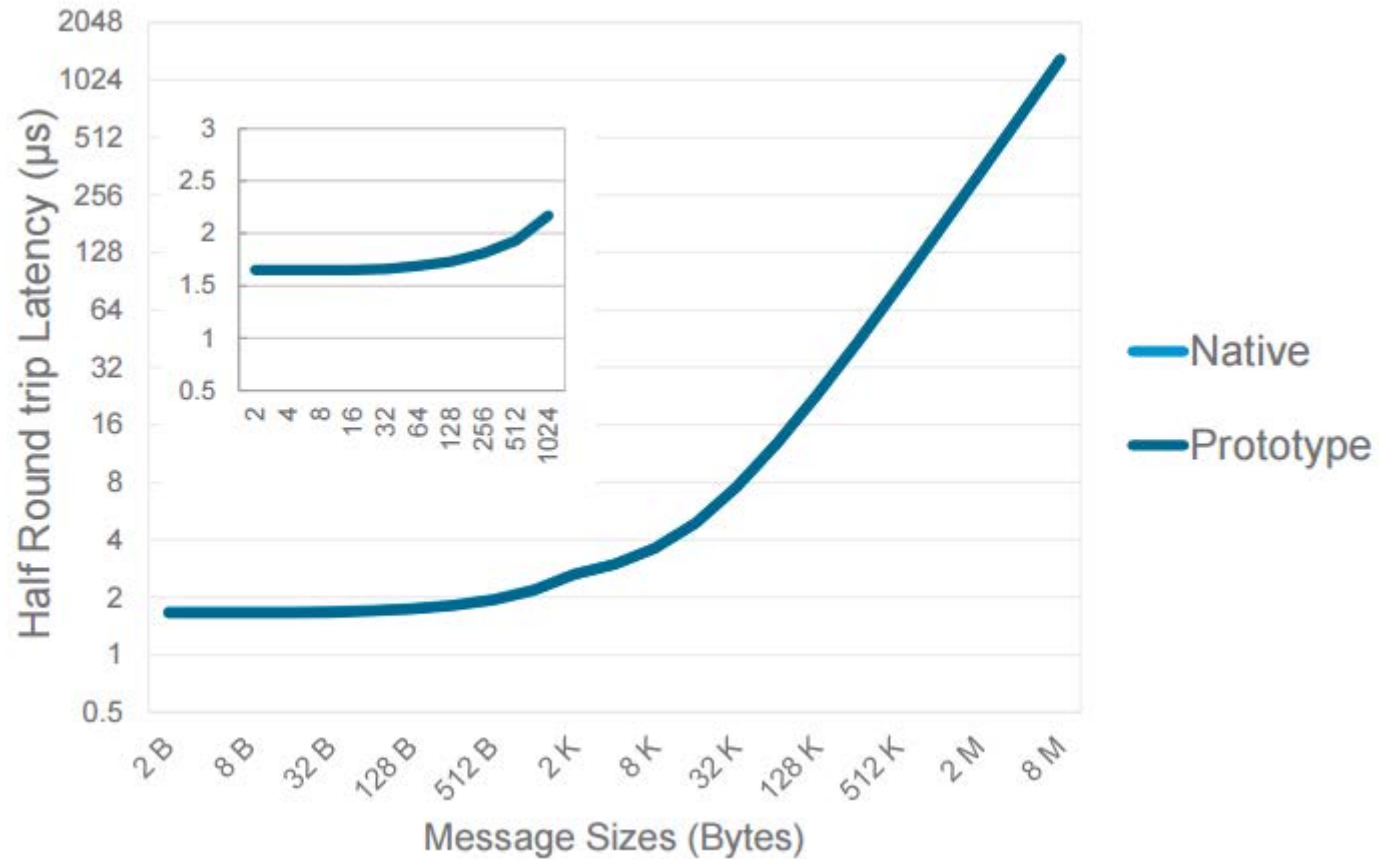# Performance - Latency

READ

WRITE

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

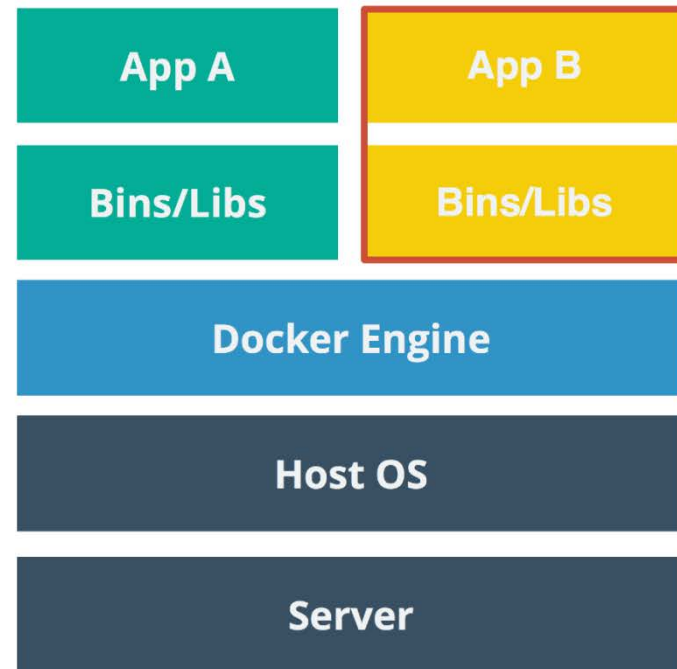 - Easy to provide and remove

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	App B
Bins/Libs	Bins/Libs
Guest OS	Guest OS
Hypervisor	
Host Kernel/OS	
Server	

