

RED HAT
SUMMIT

BOSTON, MA
JUNE 23-26, 2015

RHEV Hypervisor 7: Now and the Future

Karen Noel
Senior Software Engineering Manager
Platform Engineering – Virtualization Team
June 2015

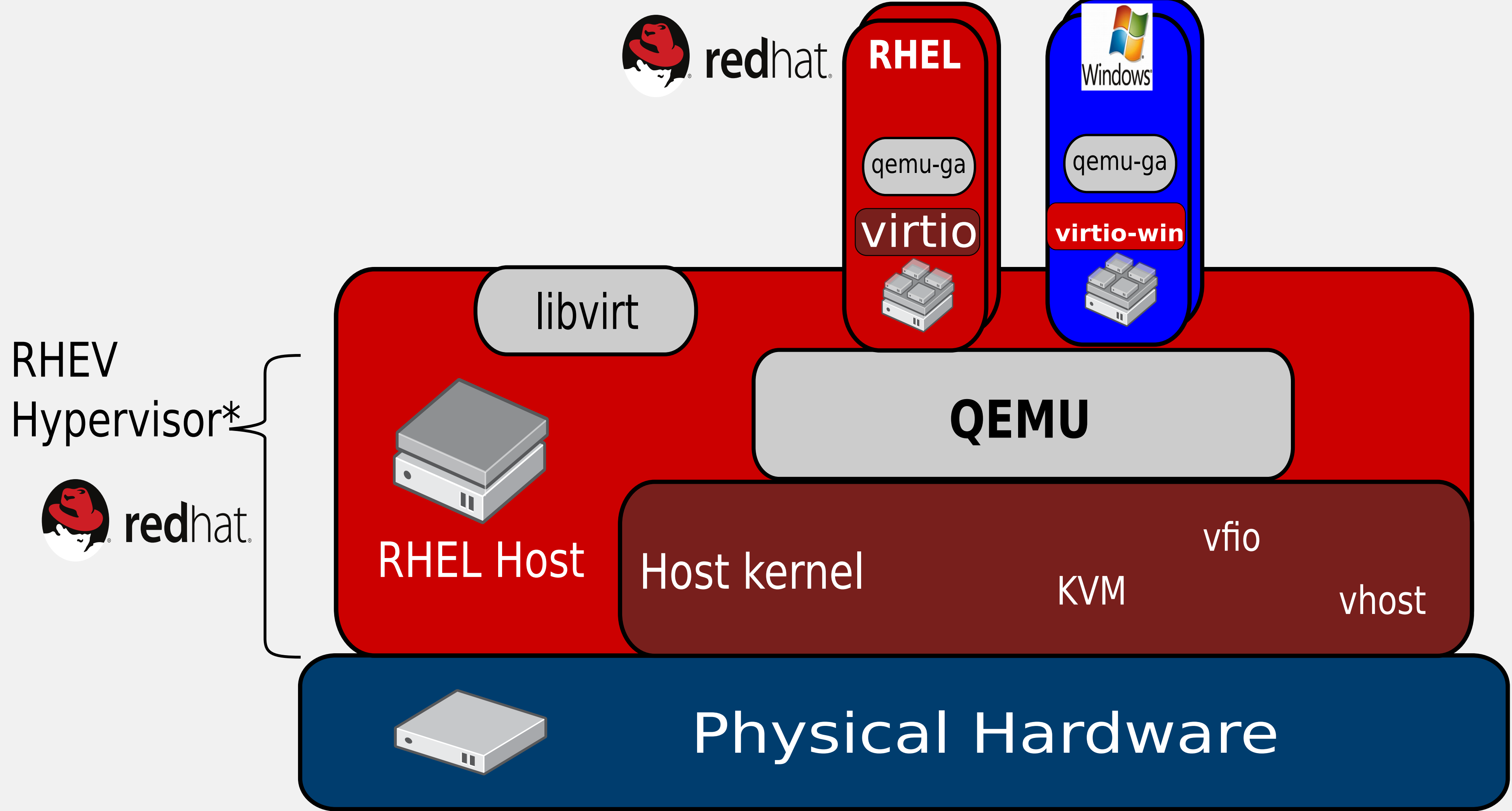
Storyline

- Engineering at Red Hat – Virt Team
- KVM Performance & Scaling
- RHEV Hypervisor 7 - Feature Roadmap

Virt Team Engineering

RHEV Hypervisor – best engineered distribution of KVM!

RHEV Hypervisor & Virt Team Contribution



*Do not confuse with RHEV-Hiso used with RHEV-M only

RHEV Hypervisor 7 - RHEL 7.1 example

- Rebase to upstream QEMU 2.1.2 for RHEL 7.1
 - Disable some devices (-)
 - Disable certain features (-)
 - Backport certain features: QEMU 2.2 (+)
 - Backport bugfixes: QEMU 2.2 (+)
- RHEV and RHEL OpenStack Platform
 - qemu-kvm-rhev-2.1.2*.rpm
 - Backport additional fixes (+)



RHEV Hypervisor 7

RHEL 7

RHEL 7.1

RHEL 7.2

RHEL 7.3

.....

qemu-kvm-rhev

QEMU 1.5.3

QEMU 2.1.2

QEMU 2.3*

QEMU 3.0*

Included in:

- RHEV
- RHEL OpenStack Platform

Unsupported
Features &
Devices
Disabled

Unsupported
Features &
Devices
Disabled

Unsupported
Features &
Devices
Disabled

Unsupported
Features &
Devices
Disabled

Backport
Bug fixes

Backport
Bug fixes

Backport
Bug fixes

Backport
Bug fixes

Backport some
Features up to
QEMU 2.0

Backport few
Features up to
QEMU 2.2

Backport few
Features up to
QEMU 2.5*

Backport few
Features up to
QEMU 3.2*

Advanced
Features - enabled

* Exact versions TBD

RHEV Hypervisor 7 Features

- Adopted by RHEV
 - Many OpenStack new features → hardened by RHEV
- Adopted by RHEL OpenStack Platform
 - KVM is the #1 hypervisor for OpenStack

RHEV Hypervisor – best engineered distribution of KVM!

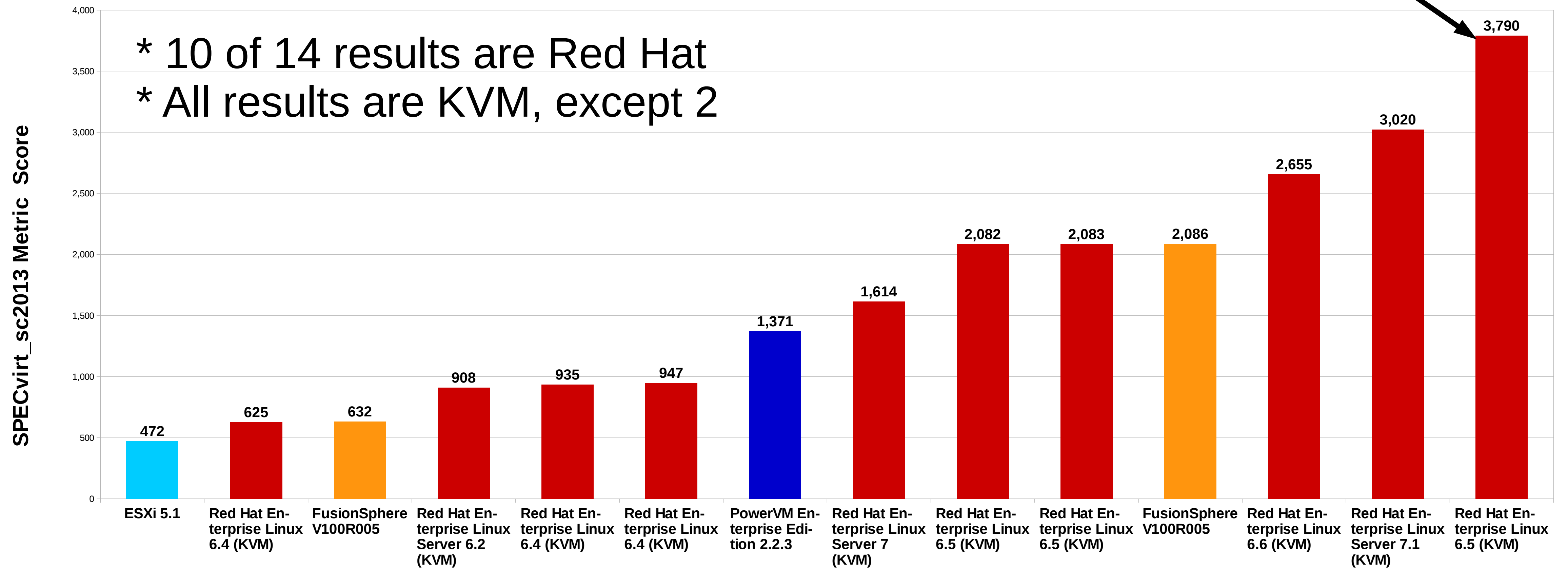
RHEV Hypervisor Performance & Scaling

RHEV Hypervisor – Fastest and Biggest!

