SUSE® Linux Enterprise Server

IBM System z – Current & Future Features

Session 13602

Mike Friesenegger
System Engineer, SUSE
MFriesenegger@suse.com

Ihno Krumreich
Project Manager, SUSE
ihno@suse.com

SUSE®
We adapt. You succeed.

Date: 2013-08-02
Distribution: PDF any

SUSE and the Attachmate Group

- **SUSE**, headquartered in Nürnberg / Germany, is an independently operating business unit of The Attachmate Group, Inc.

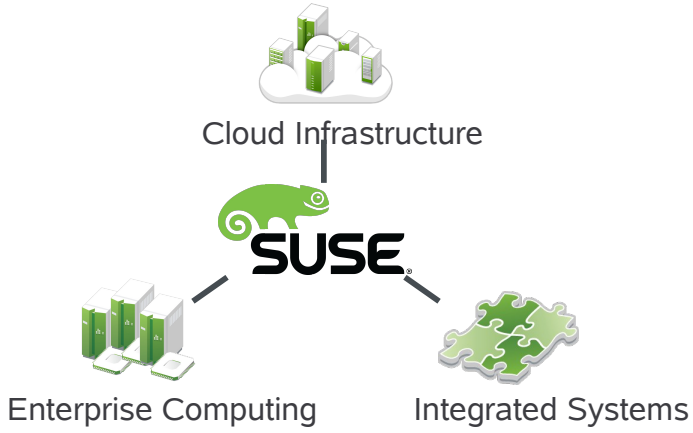
- The Attachmate Group is a privately held 1 billion+ $ revenue software company with four brands:

SUSE

Attachmate

Novell

NetIQ



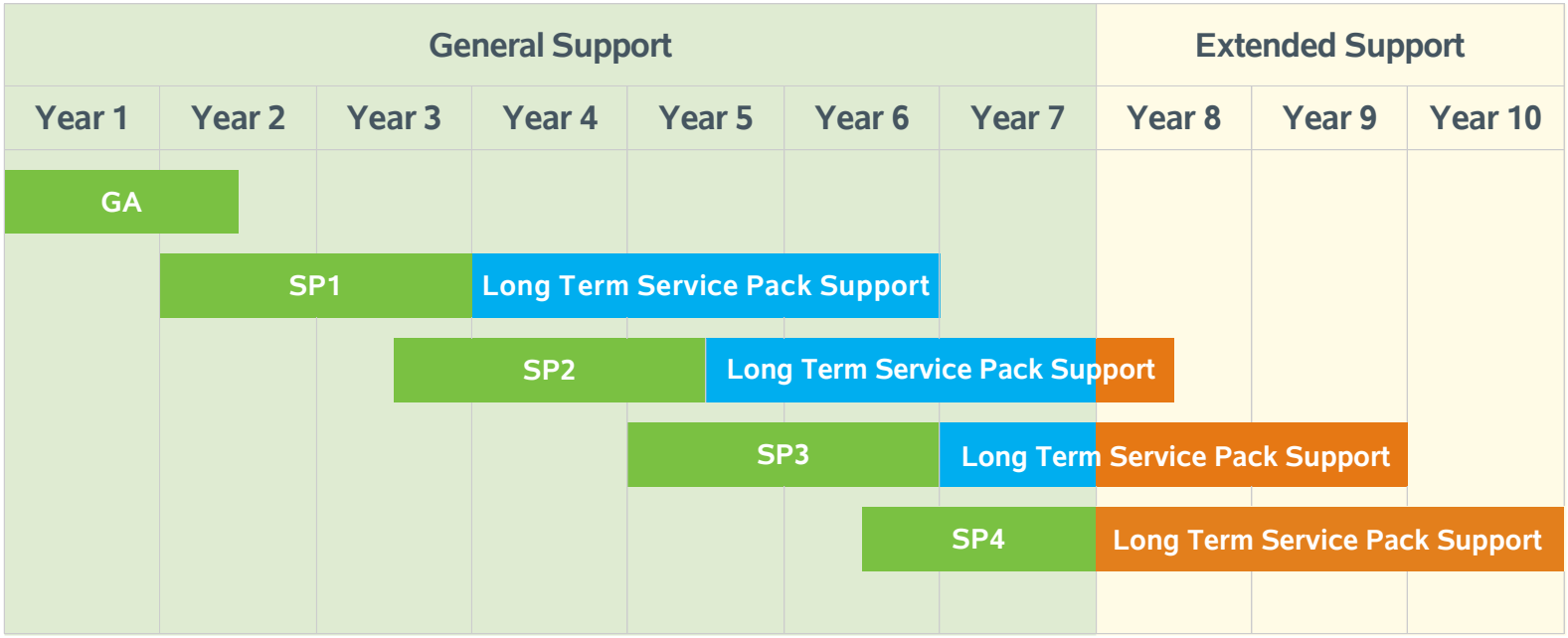
SUSE

SUSE Linux Enterprise Server

A highly reliable, scalable and secure server operating system, built to power physical, virtual and cloud-based mission-critical workloads.

