

RED HAT
ON THE MAINFRAME
THE REASONS ARE COMPELLING



Current & Future Linux on System z Technology Roadmap

Brad Hinson
Red Hat



Introductions

- Brad Hinson
 - Worldwide System z Sales, Strategy, Marketing
 - bhinson@redhat.com
 - +1 (919) 360-0443



Agenda

- Current technology through RHEL 5.5 relating to System z
- Roadmap moving forward (RHEL6)
 - New file systems
 - Kernel patches & improvements
 - Performance
- Open Dialog



RHEL 5

Current Technology

RHEL 5 Summary

- Documentation/Getting Started
 - Redbook, “z/VM and Linux on IBM System z: The Virtualization Cookbook for Red Hat Enterprise Linux 5.2”
 - <http://www.redbooks.ibm.com/abstracts/sg247492.html>
 - Release notes:
 - <http://www.redhat.com/docs/manuals/enterprise/>
 - Knowledgebase:
 - <http://kbase.redhat.com/>
 - Search “s390”
 - <http://www.redhat.com/z>

RHEL 5 Summary: Performance

- Large page support [5.2]
 - Support 1MB memory page size, changes to SIE instruction
- New package: ktune [5.3]
 - Service that sets several kernel tuning parameters to values suitable for specific system profiles.
 - Provides a profile for large-memory systems running disk-intensive and network-intensive applications.
- Include gcc 4.3 with latest z10 instruction set [5.3]
- Kernel sharing with NSS [5.4]
 - Boot from shared single copy of kernel stored in real memory

RHEL 5 Summary: Performance

- FCP performance data collection (blktrace) [5.4]
 - Track, analyze, and visualize system I/O
- CPU Node Affinity [5.3]
 - Use new interface provided by z10 exposing CPU node topology
 - Optimize Linux scheduling to increase cache hits
- Support for processor degradation [5.4]
 - Reduce processor speed in the event of cooling problems
- Dynamic CPU hotplug daemon [5.4]
 - Monitor system workload and dynamically set CPUs online/offline according to system pressure

RHEL 5 Summary: Networking

- SELinux per-packet access controls [5.3]
 - Add Secmark support to core networking
 - Allows security subsystems to place security markings on network packets (2.6.18)

- NETIUCV driver removed, deprecated [5.0]

- Samba: rebased from 3.0.28 to 3.0.32 for bugfixes
 - Now supports Windows Vista and 2008
 - Fixes for DC functionality (interoperability with Citrix and Domain trusts)

RHEL 5 Summary: Security

- 3592 tape encryption support [5.1]
 - Use IBM 3592 encryption feature to securely store data
- Crypto hardware enablement [5.2]
 - Device driver support needed for the z10 PCI crypto accelerator.
- DS8000 disk encryption [5.4]
 - Tool to display disk encryption status from Linux
- Crypto device driver use of thin interrupts [5.4]
 - Performance update: Use interrupts instead of polling
 - Eliminates wasted CPU cycles
- In-kernel exploitation of new CP assist functions
 - AES 192 / 256 and SHA 384 / 512

RHEL 5 Summary: Security

- pkinit clients can now be configured to use keys for client certificates which may not contain Kerberos-specific extensions & interoperability fixes
- nss_ldap now configured with support for paged results extension
- SELinux: enablement of New NetworkManager and Audit functionality.
- SELinux: Hundreds of AVC denial fixes.
- Improved Audit and Logging
 - TTY input audit support
 - Remote audit logging via unencrypted connection

RHEL 5 Summary: File Systems/Storage

- Block device encryption support, including support for /root partition, including configuration in anaconda installer.
- ext4 file system tech preview [5.3, with updates in 5.4]
 - Support for delayed allocation
- Ecryptfs support [5.2, with updates in 5.3]
- Improved multipath over PAV support [5.2]

RHEL 5 Summary: File Systems/Storage

- Add integrity check to cryptsetup-luks, in order to meet FIPS-140 requirements [5.4]
- File system freeze/quiesce interface added to support hardware snapshots for file systems.
- Full support for FUSE and libfuse
 - Allow end users to more easily install and use their own user space FUSE file systems.

