What's New in z/VM I/O Support?

Eric Farman
IBM, z/VM I/O Development

10 August 2011
Session 9567

Trademarks

- The following are trademarks of the International Business Machines Corporation in the United States, other countries, or both.
  - Not all common law marks used by IBM are listed on this page. Failure of a mark to appear does not mean that IBM does not use the mark nor does it mean that the product is not actively marketed or is not significant within its relevant market.
  - For a complete list of IBM Trademarks, see **www.ibm.com/legal/copytrade.shtml**
- The following are trademarks or registered trademarks of other companies.
  - Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
  - Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.
  - Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
  - Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.
  - Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
  - UNIX is a registered trademark of The Open Group in the United States and other countries.
  - Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
  - ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.
  - IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.
  - * 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.

What's New in z/VM I/O Support?

- This presentation provides insight into those “significant” updates that describe new features and facilities supported by z/VM, but that cannot occupy an entire presentation on its own.
- The contents of this presentation is expected to change over time, as new function is made available via the service stream or regular releases.

What's New in z/VM I/O Support?

- Extended Remote Copy (XRC)
- Extended Address Volumes (EAV)
- FlashCopy / Space-Efficient
- Query DASD DETAILS

Extended Remote Copy (XRC)

Extended Remote Copy (XRC)

- Premier disaster recover solution for System z storage systems
  - Applies only to ECKD disks; SCSI volumes not covered
  - Also known as z/OS Global Mirror
- Requires a combination of software and hardware capabilities
  - Asynchronous copy between primary and secondary sites
  - Consistent data set at secondary site
  - Timestamp is placed on all write I/Os, in order to reassemble non-committed changes in the event of a disaster

Extended Remote Copy (XRC)

- z/VM announced (July 22, 2010) support for XRC by December
- Support is available for z/VM 5.4.0 and 6.1.0, with PTFs for APARs VM64814 and VM64816
  - VM64814 contains bulk of support; has no impact if not enabled
  - VM64816 contains additional fixes and feature enhancements (e.g., Monitor)
- Together, the APARs provide baseline Server Time Protocol support
  - Synchronizes z/VM TOD clock with STP server at IPL
  - Maintains a delta value of TOD changes over lifetime of z/VM IPL
  - Supports STP timezone management
  - CPC must be either:
    - A member of an STP-Only CTN
    - Stratum 2 or higher member of a mixed-CTN
- Inserts timestamp on write I/O, for use by XRC-enabled hardware
  - Occurs when XRC LIC is installed on DASD and z/VM LPAR is part of STP

Extended Remote Copy (XRC)

- New FEATURES statements for SYSTEM CONFIG to enable support:
  - STP_Timestamping
    - Timestamps will be added to write channel programs issued to all DASD devices that have the XRC LIC installed.
  - STP_TIMEZone / STP_TZ
    - System time zone will be derived from the STP server.
  - XRC_OPTional
    - System behaves differently when STP is suspended – Specifically, we will stop timestamping whenever STP sync is lost but continue issuing I/O. Without this, all I/O that is to be timestamped will be deferred until STP sync is restored
  - XRC_TEST
    - Only allowed 2nd-level. This will enable STP_Timestamping without STP availability. Manually-specified TOD value is used for timestamping. Intended for vendor test support.

Extended Remote Copy (XRC) – STP State Flow

- IPL
  - If STP_Timestamping and/or STP_TZ is specified
    - Perform activation by querying STP and setting the TOD clock to match the STP TOD value
  - If STP_TZ is specified
    - Set the system timezone to match the STP timezone
  - If either of the above fail, STP will enter SUSPENDED state. Otherwise, STP activation completes successfully and STP is considered ACTIVE.
- ACTIVE
  - If STP_Timestamping is specified
    - I/O to XRC-capable DASD will be timestamped (when required)
    - I/O to non-XRC-capable DASD will be unchanged
- SUSPENDED
  - If STP_Timestamping is specified and XRC_OPT is not specified
    - I/O to XRC-capable DASD that must be timestamped will be deferred until STP becomes ACTIVE again. I/O that does not need to be timestamped will still be issued.
    - I/O to non-XRC-capable DASD will continue to be issued
  - If STP_Timestamping is specified and XRC_OPT is also specified
    - I/O to all DASD will be issued without a timestamp until STP becomes ACTIVE again

Extended Remote Copy (XRC) – STP State Changes

- ACTIVE → SUSPENDED
  - Occurs when an STP machine check is received that informs CP that the TOD value must be re-synchronized.
- SUSPENDED → ACTIVE
  - In response to machine checks / external interrupts received, CP will attempt to re-synchronize with the STP server. A successful resynch will NOT change the system TOD value, but will:
    - Update the delta between the system time and the STP TOD value (STITODOF)
    - Update the system time zone (if STP_TZ is enabled).
  - If resynchronization fails, the system remains STP suspended.
- ACTIVE->ACTIVE
  - External interrupts may be received that require CP to query STP timezone information (for example, the STP timezone was changed via the HMC). This does not cause STP to go suspended, but will cause the system timezone to change.

