#SHAREorg

SHARE
Technology · Connections · Results

Significant Enhancements in z/OS V1R13 zFS

Richard Theis
Jim Showalter
IBM

Tuesday, March 13, 2012
Session 10625

Trademarks

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

- DFS
- DFSMS
- DFSMSdss
- IBM
- MVS
- RACF
- RMF
- S/390
- z/OS
- zSeries

* Registered trademarks of IBM Corporation

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 Linus Torvalds in the United States, other countries, or both.
Microsoft, Windows and Windows NT are registered trademarks of Microsoft Corporation.
UNIX is a registered trademark of The Open Group in the United States and other countries.
SET and Secure Electronic Transaction are trademarks owned by SET Secure Electronic Transaction LLC.

* 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.

Agenda

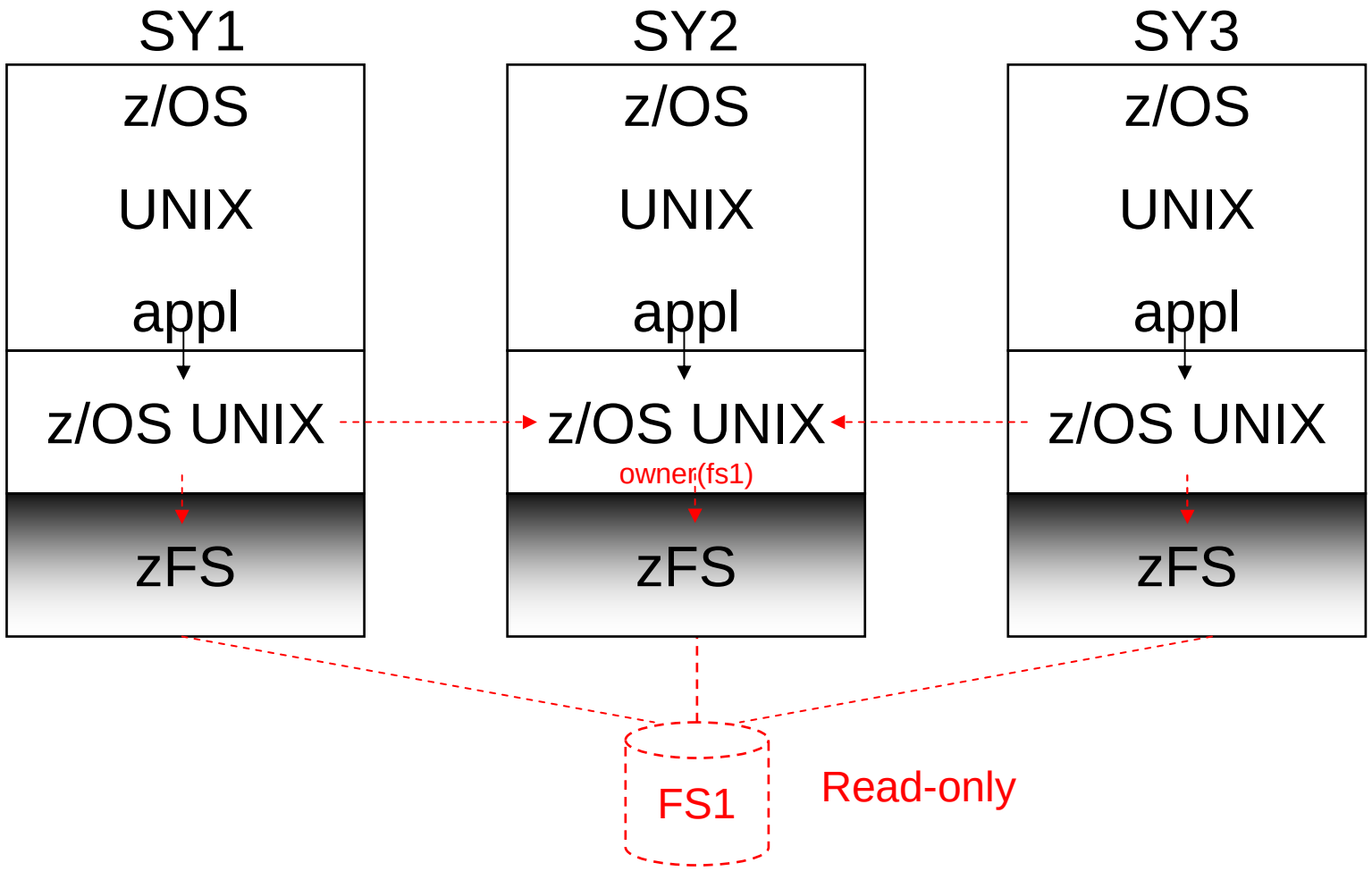
- Review of current shared file system environment support
- New support in zFS R13
  - zFS Direct I/O
  - zFS Internal Restart
  - zFS Automatic Re-enablement of disabled aggregates

