Red Hat Enterprise Virtualization Hypervisor Roadmap

Bhavna Sarathy
Senior Technology Product Manager, Red Hat
HYPERVERSOR

Virtual Machines

OEMU

USERSPACE LIBRARIES & RUNTIME

RED HAT ENTERPRISE LINUX KERNEL

KVM KERNEL MODULE
KVM is the foundation Virtualization technology in multiple Red Hat products.
Red Hat Enterprise Virtualization Hypervisor
Hypervisor Differences

<table>
<thead>
<tr>
<th>QEMU-KVM</th>
<th>QEMU-KVM-RHEV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Included in:</td>
<td>Included in:</td>
</tr>
<tr>
<td>Red Hat Enterprise Linux</td>
<td>RHEL Open Stack Platform</td>
</tr>
<tr>
<td>Red Hat Enterprise Virtualization</td>
<td></td>
</tr>
</tbody>
</table>

**Additional features**
- Live Snapshots
- Live Storage Migration
- Live Snapshot Merge
- Block I/O Throttling
- CEPH Enablement
- OpenvSwitch

**Certification etc**
- Microsoft SVVP Certification
- Support for SLES guests
- Support for unlimited VMs
Hypervisor Architecture: RHEL with KVM

Virt-Manager GUI

Virsh command line tool

libvirt API
RHEV Hypervisor Architecture: RHEV

- QEMU-KVM-RHEV
- OPENSTACK COMPUTE AGENT
- OPENVSWITCH
- USERSPACE LIBRARIES & RUNTIME
- RED HAT ENTERPRISE LINUX KERNEL
- KVM KERNEL MODULE
SELinux – Security in RHEV Hypervisor

VM separation with SELinux sandbox

When one VM is attached.....

..and compromised, host and VMs safe
Network I/O Architecture: Emulated and Virtio
Virtio

- Framework and set of drivers:
  - A hypervisor-independent, domain-independent, bus-independent protocol for transferring buffers
  - A binding layer for attaching virtio to a bus (e.g. PCI)
  - Domain specific guest drivers (networking, storage, etc.)
  - Hypervisor specific host support

- VirtIO support is included in modern Linux kernels
  - Linux guest will autodetect VirtIO enabled host and use optimized drivers
  - Included in upstream kernels since 2007. In RHEL since 4.8, 5.3 and later
Network I/O Architecture: Direct device assignment & SR-IOV

Direct Device Assignment

SR-IOV
Storage Architecture

Virtual Disks

Physical Storage
RHEV Hypervisor Themes and Features
RHEV Hypervisor Themes

- Performance
- Scalability
- Reliability
- Security
New Features in RHEL 7

- **Performance**
  - Automatic NUMA Balancing
  - Hyper-V enlightenment in KVM

- **Scalability**
  - Guest memory limit

- **Reliability**
  - Live migration improvements
  - PCI Express error containment
  - CPU hot plug
New Features in RHEL 7

- **Security**
  - Random Number Generator (virtio-rng)
  - Syscall filtering with Seccomp

- **Storage**
  - Multi-queue SCSI
  - Gluster backend integration

- **Networking**
  - Multi-queue NET

- **Graphics**
  - PCI Passthrough
Performance: Automatic NUMA Balancing

Great out of the box performance for virtualized workloads

• RHEL 6 numad - user space solution
  • Heuristics based performance
  • Automatically binding processes to NUMA nodes that have available resource capacity

• RHEL 7 kernel Automatic NUMA Balancing
  • Coordination and integration between
    • Memory management
    • CPU Scheduler

Enabled by default
Performance: Optimized Automatic NUMA Balancing

Node 0

VM1vcpu1
VM1vcpu2

Node 1

VM2vcpu1
VM2vcpu2
Best SPECvirt_sc2010 Scores by CPU Cores

(As of April 3, 2014)

The SPECvirt_sc2010 benchmark has been retired and no new results may be published.

SPEC® is a registered trademark of the Standard Performance Evaluation Corporation. For more information about the benchmark and the results, see http://www.spec.org/virt_sc2010/.

#redhat #rhsummit
Best SPECvirt_sc2013 Scores by CPU Cores

Red Hat claims the top 5 SPECvirt_sc2013 benchmark results as well as the only 4-Socket results.

