SUSE® Linux Enterprise Server

for System z - Update

Marcus Kraft
Product Manager SUSE Linux Enterprise Server for System z
mkraft@suse.com

Session 10738

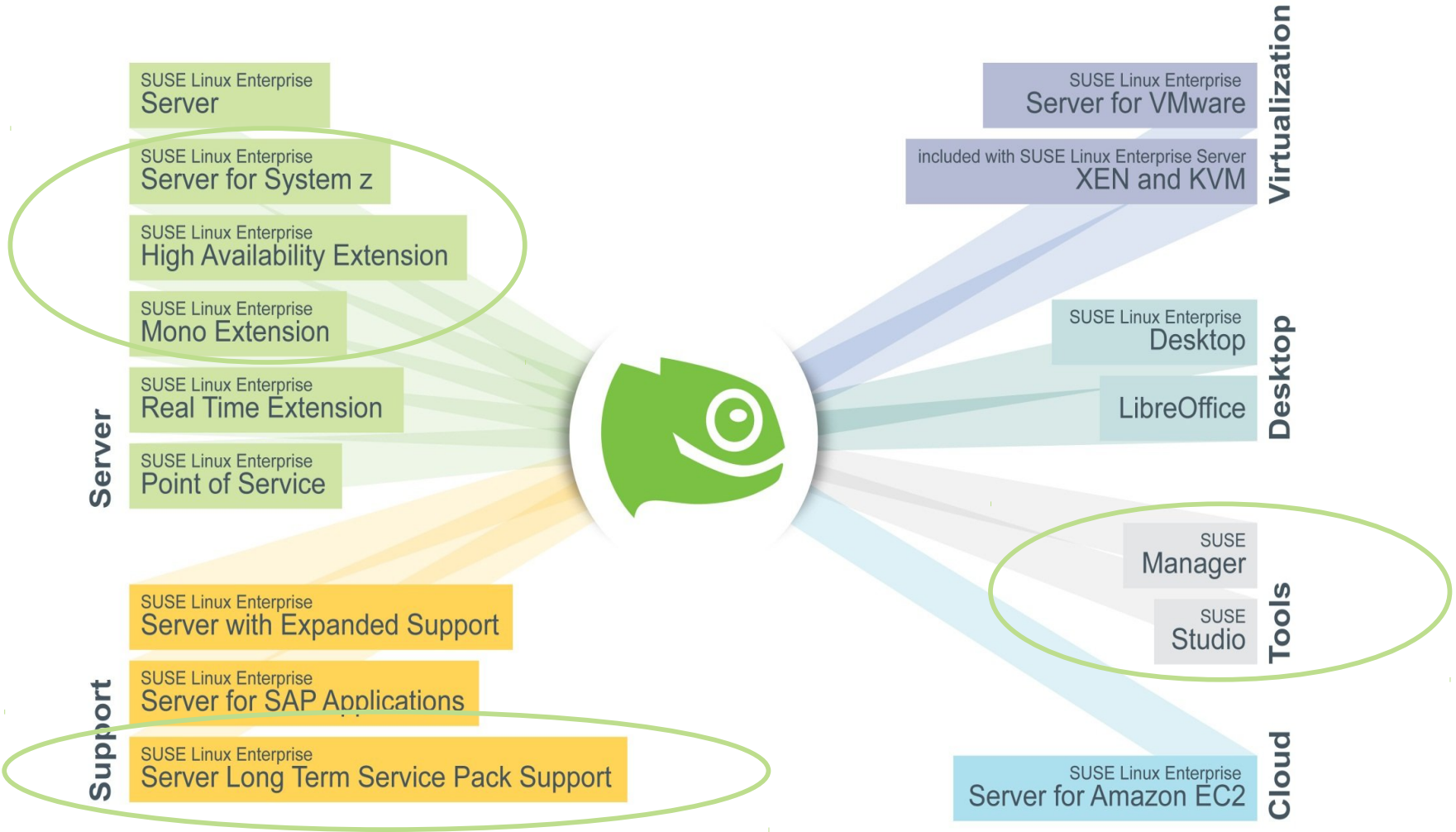
Agenda

- SUSE & Roadmap Update

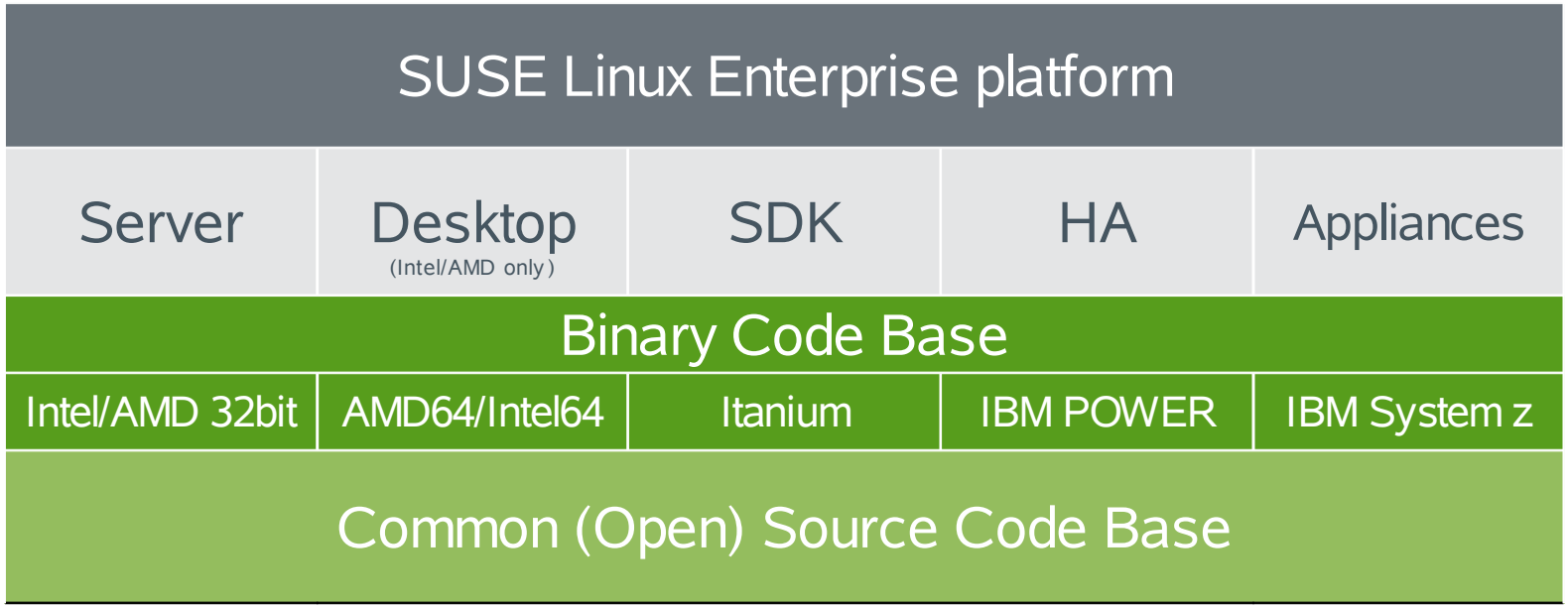
SUSE® And Attachmate Group

- SUSE, headquartered in Nürnberg / Germany, is an independently operating business unit of The Attachmate Group, Inc.
- The Attachmate Group is a privately held 1 billion+ $ revenue software company with four brands:

Comprehensive Portfolio



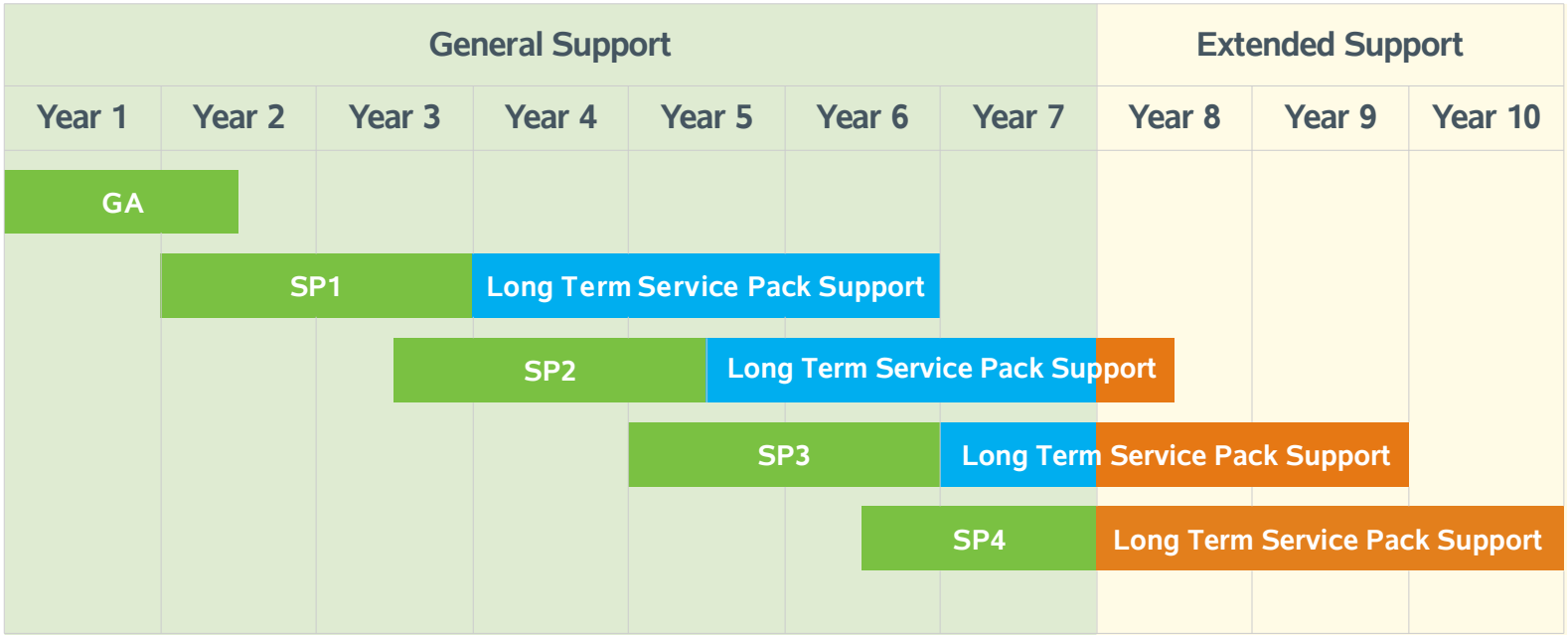
Common Code Base



- Foundation for SUSE Linux Enterprise products
- Fully supported core system
- Choose the right Architecture for your workload

SUSE® Linux Enterprise

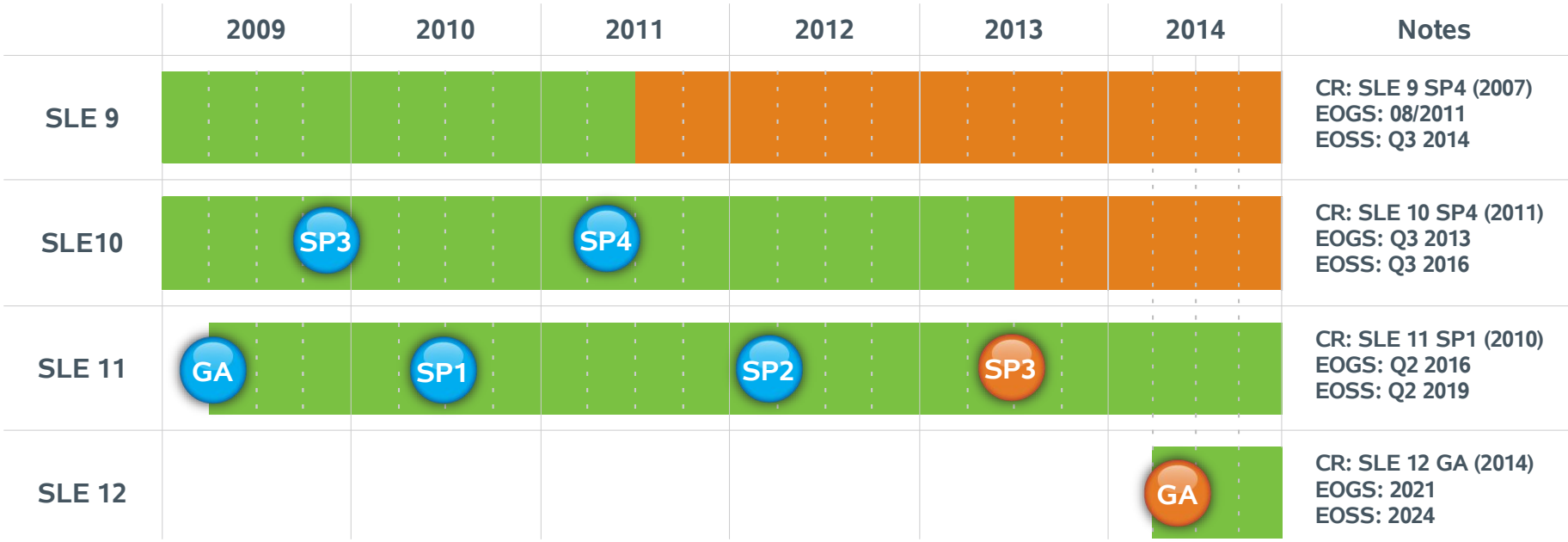
Standard Platform Lifecycle



- 10-year lifecycle (7 years general support, 3 years extended support)
- Major releases every ~4-5 years, service packs every ~18 months
- Six month upgrade window
- Long term service pack support – extend upgrade window or extend major release lifecycle

SUSE® Linux Enterprise

Current Platform Lifecycle

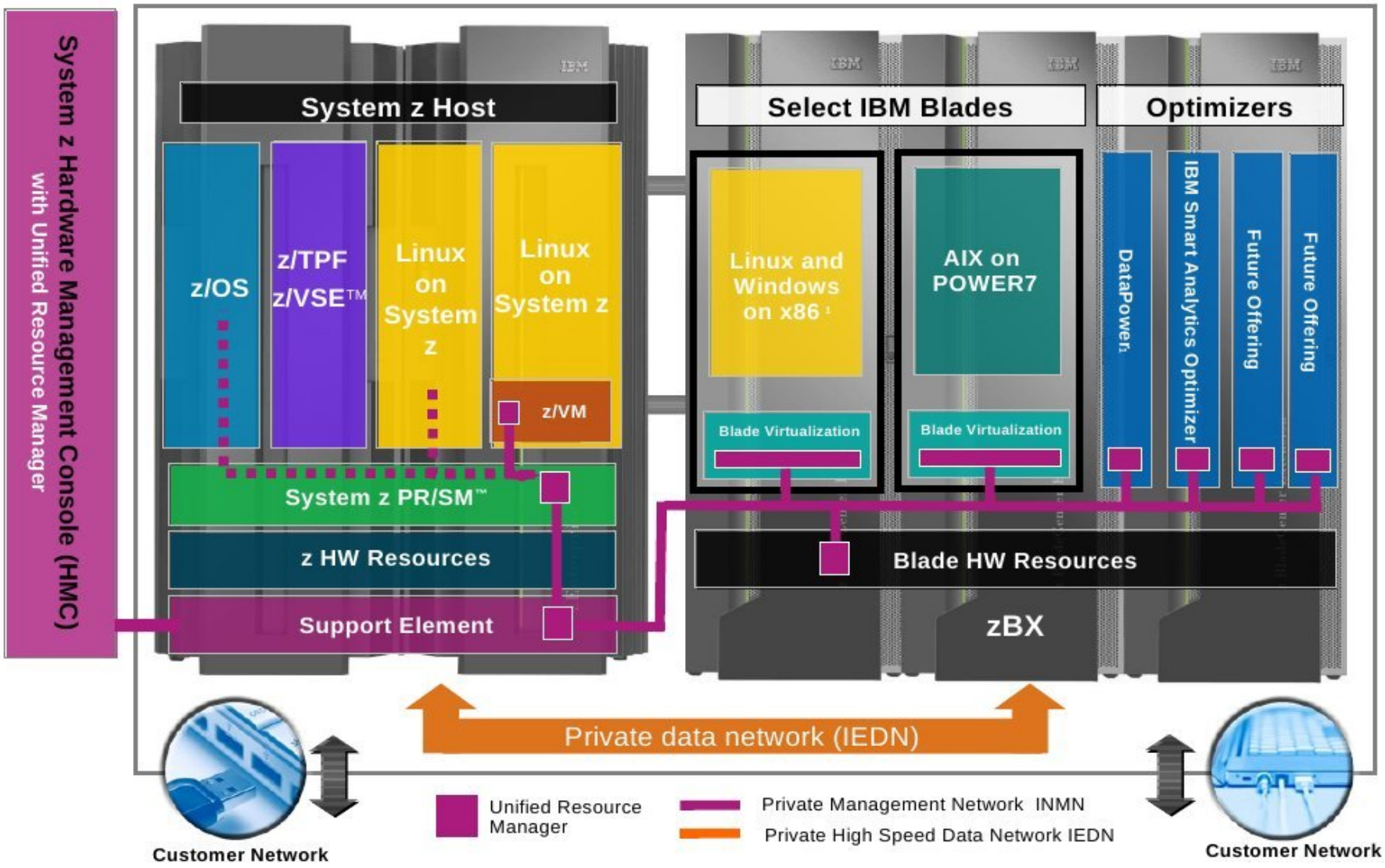


- SUSE announces service pack releases and development and product schedules to customers and partners
- Dependable release timing
- Predictability for planning rollouts and migrations
- All products at http://support.novell.com/lifecycle