All SPECvirt_sc2013 Results Sorted by Performance

(As of June 10, 2015)

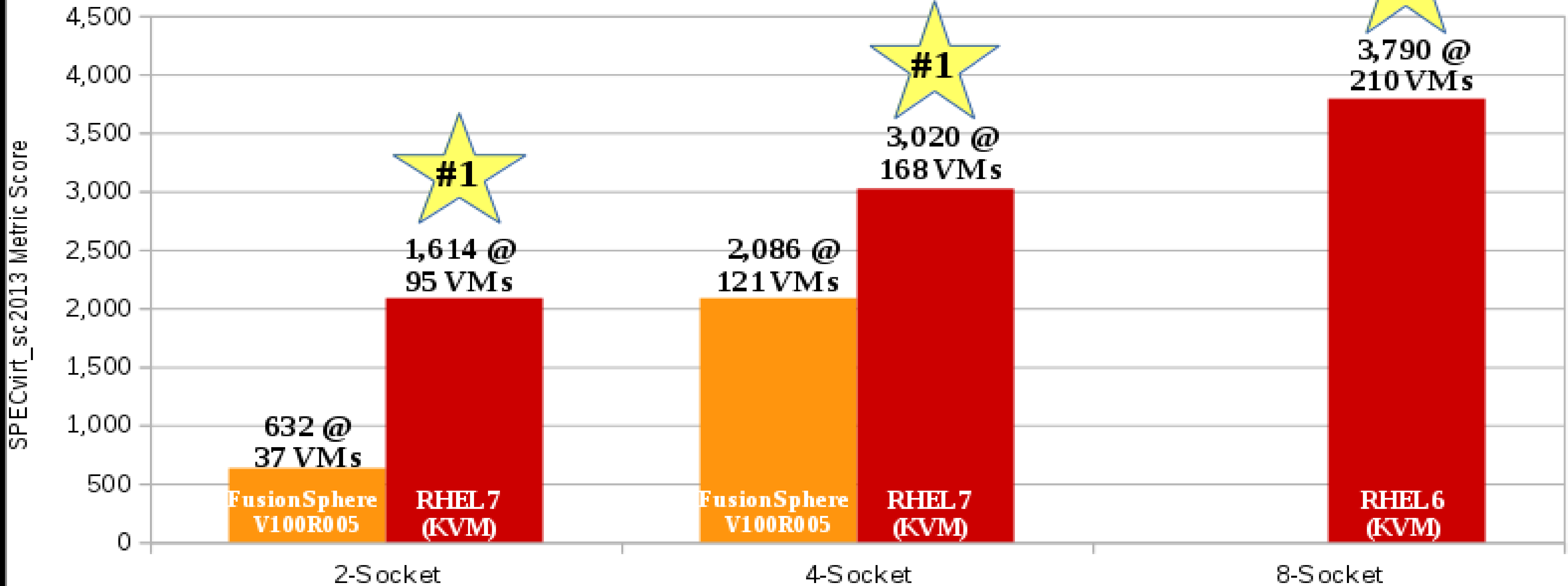
Only 8 socket result



SPEC® and SPEC virt_sc are registered trademarks of the Standard Performance Evaluation Corporation. For more information about SPEC and its benchmarks see www.spec.org. For details about the SPEC virt_sc 2013 benchmark results in this chart see www.spec.org/virt_sc2013/results.

SPEC virt_sc® 2013 Benchmark Leadership

Highest Red Hat based result versus highest non Red Hat result by socket
(As of June 10, 2015)



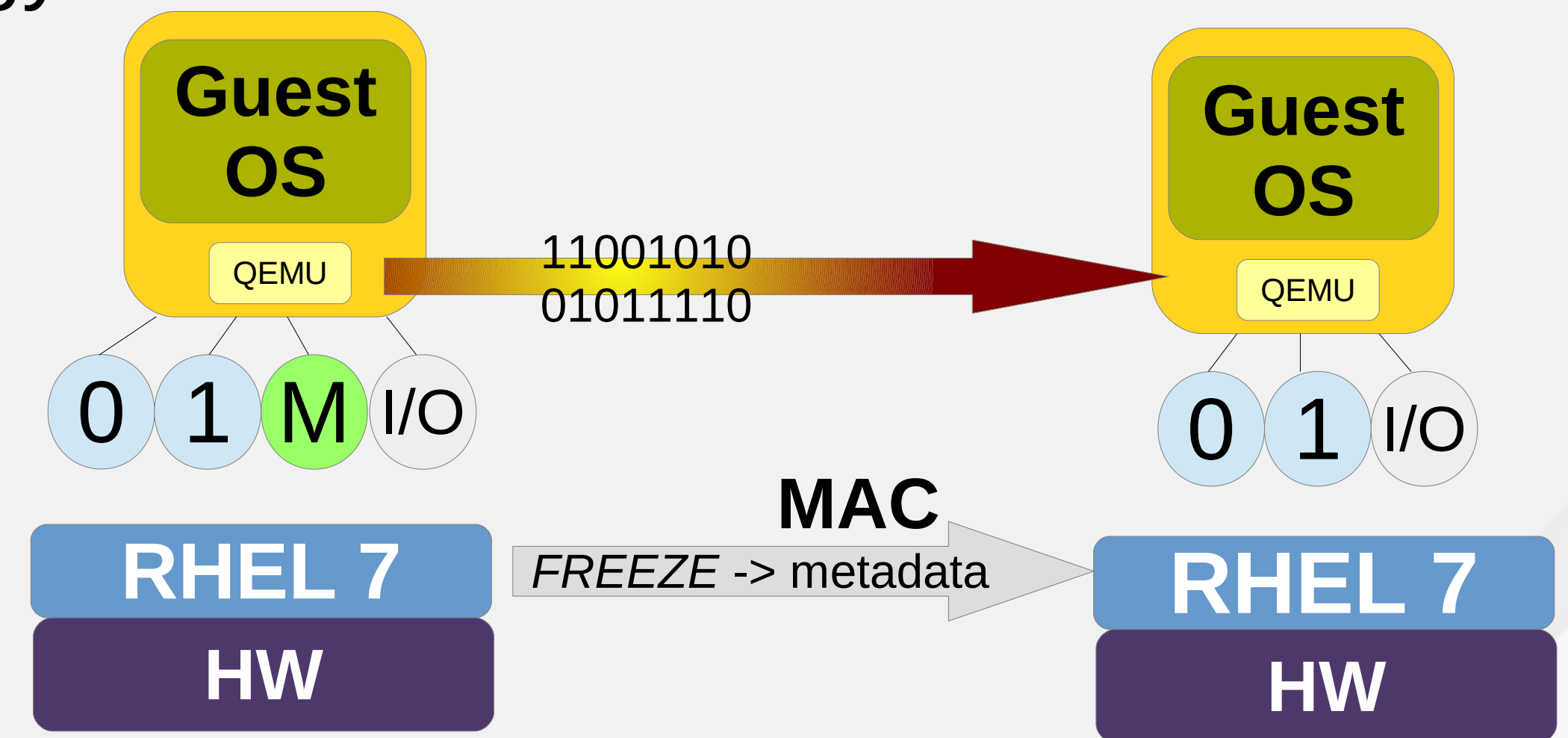
SPEC® and SPEC virt_sc are registered trademarks of the Standard Performance Evaluation Corporation. For more information about SPEC and it's benchmarks see www.spec.org. For details about the SPEC virt_sc 2013 benchmark results in this chart see www.spec.org/virt_sc2013/results.

RHEV Hypervisor 7 – Feature Roadmap

KVM, now and in the Future!

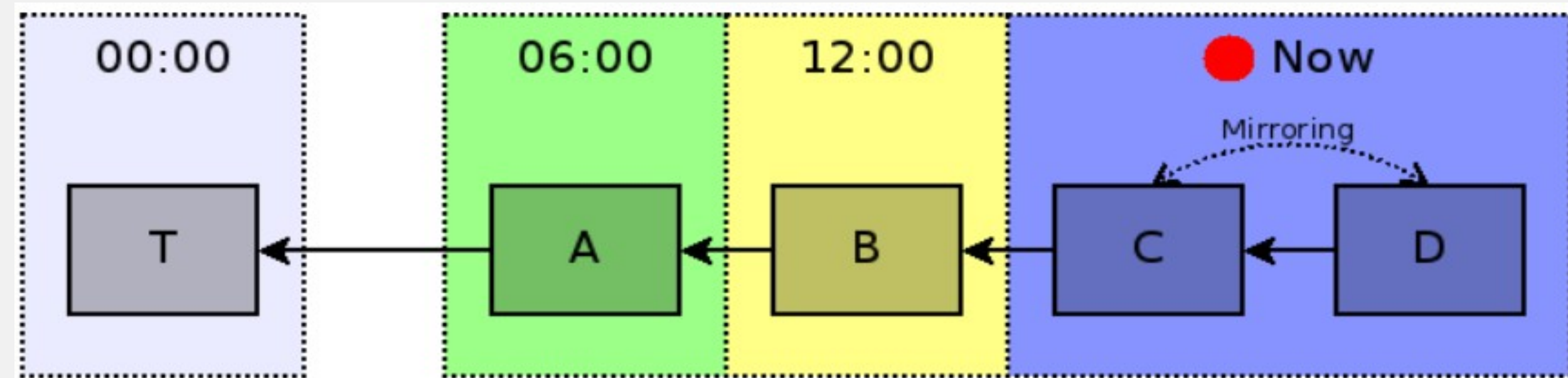
Live Migration – RHEV 3.5 & RHEL OSP 6

- **RHEL 6 → RHEL 7 hosts**
- **RHEL 7.x ↔ RHEL 7.y hosts**
- RHEL 6 machine types supported on RHEV Hypervisor 7
- Massive improvements in live migration technology!
- Joint engineering collaboration
 - Red Hat: RHEV, Nova, libvirt, QEMU, Virt QE
 - Key partners



