

# Linux on System z – Update

## Current & Future Technology



Frank Heimes – Senior IT Architect – <[frank.heimes@de.ibm.com](mailto:frank.heimes@de.ibm.com)>

# Linux Kernel Source

- The Linux kernel is **highly portable** (even if initially not developed to be) and supports the following computer architectures:
- One **single set of source code** available from “The Linux Kernel Archives”:  
<http://kernel.org> – <http://git.kernel.org>
- Linux for different architectures is all **build from the same source!**
- All **architecture dependent code** is included:
  - Linux Kernel Sources:  
/linux/arch/s390  
/linux/drivers/s390
  - Linux Kernel Headers (include/asm)
- Linux is always an **ASCII** operating system, even Linux on System z, running on the Extended Binary Coded Decimal Interchange Code (EBCDIC) encoded z/Architecture



[linux/kernel/git/torvalds/linux-2.6.git] / arch /

- Alpha architecture (alpha)
- Analog Devices (blackfin)
- **ARM architecture (arm)**
- Atmel AVR32 (avr32)
- Axis Communications' ETRAX CRIS (cris)
- Freescale's (formerly Motorola's) 68k (m68k)
- Fujitsu FR-V (frv)
- HP PA-RISC (parisc)
- H8 Renesas Technology, formerly Hitachi (H8300)
- **IBM System/390 (31-bit), z/Architecture (64-bit) (s390)**
- Intel IA-64 Itanium, Itanium II (ia64)
- **x86 architecture: 80386 ... x86\_64 (x86)**
- M32R from Mitsubishi (m32r)
- Microblaze from Xilinx (microblaze)
- MIPS architecture (mips)
- MN103 from Panasonic Corporation (mn10300)
- OpenRISC (openrisc)
- **IBM POWER architecture (powerpc)**
- **PowerPC architecture (powerpc)**
- **IBM's Cell (powerpc)**
- SPARC, UltraSPARC (sparc)
- SuperH (sh)
- S+core (score)
- Tiler (tile)
- Xtensa from Tensilica (xtensa)
- UniCore32 (unicore32)
- User Mode Linux, UML (um)

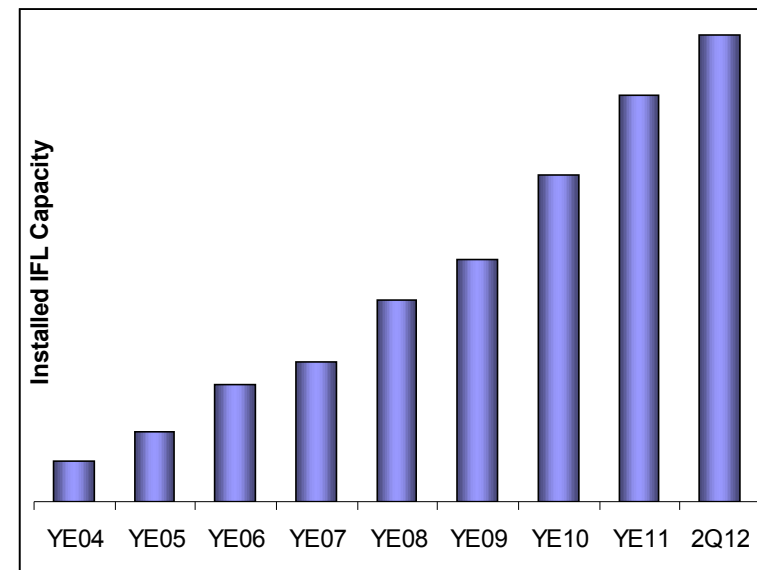


# Linux on IBM System z in 2Q2012

Installed Linux MIPS at 39% CAGR\*

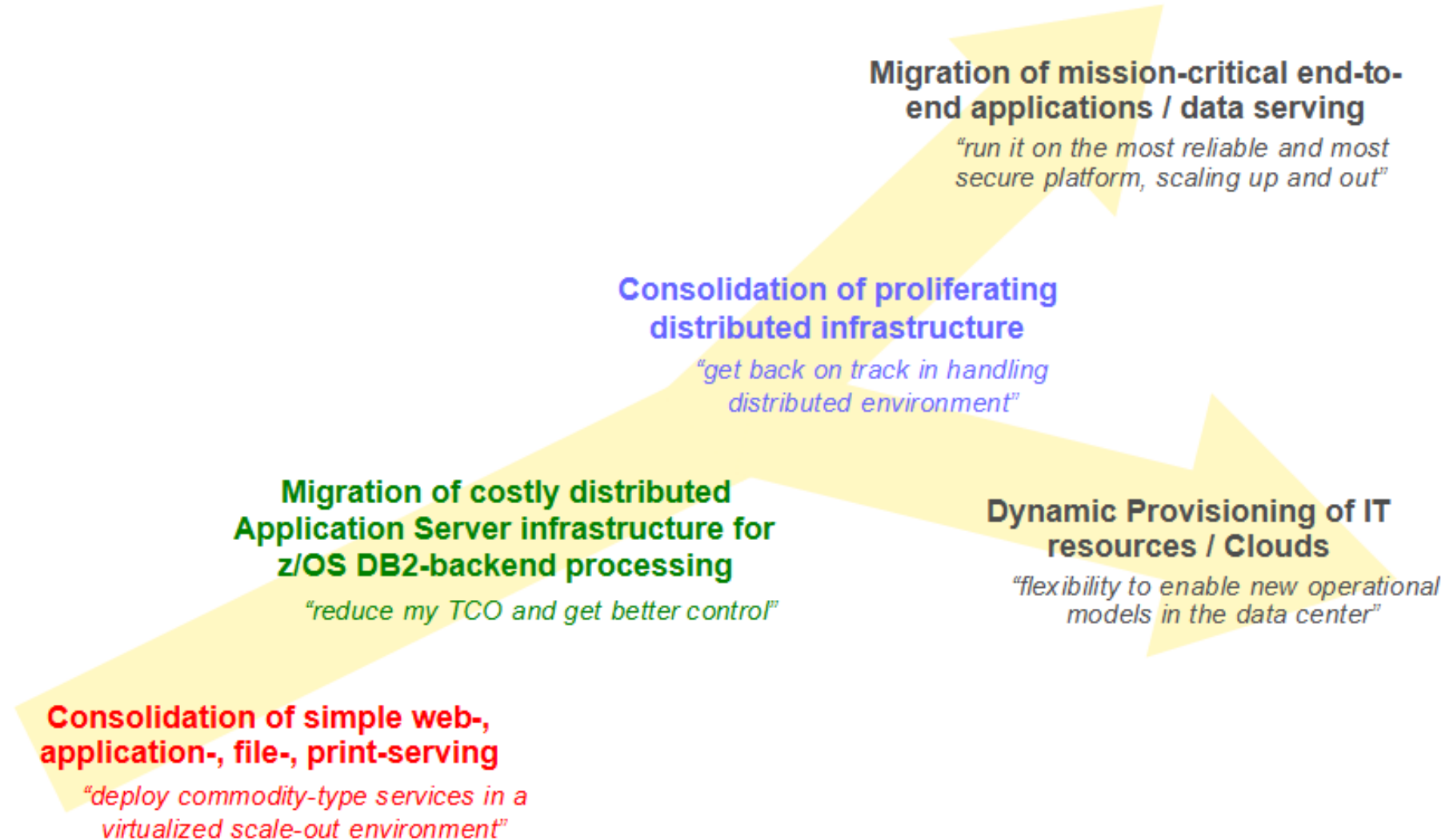
- 21% of Total installed MIPS run Linux as of 2Q12
- Installed IFL MIPS increased 26% YoY (2Q12)
- 36% of System z Customers have IFL's installed as of 2Q12
- 67 of the top 100 System z Customers are running Linux on the mainframe as of 2Q12
- 31% of all System z servers have IFLs

Installed Capacity Over Time



\* Based on YE 2004 to 2H 2012

# Linux on System z Deployment Trends

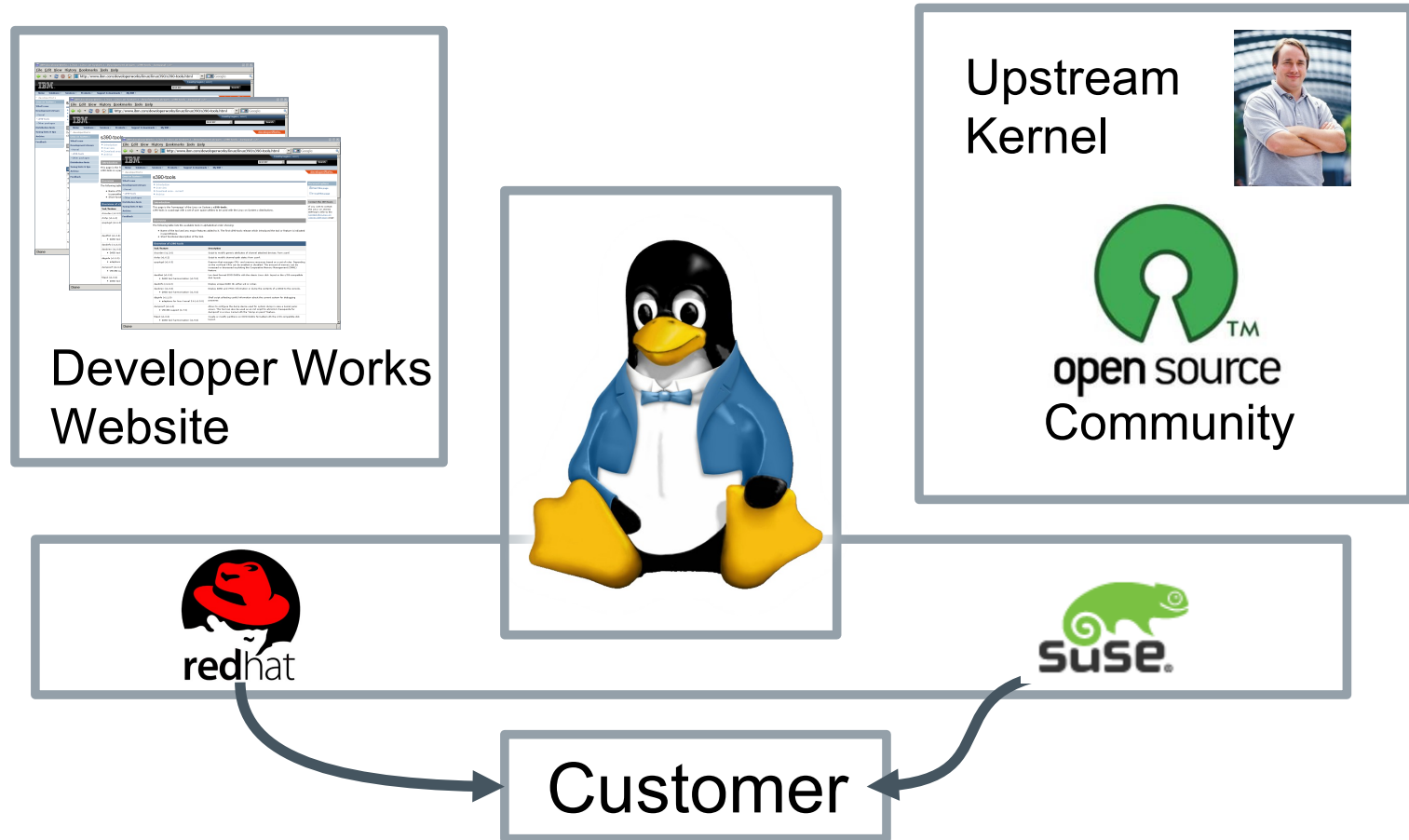


# How Linux on System z is developed

How does the “community” work?

