

**SHARE**

Technology • Connections • Results

# DFSMShsm Best Practices

Glenn Wilcock  
wilcock@us.ibm.com

August 3, 2010  
Session 8046



**SHARE** in Boston

# Legal Disclaimer

## NOTICES AND DISCLAIMERS

Copyright © 2008 by International Business Machines Corporation.

No part of this document may be reproduced or transmitted in any form without written permission from IBM Corporation.

Product information and data has been reviewed for accuracy as of the date of initial publication. Product information and data is subject to change without notice. This document could include technical inaccuracies or typographical errors. IBM may make improvements and/or changes in the product(s) and/or programs(s) described herein at any time without notice.

References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business. Consult your local IBM representative or IBM Business Partner for information about the product and services available in your area.

Any reference to an IBM Program Product in this document is not intended to state or imply that only that program product may be used. Any functionally equivalent program, that does not infringe IBM's intellectual property rights, may be used instead. It is the user's responsibility to evaluate and verify the operation of any non-IBM product, program or service.

THE INFORMATION PROVIDED IN THIS DOCUMENT IS DISTRIBUTED "AS IS" WITHOUT ANY WARRANTY, EITHER EXPRESS OR IMPLIED. IBM EXPRESSLY DISCLAIMS ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT. IBM shall have no responsibility to update this information. IBM products are warranted according to the terms and conditions of the agreements (e.g., IBM Customer Agreement, Statement of Limited Warranty, International Program License Agreement, etc.) under which they are provided. IBM is not responsible for the performance or interoperability of any non-IBM products discussed herein.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not necessarily tested those products in connection with this publication and cannot confirm the accuracy of 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.

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents or copyrights. Inquiries regarding patent or copyright licenses should be made, in writing, to:

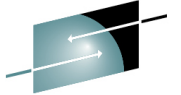
IBM Director of Licensing  
IBM Corporation  
North Castle Drive  
Armonk, NY 10504-1785  
U.S.A.

# Trademarks

**The following are trademarks of the *International Business Machines Corporation*:**

**IBM, DFSMS/MVS, DFSMSHsm, DFSMSrmm, DFSMSdss, DFSMSopt, DFSMS Optimizer, z/OS, eServer, zSeries, MVS, FlashCopy®**

The information contained in this presentation is distributed on an 'AS IS' basis without any warranty either expressed or implied, including, but not limited to, the implied warranties of merchantability or fitness for a particular purpose. The use of this information is a customer responsibility and depends on the customer's ability to evaluate and integrate it into the customer's operational environment.



# Agenda

- Control Data Sets
- Recall
- Migration
- Audit
- Recycle
- Tape
- Throughput
- Availability
- Performance
- Reporting
- Miscellaneous

# Control Data Sets

## Record Level Sharing

- To improve overall DFSMShsm performance, access the CDSs using **Record Level Sharing (RLS)**
- Customers report significant performance improvements after switching to RLS
- Actual customer data, Bank 1, comparing nonRLS and RLS, with 1 yr elapsed:

Function	Increase in GBytes moved	Decrease in Window size
Auto Backup	33%	-25%
Migrate -> ML2	18%	-36%

- Actual customer, Bank 2, AUDIT before and after:
  - Before: Couldn't complete in 24 hrs
  - After: Complete within 4 hrs
- ✓ If you tried RLS and didn't see an improvement, it is most likely a configuration problem

# Control Data Sets

## GRS Star

- *Internal* performance testing has shown a significant improvement in CDS I/O intensive functions when using **GRS Star** as opposed to **GRS Ring**
  - **GRS Star** – A parallel sysplex implementation of Global Resource Serialization
    - Resource name list is placed in the coupling facility so that any request for a resource can be resolved with a single interaction
  - **GRS Ring** – A resource request must be passed to every participating member of the sysplex (ring)

# Control Data Sets

## CDS Reorg

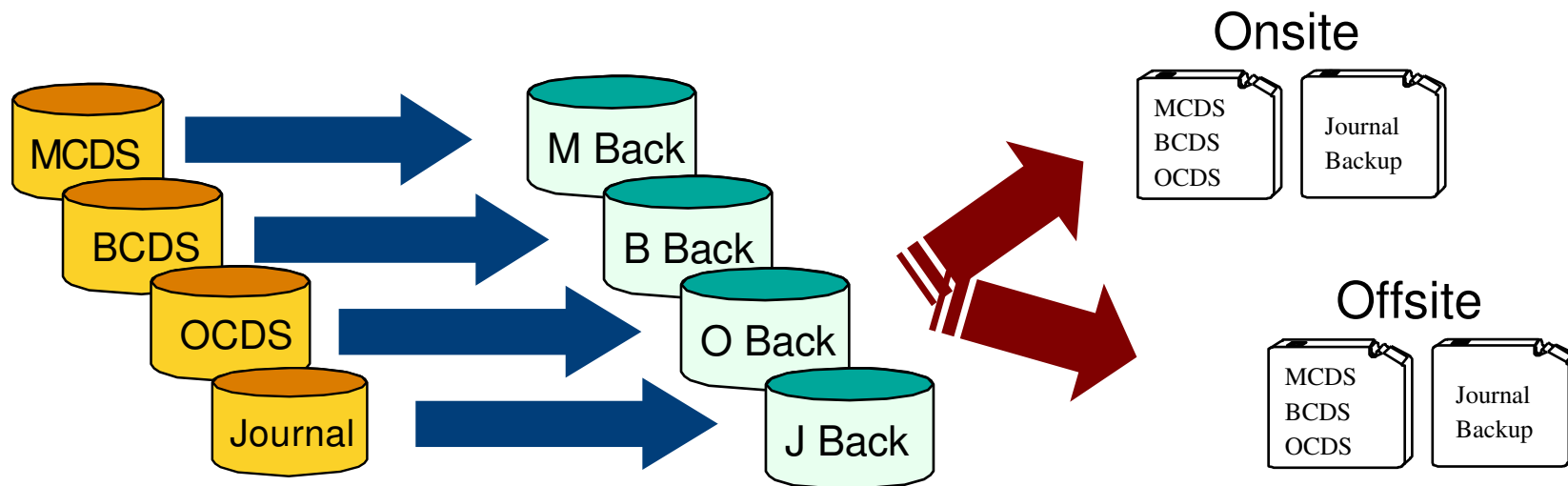
- Try to keep Reorganizing the CDSs to a minimum
  - ★V1R12 CA Reclaim
- CDS Performance will be degraded for 2-3 weeks after a REORG
  - VSAM will perform a large number of CI / CA splits to create space for record insertions
  - Don't panic when HURBA / HARBA ratio increases during first few days
- Use FREESPACE(0 0) so that VSAM can create free space where it is needed
- ! Make sure all DFSMSHsm hosts in HSMplex are shutdown
  - This is one of the leading causes of breaking the CDSs
  - Use DISP=OLD in REORG job to prevent DFSMSHsm from starting