SUSE® Linux Enterprise Server 11 SP1

- Full Dynamic Resource Handling
  - Two levels of virtualizations available: LPAR and z/VM
    - Choose the level of isolation mandated by compliance
    - Flexible resource allocation and reallocation without downtime
  - CPU, memory, I/O hotplug
    - Provide the resource where they are needed in LPAR and z/VM guest
- Abundant memory, IO bandwidth and transaction capability
  - Hipersocket support connects Linux and z/OS applications and data
  - I/O fan out and transaction workload capacity is unmatched
- RAS
  - I/O device and other performance statistics
  - Dump generation, handling and inspection tools
  - Centralized and uniform resources support DR recovery setups
  - SUSE Linux Enterprise High Availability Extension included
  - System z specific kernel messages with documentation

IBM zEnterprise System



SUSE® Linux Enterprise Server 11 SP2

- z196 / z114 + zBX = IBM zEnterprise exploitation
  - CPU topology and instruction set exploitation of z196 (SDK)
  - New CHPID support connecting both environments
- Choose the right environment for the right workload
  - ISVs application support might mandate the platform
  - SLES supported for both hardware architectures
- Improved tools and z specific support
  - Disk storage & crypto enhancements
  - Linux RAS support, s390-tools update

Unique Tools Included

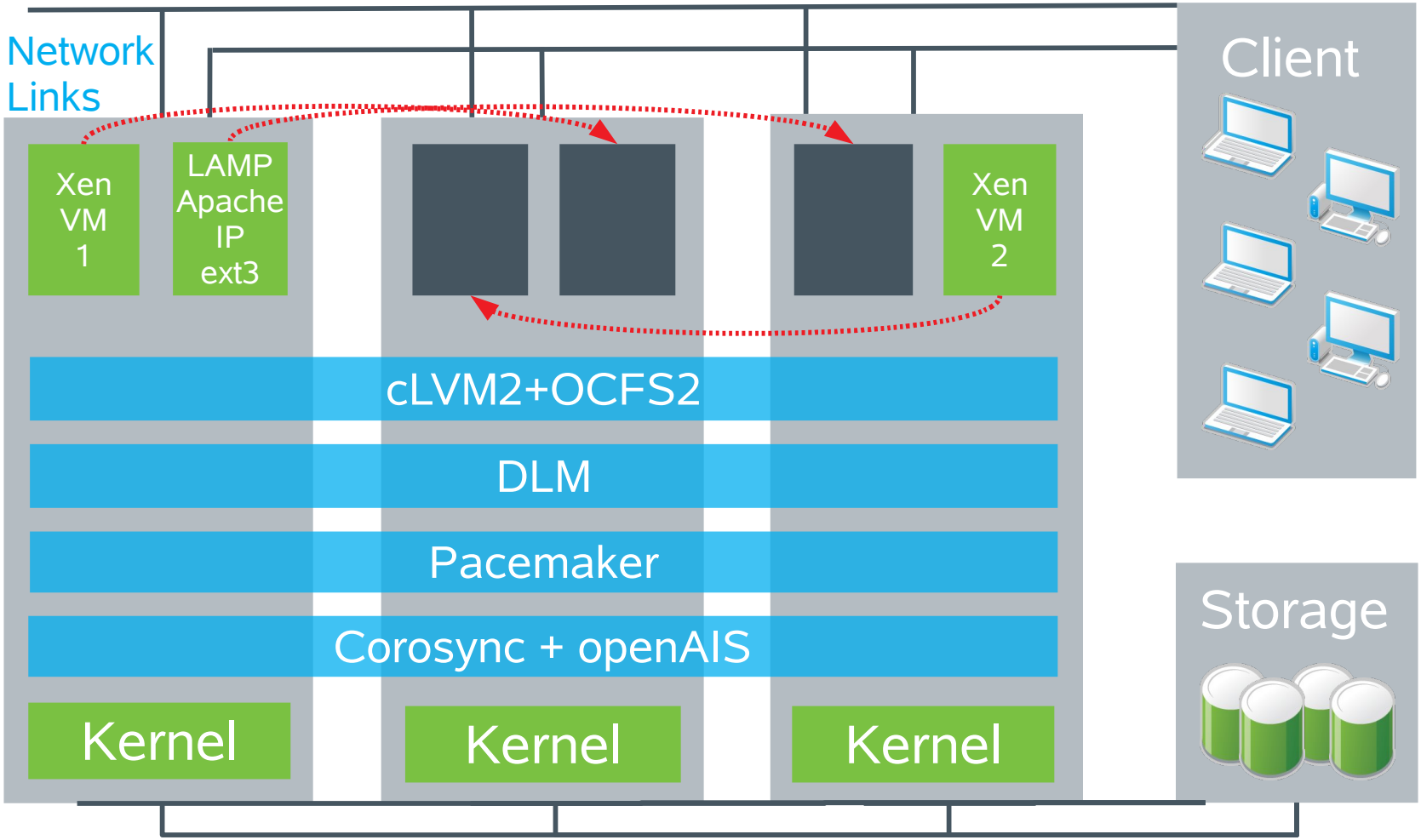
- Yast and Integrated Systems Management
  - Install, deploy, and configure every aspect of the server
- Starter System for System z
  - A pre-built installation server, depolyable with z/VM tools
- Subscription Management Tool
  - Subscription and patch management, proxy/mirroring/staging
- High Availabilty Extension for SLES
  - Cluster Framework, ClusterFS, DRBD, tools, GEO-cluster
- AppArmor Security Framework
  - Application confinement

Enhance Your Applications
Examples

- SLE HA: make your applications High Availbilty ready
  - Resource agents examples
    - /usr/lib/ocf/resource.d/heartbeat/* → example: Dummy resource agent
  - http://www.opencf.org
- AppArmor: secure your applications
  - Easy to use GUI tools with statics analysis and learing-based profile development
  - Create custom policy in hours, not days

Cluster Example

SUSE® Linux Enterprise High Availability Extension



AppArmor: usr.sbin.vsftpd

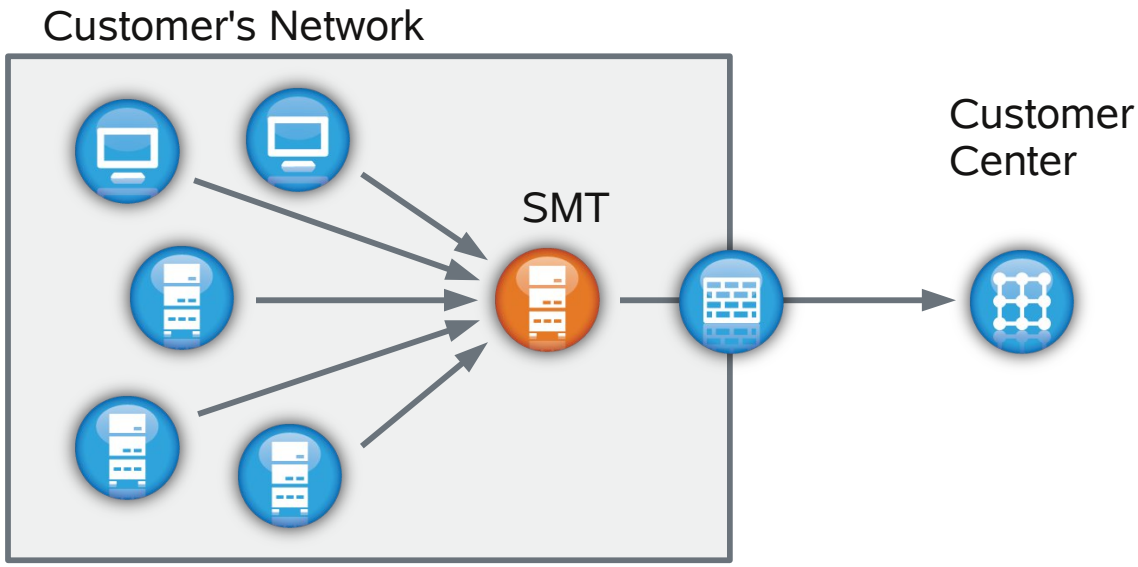
/etc/apparmor/profiles/extras/

```
#include <tunables/global>

/usr/sbin/vsftpd {
  #include <abstractions/base>
  #include <abstractions/nameservice>
  #include <abstractions/authentication>

  /dev/urandom              r,
  /etc/fstab                r,
  /etc/hosts.allow          r,
  /etc/hosts.deny           r,
  /etc/mtab                 r,
  /etc/shells               r,
  /etc/vsftpd.*             r,
  /etc/vsftpd/*             r,
  /usr/sbin/vsftpd          rmix,
  /var/log/vsftpd.log       w,
  /var/log/xferlog          w,
  # anon chroots
  /                         r,
  /pub                      r,
  /pub/**                   r,
  @{HOMEDIRS}               r,
  @{HOME}/**                rwl,
}
```

Subscription Management Tool
Overview

SMT is a proxy and auditing tool that mirrors the Customer Center and tightly integrates with it.