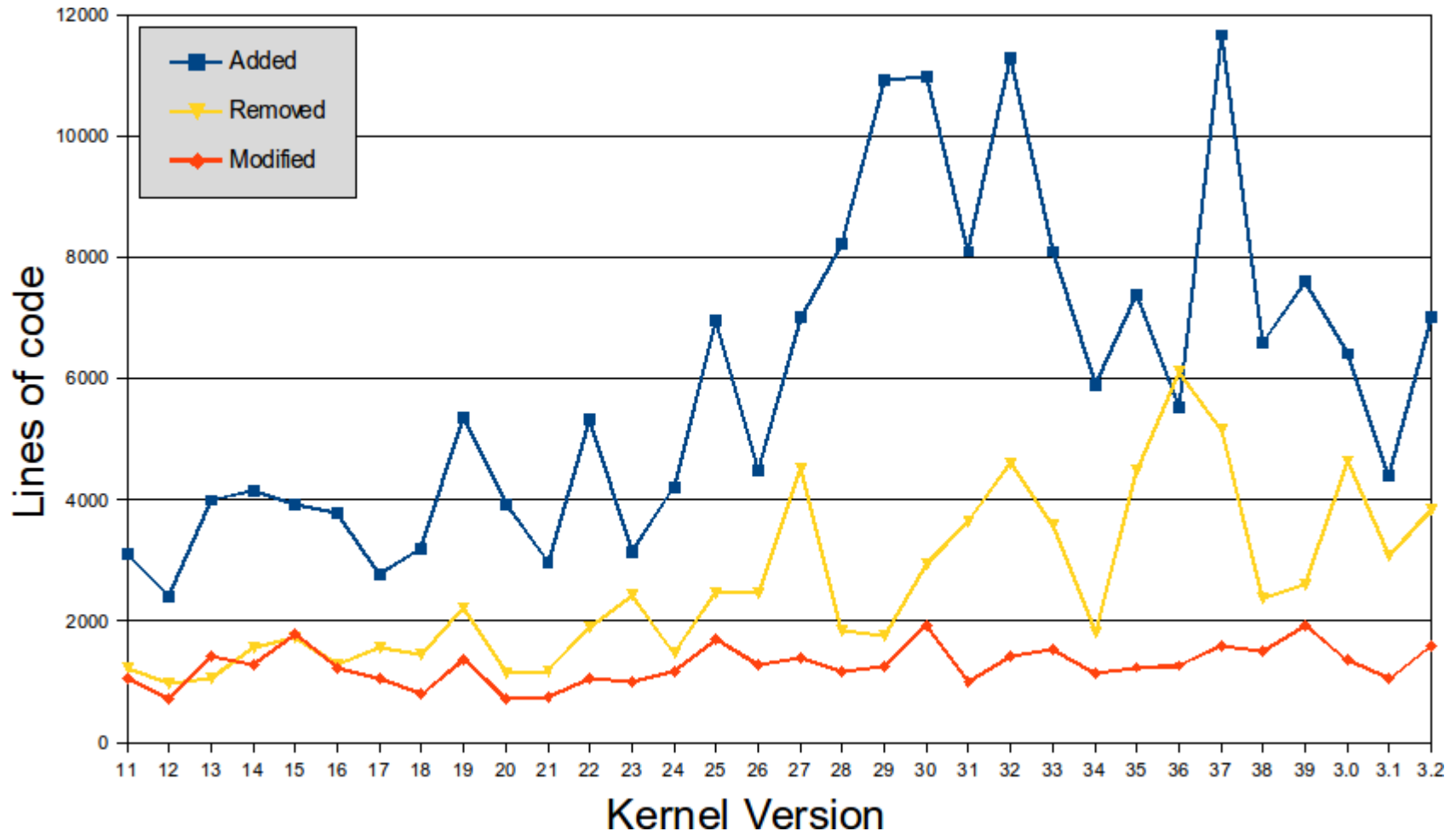
# IBM Linux on System z Development

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



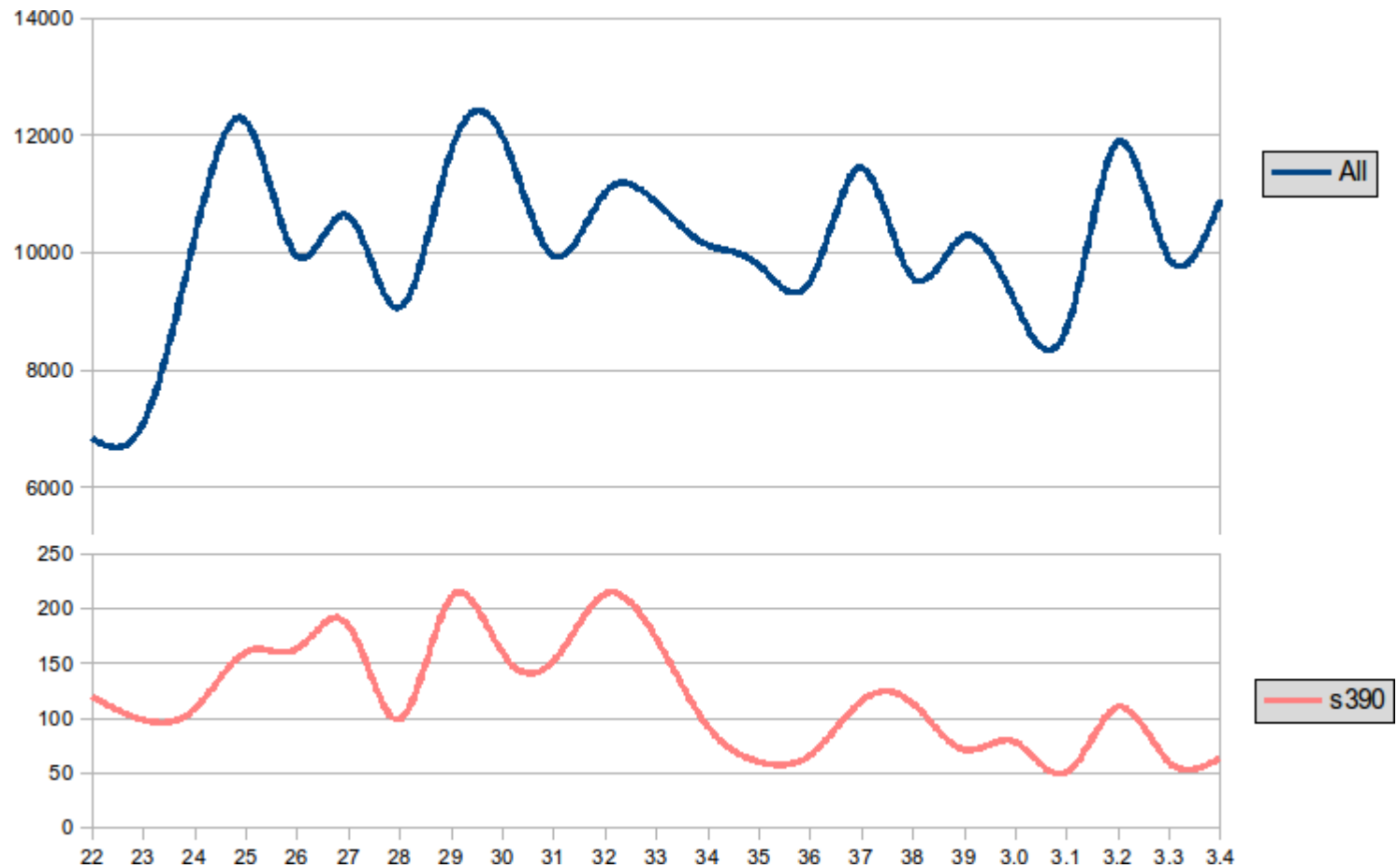
## Linux kernel development: rate of change

Average for the last 7 years (without renames): 102 days per release, 5897 lines added, 2586 lines removed and 1221 lines modified **per day**



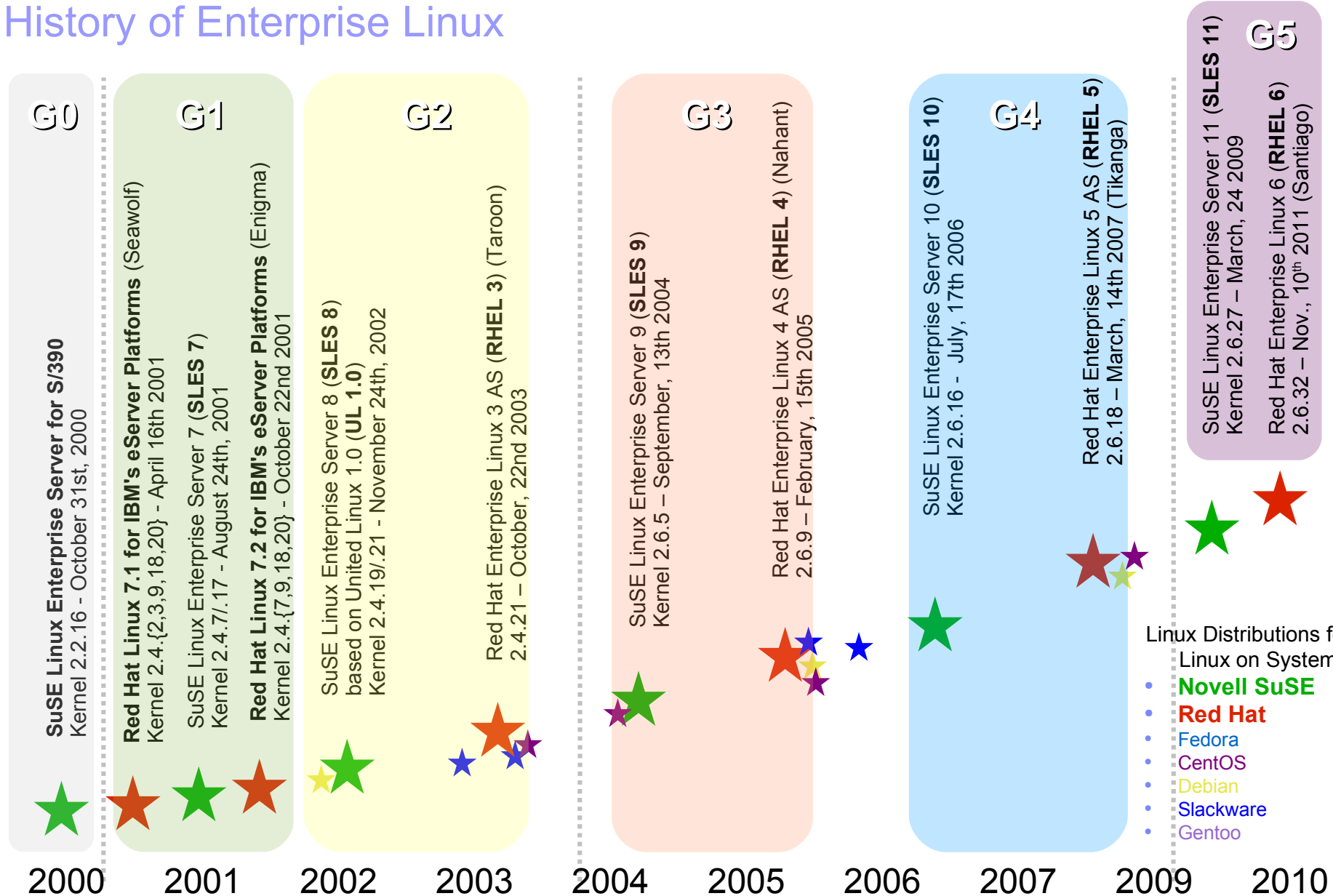
## Linux kernel development: System z contributions

Changesets per 2.6.x / 3.x kernel release





# History of Enterprise Linux



# Enterprise Linux Distributions – Tested & Supported

The table below shows IBM tested Linux environments. IBM remote technical support for these environments is provided when you obtain a [Support Line](#) contract.

