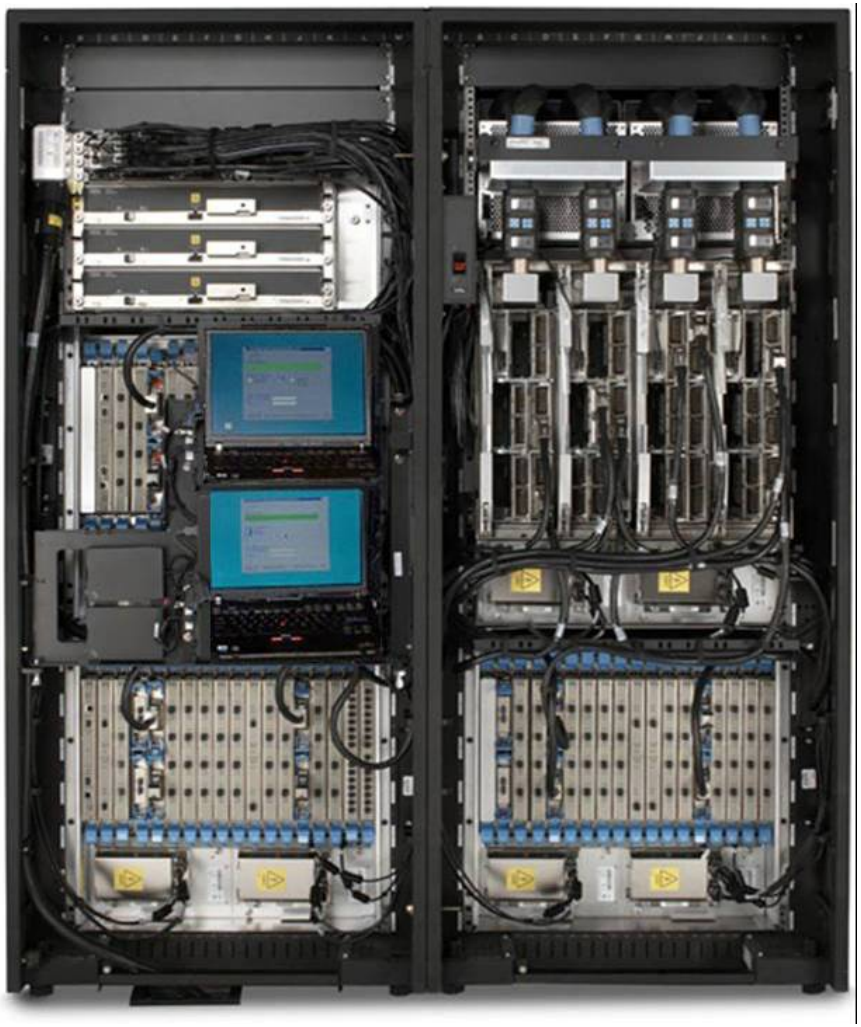
What's New
in Linux on System z

Martin Schwidefsky (schwidefsky@de.ibm.com)
Linux on System z Development
IBM Lab Boeblingen, Germany