Live Snapshot Active Merge – RHEV 3.5.1

- Live snapshots: capture disks and memory at a point in time - qcow2 volume chains
- RHEV Hypervisor 7.1
 - Block mirroring
 - Live snapshot merge, forward \Leftarrow \Rightarrow backward, internal or **active layer**
 - Live snapshot delete
- Joint engineering collaboration
 - Red Hat: RHEV-M, vds, libvirt, QEMU, field account managers
 - Customer requirements



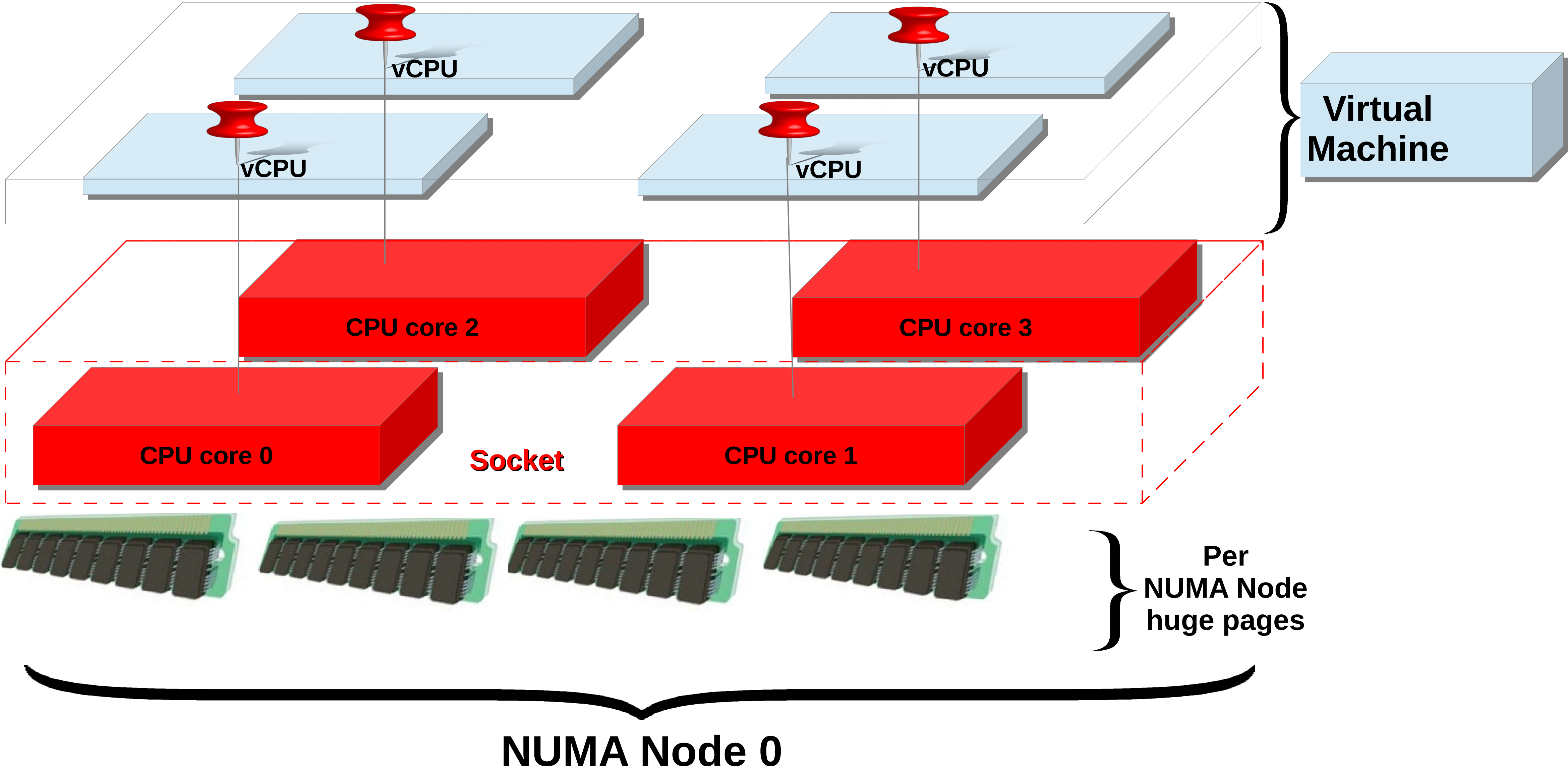
RHEV Hypervisor 7 - NFV Roadmap (now + future)

- NFV Configurations:
 - Config I – NUMA Pinning, Topology Awareness, SR-IOV, DPDK
 - Config II – DPDK with Open vSwitch, vhost-user
 - Future – In planning
 - OPNFV: https://wiki.opnfv.org/nfv_hypervisors-kvm
- Real-time KVM – low avg/max latency requirements

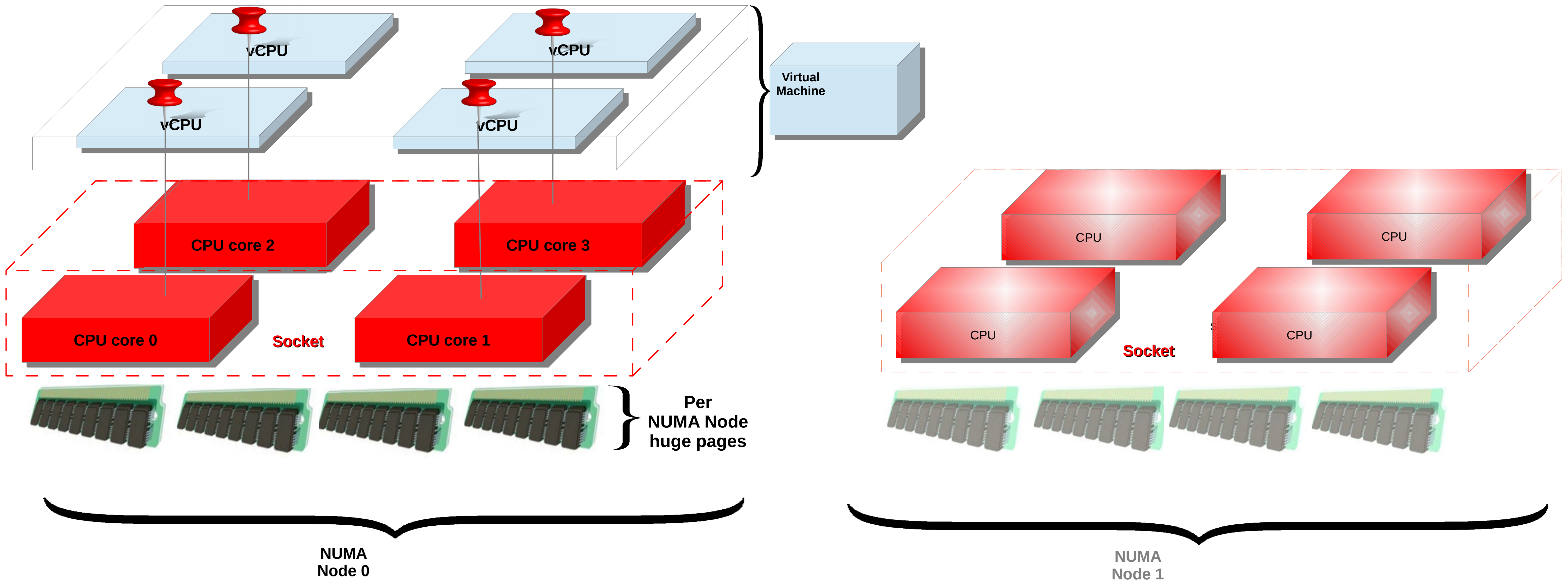
NUMA Pinning/Topology Awareness, SR-IOV

- RHEL OpenStack Platform 7 (plan)
 - Nova scheduler NUMA support – based on PCI bus location
- RHEV Hypervisor 7.1
 - VFIO device assignment (PCI passthrough)
 - Libvirt API – host NUMA topology
 - Huge pages allocated to VM – 2MB and 1GB – NUMA awareness
 - Guest NUMA node \Leftrightarrow host NUMA node alignment
- Joint engineering collaboration
 - Red Hat teams: OpenStack Nova, libvirt, QEMU, kernel, perf
 - NFV partners