Extended Remote Copy (XRC) – Externals

- Query STP
  - Server Time Protocol facility not enabled
    - The STP LIC has not been enabled on your CPC. Cannot use XRC Timestamping functionality until this is resolved.
  - Server Time Protocol available but not enabled for CP use
    - The STP LIC has been enabled, but no STP features have been enabled in the System Configuration file.
  - Server Time Protocol synchronization activated for timestamping
    - The TOD clock was synchronized with STP at IPL and timestamping is active for XRC-capable DASD.
  - Server Time Protocol synchronization suspended. I/O to XRC-capable DASD will be delayed until synchronization completes.
    - Connectivity to the STP server has been severed. I/Os to XRC-capable DASD will queue until connectivity is restored.
  - Server Time Protocol synchronization suspended. I/O to XRC-capable DASD will be issued without timestamps until synchronization completes.
    - Connectivity to the STP server has been severed. I/O to XRC-capable DASD will continue without timestamps (this is the result of specifying the XRC_OPTional feature)

Extended Remote Copy (XRC) – Externals

- Query TIMEZONE
  - Represents provided by STP
  - The boundary defines the time and date at which the next scheduled timezone change will occur
  - The inactive timezone will become the active timezone at the boundary time/date

```
Zone  Direction   Offset    Status         Boundary
GMT      ----    00.00.00  Inactive
UTC      ----    00.00.00  Inactive
EDT      West    04.00.00  Active-(STP)
EST      West    05.00.00  Inactive-(STP) 10/31/09 02:00:00
```

Extended Remote Copy (XRC)

- For more information, consult the redbook “GDPS Family – An Introduction to Concepts and Capabilities”, order number SG24-6374

Extended Address Volumes (EAV)

Extended Address Volumes (EAV)

- Mechanism for creating and using volumes larger than 65,520 cylinders (the so-called “mod-54” variant of a 3390-9), or about 48.3 GB
  - Identified as 3390-A devices, initially ~193 GB maximum

Extended Address Volumes (EAV)

- Large cylinder addressing is done by “stealing” 12 bits from head value of CCHH identifier for use as high-order bits to cylinder
  - No DASD has ever had more than 15 heads (x0000-x000E)
  - Often documented as ccccCCCh, where the complete cylinder is rearranged as CCCcccc

Cylinders (Dec)	Cylinders (Hex)	Highest ccccCCCh
3339	x0D0B	x0D0A000E
65520	xFFF0	xFFEF000E
65536	x10000	xFFFF000E
65537	x10001	x0000001E
65667	x10083	x0082001E

Extended Address Volumes (EAV)

- Two APARs for z/VM 5.4.0 and 6.1.0
  - VM64709 (CP)
  - VM64711 (CMS)
- PTFs available December 2009

Extended Address Volumes (EAV)

- EAV volumes can be attached to the SYSTEM, generally for the purposes of minidisks
- Non-PERM extents on CPOWNED EAV volumes are restricted to the first 65,520 cylinders
  - Extents that cross this line will be truncated, with message HCP138E
  - Extents that exist entirely above this line will be ignored, with message HCP139E
- CPFMTXA will enforce this boundary requirement; ICKDSF does not

Extended Address Volumes (EAV)

- z/VM support for dedicated devices, and fullpack minidisks
  - “Fullpack” is defined as 0-END, or DEVNO
  - The ending cylinder must be “END”, even if the number would equal the size of the volume
- Partial pack minidisks on 3390-A volumes are supported, provided they exist completely below cylinder 65,520

Extended Address Volumes (EAV)

- Diagnose xA8 will operate on any area of disk, provided the application is aware of EAV addressing constructs
  - New SGIOPTS byte, immediately after SGILPM, is added, with flag bit SGIEAV (x80) defined to signal this awareness
  - Any application attempting to reference above cylinder 65,520 without enabling this bit will fail with CC=1 R15=2 (unsupported device)