Agenda



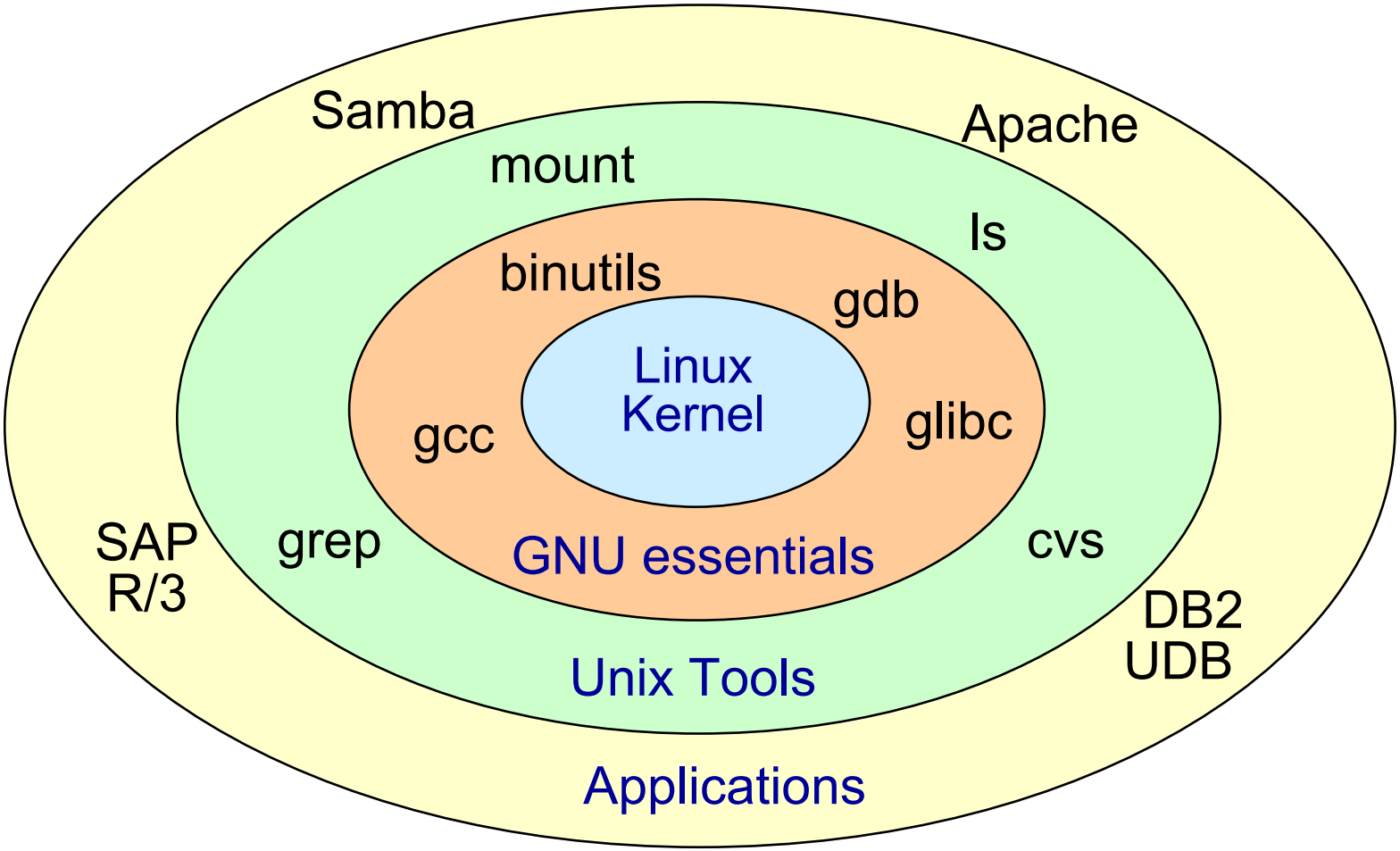
- Linux on System z Overview
- Development Process
  - Linux Kernel
  - Compiler gcc
- Distributor Support
- Linux Kernel News
- What's new n System z

Linux on System z distributions (Kernel 2.6 based)