Docker
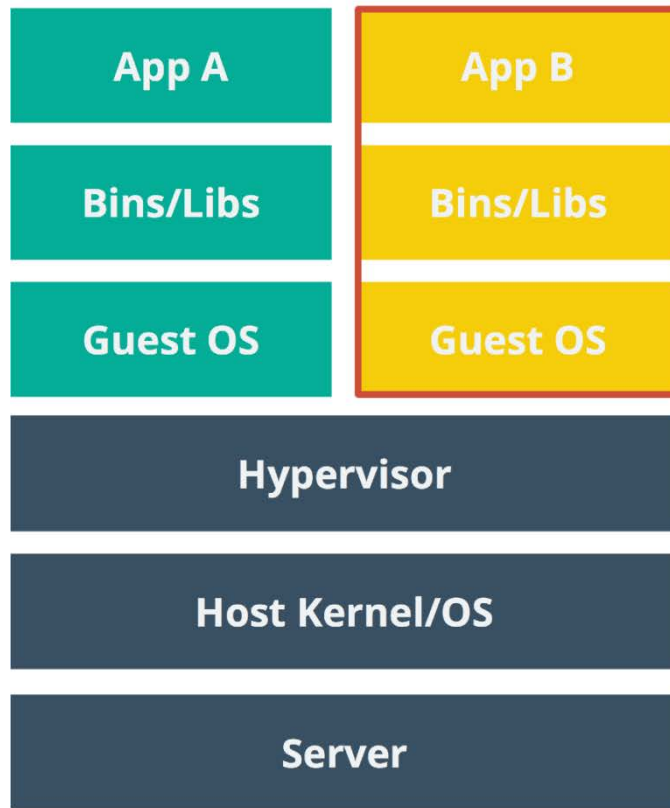
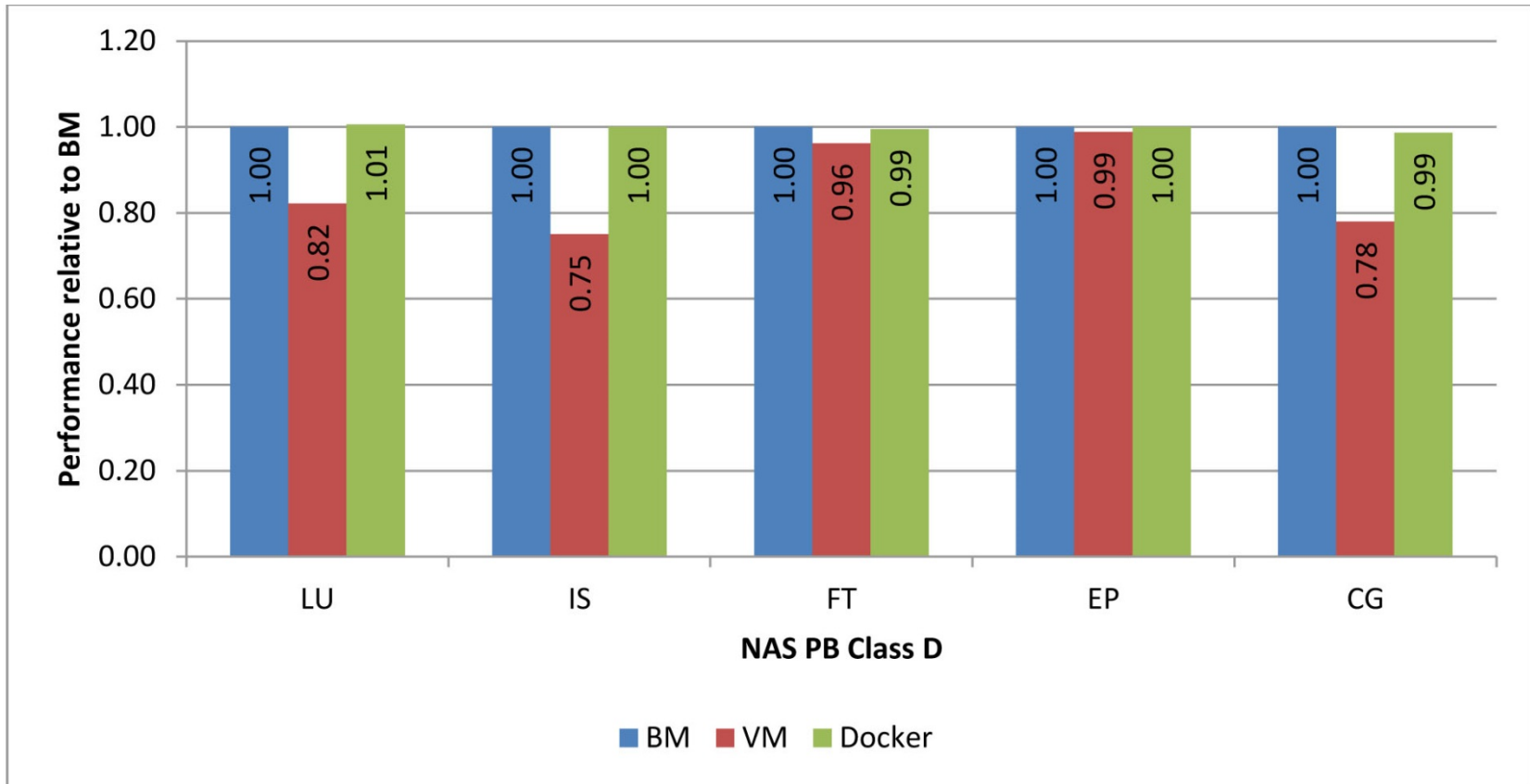| App A	App B
Bins/Libs	Bins/Libs
Docker Engine	
Host OS	
Server	

Performance

- NASA Advanced Supercomputing Parallel Benchmarks
  (Hypervisor: QEMU KVM)