Review of existing support in a shared file system environment

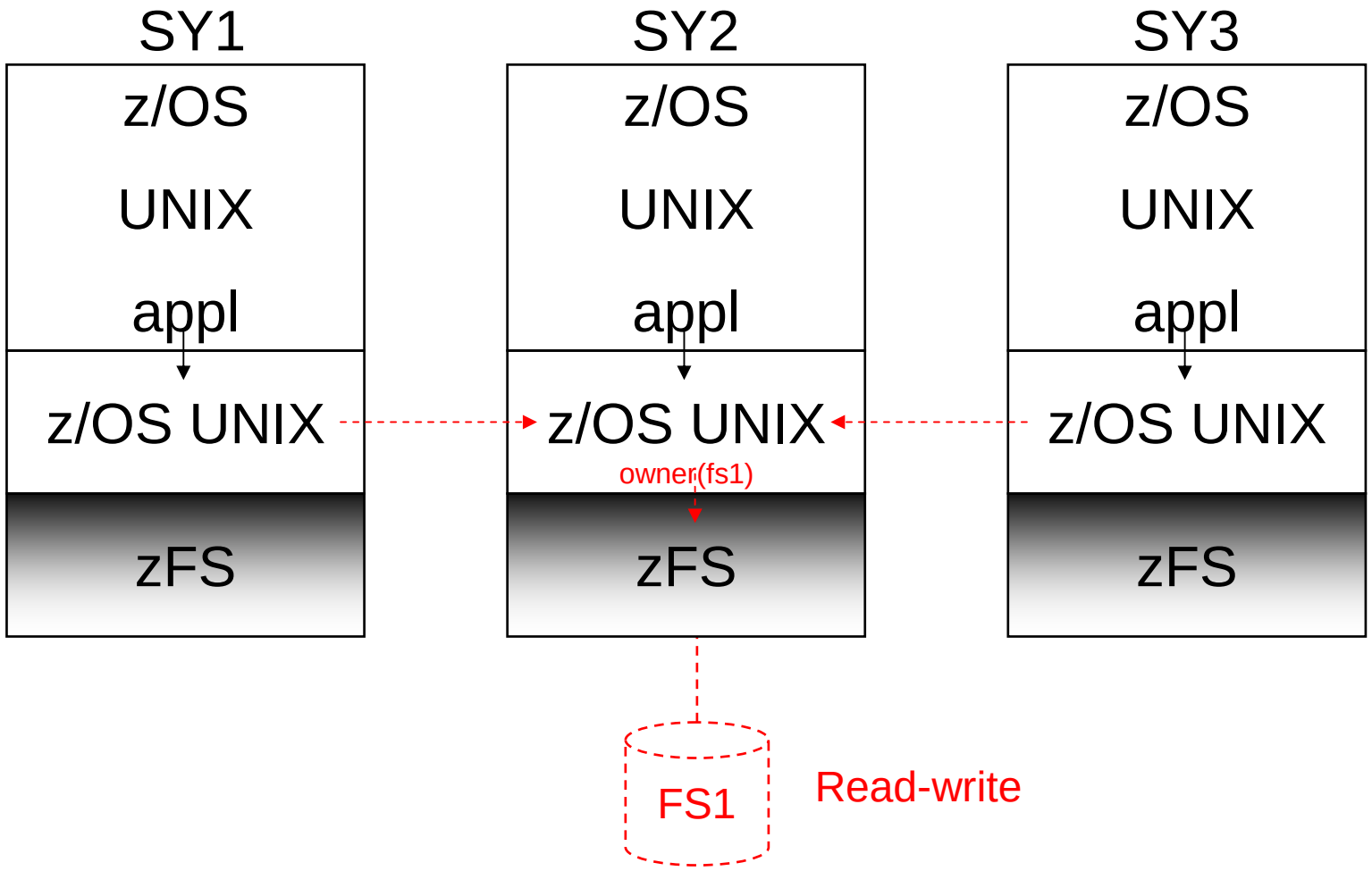
z/OS UNIX Shared File System Environment