You may also find support for these environments by contracting with a third party provider.

Hardware Platform and Operating System Software Compatibility (64-bit)					
Generation	n – 4	n – 3	n – 2	n – 1	n
Release	zSeries	System z9	System z10	z196/z114	zEC12
★ SLES 9 (*)	✓	✓	✓	✓ (6)	✗
★ SLES 10	✓	✓	✓	✓	✓ (4)
★ SLES 11	✗	✓	✓	✓	✓ (3)
★ RHEL 4 (*)	✓	✓	✓	✓ (5)	✗
★ RHEL 5	✓	✓	✓	✓	✓ (2)
★ RHEL 6	✗	✓	✓	✓	✓ (1)

✓ Indicates that the Linux distribution (version) has been tested by IBM on the hardware platform, will run on the system, and is an IBM supported environment. Updates or service packs applied to the distribution are also supported.

(1) Recommended level: RHEL 6.3

(2) Recommended level: RHEL 5.8

(3) Recommended level: SLES 11 SP2

(4) Recommended level: SLES 10 SP4 with latest maintenance updates

(5) RHEL 4.8 only. Some functions have changed or are not available with the z196, e.g. the Dual-port OSA cards support to name one of several. Please check with your service provider regarding the end of service.

(6) SLES 9 SP4 with latest maintenance updates only. Some functions have changed or are not available with the z196, e.g. the Dual-port OSA cards support to name one of several. Please check with your service provider regarding the end of service.

✗ Indicates that the distribution is not supported by IBM on this server

(\*) The distribution is out of service, extended support is required.

To retrieve interoperability support information for Enterprise Storage products when used in a supported host server environment see the **System Storage Interoperation Center**: <http://www.ibm.com/systems/support/storage/config/ssic/index.jsp>

## zEC12 – Toleration Tests and Certifications

- Certifications tests executed and published by SUSE and Red Hat:
  - SLES 10 SP4:  
<https://www.suse.com/nbswebapp/yesBulletin.jsp?BulletinNumber=138473>
  - SLES11 SP2:  
<https://www.suse.com/nbswebapp/yesBulletin.jsp?bulletinNumber=138472>
  - RHEL 5 Update 8:  
<https://hardware.redhat.com/show.cgi?id=852529>
  - RHEL 6 Update 3:  
<https://hardware.redhat.com/show.cgi?id=852525>
- Be sure to install the z/VM PTFs for zEC12 before you upgrade a system:  
[http://www-01.ibm.com/support/docview.wss?uid=isg1\\_2827DEVICE\\_2827-ZVM](http://www-01.ibm.com/support/docview.wss?uid=isg1_2827DEVICE_2827-ZVM)
- It's possible to upgrade the hardware without changing the Linux stack!
  - No compatibility package needed!
- Matrix of Tested & Supported Linux Distributions:  
[http://www.ibm.com/systems/z/os/linux/support\\_testedplatforms.html](http://www.ibm.com/systems/z/os/linux/support_testedplatforms.html)
- **RHEL4 and SLES9 are only supported with special agreements**
  - **plan for migration now!**

# Current Linux on System z Technology

*Features & Functionality contained in the Novell  
& Red Hat Distributions*

## Selected New Features for Linux on System z

- Linux Open Source Development is happening “upstream”
- Features are integrated later in distributions
  - Customer requests can help speed up this process
- Complete listing for SLES and RHEL is available here  
(by Martin Schwidefsky):  
<http://www-05.ibm.com/de/events/linux-on-z/pdf/whats-new-mar13.pdf>
- Current and Future Features SUSE Linux Enterprise Server for System z:  
(by Markus Kraft)  
[https://share.confex.com/share/118/webprogram/Handout/Session10738/SLES\\_update\\_10738.pdf](https://share.confex.com/share/118/webprogram/Handout/Session10738/SLES_update_10738.pdf)

## System z kernel features – Core

- **Improved QDIO performance statistics (kernel 2.6.33)**



- Converts global statistics to per-device statistics and adds new counter for the input queue full condition

- **Breaking event address for user space programs (kernel 2.6.35)**

- Remember the last break in the sequential flow of instructions
- Valuable aid in the analysis of wild branches



- **z196 enhanced node affinity support (kernel 2.6.37)**

- Allows the Linux scheduler to optimize its decisions based on the z196 topology



- **Performance indicator bytes (kernel 2.6.37)**

- Display capacity adjustment indicator introduced with z196 via /proc/sysinfo



## CMSFS user space file system support



6.1



11.2

- Allows to mount a z/VM minidisk to a Linux mount point
- z/VM minidisk needs to be in the enhanced disk format (EDF)
- The cmsfs fuse file system transparently integrates the files on the minidisk into the Linux VFS, no special command required

```
# cmsfs-fuse /dev/dasde /mnt/cms
# ls -la /mnt/fuse/PROFILE.EXEC
-r--r----- 1 root root 3360 Jun 26 2009 /mnt/fuse/PROFILE.EXEC
```

- By default no conversion is performed
  - Mount with '-t' to get automatic EBCDIC to ASCII conversion

```
# cmsfs-fuse -t /dev/dasde /mnt/cms
```

- Write support is work in progress, almost completed
  - use “vi” to edit PROFILE.EXEC anyone ?
- Use fusermount to unmount the file system again

```
# fusermount -u /mnt/cms
```

## Deliver z/VM CP special messages as uevent

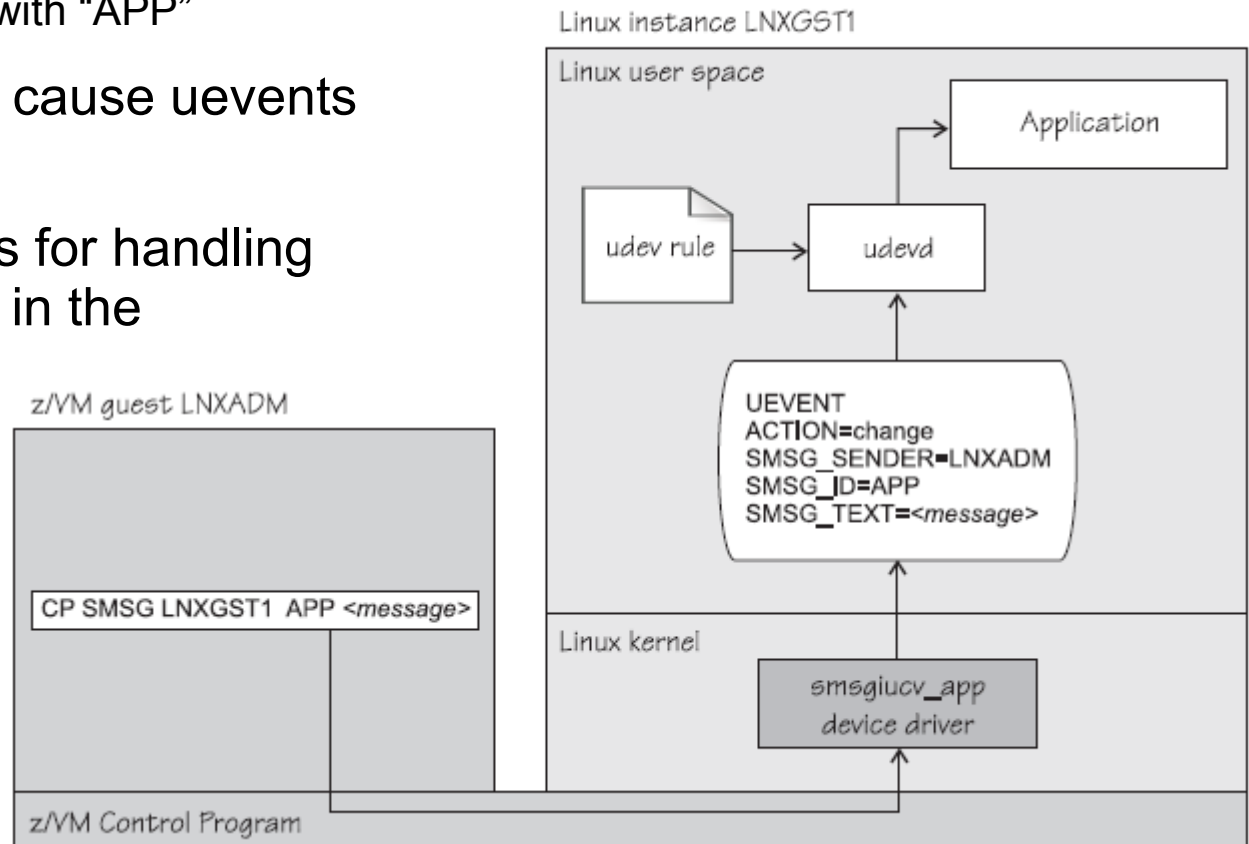


6.1



11.2

- Allows to forward SMSG messages to user space programs
  - Message needs to start with “APP”
- The special messages cause uevents to be generated
- See “Writing udev rules for handling CP special messages” in the Device Drivers Book





## System z kernel features – Usability / RAS

- **Dump on panic – prevent reipl loop (s390-tools 1.8.4)**

- Delay arming of automatic reipl after dump.
- Avoids dumps loops where the restarted system crashes immediately.



- **Add support for makedumpfile tool (kernel 2.6.34, s390-tools 1.9.0)**

- Convert Linux dumps to the ELF file format
- Use the makedumpfile tool to remove user data from the dump.
- Multi-volume dump will be removed.



- **Address space randomization (kernel 2.6.38)**

- Enable flexible mmap layout for 64 bit to randomize start address for the runtime stack and the mmap area



- **Get CPC name (kernel 2.6.39)**

- Useful to identify a particular hardware system in a cluster (GDPS/xDR)
- The CPC name and the HMC network name are provided



## System z kernel features – Networking

- **Offload outbound checksumming (kernel 2.6.35)**



6.1



11.2

- Move calculation of checksum for non-TSO packets from the driver to the OSA network card

- **OSX/OSM CHPIDs for hybrid data network (kernel 2.6.35)**



6.1

10.4  
11.2

- The OSA cards for the zBX Blade Center Extension will have a new CHPID type
- Allows communication between zBX and Linux on System z

- **Toleration of optimized latency mode (kernel 2.6.35)**



11.2

- OSA devices in optimized latency mode can only serve a small number of stacks / users. Print a helpful error message if the user limit is reached.
- Linux does not exploit the optimized latency mode

- **NAPI support for QDIO and QETH (kernel 2.6.36)**



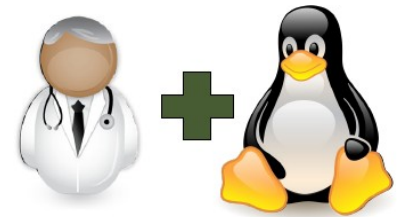
6.1



11.2

- Convert QETH to the NAPI interface, the “new” Linux networking API
- NAPI allows for transparent GRO (generic receive offload)

## LNxHC – Linux Health Checker



- The Linux Health Checker is a command line tool for Linux.
- Its purpose is to identify potential problems before they impact your system's availability or cause outages.
- It collects and compares the active Linux settings and system status for a system with the values provided by health-check authors or defined by you. It produces output in the form of detailed messages, which provide information about potential problems and the suggested actions to take.
- The Linux Health Checker will run on any Linux platform which meets the software requirements. It can be easily extended by writing new health check plug-ins.
- The Linux Health Checker is an open source project sponsored by IBM. It is released under the Eclipse Public License v1.0
  - Homepage: <http://lnxhc.sourceforge.net/>
  - IBM Linux Information Center:  
[http://publib.boulder.ibm.com/infocenter/lnxinfo/v3r0m0/topic/com.ibm.trouble.doc/lnxhc/lnxhc\\_linuxonz.htm](http://publib.boulder.ibm.com/infocenter/lnxinfo/v3r0m0/topic/com.ibm.trouble.doc/lnxhc/lnxhc_linuxonz.htm)
  - Live Virtual Class (LVC): <http://www.vm.ibm.com/education/lvc/>
  - User's Guide: <http://lnxhc.sourceforge.net/documentation.html>