It allows you to accurately register and manage an entire SUSE Linux Enterprise deployment, guaranteeing the subscription compliance and secure IT process flow organizations require.



Starter System for System z

- A pre-built installation server that can be installed on your z/VM system using CMS tools
- Eliminates the need for coordinating access to a separate Linux or UNIX system elsewhere on your network
- Minimizes the impact of network-based installation on your internal and external networks

- Next refresh is based on SLES 10 SP4 (H1 2012)
- Future refresh based on SLES 11 SP2 (H2 2012)

- *Session 10728: Best practices with SLES Starter System*

SUSE® Linux Enterprise 11 SP2

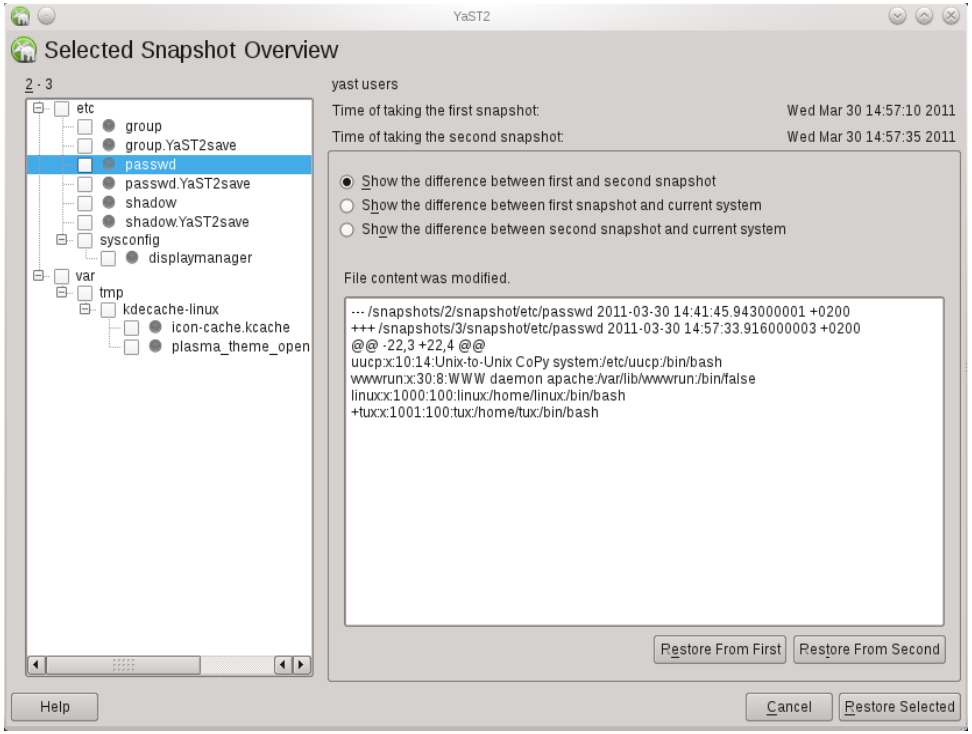
- Hardware enablement and RAS improvements
- Equivalent or exceeding proprietary Unix capabilities
  - btrfs: file system with “Copy on Write”, checksums, snapshotting, reduce cost of storage management by providing an integration of logical volume management and filesystem, checksums on data and metadata ensure data integrity
  - LXC: container support based on control groups
- Snapshot / rollback for package and configuration updates
  - YaST2 + ZYPP + btrfs
- SUSE Linux Enterprise High Availability Extension: Geo-cluster, automated and pre-configuration
- Unattended upgrade from SUSE Linux Enterprise 10 to SUSE Linux Enterprise 11

Kernel 3.0
Selected benefits

- Most recent HW enablement
- Removal of BLK (Big Kernel Lock)
- Control Groups enhancements
  - I/O throttling support for process groups
  - memory cgroup controller
- Integration of AppArmor
- More powerful firewalls based on faster packet filtering
- Transparent Huge Pages (THP)

SUSE Linux Enterprise
btrfs and Snapshot / Rollback

- SUSE Toolchain co-developed for openSUSE and SUSE Linux Enterprise
  - “snapper” command line tool
  - YaST2 integration for snapshot rollback
  - Unique functionality: selective rollback



http://lizards.opensuse.org/2011/04/01/introducing-snapper/

Why another filesystem?

- Solve Storage Challenges
  - Scalability
  - Data Integrity
  - Dynamic Resources (expand and shrink)
  - Storage Management
  - Server, Cloud – Desktop, Mobile
- Match and exceed other Operating Systems

Why btrfs?

btrfs (better fs) – Features

- Integrated Volume Management
- Support for Copy on Write
- Powerful Snapshot capabilities
- Scalability (16 EiB) including effective shrink
- Supports offline in-place migration from ext2+
- Other Capabilities:
  - Compression
  - Data integrity (checksums)
  - SSD optimization (TRIM)

Technology Overview

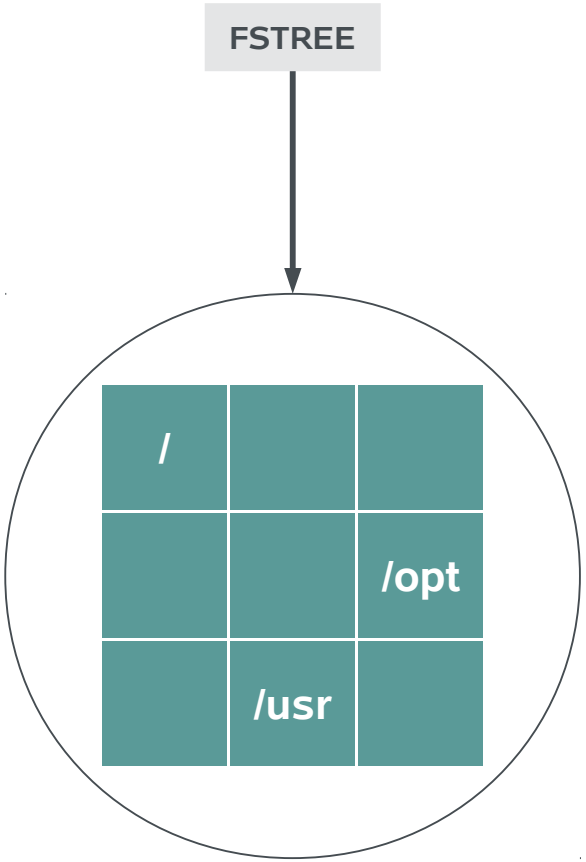
Subvolume (1)

- A complete filesystem tree
- Usually appears as a sub-directory in the “parent” fs
- Can be mounted separately, but not “just a subdirectory”
- Simliar to
  - two “foreign” filesystems, which are
  - using the same pool of data blocks (and other infrastructure)
- Benefits
  - different parts (subvolumes) of a filesystem can have different attributes, such as quotas or snapshotting rules
  - Copy on Write is possible across volumes
- Basic commandline management
  - “btrfs subvolume …”

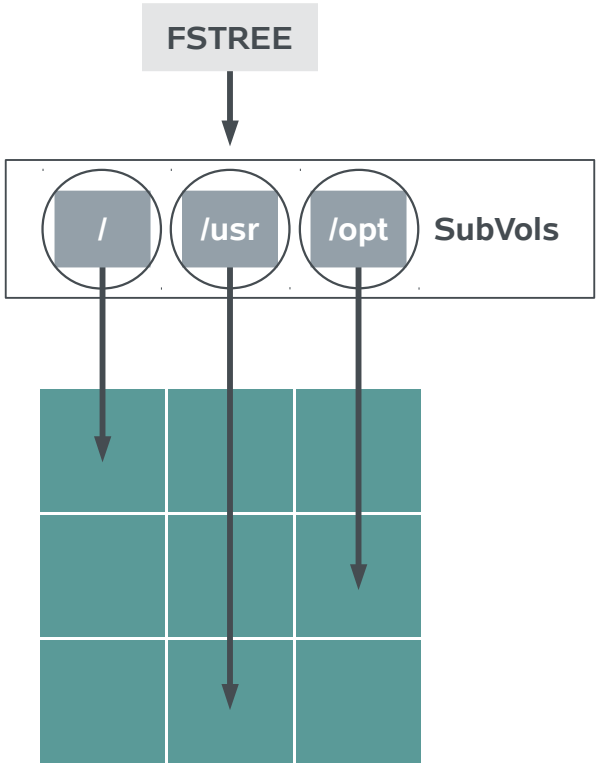
Technology Overview

Subvolume (2)

Normal Filesystem



With Subvolumes



Technology Overview

Snapshots

- Copy on Write on a
  - full subvolume tree instead of a single file only
- Every snapshot is again a subvolume of its own
- Snapshots (as subvolumes) can be mounted and accessed as every other subvolume
- Snapshots can be created read-only
- Basic command line management
  - “btrfs subvolume snapshot …”

Rollback – per Subvolume

How it works

- Instead of the original subvolume, the snapshot is mounted with the options “subvol=<name>”
  - Remember: snapshots are subvolumes
- Talking about the “/” filesystem, the “subvol” can also be hardcoded using “btrfs subvolume set-default ...”