# Control Data Sets

## Duplex CDS Backup Copies

- Create disk backup copies in parallel using PIT copy
- Use CB Exit to schedule a DFSMSDss dump job to create multiple copies of the disk backup copies

**SETSYS EXITON(CB) CDSVERSIONBACKUP(DASD)**





# Control Data Sets

## Health Checks / Journal Format

- Enable DFSMSHsm Health Checker checks
  - **HSM\_CDSB\_BACKUP\_COPIES**: Ensures that at least four CDS backup copies are being maintained
  - **HSM\_CDSB\_DASD\_BACKUPS**: When backing up to disk, ensures that all CDS Backup copies exist
  - **HSM\_CDSB\_VALID\_BACKUPS**: Determines if the number of *valid* backup copies has dropped below four
- Allocate the journal as a Large Format Sequential data set if you have to back up the CDSs more than once a day due to the journal filling up

# Control Data Sets

## CDS Recovery

- Keep journal and disk backups separate from MCDS, BCDS and OCDS
- Minimize CDS Loss
  - Dual Copy / Remote Copy
  - Raid 5 or Raid 6
- Have documented and Tested CDS Recovery Plans
- Review “Data Recovery Scenarios” in *DFSMSHsm Storage Administration* manual

# Recall Prioritization

- **RP Exit** can be used to prioritize data set Recall, Delete and Recover requests
- Priority range 0 – 100
  - Default priority is 50
- All Wait-type requests are prioritized higher than noWait-type requests
- Recall and Delete requests are on the same queue

