Hans-Joachim Picht, Linux on System z Initiatives <hans@de.ibm.com>

Linux on System z Update: Current & Future Linux on System z Technology

GSE Frühjahrstagung 2010 für z/VSE, z/VM und Linux, Würzburg
Dienstag, 20. April 2010

Trademarks & Disclaimer

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries. For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml:

IBM, the IBM logo, BladeCenter, Calibrated Vectored Cooling, ClusterProven, Cool Blue, POWER, PowerExecutive, Predictive Failure Analysis, ServerProven, System p, System Storage, System x , System z, WebSphere, DB2 and Tivoli are trademarks of IBM Corporation in the United States and/or other countries. For a list of additional IBM trademarks, please see http://ibm.com/legal/copytrade.shtml.

The following are trademarks or registered trademarks of other companies: Java and all Java based trademarks and logos are trademarks of Sun Microsystems, Inc., in the United States and other countries or both Microsoft, Windows,Windows NT and the Windows logo are registered trademarks of Microsoft Corporation in the United States, other countries, or both. Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
UNIX is a registered trademark of The Open Group in the United States and other countries or both. Linux is a trademark of Linus Torvalds in the United States, other countries, or both. Cell Broadband Engine is a trademark of Sony Computer Entertainment Inc. InfiniBand is a trademark of the InfiniBand Trade Association.

Other company, product, or service names may be trademarks or service marks of others.

NOTES: Linux penguin image courtesy of Larry Ewing (lewing@isc.tamu.edu) and The GIMP

Any performance data contained in this document was determined in a controlled environment. Actual results may vary significantly and are dependent on many factors including system hardware configuration and software design and configuration. Some measurements quoted in this document may have been made on development-level systems. There is no guarantee these measurements will be the same on generally-available systems. Users of this document should verify the applicable data for their specific environment. IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.
Information is provided “AS IS” without warranty of any kind. All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices are suggested US list prices and are subject to change without notice. Starting price may not include a hard drive, operating system or other features. Contact your IBM representative or Business Partner for the most current pricing in your geography. Any proposed use of claims in this presentation outside of the United States must be reviewed by local IBM country counsel prior to such use. The information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any

IBM collaborates with the Linux community

- ...has been an active participant since 1999
- ...is one of the leading commercial contributors to Linux
- ...has over 600 full-time developers working with Linux and open source

Linux Kernel & Subsystem Development

Kernel Base

Security

Systems Mgmt

Virtualization

Filesystems,

and more...

Expanding the Open Source Ecosystem

Apache

Eclipse

Mozilla Firefox

OpenOffice.org,

and more...

Promoting Open Standards & Community Collaboration

The Linux Foundation

Linux Standards Base

Common Criteria certification,

and more...

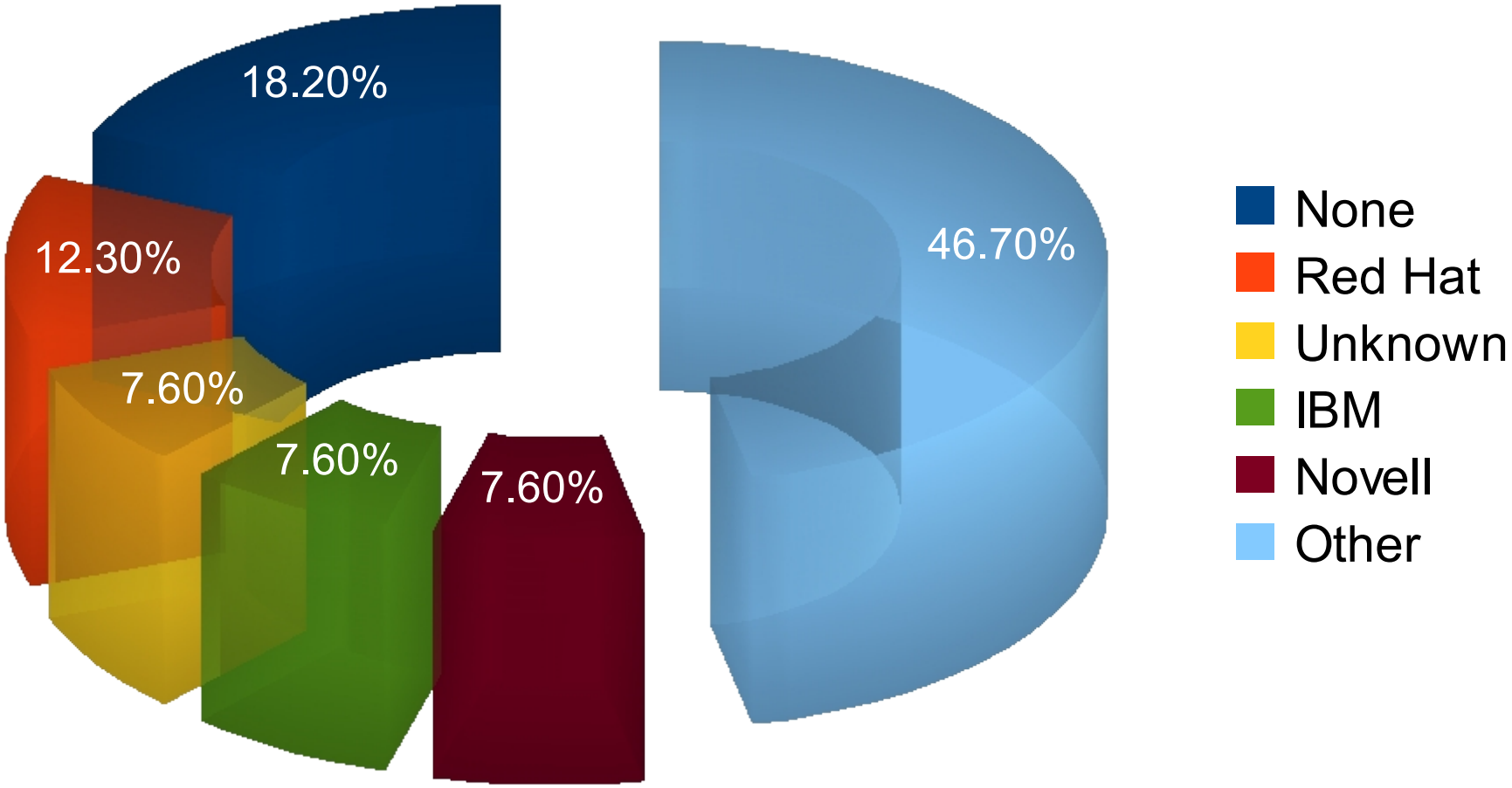
Foster and Protect the Ecosystem

Software Freedom Law Center

Free Software Foundation (FSF),

and more...