All SPECvirt_sc2013 benchmark results published as of April 2, 2014. SPEC® is a registered trademark and SPEC virt™ is a trademark of the Standard Performance Evaluation Corporation. For more information about the benchmark and the results, see http://www.spec.org/virt_sc2013/.
Performance: Windows Guest Improvements

• Extend RHEV Hypervisor to support Hyper-V Enlightenment
• Make Windows guests think they are running on Hyper-V
• Per Microsoft Hypervisor Functional Specification
  • Relaxed timing: hv-relaxed
  • Virtual APIC: hv-vapic
  • Scheduling (spinlocks): hv-spinlocks
  • Time: hv-time
Scalability: Large Guests

Large guests are necessary for efficient hardware abstraction

- **RHEV Hypervisor Host Limits**
  - KVM inherits the RHEL host limits
  - Host: 160 cores; 4TiB RAM

- **RHEV Hypervisor Guest Limits**
  - Virtual Machine CPU Limit: 160 vCPUs
  - Virtual Machine Memory Limits
    - RHEL6 4000GiB guest RAM
    - RHEL7 4 TiB guest RAM
Scalability: Guest Virtual CPU

VMware and Red Hat - Virtual CPU

Virtual CPUs

Releases

RHEL 5.5, ESX 4.1
RHEL 6.0, ESX 5.0
RHEL 6.3, ESX 5.1
RHEL 6.5, ESX 5.5

VMware ESX
Red Hat KVM
Scalability: Guest Virtual Memory

VMware and Red Hat - Virtual Memory

Virtual Machine Memory vs. Releases

- VMware ESX
- Red Hat KVM

Releases:
- RHEL 5.5, ESX 4.1
- RHEL 6.0, ESX 5.0
- RHEL 6.3, ESX 5.1
- RHEL 6.5, ESX 5.5
Reliability: Live Migration

Live migration

RHEV Hypervisor

VM

Storage Live migration

RHEV Hypervisor

VM
Reliability: Live Migration Optimizations

- Live migration threads
- Bitmap Cleanup
- Single buffer copy
- Delta Compression
- Zero page Optimization
- Migration Statistics
- Auto-convergence (tech-preview)
- RDMA Live migration (tech-preview)
Reliability: Live Migration Delta Compression

Source Host

Updated Page

Original Page

Delta Compression

Updated Page

Destination Host

Applying Delta

Original Page

Delta

Delta
Reliability: Virtual CPU hot-plug

• Dynamically add additional vCPUs to running VMs
  • Flexibility, meet SLAs
• No application downtime to adjust VM's compute capacity
Security: Cryptography

• Para-Virtual Random Number Generator (RNG)
  • Provide true randomness in the guest for cryptographic purposes
  • RHEV Hypervisor feeds entropy to the virtual machines
  • Helps alleviate entropy starvation in guests
Security: QEMU Sandboxing

• Further reduce guest attack surface in your data center or cloud!
  • Seccomp - library for application level syscall filtering
  • QEMU confined to the minimal set of syscalls
Storage: QEMU block with libgfapi access

GLUSTER FS STORAGE DOMAIN
Use case: GlusterFS volumes host VM images and data

• **QEMU – GlusterFS Integration**
  • Native integration to GlusterFS volumes (in QEMU)
  • Replaces locally mounted FUSE file system
  • Access over network to Gluster storage volumes
• **Libgfapi**: User-space library for accessing data in GlusterFS
  • Perform IO on gluster volumes directly
  • 2x performance improvement in disk IO operations while accessing gluster storage in KVM guests
Storage: Virtio-scsi Multi-queue

• **Virtio-scsi** - A virtual SCSI host bus adapter (HBA) via virtio
  • Capable of handling hundreds of devices, compared with virtio-blk which can only handle 25 devices (exhausts PCI slots).
  • Attach a virtual hard drive or CD through the virtio-scsi controller
  • Pass-through a physical SCSI device from the host to the guest via the QEMU scsi-block device
  • Supports all standard SCSI commands, and boot disks

• Adds virtio-scsi multi-queue
  • Simple queue steering algorithm
Networking: Multi-queue NIC through virtio-net

• Improves network performance in virtio-net
• Multiple-queue transmit and receive support
  • Backend and hypervisor support
  • Guest drivers
• Benefit: Improves network performance and throughput for SMP guests
New device assignment mechanism in RHEL 7