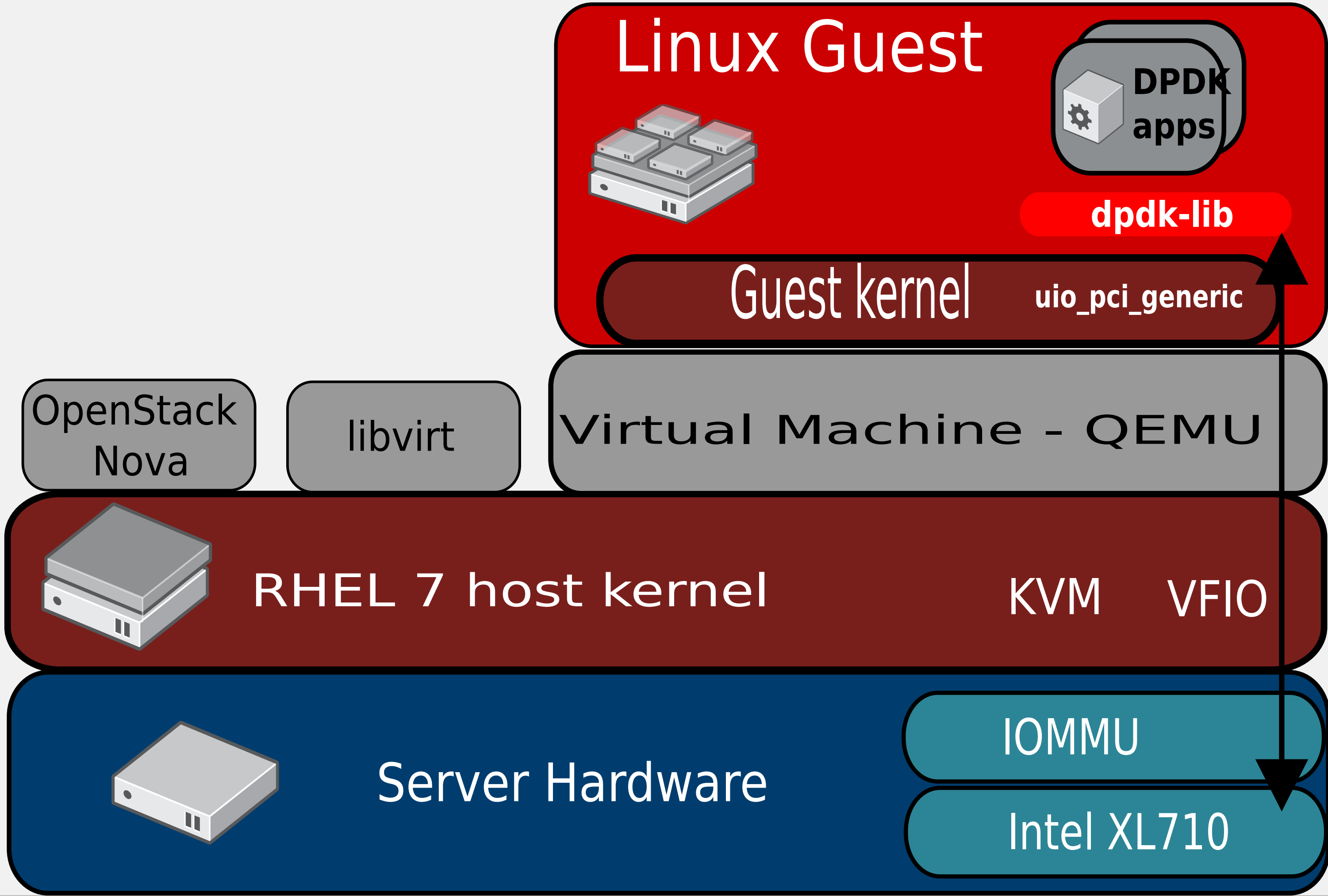
NUMA Pinning and Topology Awareness



NUMA Pinning and Topology Awareness – two sockets



Config I - DPDK with VFIO device assignment (future)

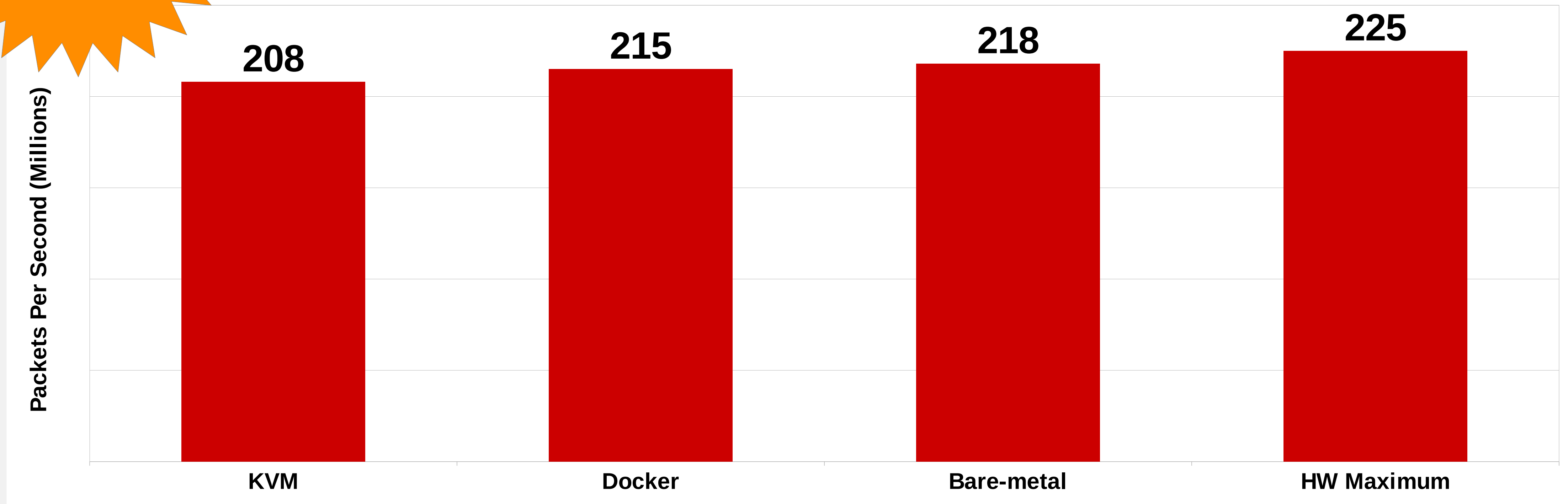


Network Function Virtualization (NFV) Throughput and Packets/sec (RHEL7.x+DPDK)