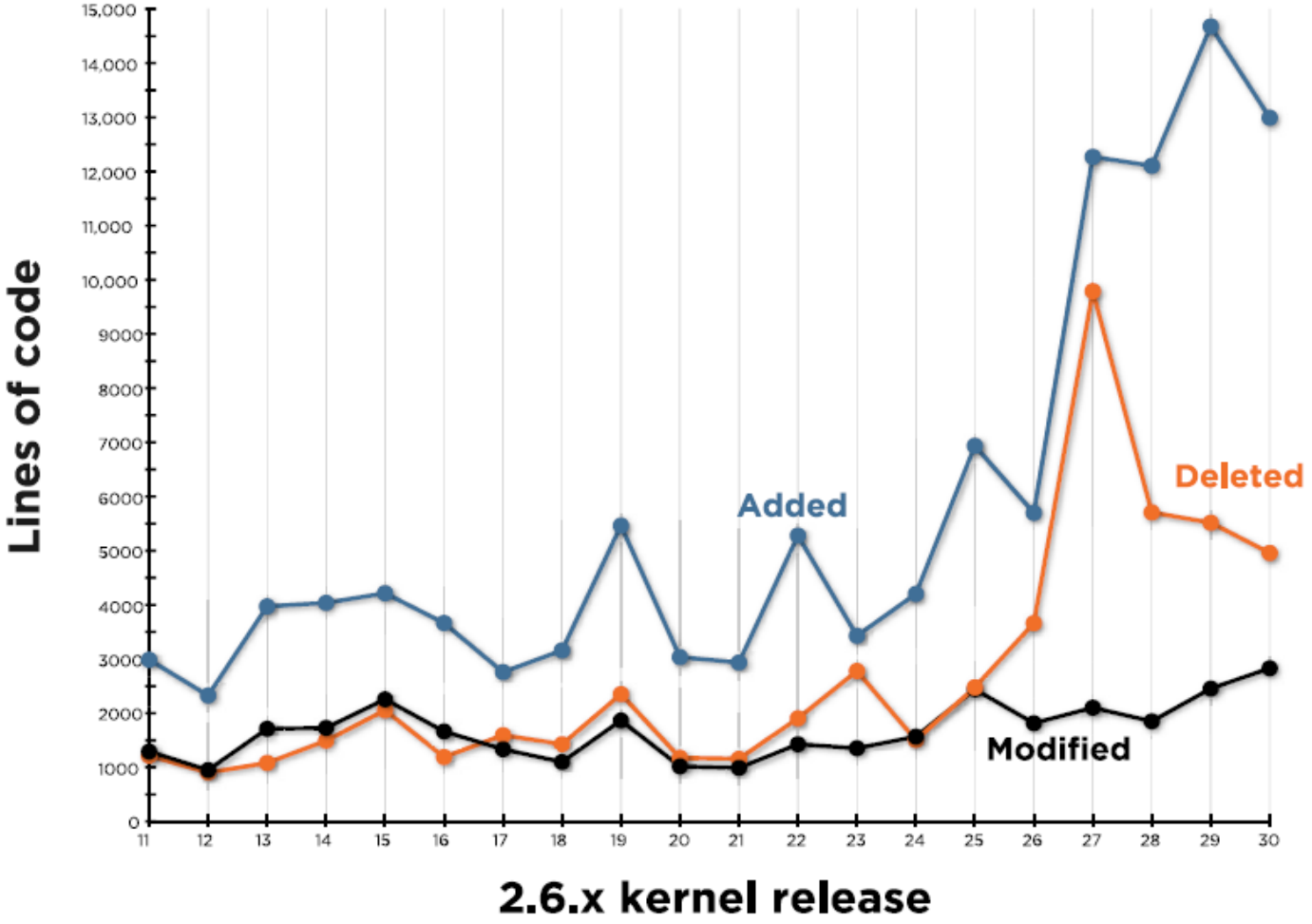
Who contributes to the Linux Kernel: Top 5 contributors



Source: Linux Foundation

Linux kernel development: rate of change

Average: 6,422 lines added, 3,285 lines removed, and 1,687 lines changed every day for the past 4 1/2 years.



Facts on Linux

- Last year, **75%** of the Linux code was developed by **programmers working for corporations**.
- **$7.37 billion:** projected cost to produce the 283 million lines of code which are contained in Linux Distribution **in a commercial environment.**
- IDC forecasts show that **Linux server revenue will grow by 85.5%** between 2008 and 2012 **in the non-x86 server space** equalling a four year compound annual growth rate of 16.7%.
- **Linux is Linux**, but …features, properties and quality differ dependent on your platform

Source: Intelligence Slideshow: 40 Fast Facts on Linux http://www.baselinemag.com/c/a/Intelligence/40-Fast-Facts-on-Linux-727574/
http://www.internetnews.com/dev-news/article.php/3659961
http://public.dhe.ibm.com/software/au/downloads/IBM_zLinux_DAG_FINAL.pdf

Kernel news – Common code

Linux version 2.6.29 (2009-03-23)
- Btrfs and squashfs filesystems
- Security module hooks for path based access control (AppArmor, Tomoyo)
- Credential records

Linux version 2.6.30 (2009-06-09)
- Reliable Datagram Sockets (RDS) protocol support
- EXOFS, a filesystem for Object-Based Storage Devices
- FS-Cache, a caching filesystem
- Filesystems performance improvements

Linux version 2.6.31 (2009-08-03)
- Performance counters
- Ftrace function tracer extensions
- Per partition blktrace