Performance

- Other high performance applications

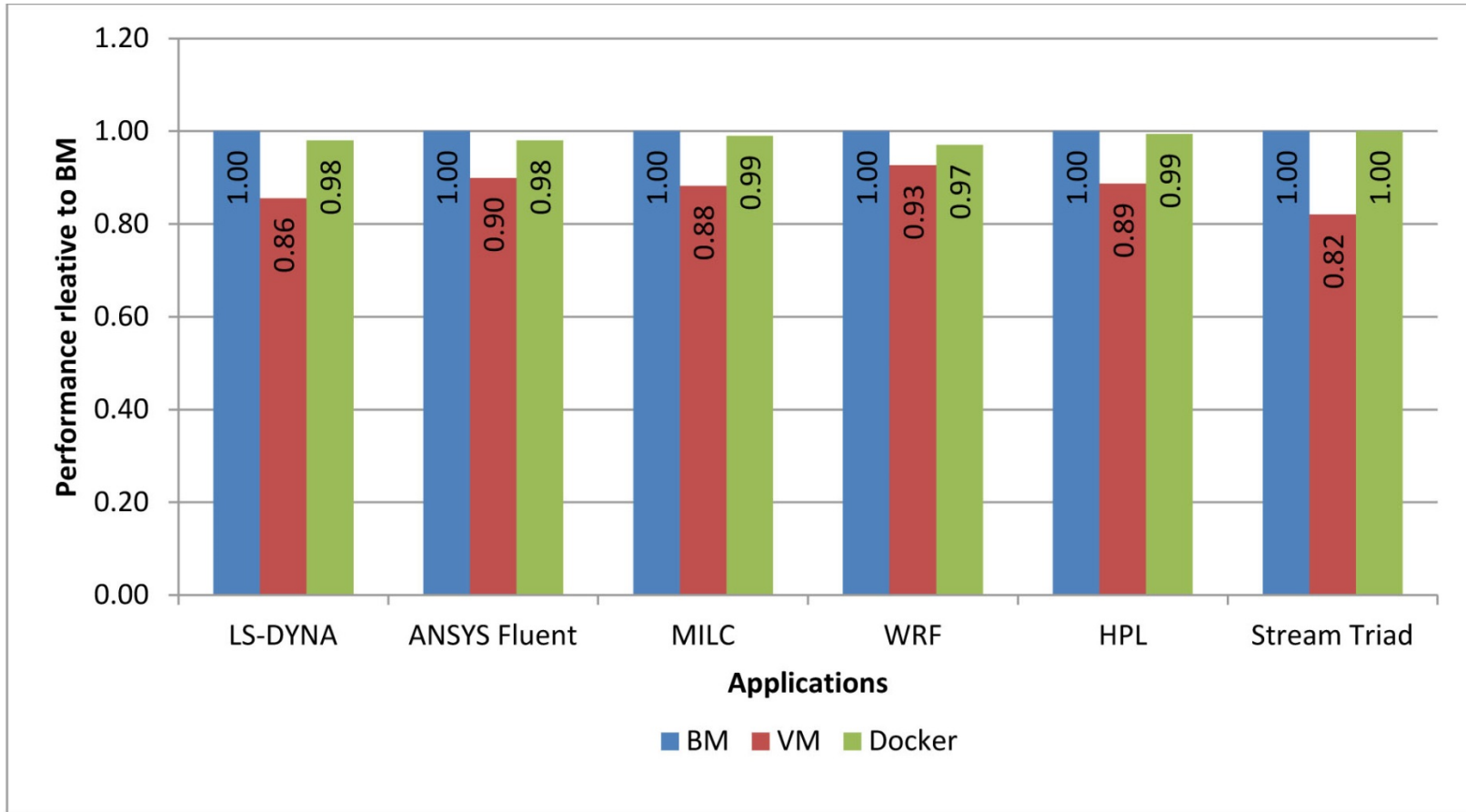# Performance