RHEL 5 Summary: Reliability

- xDR/GDPS system initialization [5.3]
 - Disaster recovery solution for DASD and FCP
 - Multipath access to PPRC (primary/mirror) volume pair
 - Hyperswap support in DASD driver

- DASD fast fail support
 - Increase failover speed in multipath setup with DASD and PAV setup

- Multipath support in installer [5.2]



RHEL 5.4

GA: September, 2009

RHEL 5.4 : Overview

- The official GA release of Red Hat Enterprise Linux 5.4 (*kernel-2.6.18-164.el5*) was released on 02-Sep-2009
- Generic (not s390 specific) updates include
 - **Virtualization:** full support for the Kernel-based Virtual Machine (KVM) hypervisor only on x86_64 while Xen only x86 & Power based virtualization is still available & supported
 - **Network:** Kernel & Userspace update to support Generic Receive Offload (GRO) which increases the performance of inbound network connections by reducing the amount of processing done by the CPU. Furthermore Netfilter Framework & Bind Updates
 - **Storage:** Support for the XFS file system has also been added to the kernel as a Technology Preview.
 - **Tools:** SystemTap is now fully supported, and has been re-based to the latest upstream version.

RHEL 5.4: System Services

- Rebased version of CUPS print server, now fully Kerberized
- Package upstream rebases to the following utilities:
 - ksh, lm-sensors, lftp, net-snmp, openIPMI-tool, openldap, openmotif, python-urlgrabber, openPegasus, VNC
 - RPM to Fedora 9 version, which includes numerous bugfixes
 - yum and yum-utils primarily for speed improvements
 - totem, rb, and gstreamer rebased to enable modular codecs addition

s390-tools package rebased to version 1.8.1

- The s390utils package has been rebased to version 1.8.1.
- This package provides **the** essential tool chain for Linux on System z. It contains everything from the boot loader to dump related tools for a system crash analysis .
- News Features (excerpt)
 - DASD related tools: Add Large Volume Support for ECKD DASDs
 - IPL tools: Can be used to change the reipl & shutdown behaviour
 - ziomon tools: Set of tools to collect data for zfcpl performance analysis.
 - Isluns: List available SCSI LUNs depending on adapter or port.
 - Iszcrypt/chzcrypt: Show/modify information about zcrypt devices and configuration.
 - cpuplugd: Daemon that manages CPU- and memory-resources based on a set of rules. Depending on the workload CPUs can be enabled or disabled. The amount of memory can be increased or decreased exploiting the Cooperative Memory Management (CMM1) feature.
 - Ischp/chchp: Tool to show/modify channel-path states and available channel paths.
 - mon_procd: Daemon that writes process information data to the z/VM monitor stream.
 - vmur: Tool to work with z/VM spool file queues (reader, punch, printer).
 - zfcpldump_v2: Version 2 of the zfcpldump tool. Now based on the upstream Linux kernel 2.6.23.
- Plus various bug fixes

Kernel

■ **Control Program Identification (CPI)**

- If your RHEL5.4 Linux instance runs in LPAR mode, you can now use the extended control program identification (CPI) module, `sclp_cpi` and the sysfs interface `/sys/firmware/cpi` to assign names to your Linux instance
- The names are used, for example, to identify the Linux instance on the HMC.
- *This feature is only available while running in LPAR*

■ **Extra kernel parameter via VMPARM**

- Modify the IPL records to append extra parameters specified with the z/VM VMPARM option to the kernel command line.

■ **Support for processor degradation**

- Adds support for processor degradation, which allows processor speed to be reduced in some circumstances (i.e. system overheating). This new feature allows automation software to observe the machine state.