## hyptop: Display Hypervisor Utilization

- The hyptop command is a top-like tool that displays a dynamic real-time view of the hypervisor environment
  - It works with both the z/VM and LPAR hypervisor
  - Depending on the available data it can display information about CPU and memory
  - running LPARs or z/VM guest operating systems
- The following is required to run hyptop
  - The debugfs file system must be mounted
  - The hyptop user must have read permission for the required debugfs files:
    - z/VM: <debugfs mount point>/s390\_hypfs/diag\_2fc
    - LPAR: <debugfs mount point>/s390\_hypfs/diag\_204
  - To monitor all LPARs or z/VM guests your instance requires additional privileges
    - For z/VM: The user ID requires privilege class B
    - For LPAR: The global performance data control box in the LPAR activation profile needs to be selected

# hyptop: Display Hypervisor Utilization – z/VM

## Example of z/VM utilization data

10:11:56 CPU-T: UN(16) ?							
system	#cpu	cpu	Cpu+	online	memuse	memmax	wcur
T6360003	6	<u>506.92</u>	3404:17	44:20:53	7.99	8.00	100
T6360017	2	<u>199.58</u>	8:37	29:23:50	0.75	0.75	100
T6360004	6	<u>99.84</u>	989:37	62:00:00	1.33	2.00	100
T6360005	2	<u>0.77</u>	0:16	5:23:06	0.55	2.00	100
T6360015	4	<u>0.15</u>	9:42	18:23:04	0.34	0.75	100
T6360035	2	<u>0.11</u>	0:26	7:18:15	0.77	1.00	100
T6360027	2	<u>0.07</u>	2:53	62:21:46	0.75	0.75	100
T6360049	2	<u>0.06</u>	1:27	61:17:35	0.65	1.00	100
T6360010	6	<u>0.06</u>	5:55	61:20:56	0.83	1.00	100
T6360021	2	<u>0.06</u>	1:04	48:19:08	0.34	4.00	100
T6360048	2	<u>0.04</u>	0:27	49:00:51	0.29	1.00	100
T6360016	2	<u>0.04</u>	6:09	34:19:37	0.30	0.75	100
T6360008	2	<u>0.04</u>	3:49	47:23:10	0.35	0.75	100
T6360006	2	<u>0.03</u>	0:57	25:20:37	0.54	1.00	100
NSLCF1	1	<u>0.01</u>	0:02	62:21:46	0.03	0.25	100
VTAM	1	<u>0.00</u>	0:01	62:21:46	0.01	0.03	100
T6360023	2	<u>0.00</u>	0:04	6:21:20	0.46	0.75	100
PERFSVM	1	<u>0.00</u>	2:12	7:18:04	0.05	0.06	0
AUTOVM	1	<u>0.00</u>	0:03	62:21:46	0.00	0.03	100
FTPSERVE	1	<u>0.00</u>	0:00	62:21:47	0.01	0.03	100
TCPIP	1	<u>0.00</u>	0:01	62:21:47	0.01	0.12	3000
DATAMOVE	1	<u>0.00</u>	0:06	62:21:47	0.00	0.03	100
VMSEVRU	1	<u>0.00</u>	0:00	62:21:47	0.00	0.03	1500
OPERSVMP	1	<u>0.00</u>	0:00	62:21:47	0.00	0.03	100

# hyptop: Display Hypervisor Utilization – LPAR

## Example of single LPAR utilization data

```
10:16:59 H05LP30 CPU-T: IFL(18) CP(3) UN(2) ?
=help
```

(#)	(str)	(%)	(%)	(vis)
0	IFL	29.34	0.72	#####
1	IFL	28.17	0.70	#####
2	IFL	32.86	0.74	#####
3	IFL	31.29	0.75	#####
4	IFL	32.86	0.72	#####
5	IFL	30.94	0.68	#####
6	IFL	0.00	0.00	
7	IFL	0.00	0.00	
8	IFL	0.00	0.00	

## Out of Order Detail

**Out of order** yields significant performance benefit through

### 1) Re-ordering instruction execution

- Instructions stall in a pipeline because they are waiting for results from a previous instruction or the execution resource they require is busy
- In an in-order core, this stalled instruction stalls all later instructions in the code stream
- In an out-of-order core, later instructions are allowed to execute ahead of the stalled instruction

### 2) Re-ordering storage accesses

- Instructions which access storage can stall because they are waiting on results needed to compute storage address
- In an in-order core, later instructions are stalled
- In an out-of-order core, later storage-accessing instructions which can compute their storage address are allowed to execute

### 3) Hiding storage access latency

- Many instructions access data from storage
- Storage accesses can miss the L1 and require 10 to 500 additional cycles to retrieve the storage data
- In an in-order core, later instructions in the code stream are stalled
- In an out-of-order core, later instructions which are not dependent on this storage data are allowed to execute

## System z toolchain (1)

### ■ zEnterprise 196 / z196 (gcc 4.6)

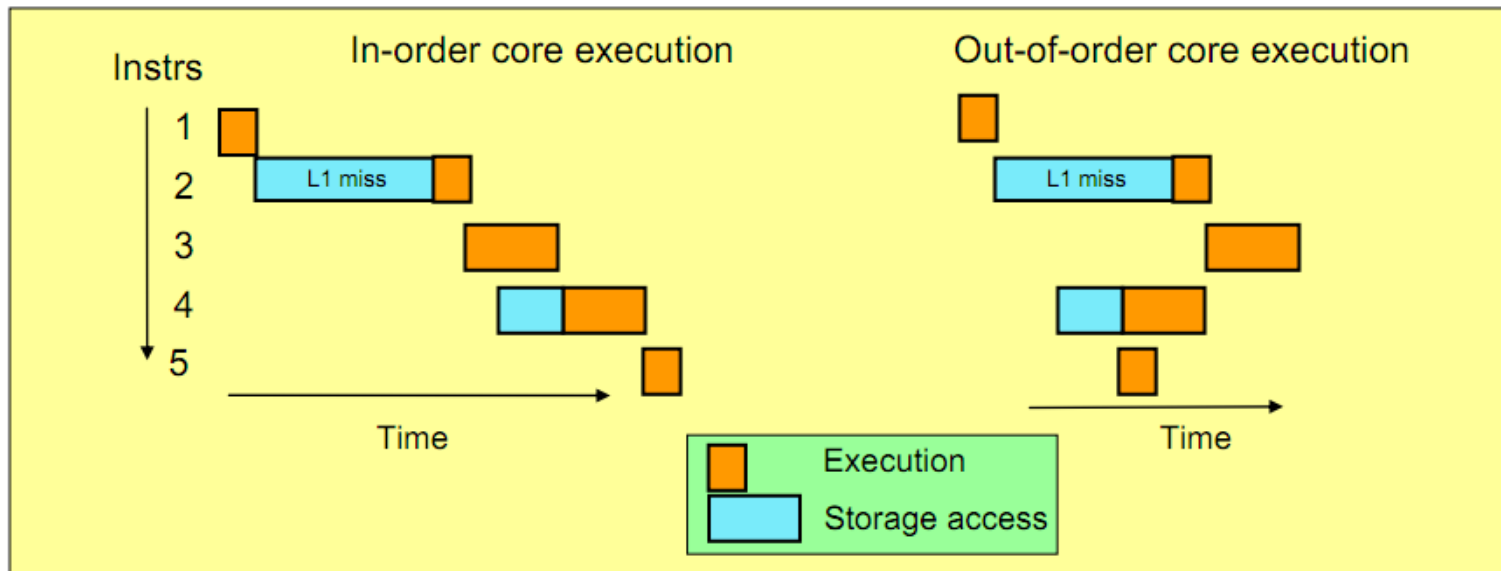
- Exploitation of new z196 processor instructions
- Optimized alignment of code (out-of-order pipeline architecture, conditional load/store instructions, new 3 register operand instructions, new atomic instructions, etc.)
- Use option `-march=z196` to utilize the new instructions added with z196
- Use `-mtune=z196` to schedule the instruction appropriate for the new out-of-order pipeline of z196
- Re-compiled code/apps get further performance gains through 110+ new instructions
- <http://gcc.gnu.org/gcc-4.6/changes.html>



6.1



11.2

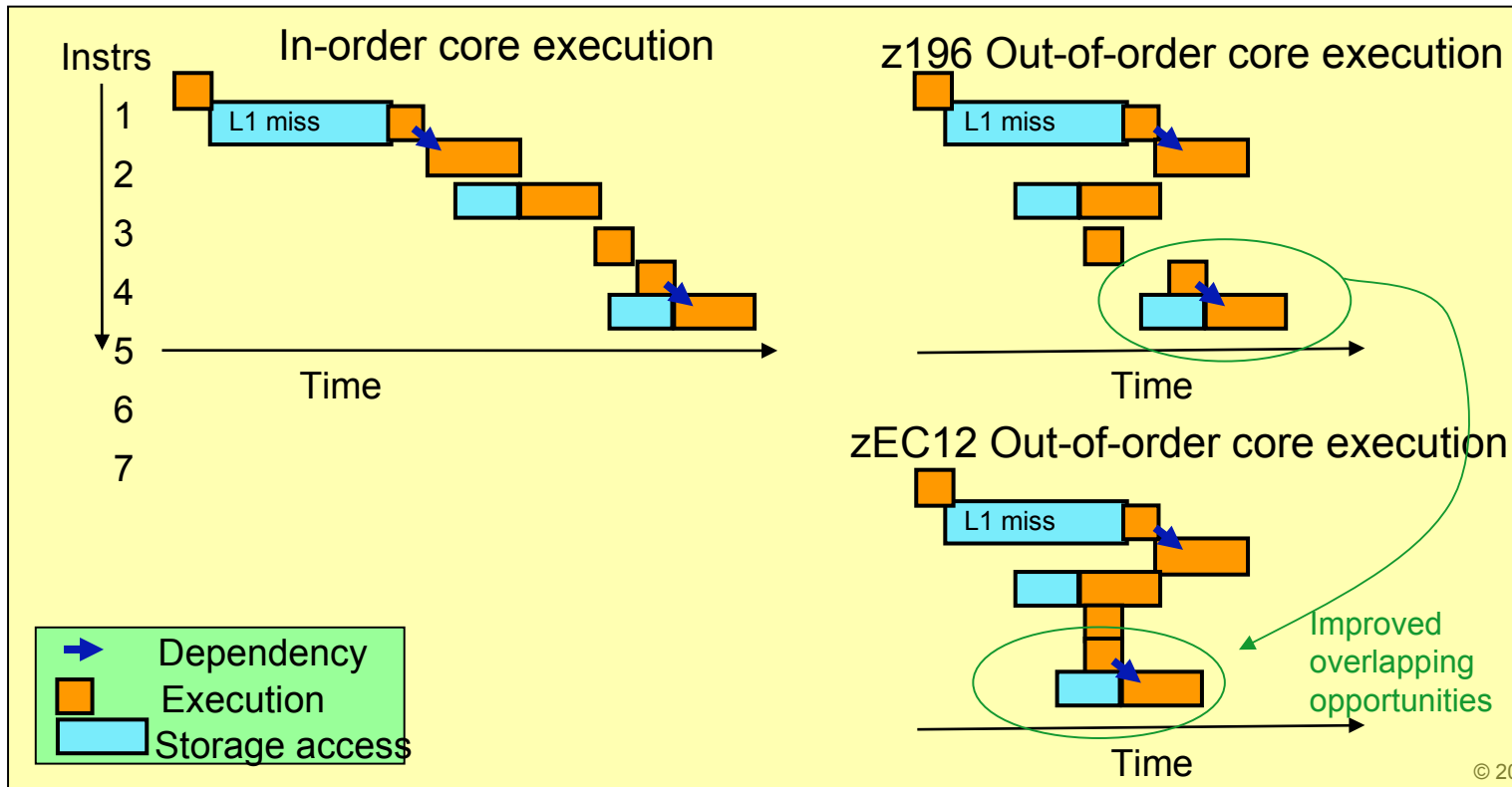




## System z toolchain (2)

### ■ zEC12 exploitation (gcc 4.x)

- Can decode up to 3 instructions and initiate the execution of up to 7 instructions - per cycle
- Use option `-march=zEC12` to utilize the new instructions added with zEC12
- Use `-mtune=zEC12` to schedule the instruction appropriate for the improved out-of-order pipeline of zEC12
- <http://gcc.gnu.org/ml/gcc-patches/2012-10/msg00943.html>
- <http://sourceware.org/ml/binutils/2012-10/msg00040.html>



