IBM and Linux: Community Innovation for your Business

Linux on System z – Whats New?

Dortmund, 28.04.2009

IBM Corporation
Hans-Joachim Picht
hans@linux.vnet.ibm.com

© 2009 IBM Corporation

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States, other countries, or both.

Not all common law marks used by IBM are listed on this page. Failure of a mark to appear does not mean that IBM does not use the mark nor does it mean that the product is not actively marketed or is not significant within its relevant market.

Those trademarks followed by ® are registered trademarks of IBM in the United States; all others are trademarks or common law marks of IBM in the United States.

For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml:

*, AS/400®, e business(logo)®, DBE, ESCO, eServer, FICON, IBM®, IBM (logo)®, iSeries®, MVS, OS/390®, pSeries®, RS/6000®, S/30, VM/ESA®, VSE/ESA, WebSphere®, xSeries®, z/OS®, zSeries®, z/VM®, System i, System i5, System p, System p5, System x, System z, System z9®, BladeCenter®

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.
Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
Microsoft, Windows, Windows NT, and the Windows logo are 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.
Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.
IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

* All other products may be trademarks or registered trademarks of their respective companies.

Notes:
Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.
IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.
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 subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

© 2009 IBM Corporation

Agenda

* Linux on System z Distributions
* Linux Common Code news
* What's New in System z
  - Kernel
  - GCC
  - s390-tools
  - Kuli
* Where to find more Information

© 2009 IBM Corporation

Linux on System z distributions (Kernel 2.6 based)

- **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 2 (GA 05/2008)
- **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 7 (GA 07/2008)
- **Red Hat Enterprise Linux AS 5 (GA 03/2007)**
  - Kernel 2.6.18, GCC 4.1.0, Update 3 (GA 01/2009)
- **Others**
  - Debian, Slackware, ...
  - Support may be available by some third party

© 2009 IBM Corporation

Supported Linux Distributions

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	✓	✓	✓

✓ 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.

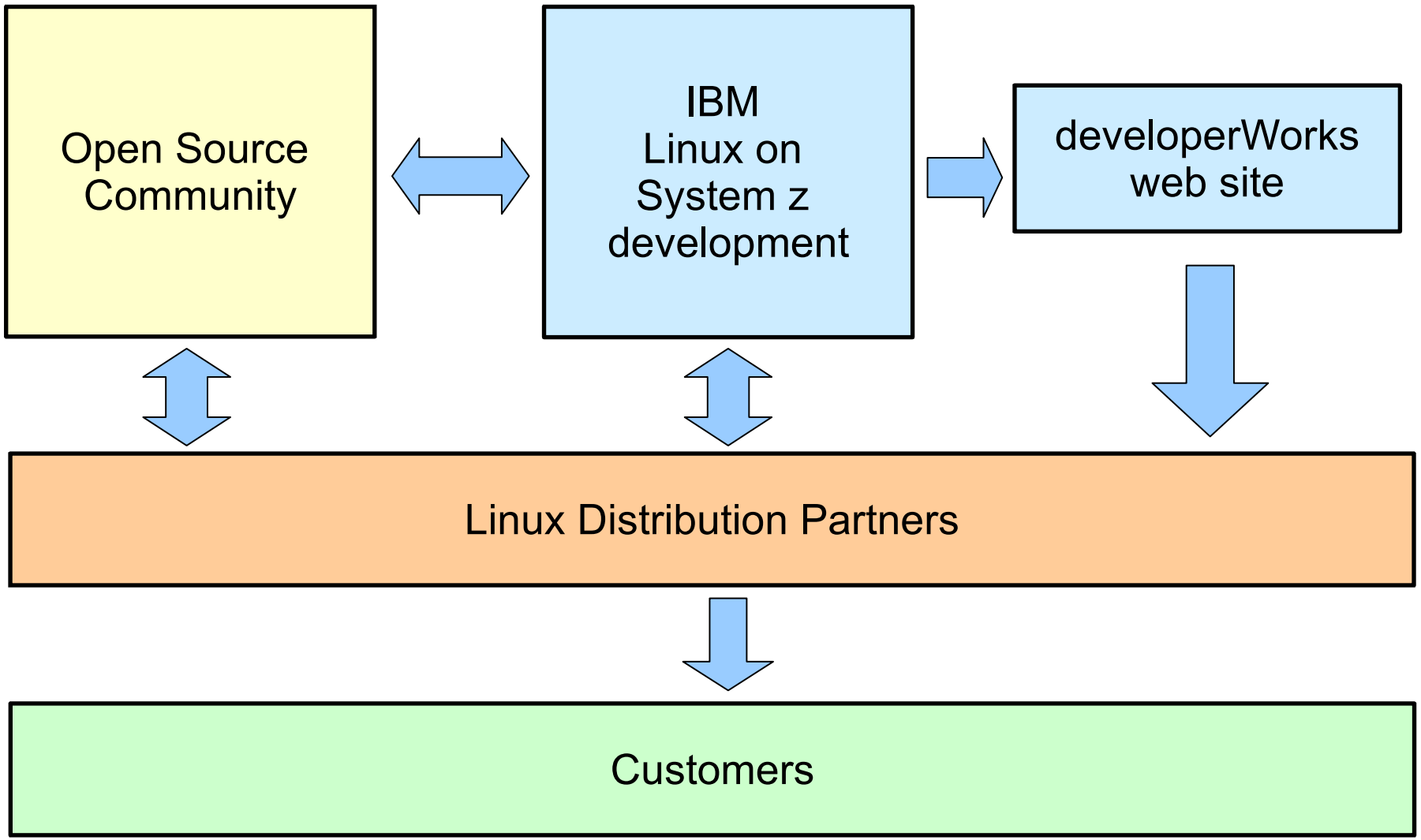
✗ Indicates that the distribution is not supported by IBM.