Extended Address Volumes (EAV)

```
*** SGIOP - Synchronous General I/O Parameters
*
*     +-------------+------+------+---------------------------+
*   0 |  SGIDEVNO   |SGIKEY|SGIFLG|         SGIRESV1          |
*     +-------------+------+------+---------------------------+
*   8 |          SGICPA           |         SGIRESV2          |
*     +---------------------------+------+------+-------------+
*  10 |         SGICCWA           |:DEVST|:SCHST|  SGIRESCT   |
*     +------+------+-------------+------+------+-------------+
*  18 |SGILPM|:OPTS |  SGIRESV3   |  SGIRESV4   |  SGISNSCT   |
*     +------+------+-------------+-------------+-------------+
*  20 |         SGIRESV5          |         SGIRESV6          |
*     +---------------------------+---------------------------+
*  28 |         SGIRESV7          |         SGIRESV8          |
*     +---------------------------+---------------------------+
*  30 |         SGIRESV9          |         SGIRESVA          |
*     +---------------------------+---------------------------+
*  38 |                                                       |
*     =                       SGISDATA                        =
*     |                                                       |
*     +-------------------------------------------------------+
*  58
*
*** SGIOP - Synchronous General I/O Parameters
```

Extended Address Volumes (EAV)

- Diagnose x210 returns the size of a volume in field VRDCPRIM, but this field cannot hold anything larger than 16-bits (65,535)
  - Field practical maximum is 65,520 (xFFF0)
- If VRDCPRIM contains xFFFE, indicates the field has overflowed (large, EAV volume)
- New 32-bit field VRDCCYLS at offset x4C
  - Always contains cylinder size

Extended Address Volumes (EAV)

```
*** VRDCBLOK - VIRTUAL/REAL DEVICE CHARACTERISTICS BLOCK
*
*     +-------------+-------------+------+------+------+------+
*   0 |  VRDCDVNO   |  VRDCLEN    |:CVCLA|:CVTYP|:CVSTA|:CVFLA|
*     +------+------+------+------+------+------+------+------+
*   8 |:CRCCL|:CCRTY|:CCRMD|:CCRFT|:CUNDV|:CRDAF|  VRDCRSVD   |
*     +------+------+------+------+------+------+-------------+
*  10 |  VRDCCUTY   |:CCUMD|  VRDCDVTY   |:CDVMD| VRDCDVFE-   |
*     +------+------+------+------+------+------+-------------+
*  18 |-(016)|:CSDFE|:CDVCL|:CDVCO|  VRDCPRIM   |  VRDCTRKC   |
*     +------+------+------+------+-------------+------+------+
*  20 |:CSECT|     VRDCTOTR       |   VRDCHA    |:CMODE|:CMDFR|
*     +------+------+-------------+-------------+------+------+
*  28 |  VRDCNKOV   |  VRDCKOVH   |  VRDCALTC   |  VRDCALTR   |
*     +-------------+-------------+-------------+-------------+
*  30 |  VRDCDIG    |  VRDCDIGN   |  VRDCDVCY   |  VRDCDVTR   |
*     +------+------+------+------+-------------+-------------+
*  38 |:CMDR |:COBR |:CCUID|//////////////////////////////////|
*     +------+------+------+//////////////////////////////////|
*     |///////////////////////////////////////////////////////|
*     +------+--------------------+---------------------------+
*  48 |:RCUC |////////////////////|         VRDCCYLS          |
*     +------+--------------------+---------------------------+
*  50 |                       VRDCPGID                        |
*     |                    +----------------------------------+
*  58 |                    |//////////////////////////////////|
*     +--------------------+----------------------------------+
```

Extended Address Volumes (EAV)

- Other Diagnoses (x18, x20, xA4, x250, and *BLOCKIO) cannot operate on devices that exist wholly or partially above cylinder 65,520
  - Attempts to do so will be rejected, with the proper “unsupported device” return/condition code combination for the Diagnose

Extended Address Volumes (EAV)

- New DDR fully supports EAV volumes for backup purposes
- Updates to FlashCopy ensure it fully supports EAV volumes
- New fields in MRSEKSEK (Domain 7, Record 1) monitor record less likely to wrap for larger volumes
  - CALCURCY, CALSKCYL, IORPOSSM, CALECYL

Extended Address Volumes (EAV)