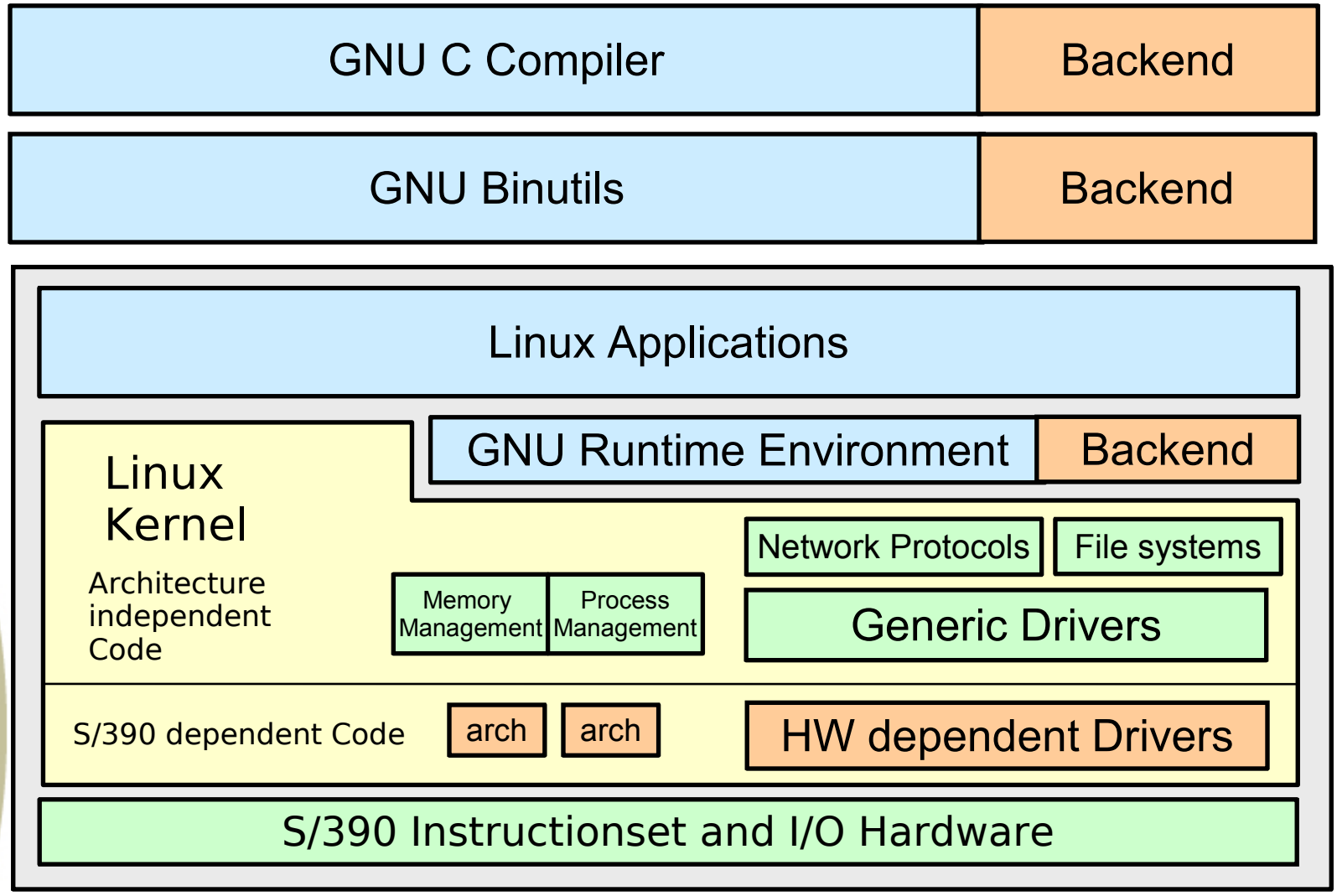
SHARE
Technology · Connections · Results

- **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 1 (GA 06/2007)
- **Red Hat Enterprise Linux AS 4 (GA 02/2005)**
  - Kernel 2.6.9, GCC 3.4.3
  - Update 6 (GA 11/2007)
- **Red Hat Enterprise Linux AS 5 (GA 03/2007)**
  - Kernel 2.6.18, GCC 4.1.0
  - Update 1 (GA 11/2007)
- **Others**
  - Debian, Slackware, ...
  - Support may be available by some third party