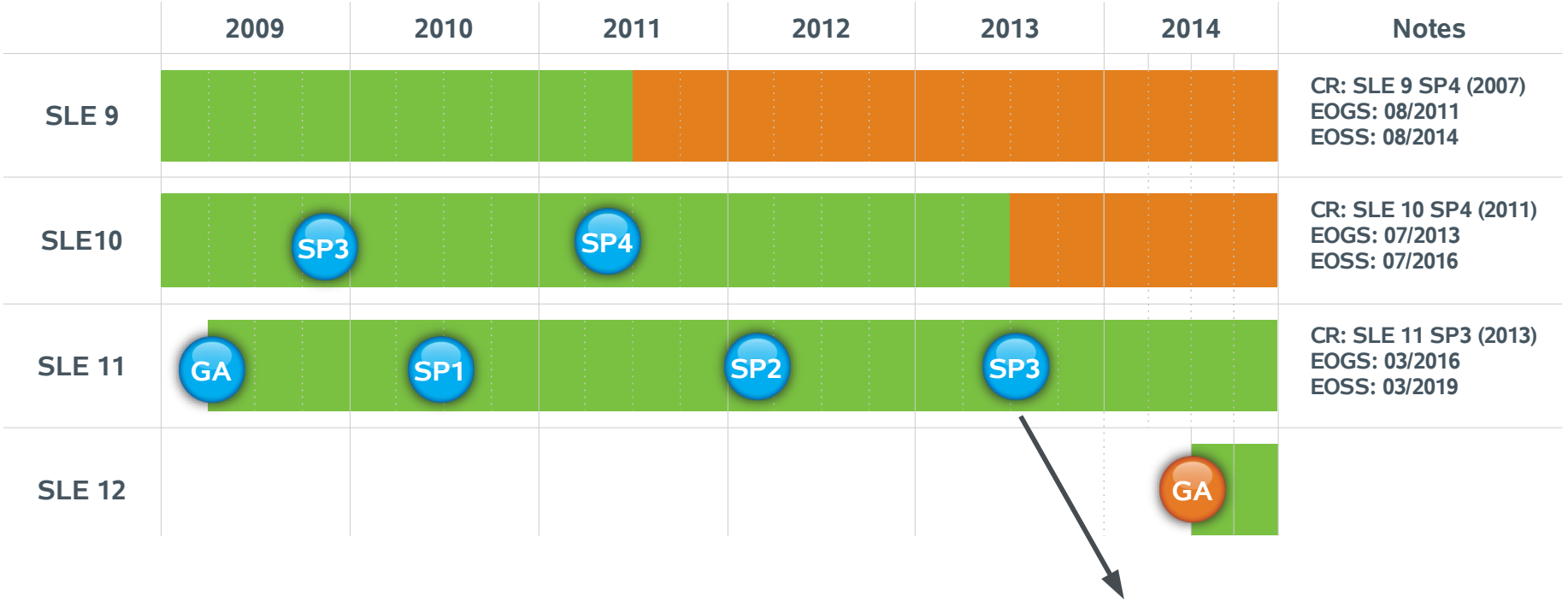
Lifecycle & Support

Generic Product Lifecycle



- 10-year lifecycle (7 years general support, 3 years extended support)
- Service Packs are released every ~18 months
  - 5 years lifetime with
  - ~2 years general support per Service Pack
  - 6 month upgrade window after release of the next Service Pack
- Long Term Service Pack Support (LTSS) option
  - Extend upgrade window or extend major release lifecycle

Current SUSE® Linux Enterprise Streams



- Dependable release timing
- Predictability for planning rollouts and migrations
  - Service Pack releases, development and product schedules announced to customers and partners
- Major releases every 4-5 years.

Service Pack Migration

Systems Management Today

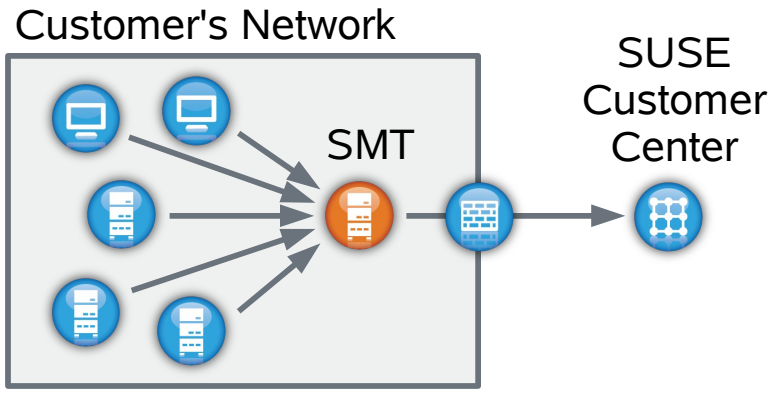
- YaST – unique, highly integrated local management tool
  - Ease of use, effective learning curve; reduces training efforts
  - Automation via AutoYaST for data center mass deployments
- Fastest open source update stack (ZYpp)
  - Reduce management time, effort and cost
  - Improve reliability and availability by reducing downtimes
  - ZYpp handles multiple installed package versions (e.g. Kernel)
- Build in Installation Server
  - Easy setup, allows for internal high speed repository serving
  - Allows to speed up and automated release and SP migrations
  - Can be combined with SMT to serve multiple SUSE products

SUSE® Linux Enterprise 11

Systems Management Today

- Unattended migration reduces cost and downtime
  - SUSE Linux Enterprise 10 SP4 to SUSE Linux Enterprise 11 SP3
  - SUSE Linux Enterprise 11 SP2 to SUSE Linux Enterprise 11 SP3
- Example: http://www.suse.com/documentation/sles11/book_sle_deployment/?page=/documentation/sles11/book_sle_deployment/data/cha_update_auto.html

- Subscription Management Tool
  - SMT is a proxy and auditing tool that mirrors the Customer Center
  - Tightly integrates with it
  - Accurately registers and manages deployments
  - Guarantees subscription compliance
  - Secures IT process flow



Repository & Channels SLES 11 SP2

Only a few are mandatory channels

```
# zypper lr -u
#  | //  | Name                              | Enabled | Refresh |
---+-----+-----------------------------------+---------+---------+
 1 | //  | SLES11-SP1-Pool                   | Yes     | Yes     |
 2 | //  | SLES11-SP1-Updates                | Yes     | Yes     |
 3 | //  | SLES11-SP2-Core                   | Yes     | Yes     |
 4 | //  | SLES11-SP2-Updates                | Yes     | Yes     |
 5 | //  | SLE11-SP1-Debuginfo-Pool          | No      | Yes     |
 6 | //  | SLE11-SP1-Debuginfo-Updates       | No      | Yes     |
 7 | //  | SLE11-SP2-Debuginfo-Core          | No      | Yes     |
 8 | //  | SLE11-SP2-Debuginfo-Updates       | No      | Yes     |
 9 | //  | SLE11-WebYaST-SP2-Pool            | No      | Yes     |
10 | //  | SLE11-WebYaST-SP2-Updates         | No      | Yes     |
11 | //  | SLES11-Extras                     | No      | Yes     |
12 | //  | SLES11-SP2-Extension-Store        | Yes     | Yes     |
```

Required SLES 11 SP2 channels for installation and updates.