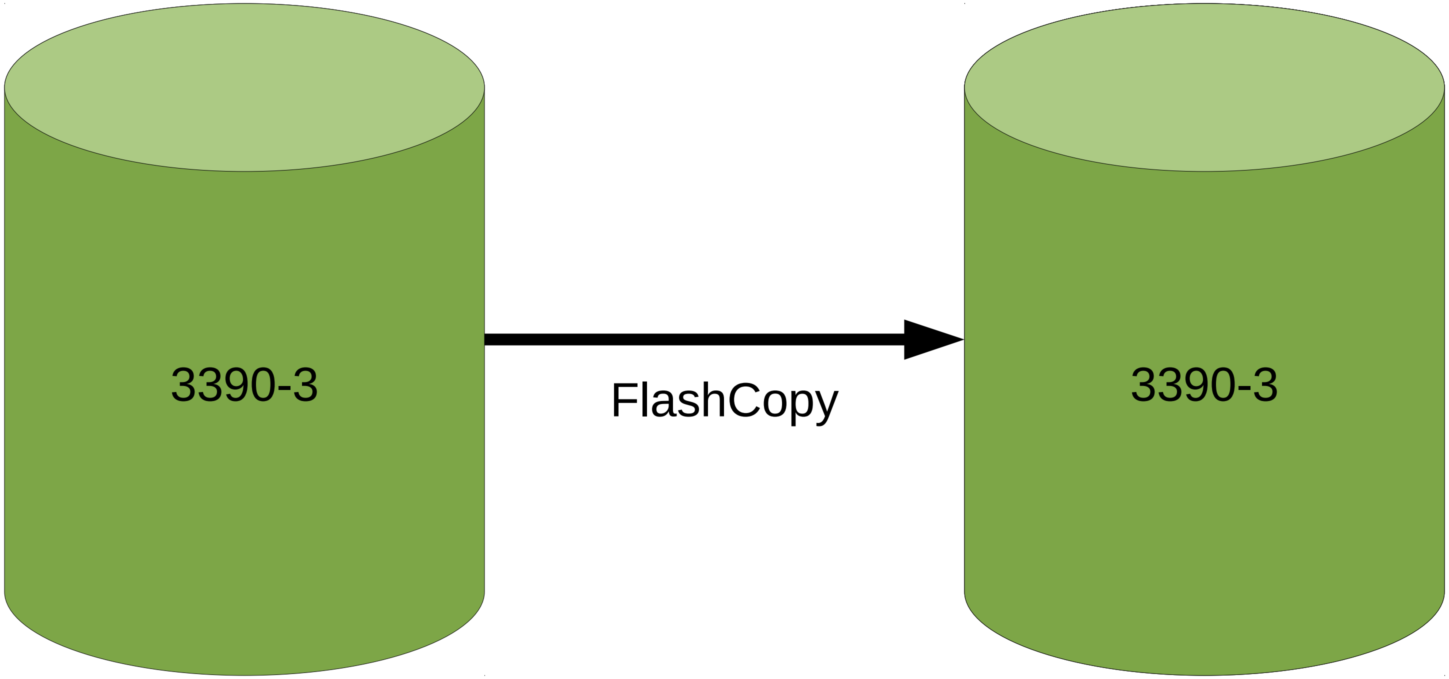
- CMS currently supports volumes up to 32,767 cylinders in size
- With aforementioned APAR, CMS is updated to support, via FORMAT and ACCESS, volumes and minidisks up to 65,520 cylinders in size
  - Since this does not provide support for EAV-sized volumes, can be applied separately from CP APAR
- Still, does not eliminate the requirement of file status and control information that is stored below the 16MB line of the CMS virtual machine
  - If volumes larger than 32,767 cylinders in size are used, care should be taken to avoid large numbers of small files on the disk. Otherwise, the CMS file system may encounter problems accessing and using the device.