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

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

© 2009 IBM Corporation

Linux on System z development process



© 2009 IBM Corporation

Current Linux Kernel Development

Most active 2.6.26 employers

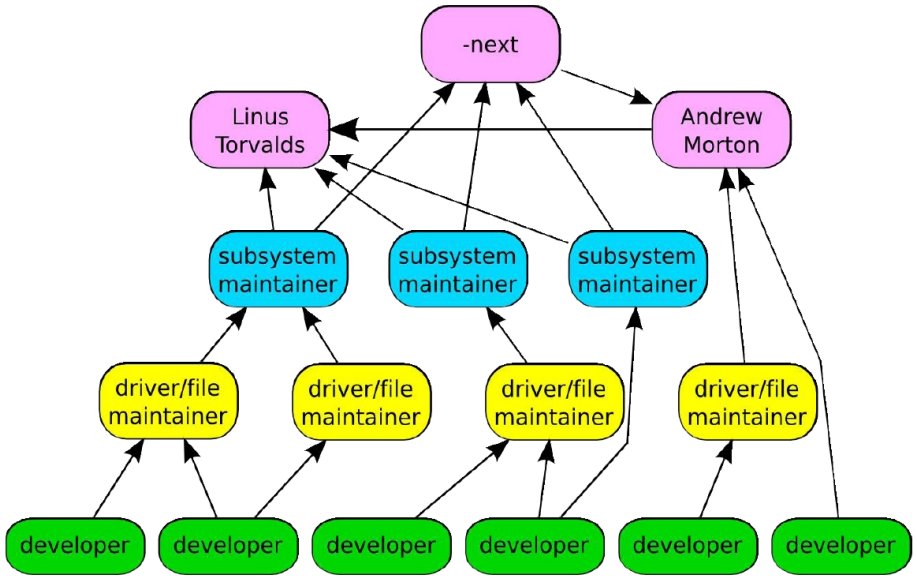
By changesets			By lines changed		
(None)	2085	20.6%	(None)	111703	15.7%
Red Hat	1130	11.2%	IBM	73601	10.3%
(Unknown)	906	8.9%	Red Hat	56331	7.9%
IBM	609	6.0%	Intel	50297	7.1%
Novell	597	5.9%	(Unknown)	44699	6.3%
Intel	469	4.6%	Vyatta	41835	5.9%
Parallels	312	3.1%	Novell	33745	4.7%
SGI	211	2.1%	Movial	28632	4.0%
Movial	180	1.8%	Hauppauge	20234	2.8%
Oracle	142	1.4%	Analog Devices	18363	2.6%
Analog Devices	134	1.3%	(Consultant)	16397	2.3%
HP	124	1.2%	Solarflare	15585	2.2%
MontaVista	122	1.2%	Freescale	15090	2.1%
(Consultant)	116	1.1%	MontaVista	14013	2.0%
Freescale	109	1.1%	QLogic	13327	1.9%
QLogic	97	1.0%	SGI	10351	1.5%
Fujitsu	95	0.9%	Marvell	7881	1.1%
Google	94	0.9%	Wind River	7770	1.1%
(Academia)	89	0.9%	Oracle	7680	1.1%
Marvell	88	0.9%	Pengutronix	7334	1.0%