- SLES11-SP1-Pool — static, copied media packages
- SLES11-SP1-Updates — receives updates related to SLES11-SP1-Pool
- SLES11-SP2-Core — static, SP2 packages not present in SLES11-SP1-Pool
- SLES11-SP2-Updates — receives updates for SP2

- Note: removing these channels disables receiving updates for SP2.

Repository & Channels SLES 11 SP3

Only a few are mandatory channels

```
# zypper lr -u
#  | // | Name                            | Enabled | Refresh |
---+----+---------------------------------+---------+---------+
 1 | // | SLES11-SP1-Pool                 | No      | Yes     |
 2 | // | SLES11-SP1-Updates              | No      | Yes     |
 3 | // | SLES11-SP2-Core                 | No      | Yes     |
 4 | // | SLES11-SP2-Updates              | No      | Yes     |
 5 | // | SLES11-SP3-Pool                 | Yes     | Yes     |
 6 | // | SLES11-SP3-Updates              | Yes     | Yes     |
 7 | // | SLE11-SP1-Debuginfo-Pool        | No      | Yes     |
 8 | // | SLE11-SP1-Debuginfo-Updates     | No      | Yes     |
 9 | // | SLE11-SP2-Debuginfo-Core        | No      | Yes     |
10 | // | SLE11-SP2-Debuginfo-Updates     | No      | Yes     |
11 | // | SLE11-SP3-Debuginfo-Pool        | No      | Yes     |
12 | // | SLE11-SP3-Debuginfo-Updates     | No      | Yes     |
13 | // | SLES11-Extras                   | No      | Yes     |
14 | // | SLES11-SP2-Extension-Store      | No      | Yes     |
15 | // | SLES11-SP3-Extension-Store      | No      | Yes     |
```

Required SLES 11 SP3 channels for installation and updates.

- SLES11-SP3-Pool — static, copied media packages
- SLES11-SP3-Updates — receives updates related to SLES11-SP1-Pool
- *All other channels are configured with the system for convenient activation*

- Note: removing the Pool and Updates channels disables receiving updates for SP3.

SLES for System z

Unique Tools Included

- YaST2 systems management
  - Install, deploy, and configure every aspect of the server
- AppArmor Security Framework
  - Application confinement
- Subscription Management Tool
  - Subscription and patch management, proxy/mirroring/staging
- Starter System for System z
  - A pre-built installation server, deploy with z/VM tools

- ***Free*** High Availability Extension
  - Cluster Framework, Cluster FS, DRBD, (GEO-cluster*)
- Brtfs – next gen file system
  - Scalability, plus snapshot & rollback options