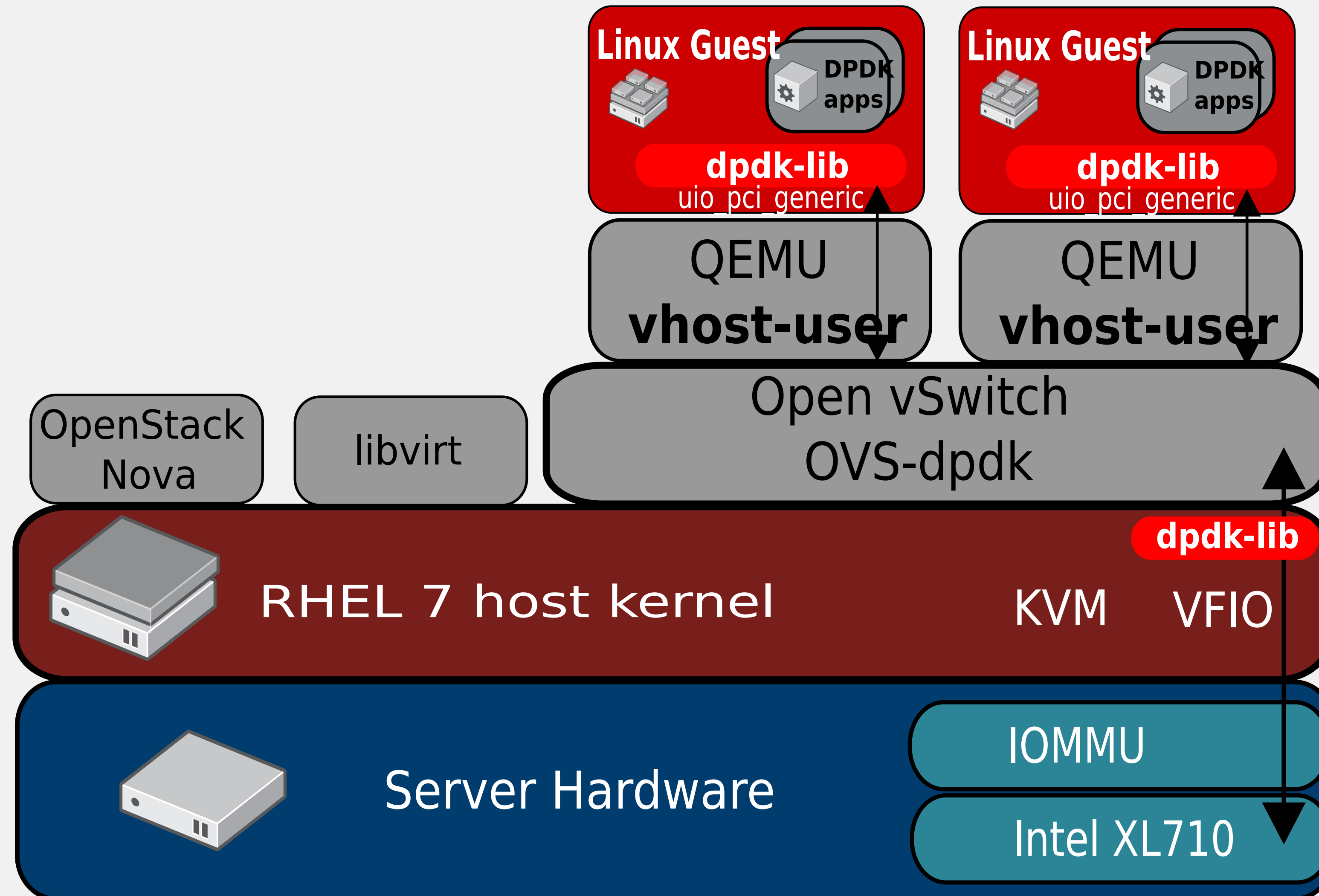
208Mpps+

NFV: Millions of Packets Per Second

RHEL7.x, L2 Forwarding, 12 x 40Gb NICs

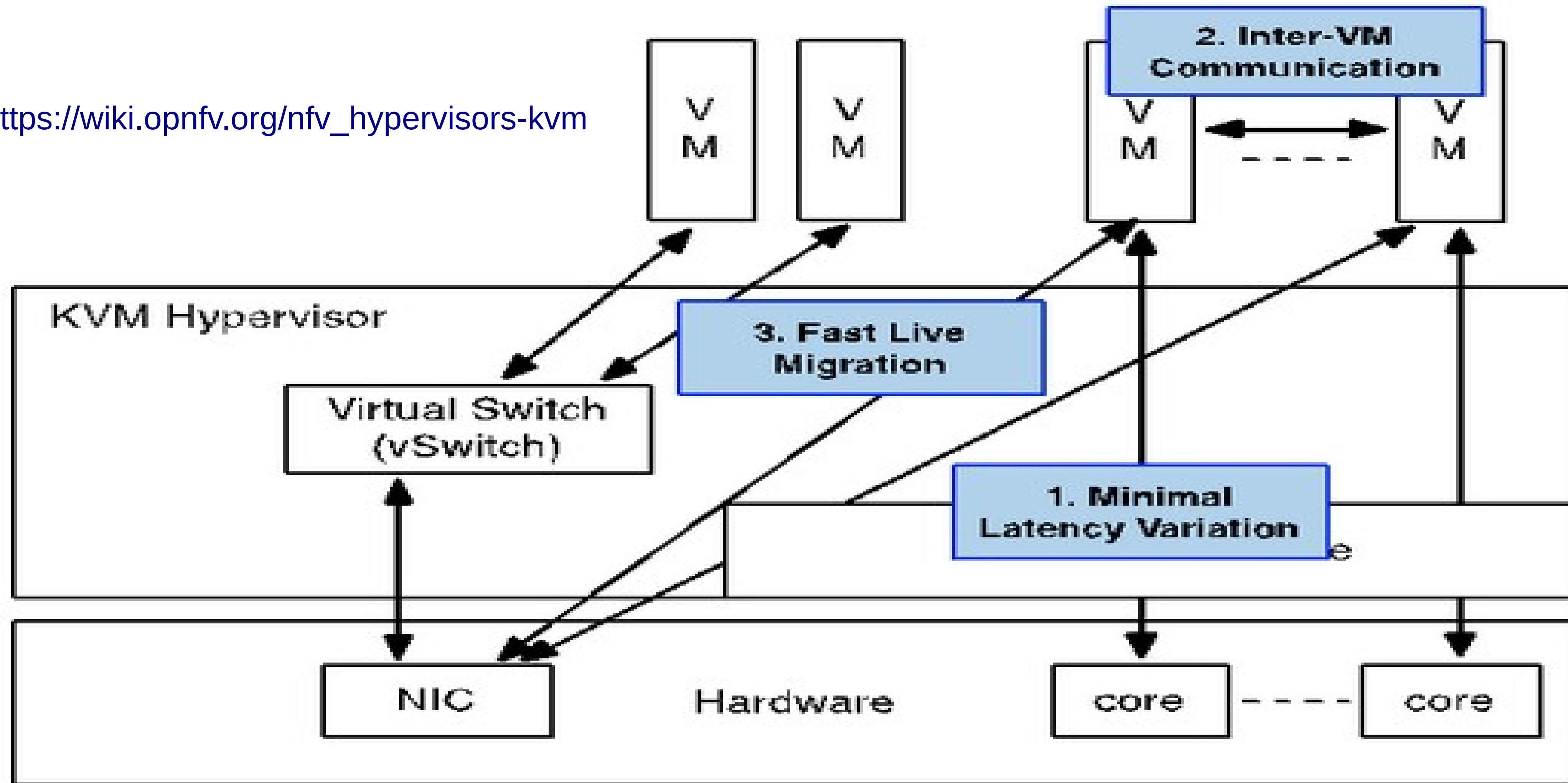


Config II - DPDK w/Open vSwitch, vhost-user (future)



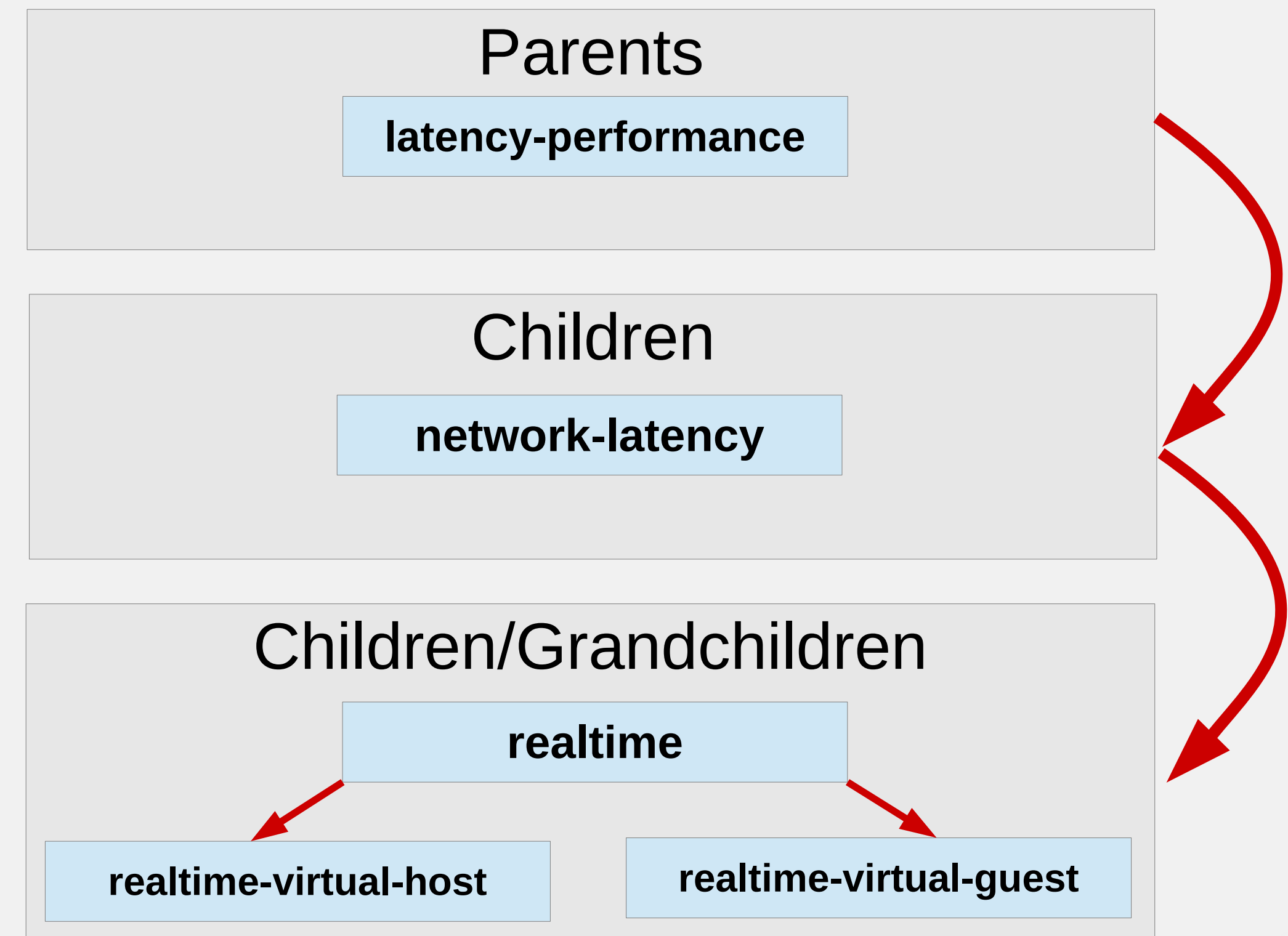
OPNFV KVM Hypervisor Project (future)