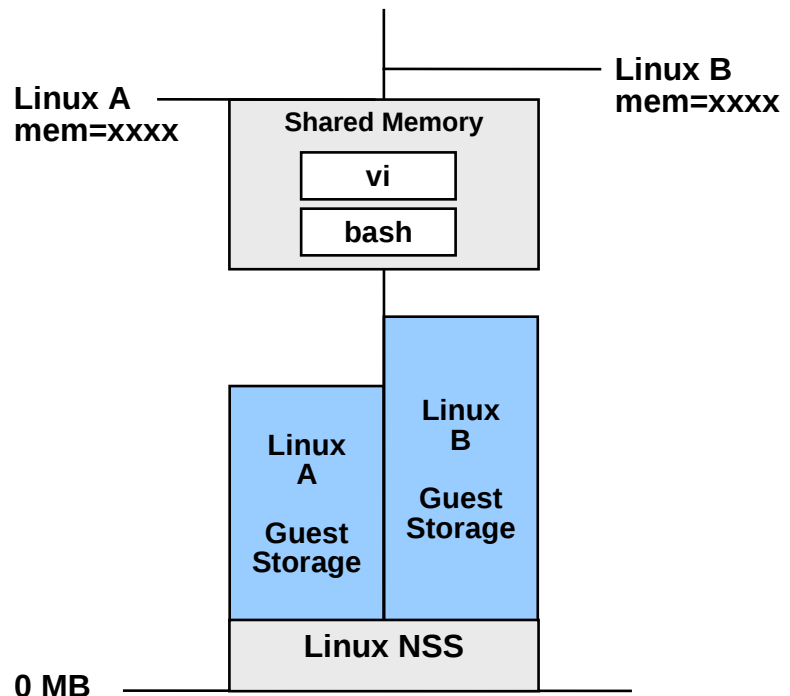
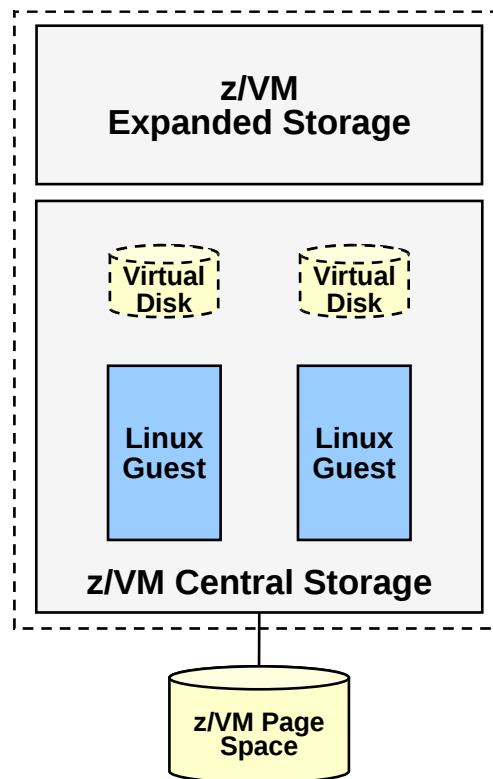
■ **TTY terminal server over IUCV**

- Provide central access to the Linux console for the different guests of a z/VM.
- The terminal server connects to the different guests over IUCV.
- The IUCV based console is ASCII based.
- Fullscreen applications like `vi` are usable on the console.

Virtual Server

Named Saved Segments (NSS)

- Using NSS the z/VM hypervisor makes operating system code in shared real memory pages available to z/VM guest virtual machines.
- With this update, Linux guest operating systems using z/VM can boot from the NSS and be run from a single copy of the Linux kernel in memory.



Networking

■ **HiperSockets Layer3 Support for IPv6**

- How IPv6 support for HiperSockets devices running in layer 3 mode is available
- IPv6 is supported on:
 - Ethernet interfaces of the OSA-Express adapter running in QDIO mode.
 - HiperSockets layer 2 and layer 3 interfaces
 - z/VM guest LAN interfaces running in QDIO mode.
- IPv6 is not supported on the OSA-Express Token Ring and ATM features.

RAS

■ Multi volume dump support for DASDs

- Added the ability to dump on multiple ECKD DASD devices, which can be necessary, if the system memory size is larger than the size of a single DASD device.