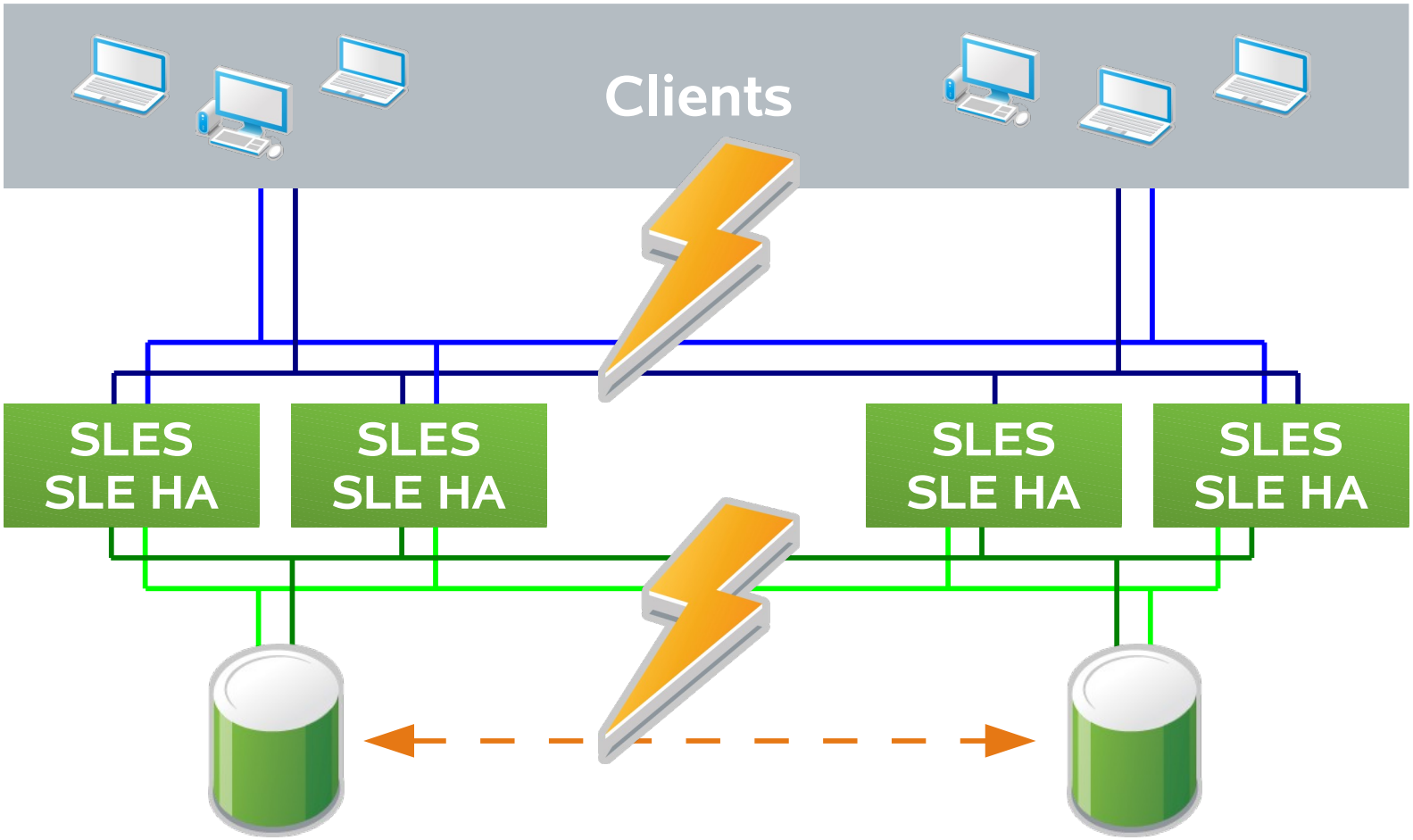
High Availability

Reasons for

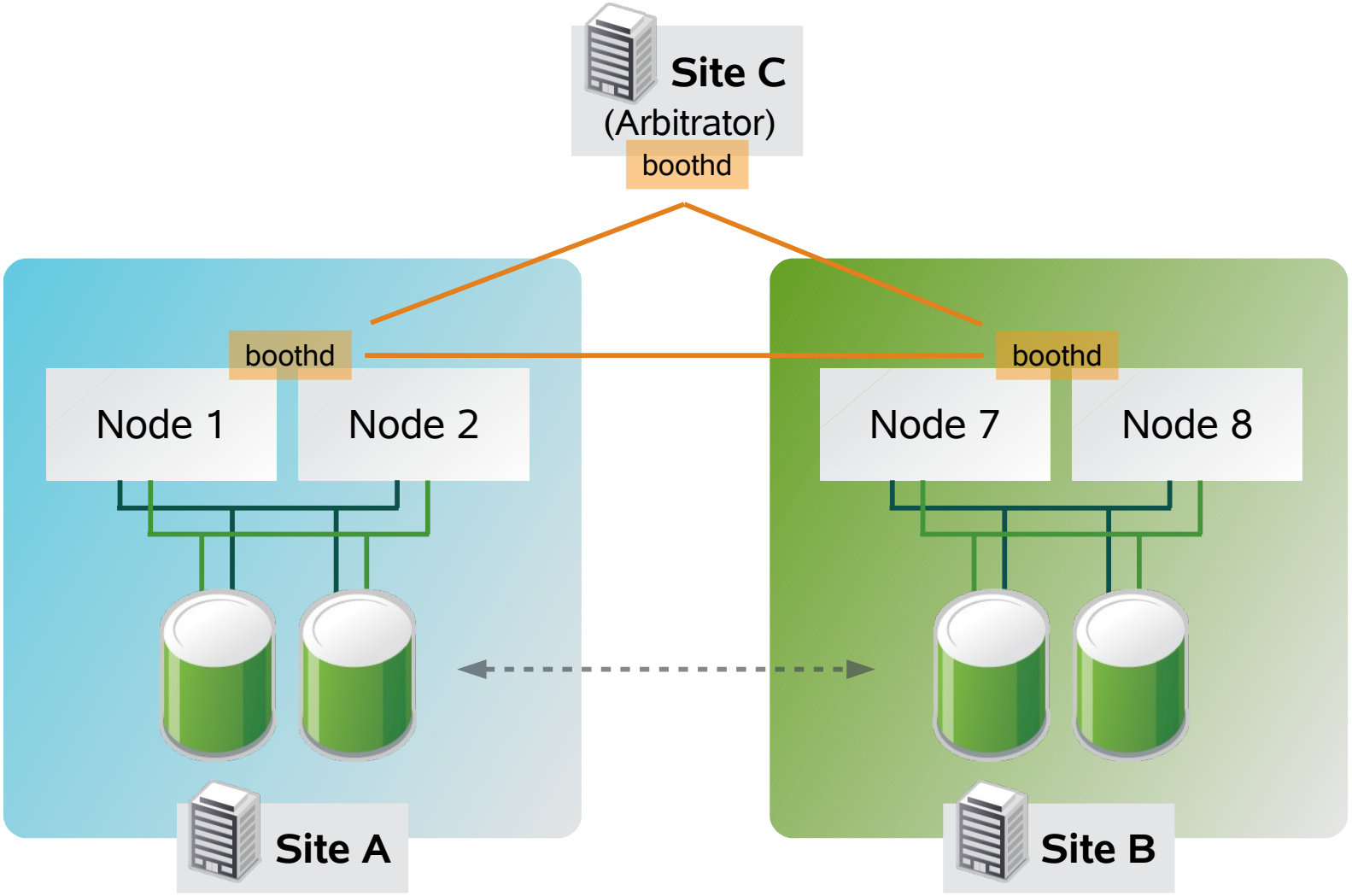
SUSE® Linux Enterprise High Availability

- Long history track record
- Most up-to-date Open Source High Availability stack
- Deep OS integration
- Ready for Virtualization
- Integrated Data Replication
- Superior Cluster File System
- Geo cluster support

Local & Stretched Cluster



Geo Cluster – Setup



btrfs

Btrfs (better fs) – Features

- Scalability (16 EiB) including effective shrink
- Supports offline in-place migration from ext2, ext3
- Support for Copy on Write
- Powerful Snapshot capabilities
- Other Capabilities:
  - SSD optimization (TRIM support)
  - Data integrity (checksums)