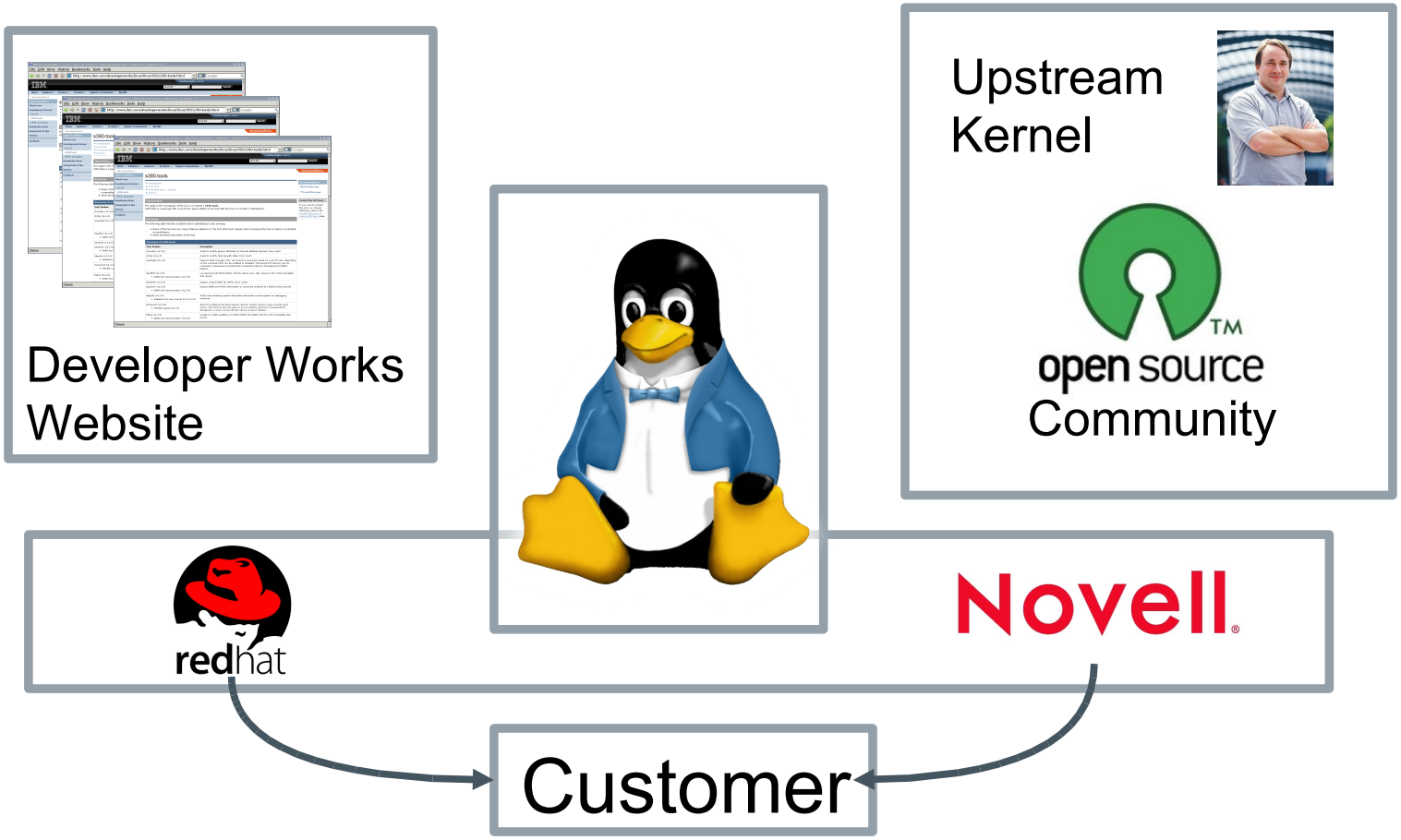
Linux version 2.6.32 (2009-08-03)
- Per-backing-device based writeback (pdflush replaced by flush <major>)
- Kernel Samepage Merging (memory deduplication)
- CFQ I/O scheduler low latency mode
- S+core architecture support

Linux version 2.6.33 (2010-02-24)
- DRDB (Distributed Replicated Block Device)
- TCP Cookie Transactions for DNSSEC protocol
- Swappable KSM pages
- Compcache: memory compressed swapping

Linux version 2.6.34 (TBD)
- Ceph: Distributed, replicated network file system for cluster

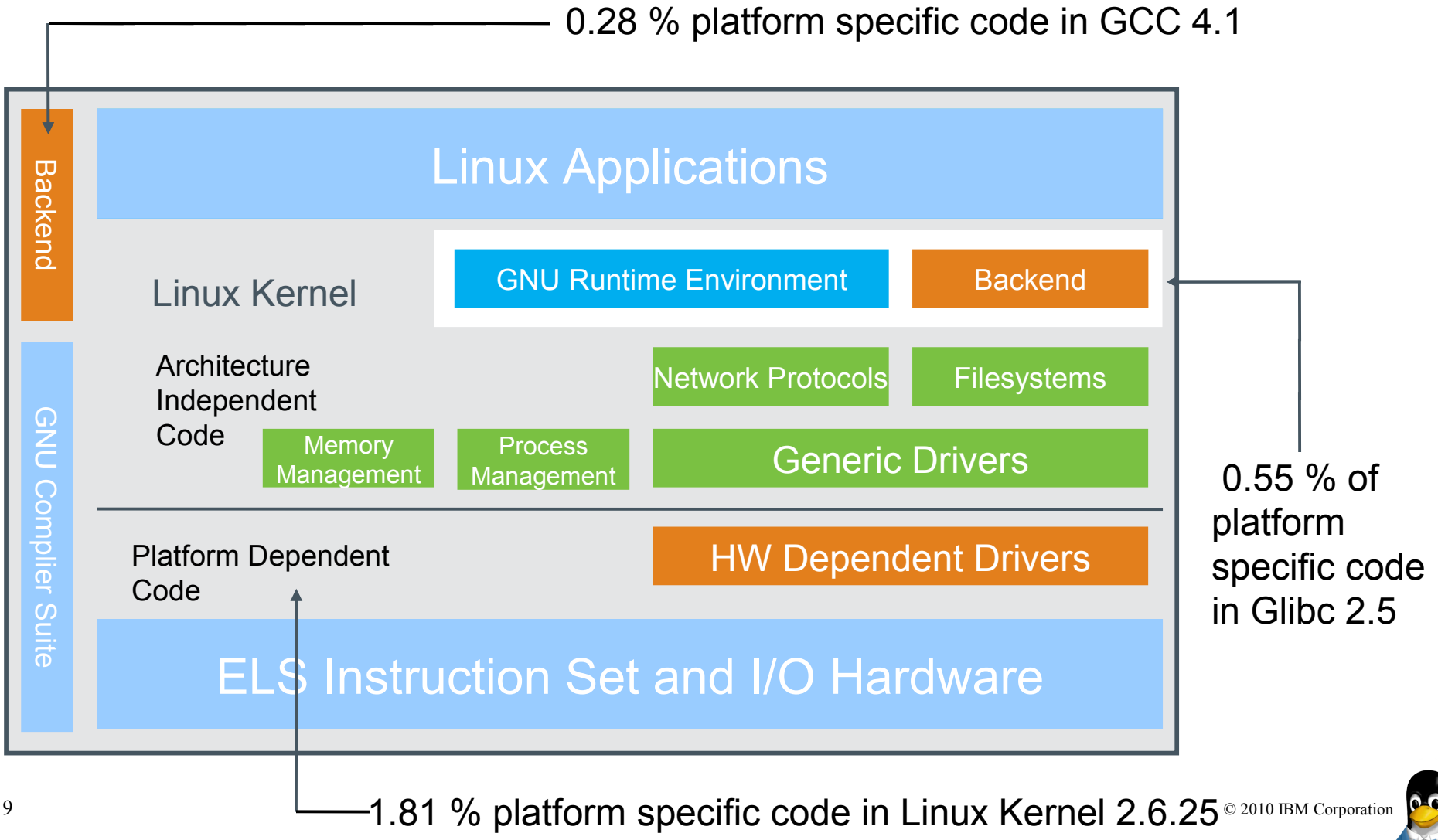
IBM Linux on System z Development

IBM Linux on System z Development contributes in the following areas: Kernel, s390-tools, Open Source Tools (e.g. eclipse, ooprofile), GCC, GLIBC, Binutils