Source:http://lwn.net/Articles/288233/

4.300 lines added
1.800 lines removed
1.500 lines modified

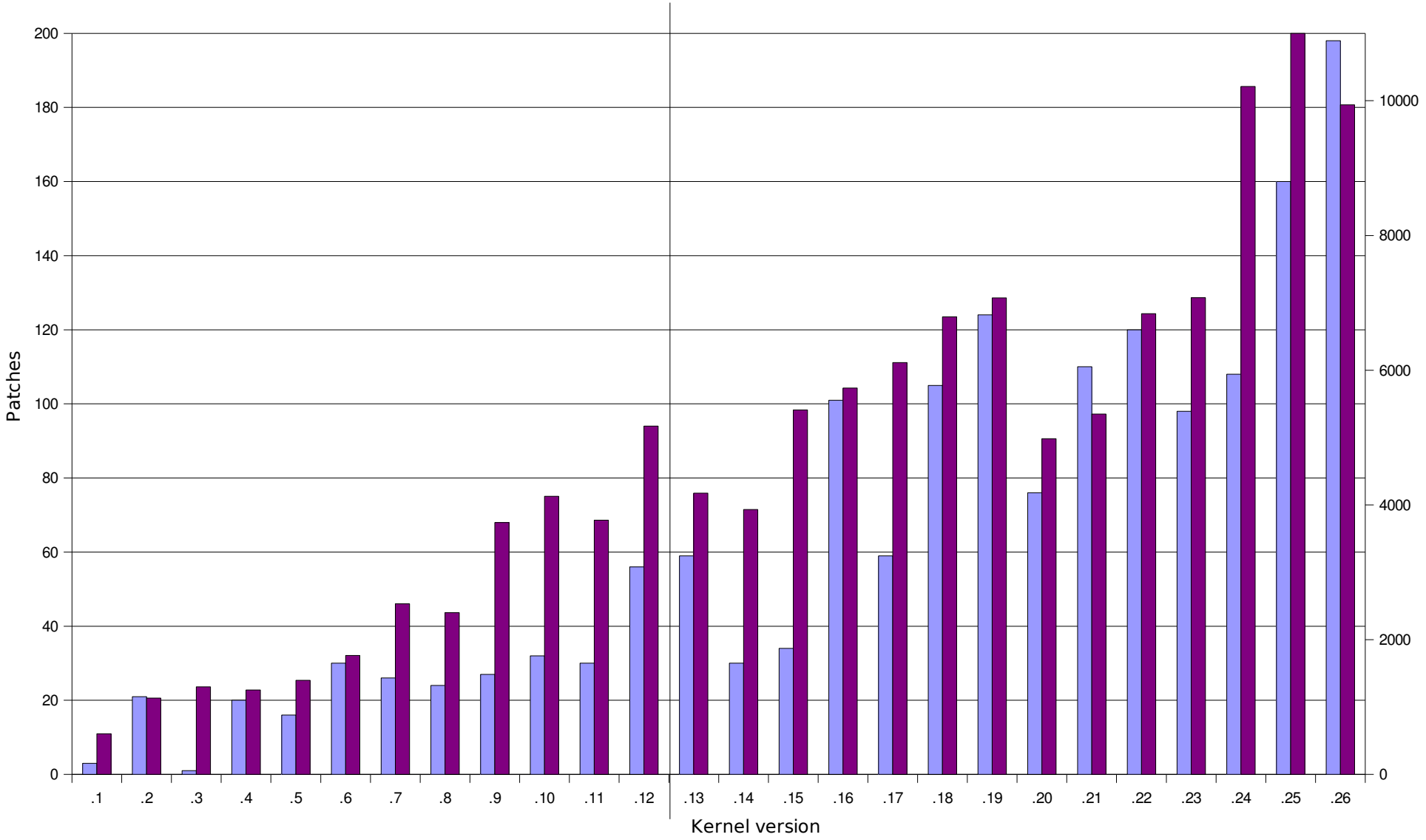
per day 2007-2008

Source: Greg KH



© 2009 IBM Corporation
© 2008 IBM Corporation

Linux kernel – System z contributions



© 2009 IBM Corporation

Common Code Kernel News

* Linux version 2.6.23 (2007-10-09)
  - Completely Fair Scheduler (CFS)
  - Variable argument length (no more “arg list too long”)
* Linux version 2.6.24 (2008-01-24)
  - CFS improvements: performance, fair group scheduling, guest time
  - Anti-fragmentation patches
* Linux version 2.6.25 (2008-04-16)
  - Latencytop
* Linux version 2.6.26 (2008-07-13)
  - Kgdb
  - read-only bindmounts
  - KVM enhancements (ia64, s390, ppc)
* Linux version 2.6.27 (2008-10-09)
  - Lockless page
  - Ubifs filesystem for flash media
  - Multiqueue networking
  - Ftrace function tracer
* Linux version 2.6.28 (2008-12-24)
  - Ext4 filesystem
  - Memory management scalability improvements
  - Unified trace buffer for LLT ng, ftrace, etc.
* Linux version 2.6.29 (2009-03-23)
  - Btrfs and squashfs filesystems
  - Security module hooks for path based access control (AppArmor, Tomoyo)

© 2009 IBM Corporation

Linux Kernel Directions

* Diversity: now 20 architectures (26 w/o unification)
  alpha (64 bit), arm (32 bit), avr32 (32 bit), blackfin (32 bit), cris (32 bit), frv (32 bit), h8300 (32 bit), ia64 (64 bit), m32r (32 bit), m68k (32 bit), m68knommu (32 bit), mips (32 bit), mips (64 bit), mn10300 (32 bit), pa-risc (32 bit), powerpc (32 bit), powerpc (64 bit), s390 (32 bit), s390 (64 bit), sh (32 bit), sh (64 bit), sparc (32 bit), sparc (64 bit), x86 (32 bit), x86 (64 bit), xtensa (32 bit)
* Emphasis on larger, more powerful machines
* Virtualization (KVM, paravirt, XEN), continues to attract a lot of attention