**DSN: CUSTMR.DS1**  
Priority: 100

**DSN: PROD05.DS1**  
Priority: 80

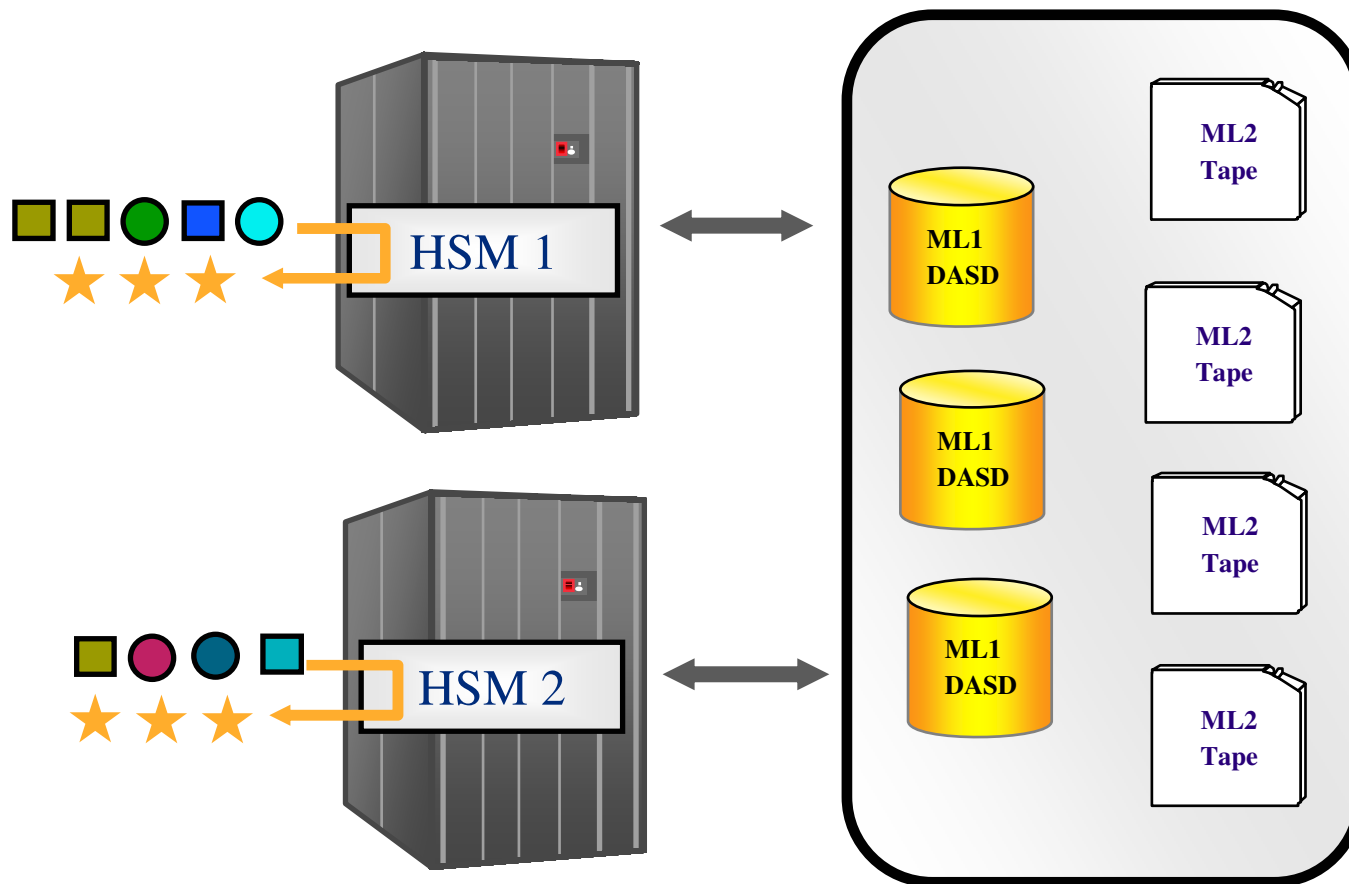
**DSN: USERME.DS9**  
Priority: 50

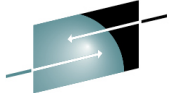
**DSN: UTIL01.DS2**  
Priority: 20

**DSN: CLEANUP.DS7**  
Priority: 10

# Control Data Sets Common Recall Queue

- NonCRQ environment – Each host processes own requests

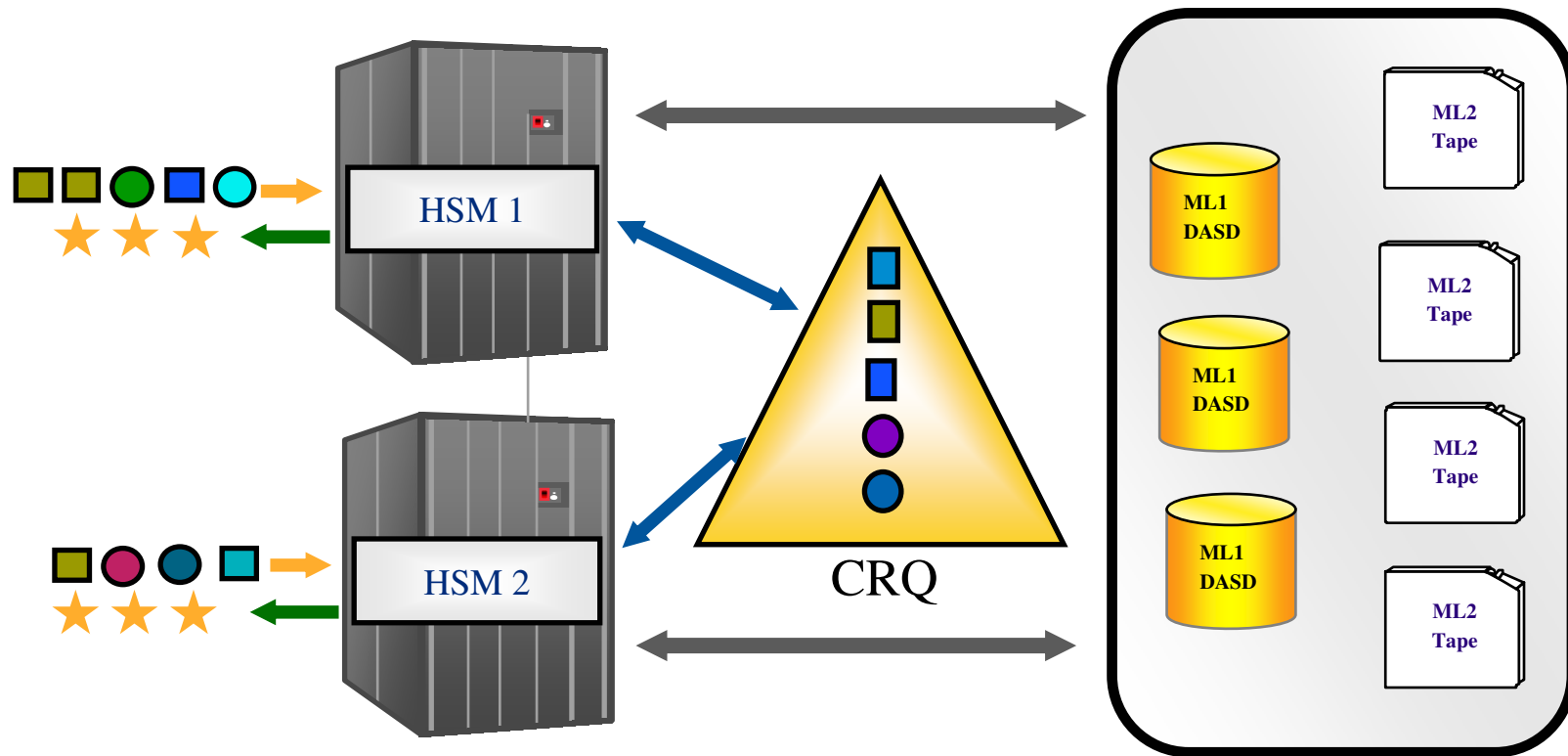




# Recall

## Common Recall Queue (cont)

- CRQ - All requests are placed onto a shared queue from which all hosts can select requests for processing
  - Implemented using a Coupling Facility List Structure