SUSE Linux Enterprise 11 SP3

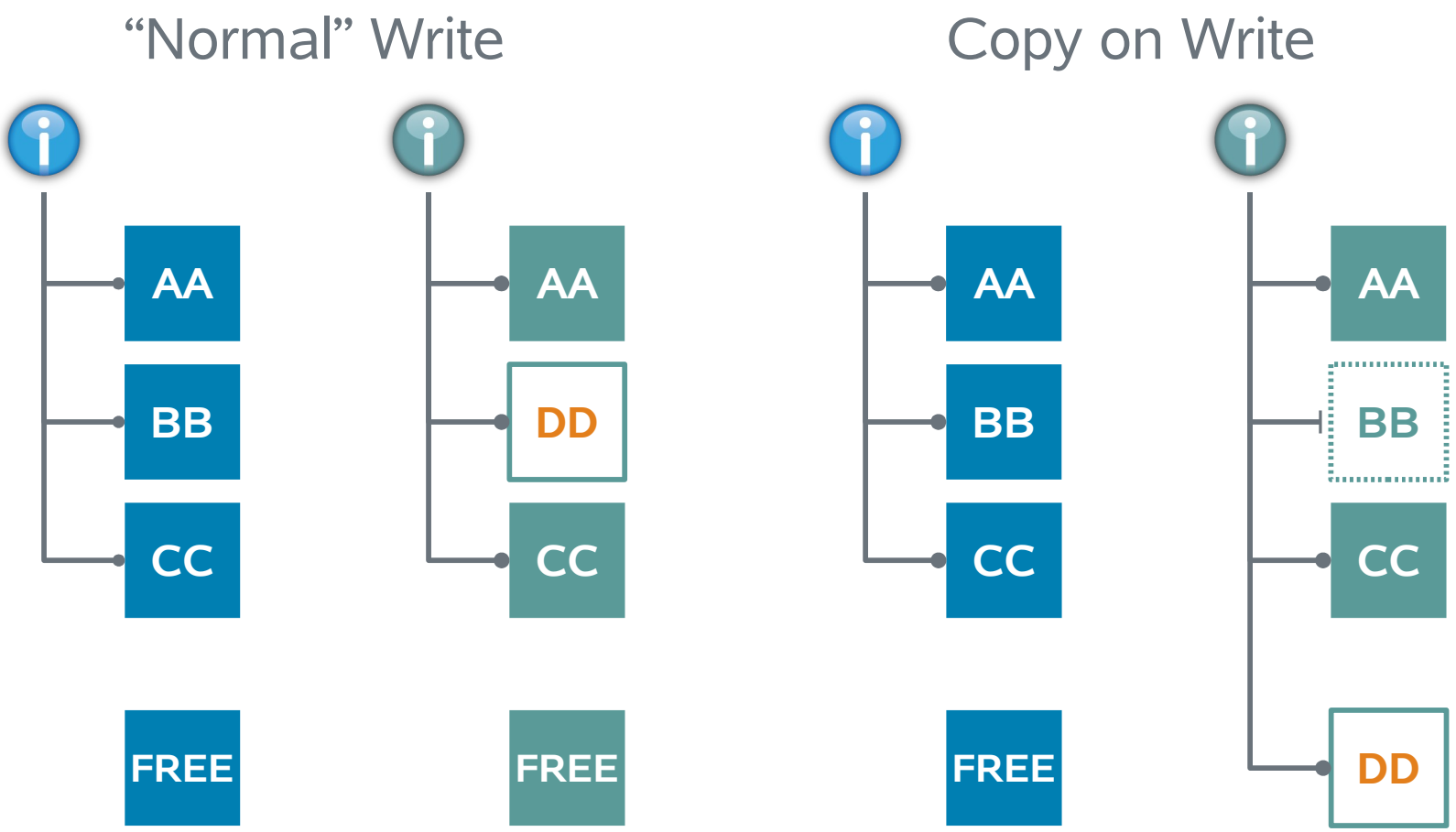
Btrfs Support

- btrfs is supported
  - As “/” filesystem
  - For migration from ext2/3/4, except “/”
  - Snap shots
  - Subvolume quota support
  - Everything you can do with the YaST partitioner
- Exceptions (not supported)
  - “/boot”
  - RAID, Integrated Volume Management
  - Compression and Encryption
- Recommendation for data volumes: xfs
  - Performance and scalability are proven for 10+years
- Full text in the release notes:

  http://www.suse.com/releasenotes/s390x/SUSE-SLES/11-SP3 → btrfs

Technology Overview

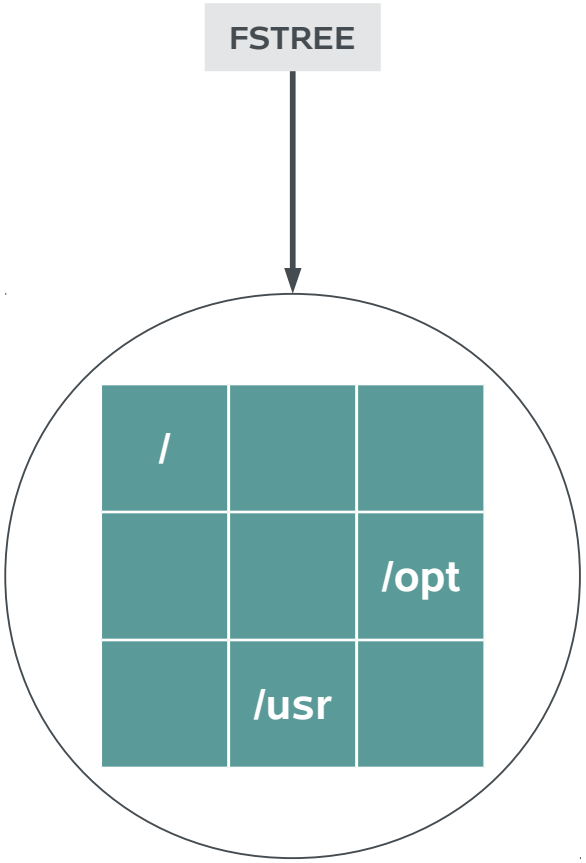
Copy on Write



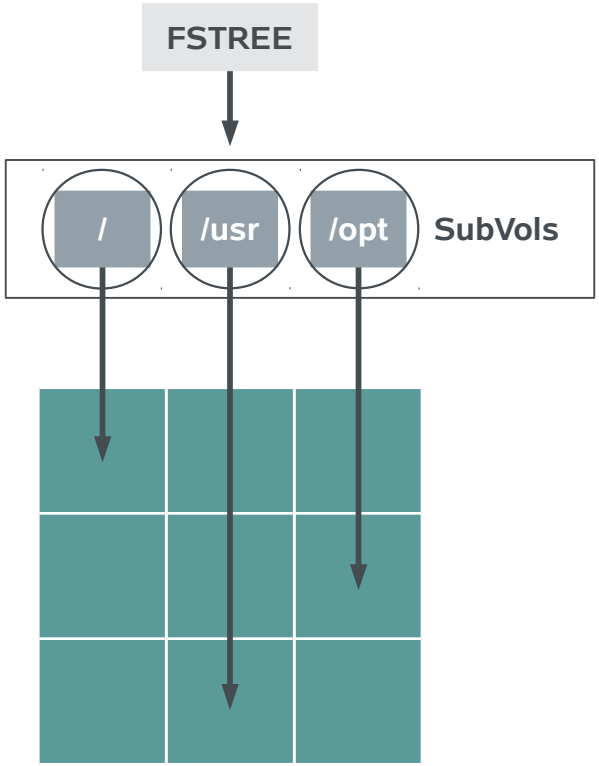
Technology Overview

Subvolume

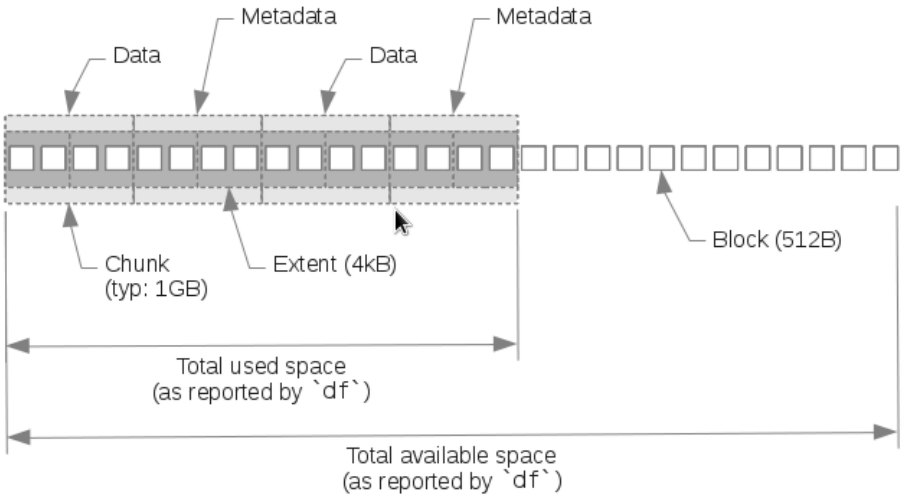
Normal Filesystem



With Subvolumes



Btrfs Disk Space And Extents



In case of a Btrfs filesystem on a single underlying block device

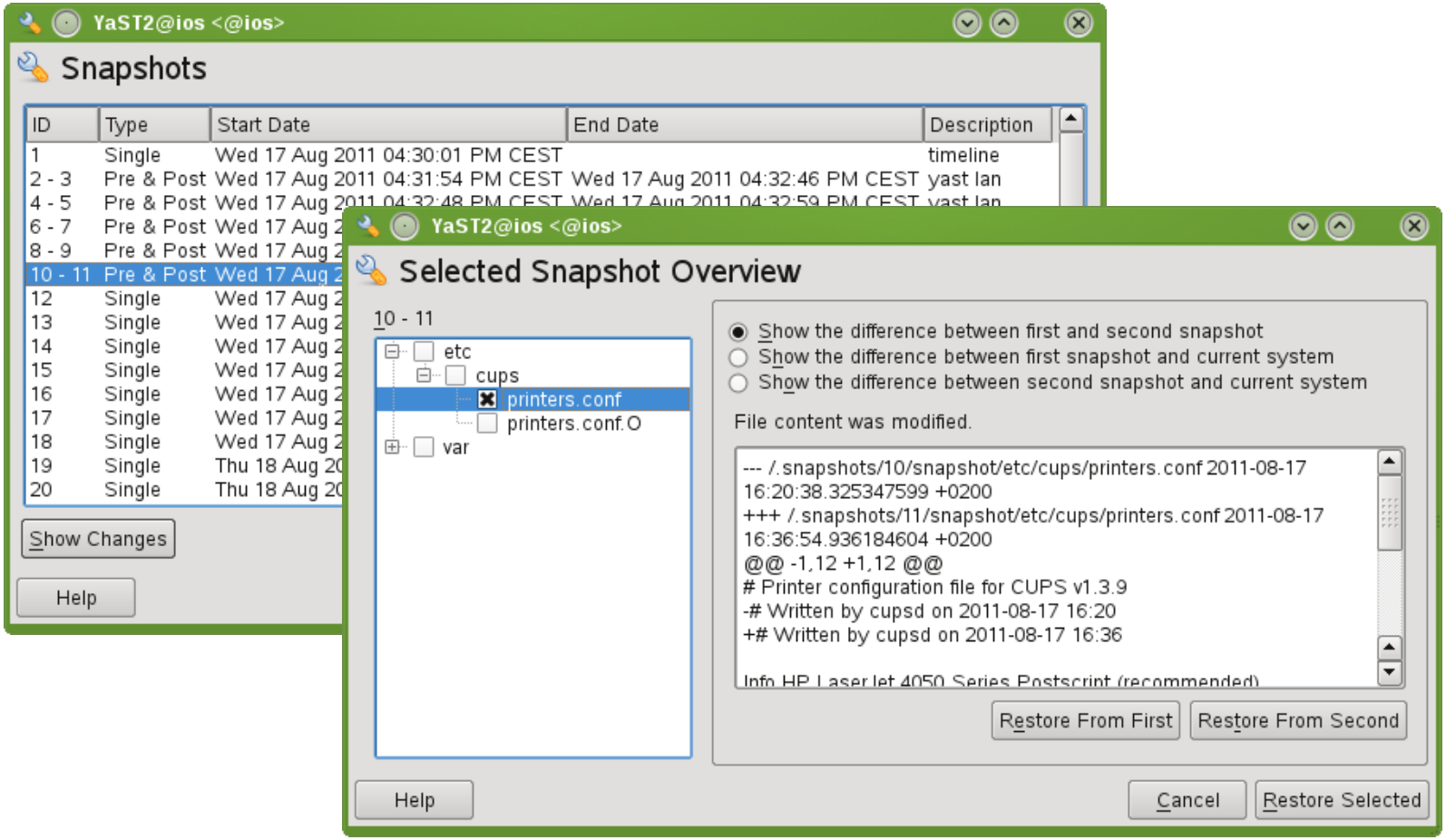
```
# btrfs filesystem df /
Data: total=14.50GB, used=12.20GB
System, DUP: total=8.00MB, used=12.00KB
System: total=4.00MB, used=0.00
Metadata, DUP: total=1.75GB, used=904.11MB

# df -h /
Filesystem      Size  Used Avail Use% Mounted on
/dev/sda7        20G   14G  4.3G  77% /
#
```