- A shared file system environment means (at least)
  - Multiple systems with shared DASD and XCF communications between the systems
  - An IEASYSxx member that specifies PLEXCFG=MULTISYSTEM
  - A z/OS UNIX Couple Data Set
  - A BPXPRMxx member the specifies SYSPLEX(YES) and VERSION('nnnn')

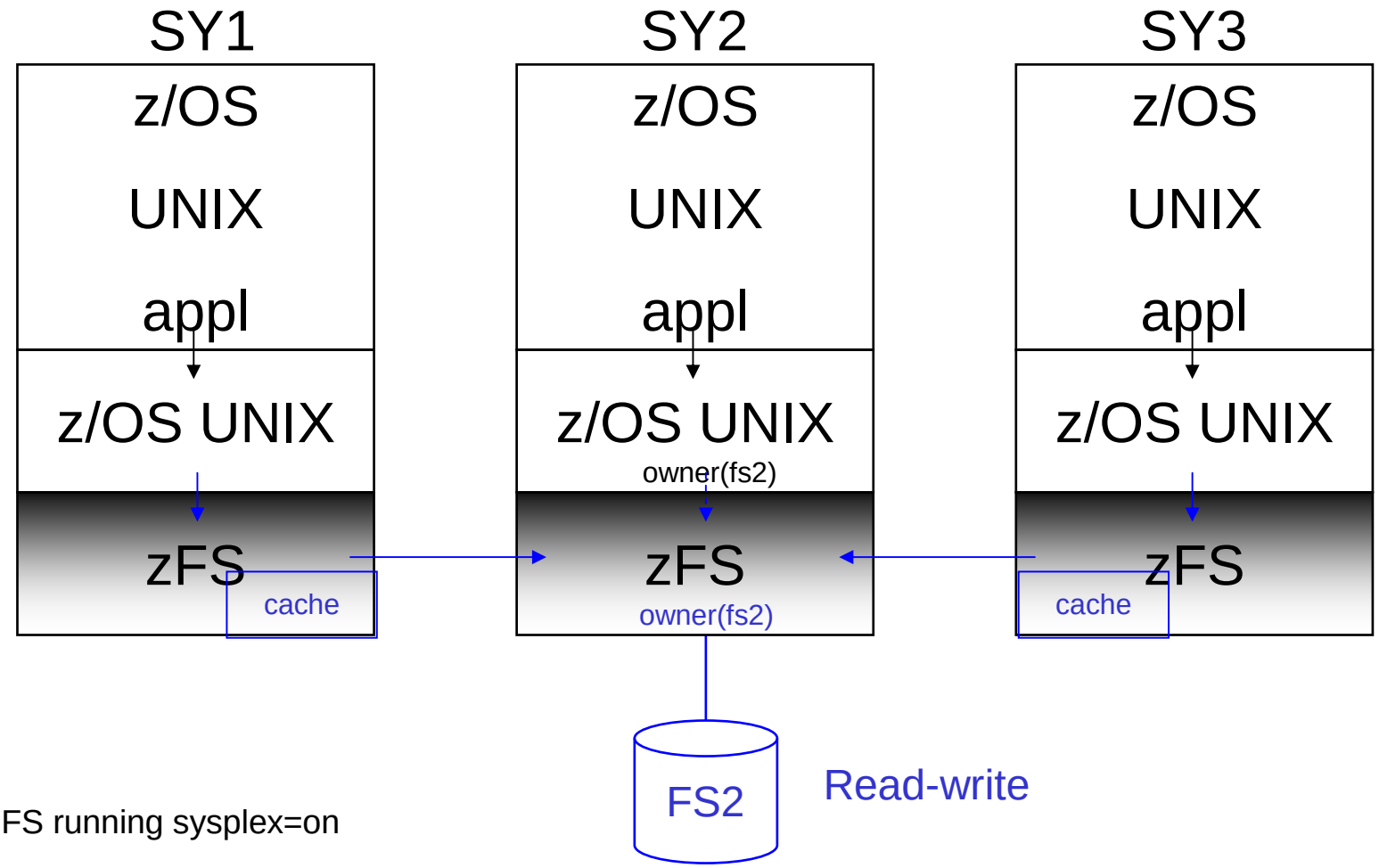
Read-only mounted file system in a shared file system environment