- **Linux is Linux, but**
  - Features, properties and quality differ dependent on your platform

© 2009 IBM Corporation

The developerWorks “Development Stream”

- The developerWorks "Development stream" describes Linux on System z contributions and functionality against recent upstream kernel and other upstream Linux components which can be used to build Linux on System z distributions.
- Currently, (2008-11-25) the "Development stream" consists of and has been tested with:
  - Kernel 2.6.27
  - Toolchain:
    - GCC 4.3.2 with patch for z10 exploitation
    - GNU Binutils - binutils 2.18.50.0.7
    - GNU C Library - glibc 2.8 with patch for utmp-support
  - Utilities:
    - s390-tools 1.8.0
    - zfcp HBA API 2.0
  - Debug Tools:
    - GDB 6.8, see http://sources.redhat.com/gdb/download/
    - strace 4.5.18, see: http://sourceforge.net/projects/strace/
    - lkcdutils (LKCD) SVN HEAD, see: http://sourceforge.net/projects/lkcd/
    - crash, latest available, see: http://people.redhat.com/anderson/

© 2009 IBM Corporation

The developerWorks “Development Stream” (cont.)

* Exploitation of features introduced with IBM System z10:
  - Toolchain support for z10 instructions with:
    - GCC 4.3.2 patch for z10 exploitation
    - Binutils 2.18.50.0.7
  - Automatic CPU detection (kernel 2.6.27 – upstream)
* Exploitation of features introduced with IBM System z9:
  - Server time protocol (STP) support for clock synchronization (kernel 2.6.27 – upstream)
  - HiperSockets IPv6 support for Layer 3 - formerly, IPv6 support was only available when using HiperSockets Layer 2 (kernel 2.6.27 - patch 01)
* Exploitation of other IBM System z features:
  - Enable to attach and use standby memory that is configured for a logical partition or z/VM guest (requires z/VM 5.4 plus the PTF for APAR VM64524) (kernel 2.6.27 – upstream)
  - Dynamic memory attach/detach (requires z/VM 5.4 plus the PTF for APAR VM64524) (kernel 2.6.27 – upstream)

© 2009 IBM Corporation

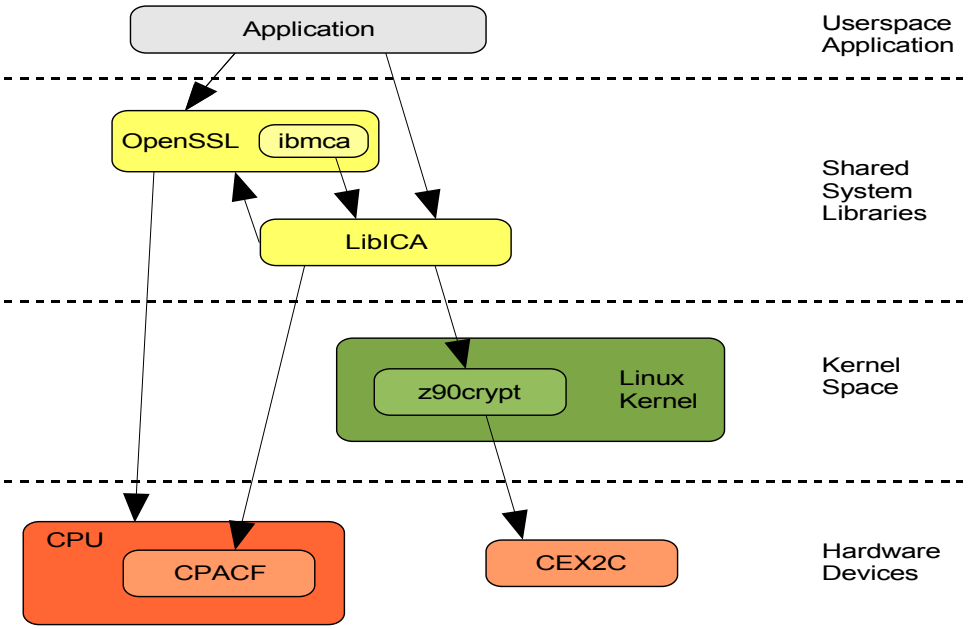
The developerWorks “Development Stream” (cont.)

* Exploitation of z/VM 5.4 features:
  - Expanded shared memory addressability: Linux on System z can now use Discontiguous Saved Segments (DCSS) above 2047 MB (2G) of virtual storage (kernel 2.6.27 - patch 01)
  - Capability to dump Linux guests to SCSI disks: Requires zfcpdump-support for memory holes because z/VM allows to have discontinuous memory (s390-tools 1.8.0)
* Other enhancements:
  - Processor-type safety-check, preventing a kernel to run a processor if it was compiled to exploit instructions of a newer machine (kernel 2.6.27 - patch 01)
  - New IPL tools (s390-tools 1.8.0)
  - zipl can dump on multiple ECKD DASD devices (s390-tools 1.8.0)
  - Enhanced zfcp trace facility (kernel 2.6.27 – upstream)
  - zfcp performance data collection (incl. z9 or later FCP adapter statistics) via zfcp (patch 01), ziomon-tool (s390-tools 1.8.0), and blktrace-1.0.0
  - zfcp Host Bus Adapter application programming interface (zfcp HBA API 2.0)