https://wiki.opnfv.org/nfv_hypervisors-kvm



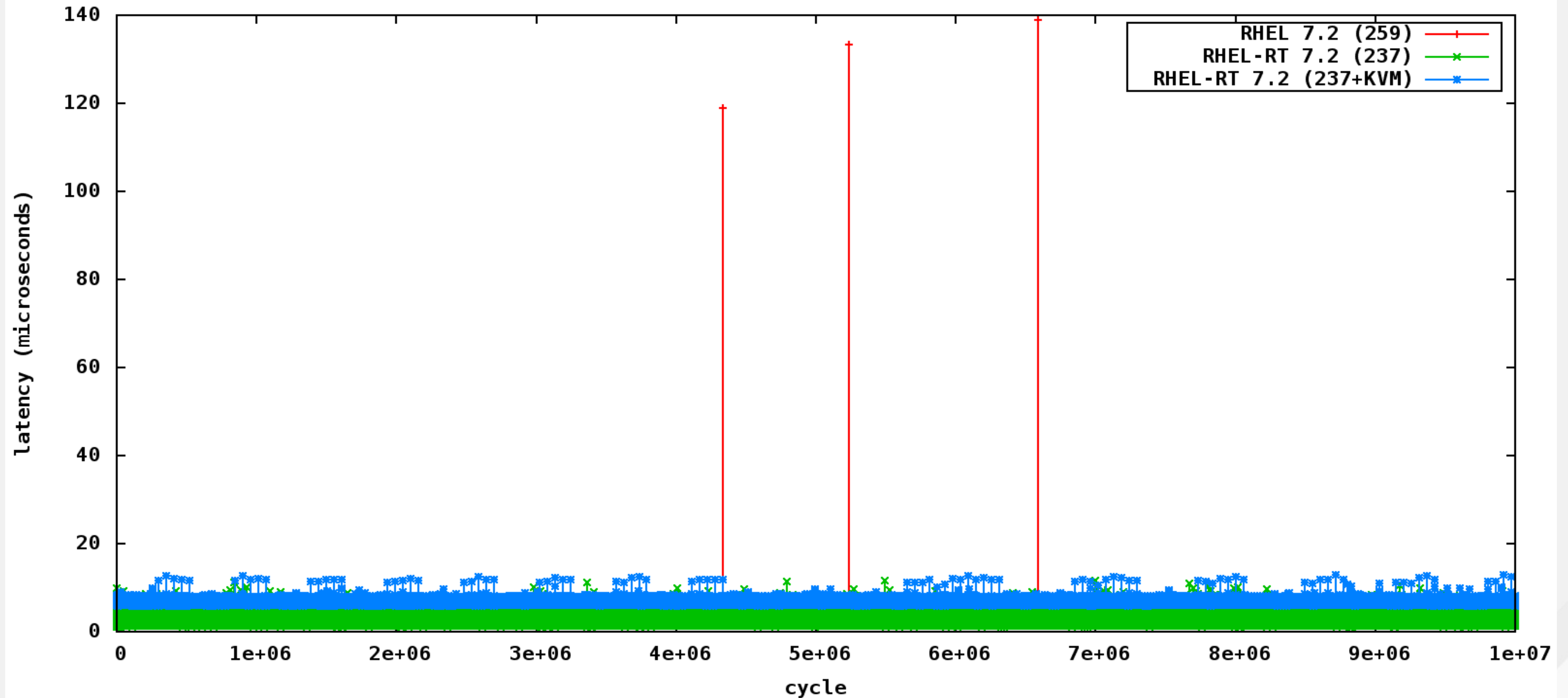
Real-Time KVM for NFV (future)

- Low average and max latency requirements – **no latency spikes!**
- RHEL-RT in host and guests
- New tuned profiles in host and guest
- RHEV Hypervisor 7.2 (plan)
 - KVM and kernel patches
 - Libvirt support
- Joint engineering collaboration
 - Red Hat: Nova, libvirt, KVM, kernel
 - DPDK, Open vSwitch, Real-time
 - performance, tuned
 - NFV Partners



RHEL7.x Real-time Scheduler Latency Jitter Plot

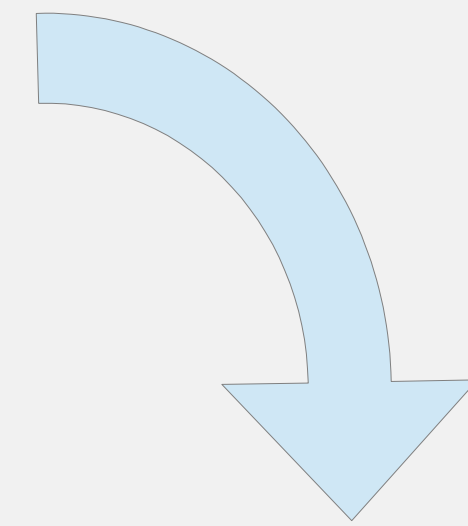
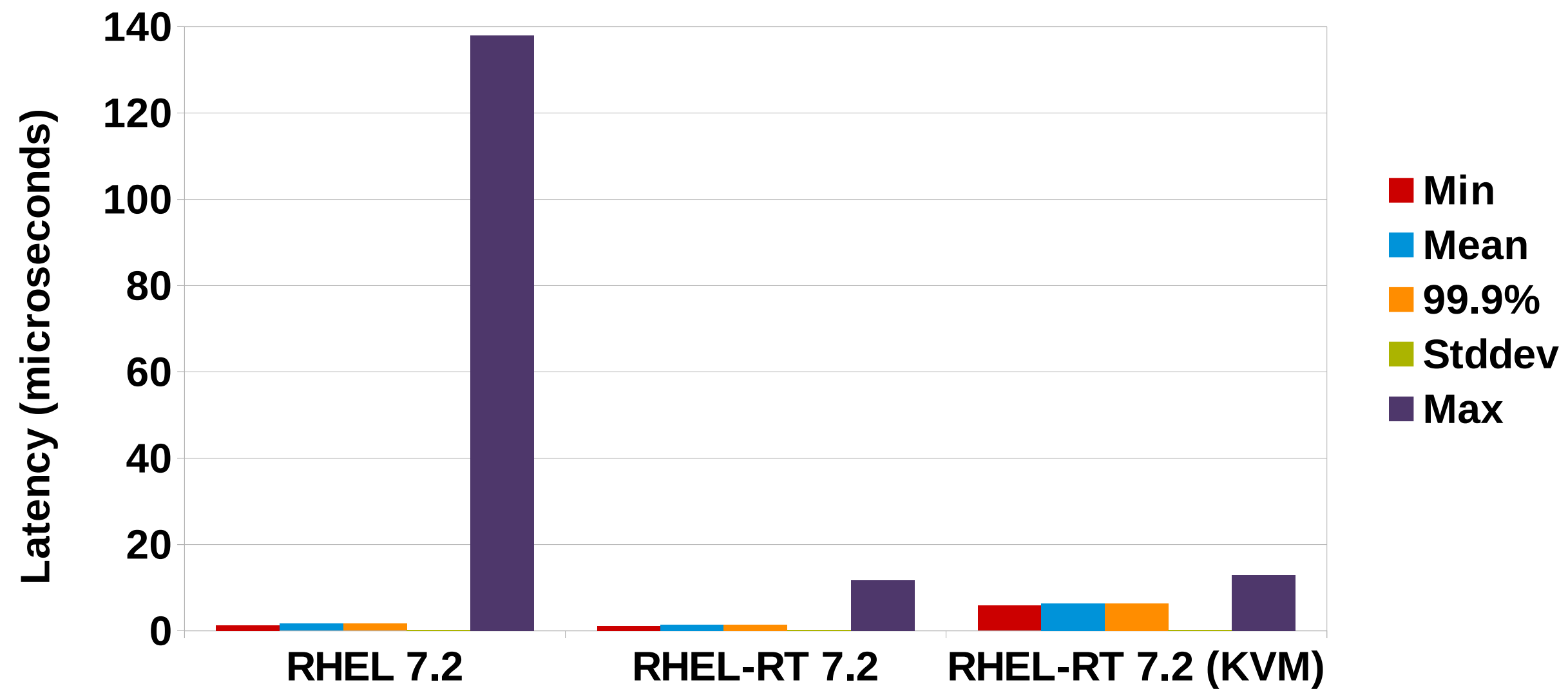
cyclictest -m -n -N -q -v -p95 -h60 -i 200 -D 1h