Read-write mounted file system in a shared file system environment



Sysplex-aware read-write mounted file system in a shared file system environment (R11 zFS)

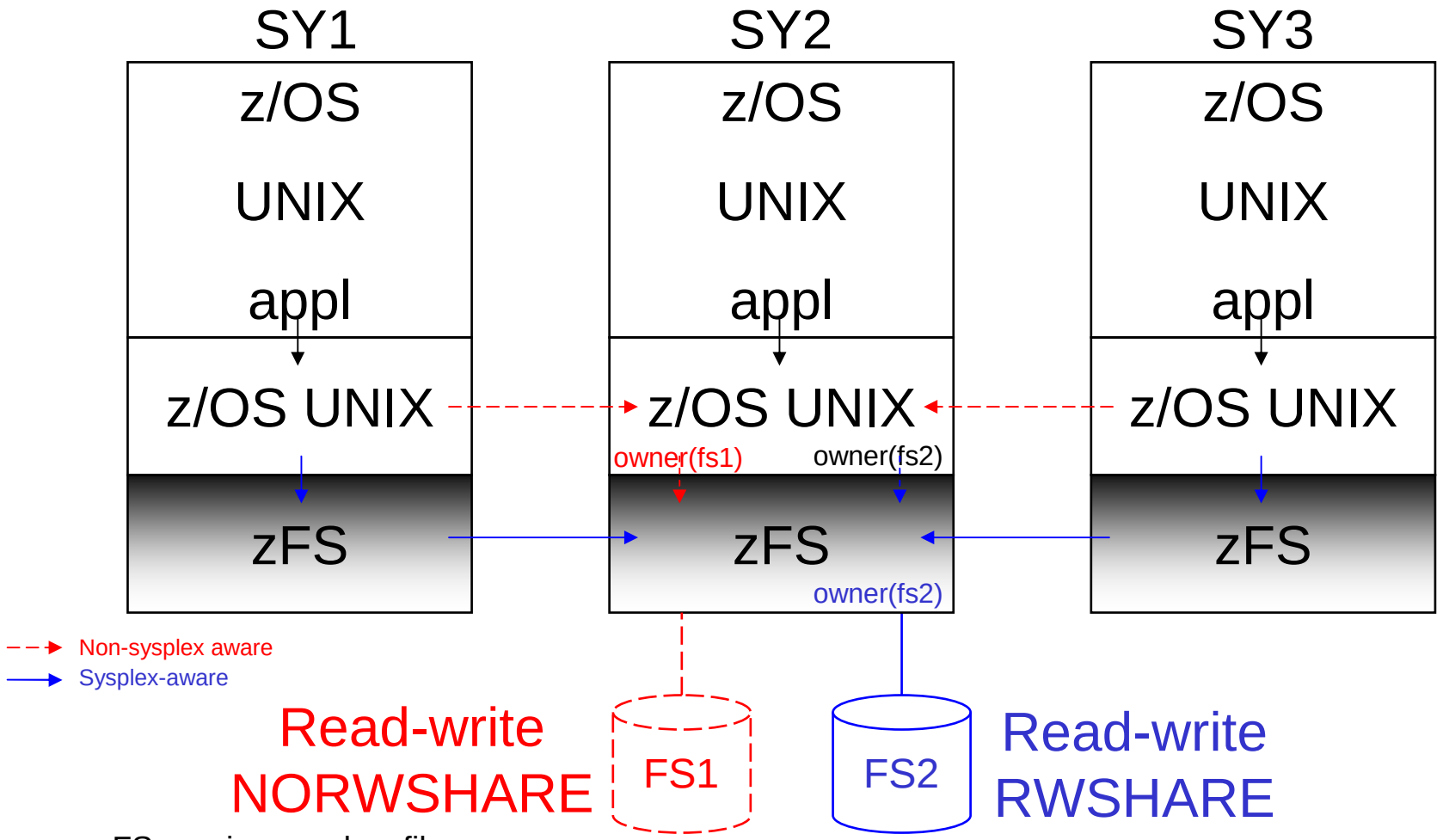


zFS running sysplex=on

Sysplex-aware read-write mounted file system in a shared file system environment

- Local mount
  This means that the mount request was sent by z/OS UNIX to the local zFS. The local zFS has the data set open. A **df -v** command for the file system would show Client=N. A file system is always locally mounted on the z/OS UNIX owning system.
- Global mount