© 2009 IBM Corporation

Crypto development

* **Cleanup of libICA crypto library**
  - Version 2.0 of the libICA library has been published with a set of simpler functions for the existing interfaces.
  - See http://sourceforge.net/projects/opencryptoki/



© 2009 IBM Corporation

SLES 11

- **Vertical CPU management (kernel 2.6.25)**
  - With this feature it is possible to switch between horizontal and vertical CPU polarization via a sysfs attribute.
  - If vertical CPU polarization is active then the hypervisor will dispatch certain CPUs for a longer time than other CPUs for maximum performance.
  - This support is available only on z10, running Linux on System z in an LPAR.
- **I/O configuration support (kernel 2.6.27)**
  - Adds the infrastructure to allow Linux system to change the I/O configuration of a System z system.
  - Operations are addition, removal and reconfiguration/reassignment of I/O channels, control units and subchannels.
  - This support is available only when running Linux on System z in an LPAR
- **DASD HyperPAV support (kernel 2.6.25)**
  - Parallel access volumes (PAV) is a storage server feature, that allows to start multiple channel programs on the same DASD in parallel.
- **FCP automated port discovery (kernel 2.6.25)**
  - Scan the connected fiber channel SAN and automatically activate all available and accessible target ports. This requires a proper SAN setup with zoning.

© 2009 IBM Corporation

SLES 11 (cont.)

- **Extra kernel parameter via VMPARM (kernel 2.6.27)**
  - Modify the IPL records to append extra parameters specified wit the z/VM VMPARM option to the kernel command line.
- **TTY terminal server over IUCV (kernel 2.6.29)**
  - 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.
- **Support for enhanced z/VM DASD UIDs (kernel 2.6.27)**
  - When z/VM provides two virtual devices (minidisks) that reside on the same real device, both will receive the configuration data from the real device and thus get the same uid. To fix this problem, z/VM provides an additional configuration data record that allows to distinguish between minidisks. z/VM APAR VM64273 needs be installed to enable enhanced DASD IUDs.
- **Add vmconvert option to vmur tool (s390-tools 1.8.0)**
  - Simplify the copy of a z/VM dump from the reader to Linux
- **FCP performance data collection: adapter statistics (2.6.26)**
  - The zFCP adapter collects a number of statistics about the virtual adapter. This information is fetched by the driver and is exported to user space via sysfs. This support is available only on IBM System z9 or later.

© 2009 IBM Corporation

SLES 11 (cont.)

- **QETH componentization (kernel 2.6.25)**
  - The qeth driver module is split into a core module and layer2-/layer3-specific modules.
- **Generic algorithm fallback (kernel 2.6.25)**
  - Use software implementation of the in-kernel crypto library for key lengths not supported by hardware. Without the fallback support it is not possible to use in-kernel crypto with a key length that is not supported by the hardware module.
  - **Standby CPU activation/deactivation (kernel 2.6.25)**
  - With this feature it is possible to make use of standby CPUs for instruction execution.
  - A CPU can be in one of the states "configured", "standby", or "reserved".
  - This support is available only on IBM System z10, when running Linux on System z in an LPAR.
- **Shutdown Actions Interface (kernel 2.6.25)**
  - 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.

© 2009 IBM Corporation

SLES 11 (cont.)

* **Dynamic memory add / remove (kernel 2.6.27)**
  - Use the SCLP interface to attach and detach storage elements to the image.
  - Provide the platform support for Linux memory add / remove interface.

* **Struct page elimination (kernel 2.6.26)**
  - Remove the need to allocate a “struct page” structure for pages of a DCSS.
  - No more “mem=” to include the memory areas of the DCCS segments in the memory map.

* **STP Support (kernel 2.6.27)**
  - Support for clock synchronization using the server time protocol (STP)
  - This support is available only when running Linux on System z in an LPAR.

© 2009 IBM Corporation

RHEL 5.3 & SLES 10.2

* **CPU node affinity (kernel 2.6.25)**
  - With this feature the kernel uses CPU topology information as supplied by the IBM System z10. This information is used by the scheduler to build scheduling domains and should increase overall performance on SMP machines.
  - This support is available only on IBM System z10, when running Linux on System z in an LPAR.

* **DASD: system information messages (kernel 2.6.25)**
  - With this feature the system reports system information messages (SIM) to the user. The System Reference Code (SRC), which is part of the SIM, is reported to the user and allows to look up the reason of the SIM online in the documentation of the storage server.

* **Support two OSA ports per CHPID - Four-port exploitation (kernel 2.6.25)**
  - Exploit next OSA feature which offers two ports within one CHPID The additional port number 1 can be specified with the qeth sysfs-attribute "portno".
  - This support is available only for OSA-Express3 GbE SX and LX on z10, running Linux on System z in an LPAR or as a VM guest (PTF for z/VM APAR VM64277 required).