- SYSBENCH OLTP with Red Hat 7 & Docker
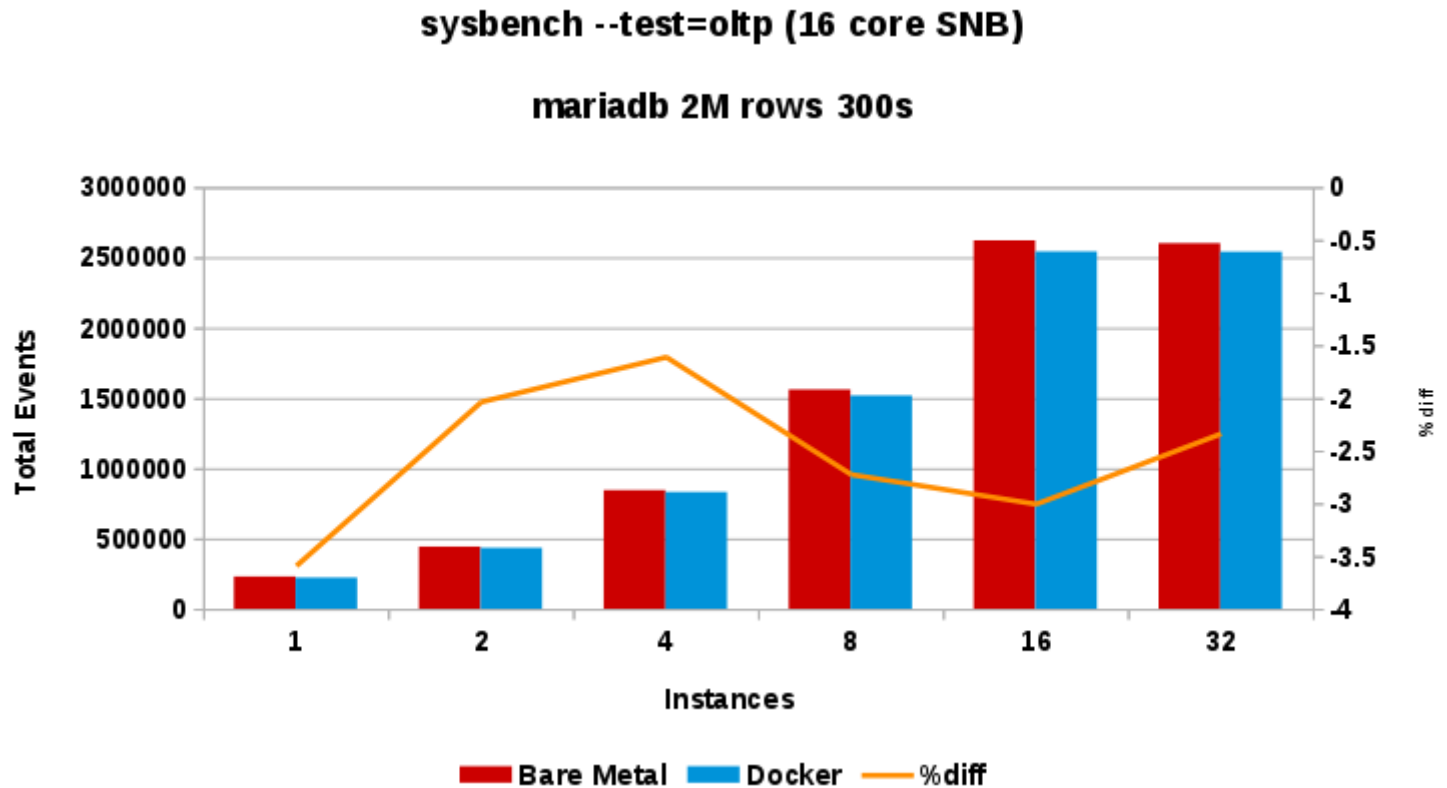
Celebrities