  This means that the file system is available from any system in the shared file system environment. The file system is locally mounted on the z/OS UNIX owning system and may or may not be locally mounted on the other systems. A file system mounted in a shared file system environment is always globally mounted.

Sysplex-aware and non-sysplex aware read-write mounted file systems in a shared file system environment (R11 zFS APAR OA29619)



zFS running sysplex=filesys
(this is the preferred option)

New support in zFS R13

New support in zFS R13

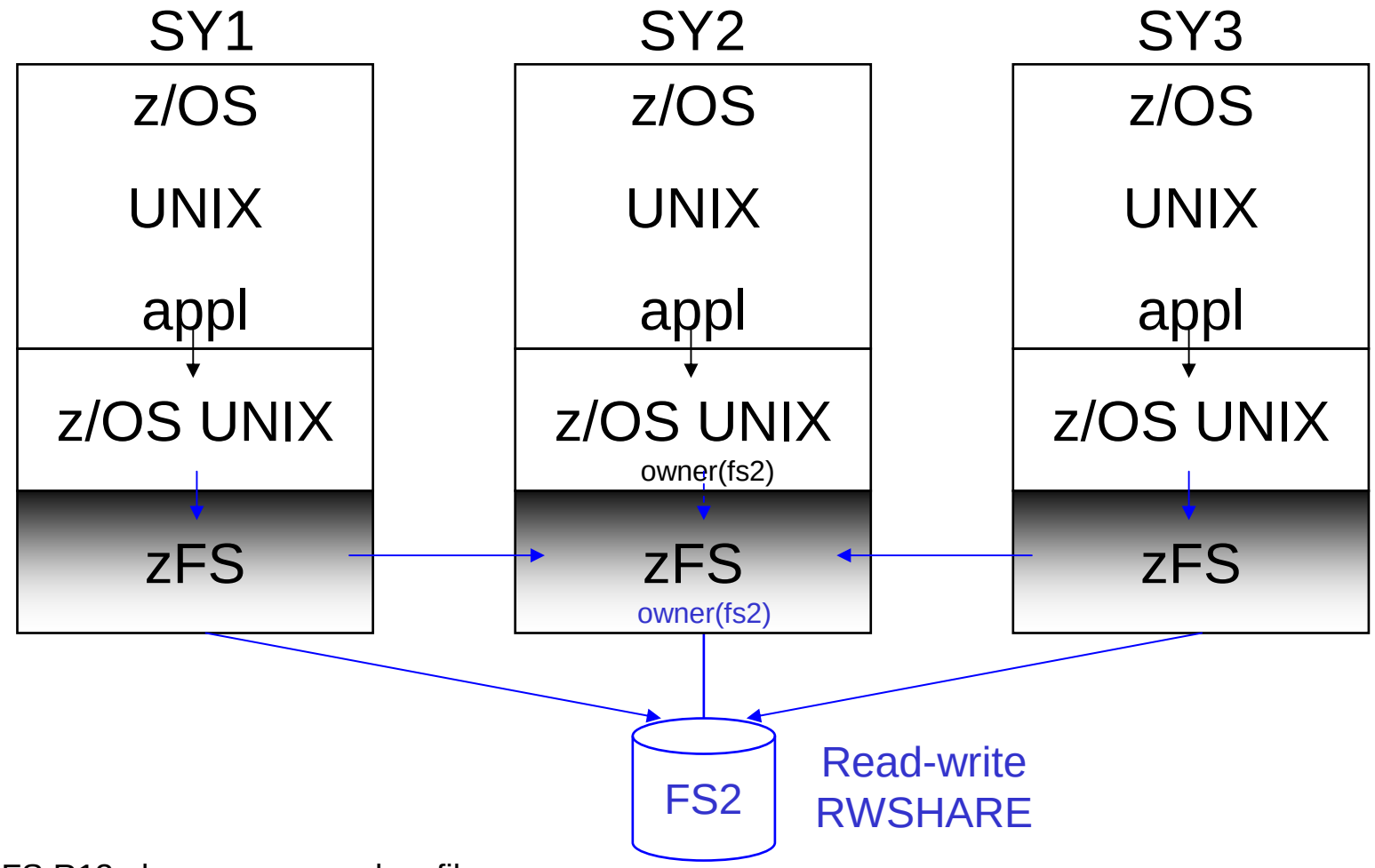
- zFS Direct I/O
- zFS internal restart
- zFS automatic re-enablement of disabled aggregates