# Recall

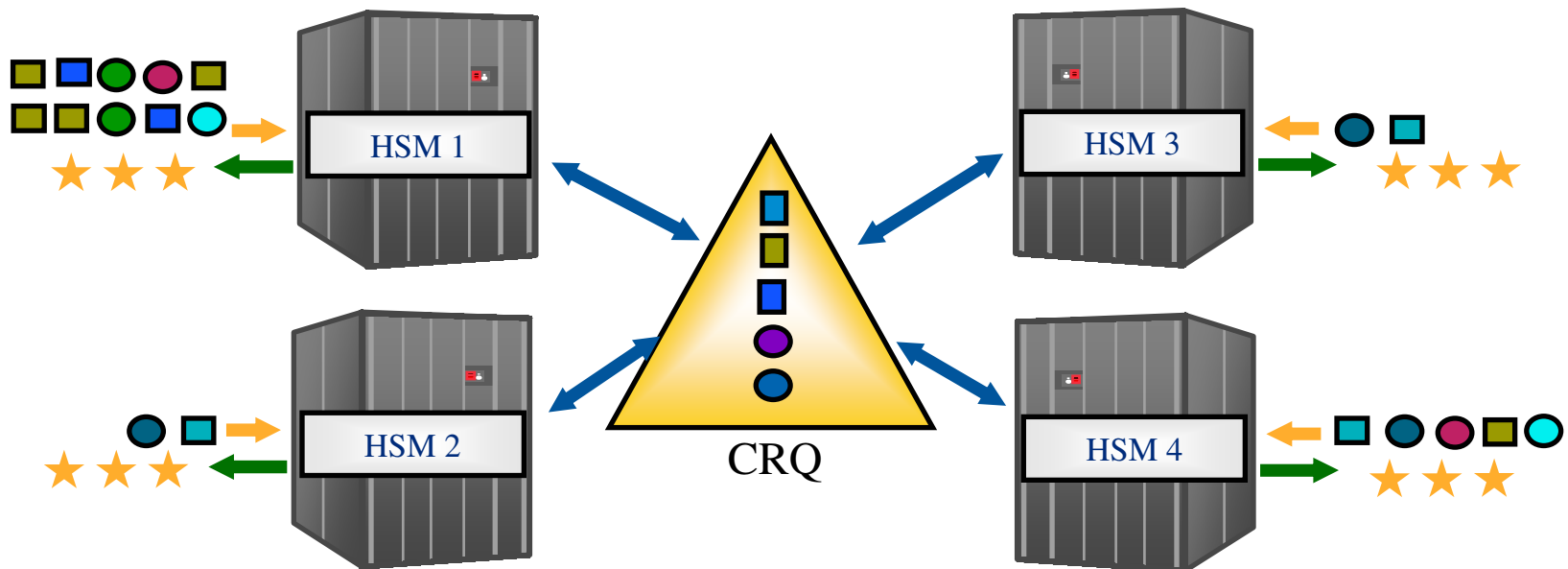
## Common Recall Queue *(cont)*

- **Advantages of CRQ**
  - Workload balancing
  - Tape mount optimization
  - Quiesce Activity w/o impacting Recall
  - Priority optimization
  - Flexible configurations
  - Request persistence

# Recall

## Common Recall Queue (cont)

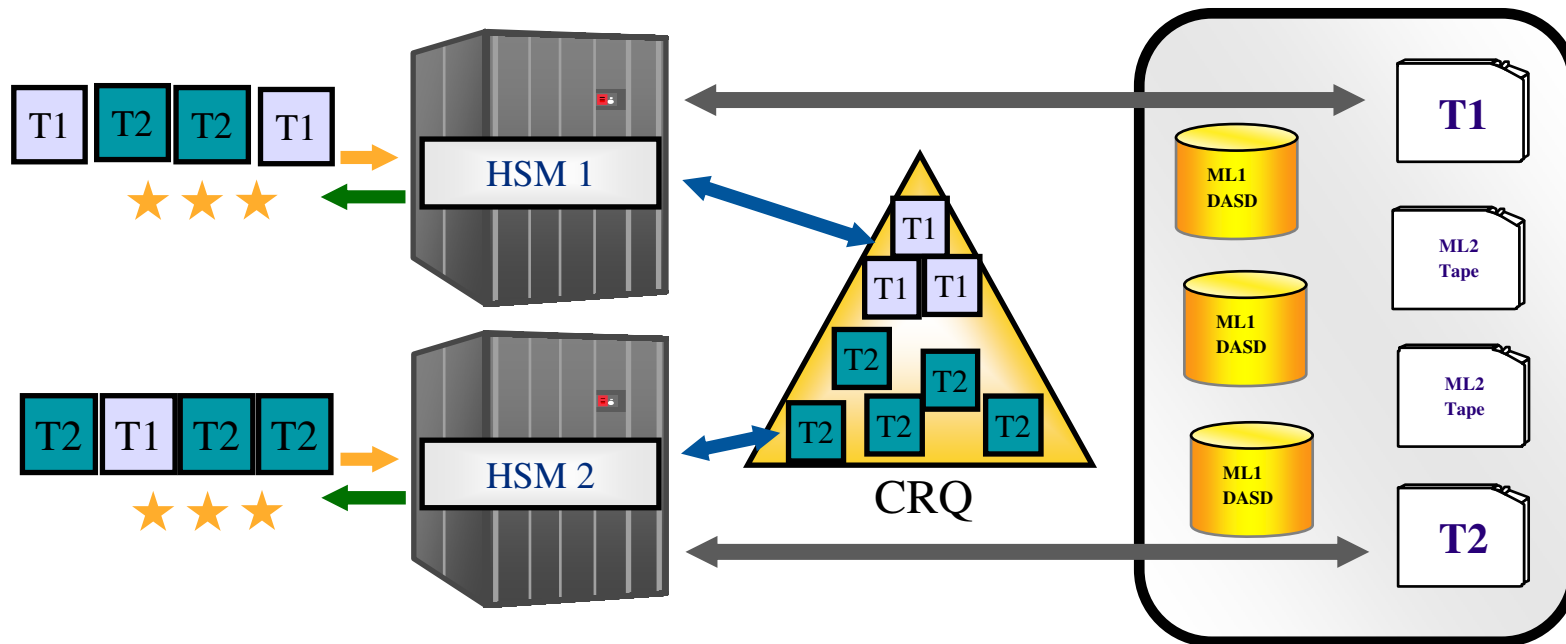
- Workload balancing
  - Requests are evenly distributed among hosts until the maximum tasking level has been reached