■ Service Levels of Hardware & Hypervisor

- A new Interface which provides service levels of hardware and z/VM service-levels to the Linux userspace. Interface: */proc/service_levels*

■ Lstape support for SCSI Tapes

- With this feature it is now possible to list installed FCP-attached tape devices (SCSI tapes) besides channel attached tapes using the *lstape* command

■ Shutdown Actions Interface

- The new shutdown actions interface allows to specify for each shutdown trigger (halt, power off, reboot, panic) one of the five available shutdown actions (stop, ipl, reipl, dump, vmcmd).
- A sysfs interface under */sys/firmware* is provided for that purpose.
- Possible use cases are e.g. to specify that a vmdump should be automatically triggered in case of a kernel panic or the z/VM logoff command should be executed on halt.

■ Automatic IPL after dump

- The new shutdown action *dump_reipl* is introduced. It combines the actions *dump* and *re-ipl*, first a dump is taken, then a re-ipl of the system is triggered

Storage

■ **FCP performance data collection & reports:**

- Fibre Channel Protocol (FCP) performance data can now be measured. Metrics that are collected and reported on include:
- Performance relevant data on stack components such as Linux devices, Small Computer System Interface (SCSI) Logical Unit Numbers (LUNs) and Host Bus Adapter (HBA) storage controller information.
- Per stack component: current values of relevant measurements such as throughput, utilization and other applicable measurements.
- Statistical aggregations (minimum, maximum, averages and histogram) of data associated with I/O requests including size, latency per component and totals.

■ **DS8K Encryption Support**

- This feature enhances s390-tools to be able to display if the Storage has its disk encrypted or not.

■ **Kernel support to issue Control I/O to dasd on EMC Symmetrix arrays**

- Support has been added to the kernel to issue EMC Symmetrix Control I/O. This update provides the ability to manage EMC Symmetrix storage arrays.



RHEL 5.5

GA: March, 2010

RHEL 5.5: New Features: Kernel

- **Feature: DS8000 Large Volume Support (kernel 2.6.30)**
- **Summary:** Implement support for DS8k volumes, which have more than 64k cylinders. This allows DASD/ECKD larger than 50GB, reducing the need for volume management of small disks.

- **AF_IUCV SOCK_SEQPACKET Support (kernel 2.6.31)**
- **Summary:** Provides a sequenced, reliable, two-way connection-based data transmission path for datagrams of fixed maximum length.

- **Feature: Digitally sign s390x kernel**
- **Summary:** Digitally sign all modules in the kernel, so that authenticity can be checked immediately by the “tainted” flag.

RHEL 5.5: New Features: Installer

- **Feature: Provide CMS Script for Initial IPL under z/VM**
- **Summary:** In order to begin the install process, the user must create a REXX EXEC to IPL the kernel, initial RAMdisk, and PARM file. A sample EXEC file is now provided on the installation media, so this no longer needs to be manually created with CMS XEDIT.

- **Feature: Enhanced ReIPL Support from Alternate Location**
- **Summary:** Take advantage of the ReIPL capabilities in the kernel, enabling the installer to boot into a non-default location. Works in all cases, including:
 - LPAR on machine older than z9
 - LPAR on z9 or newer (required to reIPL from FCP)
 - z/VM 5.2 or older
 - z/VM 5.4 or later (required to reIPL from FCP)

RHEL 5.5: New Features: Software

- **Update Java to version 1.5 to SR10**
 - Critical Java updates

- **SBLIM updates**
 - Standards Based Linux Instrumentation for Manageability
 - Add SFCB/SFCC to allow for common CIM/SEBM infrastructure across various Linux guests
 - Update Java CIM Client to version 1.3.9.1
 - Update versions of packages from sblim.org (27 updated packages total)

- **Rebase Systemtap to version 1.1**
 - Instrumentation system for the 2.6 kernel used to collect data on the operation of the Linux guest