* **Support for large random numbers (kernel 2.6.25)**
* Allow user space applications to access large amounts of truly random data. The random data source is the built-in hardware random number generator on the CEX2C cards.

© 2009 IBM Corporation

Kernel enhancements > 2.6.29

* **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, s390-tools support for large volumes is required.
* **DASD High Performance FICON (> kernel 2.6.29)**
  - Support access to a storage server attached using the I/O subsystem in transport mode.
  - This features is available with DS8000 R4.
* **DASD Format Record 0 (> kernel 2.6.29)**
  - Allows to initialized unformatted disks on EMC storage arrays
* **Shutdown action IPL after dump (> kernel 2.6.29)**
  - 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.
* **Kernel vdso support (kernel 2.6.29)**
  - Kernel provided shared library to speed up a few system calls (gettimeofday, clock_getres, clock_gettime)

© 2009 IBM Corporation

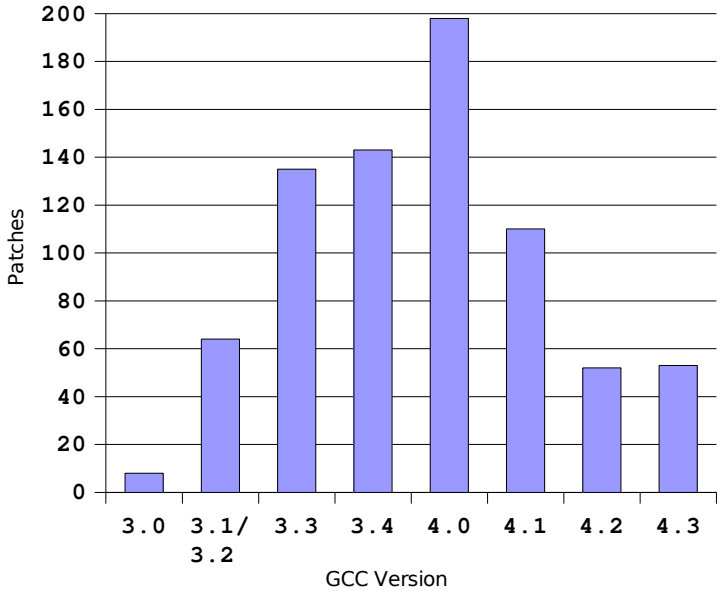
Other packages

* Experimental (unsupported) userspace sample kuli 1.0.0 demonstrating KVM usage (2008-07-04)
  - kuli" is an experimental (unsupported) 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.

© 2009 IBM Corporation

GNU Compiler

* **General optimizer improvements**
  - New data flow analyzer framework (GCC 4.3)
* **System z machine support**
  - System z10 processor support (GCC 4.4)
    - Exploit instruction new to z10
    - Selected via -march=z10 / -mtune=z10
  - Decimal floating point support (GCC 4.3)
    - For newer machines with hardware DFP support
  - 64 bit registers for 31 bit applications (> GCC 4.4)
    - Work in progress, harder than it looks



* **System z compiler performance**
  - Overall enhancement > 10% on z9 with industry-standard integer benchmark
    - 8% comparing GCC 3.4 and GCC 4.1
    - 5.9% comparing GCC 4.1 and GCC 4.2
    - 0.5% comparing GCC 4.2 and GCC 4.3

© 2009 IBM Corporation

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 .
- The current version is 1.8.0 and was released in November 2008
- A new version will be available in May 2009
- This software package is contained in all major (and IBM supported) 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

© 2009 IBM Corporation

NEW

Whats new with version 1.8.0

* New tools
  - chreipl: Change reipl device settings.
  - chshut: Change actions which should be done in case of halt, poff, reboot or panic.
  - lsreipl: List information of reipl device.
  - lsshut: List actions which will be done in case of halt, poff, reboot or panic.
  - ziomon tools: Set of tools to collect data for zfcp performance analysis.
  - lsluns: List available SCSI LUNs depending on adapter or port.
  - lszcrypt: Show information about zcrypt devices and configuration.
  - chzcrypt: Modify zcrypt configuration
* Changes to existing tools
  - ip_watcher: New qeth driver support.
  - lscss: Show non I/O subchannels.
  - lstape: Add SCSI tape support.
  - osasnmpd: New qeth driver support.
  - zfcpdump_v2: Add support for memory holes
  - zipl: Support for virtio devices.
* Bugfixes
  - cpuplugd
  - lsdasd
  - mon_statd
  - zfcpdump_v2
  - zipl dump tools