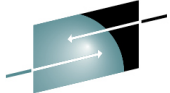
# Recall

## Common Recall Queue (cont)

- Tape Mount Optimization
  - A recall task will process all requests in the CRQ that require the same tape
  - ★ Only a single tape mount is required



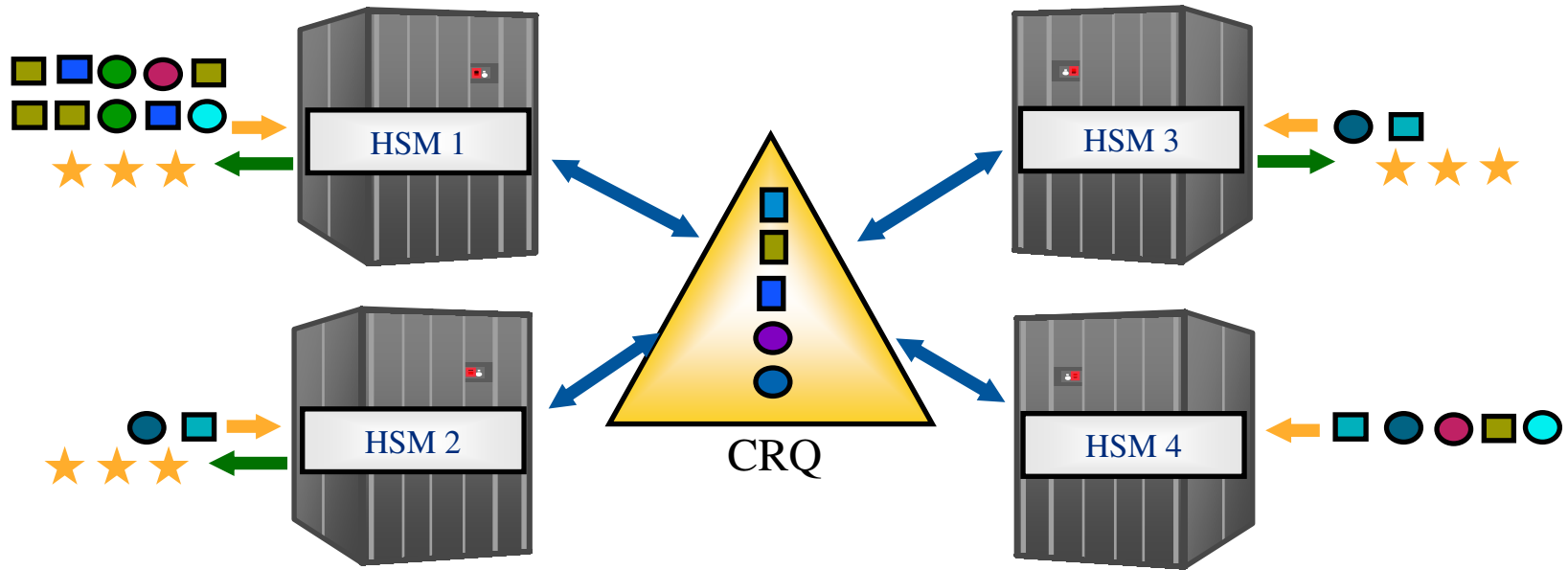




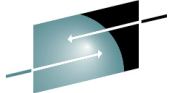
# Recall

## Common Recall Queue (cont)

- Quiesce activity in preparation for a shutdown without holding Recall Activity
  - **HOLD CQ(RECALL(SELECTION))**
  - Places Recalls onto CRQ, but does not process any