Disk utilization
12,2GB + 2x 0,9GB + = 14 GB

Snapshots in SUSE® Linux Enterprise 11 SP3

YaST2 Management



Session 13457: Alternatives To Solaris Containers And ZFS For Linux On System z

SLES for System z 11 SP3

SUSE® Linux Enterprise Server for System z 11 SP3

- zEC12 + zBX = IBM zEnterprise exploitation continued
  - **zBC12, z/VM 6.3,** zBX HX5 support (blade center extension)
  - z9 EC, z10 EC, z196 EC, z9 BC, z10 BC, z114 BC support
  - Java 7 and supportive kernel enhancements
  - Flash Express SC Memory support (/dev/scm)
  - GCC 4.7 for applications targeting zEC12 processor
- Improved RAS tools and System z specific support
  - 2 stage dump & network storage sharing with compression
  - Robust disk mirroring for large pools of DASDs (MD RAID10)
  - Enhanced DASD statistics for PAV & HPF
  - IUCV terminal server client & server setup support
  - s390-tools update

zEC12 Exploitation

- Kernel support to improve Java performance (Transactional Execution)
  - Middleware & applications using Java will benefit
  - Inclusion of latest Java 7
- Storage class memory – Flash Express
  - Support new storage device: /dev/scm
  - IPL save 'disk storage' with low latency and high throughput
- Support for Crypto Express 4S cards
- Leverage Cross Memory Attach Functionality
  - Speedy middleware data exchange via shared main storage
- Backport GCC 4.7.x patches (SDK)
  - Add new instructions to the compiler (z196, zEC12)
  - Added new pipeline description to generate optimized code

Enhanced Dump Capabilities

- Two Stage Dumper framework
  - More flexible and efficient handling of dumps
- Compression of kernel dumps
  - More efficient use of disk storage, lower space requirements
- Fuzzy live dump
  - Extract current memory state of the kernel for debugging
- Allow to compare dump system with boot system
  - Did the dump occurred on the system it was IPLed ?
- Add option to mkdumprd to clean up older initrd's
  - Dump and initrd handling in /boot
- FICON DASD sanity check
  - Detect path connection errors

Misc

- Enhanced DASD statistics for PAV and HPF
  - Improved diagnosis and analysis
  - Supports recommendations on the use of eg aliases
- Optimized compression library zlib
  - Enhanced to speed up Java, report generation, backup and installation
- ZYpp transaction auditing
  - Track transaction id also for client side
- libhugetlbfs support
  - Allow applications to benefit from hugetbls w/o recompile
- Enable larger shm segments than 256GB
  - Allows data bases to share larger areas

Technical Preview: KVM for s390x

- Kernel Based Virtual Machine
  - KVM (for Kernel-based Virtual Machine) is a virtualization solution for Linux on x86, POWER and z/Architecture hardware containing virtualization extensions.
  - It consists of a loadable kernel module, kvm.ko, that provides the core virtualization infrastructure and a processor specific module (eg. kvm-intel.ko or kvm-amd.ko)
  - KVM also requires a modified QEMU to connect to the I/O world of the hosting system.
  - Lowers the entry barrier for non-mainframe, but Linux skilled users to explore hardware and virtualization options of the mainframe

zEC12: Kernel support for Transactional Execution

Fate 314142 / [LTC 79946]

www.redbooks.ibm.com/redpieces/pdfs/sg248050.pdf → Transactional Execution (TX) facility

- **Description:** Allow user-space processes to use transactional execution (TX) and runtime instrumentation (RI). Supports hardware runtime instrumentation, that improves analysis of and optimization of the code generated by the new IBM JVM. Software locking overhead is minimized and scalability and parallelism is increased.
- **Customer benefit**

technical	business
• improved Java performance • useful for lockless data structures and speculative compiler optimizations offering increased scalability and parallelism to drive higher throughput	• Java based workload consolidation

SLES	10	11
GA	-	-
SP1+2	-	-
SP3	-	**yes**
SP4	-	n/a

Support up to 4 TiB of storage (main memory)

Fate 314338

http://www.vm.ibm.com/

https://www.suse.com/products/server/technical-information/ (needs update)

- **Description:** Max. 4 TB of storage is supported. This may however be reduced by limitations of the underlaying certified hardware. zEC12 offers up to 3.75 TiB of storage, supports up to 1TiB per LPAR, and z/VM 6.3 1TiB real memory, z/VM 6.2 256GB

  .

- **Customer benefit**

technical	business
• Larger amount of memory for applications • More LPARs with large amount of memory	• Larger consolidation scenarios • More workloads per physical server

SLES	10	11
GA	-	-
SP1+2	-	-
SP3	-	**yes**
SP4	-	n/a

Support for Flash Express Storage Class Memory

Fate 314095 / [LTC 79955]

Device Drivers, Features, and Commands on SUSE Linux Enterprise Server 11 SP3
http://public.dhe.ibm.com/software/dw/linux390/docu/les3dd03.pdf → Chap. 6, p97

- **Description:** Flash Express SCM is accessed as storage-class memory increments as block devices /dev/scm, which can help to improve the paging rate and access performance for temporary storage.