Benefits

- “atomic” operation
- Very fast

Disadvantages

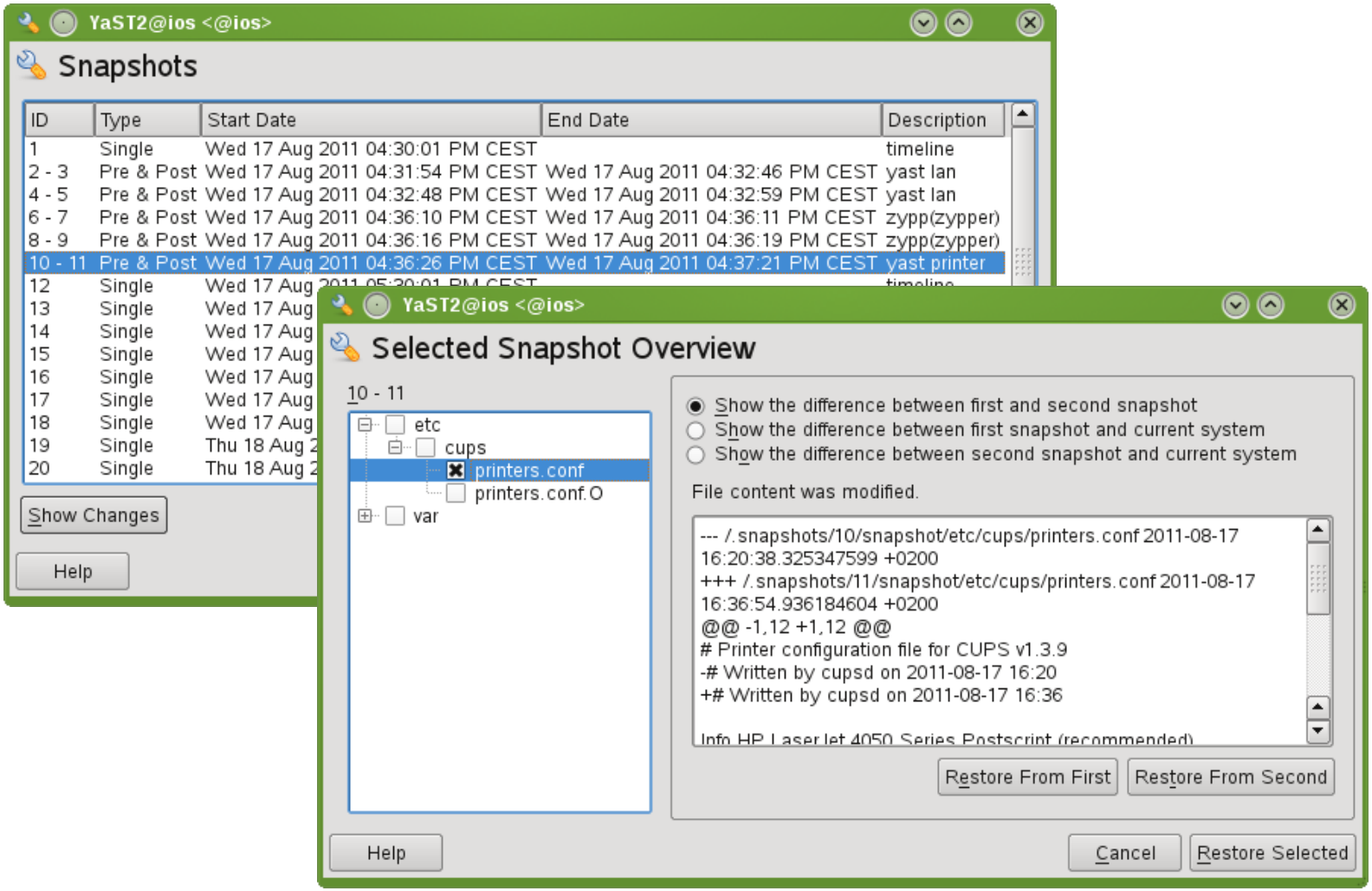
- Additional complexity
  - May require explicit mounting of subvolumes
- No “rollback” per single file

Snapshotting “/”

- We have decided to go the way of “/” in a subvolume
- Disadvantages of this model are mitigated by
  - support from the YaST2 Partitioner to install and configure
  - using “set-default” for the root filesystem to
    - make migration as smooth as possible
    - enable use of “normal” rescue systems

Snapshots in SUSE® Linux Enterprise 11 SP2

YaST2 Management



What Are Control Groups?

Control Groups provide a mechanism for aggregating and partitioning sets of tasks, and all their future children, into hierarchical groups with specialized behavior.

- cgroup is another name for Control Groups
- Partition tasks (processes) into a one or many groups of tree hierarchies
- Associate a set of tasks in a group to a set of subsystem parameters
- Subsystems provide the parameters that can be assigned
- Tasks are affected by the assigning parameters

Example of the Capabilities of a cgroup

Consider a large university server with various users - students, professors, system tasks etc. The resource planning for this server could be along the following lines:

CPUs

Top cpuset (20%)

/ \

CPUSet1 CPUSet2

| |

(Profs) (Students)

60% 20%

Memory

Professors = 50%

Students = 30%

System = 20%

Disk I/O

Professors = 50%

Students = 30%

System = 20%

Network I/O

WWW browsing = 20%

/ \

Prof (15%) Students (5%)

Network File System (60%)

Others (20%)

Source: /usr/src/linux/Documentation/cgroups/cgroups.txt

Control Group Subsystems

Two types of subsystems

- Isolation and special controls
  - cpuset, namespace, freezer, device, checkpoint/restart
- Resource control
  - cpu(scheduler), memory, disk i/o, network

Each cgroup subsystem can be mounted independently

- mount -t cgroup -o cpu none /cpu
- mount -t cgroup -o cpuset none /cpuset

or all at once

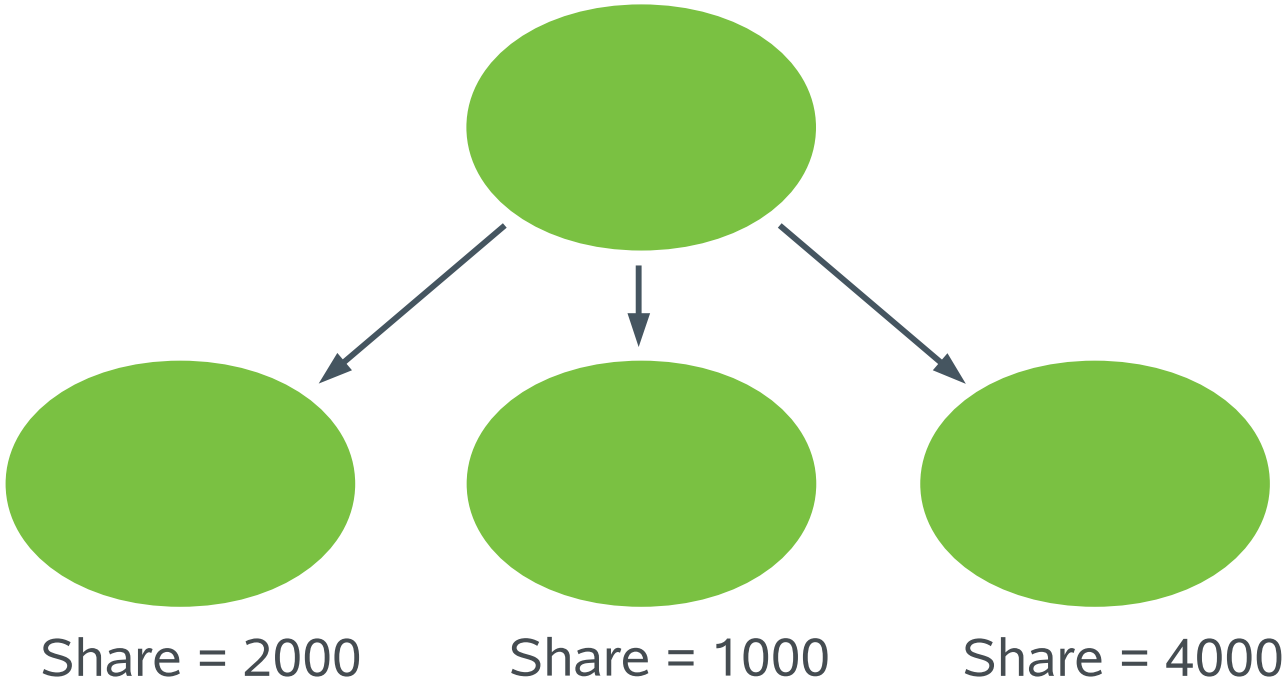
- mount -t cgroup none /cgroup

Source: http://jp.linuxfoundation.org/jp_uploads/seminar20081119/CgroupMemcgMaster.pdf