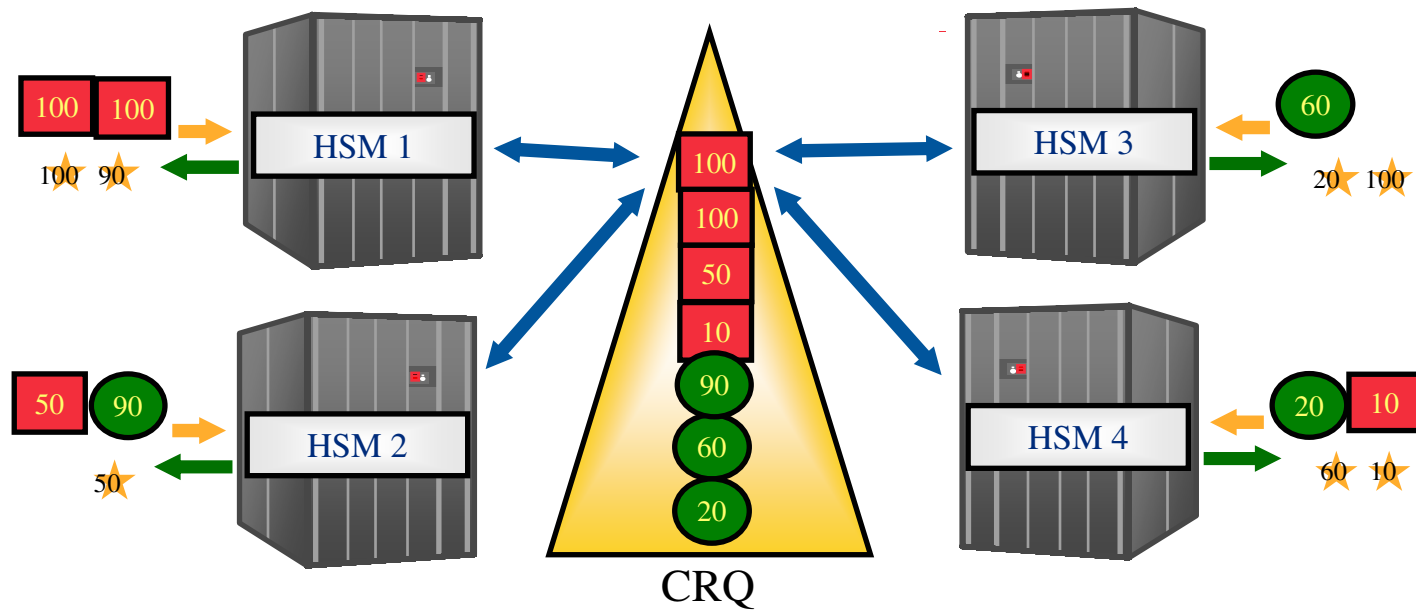
**HOLD CQ(RECALL(SELECTION))**



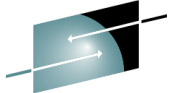
# Recall

## Common Recall Queue (cont)

- Priority Optimization
  - Highest priority requests in the HSMplex are processed first



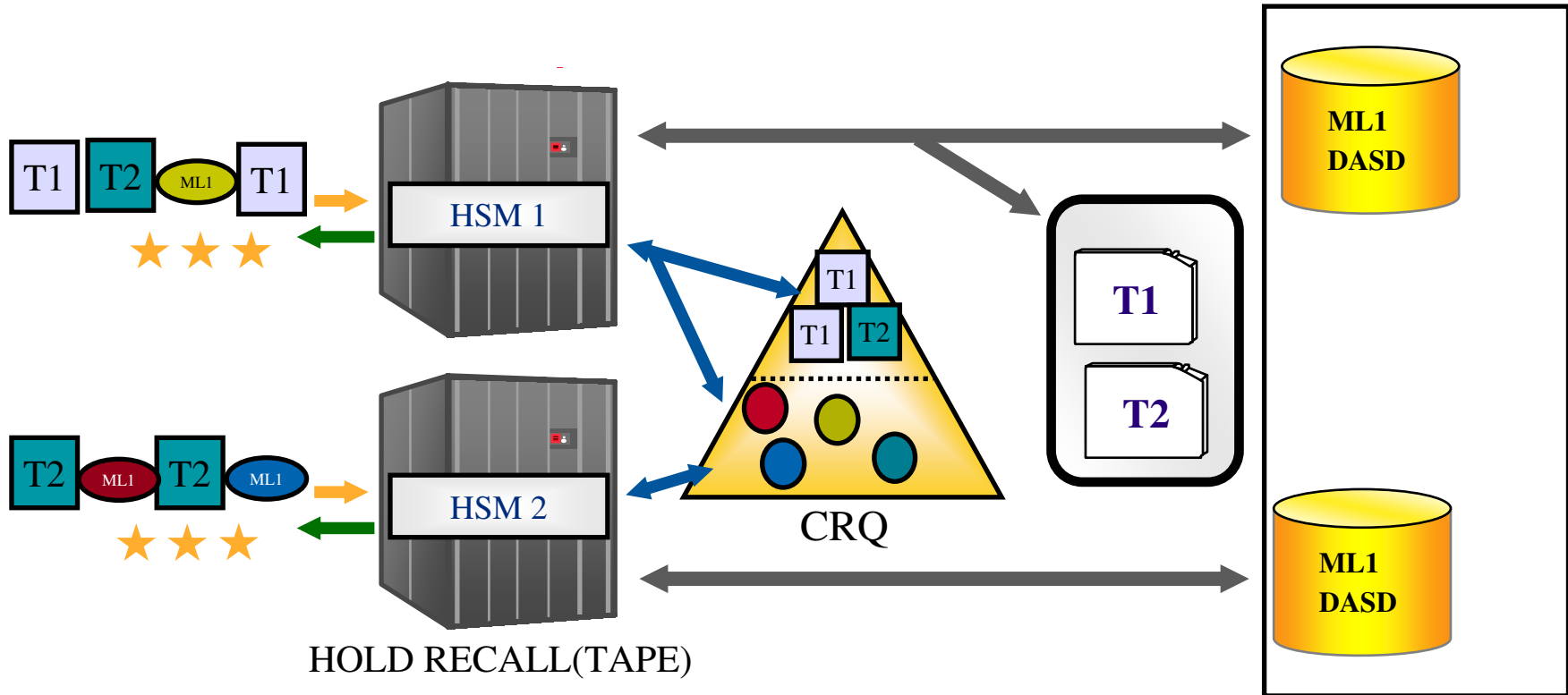
■ = Wait ● = Nowait 100 = Highest 0 = Lowest

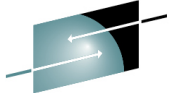


# Recall

## Common Recall Queue (cont)

- Flexible Configurations
  - Hosts not connected to tape drives can be configured to only select non-tape requests