FlashCopy / Space-Efficient

FlashCopy

- Point-in-time copy technology, good for duplicating volumes where differences presented over time are acceptable
- Target volume is an exact replica of the source upon completion of the copy
- Both source and target volume can be used for applications when FlashCopy is finished

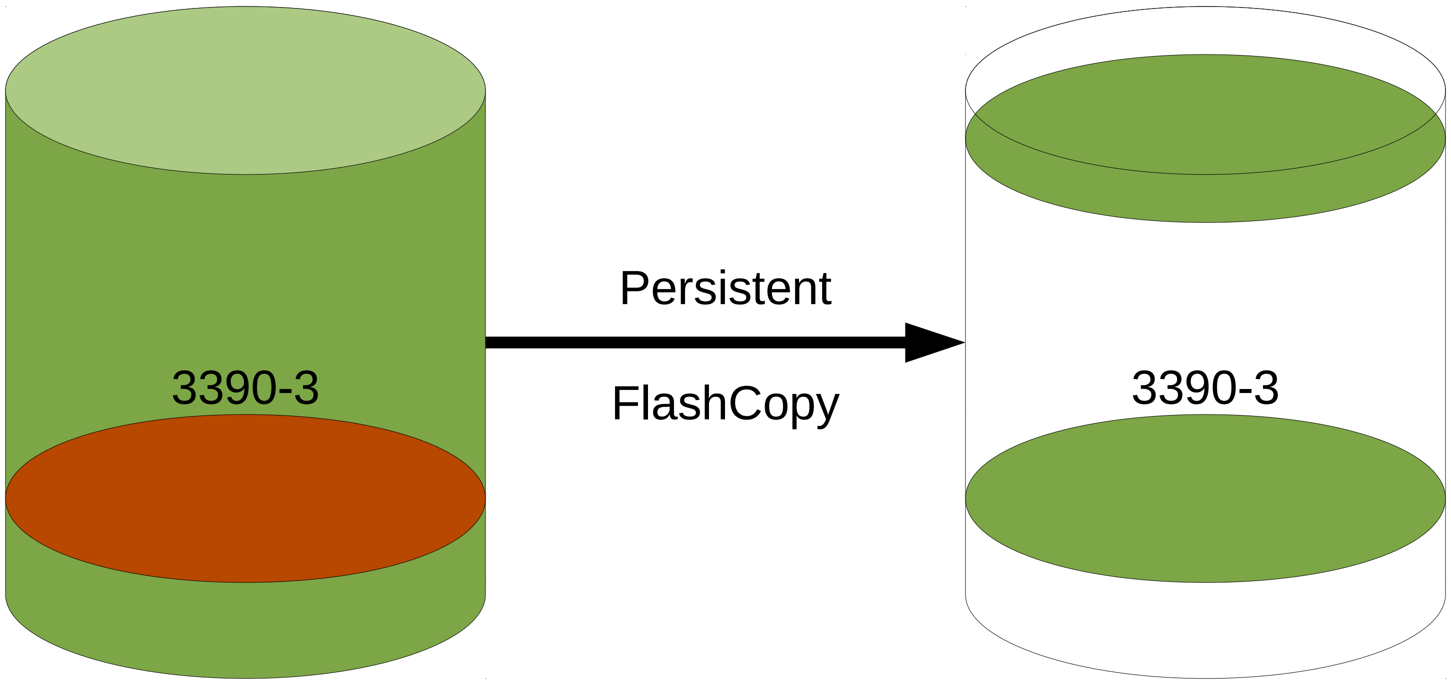
FlashCopy



FlashCopy

- Persistent FlashCopy relationships are used to maintain pointers between source and target volumes, and changes made to either one
- Still a point-in-time copy; does not suggest a mirroring technology in any fashion
- The no-background-copy option (typically NOCOPY) is used to prevent tracks being written to target device until they are actually needed by storage subsystem.

Persistent FlashCopy with NOCOPY



FlashCopy / Space-Efficient

- Space-Efficient volumes are those created solely for the purpose of being targets of persistent FlashCopy relationships
  - A special LIC feature of DS8000
- Tracks for the volume are only allocated as needed to maintain the point-in-time image as source/volume change
  - Still have the look-and-feel of a fully populated device to the user

FlashCopy / Space-Efficient

- z/VM APAR VM64449 (December 2008) was released to add persistent FlashCopy, and by extension, FlashCopy/SE, support to the z/VM FLASHCOPY command
  - Several APARs released in 2009; check for related!
- Several new commands are added to create and manage these relationships

FlashCopy / Space-Efficient

- FLASHCOPY ESTABLISH
  - Create Persistent Relationships
  - Needed for SE volumes, but can be used with traditional volumes, if desired
  - REVERSIBLE option is combination of CHGRECORD and NOTGTWRITE, and ensures targets is not SE
  - LABEL and SAVELABEL options also exist on traditional (non-persistent) FLASHCOPY command