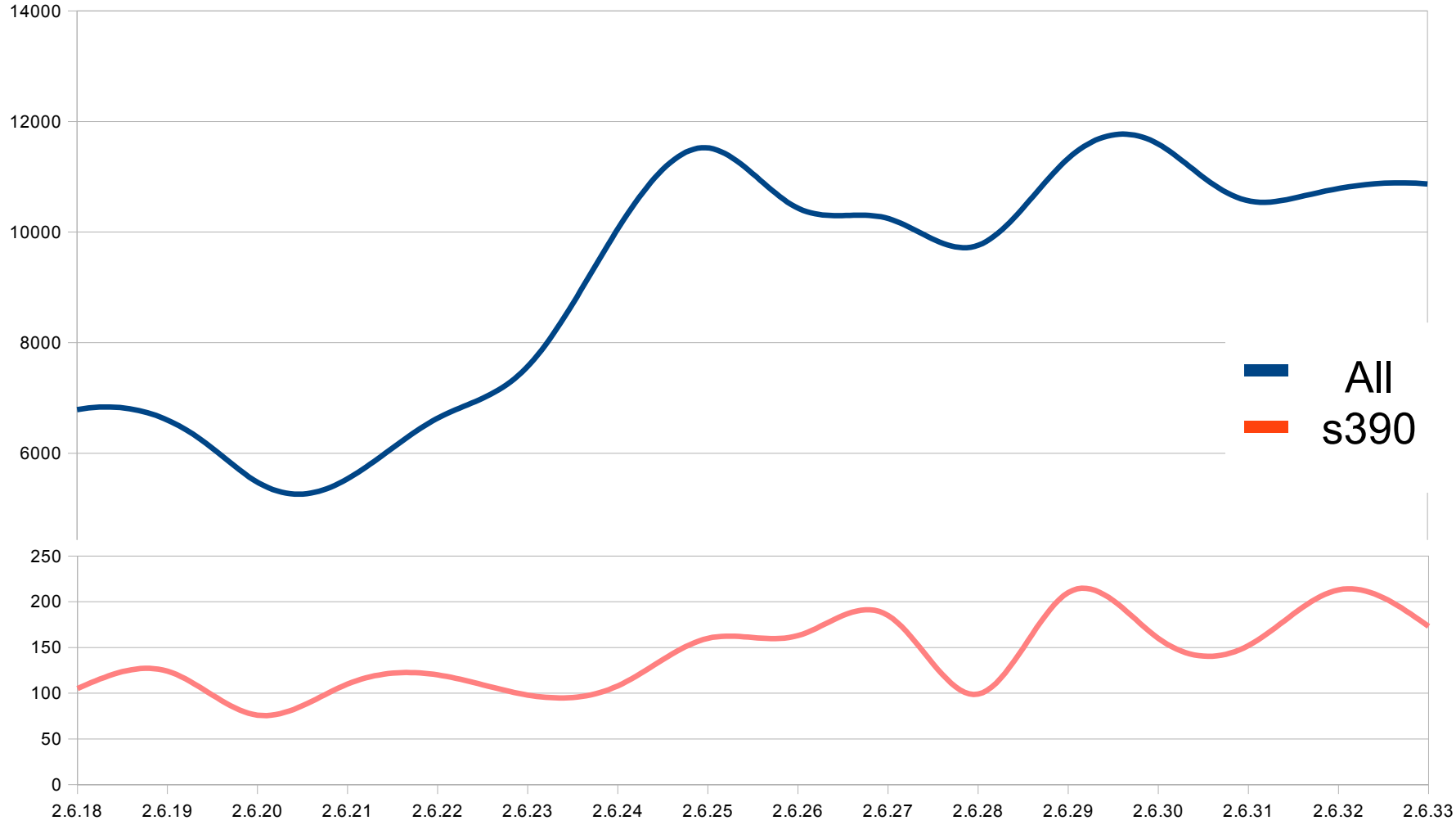


Structure of Linux on System z

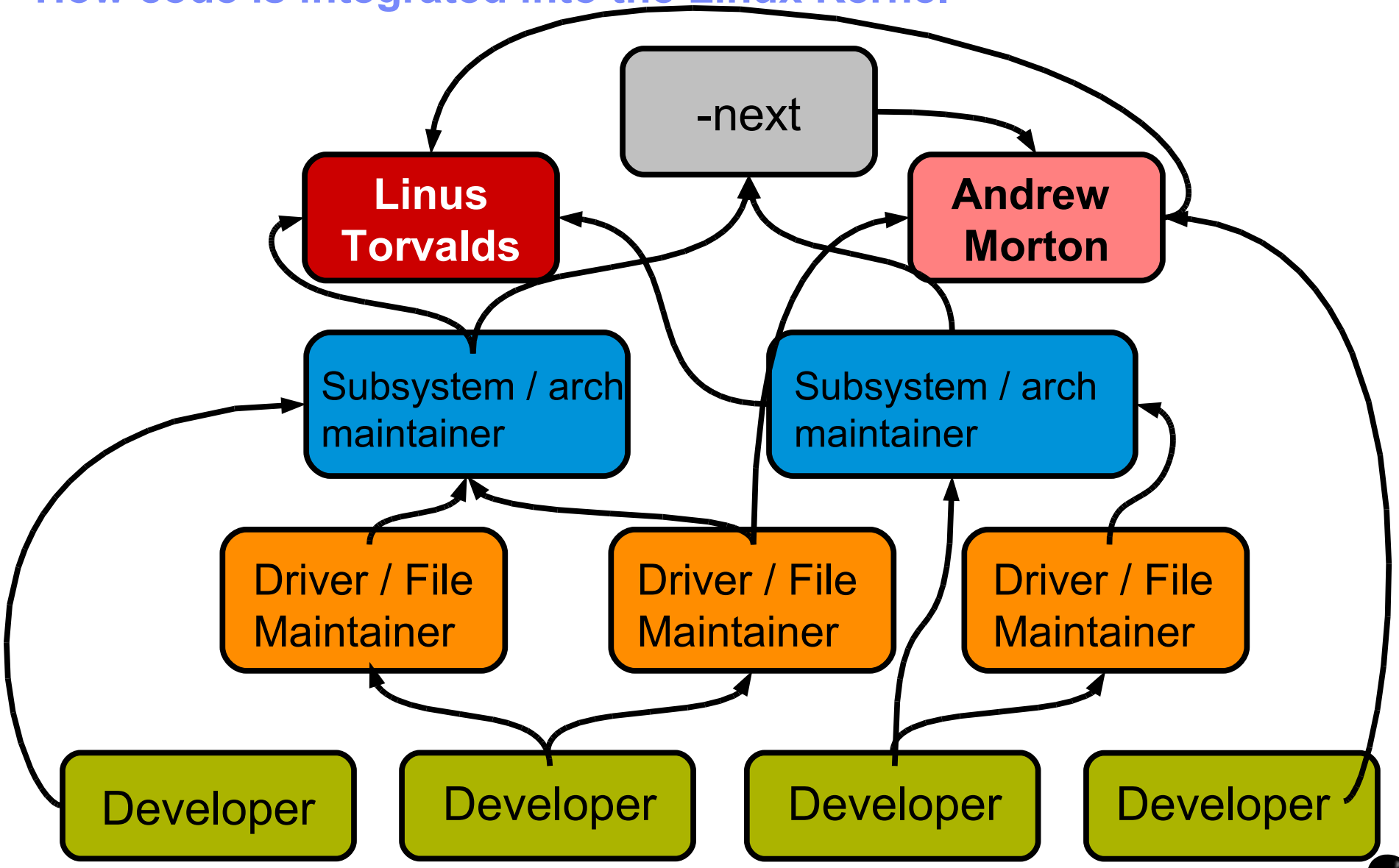
Many Linux software packages did not require any code change to run on Linux on System z