RHEL 5.5: Bug Fixes

- **[Kernel] Honor vmhalt, vmpanic, vmpoff, vmreboot parameters**
 - Bugzilla: https://bugzilla.redhat.com/show_bug.cgi?id=518229
- **[Kernel] Panic found under specific installer/network configurations**
 - Bugzilla: https://bugzilla.redhat.com/show_bug.cgi?id=548653
- **[Kernel] Optimize storage key operations for anonymous pages**
 - Bugzilla: https://bugzilla.redhat.com/show_bug.cgi?id=519977
 - Fix poor performance when anonymous mappings are removed, by removing the check for anonymous pages that aren't swap cache backed.

RHEL 5.5: Bug Fixes

- **[OpenCryptoki] CCA Token not being recognized properly**
 - Bugzilla: https://bugzilla.redhat.com/show_bug.cgi?id=541028
- **[Qeth] Handle error codes from z/VM VSWITCH port isolation (kernel 2.6.31)**
 - Bugzilla: https://bugzilla.redhat.com/show_bug.cgi?id=503232
 - Usually z/VM guests are allowed to change the MAC address of a VSWITCH device. This is no longer true, if z/VM Virtual Switch Port Isolation is active (see z/VM Apars VM64281 and VM64463).
- **[Qeth] Improve checksum handling for Layer 3 traffic**
 - Bugzilla: https://bugzilla.redhat.com/show_bug.cgi?id=503238

RHEL 5.5: Bug Fixes

■ [CIO] Number of internal I/O retries

- Bugzilla: https://bugzilla.redhat.com/show_bug.cgi?id=519814
- Increase number of retries before device is considered nonfunctional

■ [CIO] Online/Offline processing

- Bugzilla: https://bugzilla.redhat.com/show_bug.cgi?id=523323
- Unit testing for setting DASD online/offline to increase reliability during disaster recovery situations.

■ [DASD] Failfast flag cannot be set on

- Bugzilla: https://bugzilla.redhat.com/show_bug.cgi?id=503222
- Improved device I/O recovery when the connection between host and storage server is lost.

RHEL 5.5: Bug Fixes

- **[Installer] Add “mpath” to default generic.prm for SCSI multipath installation**
 - Bugzilla: https://bugzilla.redhat.com/show_bug.cgi?id=538129
- **[Installer] Prompts for network information regardless of contents of PARM/CONF**
 - Bugzilla: https://bugzilla.redhat.com/show_bug.cgi?id=506898
 - Occurred on RHEL 5.4 with 1 virtual CPU and/or < 2G virtual STOR
- **[Installer] Problem reaching default gateway when IP is changed**
 - Bugzilla: https://bugzilla.redhat.com/show_bug.cgi?id=506742
 - Affected first-time installations, OSA driver timing issue
- **[Installer] Correctly handle partitioning from previous installation**
 - Bugzilla: https://bugzilla.redhat.com/show_bug.cgi?id=563266