CPU Subsystem

Resource Control

- Share CPU bandwidth between groups by using the group scheduling function of CFS (the scheduler)



Memory Subsystem

Resource Control

- For limiting memory usage of user space processes.
- Limit LRU (Least Recently Used) pages
  - Anonymous and file cache
- No limits for kernel memory
  - Maybe in another subsystem if needed
- Note: cgroups need ~2% of (resident) memory
  - can be disable at boot time with kernel paramenter "cgroup_disable=memory"

Source: http://jp.linuxfoundation.org/jp_uploads/seminar20081119/CgroupMemcgMaster.pdf

Device Subsystem

Isolation

A system administrator can provide a list of devices that can be accessed by processes under cgroup

- Allow/Deny Rule
- Allow/Deny : READ/WRITE/MKNOD

Limits access to device or file system on a device to only tasks in specified cgroup

Source: http://jp.linuxfoundation.org/jp_uploads/seminar20081119/CgroupMemcgMaster.pdf

Tools / SDK

zPDT

IBM System z Personal Development Tool

https://www.ibm.com/partnerworld/page/pw_com_zpdt

- zPDT is a software-based application tool
  - Low cost IBM System z platform for ISV application development, testing, demo
  - A virtual System z architecture environment that allows select mainframe operating systems, middleware and software to run unaltered on x86 processor-compatible platforms.
  - Portable System z platform for training & education of applications and operating system environments
  - Supports openSUSE 10.3+, SLES11SP1 x86_64, and others
  - SUSE's evaluation versions for x86_64 and s390x available at http://www.suse.com/products/server/eval.html

Tools

Dynamic analysis tools

- valgrind
  - Memcheck
  - Cachegrind
  - Massif
  - Helgrind
  - DRD
  - None
  - Exp-ptrcheck
  - Callgrind
- http://valgrind.org

Tools
cachegrind

- Analysis of cache behaviour of applications
  - z10 cache sizes used as default, changeable (eg. z9, z196)
  - Two cache levels (1st and last level) for instructions & data
  - Writes cachegrind.out.<pid> files

```
r1745045:~ # valgrind --tool=cachegrind ls
==21487== Cachegrind, a cache and branch-prediction profiler
==21487== Copyright (C) 2002-2010, and GNU GPL'd, by Nicholas Nethercote et al.
==21487== Using Valgrind-3.6.1 and LibVEX; rerun with -h for copyright info
==21487== Command: ls
==21487==
--21487-- Warning: Cannot auto-detect cache config on s390x, using one or more defaults
bin  inst-sys  repos  testtools
==21487==
==21487== I   refs:      656,270
==21487== I1  misses:        792
==21487== LLi misses:        656
==21487== I1  miss rate:    0.12%
==21487== LLi miss rate:    0.09%
==21487==
==21487== D   refs:      453,124  (361,066 rd   + 92,058 wr)
==21487== D1  misses:      1,869  (  1,589 rd   +    280 wr)
==21487== LLd misses:      1,313  (  1,061 rd   +    252 wr)
==21487== D1  miss rate:     0.4% (    0.4%     +    0.3%  )
==21487== LLd miss rate:     0.2% (    0.2%     +    0.2%  )
==21487==
==21487== LL refs:         2,661  (  2,381 rd   +    280 wr)
==21487== LL misses:       1,969  (  1,717 rd   +    252 wr)
==21487== LL miss rate:      0.1% (    0.1%     +    0.2%  )
```

z196 exploitation via alternate GCC 4.6

Fate 311859 / [LTC 66797]

http://www.ibm.com/developerworks/linux/linux390/ -> toolchain (pending)

http://gcc.gnu.org/gcc-4.6/changes.html -> z196

- **Performance improvement for applications:** exploitation of new z196 processor instructions and optimized alignment of code (out-of-order pipeline architecture, conditional load/store instructions, new 3 register operand instructions, new atomic instructions, etc)

- **Customer benefit**

technical	business
• Hardware exploitation of the z196 instruction set for user land applications (ISV and self compliled applications), recompile programs with --march=z196 and/or –mtune=z196	• z196 optimized code and efficient excecution use less time and cycles for same workload • Increase of application workload density per system

SLES	10	11
GA	-	-
SP1	-	-
SP2+3	-	yes
SP4	-	n/a

Exploitation of z10 prefetching instructions in GCC

Fate 311845 / [LTC 66745]

http://www.ibm.com/developerworks/linux/linux390/ -> gcc 4.5.1 (upstream)
http://gcc.gnu.org/gcc-4.5/

- **Toolchain based performance improvement for applications:** z10 introduced pre-fetching instructions to enhance memory access like copying memory, zeroing out memory and exploiting predictable loops by help of the compiler.

- **Customer benefit**

technical	business
• Hardware exploitation of the z10 and z196 instruction set for user land applications (ISV and self compliled applications)	• z10 and z196 optimized code and efficient excecution use less time and cycles for same workload • Increase of application workload density per system

SLES	10	11
GA	-	-
SP1	-	-
SP2+3	-	yes
SP4	-	n/a

OSX (OSM) chpids for hybrid data (management) network

Fate 311898 / [LTC 66966]

http://www.ibm.com/developerworks/linux/linux390/kernel-2.6.35.html → OSX

http://www.ibm.com/developerworks/linux/linux390/s390-tools-1.10.0.html → znetconf

- **z196 and zBX exploitation:** enhancement in the network device configuration tool znetconf (s390-tools) by updating internal tables to handle OSX and OSM CHPIDs.

- **Customer benefit**

technical	business
• Hardware exploitation of the z196 and zBX for hybrid computing	• Fit-to-purpose workload placement support

SLES	10	11
GA	-	-
SP1	-	-
SP2+3	-	**yes**
SP4	-	n/a

cio: handle channel path description changes

Fate 311913 / 311911 / [LTC 69631]

http://www.ibm.com/developerworks/linux/linux390/kernel-3.0.html -> dynamic IODF

- **Dynamic resource allocation:** the common I/O layer handles dynamic IODF changes that result in changed capabilities of channel paths. Applies for LPAR installations only, since the required channel subsystem notifications are not supported on current z/VM versions.

- **Customer benefit**

technical	business
• More flexible I/O configuration for Linux running in LPARs	• Non disruptive change of IO configuration

SLES	10	11
GA	-	-
SP1	-	-
SP2+3	-	**yes**
SP4	-	n/a

FICON Dynamic PAV toleration

Fate 311760 / [LTC 66751]

http://www.ibm.com/developerworks/linux/linux390/kernel-2.6.35.html → dynamic PAV

- **Dynamic PAV:** the DASD device driver tolerates dynamic Parallel Access Volume (PAV) changes for base PAV. PAV changes in the hardware configuration are detected and the mapping of base and alias devices in Linux is adjusted accordingly.
- **Customer benefit**

technical	business
• if the mapping of an alias to a base device is changed another device the DASD driver will tolerate this change • change in the base/alias mapping is automatically discovered by the DASD device driver	• improve the flexibility and availability of SLES for System z, by allowing to tolerate changes in the PAV infrastructure without need to restart the system

SLES	10	11
GA	-	-
SP1	-	-
SP2+3	-	**yes**
SP4	-	n/a

FICON Multi-Track extensions for High Performance

Fate 311870 / [LTC 66846]

http://www.ibm.com/developerworks/linux/linux390/kernel-2.6.38.html → multitrack

- **Hardware exploitation:** exploit DS8000 storage systems support for multi-track High Performance FICON requests (read or write data to more than one track).