## 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*

## System z kernel features – Storage FICON

- **DASD sanity check to detect path connection errors (kernel 3.3)**
  - An incorrect physical connection between host and storage server which is not detected by hardware or microcode can lead to data corruption
  - Add a check in the DASD driver to make sure that each available channel path leads to the same storage server
- **Extended DASD statistics (kernel 3.1)**
  - Add detailed per-device debugging of DASD I/Os via debugfs
  - Useful to analyze problems in particular for PAV and HPF

## Extended DASD statistics

- Start data collection

```
# dasdstat -e dasda 0.0.1234
```

- Reset statistics counters

```
# dasdstat -r dasda
```

- Read summary statistics

```
# dasdstat
statistics data for statistic: 0.0.6527
start time of data collection: Fri Feb 24 16:00:19 CET 2012

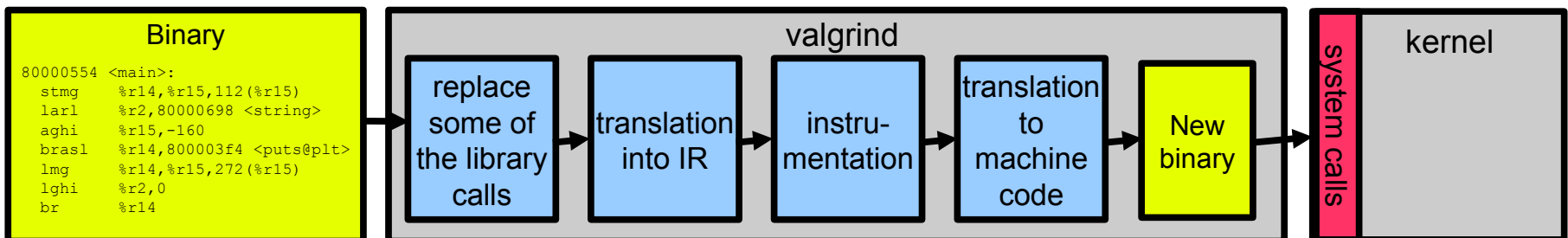
1472 dasd I/O requests
with 14896 sectors(512B each)
0 requests used a PAV alias device
0 requests used HPF
  <4    8   16   32   64  128  256  512  1k   2k   4k   8k  16k  32k  64k 128k
  256  512  1M  2M  4M  8M  16M  32M  64M 128M 256M 512M 1G   2G   4G  >4G
Histogram of sizes (512B secs)
    0    0 1441    8   13    5    2    2    0    1    0    0    0    0    0    0
    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
Histogram of I/O times (microseconds)
    0    0    0    0    0    0    1 1160   49   52   61  142    7    0    0    0
    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
```

## System z toolchain

- **Oprofile support for hardware sampling introduced with z10 (2.6.39)**
  - Provide CPU measurement data to applications for performance tuning
  - Based on hardware counters and samples built into the CPU
  - Use oprofile to communicate the information to user space programs
- **Oprofile z196 hardware customer mode sampling (kernel 3.3)**
  - Extend the hardware sampling to support z196.
- **Valgrind System z support**
  - Valgrind is a generic framework for creating dynamic analysis tools and can be used for memory debugging, memory leak detection and profiling (e.g. cachegrind)
  - Valgrind is in essence a virtual machine using just-in-time (JIT) compilation techniques
  - Memory debugging is available with Valgrind version 3.7.0

## Valgrind System z support

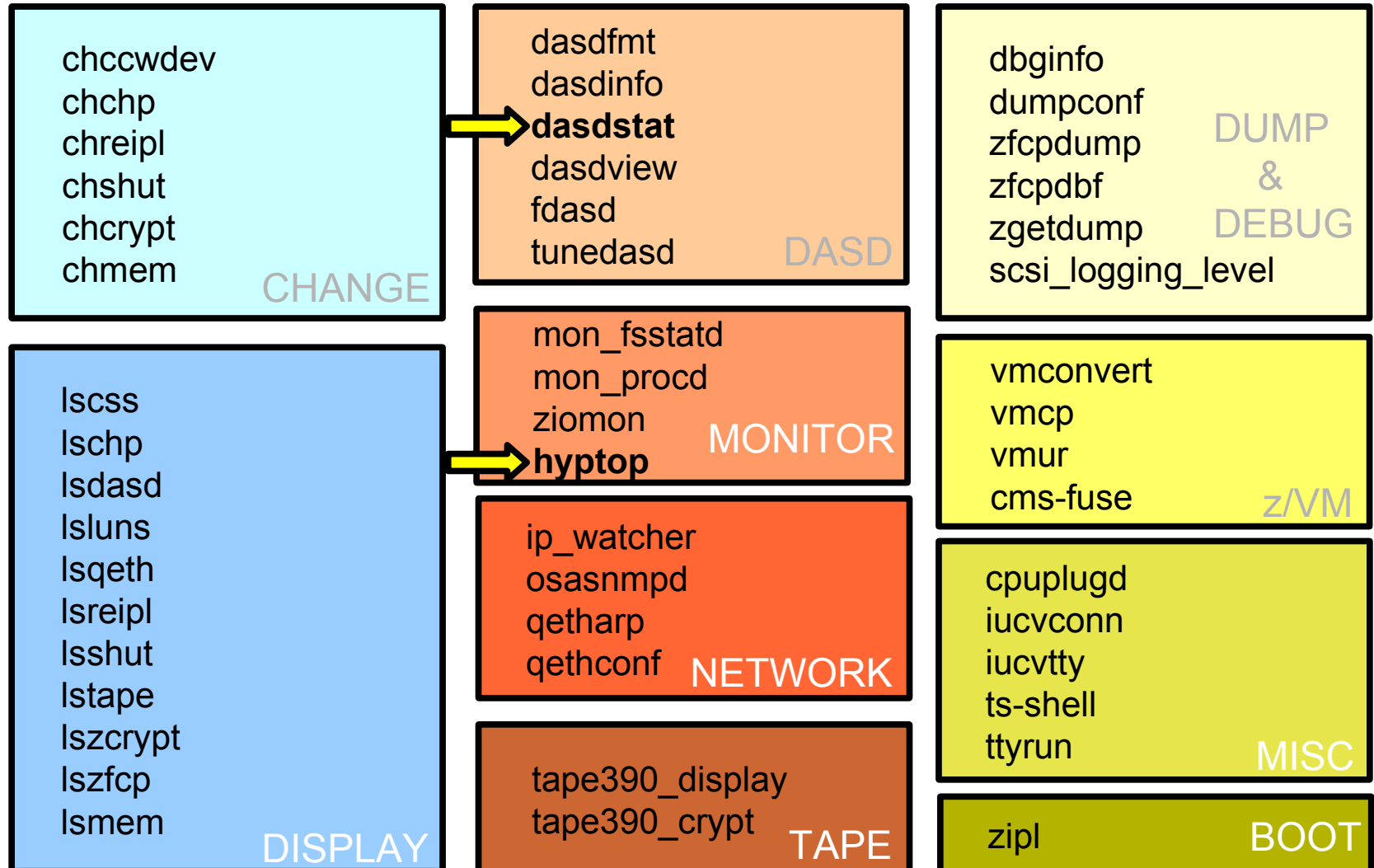
- **Memcheck**: checks for uninitialized memory, unallowed memory read/writes after it has been free'd or off the end of malloc'd blocks, memory leaks
- **Massif**: heap profiler
- **Helgrind / DRD**: detecting race conditions in multithreaded code
- **Cachegrind**: cache profiler
- **Callgrind**: extension to Cachegrind for callgraphs
- **exp-sgcheck** (former exp-ptrcheck): find stack and global array overrun errors (beyond Memcheck), etc.
- **valgrind --tool=memcheck [--leak-check=full] [--track-origins] <program>**
  - Detects if your program accesses memory it shouldn't
  - Detects dangerous uses of uninitialized values on a per-bit basis
  - Detects leaked memory, double frees and mismatched frees
- **valgrind --tool=cachegrind**
  - Profile cache usage, simulates instruction and data cache of the cpu
  - Identifies the number of cache misses
  - Needs cache line size, Extract Cache Attributes (ECAG) instruction introduced with z10
- **valgrind --tool=massif**
  - Profile heap usage, takes regular snapshots of program's heap
  - Produces a graph showing heap usage over time



## s390-tools package: what is it?

- 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 .
- 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
  - RedHat Enterprise Linux 6
  - SuSE Linux Enterprise Server 9
  - 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](mailto:linux390@de.ibm.com)

## s390-tools package: the content





## zEC12 Technology

*Planned exploitation of features, introduced with  
zEnterprise EC12 (zEC12).*

# zEnterprise EC12 (zEC12) – New Technology



the new **Crypto Express4S**-security that lets you sleep  
**Secure Enterprise PKCS #11**

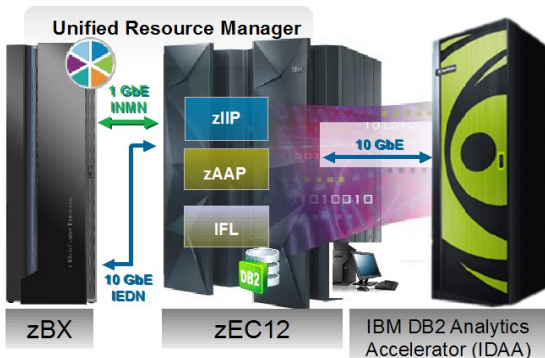


**IBM zAware**  
Discover issues  
**before**  
they become  
problems

powered by **120** of the world's  
**most**  
**powerful microprocessors**  
running at **5.5 GHz**



- Up to **25%** **Increase of Single Processor Performance** over z196
- Up to **50%** **Total Capacity Improvement** over z196
- Up to **45%** **Improvement for Java workloads**
- Up to **3TB** **RAIM Redundant array of independent memory**