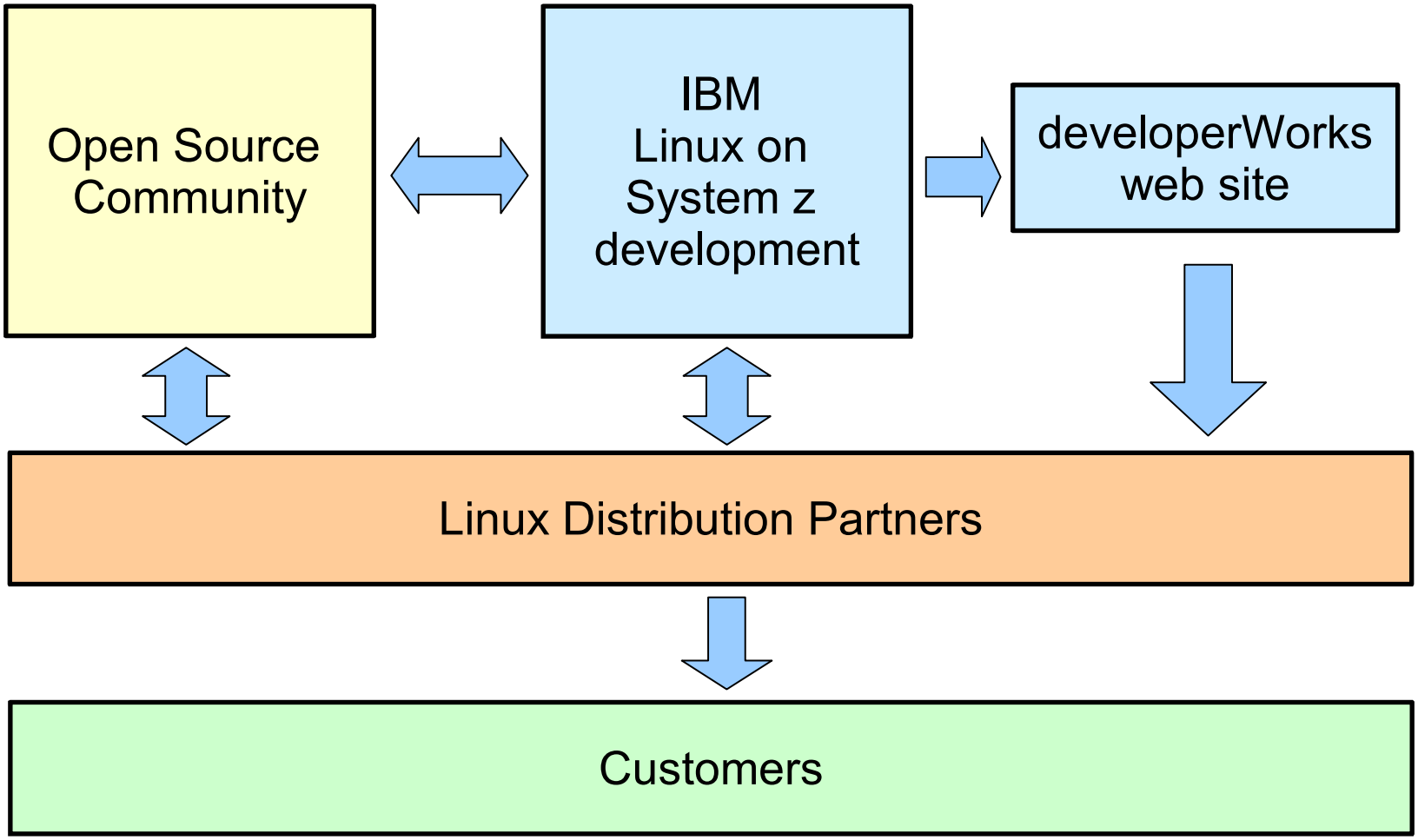
Linux system components



Linux on System z system structure

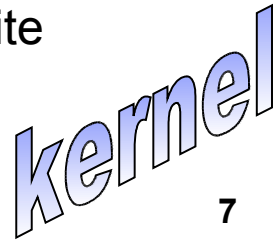


Linux on System z development process

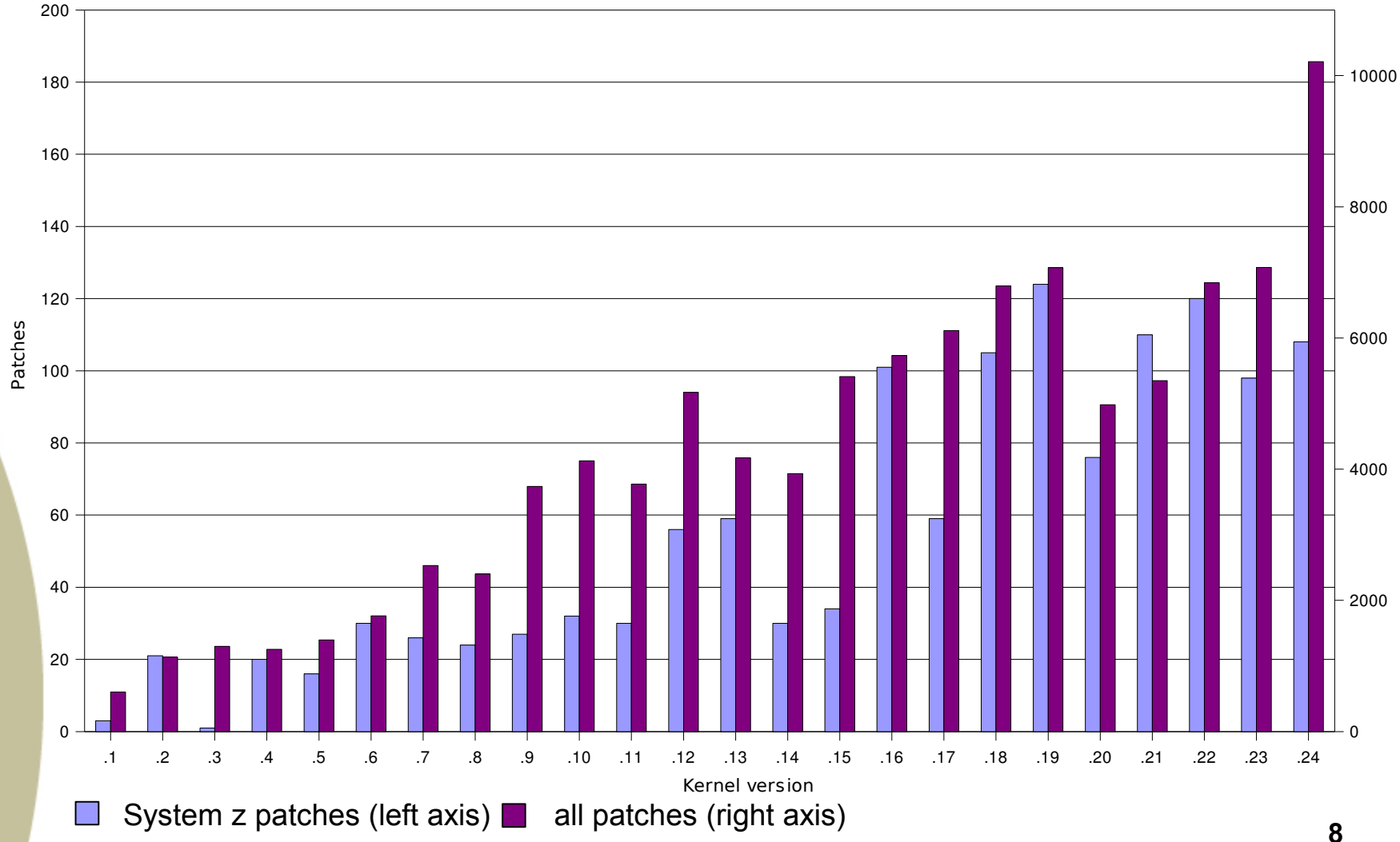


Open Source development process Linux Kernel

- **Distributed development model**
  - Source code control tool: git
  - 'Master' repository maintained by Linus Torvalds
  - 'Experimental' repository maintained by Andrew Morton
  - Secondary repositories maintained by subsystem maintainers
  - Flow of code tracked via “Signed-Off” and “Acked-By” statements
- **Release process**
  - New 2.6.x version released every 2-3 months by Linus
  - First two weeks to merge new features, leading to first -rc
  - Sequence of multiple release candidates to stabilize
- **System z integration**
  - Platform subsystem maintainer: Martin Schwidefsky, Heiko Carstens
  - git repository for System z features hosted on non-IBM site
    - Staging area for IBM and third-party System z patches
    - [Experimental System z features]