- **Customer benefit**

technical	business
• Provides a new cio layer function using an interface to get the maximum usable data size for zHPF requests on a given subchannel	• Maximize I/O performance with FICON, zHPF and DS8000 storage servers

SLES	10	11
GA	-	-
SP1	-	-
SP2+3	-	**yes**
SP4	-	n/a

Crypto CP ACF exploitation

Fate 311914 / 311924 311091 / [LTC 69628 etc]

http://opencryptoki.git.sourceforge.net/ -> libica

http://www.ibm.com/developerworks/linux/linux390/kernel-2.6.38.html → zcrypt

- **Cryptography:** hardware based acceleration of complex cryptographic algorithms, support for 4096 bit RSA FastPath (support zEnterprise Crypto Express3 card RSA mod expo operations with 4096-bit RSA keys in ME (Modulus Exponent) and CRT (Chinese Remainder Theorem) format)
- **Customer benefit**

technical	business
• Cryptographically secured connections • devices maximum request size is adjusted based on a test	• Improve security for data transfers over the network • reduce the cost of SSL acceleration replacing expensive and MIPS intensive mathematical calculations in software

SLES	10	11
GA	-	-
SP1	-	-
SP2+3	-	**yes**
SP4	-	n/a

cmsfs support for kernel 2.6

Fate 311847 / 311858 / [LTC 60032] / [LTC 66799]

Device Drivers, Features, Commands on SUSE Linux Enterprise Server 11 SP2, p.441

- **s390-tools:** read and write configuration files stored on CMS disks directly from Linux. CMS disk can be mounted so the files on the disk can be accessed by common Linux tools

- **Customer benefit**

technical	business
• cmsfs-fuse tool translates the record-based EDF file system on the CMS disk to UNIX semantic • Text files can be automatically converted from EBCDIC to ASCII	• Access data (config, files, dumps,...) from the z/VM CMS filesystem, from Linux during operation (no shutdown required)

SLES	10	11
GA	-	-
SP1	-	-
SP2+3	-	**yes**
SP4	-	n/a

Get CPC name (xDR)

Fate 311920 / [LTC 69632]

- **RAS/DR:** enables for dynamic changes in the GDPS environment definition to. now changed to retrieve CPC and LPAR information dynamically. With the new function, GDPS always resets exactly the LPAR in which the OS is running.
- **Customer benefit**

technical	business
• avoid possible failures from manual or forgotten changes • avoid resetting a LPAR due to incorrect configuration definitions GDPS	• additional protection against outages

SLES	10	11
GA	-	-
SP1	-	-
SP2+3	-	**yes**
SP4	-	n/a


IBM
IBM
zEnterprise 196
z BladeCenter Extension

SUSE® Linux Enterprise Server for System z11 SP2

Available today

- IBM zEnterprise System exploitation
- Enhanced tools and z support
- Choose the right environment for th right workload

Corporate Headquarters
Maxfeldstrasse 5
90409 Nuremberg
Germany

+49 911 740 53 0 (Worldwide)
www.suse.com

Join us on:
www.opensuse.org

SUSE® Linux Enterprise

Documentation and Release Notes

- Product Pages
  - http://www.suse.com/products/server/
  - http://www.suse.com/products/highavailability/
  - http://www.suse.com/products/realtime/
- Unix to Linux Migration
  - http://www.novell.com/linux/unixtolinux/
- Documentation
  - http://www.novell.com/documentation/suse.html
- Product Life-cycle
  - http://support.novell.com/lifecycle/linux.html
- Release Notes
  - http://www.novell.com/linux/releasenotes/

More Resources

- Product Homepage
  - http://www.suse.com/products/suse-manager/
- Documentation
  - http://www.suse.com/documentation/suse_manager/index.html
- Wiki
  - http://wiki.novell.com/index.php/SUSE_Manager

Resources

- Product website
  www.suse.com/products/systemz
- Customer References
  www.suse.com/success → extended search for SUSE Linux Enterprise Server for System z

- Download SUSE Linux Enterprise Server for System z
  www.suse.com/products/server/eval.html
- Promotion Website
  www.novell.com/products/systemz/els.html
- Partner Website
  www.suse.com/mainframe
- Starter System for System z
  www.suse.com/partner/ibm/mainframe/startersystem.html

SUSE® Linux Enterprise

Documentation and Release Notes

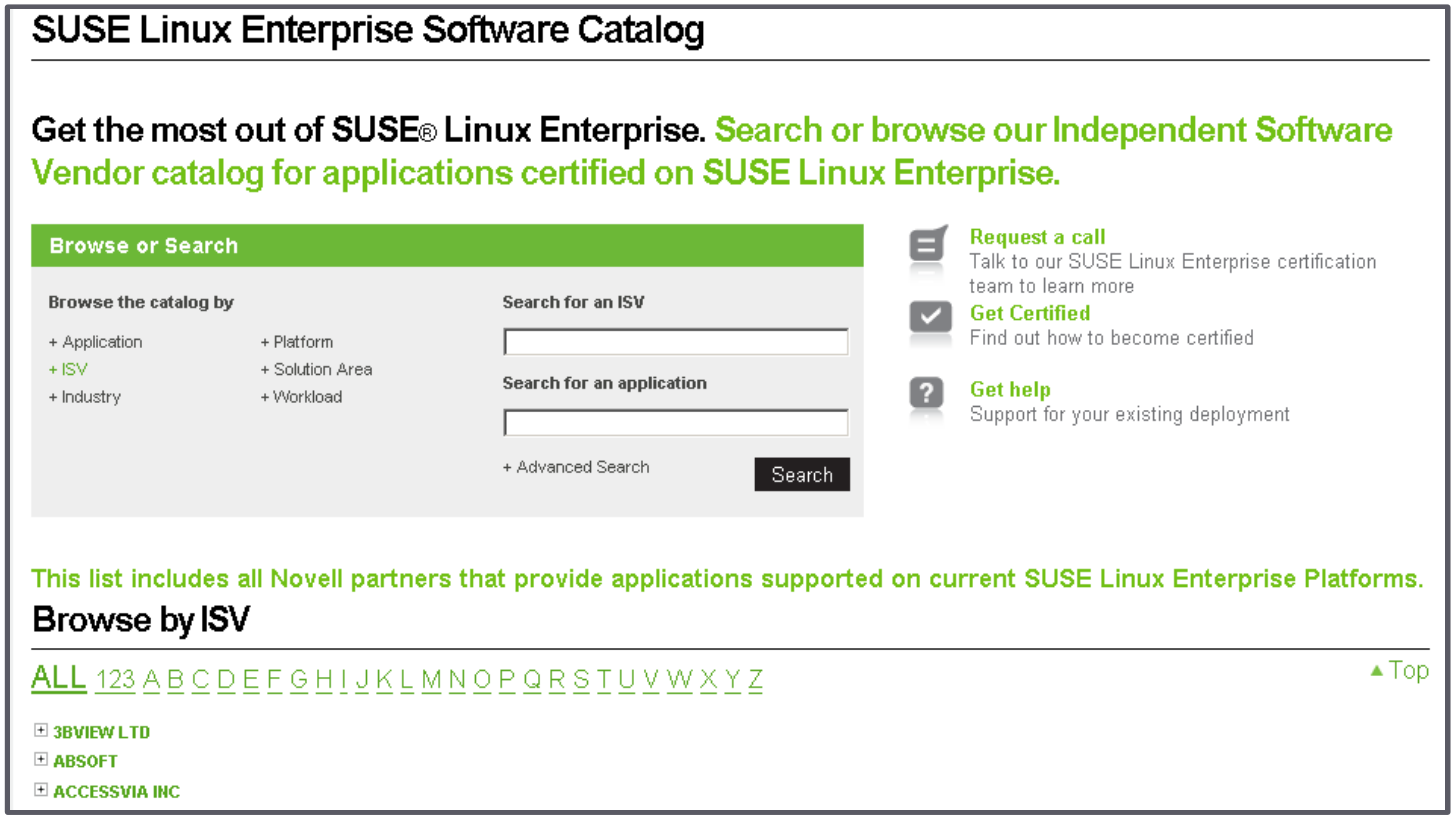
- Product Pages
  - http://www.suse.com/products/server/
  - http://www.suse.com/products/highavailability/
- Release Notes
  - http://www.novell.com/linux/releasenotes/
- Product Life-cycle
  - http://support.novell.com/lifecycle/linux.html
- ISV Partner Program
  - suse_isv@suse.com
- Unix to Linux Migration
  - http://www.novell.com/linux/unixtolinux/
- Documentation
  - http://www.novell.com/documentation/suse.html

Resources

- SUSE Linux Enterprise Server and IBM zEnterprise
  http://www.novell.com/docrep/2010/11/suse_linux_enterprise_server_and_ibm_zenterprise_system.pdf
- zBX entitlement for SUSE Linux Enterprise Server offering
  http://www.suse.com/promo/zbx.html
- SUSE Linux Enterprise Server for System z
  http://www.suse.com/products/systemz/
- IBM zEnterprise Success Story: Sparda-Datenverarbeitung eG
  http://www.novell.com/success/sparda.html
- Chalk Talk: Server consolidation on IBM System z
  http://www.novell.com/media/content/chalktalk-server-consolidation-on-system-z.html
- SUSE Manager
  http://www.suse.com/products/suse-manager
- SUSE Studio
  http://www.susestudio.com