zFS Direct I/O

- zFS R13 can now directly read and write **user data** for zFS sysplex-aware file systems
- This provides a significant performance improvement especially for zFS client systems
- Metadata[1] is still read and written by XCF communication to the zFS owning system

[1] Metadata is information about the file – for example, the file owner, the file permissions, the size of the file, the pointers to the data blocks, etc.

zFS Direct I/O



zFS R13 always runs sysplex=filesys

zFS Direct I/O performance

- Performance
  - zFS processing has been redesigned to allow all members of a Parallel Sysplex to perform zFS file system read and write I/O operations for shared file systems. In IBM laboratory tests, this new design yielded substantial file processing performance gains for all systems, both those that would have been zFS owning systems in prior releases and those that would not, and for single systems as well as those which were members of a Parallel Sysplex. The performance improvements observed in file processing varied over a broad range, from small to over 900% (over 10X) improvement, depending on the environment (monoplex or Parallel Sysplex) and the type of file processing being done. For most measured environments, the range of improvement observed fell between 50% (1.5X) and 150% (2.5X).
- Scalability improved, less storage needed on zFS owning system
- Application location (and zFS file system ownership) is less important

Changes to zFS IOEFSPRM options

- dir_cache_size is no longer used (it is ignored)
- nbs (new block security) is no longer used
  (zFS always runs with nbs)
- client_cache_size is not used for zFS Direct I/O clients
  (user_cache_size is used)
- sysplex is ignored (always runs sysplex=filesys)

Migration to zFS R13

- zFS R13 always runs sysplex=filesys (and it needs all other systems in the shared file system environment to be running sysplex=filesys)
- Migration to zFS R13 is a two step process
  - Install toleration APAR OA32925 (PTF UA55765) on all zFS R11 and R12 systems and make it active with a rolling IPL.
  - Change your zFS IOEFSPRM file to sysplex=filesys on all systems and make it active with a rolling IPL.
    (Default is that all zFS read-write file systems will be non-sysplex aware - NORWSHARE)
- There is a zFS migration health check to verify the sysplex=filesys option in prior releases (ZOSMIGV1R13_ZFS_FILESYS)
- Recognize that zFS R13 can use more DASD space than prior releases of zFS (zFS R13 does not use 1K fragments anymore).