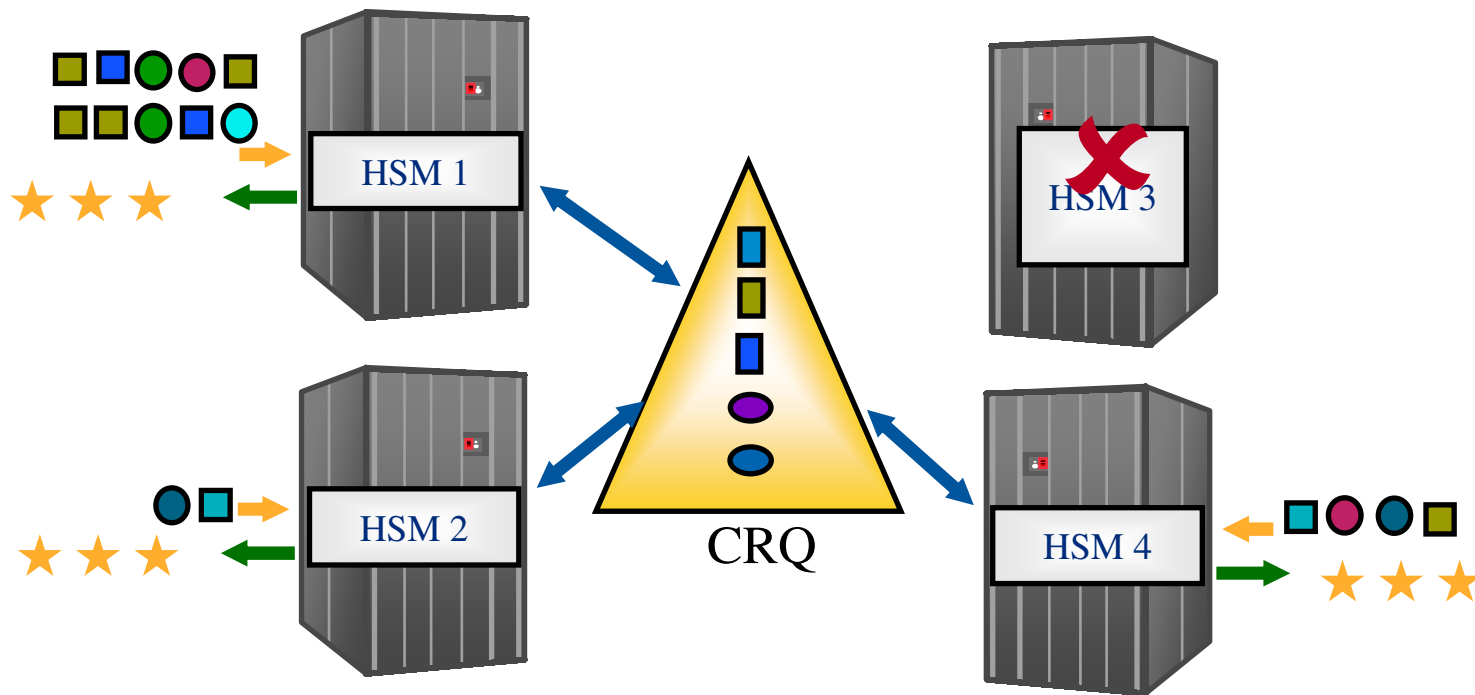




# Recall

## Common Recall Queue (cont)

- Request Persistence
  - Outstanding Recall requests from unavailable hosts are processed by available hosts