- 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
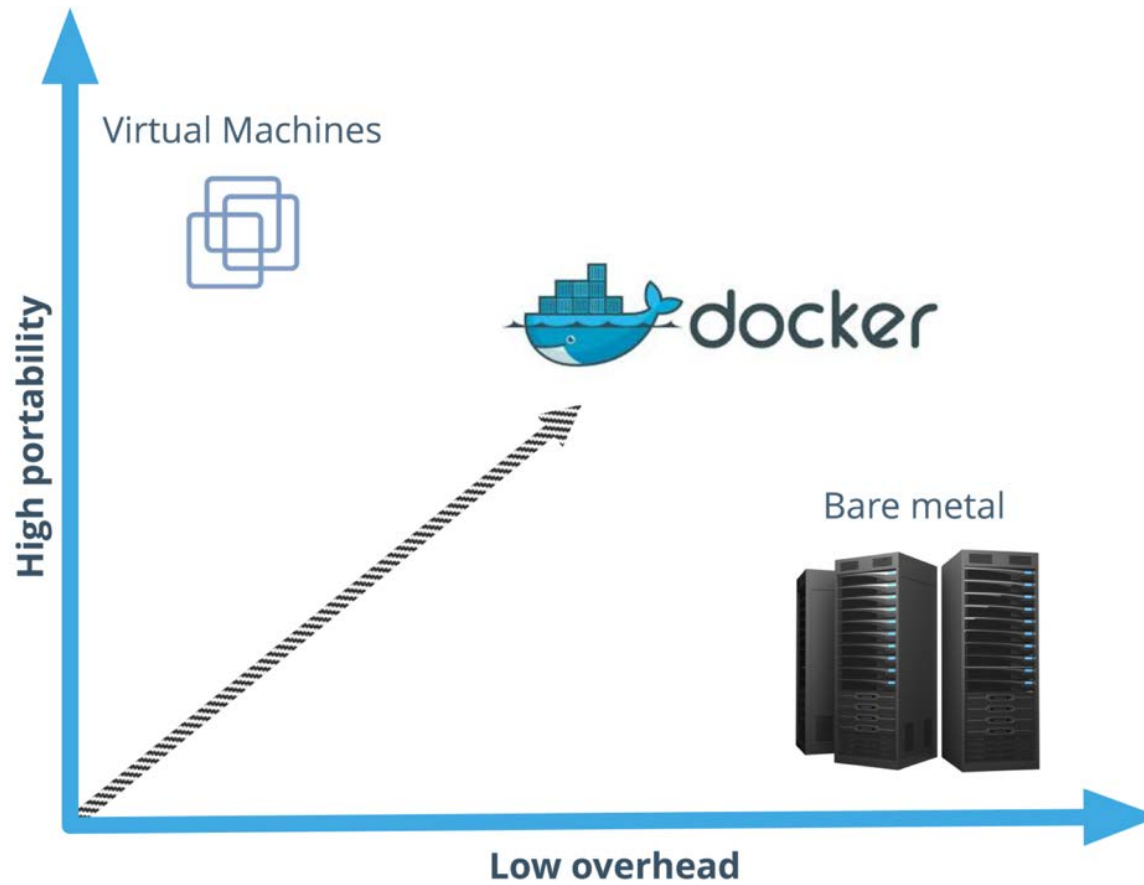
Conclusion – a fusion?