RHEL7.x Scheduler Latency (cyclicttest)

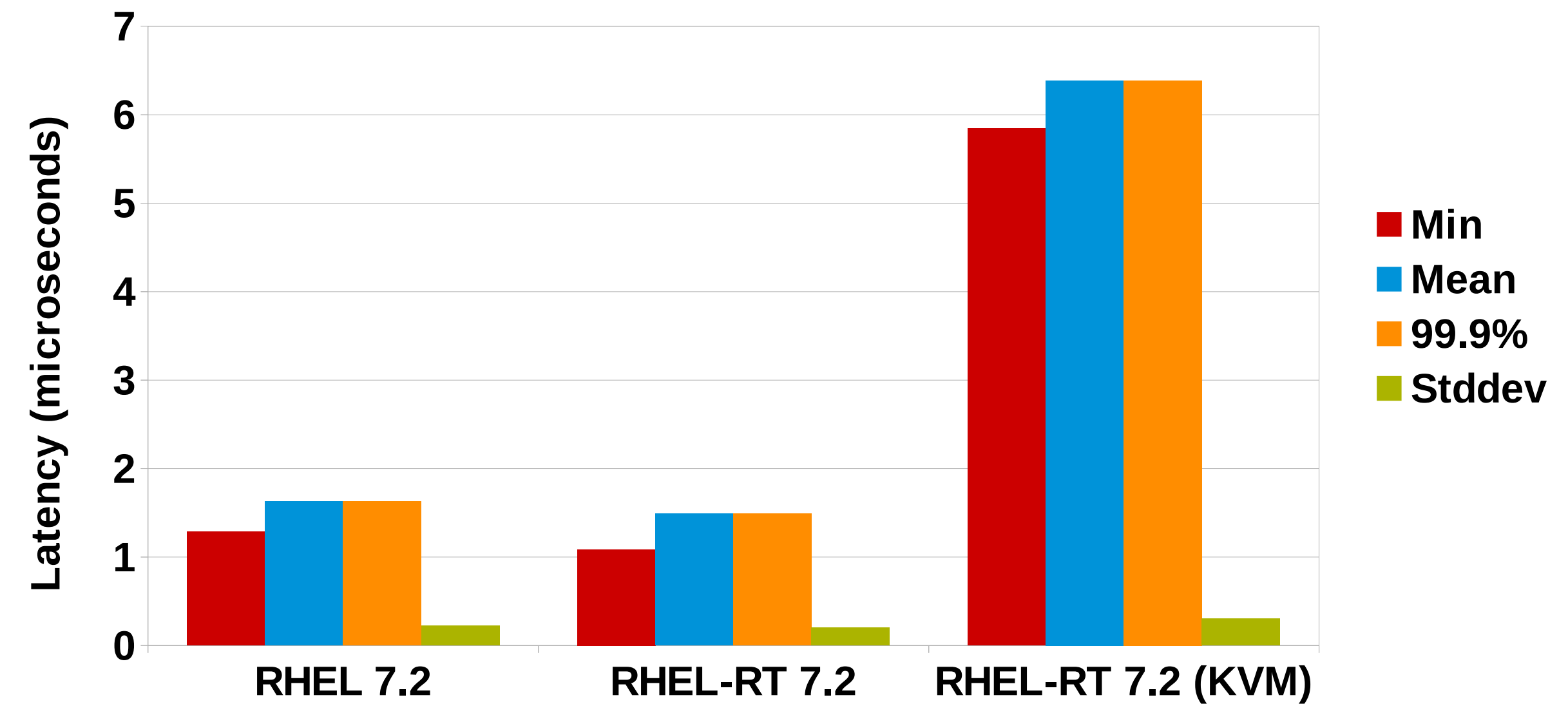
Intel Ivy Bridge 2.4 Ghz, 128 GB mem

Cyclicttest Latency



Remove maxes to zoom in

Cyclicttest Latency



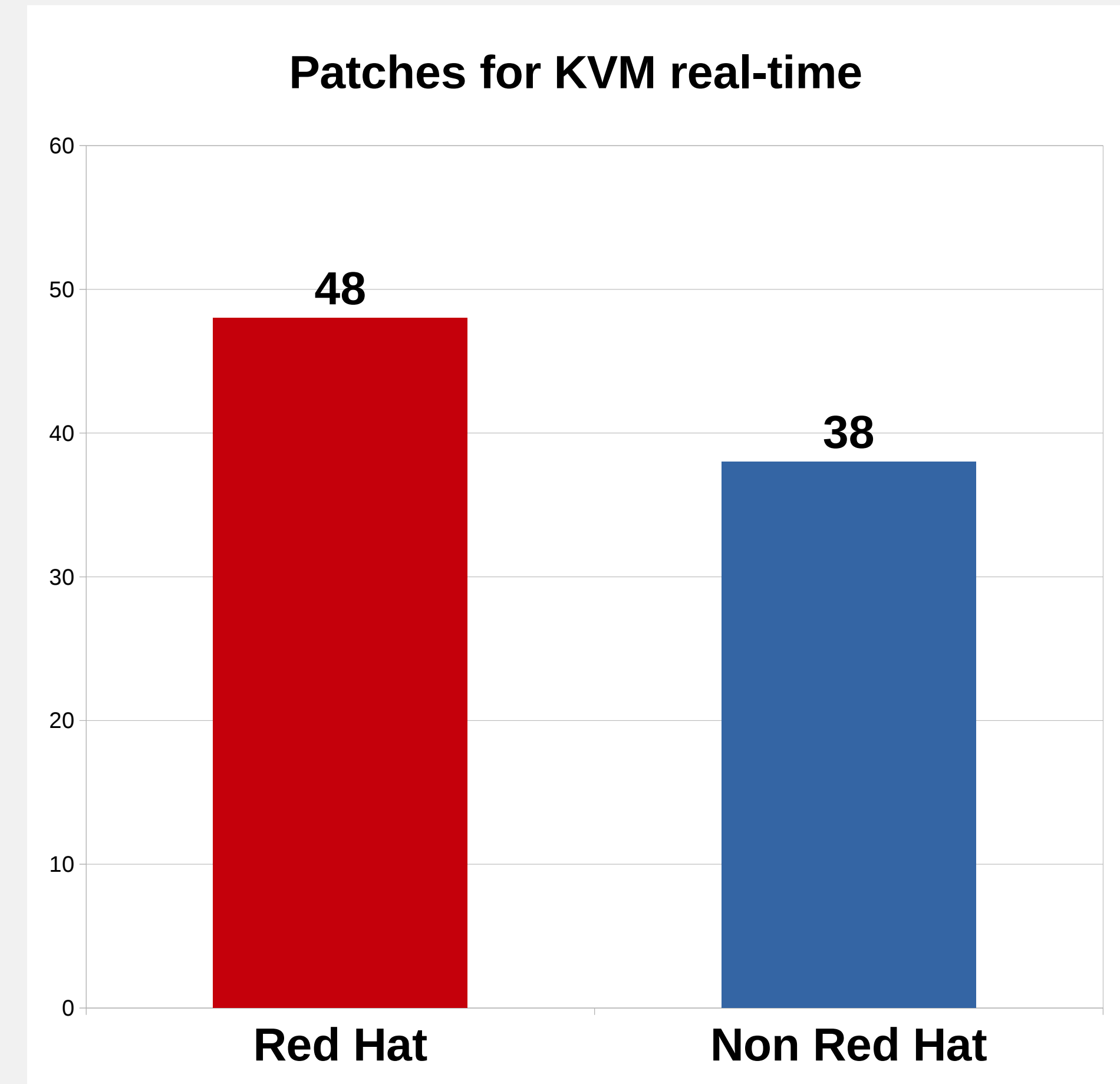
KVM Real-Time Patches (as of 18-Jun-2015)

- Kernel patches

- vmstat: 6
- timer: 4
- sched: 45
- rcu, nohz, kvm: 12
- cpusets, isolcpus: 2
- workqueue: 4
- kvm: 4
- kernel-rt: 4