- **Customer benefit**

technical	business
• Improves the paging rate and access performance for temporary storage • Fast, low latency Linux disk storage • Data storage device are combining properties of both storage and memory	• Database acceleration (eg. transaction log file), datawarehouse performance

SLES	10	11
GA	-	-
SP1+2	-	-
SP3	-	**yes**
SP4	-	n/a

Robust disk mirroring for large pools of DASDs

Fate 311379

SUSE document: How to operate a disk mirrored target with SUSE

- **Description:** Refurbished MD RAID10. Improve storage operation to enable continuous operation even in case of a temporary DS8000/ESS failure or timeout. This is a disk mirroring solution with real-time enhancement, based on LVM2/Device-mapper, with Linear-Mirror and Stripe-Mirror support.
- **Customer benefit**

technical	business
• Migration paths from previous versions • High Availability to bridge temp timeouts	• Continuous storage availability

SLES	10	11
GA	-	-
SP1+2	-	-
SP3	-	**yes**
SP4	-	n/a

Two Stage Dump Framework

Fate 314079 / [LTC 74309]

Using the Dump Tools on SUSE Linux Enterprise Server 11 SP3
http://public.dhe.ibm.com/software/dw/linux390/docu/les3dt02.pdf

- **Description:** kdump can be used to create system dumps, reduces both dump time and dump size and facilitates dump disk storage sharing. A setup GUI is provided by YaST.

- **Customer benefit**

technical

- Dump time and size reduction
- Dump disk space sharing
- Improved easy of use
- Enhanced debugging capabilities, potential reduction for problem resolution time

business

- Improve serviceability
- Enable serviceability for customers that run huge images

SLES	10	11
GA	-	-
SP1+2	-	-
SP3	-	**yes**
SP4	-	n/a

Support for crypto hardware zEC12 Crypto Express4S

Fate 314097 / [LTC 79958]

http://www-03.ibm.com/systems/z/advantages/security/zec12cryptography.html
http://www-03.ibm.com/systems/z/hardware/zenterprise/zec12_specs.html → Crypto Express 4S
Device Drivers, Features, and Commands on SUSE Linux Enterprise Server 11 SP3
http://public.dhe.ibm.com/software/dw/linux390/docu/les3dd03.pdf → Chap 34, p319

- **Description:** z90crypt device driver supports the Crypto Express 4 (CEX4) adapter card, which represents the newest-generation cryptographic feature and is designed to complement the cryptographic capabilities of the CPACF.

- **Customer benefit**

technical	business
• New modes for DES, 3DES, AES	• Enhanced security

SLES	10	11
GA	-	-
SP1+2	-	-
SP3	-	**yes**
SP4	-	n/a

Tools / SDK

Developer Tools

Dynamic analysis tools

- valgrind
  - Cachegrind
  - Memcheck
  - 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%  )
```

zPDT

IBM System z Personal Development Tool

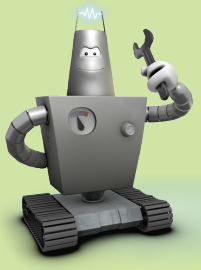
https://www.ibm.com/partnerworld/page/pw_com_zpdt

- 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 11+, SLES 11 SP2 x86_64, and others
  - SUSE's evaluation versions for x86_64 and s390x available at http://www.suse.com/products/server/eval.html


zEnterprise 196
z BladeCenter Extension

How to build a SUSE environment

BUILD
your workloads



SUSE Studio
Build workloads for any platform and the cloud

MANAGE
your environment

SUSE Manager
Manage Linux workloads across platforms

Come and see us at the booth.

Additional SUSE Sessions at SHARE Boston:

13437: How To Make Databases On Linux On System z High Available
13457: Alternatives To Solaris Containers And ZFS For Linux On System z

Thank you.

SUSE
We adapt. You succeed.

Appendix

SUSE to Go

Mobile Enablement App

ADownload from the iTunes App Store or Google Play or point your device to: www.suse.com/susetogo

SUSE® Linux Enterprise

Documentation and Release Notes

- Product Pages
  - http://www.suse.com/products/server/
  - http://www.suse.com/products/sles-for-sap/
  - http://www.suse.com/products/highavailability/
  - http://www.suse.com/products/realtime/
- Unix to Linux Migration
  - http://www.suse.com/solutions/enterprise-linux-servers/unixtolinux.html
- Documentation
  - http://www.suse.com/documentation/
- Release Notes
  - http://www.suse.com/releasenotes/

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
  https://www.suse.com/partners/alliance-partners/ibm/mainframe/zbx.html
- SUSE Linux Enterprise Server for System z
  http://www.suse.com/products/systemz/
- IBM zEnterprise Success Story
  https://www.suse.com/success/#suselinuxenterpriseserverforsystemz
- IBM System z Personal Development Tool (zPDT) Live Demo
  https://www.suse.com/media/content/ibm-system-z-personal-development-tool-zpdt-live-demo.html
- SUSE Manager
  http://www.suse.com/products/suse-manager
- SUSE Studio
  http://www.susestudio.com

Specs

SUSE® Linux Enterprise 11 SP3

Kernel Capabilities

SLE 11 SP 3	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/ 1.5 TiB
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 SP3

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.

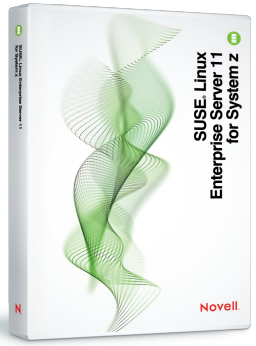
[2] 1024 Bytes = 1 KiB; 1024 KiB = 1 MiB; 1024 MiB = 1 GiB; 1024 GiB = 1 TiB; 1024 TiB = 1 PiB; 1024 PiB = 1 EiB (see also http://physics.nist.gov/cuu/Units/binary.html)

[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.

SUSE® Linux Enterprise Server

- SUSE Linux Enterprise Server 10/2000
- SUSE Linux Enterprise Server 7 08/2001
- SUSE Linux Enterprise Server 8 10/2002
- SUSE Linux Enterprise Server 9 08/2004
- SUSE Linux Enterprise Server 10 07/2006
- SUSE Linux Enterprise Server 11 03/2009
- SUSE Linux Enterprise Server 12 ~2014



Corporate Headquarters
Maxfeldstrasse 5
90409 Nuremberg
Germany

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

Join us on:
www.opensuse.org