- basically possible

- **but**

  - advantages gets lost

  - **bottleneck effect**

Conclusion

Fig: http://blogs.vmware.com/performance/files/2014/10/linpack.png
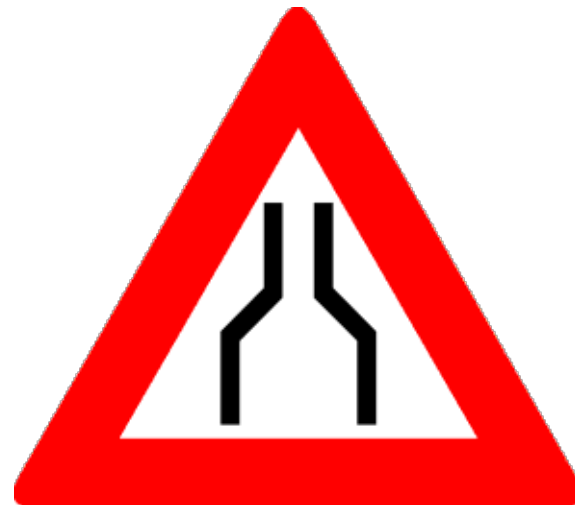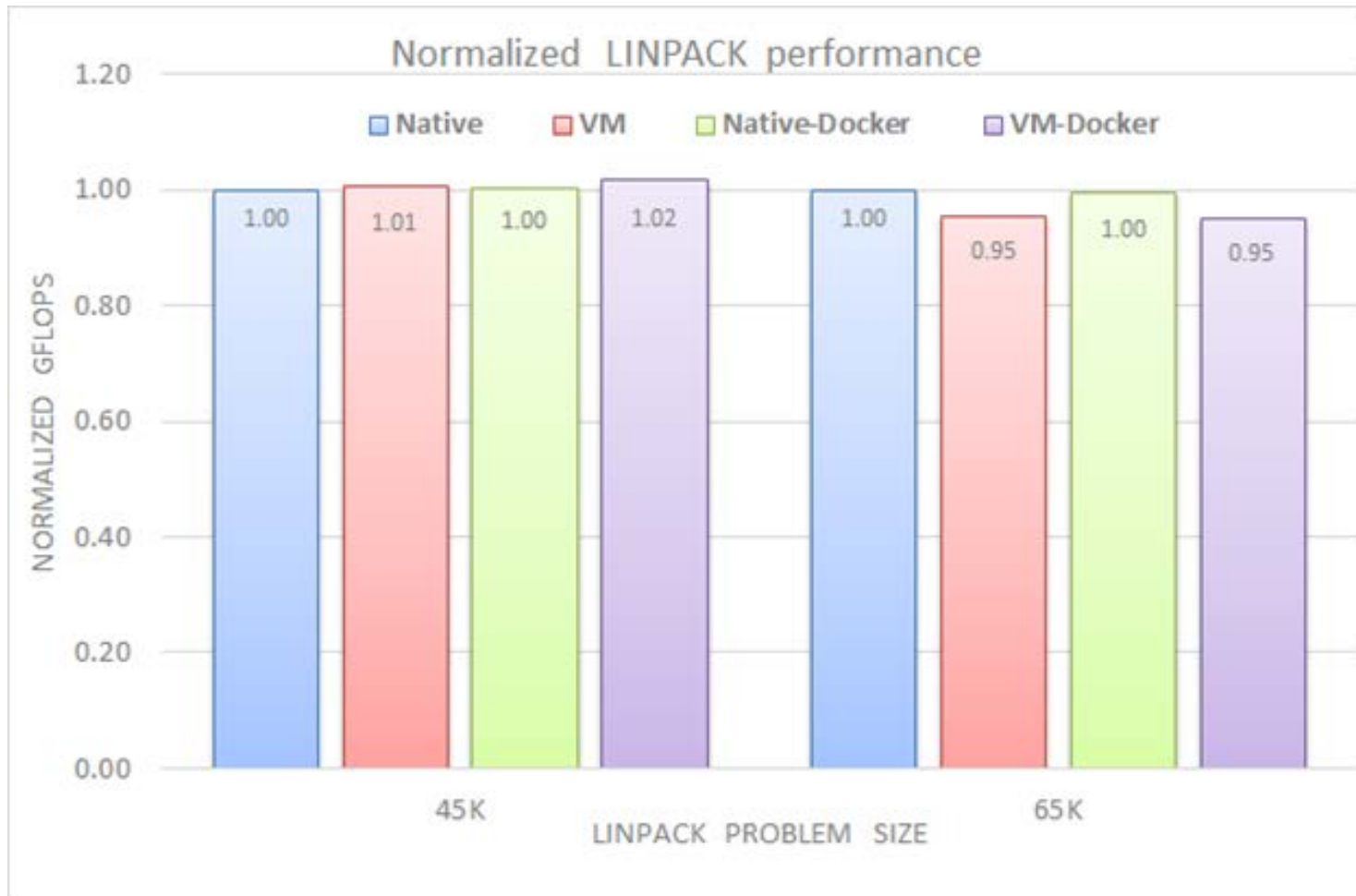
Customers
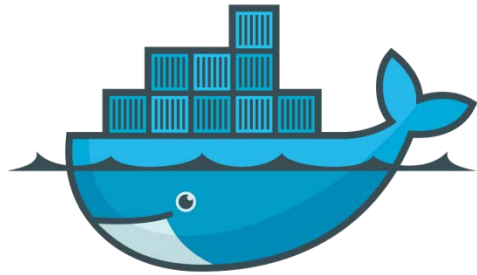
- PayPal
- Groupon
- Uber
- eBay
- Spotify
- BBC News
- yelp