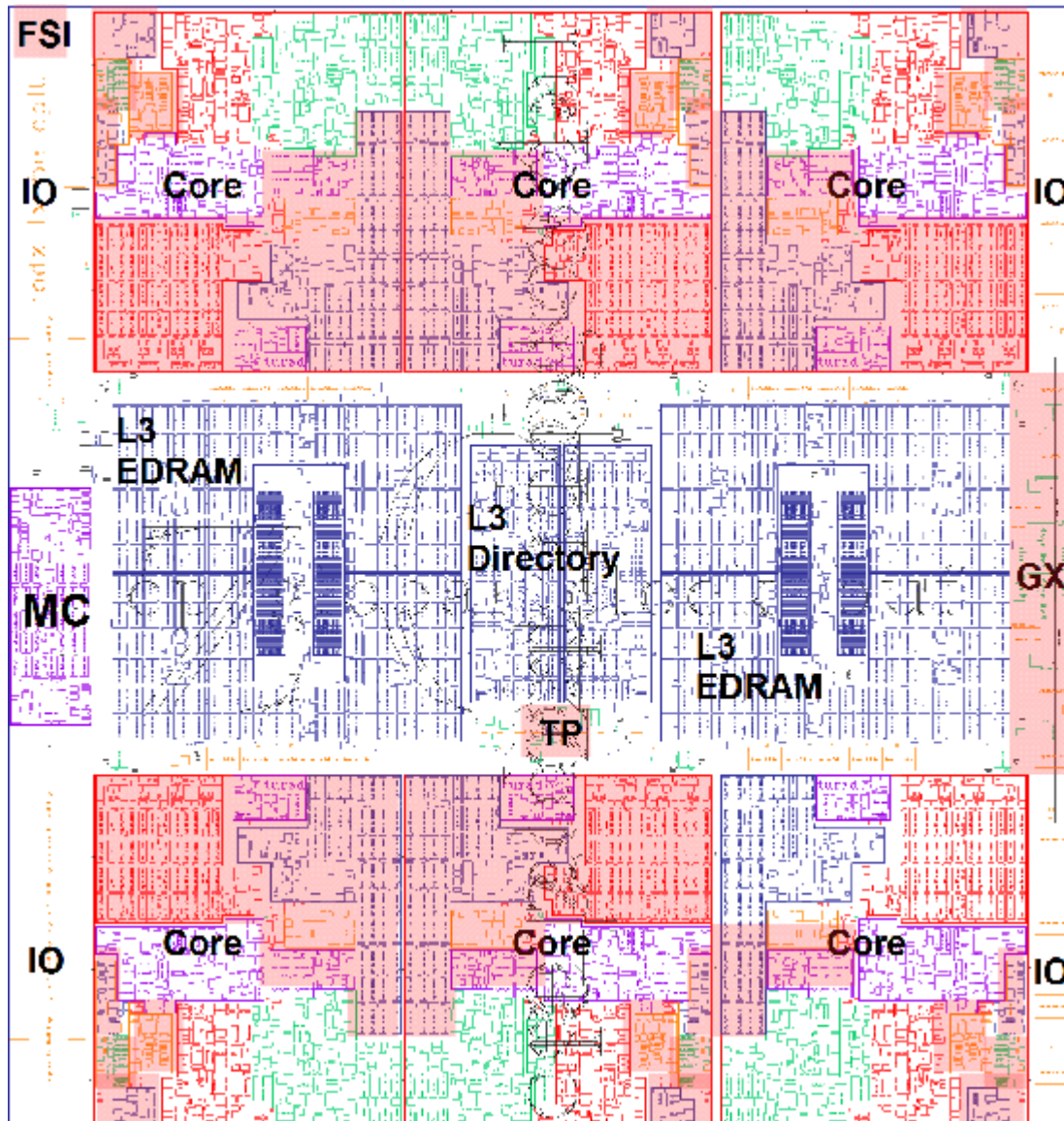


**Transactional Execution**  
and **Runtime Instrumentation**  
enable significantly more efficient  
software



all new **Flash Express**  
speed-up z/OS for  
paging

# zEnterprise EC12 (zEC12) – Processor Features, CP



**UNITS  
OWNED IN  
GERMANY**

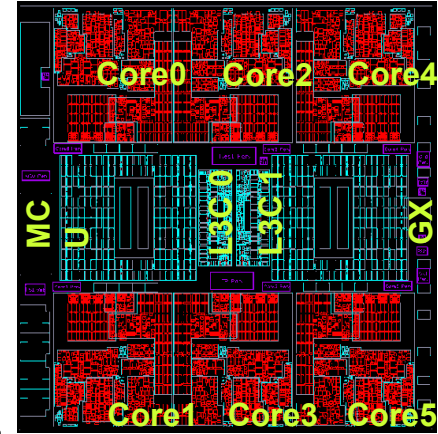
**CP/PU Chip – Compute Processor**  
**2.75 Billion Transistors**  
**48 MB L3**  
**598 mm<sup>2</sup>**

**50% More Cores**  
 + 50% Capacity Increase  
 + Same Total Power  
 + More Throughput  
 + More Redundancy

**5.5 GHz Out-of-Order Cores**  
 + +25% Single-Thread Performance  
 + Transactional Execution (TX)  
 + Advanced Crypto Engine  
 + Runtime Instrumentation (RI)

## zEC12 – Optimization and scale improvements Starting at the Core

- **New 5.5 GHz 6-Core Processor Chip continues our leadership in microprocessor design with a boost in performance for all workloads**
  - Second generation out of order execution design
- **Larger caches to optimize data serving environments**
- **New hardware functions optimized for software performance**
  - **Transactional Execution Facility** for parallelism and scalability (TX)
  - **Runtime Instrumentation Facility** helps reduce JIT / JVM overhead (RI)
  - **2 GB page frames** are intended to offer performance improvements for DB2 buffer pools and Java heaps (EDAT-2)
  - New IBM Enterprise PL/I compiler is planned to exploit and get a performance boost from **decimal format conversions facility**
  - **Unicode UTF8<->UTF16** conversion support via Dedicated Co-Processor
- **Integrated cryptographic function available on each core characterized as a CP or IFL with enabling microcode**

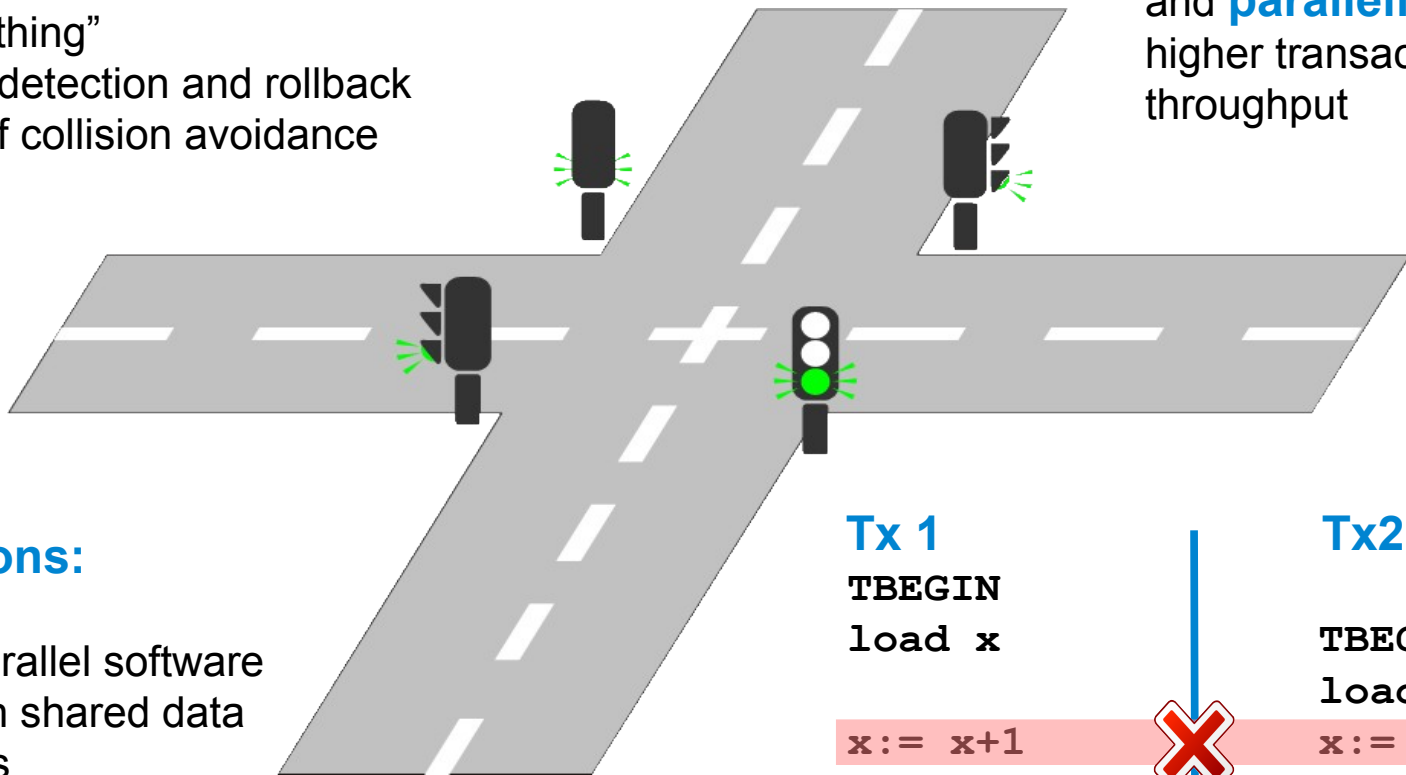


# Transactional Execution (TX) Facility Hardware Transactional Memory (HTM)

## Transaction:

- Atomic memory accesses for set of instructions
- “All or nothing”
- Collision detection and rollback instead of collision avoidance

Increases **scalability** and **parallelism** for higher transaction throughput



## Applications:

- Java
- Highly parallel software relying on shared data structures

**Tx 1**

**TBEGIN**

load x

**x := x+1**

store x

**TEND**

**Tx2**

**TBEGIN**

load x

**x := x+1**

store x

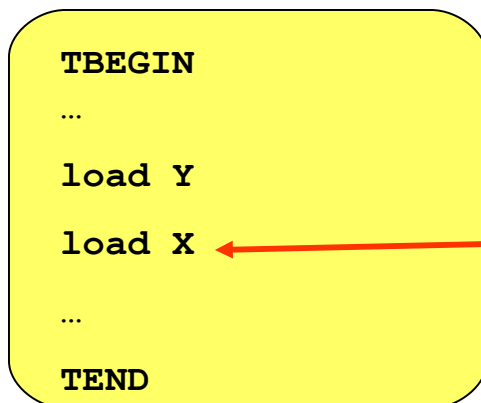
**TEND**



# Hardware Transactional Memory (HTM)

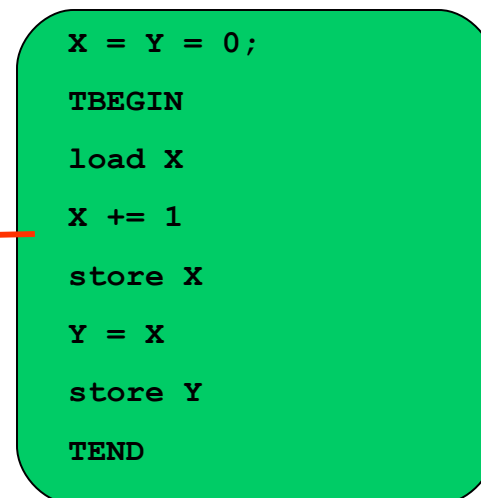
- **Allow lockless interlocked execution of a block of code called a 'transaction'**
  - **Transaction:** Segment of code that appears to execute 'atomically' to other CPUs
    - Other processors in the system will either see **all-or-none** of the storage up-dates of transaction
- **How it works:**
  - TBEGIN instruction starts speculative execution of 'transaction'
  - Storage conflict is detected by hardware if another CPU writes to storage used by the transaction
  - Conflict triggers hardware to roll-back state (storage and registers)
    - transaction can be re-tried, or
    - a fall-back software path that performs locking can be used to guarantee forward progress
  - Changes made by transaction become visible to other CPUs after TEND