© 2009 IBM Corporation

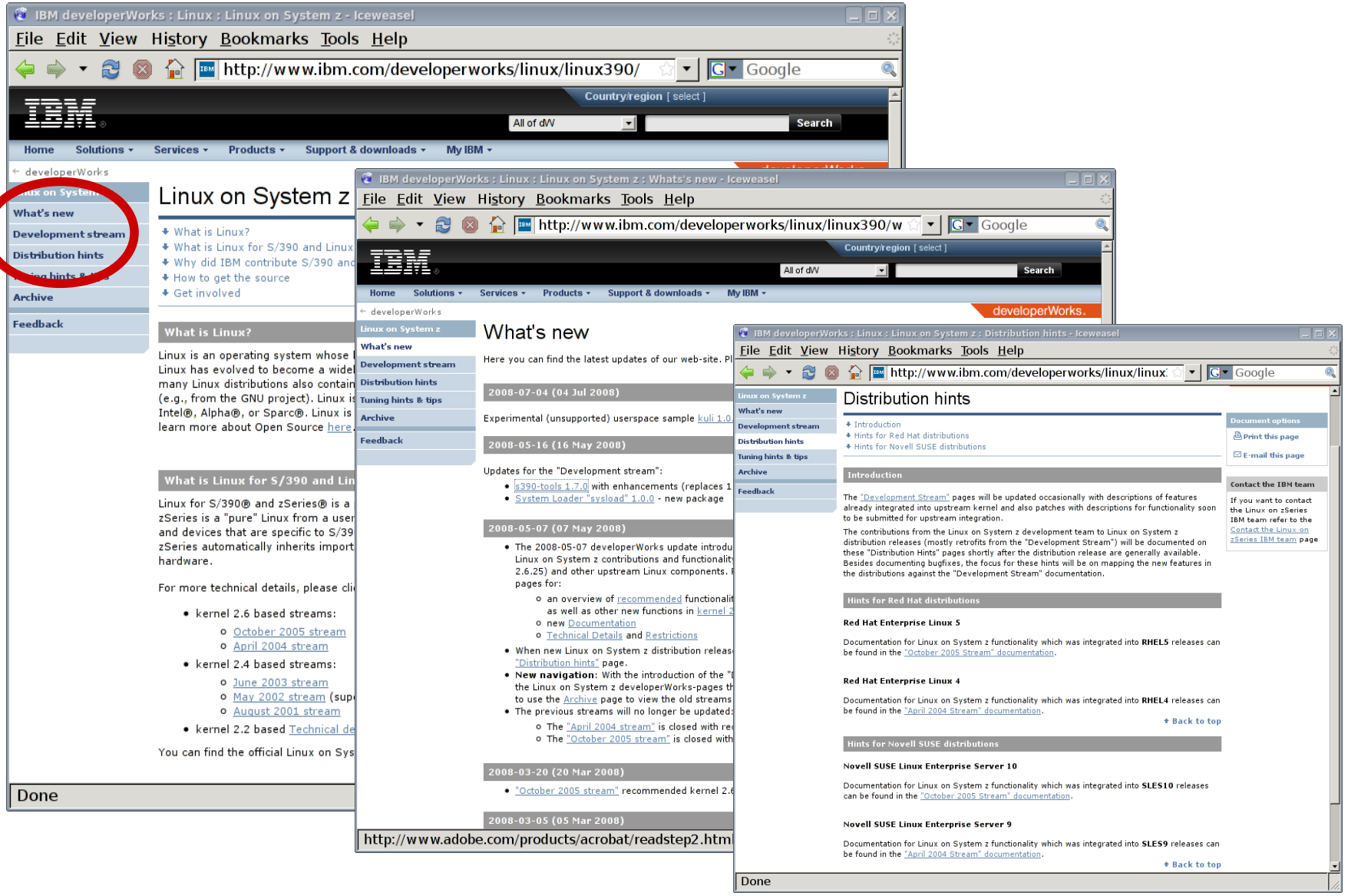
System z kernel features – message documentation

* Document all System z related kernel messages
  - Cleanup messages in System z related code (kernel 2.6.27, >kernel 2.6.29)
  - Script to generate a man page for every kernel message (upstream rejected)
  - Distributors generate man pages for their distributions (up to distributor)

```
xpram.1(9)                                                    xpram.1(9)

Message
       xpram.1: %d is not a valid number of XPRAM devices

Severity
       Error

Parameters
       @1: number of partitions

Description
       The  number  of  XPRAM partitions specified for the 'devs' module parameter or with
       the 'xpram.parts' kernel parameter must be an integer in the range  1  to  32.  The
       XPRAM  device  driver created a maximum of 32 partitions that are probably not con-
       figured as intended.

User action
       If the XPRAM device driver has been complied as a separate module, unload the  mod-
       ule  and  load  it  again with a correct value for the into the kernel, correct the
       'xpram.parts' parameter in the kernel parameter line and restart Linux.

LINUX                          Linux Messages                 xpram.1(9)
```