Linux kernel development: System z contributions



How code is integrated into the Linux Kernel



IBM Supported Linux Distributions for System z

Hardware platform and operating system software compatibility

64-bit environment

Distribution	System z10	System z9	zSeries
RHEL 5	✓	✓	✓
RHEL 4	✓	✓	✓
RHEL 3	–	*	✓
SLES 11	✓	✓	✗
SLES 10	✓	✓	✓
SLES 9	✓	✓	✓

31-bit environment

Distribution	System z10	System z9	zSeries
RHEL 5 (1)	–	–	–
RHEL 4	✓	✓	✓
RHEL 3	–	*	✓
SLES 11 (1)	–	–	–
SLES 10 (1)	–	–	–
SLES 9	✓	✓	✓

(1) A 64-bit distribution does not run in a 31-bit environment; note that 31-bit applications can be run on a 64-bit distribution using the 31-bit emulation layer.

✓ Indicates that the distribution (version) has been tested by IBM in the environment, will run on the system, and is an IBM supported environment. Updates or service packs applied to the distribution are also supported. New distributions are not supported unless they are listed here.

✗ Indicates that the distribution is not supported by IBM.

– Indicates that the distribution has not been tested by IBM.

* Provided on customer request for existing zSeries workloads only. No System z9 feature exploitation.

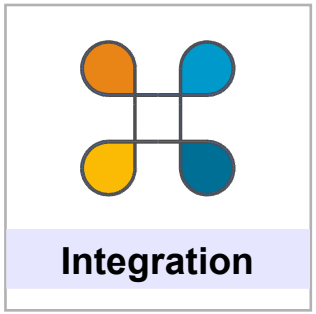
For information on which hardware is supported by

- Novell SUSE, please visit the "YES Certfied Bulletin Search" information
- Red Hat, please visit the "Certified Hardware" information

To retrieve interoperability support information for Enterprise Storage products when used in a supported host server environment see the IBM System Storage Interoperation Center.

SUSE Linux Enterprise Server 9 (GA 08/2004):
Kernel 2.6.5, GCC 3.3.3, Service Pack 4 (GA 12/2007)
SUSE Linux Enterprise Server 10 (GA 07/2006)
Kernel 2.6.16, GCC 4.1.0, Service Pack 3 (GA 09/2009)
SUSE Linux Enterprise Server 11 (GA 03/2009)
Kernel 2.6.27, GCC 4.3.3
Red Hat Enterprise Linux AS 4 (GA 02/2005)
Kernel 2.6.9, GCC 3.4.3, Update 8 (GA 05/2009)
Red Hat Enterprise Linux AS 5 (GA 03/2007)
Kernel 2.6.18, GCC 4.1.0, Update 5 (GA 04/2010)

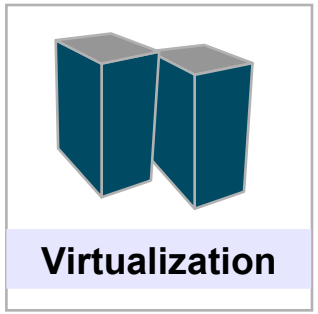
Linux on System z Development Focus



Integration

Application Serving
- z/OS integration

Data Hub
- Database Consolidation

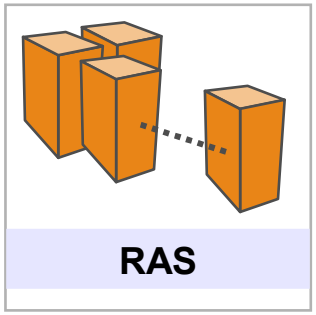


Virtualization

Virtualization &
Virtualization Management
- Ease of Use
- Serviceability
- Hosting capacity

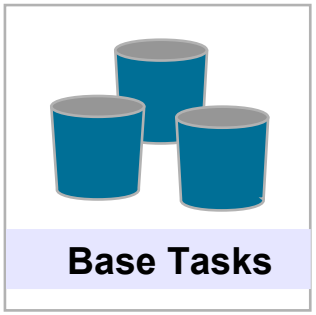
Security

Security
- Certifications
- Data security & privacy



RAS

Continuous Availability
& Data Replication
- RAS
- Differentiation for mission critical workloads



Base Tasks

Customer Requirements
- Address customer observed deficiencies

Competitiveness
- Close competitive gaps
- Differentiation / innovation that matters

Hardware Support
- Exploitation of new System z HW
- Storage exploitation

Linux
- Maintainership & code currency

Future Linux on System z Technology

Software which has already been developed and integrated into the Linux Kernel – but is **not** yet available in any Enterprise Linux Distribution

Virtualization

- **Extra kernel parameter for SCSI IPL (kernel 2.6.32)**
  - Modify the SCSI loader to append append extra parameters specified with the z/VM VMPARM option to the kernel command line.

- **hvc_iucv: Provide IUCV z/VM user ID filtering (kernel 2.6.29)**
  - Introduces the kernel parameter "hvc_iucv_allow=" that specifies a comma-separated list of z/VM user IDs.
  - If specified, the z/VM IUCV hypervisor console device driver accepts IUCV connections from listed z/VM user IDs only.

Note: this list comprises selected items only

Security

- **HiperSockets Network Traffic Analyser (> kernel 2.6.33)**
  - Trace HiperSockets network traffic for problem isolation and resolution. Supported for layer 2 and layer 3
- **OSA QDIO Data Connection Isolation (> kernel 2.6.33)**
  - Feature (available for OSA-Express2 cards since early 2009) allows an operating system to configure the OSA adapter to prevent any direct package exchange between itself and other operating system instances that share the same OSA adapter.
  - The configuration of the isolation level is done through sysfs.
  - Starting with s390-tools 1.8.4 lsqeth indicates connection isolation in qeth attribute 'isolation'.
- **Crypto Express 3 (kernel 2.6.33)**
  - Support for Crypto Express3 Accelerator (CEX3A) and Crypto Express3 Coprocessor (CEX3C)
  - z/VM 6.1, 5.4, or 5.3 with the PTF for APAR VM64656 is required for

Note: this list comprises selected items only

RAS

- **Suspend / resume support (kernel 2.6.31)**
  - Add the ability to stop a running Linux system and resume operations later on.
  - The image is stored on the swap device and does not use any system resource while suspended.
  - Only suspend to disk is implemented, suspend to RAM is not supported.
- **Add Call Home data on halt and panic if running in LPAR (kernel 2.6.32)**
  - Report system failures (kernel panic) via the service element to the IBM service organization.
  - Improves service for customers with a corresponding service contract. (by default this features is deactivated)
- **cio, dasd: Improved DASD error recovery (2.6.33)**
  - Improved the DASD error recovery procedures used in the early phases of IPL and DASD device initialization.