- Virtual Function I/O (VFIO)
  - New user space driver interface – used by QEMU
  - Allows for cleaner PCI device assignment architecture
  - Replaces legacy code in RHEL6
  - More maintainable, better architected, more secure/isolated
- GPU device assignment to Virtual Machines
  - For both graphics and compute
Graphics: Dedicated GPU device assignment

Virtual Machine

Guest OS
- Application
- Graphics driver

QEMU
VFIO
KVM

RHEL Kernel

GPU card
RHEV Hypervisor Future Features
Graphics Futures with RHEV Hypervisor
RHEV Hypervisor Future Features

- Memory hot-plug
- Virtio-blk data-plane
- USB 3.0
- Live migration Improvements
- PCI Express Bus
- UEFI
Telco Futures with RHEV Hypervisor
Key Takeaways

• KVM foundational hypervisor to RHEL, RHEV and RHEL-OSP
  • Use cases: hardware abstraction, large guests, single host
• RHEV Hypervisor support additional features
  • Data center use case with Red Hat Enterprise Virtualization
  • Cloud use case with Red Hat Enterprise Linux Open Stack Platform
• Two KVM streams
  • RHEL - bug fixes and new HW support
  • Fast moving in RHEV and RHEL-OSP (increased feature cadence)
Questions?
Backup
<table>
<thead>
<tr>
<th>RHEL</th>
<th>QEMU 0.12</th>
<th>Backported Features</th>
<th>Feature Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHEL 6</td>
<td>QEMU 0.12</td>
<td>Backported Features</td>
<td>• Rebased QEMU 0.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Vhost-net Qcow2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Common Access Card (CAC) support</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Sound, Intel Block IO tunables in libvirt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• NIC model type definition</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Efficient IO barriers</td>
</tr>
<tr>
<td>RHEL 6.1</td>
<td>QEMU 0.12</td>
<td>Backported Features</td>
<td>• Vhost-net Qcow2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Common Access Card (CAC) support</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Sound, Intel Block IO tunables in libvirt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• NIC model type definition</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Efficient IO barriers</td>
</tr>
<tr>
<td>RHEL 6.2</td>
<td>QEMU 0.12</td>
<td>Backported Features</td>
<td>• Vhost-net Qcow2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Common Access Card (CAC) support</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Sound, Intel Block IO tunables in libvirt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• NIC model type definition</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Efficient IO barriers</td>
</tr>
<tr>
<td>RHEL 6.3</td>
<td>QEMU 0.12</td>
<td>Backported Features</td>
<td>• Vhost-net Qcow2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Common Access Card (CAC) support</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Sound, Intel Block IO tunables in libvirt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• NIC model type definition</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Efficient IO barriers</td>
</tr>
<tr>
<td>RHEL 6.4</td>
<td>QEMU 0.12</td>
<td>Backported Features</td>
<td>• Vhost-net Qcow2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Common Access Card (CAC) support</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Sound, Intel Block IO tunables in libvirt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• NIC model type definition</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Efficient IO barriers</td>
</tr>
<tr>
<td>RHEL 6.5</td>
<td>QEMU 0.12</td>
<td>Backported Features</td>
<td>• Vhost-net Qcow2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Common Access Card (CAC) support</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Sound, Intel Block IO tunables in libvirt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• NIC model type definition</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Efficient IO barriers</td>
</tr>
</tbody>
</table>

* Technology Preview
<table>
<thead>
<tr>
<th>RHEL 6</th>
<th>RHEL 6.1</th>
<th>RHEL 6.2</th>
<th>RHEL 6.3</th>
<th>RHEL 6.4</th>
<th>RHEL 6.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>QEMU 0.12</td>
<td>QEMU 0.12</td>
<td>QEMU 0.12</td>
<td>QEMU 0.12</td>
<td>QEMU 0.12</td>
<td>QEMU 0.12</td>
</tr>
<tr>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
</tr>
<tr>
<td>Backported Features</td>
<td>Backported Features</td>
<td>Backported Features</td>
<td>Backported Features</td>
<td>Backported Features</td>
<td>Backported Features</td>
</tr>
</tbody>
</table>