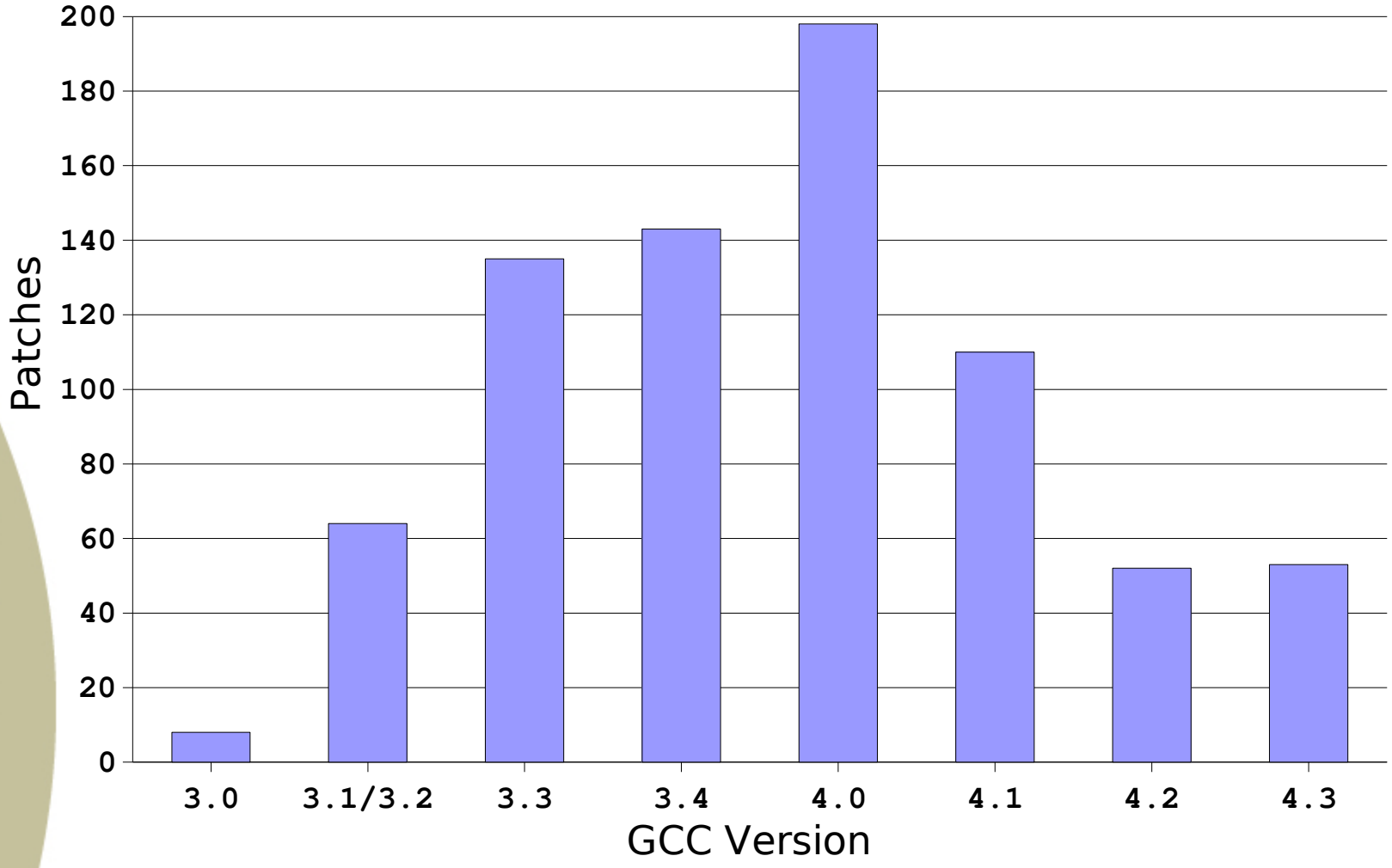
Linux kernel – System z contributions



Open Source development process GCC

- **Centralized development model**
  - Source code control tool: subversion
  - Master repository hosted by the Free Software Foundation
    - Read access to the general public, write access to maintainers
    - All copyright owned by / transferred to the FSF
  - GCC Steering Committee oversees the project
  - SC delegates design/development to maintainers
    - Global maintainers (ca. 12), Subsystem maintainers (ca. 130)
- **Release process**
  - New major release every 8-12 months
  - Stages: Major changes, minor changes, bugs, regressions
  - “Dot releases” every 2 months containing regression fixes only
- **System z integration**
  - Back-end maintainers:
    Ulrich Weigand, Hartmut Penner, Andreas Krebbel
  - Common code reload maintainer: Ulrich Weigand
  - GDB head maintainer: Ulrich Weigand

GNU Compiler Collection - System z contributions



How to get new features into distributions

- **Upstream feature (ideal case)**
  - Develop feature against mainline kernel, accepted in kernel version 2.6.x
  - Distribution release based on 2.6.x or later will usually include feature
- **Backport of upstream feature (usually acceptable)**
  - Code already accepted in some kernel version 2.6.x
  - Develop back-port against previous kernel release, provide on developerWorks and/or to distributor
  - Distribution release/update based on earlier kernel may add the feature as additional patch
- **Feature not upstream (difficult)**
  - Code provided only on developerWorks and/or to distributor, not yet accepted in any upstream kernel
  - Distributors are generally reluctant to add such features as additional patches due to maintenance concerns

Object-code only kernel modules

- **Issues**
  - OCO modules need to be re-built with every kernel change
  - Distributors reluctant to include OCO modules
- **Currently, we have no OCO module**
  - lcs: open source since 2002-03-04, upstream in 2.4.x
  - z90crypt: open source since 2002-07-31, upstream in 2.4.x
  - qdio: open source since 2002-09-13, upstream in 2.4.x
  - qeth: open source since 2003-06-30, upstream in 2.4.x
  - tape_3590: open source since 2006-03-28, upstream in 2.6.17
- **Future strategy: No more OCO modules!**

Kernel news - Linux version 2.6.20 (2007-02-04)

- Kernel Virtual Machine (KVM)
- Relocatable kernel images (i386)
- Asynchronous SCSI scanning
- Multithreaded USB probing
- I/O Accounting
- Relative atime support
- Bus event notifications
- ...
- [tons of architecture and driver updates]

Kernel news - Linux version 2.6.21 (2007-04-25)

- Virtual Machine Interface (VMI)
- KVM updates
- Dynticks and Clockevents
- ALSA System on Chip (ASoC)
- Dynamic kernel command-line
- Optional ZONE_DMA
- GPIO API
- ...
- [tons of architecture and driver updates]

Kernel news - Linux version 2.6.22 (2007-07-08)

- SLUB in kernel memory allocator
- New Wireless stack
- New FireWire stack
- Signal/timer events through file descriptors
- Blackfin architecture
- Unsorted Block Images (UBI)
- Secure RxRPC sockets
- Process footprint measurement facility
- ...
- [tons of architecture and driver updates]