Note: this list comprises selected items only

Suspend / resume support

- Ability to stop a running Linux on System z instance and later continue operations
- Memory image is stored on the swap device specified with a kernel parameter: **resume=/dev/dasd<x>**
- Lower the swap device priority for the resume partition

```
root@larsson:~> grep swap /etc/fstab
/dev/dasdb1 swap swap pri=-1 0 0
/dev/dasdc1 swap swap pri=-2 0 0
```

- Suspend operation is started with a simple echo:

```
root@larsson:~> echo disk > /sys/power/state
```

- Resume is done automatically on next IPL
- Use signal quiesce to automatically suspend a guest

```
ca::ctrlaltdel:/bin/sh -c "/bin/echo disk > \
/sys/power/state || /sbin/shutdown -t3 -h now"
```


New Linux on System z Storage Features

- **DASD: Add support for large volume (EAV) (kernel 2.6.29)**
  - Extended Address Volume (EAV) is available for IBM System Storage DS8000 since R4.0. Support for EAV is also known as "large volume support".
  - The dasd device driver will now support ECKD devices with more than 65,520 cylinders.
  - EAV support is available for Linux on System z running as a z/VM-guest if you are using z/VM 5.4 or z/VM 6.1 with the PTFs for APARs VM64709 (CP) and VM64711 (CMS).
- **DASD High Performance FICON (kernel 2.6.30)**
  - Adds support for the zHPF protocol to the DASD driver
    zHPF provides a much simpler link protocol than FICON: Promises increased I/O bandwidth due to better channel utilization
  - This features is available with DS8000 R4.1
- **FCP SCSI error recovery hardening (kernel 2.6.30)**
  - Avoid SCSI error recovery escalation in case of concurrent zfcp and SCSI error recovery.

Note: this list comprises selected items only

Miscellaneous Linux on System z Features

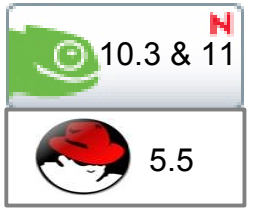
- **Kernel vdso support (kernel 2.6.29)**
  - Kernel provided shared library to speed up a few system calls (gettimeofday, clock_getres, clock_gettime)
- **Kernel image compression (> kernel 2.6.33)**
  - The kernel image size can be reduced by using one of three compression algorithms: gzip, bzip2 or lzma.
- **KULI (2009-06-24)**
  - kuli is experimental userspace sample to demonstrate that KVM can be used to run virtual machines on Linux on System z.
  - This experimental proof of concept is unsupported and should not be used for any production purposes.
- **Oprofile**
  - Starting with version 0.9.4, oprofile supports sampling of Java byte code applications for Linux on System z.
- **Eclipse 3.3**
  - Starting with Eclipse 3.3, Linux on System z is officially supported.

Current Linux on System z Technology

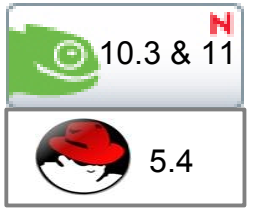
Features & Functionality contained in the Novell & Red Hat Distributions

Integration

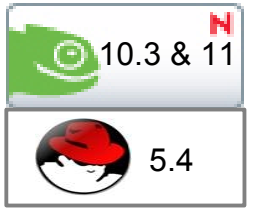
- **AF_IUCV SOCK_SEQPACKET support**
  - Introduce AF_IUCV sockets of type SOCK_SEQPACKET that map read/write operations to a single IUCV operation.
  - The socket data is not fragmented.
  - The intention is to help application developers who write applications using the native IUCV interface, e.g. Linux to z/VSE.



- **HiperSockets Layer3 support for IPv6**
  - Providing Layer3 IPv6 communication, for communication to z/OS



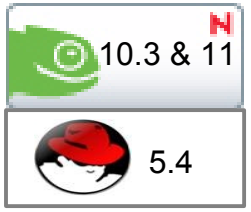
- **Linux to add Call Home data if running in LPAR**
  - Also referred to as Control Program Identification (CPI) or SCLP_CPI
  - Allows the user to set information about the LPAR which will be displayed on the HMC/SE



Note: this list comprises selected items only

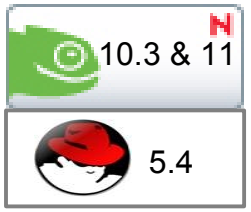
Virtualization

- **TTY terminal server over IUCV**
  - Provide central access to the Linux console for the different guests of a z/VM.
  - Fullscreen applications like *vi* are usable on the console.
  - Access Linux instances with no external network because IUCV is independent from TCP/IP

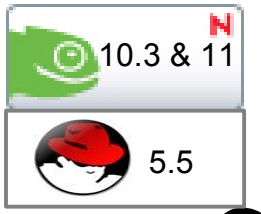


- **Dynamic memory attach/detach**
  - Allows to attach/detach memory for Linux as a guest without needing to reipl.

- **Extra kernel parameter via VMPARM**
  - Allows to use z/VM VMPARM variable to add or substitute the kernel command line.

