A TECHNICAL INTRODUCTION TO RED HAT ENTERPRISE LINUX 5.4

THE ENTERPRISE LINUX TEAM

PURPOSE

In September 2009, during the Red Hat Summit and JBoss World in Chicago, Red Hat delivered the fourth update in the Red Hat Enterprise Linux 5 family, Red Hat Enterprise Linux 5.4. In keeping with Red Hat's goals of providing a stable, scalable, high-performance operating system, Red Hat Enterprise Linux 5.4 brings together the innovations, patches, and security enhancements on which you can confidently run your applications. As with our previous updates, Red Hat delivers enhancements to the hardware ecosystem, core operating system, developer tools, and virtualization technology, while maintaining the Red Hat Enterprise Linux 5 ABI/API stability that is the hallmark of a Red Hat subscription. With the delivery of Red Hat Enterprise Linux 5.4, Red Hat provides the foundation for the Red Hat Enterprise Virtualization portfolio of solutions, including a standalone hypervisor as well as server and desktop management solutions.

This paper will highlight the new features within Red Hat Enterprise Linux 5.4 that improve system performance, simplify application development, and transform virtualization technology. Since the previous update to Red Hat Enterprise Linux, Red Hat, our partners, and our community have made advances in I/O performance and hardware, including virtualization standards, that are supported in this release. These include Single Root Virtualization (SR-IOV) on both AMD and Intel architectures, as well as enhanced FcoE and iSCSI devices and protocols. Developers and production administrators rely on the tools within Red Hat Enterprise Linux to understand application behavior and system performance. With this release, we expand the breadth and depth of the system analysis including tracepoints for SystemTap, enhanced remote management and new flexible file system features.

Red Hat Enterprise Linux 5.4 commercially releases the next generation of virtualization technology through incorporation of Kernel-based Virtual Machine (KVM) technology with the operating system, providing the next step in delivery of the Red Hat Enterprise Virtualization portfolio. KVM is integrated into the Linux kernel and takes advantage of all of the hardware and software support already available for Red Hat Enterprise Linux 5. Designed to leverage the latest advances in x86 virtualization support from Intel and AMD, this next-generation virtualization technology reduces the overall overhead of virtualization. Development continues on the Xen-based virtualization technology, which will be supported throughout the Red Hat Enterprise Linux 5 release cycle until at least 2014. Libvirt, the open standard API for managing virtual machines has been extended to manage both virtualization technologies. With the inclusion of Xen, KVM, and libvirt technologies, this release simplifies the support, deployment, and certification of virtual systems.

Before describing the new features included in Red Hat Enterprise Linux 5.4, it is important to keep in mind that the key values of any update Red Hat Enterprise Linux release are the hundreds of fixes that won't get detailed mention in this whitepaper. These fixes comprise the majority of the maintenance work that makes Red Hat Enterprise Linux the leading enterprise distribution. They address widely diverse bugs discovered during our ongoing proactive testing initiatives as well as through customer reports. Of equal importance is Red Hat's ongoing vigilance to plug all known security vulnerabilities. These software updates to Red Hat Enterprise Linux are delivered via errata advisories in a single stream and applied incrementally. Each errata is released individually on an as-needed basis or the errata are aggregated as a minor release. This Red Hat Enterprise Linux update release aggregates the full collection of features, bug fixes, and security hardening into a fully integrated and tested release distribution for ease of deployment. This work is the cornerstone of Red Hat Enterprise Linux. The remainder of this whitepaper focuses on the new features which debut in Red Hat Enterprise Linux 5.4.

CAPABILITIES

The feature details in this paper are organized into two functional topics. The feature functional areas include:

- Systems enablement – including performance enhancements and platform features, and
- Virtualization – new hypervisor support and and scalability enhancements.

Changes in both of these areas will be described in the upcoming sections. Please keep in mind that this paper contains only high-level summary information and is not comprehensive. In addition to the technology update, this update also includes new documentation. The Release Notes, which include all new features, has been expanded to include an additional document, Technical Notes, that addresses all changes in the this release. The Technical Notes document is designed to help customers understand all of the changes from new features to security and bug fixes that have occurred during this update.

SYSTEMS ENABLEMENT

Red Hat Enterprise Linux 5 update releases include new systems enablement, including: new hardware support, system tools, and general features.

HARDWARE ENABLEMENT

One of the core reasons for creating Red Hat Enterprise Linux update releases is to enable new hardware such as HBAs, NICs, and chipsets. In addition to the I/O device updates scheduled for all Red Hat Enterprise Linux 5 update releases, Red Hat Enterprise Linux 5.4 includes new support for SR-IOV along with Intel's VT-d and AMD's IOMMU.