Advantages

- Stable base platform
- Cherry pick select features from upstream
- Take advantage of new features without risking leading/bleeding edge
- Proven model based on tried and tested RHEL best practices
<table>
<thead>
<tr>
<th>RHEL 6</th>
<th>RHEL 6.1</th>
<th>RHEL 6.2</th>
<th>RHEL 6.3</th>
<th>RHEL 6.4</th>
<th>RHEL 6.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>QEMU 0.12</td>
<td>QEMU 0.12</td>
<td>QEMU 0.12</td>
<td>QEMU 0.12</td>
<td>QEMU 0.12</td>
<td>QEMU 0.12</td>
</tr>
<tr>
<td>Backported</td>
<td>Backported</td>
<td>Backported</td>
<td>Backported</td>
<td>Backported</td>
<td>Backported</td>
</tr>
<tr>
<td>Bug fixes</td>
<td>Bug fixes</td>
<td>Bug fixes</td>
<td>Bug fixes</td>
<td>Bug fixes</td>
<td>Bug fixes</td>
</tr>
<tr>
<td>Backported</td>
<td>Backported</td>
<td>Backported</td>
<td>Backported</td>
<td>Backported</td>
<td>Backported</td>
</tr>
<tr>
<td>Features</td>
<td>Features</td>
<td>Features</td>
<td>Features</td>
<td>Features</td>
<td>Features</td>
</tr>
</tbody>
</table>

Disadvantages

- Increasing challenging to backport features as upstream codebase moves on
- Very rapid pace of feature innovation compared to other RHEL components
- Data center and cloud virtualization products require rapid feature development

Need to balance enterprise stability with rapid feature cadence
## QEMU-KVM

**Included in**
- Red Hat Enterprise Linux

### Advanced Features

- Backported Bug fixes
- Backported Features

---

## QEMU-KVM-RHEV

**RHEV Hypervisor**

**Included in**
- Red Hat Enterprise Virtualization
- RHEL OpenStack Platform

### Advanced Features

- Backported Bug fixes
- Backported Features

---

<table>
<thead>
<tr>
<th>RHEL 6</th>
<th>RHEL 6.1</th>
<th>RHEL 6.2</th>
<th>RHEL 6.3</th>
<th>RHEL 6.4</th>
<th>RHEL 6.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>QEMU 0.12</td>
<td>QEMU 0.12</td>
<td>QEMU 0.12</td>
<td>QEMU 0.12</td>
<td>QEMU 0.12</td>
<td>QEMU 0.12</td>
</tr>
<tr>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
</tr>
<tr>
<td>Backported Features</td>
<td>Backported Features</td>
<td>Backported Features</td>
<td>Backported Features</td>
<td>Backported Features</td>
<td>Backported Features</td>
</tr>
</tbody>
</table>

---

#redhat  #rhsummit
<table>
<thead>
<tr>
<th></th>
<th>RHEL 7</th>
<th>RHEL 7.1</th>
<th>RHEL 7.2</th>
<th>RHEL 7.3</th>
<th>RHEL 7.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>QEMU-KVM</td>
<td>QEMU 1.5</td>
<td>QEMU 1.5</td>
<td>QEMU 1.5</td>
<td>QEMU 1.5</td>
<td>QEMU 1.5</td>
</tr>
<tr>
<td></td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>QEMU-KVM-RHEV</th>
<th>QEMU 1.5</th>
<th>QEMU 2.1</th>
<th>QEMU 2.4*</th>
<th>QEMU 2.7*</th>
<th>QEMU 3.1*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RHEV HYPERVERSOR</td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
<td></td>
</tr>
</tbody>
</table>

- Red Hat Enterprise Linux
- Red Hat Enterprise Virtualization
- RHEL OpenStack Platform

* Exact rebase versions TBD
### RHEV Hypervisor

**Included in**
- Red Hat Enterprise Virtualization
- RHEL OpenStack Platform

### Beginning with Red Hat Enterprise Linux 7

- Two distinct streams for QEMU, libvirt and other virtualization components
- Rebasing as needed based on upstream features and upstream stability
- More rapid feature cadence
- Ability to update asynchronously if we need to align with RHEL-OSP and RHEV schedules

<table>
<thead>
<tr>
<th>QEMU-KVM-RHEV</th>
<th>RHEL 7</th>
<th>RHEL 7.1</th>
<th>RHEL 7.2</th>
<th>RHEL 7.3</th>
<th>RHEL 7.4</th>
<th>......</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QEMU 1.5</td>
<td>QEMU 2.1</td>
<td>QEMU 2.4</td>
<td>QEMU 2.7</td>
<td>QEMU 3.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
<td>Backported Bug fixes</td>
<td></td>
</tr>
</tbody>
</table>

---

#redhat #rhsummit