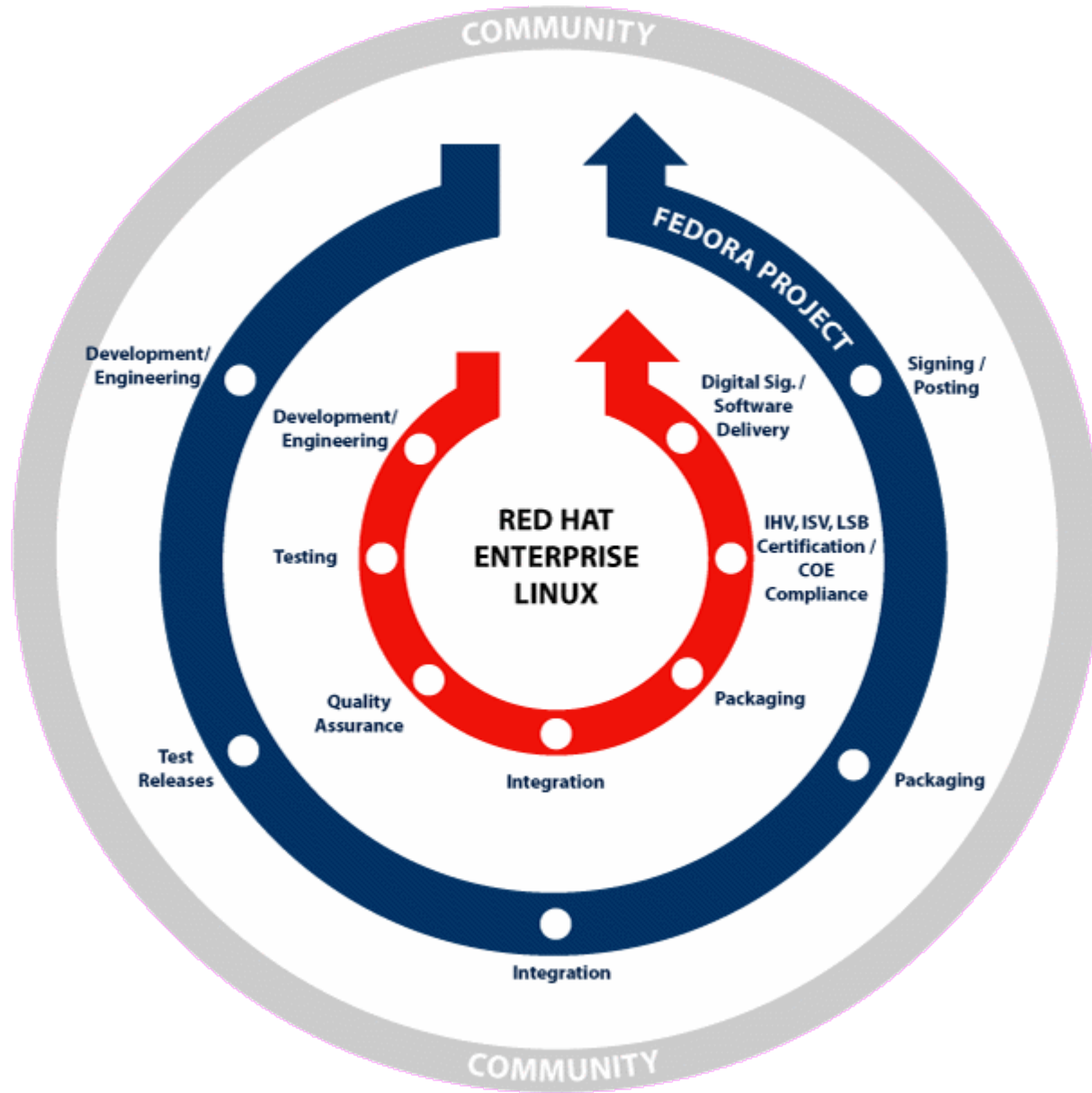
Future Linux on System z Technology

RHEL 5.6: Features

- (coming soon)



Development Model



Advanced Virtualization

- **Dynamic Memory Add/Remove** (kernel 2.6.27)
 - Enable to attach and use standby memory that is configured for a logical partition or z/VM guest.
 - Memory Attach & Detach requires running Linux on System z as a VM-guest requires z/VM 5.4 plus the PTF for APAR VM64524.

- **Standby CPU activation/deactivation** (kernel 2.6.25)
 - Allow standby CPUs to be activated / deactivated

- **Suspend / Resume** (kernel 2.6.31)
 - With suspend and resume support, you can stop a running Linux on System z instance and later continue operations.
 - When Linux is suspended, data is written to a swap partition. The resume process uses this data to make Linux continue from where it left off when it was suspended.
 - A suspended Linux instance does not require memory or processor cycles.