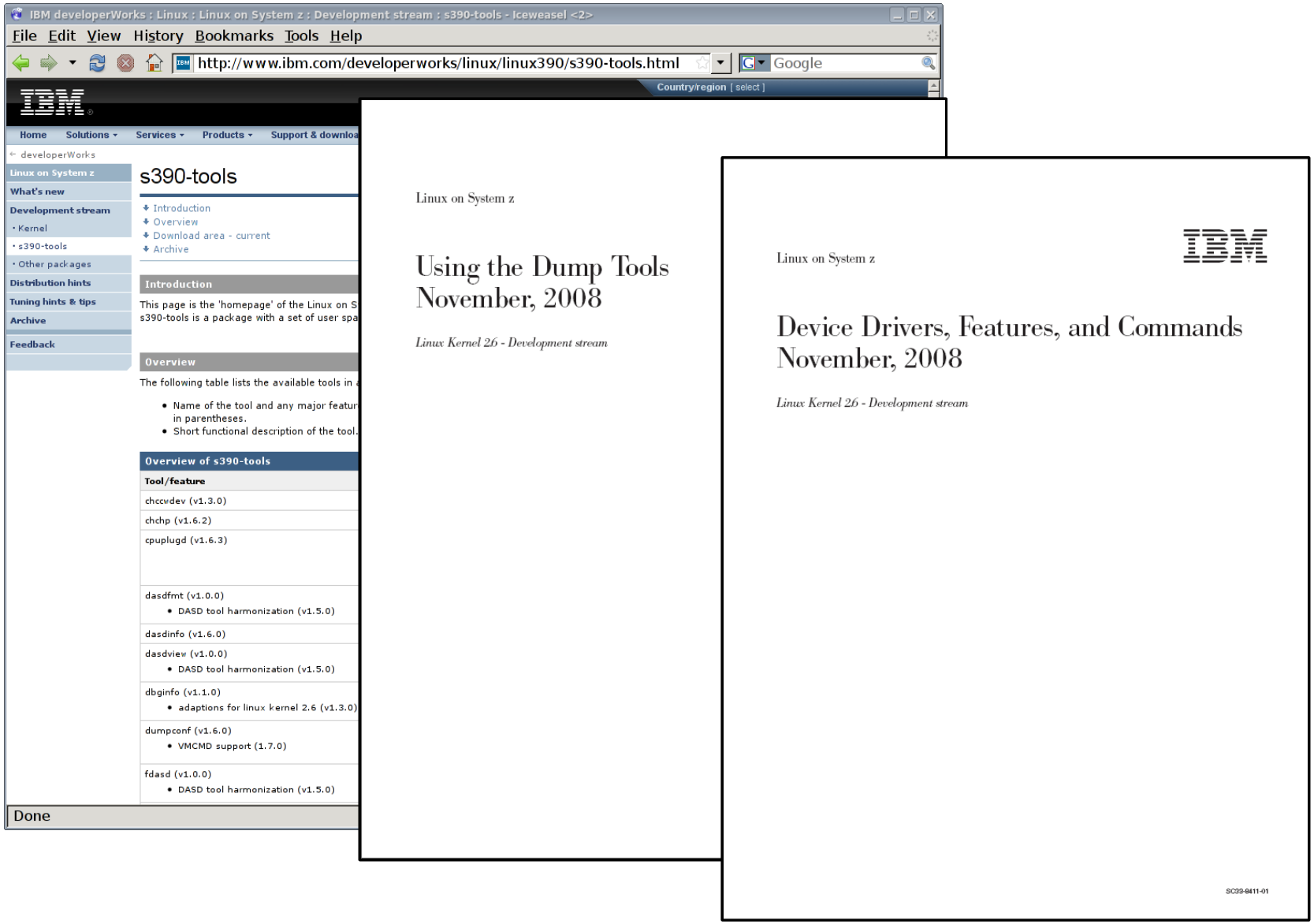

© 2009 IBM Corporation

IBM Linux on System z Website



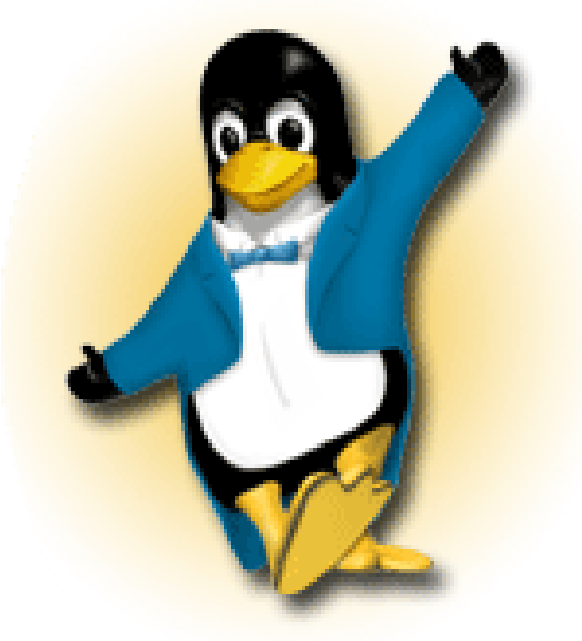
© 2009 IBM Corporation

More Information



© 2009 IBM Corporation

Questions?



Hans-Joachim Picht

Linux Technology Center

Linux on System z Kernel Development & Red Hat Liaision

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

Phone +49 (0)7031-16-1810
Mobile +49 (0)175 - 1629201
hans@linux.vnet.ibm.com

© 2009 IBM Corporation

© 2008 IBM Corporation