Migration to zFS R13 (DASD space) …

- zFS R13 file data and directory names are no longer stored in fragments
  - When a fragmented file is updated, it is moved to an 8K block
  - New files and directories are stored in 8K blocks
  - Existing fragmented files remain fragmented if only read
- zFS R13 can use more DASD space than prior releases
  - For example, 1000 1K files can use up to 10 cylinders more of 3390 space in zFS R13 than prior releases
- Also, default for aggrgrow changes from aggrgrow=off to **aggrgrow=on**

zFS Internal Restart

- When zFS detects an internal failure, it will
  - Take a storage dump, and
    - Continue, or
    - Disable a zFS file system, or
    - Stop zFS and request that z/OS UNIX automatically **restart** it

      In the **restart** case, zFS recovers from the failure, but some zFS file systems may be unmounted

zFS Internal Restart …

- In zFS R13, when zFS would have stopped zFS and requested that z/OS UNIX automatically restart zFS, it will now execute an **internal restart**. It will
  - Take a storage dump
  - Suspend incoming requests
  - The zFS controller task stops and detaches the zFS kernel sub-task
  - The zFS controller task attaches the zFS kernel sub-task
  - The zFS kernel sub-task initializes
  - The zFS controller task issues internal mounts
  - Incoming requests are resumed

zFS automatic re-enablement of a disabled aggregate

- When a zFS file system is disabled by zFS, prior to R13, it would require an explicit unmount and remount to recover
- Now, in zFS R13, zFS will attempt re-enable the disabled file system by either
  - An internal remount samemode, or
  - A request to another zFS R13 system to takeover zFS ownership
- Even though zFS file system is re-enabled, you should run the zFS Salvager (IOEAGSLV) against the file system at your earliest convenience to ensure it is not corrupted