- SR-IOV or Single Root I/O Virtualization is a PCI SIG specification that allows virtual functions (VF) to be created that share a physical device. In essence, it allows a single PCI device to be shared amongst multiple virtual machines and removes the need for direct PCI pass through if the devices support SR-IOV. This will reduce the need to fill the server with PCI devices dedicated to virtual machines.
- Virtual machine I/O performance can also be improved through the use of VT-d and IOMMU implementations for remapping of I/O DMA transfers. VT-d and IOMMU also provide memory management for devices and helps prevent a device mapped to a guest from being written on by other guests and hosts.

- FibreChannel over Ethernet (FCoE) is now supported over standard network interfaces. The FCoE support includes a SNIA standard library allowing third-party tools to effectively manage different FCoE interfaces.
- As 10GigE becomes more prevalent in datacenter designs, new drivers for support 10GigE NICs are included.
- Generic Receive Offload (GRO) aggregates packets before they are processed by the rest of the networking stack. This allows TCP performance to be greatly enhanced at high speeds, in particular, it is crucial for good 10GbE performance. With GRO enabled, you should be able to observe higher throughput and / or lower CPU utilization of network traffic, especially with smaller message sizes. The kernel in Red Hat Enterprise Linux 5.4 allows users to manually control whether generic receive offload is enabled for supported ethernet adapters. An updated ethtool provides a command-line interface for setting and querying that flag. A full list of drivers supporting GRO is available in the Release Notes.

SYSTEM TOOLS

Some new and updated tools have been included in Red Hat Enterprise Linux 5.4 for ease of administration, including:

- blktrace: a block layer IO tracing mechanism that provides detailed information about request queue operations in block layer I/O. This allows operators a detailed view into I/O transactions from processes to the block devices.
- I/O Accounting: Per process I/O accounting data is now available at /proc/<pid>/io. This provides real storage I/O statistics on a per process basis and is not impacted by I/O to cached data.
- SystemTap: The SystemTap implementation now includes support for tracepoints. Tracepoints are static markers in the code that support probing of pre-defined events without the need to install any debuginfo packages. Kernel tracepoints have been added in the networking, virtual memory and nfs subsystems as a Technology Preview[1]. Using kernel tracepoints with systemtap removes the need for a full kernel debuginfo to be present on the target system. A full list of available tracepoints can be found via:

  ```
  stap -L'kernel.trace("*")'
  ```

  SystemTap can now be used as a flight recorder with data being collected to a file on disk instead of in memory. This allows SystemTap scripts to run in the background and collect large amounts of data without the memory requirements. Options allow log size limits and log rotation.
- FIEMAP access: File Extent MAP ioctl is a new interface that returns logical to physical mapping for the extent that contains the specified logical byte address. FIEMAP support is now enabled on ext2, ext3, ext4, and GFS2, and allows applications efficient access to sparse files.
- Hugepage coredump: Hugepages can now be included in a coredump allowing developers more data to debug memory intensive applications.

1 Technology preview, tech preview. The majority of new features provided in Red Hat Enterprise Linux releases are fully supported, however it is also common for releases to include some new features which are not sufficiently mature to be ready for fully supported production deployment. To provide customers with early access to these promising new features, and to enable them to participate in validation testing, these features are designated as tech preview.

GENERAL FEATURES

A number of new features are also being introduced in this update to Red Hat Enterprise Linux 5, including:

- FUSE: Filesystem In User Space (FUSE) allows regular users to mount any FUSE-based filesystems. There is no need for each FUSE-based filesystem to be enabled in the kernel, allowing developer the flexibility to try new filesystems. The update includes both kernel infrastructure and user space utilities to manage FUSE filesystems and is fully supported. However, it does not include any particular FUSE filesystems. A number of FUSE filesystems are listed on Wikipedia.
- Clustered CIFS: An updated version of Samba 3.3 and a corresponding CTDB back end are available as a Technology Preview on Red Hat Enterprise Linux Advanced Platform on the x86_64 architecture. This update enables a new feature, which when configured with GFS2 and Cluster Suite, can provide a scalable active-active CIFS filesystem from a number of cluster nodes. This improves filesystem availability support for Microsoft Windows™ [2] operating systems.
- SAP Application Availability: Red Hat Enterprise Linux Advanced Platform includes updated resource agents for improved flexibility in managing high-availability for SAP applications and databases.
- Updated remote system management utilities, including: OpenHPI (Hardware Platform Interface) for new OEM platform support; OpenIPMI for IPMI support; updated tog-pegasus and new SBLIM components for enhanced CIM support to support standard management frameworks.
- Technology Preview of a new malloc() implementation. This implementation needs to be explicitly enabled via a environment variable. Once enabled, the new malloc() uses per thread memory pools for reduction in memory contention.
- Technology Preview of the next GCC compiler has been updated to GCC 4.4. Please see the Technical Notes for a complete description.