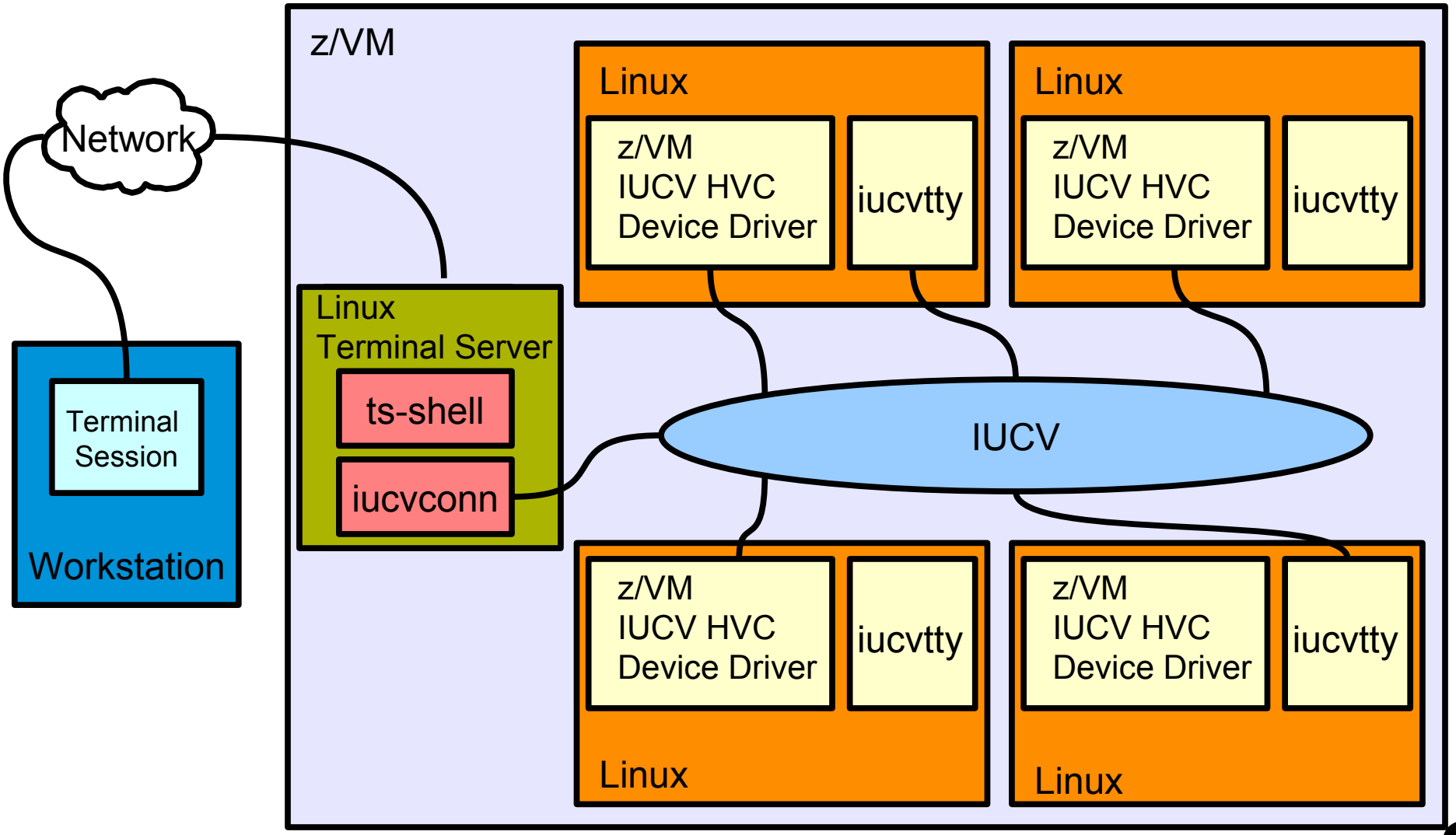


- **Provide CMS script for initial IPL**
  - Avoids having to create an script to start a new installation under z/VM.



Note: this list comprises selected items only

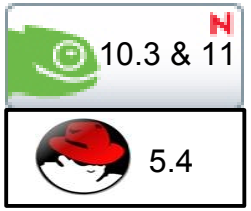
IUCV terminal environment



Virtualization (cont.)

- **Exploitation of DCSSs above 2G**
  - Solves restriction to use DCSS above or greater than 2GB.

- **Provide service levels of HW & Hypervisor in Linux**
  - Improves serviceability by providing uCode and z/VM levels via /proc interface

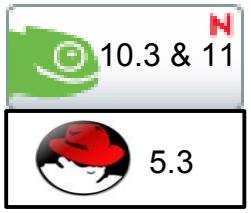



```
root@larsson:~> cat /proc/service_levels
VM: z/VM Version 5 Release 2.0
service level 0801(64-bit)
qeth: 0.0.f5f0 firmware level 087d
```


Note: this list comprises selected items only

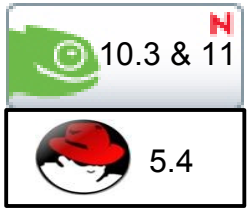
Security

- **Long Random Numbers Generation**
  - Provide access to the random number generator feature on the Crypto card (high volume random number generation, compared to a CPU based solution)



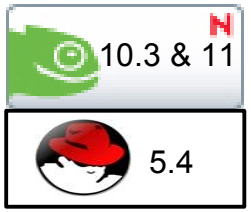
- **Crypto Express3 cards enablement**
  - Support for Crypto Express3 Accelerator (CEX3A) and Crypto Express3 Coprocessor (CEX3C)

- **Crypto device driver use of thin interrupts**
  - Provides better performance and lower CPU consumption.

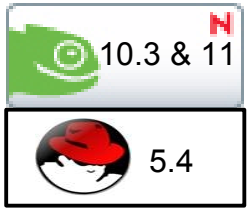


Note: this list comprises selected items only

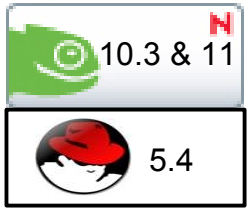
RAS



- **Shutdown Actions Interface**
  - The shutdown actions interface allows the specification of a certain shutdown action (stop, ipl, reipl, dump, vmcmd) for each shutdown trigger (halt, power off, reboot, panic)
  - 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 introduces a system configurations which allows to create a dump in case of a Linux panic, followed by a re-ipl of the system, once the dump was successfully created.
  - Allows to configure system to re-ipl after a dump is taken.

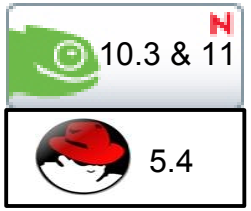


- **Large image dump on DASD**
  - Solves restriction to dump only 48GB of memory to DASD.
  - Now up to 32 ECKD DASDs can be used in a multiple volume configuration

Note: this list comprises selected items only

Storage

- **FICON DS8000 Large Volume (EAV) Support**
  - Large Volume Support is a feature that allows to use ECKD devices with more than 65520 cylinders (>50GB).
  - This features is available with DS8000 R4.0 Allows to exploit
- **DS8000 Disk Encryption Support**
  - Shows the encryption status of the DS8000 Storage.
- **EMC Symmetrix DASD Format Record 0**
  - Allows to initialized unformatted disks on EMC storage arrays
- **FCP LUN discovery tool**
  - New LUN discovery tool: lsluns (e.g. used by yast)
- **FCP performance data collection & reports:**
  - Fibre Channel Protocol (FCP) performance data can now be measured.



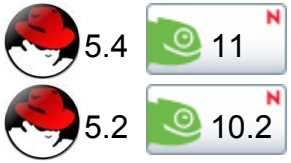
Note: this list comprises selected items only

Red Hat Enterprise Linux 5 Update 5

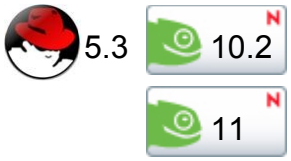
- **GA since 03/30/2010**
  - Kernel GA: 2.6.18-194
- **New Features:**
  - ***FICON DS8000 Large Volume (EAV) Support:*** Allows to exploit DS8000 Storage feature to use DASD volumes >50GB.
  - ***AF_IUCV SOCK_SEQPACKET support:*** Enhances existing AF_IUCV to allow customer to develop using SOCK_SEQPACKET.
  - ***Provide CMS script for initial IPL:*** Avoids having to create an script to start a new installation under z/VM.
  - ***Installer re-IPL support:*** Solves past restriction and allows the installer to direct reboot in the installed system right after installation
- **Bugfixes**

Overview: Linux on System z10 Features

- New capabilities through new instruction support
  - Compiler supports *Decimal Floating Point* (DFP)
  - Kernel and user space libraries support *new CPU crypto algorithms* like SHA-512/384 and AES192/256