Publications

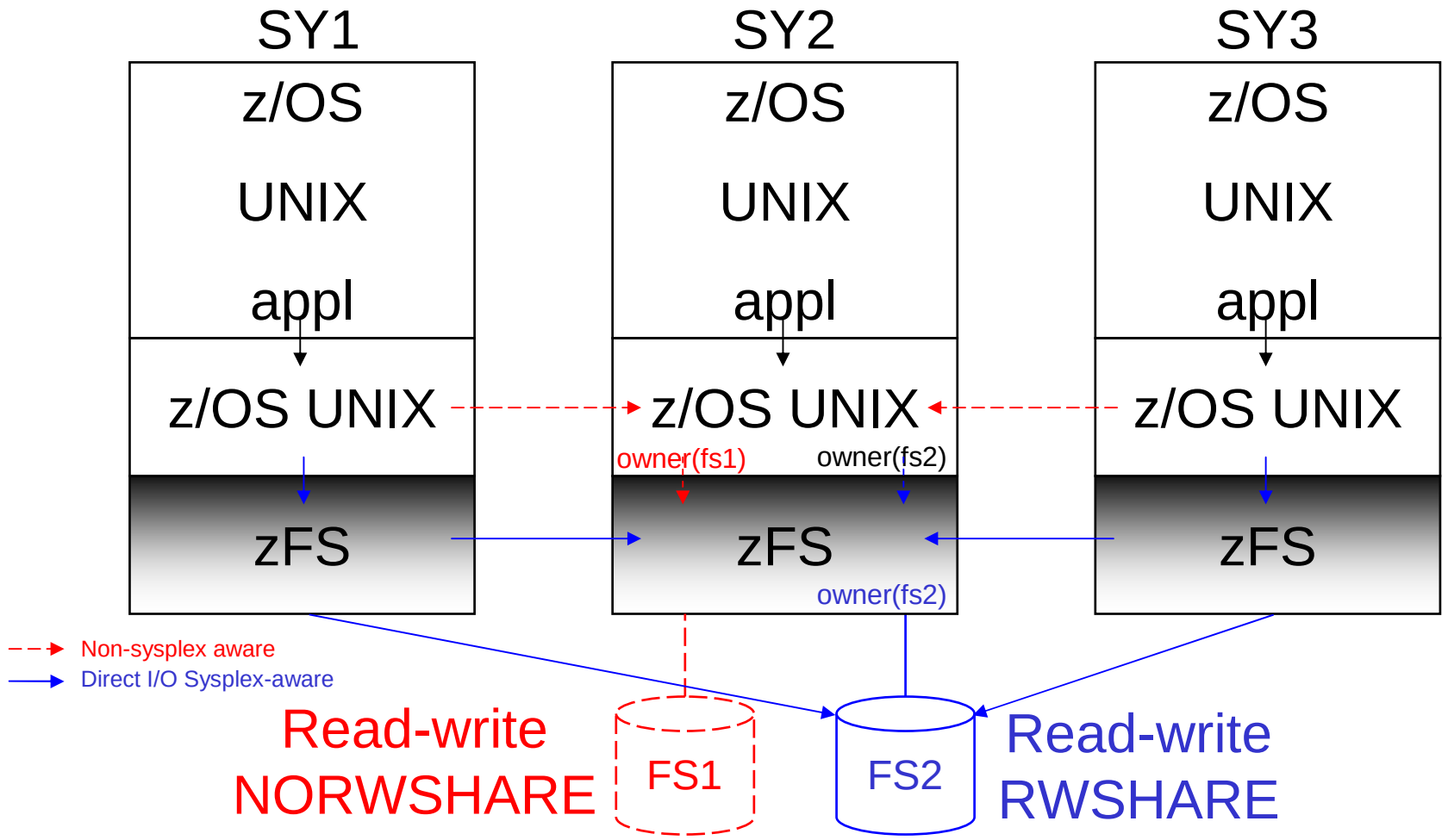
- z/OS UNIX System Services Planning (GA22-7800)
  General Administration of z/OS UNIX file systems
- z/OS UNIX Command Reference (SA22-7802)
  confighfs command for HFS
- z/OS MVS System Messages Volume 9 (IGF-IWM) (SA22-7639)
  IGWxxxt messages for HFS
- z/OS UNIX System Services Messages and Codes (SA22-7807)
  z/OS UNIX return codes, z/OS UNIX reason codes, X'5Bxxrrrr' reason codes for HFS
- z/OS Distributed File Service zSeries File System Administration (SC24-5989)
  zFS Concepts and zfsadm command for zFS
- z/OS Distributed File Services Messages and Codes (SC24-5917)
  IOEZxxxt messages and X'EFxxrrrr' reason codes for zFS
- z/OS Distributed File Service zSeries File System Implementation (SG24-6580)
  - Redbook available (updated February 2010 to include z/OS V1R11)
  - http://www.redbooks.ibm.com/abstracts/sg246580.html?Open
- z/OS Version 1 Release 8 Implementation (SG24-7265)
  - Redbook available (contains zFS updates for z/OS V1R8)
  - http://www.redbooks.ibm.com/abstracts/sg247265.html?Open
- z/OS DFSMS™ Access Method Services for Catalogs (SC26-7394)
  IDCAMS utility
- z/OS DFSMS™ Storage Administration Reference (SC26-7402)
  ADRDSSU utility for backup

Other z/OS UNIX related SHARE presentations

- Session 10646: z/OS 1.13 UNIX System Services Latest Status and New Features
  Monday, March 12, 2012 3:00 -4:00 PM
- Session 10637: New: sudo for z/OS
  Tuesday, March 13, 2012 3:00-4:00 PM
- Session 10866: Cloning for z/OS UNIX Service in a Shared File System Environment
  Wednesday, March 14, 2012 8:00 AM-9:00 AM
- Session 10979: UNIX Systems Services in Today's Mainframe Software
  Wednesday, March 14, 2012 9:30-10:30 AM
- Session 10983: z/OS UNIX for all
  Thursday, March 15, 2012 3:00-4:00 PM

Backup

Direct I/O sysplex-aware read-write mounted file system in a shared file system environment



zFS ownership movement of a sysplex-aware read-write mounted file system in a shared file system environment