2 Windows is a registered trademark of Microsoft Corporation in the United States and other countries.

VIRTUALIZATION

Red Hat Enterprise Linux first incorporated virtualization technology with the delivery of Red Hat Enterprise Linux 5 in March 2007. After acquiring Qumranet and its technologies in September 2008, Red Hat announced plans to deliver the Red Hat Enterprise Virtualization portfolio, based upon KVM virtualization technology, including the incorporation of this next-generation virtualization in the fourth update to Red Hat Enterprise Linux 5. The Red Hat Enterprise Linux 5.4 update introduces full support for the KVM hypervisor, providing it alongside of the Xen hypervisor. KVM is a loadable module in the Red Hat Enterprise Linux kernel and benefits from all the features and the stability of the Red Hat Enterprise Linux kernel, such as process scheduling and NUMA support. KVM leverages the presence of virtualization-enabled CPU architectures like Intel VT and AMD-V to deliver remarkable performance, security, and scalability. KVM itself consists of two major components:

- A device driver managing virtualization hardware, and
- A user-space component for emulating PC hardware; this is a lightly modified qemu process.

Since a virtual machine is simply a process, all of the standard Linux process management tools apply. Permissions are handled by the normal Linux method, the virtual machine belongs to the user who started it. Future enhancements on the permission model will provide out-of-box SELinux policies for virtual machines and integration via libvirt. For more information on installing and configuring KVM, please consult the Virtualization Guide in the Red Hat Enterprise Linux documentation.

In addition to the introduction of a new hypervisor, Red Hat Enterprise Linux 5 adds new virtualization support features such as hugepages, and improved scaling for hypervisors:

- VMware: Improved timer clock synchronization for enhanced VMware guest operation. There is new hypervisor detection code that allows a Red Hat Enterprise Linux 5 guest to detect a VMware hypervisor and select appropriate clock interfaces to the hypervisor.
- Hugepages: Red Hat Enterprise Linux has traditionally supported a fixed hugepage size of 2MB to improve performance of applications that are memory intensive such as databases and virtual machines. This release enables 1 GB hugepages in addition to 2MB hugepages on the x86_64 platforms that have GB hugepage support. Hugepage enablement is required at boot time, but provides improved performance due to:
    1. reducing TLB misses, and
    2. reducing nested page table walk.
- Scaling: Red Hat continues to test the limits of Red Hat Enterprise Linux with commercially available hardware. While maintaining the virtual machine limits, the Red Hat Enterprise Linux 5 hypervisor has shown increased scalability from 126 cpus to 192 cpus[3]. For production use, please consult the Hardware Catalog for a current list of certified systems and limits.

3 System scalability limits describe the practical software support limits. Certification of hardware configurations is dependent on Red Hat's requirements of passing the certification test suite and availability of systems for our development and test teams to have direct access to for testing purposes.

With the introduction of the new hypervisor, libvirt, the core virtualization management interface, has been updated to provide a seamless interface between the two options. The API has been expanded to support new features, including:

- NPIV: N_Port ID Virtualization allows multiple N_Port IDs to share a single physical N_Port. New HBAs provide this capability and libvirt has been extended to support NPIV configuration. Using NPIV storage administrators should be able track storage allocation and usage by virtual machine.
- PCI Pass Through: Enable libvirt-defined hosts to manage PCI pass through configuration for direct access between virtual machines and named hardware.
- Serial port redirect: Enable the serial port to be redirected to a file or a TCP connection from a virtual machine. This can be used for traditional logging over serial port consoles.

CONCLUSION

Red Hat Enterprise Linux 5 continues to deliver on the reliable Red Hat subscription model by providing ABI and API stability, while continuing to add new hardware enablement and software features. The Red Hat Enterprise Linux 5.4 release marks the fourth cumulative enhancement and bug fix release in Red Hat Enterprise Linux 5, and provides the foundational piece for the Red Hat Enterprise Virtualization portfolio.

RED HAT SALES AND INQUIRIES

NORTH AMERICA
1-888-REDHAT1
www.redhat.com

EUROPE, MIDDLE EAST AND AFRICA
00800 7334 2835
www.europe.redhat.com
europe@redhat.com

ASIA PACIFIC
+65 6490 4200
www.apac.redhat.com
apac@redhat.com

LATIN AMERICA
+54 11 4341 6200
www.latam.redhat.com
info-latam@redhat.com

www.redhat.com
#1256889_0809