Storage Support

- **HyperPav** (kernel 2.6.25)
 - HyperPav is addressing the need to access more data with good performance and high availability!
 - This feature, which required a IBM DS8000™ disk storage system in average leads to a higher utilization, resulting in I/O transfer rates.
 - Activated automatically when the necessary prerequisites are there (DS8000 with HyperPAV LIC, z/VM 5.3). Transparent for the Linux on System z guest
- **DASD Large Volume Support** (> kernel 2.6.29)
 - Large Volume Support is a feature that allows to use ECKD devices with more than 65520 cylinders. This features is available with DS8000 R4.0
- **High Performance FICON (HPF)** (kernel 2.6.29)
 - Added HPF support to the DASD Device Driver
 - HPF is an extension to the FICON architecture and is designed to improve the execution of small block I/O requests.
 - HPF streamlines the FICON architecture and reduces the overhead on the channel processors, control unit ports, switch ports, and links by improving the way channel programs are written and processed.

Usability & Serviceability

▪ **Automatic IPL After Dump** (kernel 2.6.30)

- Extension to the shutdown action interface which combines the actions dump and re-ipl, first a dump is taken, then a re-ipl of the system is triggered

▪ **Compiler Improvements** (gcc 4.3/4.4)

- The latest compiler enhancements allow a customer to recompile existing applications which can be optimized for the latest hardware generation without any changes to the source code.
- This can lead up to a > 10 % performance improvement.

▪ **Large Page Support** (kernel 2.6.25)

- Support for a new access method to allocate larger chunks of memory, resulting in performance improvements, especially in Java based environments
- This feature exploits z10 hardware features and provides a software emulation for older systems.

Miscellaneous

- **STP/ETR Support** (kernel 2.6.27)
 - Support for clock synchronization using the server time protocol (STP) or an external time reference (ETR).
- **Kernel vdso support** (kernel 2.6.29)
 - Kernel provided shared library to speed up a few system calls (gettimeofday, clock_gettime, clock_getres)

Filesystems

- Ext4 – main, default filesystem. Incremental scalability (ie 100TB files) & performance enhancements over ext3.
- NFS
 - NFS4.0 – as default
 - NFS4.1 – referrals, delegation & failover.
 - Secure NFS, selinux labels over the wire – enabler for secure virt – timeframe likely post-GA update
- Fuse kernel portion as enabler.
- GFS2
 - Targeting HA clusters of up to 16 nodes.
 - Primary incremental focus is stabilization and performance
 - Possible improvements in ability to shrink/defragmentation
- BTRFS – next generation, enhanced data integrity, ease of use, & scalability. Likely tech preview at GA, non root/boot.

Kernel Scalability

- Scalability limits – Maximum values
 - Max CPUs – dynamic allocation of CPU structs (2.6.29) allows supported limit of 4096 and theoretical limit of 64K.
 - Max IRQs – dynamic allocation (2.6.28) allows limit of 256
 - Max memory 48-bit addressing (256TB) requires pending patch incorporation
 - Max # processes – 4 million on 64 bit kernels
 - Max threads per process remains at 32000 (same as RHEL5)
 - Filesystem limits for ext4 – 100T (practical target – bounded by fsck time)

- Scalability features
 - Split LRU VM – different eviction policies for file backed vs swap backed

IPA Client (Identity, Policy, Audit)

- IPA client in core RHEL for centralized security management
 - Kerberos authentication with host based access control
 - Provides central storage of extended user attributes
 - Enables centralized policy for applications, including SELinux policy
 - Audit log aggregation services & search capabilities

Security enhancements

- Virtualization isolation in conjunction with SELinux
 - May be post-GA update
 - Labeled NFS for filesystem isolation
 - Cut-and-paste window controls per authorization level
 - Guest confinement via SELinux policy enhancements
- NSS crypto
 - Broaden the core services which utilize NSS crypto libraries
 - Allows cheaper implementation of new features, ie TPM & centralized key store
 - Incremental targeted conversion of: Openswan, openldap, glibc
 - Add new crypto GUI for key import & establishment of trust
- Volume encryption
 - Basic operation already present in RHEL5 – incremental centralized key management for RHEL6
- Sectool
 - Compliance checking / intrusion detection utility – validates system admin config: file permissions, valid UIDs, reasonable passwords, etc



Q & A