FlashCopy / Space-Efficient

```
>>--FLASHCopy--ESTABLISH----------------------------------------------------->

                                  <-------------< (1)  <--------------<
>--.-SOURCE--| fullext |--TARGET----| fullext |---------.------------.--.---->
   |                                                    |-CHGRECORD--|  |
   |                                                    |-NOTGTWRITE-|  |
   |                                                    '-REVERSIBLE-'  |
   |                              <-------------< (2)                   |
   '-SOURCE--| miniext |--TARGET----| miniext |-------------------------'

>--.--------.--.-------------.--.------------.--.---------------.-----------><
   '-NOCOPY-'  '-FAILNOSPACE-'  '-NOSETARGET-'  |-LABEL--volser-|
                                                '-SAVELABEL-----'
```

FlashCopy / Space-Efficient

- FLASHCOPY WITHDRAW
  - Removes Persistent FlashCopy relationship between paired volumes
  - FORCE option required when background copy has not completed, as a result of NOCOPY option on Establish
    - FLASHCOPY BACKGNDCOPY can be used to initiate background copy without withdrawing relationship, but cannot be used on space-efficient targets

FlashCopy / Space-Efficient

- FLASHCOPY RESYNC
  - For persistent relationships, will establish a new point-in-time copy between source and target volume(s)
- FLASHCOPY TGTWRITE
  - Disables the write-inhibit option on target volume, if specified on original Establish

FlashCopy / Space-Efficient

- QUERY FLASHCOPY
  - Privileged command to interrogate status of FlashCopy relationship(s) on a real device
  - State on the hardware, and options for retrieving information about a relationship created within a z/VM session during current IPL

```
                                  <---------------<
>>--Query--FLASHCopy--.-HARDWARE----.-----------.---.----------------------><
                      |             |-rdev------|   |
                      |             '-rdev-rdev-'   |
                      |-TABLEd----------------------|
                      |-CREATed--userid|*-----------|
                      |-OWNer--userid|*-------------|
                      |-SEQUENCE--hhhhhhhh----------|
                      |-VOLume--volser--------------|
                      '-DEVice--rdev----------------'
```

FlashCopy / Space-Efficient

```
q flashcopy hardware
              --------SOURCE-------- --------TARGET--------
SEQUENCE FLGS RDEV VOLSER CC...CC/HH RDEV VOLSER CC...CC/HH REMAINING/TOTAL
4B145C34 0800 5100 PACK01     100/00 5101 PACK02     100/00 0/1500
4B14700A 0800 5100 PACK01     200/00 5101 PACK02     200/00 0/1500
4B145C34 8800 5100 PACK01     100/00 5101 PACK02     100/00 0/1500
4B14700A 8800 5100 PACK01     200/00 5101 PACK02     200/00 0/1500

q flashcopy tabled
SEQUENCE ---DATE--- --TIME--  RDEV  VOLSER CREATOR  OWNER    VDEV
4B145C34 2008-04-30 15:19:55  5100> PACK01 RWS      RWS      0200
4B145C34 2008-04-30 15:19:55 >5101  PACK02 RWS      RWS      0300
4B14700A 2008-04-30 15:20:00  5100> PACK01 RWS      RWS      0210
4B14700A 2008-04-30 15:20:00 >5101  PACK02 RWS      RWS      0310
Ready; T=0.01/0.01 15:20:40
```

Query DASD DETAILS

Query DASD DETAILS

- New microcode on DS8000 boxes allow for encrypted DASD and/or solid-state drives
- z/VM APAR VM64650 (May 2009) provides updates to Query DASD DETAILS output to indicate the presence of these facilities
  - No new output is displayed if neither exists

Query DASD DETAILS

```
q dasd details 521d
521D  CUTYPE = 2107-E8, DEVTYPE = 3390-0A, VOLSER = ERF001, CYLS = 3339
      CACHE DETAILS:  CACHE NVS CFW DFW PINNED CONCOPY
           -SUBSYSTEM   Y    Y   Y   -    N       N
              -DEVICE   Y    -   -   Y    N       N
      DEVICE DETAILS: CCA = 1D, DDC = --, DED = YES, SSD = NO
      DUPLEX DETAILS: --
      CU DETAILS: SSID = 0102, CUNUM = 5200

q dasd details dead
DEAD  CUTYPE = 6310-80, DEVTYPE = 9336-10, VOLSER = ERF105, CYLS = 9446
      BLKS = 7340032
      DEVICE DETAILS: DED = YES, SSD = NO
```

FIN