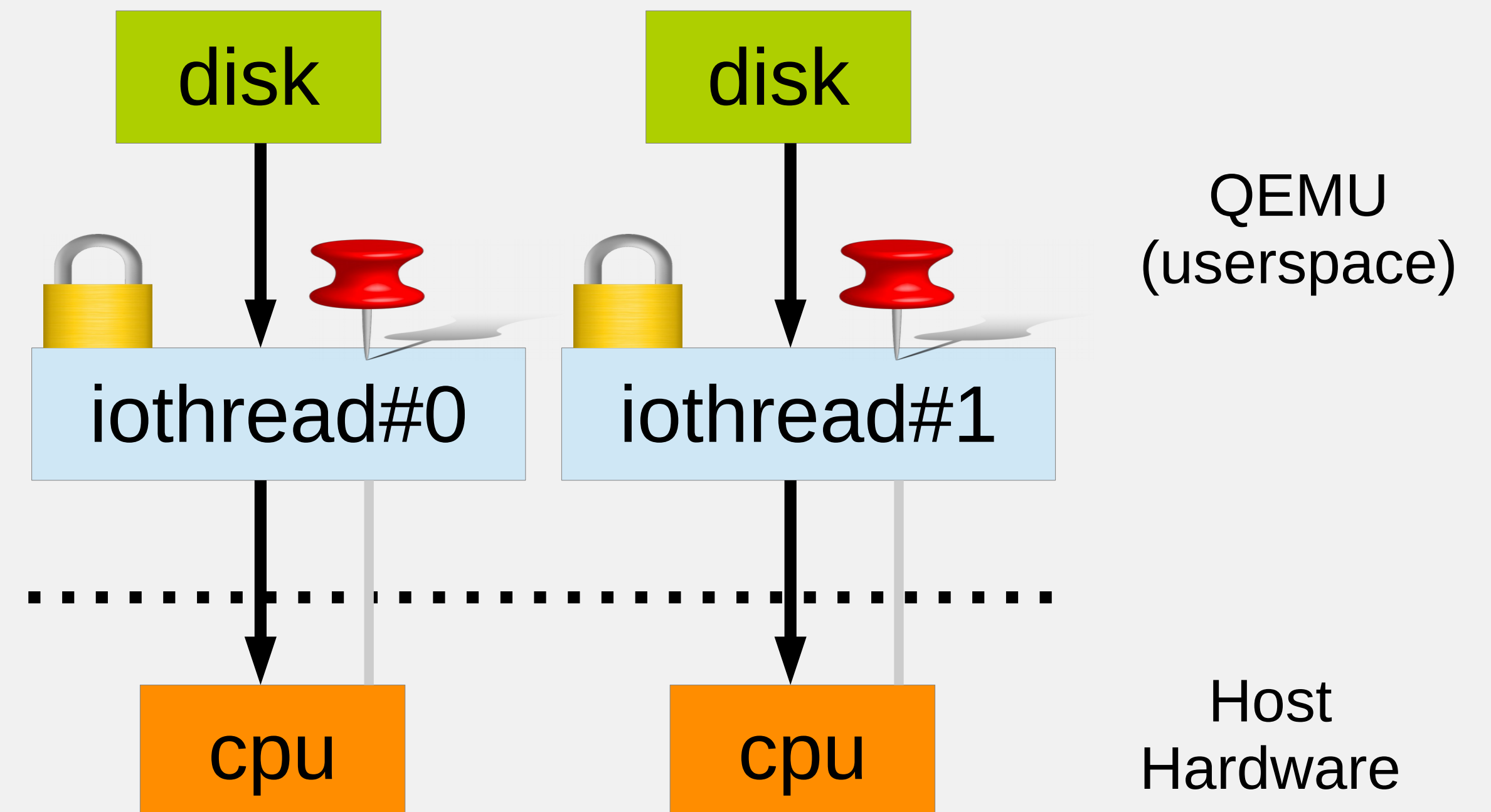
- Non-kernel patches

- irqbalance: 3
- tuned profile: 9
- tuned adm: 1
- python-linux-procfs: 1
- libvirt: 6

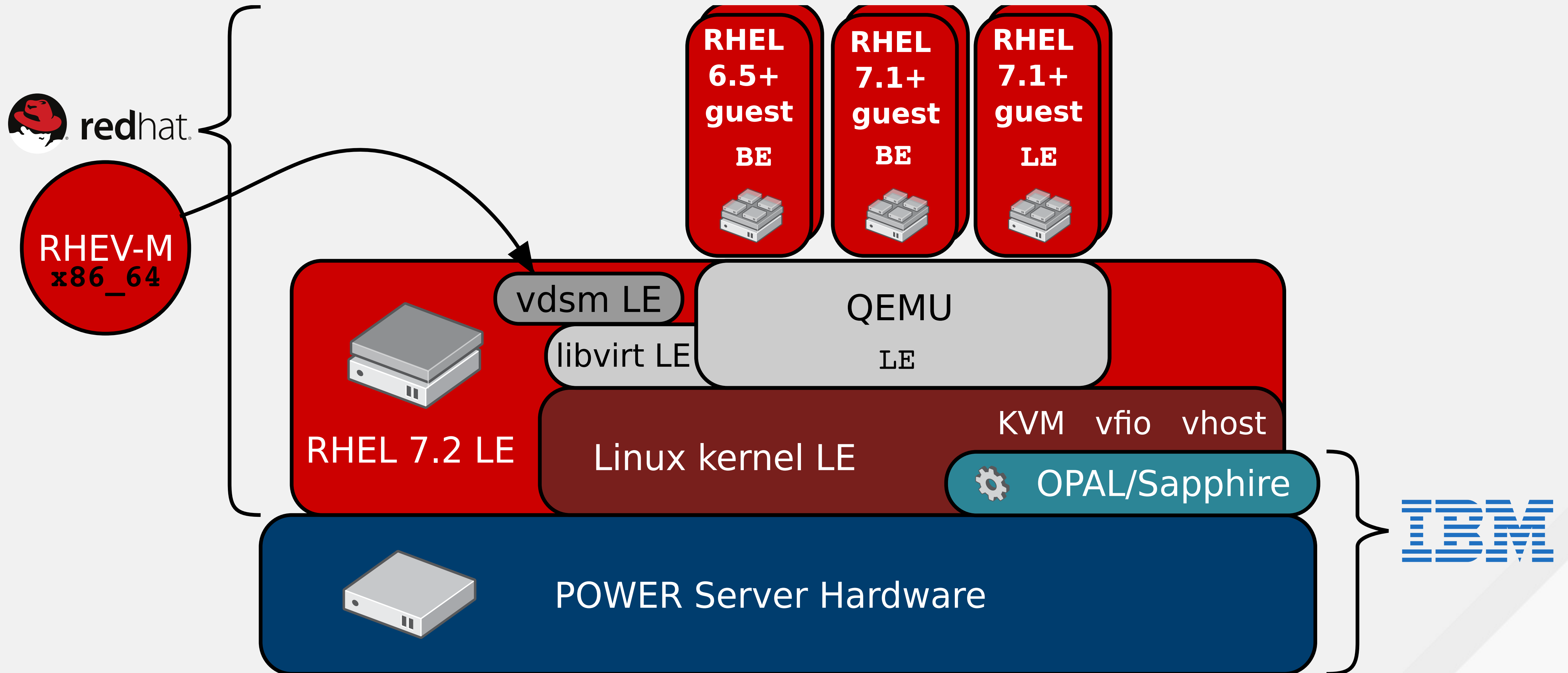


Virtual Storage Improvements (future)

- Single VM scaling w/IOthreads:
 - fine-grained locking in QEMU
 - virtio data-plane
 - IOthread pinning
 - thread per disk/initiator or host CPU
- Multi-queue QEMU block layer
- Fine grained block layer config



RHEV for Power Little Endian (future)



RHEV Hypervisor 7 - highlights

RHEV host	Integration	Feature
RHEL 7+	RHEV and OSP	Live migration from RHEL 6 to RHEL 7 host
RHEL 7+	RHEV and OSP	Live migration improvements
RHEL 7.1	RHEV 3.5.1	Live snapshot active merge
RHEL 7.1	RHEL OSP 7*	NFV: huge pages, NUMA, pinning, SR-IOV
RHEL 7.2*	RHEL OSP 8*	NFV: vhost-user – for OVS and DPDK
RHEL 7.2*	RHEL OSP 8*	NFV: real-time KVM
RHEL 7.2*	RHEV 3.6*	IOthreads – virtio-blk data-plane
RHEL 7.2*	RHEV 3.5/6*	RHEV for Power, RHEV Hypervisor LE Host

*RHEV Hypervisor, RHEV and OSP versions are specified for planning purposes only

RHEV Hypervisor – future plans

RHEV host	Integration	Feature
RHEL 7.2*	RHEV 3.6*	Flexibility: Memory hot-plug
RHEL 7.2*	RHEV and OSP	Windows 10, Server 2016 guest support
RHEL 7.3*	RHEV	IOthreads – virtio-scsi
RHEL 7.3*	RHEV	Error reporting/recovery: pci-express
RHEL 7.3*	RHEV	OVMF with UEFI secure boot
RHEL 7.3*	RHEV	Post-copy live migration
RHEL.next	RHEL OSP	OpenStack on ARM
RHEL.next	RHEV and OSP	Nested KVM – tech preview
RHEL.next	RHEV and OSP	Virtual GPU (compute and graphics)

*RHEV Hypervisor, RHEV and OSP versions are specified for planning purposes only



Speakers:

- Red Hat
- Linaro
- Google
- INRIA
- ARM
- Siemens
- GE
- IBM
- SUSE
- Fujitsu
- Intel
- Huawei
- Xilinx
- Samsung
- Montavista
- op5

Communities:

- KVM
- QEMU
- libvirt
- Kernel
- OVMF
- OpenStack Nova
- oVirt

RED HAT
SUMMIT

LEARN. NETWORK.
EXPERIENCE OPEN SOURCE.