- 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

- …

References

- https://en.wikipedia.org/wiki/Hypervisor Slides 5 - 8 | 11.2015

- https://en.wikipedia.org/wiki/Virtual_machine Slides 5 - 8 | 11.2015

- http://www.tecchannel.de/server/virtualisierung/2029842/faq_alles_ueber_virtualisierung_varianten_und_unterschiede/ Slides 5 - 8 | Bertram Wöhrmann, 03.09.2012

- http://www.infoworld.com/article/2621446/server-virtualization/server-virtualization-top-10-benefits-of-server-virtualization.html Slides 5 - 8 | David Marshall, 02.11.2011

- https://www.vmware.com/de/products/vsphere/features/vmfs Slide 6 | 11.2015

- https://www.vmware.com/de/business-continuity/disaster-recovery Slide 8 | 11.2015

- http://glennklockwood.blogspot.de/2013/12/high-performance-virtualization-sr-iov.html
  Slide 11 | Glenn K. Lockwood, 03.12. 2013

References

- http://blogs.vmware.com/cto/hpc-update/ | Josh Simons, 31.10.2014

- http://blogs.vmware.com/cto/files/2014/11/ndm2014-sc14-simons.pdf Slides 10, 16, 17 | Josh Simons, 11.2014

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

- https://blogs.vmware.com/cto/running-hpc-applications-vsphere-using-infiniband/
  Slides 13 - 15 | Josh Simons, 22.12.2014

- http://www.rdmamojo.com/2014/03/31/remote-direct-memory-access-rdma/ | Dotan Barak, 31.03.2014

- https://en.wikipedia.org/wiki/Red_Flag_Linux Slide 19 | 11.2015

- http://www.spantree.net/blog/2015/04/29/10-things-to-know-about-docker.html
  Slides 22, 23, 31 | Cedric Hurst, 29.04.2015

References

- https://en.wikipedia.org/wiki/UnionFS Slide 22 | 11.2015

- http://sleekd.com/servers/docker-vs-virtualization/ Slide 22 | Razvan, 29.09.2014

- https://access.redhat.com/documentation/en-US/Red_Hat_Enterprise_Linux/6/html/Resource_Management_Guide/ch01.html Slide 22 | 11.2015

- http://www.nas.nasa.gov/publications/npb.html Slide 24 | 11.2015

- http://en.community.dell.com/techcenter/high-performance-computing/b/general_hpc/archive/2014/11/04/containers-docker-virtual-machines-and-hpc
  Slides 24, 25 | Nishanth Dandapanthula, 04.11.2014

- http://developerblog.redhat.com/2014/08/19/performance-analysis-docker-red-hat-enterprise-linux-7/
  Slide 26 | Jeremy Eder, 19.08.2014

References

- https://de.wikipedia.org/wiki/Online_Transaction_Processing Slide 26 | 11.2015

- http://blogs.vmware.com/performance/2014/10/docker-containers-performance-vmware-vsphere.html
  Slide 33 | Banit Agrawal, 15.10.2014

- http://www.top500.org/project/linpack/ Slide 33 | 11.2015

- http://www.vmware.com/a/customers/customer Slide 34 | 11.2015

- https://www.docker.com/customers Slide 34 | 11.2015

- http://www.mellanox.com/page/virtualization | 11.2015

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