Kernel news - Linux version 2.6.23 (2007-10-09)

- Completely Fair Scheduler (CFS)
- On-demand read-ahead (readahead trashing x3)
- fallocate system call to preallocate space in a file system
- lguest and Xen
- Variable argument length (no more “arg list too long”)
- Movable Memory Zone
- UIO
- Use splice for sendfile
- XFS and ext4 improvements
- ...
- [tons of architecture and driver updates]

Kernel news - Linux version 2.6.24 (2008-01-24)

- CFS improvements: performance, fair group, guest time
- Tickless support for x86-64, PPC, UML, ARM, MIPS
- New wireless drivers and configuration interface
- Anti-fragmentation patches
- Per-device dirty memory thresholds
- PID and network namespaces
- Large Receive Offload (LRO)
- Task Control Groups
- Read-only bind mounts
- x86-32/64 arch unification
- [tons of architecture and driver updates]

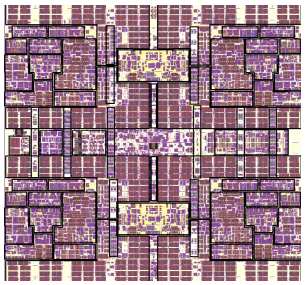
Kernel directions

- **Diversity: now 24 architectures (blackfin +1 unification -2)**
- **Bigger servers (large SGI machines, Mainframes, ...)**
- **Embedded systems, real-time (Cell-phones, PDAs)**
- **Appliances (network router, digital video recorder)**
- **Virtualization (KVM, paravirt, XEN), stronger than ever**

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

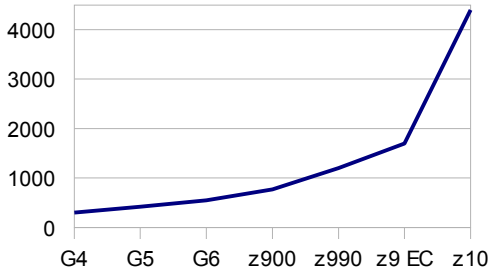
System z kernel features - CPU

- **New hardware support – System z10**
  - CPU node affinity (> 2.6.24, DW 2Q08)
  - Vertical CPU management (> 2.6.24, DW 2Q08)
  - STSI changes for capacity provisioning (> 2.6.24, DW 2Q08)
- **Dynamic configuration**
  - Standby CPU activation / deactivation (> 2.6.24, DW 2Q08)
- **User space tooling**
  - Dynamic CPU hotplug daemon (user space, DW 2Q08)
  - Support for processor degradation (in 2.6.22, DW 4Q07)



System z kernel features - Performance

- **New hardware support – System z10 processor**
  - Large page support (> 2.6.24, DW 2Q08)
- **DASD performance**
  - Hyper PAV enablement (> 2.6.24, DW 2Q08)
  - 4G FICON Express support for DASD (test only, no DW)
- **Network performance**
  - Support for skb scatter-gather (in 2.6.23, DW 4Q07)
- **FCP performance**
  - FCP performance data collection: I/O statistics (> 2.6.24, DW 4Q07)
  - FCP performance data collection: adapter statistics (> 2.6.24, DW 2Q08)
  - FCP performance enhancements: qdio rate improvement. (test only, no DW)
  - 4G FICON Express support for FCP (test only, no DW)



System z kernel features - Security

- **New hardware support – System z10 processor**
  - Support user-space AES 192/256, SHA 384/512 (> 2.6.24, DW 2Q08)
  - Support in-kernel AES 192/256, SHA 384/512 (> 2.6.24, DW 2Q08)
- **Generic algorithm fallback**
  - Use software implementation for key lengths not supported by hardware (> 2.6.24, no DW)
- **Crypto driver**
  - Support for long random numbers (> 2.6.24, DW 2Q08)
  - Capability for dynamic crypto device add (in 2.6.19, no DW)

System z kernel features - z/VM, networking

- **z/VM APPLDATA enhancements**
  - Linux process data in monitor APPLDATA (user space, DW 4Q07)
- **z/VM integration**
  - Unit record device driver (in 2.6.22, DW 4Q07)
  - IUCV access to z/VM services (user space netcat, no DW)
- **QETH network driver**
  - HiperSockets MAC layer routing (> 2.6.24, DW 2Q08)
  - QETH componentization (> 2.6.24, DW 2Q08)
  - OSA 2 Port per CHPID support (> 2.6.24, DW 2Q08)