CPU 0: Tran A



**Storage conflict:**  
Tran A will abort  
Tran B will commit  
changes to X and Y

CPU 1: Tran B



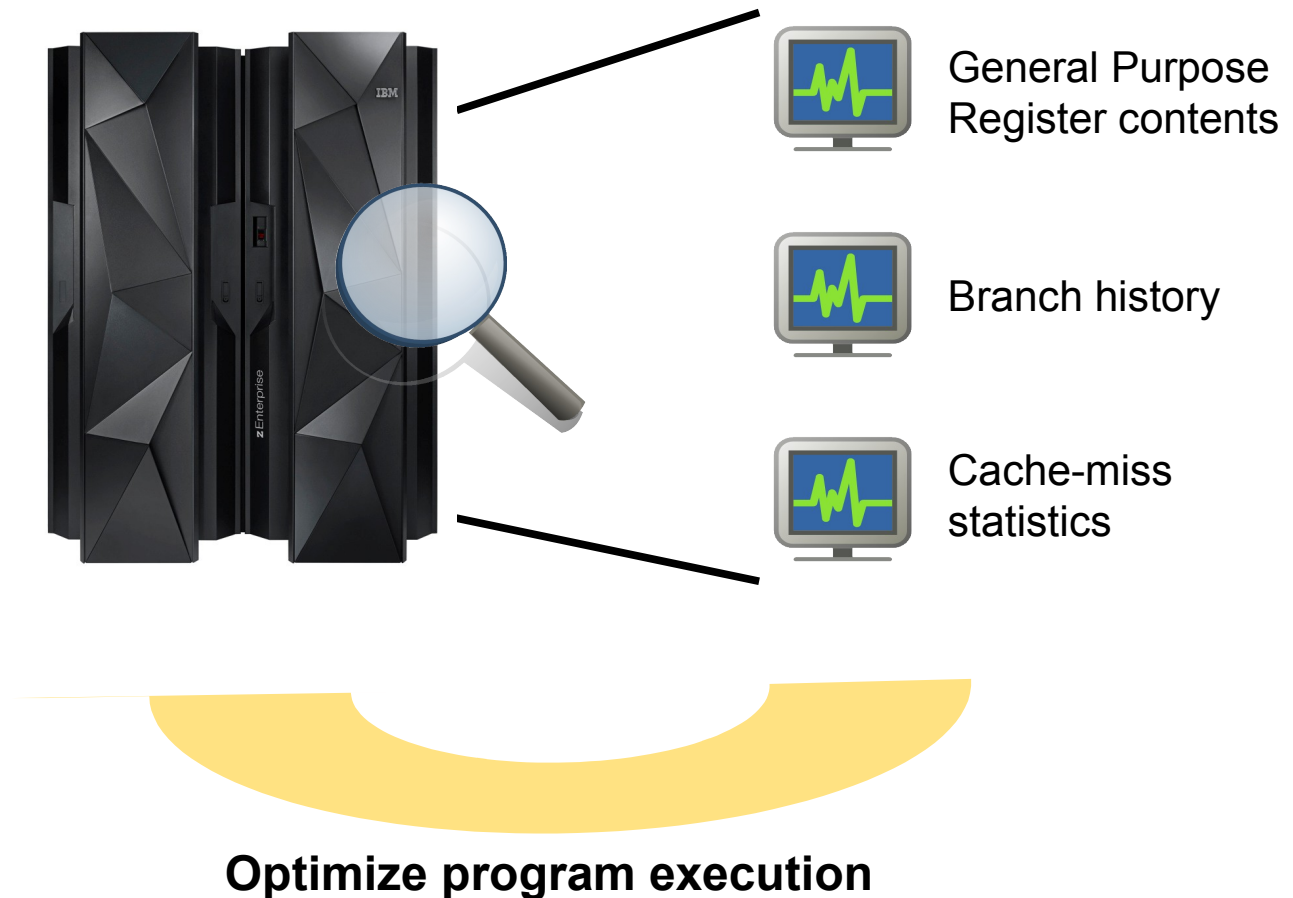
CPU 0 can only see (X=Y=0) or (X=Y=1),  
cannot see (X=1, Y=0)

# Runtime Instrumentation (RI) Facility

- Allows collection of **real-time data** about program execution
- Enables **deep inspection** of program behavior

## Application:

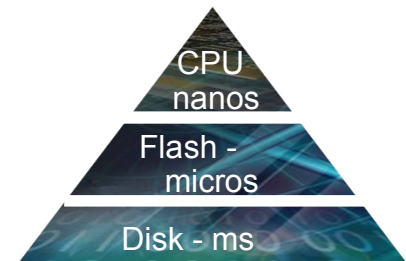
- Dynamic adjustment of Java just-in-time-compiled code (JIT)





## IBM Flash Express – Helps to improve availability and to maximize serviceability

- Companies competing for the highest quality of service want **outstanding availability** to deliver superior SLAs
- **Flash Express** is a new adapter card that **fits in the PCIe I/O drawer** and is designed to help improve availability and performance
  - Automatically improve availability for key workloads at critical processing times
  - Drive performance for critical business workloads that cannot tolerate paging spikes
  - Slashes latency for critical application processing such as diagnostics collection
- **Secure, resilient, easily configurable** for immediate use solution
- IBM is working with its Linux distribution partners to include support in future **Linux on System z** distribution releases as block device.
- Flash Express will be first exploited by z/OS
  - **DB2 for z/OS V10 and Java SDK7 SR3** will support pageable large pages and be able to take advantage of Flash Express when the z/OS support is available
  - **IMS V13 beta program** (announced 4Q12) announced the intent to support pageable large pages



**IBM Flash Express**

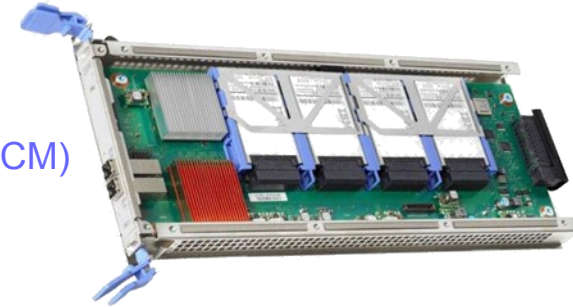




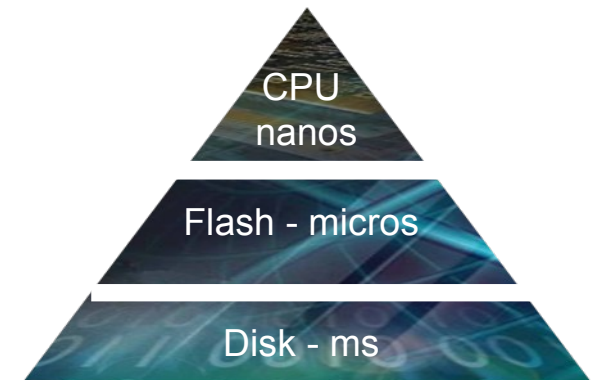
# Flash Express – What is it?

## FLASH Express

- Physically comprised of internal storage on Flash SSDs
- Used to deliver a new tier of memory - **storage class memory (SCM)**
- Supported on z/OS V1.13 plus web deliverable (PTF)
- Uses **PCIe I/O drawer**
- Sized to accommodate *all LPAR paging*
  - Each **card pair** provides **1.6 TB** usable storage (3.2 TB total)
  - Maximum 4 card pairs (4 X 1.6 = 6.4 TB)
- Immediately usable
  - No capacity planning needed
  - No intelligent data placement needed
- Robust design
  - Delivered as a **RAID10** mirrored pair
  - Designed for long life
  - Designed for concurrent firmware upgrade
- Secured
  - Flash Express adapter is protected with 128-bit AES encryption.
  - Key Management provided based on a Smart Card

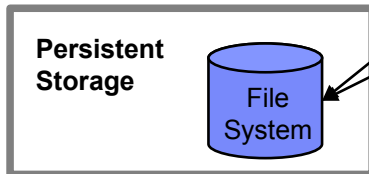
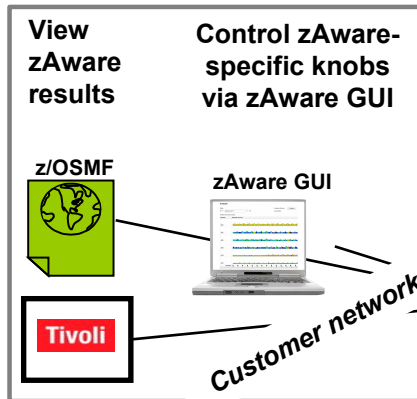


**One Flash Express Card**



# zAware - Detect issues before they become problems

Analytics engine  
from research lab



Sick but  
not dead

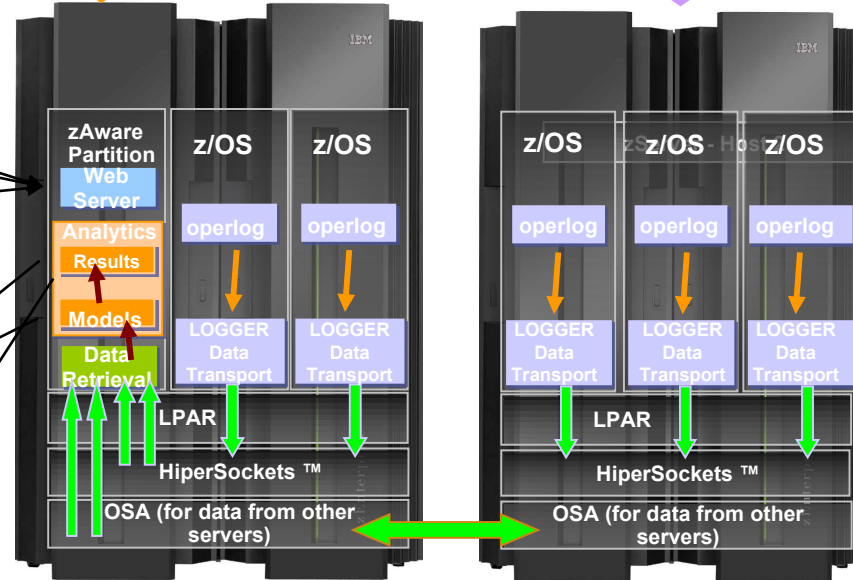


Recognize  
patterns

zAware Partition  
Shipped as firmware



Unveils anomalous  
situations



SWG/SOCC  
collaboration

Early alerts



# Secure Enterprise PKCS#11 with Crypto Express 4S



many **SmartCards**

YESTERDAY



**TODAY**



**Crypto Express 4S**

**OPEN  
STANDARD**

**Public-Key** Cryptography  
Standard **PKCS#11**

**BUILD  
FOR**

Common Criteria  
**Certification**

**NEW  
MARKETS**

Digital Signature  
**Public Sector**

# Linux on System z Support



	z10 EC WdfM	z10 BC WdfM	z196	z114	zEC12	Availability Date
RHEL 5	x	x	x	x	x	03/2007
RHEL 6	x	x	x	x	x	11/2010
SLES 10	x	x	x	x	x	08/2006
SLES 11	x	x	x	x	x	03/2009

	End of Production Ph 1	End of Production Ph 2	End of Production Ph 3
RHEL 5 support*	4Q 2011	4Q 2012	03/31/2014
RHEL 6 support*	4Q 2014	4Q 2015	11/30/2017

	General support	Extended support	Self support
SLES 10 support*	07/31/2013	07/31/2016	07/31/2016
SLES 11 support*	03/31/2016	03/31/2019	03/31/2019

• **For latest information and details contact your Linux distributor**

• Recommendation: use RHEL 6 or SLES 11 for new projects

• RHEL 4 and SLES 9 are not supported on zEC12

• For latest information about supported Linux distributions on System z refer to:

<http://www.ibm.com/systems/z/os/linux/resources/testedplatforms.html>

\* SLES = SUSE Linux Enterprise Server  
RHEL = Red Hat Enterprise Linux  
Support dates may be changed by Linux distributors

## z/VM zEC12 Compatibility Support

- Compatibility → z/VM support for host / guests on zEC12 at the z196 functional level with limited exploitation of new functions (some transparent)
  - Support available as PTFs concurrently with the availability of zEC12.  
Be sure to install these z/VM PTFs for zEC12 before you upgrade a system:  
[http://www-01.ibm.com/support/docview.wss?uid=isg1\\_2827DEVICE\\_2827-ZVM](http://www-01.ibm.com/support/docview.wss?uid=isg1_2827DEVICE_2827-ZVM)
  - Includes PTFs for EREP, IOCP, HCD, HLASM, OSASF, and Performance Toolkit
  - For z/VM 5.4, 6.1, and 6.2
- Compatibility Support for several new architectural facilities
- Support for the PCIe I/O Drawer cage and new OSA-Express4S 1000BASE-T and Crypto Express4S features
- Supported releases:
  - z/VM 5.4 EoS 9/2013
  - z/VM 6.1 EoS 4/2013, requires a z10 ALS (Advanced Lifecycle Support)
  - z/VM 6.2 EoS 4/2015, requires a z10 ALS

## Operating System Support for zEC12

- Currency is key to operating system support and exploitation of future servers
- The following releases of operating systems are supported on zEC12  
(refer to PSP buckets for any required maintenance):

Operating System	Supported levels
<b>z/OS</b>	<ul style="list-style-type: none"> <li>• z/OS V1R13 with PTFs (Exploitation)</li> <li>• z/OS V1R12 with PTFs (Exploitation)</li> <li>• z/OS V1R11 with PTFs (Toleration, Lifecycle extension required after 09/12)</li> <li>• z/OS V1R10 (Toleration, Lifecycle Extension Required)</li> </ul>
<b>Linux on System z</b>	<ul style="list-style-type: none"> <li>• Planned with the following distributions: <ul style="list-style-type: none"> <li>• SUSE SLES 10 and SLES 11</li> <li>• Red Hat RHEL 5 and RHEL 6</li> </ul> </li> </ul>
<b>z/VM</b>	<ul style="list-style-type: none"> <li>• z/VM 5.4, 6.1, and 6.2 will support zEC12 with PTFs</li> </ul>
<b>z/VSE</b>	<ul style="list-style-type: none"> <li>• z/VSE compatibility with PTFs, support for: <ul style="list-style-type: none"> <li>• z/VSE 4.3</li> <li>• z/VSE 5.1</li> </ul> </li> </ul>
<b>z/TPF</b>	<ul style="list-style-type: none"> <li>• V1.1</li> </ul>

- Support for p Blades in zBX Model 003
  - AIX 5.3 Technology Level 12 or higher, AIX 6.1 Technology Level 5 or higher, AIX 7.. All with PowerVM™ Enterprise Edition
- Support for Linux and Windows\* environments on select System x blades in zBX Model 003
  - 64 bit version support only
  - Red Hat RHEL 5.5 and higher, 6.0 and higher and SLES 10 (SP4), 11 SP1 and higher
  - Microsoft Windows Server 2008 R2 and Microsoft Windows Server 2008 (SP2) (Datacenter Edition recommended)

# Linux on System z Development Focus



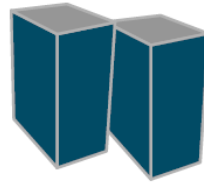
## Integration

### Application Serving

- z/OS & zVSE Integration

### Data Hub

- Database Consolidation



## Virtualization

### Virtualization and Virtualization Management

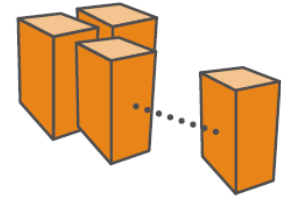
- Ease of Use
- Serviceability
- Hosting Capacity



## 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 Hardware
- Storage exploitation

### Linux

- Maintainability & code currency



# Thank you – Questions ?

## Obrigado

Portuguese

## Merci

French

## Thank You

English

## Gracias

Spanish

## Danke

German



[ibm.com/linux](http://ibm.com/linux)

Linux and IBM:  
In-demand skills for an on demand world.

**IBM**

[ibm.com/education/students](http://ibm.com/education/students)



# LIBERTE



# 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](http://www.ibm.com/legal/copytrade.shtml): AS/400, DB2, e-business logo, ESCON, eServer, FICON, IBM, IBM Logo, iSeries, MVS, OS/390, pSeries, RS/6000, S/390, System Storage, System z9, VM/ESA, VSE/ESA, WebSphere, xSeries, z/OS, zSeries, z/VM.

The following are trademarks or registered trademarks of other companies

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 in the United States and other countries. 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.

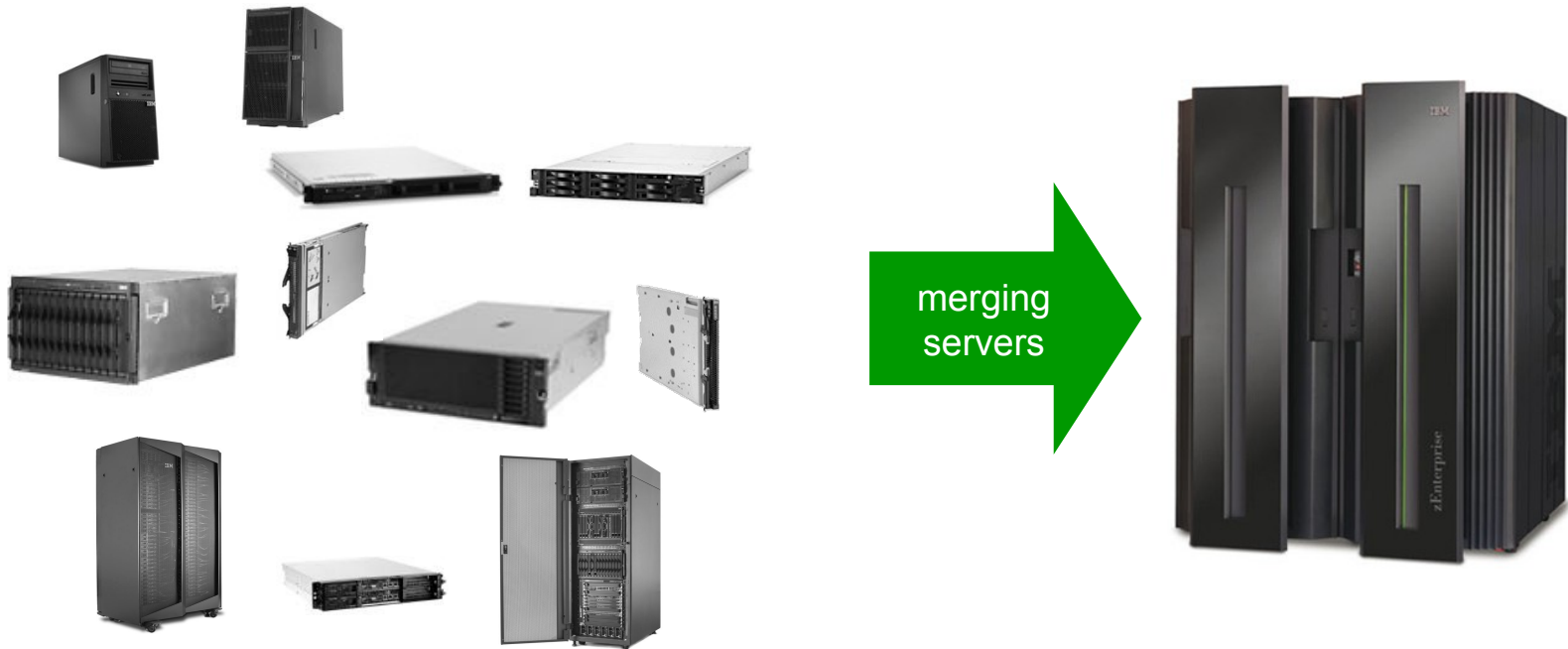
NOTES: Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's 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.

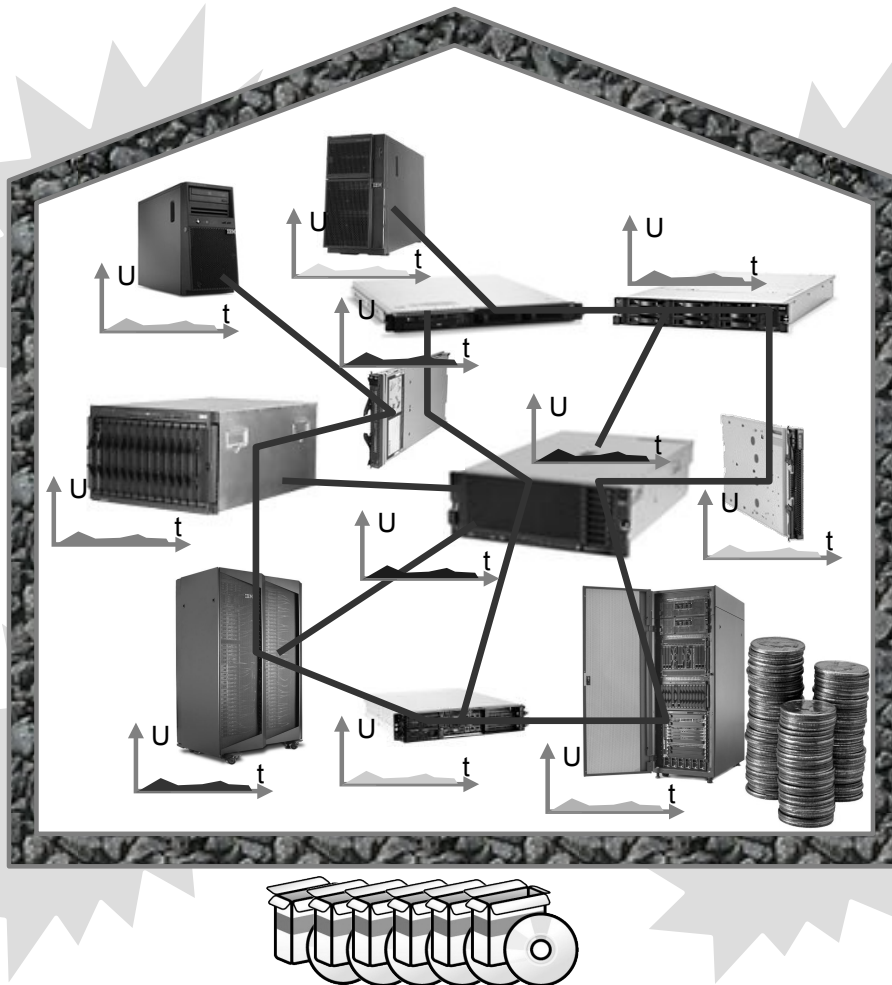
Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography. References in this document to IBM products or services do not imply that IBM intends to make them available in every country. Any proposed use of claims in this 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.

## Consolidation is ... ... merging multiple servers



Yes, but there is more ...

## But consolidation is so much more ...



- Merging multiple servers
- Floor-space / real estate reduction
- Energy (power and cooling) reduction
- Network virtualization and reduction
- Hardware maintenance simplification and reduction
- Operational and management reduction
- Hardware acquisition cost reduction
- Software acquisition and licensing cost reduction
- Merging multiple low utilized servers
- Maximizing utilization
- collocation of data and applications
- flexible security
- flexible scalability
- technology refresh
- DR improvements