- z10 exploitation for improved LPAR performance
  - *Node affinity* aligns process scheduling to book boundaries
  - *Vertical CPU Management* concentrates workload on fewer physical CPUs instead of spreading virtual CPs over physical CPs



- New functions with z10
  - *Large Page Support* minimizes lookup overhead into areas of large memory, with large page emulation on older hardware
  - *HiperSockets Layer 2* support for simplification of Linux environments
  - *Crypto Express 3* Enablement and *performance improvements*

Note: this list comprises selected items only

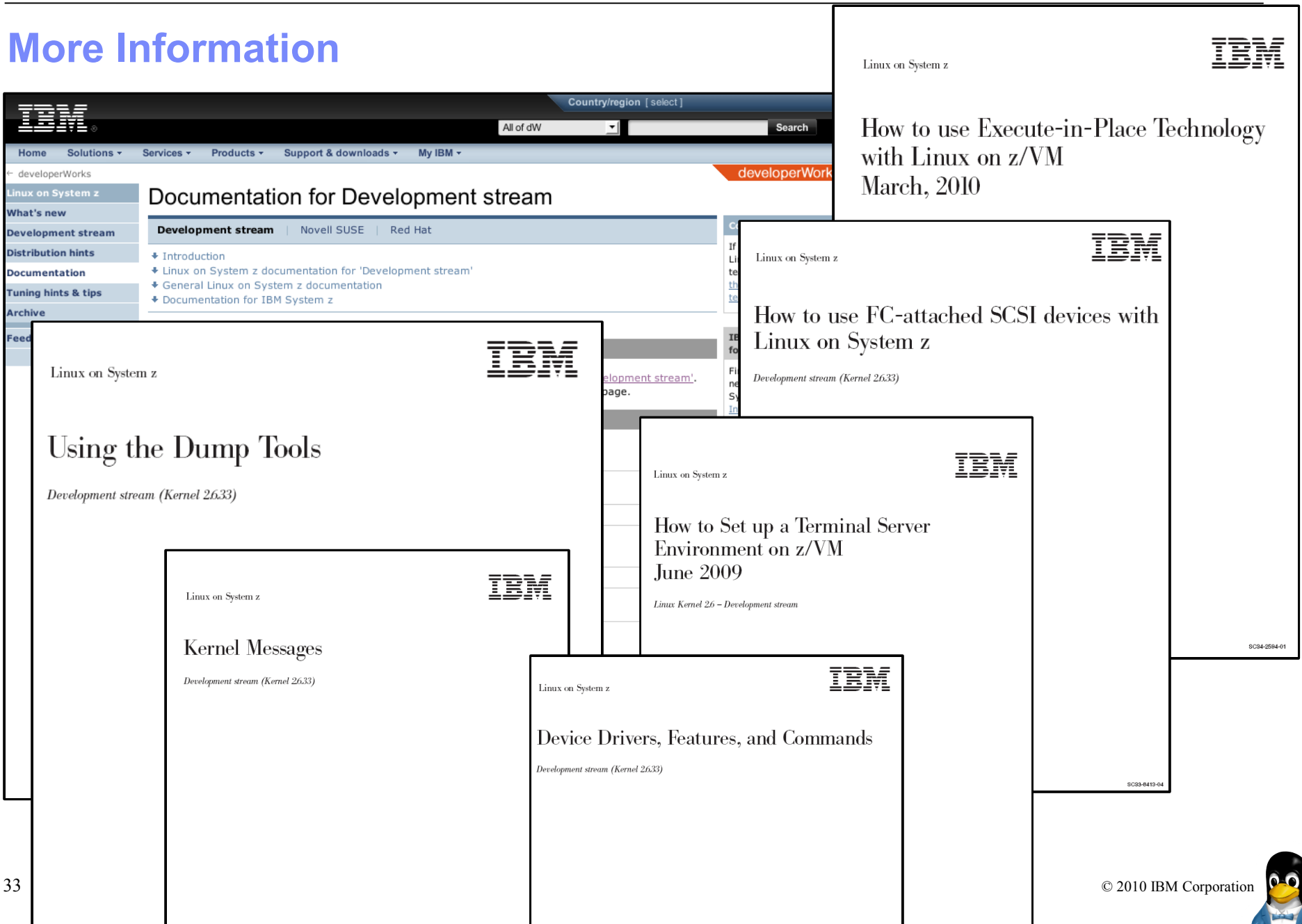
The s390-tools package

- s390-tools is a package with a set of user space utilities to be used with the Linux on System z distributions.
  - It is **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 .
- Version 1.8.4 was released on 2010-03-12
- This software package is contained in all major (and IBM supported) enterprise Linux distributions which support s390
  - RedHat Enterprise Linux 4
  - RedHat Enterprise Linux 5
  - SuSE Linux Enterprise Server 10
  - SuSE Linux Enterprise Server 11
- Website: http://www.ibm.com/developerworks/linux/linux390/s390-tools.html
- Feedback: linux390@de.ibm.com

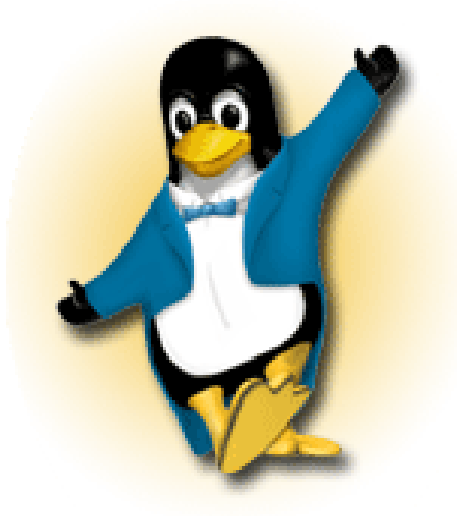
s390-tools 1.8.4

- **New tools**
  - *60-readahead.rules:* udev rules to set increased "default max readahead". This will increase sequential read performance up to 40%.
- **Changes to existing tools**
  - *lsqeth:* Introduce "lay'2" column for "lsqeth -p" and new qeth attribute "isolation".
  - *dumpconf:* Prevent re-IPL loop for dump on panic.
  - *qethconf:* Indicate command failure and show message list.
  - *zipl:* Improve I/O error recovery during IPL & Automatically calculate the ramdisk address dependent on the kernel image size.
- **Bug Fixes**
- **New in 1.8.3:**
  - *zipl:* Add support for device mapper devices: zipl now allows installation of and booting from a boot record on logical devices, i.e. devices managed by device mapper (or similar packages), e.g. multipath devices.

More Information



Questions?



Hans-Joachim Picht

Linux on System z Initiatives

IBM Deutschland Research & Development GmbH
Schönaicher Strasse 220
71032 Böblingen, Germany

Mobile +49 (0)175 - 1629201
hans@de.ibm.com

Your Linux on System z Requirements?

Are you missing a certain feature, functionality or tool? **We'd love to hear from you!**

We will evaluate each request and (hopefully) develop the additional functionality you need.