System z kernel features - Usability and RAS

- **IPL**
  - IPL through IFCC / multipath IPL (s390-tools, DW 2Q08)
  - Shutdown actions interface (> 2.6.24, DW 2Q08)
  - Linux system loader (user space, DW 2Q08)
- **System dump**
  - Intuitive dump device configuration (distro, no DW)
  - Cleanup SCSI dumper for upstream integration (in 2.6.23, no DW)
- **DASD sense data reporting**
  - SIM/MIM handling for ECKD DASD (> 2.6.24, DW 2Q08)
- **Channel subsystem**
  - Dynamic CHPID reconfiguration via SCLP (in 2.6.22, DW 4Q07)

Compiler – Common features

- **General optimizer improvements**
  - SSA-based common optimization infrastructure (GCC 4.0)
  - Inter-procedural optimization infrastructure (GCC 4.1)
  - New data flow analyzer framework (GCC 4.3)
- **Languages and language features**
  - Fortran 95 front end (GCC 4.0)
  - Decimal Floating Point support (GCC 4.2)
  - OpenMP support for C/C++/Fortran (GCC 4.2)
- **Other improvements**
  - Stack Protector feature (GCC 4.1)
  - Builtins for atomic operations (GCC 4.1)

Compiler – System z machine support

- **System z10 processor support (> GCC 4.3)**
  - Exploit instruction new to z10
  - Selected via `-march=z10-ec/-mtune=z10-ec`
- **System z9 109 processor support (GCC 4.1)**
  - Exploit instructions provided by the e*xtended immediate facility*
  - Selected via `-march=z9-109/-mtune=z9-109`
- **Support for 128-bit IEEE “long double” data type (GCC 4.1)**
  - Provide extended range of floating point exponent and mantissa
  - Selected via `-mlong-double-128`
- **Support for atomic builtins**
  - __builtin_compare_and_swap and friends
- **Decimal floating point support (GCC 4.3)**
  - For newer machines with hardware DFP support
  - Selected via `-march=z9-ec, -mhard-dfp/-mnohard-dfp`

Compiler - System z features

- **Software dfp support (GCC 4.2)**
  - For older machines without hardware DFP support
- **Kernel stack overflow avoidance/detection (GCC 4.0)**
  - Compile time detection:
    `-mwarn-framesize` / `-mwarn-dynamicstack`
  - Run-time detection:
    `-mstack-size` / `-mstack-guard`
  - Stack frame size reduction:
    `-mpacked-stack`
- **GCC support for the z/TPF OS (GCC 4.0/4.1)**
  - z/TPF uses Linux / GCC as cross-build environment
  - New target `s390x-ibm-tpf`

Compiler - System z performance

- **Compiler back-end improvements**
  - Improved condition code handling (GCC 4.0)
  - Improved function prologue/epilogue scheduling (GCC 4.0)
  - Improved use of memory-to-memory instructions (GCC 4.0)
  - Added sibling call support (GCC 4.0)
  - Enhanced use of string instructions (SRST, MVST, ...) (GCC 4.1)
  - More precise register tracking (r13, r6, ...) (GCC 4.1)
  - Use LOAD ZERO (GCC 4.1)
  - ICM/STCM, BRCT, vararg enhancements (GCC 4.1)
  - More small optimizations / improvements (GCC 4.3)
- **Performance enhancements on z9**
  - Industry-standard integer performance benchmark
  - 8% comparing GCC 3.4 and GCC 4.1 on System z
  - 5.9% comparing GCC 4.1 and GCC 4.2 (-march= z990 vs z9-109)
  - 0.5% comparing GCC 4.2 and GCC 4.3 (-march= z990 vs z9-109)

Outlook

- New hardware exploitation
- Enhanced Linux – z/VM synergy
- Basic support for KVM virtualization
- Keep current with open source

Trademarks

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: AS/400, DBE, e-business logo, 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

The following are trademarks or registered trademarks of other companies

Lotus, Notes, and Domino are trademarks or registered trademarks of Lotus Development Corporation
Java and all Java-related trademarks and logos are trademarks of Sun Microsystems, Inc., in the United States and other countries
LINUX is a registered trademark of Linux Torvalds
UNIX is a registered trademark of The Open Group in the United States and other countries.
Microsoft, Windows and Windows NT are registered trademarks of Microsoft Corporation.
SET and Secure Electronic Transaction are trademarks owned by SET Secure Electronic Transaction LLC.
Intel is a registered trademark of Intel Corporation
* 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.

References in this document to IBM products or services do not imply that IBM intends to make them available in every country.

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 time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.