How to Find Certified Applications

Visit our catalog:
www.suse.com/partner/isv/isvcatalog



SUSE® Linux Enterprise 11 SP2

Kernel Capabilities

SLE 11 SP 1 (2.6.32)	x86	ia64	x86_64	s390x	ppc64
CPU bits	32	64	64	64	64
max. # logical CPUs	32	up to 4096	up to 4096	64	up to 1024
max. RAM (theoretical/practical	64/ 16 GiB	1 PiB/ 8+ TiB	64 TiB/ 16TiB	4 TiB/ 256 GiB	1 PiB/ 512 GiB
max. user-/ kernelspace	3/1 GiB	2 EiB/ϕ	128 TiB/ 128 TiB	ϕ/ϕ	2 TiB/ 2 EiB
max. swap space	up to 31 * 64 GB				
max. #processes	1048576				
max. #threads per process	tested with more than 120000; maximum limit depends on memory and other parameters				
max. size per block device	up to 16 TiB	and up to 8 EiB on all 64-bit architectures			

Supported on certified hardware only

SUSE® Linux Enterprise 11 SP2

Filesystems

Feature	Ext 3	reiserfs	XFS	OCFS 2	btrfs
Data/Metadata Journaling	•/•	○/•	○/•	○/•	N/A [3]
Journal internal/external	•/•	•/•	•/•	•/○	N/A
Offline extend/shrink	•/•	•/•	○/○	•/○	•/•
Online extend/shrink	•/○	•/○	•/○	•/○	•/•
Inode-Allocation-Map	table	u. B*-tree	B+-tree	table	B-tree
Sparse Files	•	•	•	•	•
Tail Packing	○	•	○	○	•
Defrag	○	○	•	○	•
ExtAttr / ACLs	•/•	•/•	•/•	•/•	•/•
Quotas	•	•	•	•	•
Dump/Restore	•	○	•	○	○
Blocksize default	4KiB				
max. Filesystemsize [1]	16 TiB	16 TiB	8 EiB	4 PiB	16 EiB
max. Filesize [1]	2 TiB	1 EiB	8 EiB	4 PiB	16 EiB
Support Status	SLES	SLES	SLES	SLE HA	SLES

SUSE® Linux Enterprise was the first enterprise Linux distribution to support journaling filesystems and logical volume managers back in 2000. Today, we have customers running XFS and ReiserFS with more than 8TiB in one filesystem, and the SUSE Linux Enterprise engineering team is using our 3 major Linux journaling filesystems for all their servers. We are excited to add the OCFS2 cluster filesystem to the range of supported filesystems in SUSE Linux Enterprise. For large-scale filesystems, for example for file serving (e.g., with with Samba, NFS, etc.), we recommend using XFS. (In this table "+" means "available/supported"; "-" is "unsupported")

[1] The maximum file size above can be larger than the filesystem's actual size due to usage of sparse blocks. It should also be noted that unless a filesystem comes with large file support (LFS), the maximum file size on a 32-bit system is 2 GB (2^{31} bytes). Currently all of our standard filesystems (including ext3 and ReiserFS) have LFS, which gives a maximum file size of 2^{63} bytes in theory. The numbers given in the above tables assume that the filesystems are using 4 KiB block size. When using different block sizes, the results are different, but 4 KiB reflects the most common standard.

[2] 1024 Bytes = 1 KiB; 1024 KiB = 1 MiB; 1024 MiB = 1 GiB; 1024 GiB = 1 TiB; 1024 TiB = 1 PiB; 1024 PiB = 1 EiB (see also http://physics.nist.gov/cuu/Units/binary.html)

[3] Btrfs is a copy-on-write logging-style file system, so rather than needing to journal changes before writing them in-place, it writes them in a new location, and then links it in. Until the last write, the new changes are not "committed."

[4] Btrfs quotas will operate differently than traditional quotas. The quotas will be per-subvolume rather than operating on the entire filesystem at the user/group level. They can be made functionally equivalent by creating a subvolume per- user or group.

z196 enhanced node affinity support

Fate 311860 / [LTC 66807]

http://www.ibm.com/developerworks/linux/linux390/ -> affinity

Device Drivers, Features, and Commands (Kernel 2.6.38) – Chapter 27, p.292

- **CPU node affinity support for z196:** allowing the Linux kernel scheduler to optimize its decisions based on the z196 processor, cache and book topology.

- **Customer benefit**

technical	business
• Hardware exploitation z196 processor topology and cache hierachy, increase cache hit ratio and therefore overall performance	• Increase of application workload density per system • Increased performance

SLES	10	11
GA	-	-
SP1	-	-
SP2+3	-	yes
SP4	-	n/a

Spinning mutex performance enhancement

Fate 312075 / [LTC 70029]

http://www.ibm.com/developerworks/linux/linux390/kernel-2.6.38.html → spinning mutex

- **Performance:** The status of a thread owning a locked mutex is examined and waiting threads are not scheduled unless the first is scheduled on a virtual *and* physical processor.

- **Customer benefit**

technical	business
• New sophiticated handling of mutexes and scheduler decisions to improve performance also for z/VM based workloads	• Performances benefits for workloads making usage of parallel processing in an SMP environment of virtual CPUs

SLES	10	11
GA	-	-
SP1	-	-
SP2+3	-	**yes**
SP4	-	n/a

cio resume handling for reordered devices

Fate 311876 / [LTC 66907]

Device Drivers, Features, Commands on SUSE Linux Enterprise Server 11 SP2, p.368

- **Usability:** Improves cio resume handling to cope with devices that were attached on different subchannels prior to the suspend operation.

- **Customer benefit**

technical	business
• If the subchannel changes for disk device, the configuration is changed to reflect the new subchannel. This change is accomplished without de-registration. Device name and device configuration are preserved.	• Optimized or no downtime when resuming a Linux instance

SLES	10	11
GA	-	-
SP1	-	-
SP2+3	-	**yes**
SP4	-	n/a

FICON DS8k support – solid state disk flag

Fate 311756 / [LTC 60095]

http://www.ibm.com/developerworks/linux/linux390/s390-tools-1.8.2.html → solid

- **Solid State Drive support:** transparent to the DASD device driver, no change is needed to use solid state disks. A new flag in the device characteristics will show the administrator if a device is a solid state disk.

- **Customer benefit**

technical	business
• Storage servers can be queried if they provide solid state disks • Device characteristics are already exported per ioctl and can be read as binary data with the dasdview tool.	• Workloads can be placed on storage which support best their I/O characteristics • Acceleration of random I/O, cost effective placement of sequential I/O

SLES	10	11
GA	-	-
SP1	-	-
SP2+3	-	**yes**
SP4	-	n/a

Access to raw ECKD data from Linux (DASD)

Fate 311973 / [LTC 66951]

- **Interoperabilty:** allows to access ECKD disks in raw mode. Linux dd command can copy the disk level content of an ECKD disk to a Linux file, and vice versa. Works independent of the operating system or file system that is on the ECKD disk.

- **Customer benefit**

technical	business
• Includes Linux ECKD disks used with LVM, Linux ECKD disks that are used directly, and z/OS ECKD disks	• Use case for Linux by eliminating the need for data transfers from z/OS to Linux via network

SLES	10	11
GA	-	-
SP1	-	-
SP2+3	-	**yes**
SP4	-	n/a

Store I/O Operation Status and initiate logging (SIOSL)

Fate 311917 / [LTC 66847]

http://www.ibm.com/developerworks/linux/linux390/kernel-2.6.36.html

- **Description:** interface for the store-I/O-operation-status-and-initiate-logging (SIOSL) CHSC command and its exploitation by the FCP device driver

- **Customer benefit**

technical	business
• Can be used to synchronize log gathering between the operating system and the channel firmware.	• Concurrent data collection for problem resolution, minimizing customer operation impact

SLES	10	11
GA	-	-
SP1	-	-
SP2+3	-	**yes**
SP4	-	n/a

Intuitive dump device configuration

Fate 304024 / [LTC 201624]

- **Description:** provide a yast dialog to prepare I/O devices for dump, during the installation and post-installation.

- **Customer benefit**

technical	business
• Guided setup and configuration of a suitable dump device	• Improved serviceability

SLES	10	11
GA	-	-
SP1	-	-
SP2+3	-	**yes**
SP4	-	n/a

More features ...

- About ~ 80 more z specific features included in SP2
- Contact Marcus Kraft <mkraft@suse.de> to request a full and detailed list