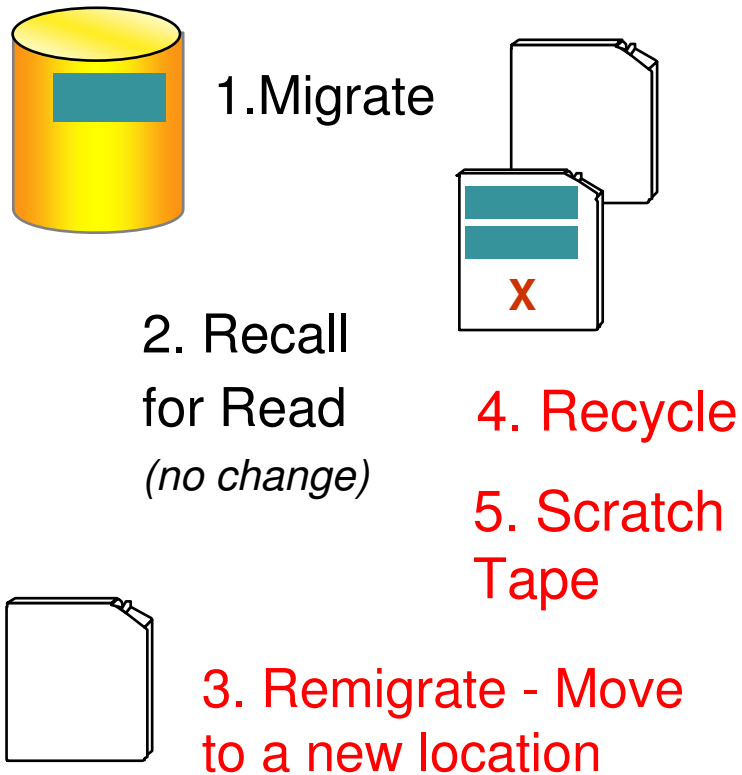


# Migration

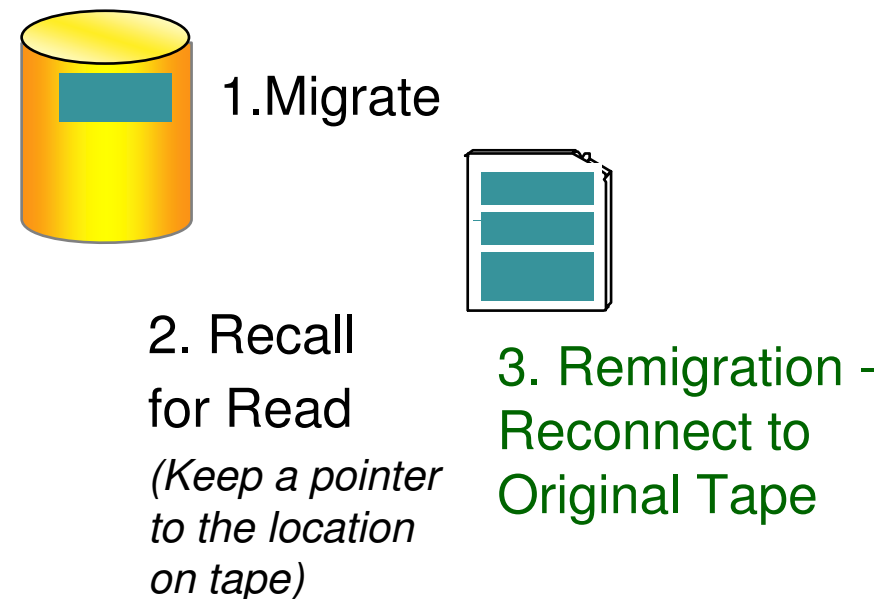
## Fast Subsequent Migration *(cont)*

*Remigrating a data set to tape that was not updated since the Recall...*

### Without FSM



### With FSM



# Migration

## Fast Subsequent Migration *(cont)*

### ★ Advantages

- No actual data movement for reconnection
  - Recycle work load reduced
  - SMS, nonSMS data sets supported
  - Reconnection done automatically and is transparent to user
- 
- DFSMSHsm V1R7 updated this support to no longer rely on Data Set Change Indicator in VTOC to be OFF

SETSYS TAPEMIGRATION(RECONNECT(NONE | ALL | ML2DIRECT))

- **ALL** – Reconnect when data is eligible for either ML1 or ML2
- **ML2DIRECT** – Only reconnect when data is eligible for ML2

# Migration Duplex Tape Error Handling

```
SETSYS DUPLEX(MIGRATION(Y ERRORALTERNATE(CONT | MARKFULL)))
```

- For duplexing of migration tapes, both the original and the alternate will be marked full and two new tapes will be mounted
- ★ Ensures that the original and alternate tapes are always identical
  - Greatly reduces the need for Tape Copies
  - No delay in creating the alternate copy
  - Certain abends require a tape copy to be created

# Migration Management Class Values

- In the Management class, pay attention to “Primary Days Non-usage” and “Level 1 Days Non-usage”
- “Level 1 Days Non-usage” includes time spent on Primary
- Example
  - Primary Days Non-usage = 4
  - Level 1 Days Non-usage = 4
- This results in data sets being migrated **directly to ML2** after 4 days
- In this case, “Level 1 Days Non-usage” should be ‘8’



# Audit Performance

- IBM tape control units dramatically improve the performance of **AUDIT MEDIACONTROLS**
  - Implements a *Hardware Assisted Search* to improve performance
  - Performance will vary depending on data set sizes on tape. Larger data set sizes get more benefit, smaller will get less.
  - CPU utilization is drastically reduced since fewer blocks are read from tape
  - IBM J70 tape control units or newer
    - If required tape CU and microcode not available, then legacy AUDIT code is used

# Audit Mediacontrols



- Audit Mediacontrols can resume processing of migration or backup tape if:
  - AUDIT MEDCTL of a volume is held
  - DFSMSHsm is stopped
  - SETSYS EMERGENCY has been specified
- **RESUME** parameter of AUDIT MEDCTL VOLUMES(*tapevolser*) FIX command
  - AUDIT cannot resume after ABENDS or I/O errors
- **RESUME** only valid when auditing a tape volume
- Valid only when **FIX** parameter is specified

**AUDIT MEDCTL VOLUMES(A00342) RESUME FIX ODS('HSM.FIX')**

# Recycle Limiting Workload


- Use the **LIMIT** parameter to match RECYCLE workload to your scratch tape needs:
  - **LIMIT(50)**: Process enough input tapes to return a net gain of 50 scratch tapes
  - Example: Read 60 input, create 10 output
- Use **PERCENTVALID(0)** to reclaim empty tapes when no drives available

- In a VTS environment
  - Disk backed by Tape:
    - **SETSYS PARTIALTAPE** should be **MARKFULL**
    - **REUSE**
      - *Causes the complete virtual volume to be staged when DFSMSHsm reuses it*
      - *Perhaps worse yet, it causes a 'hole' in the physical tape from which the virtual tape came*
    - **MARKFULL**
      - *Only used portion of virtual volume de-staged to back-end tape*
      - *Does increase the number of virtual volumes required by DFSMSHsm*
  - Disk only:
    - **SETSYS PARTIALTAPE** should be **REUSE**
      - *Above issues do not exist*

# Tape VTS *(cont)*

- Volume size considerations
  - Using a larger size (4GB)
    - Multiple concurrent Recall requests from same volume are single threaded
      - *Can slow down overall throughput*
      - *Reduces mounts, which is positive when the tape has moved to physical tape*
    - Multitasking RECYCLE may be limited if there are fewer larger tapes to recycle
    - AUDIT and TAPECOPY/TAPEREREPLACE are not multitasked, so no impact
    - Reduce instances of reaching 40 volume limit

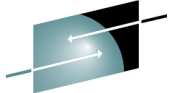
# Tape Connected Sets

- **Connected Set** - sequence of tape volumes connected by valid spanning data
    - Slows down recall and recycle activity
    - More difficult for tape library ejections
- 
- You can minimize the occurrence of connected sets with the judicious use of **SETSYS TAPEUTILIZATION(PERCENTFULL)** and **TAPESPANSIZE** parameters
    - **Never** use TAPEUTILIZATION(NOLIMIT)
    - For TAPESPANSIZE, a larger absolute value is needed to represent the same amount of unused capacity on a percentage basis when the tape has a larger total capacity. eg:
      - If you allow 2% of unused tape to avoid tape spanning for a 3590-Hxx device using enhanced media, you specify a TAPESPANSIZE of 1200 MB.
      - To allow 2% unused tape for a MEDIA5 tape on a 3592 Model J device (no performance scaling), you specify a TAPESPANSIZE of 6000 MB.
    - All size calculations for scaled tapes are based upon the scaled size and not the unscaled size.

# Tape

## Connected Sets *(cont)*

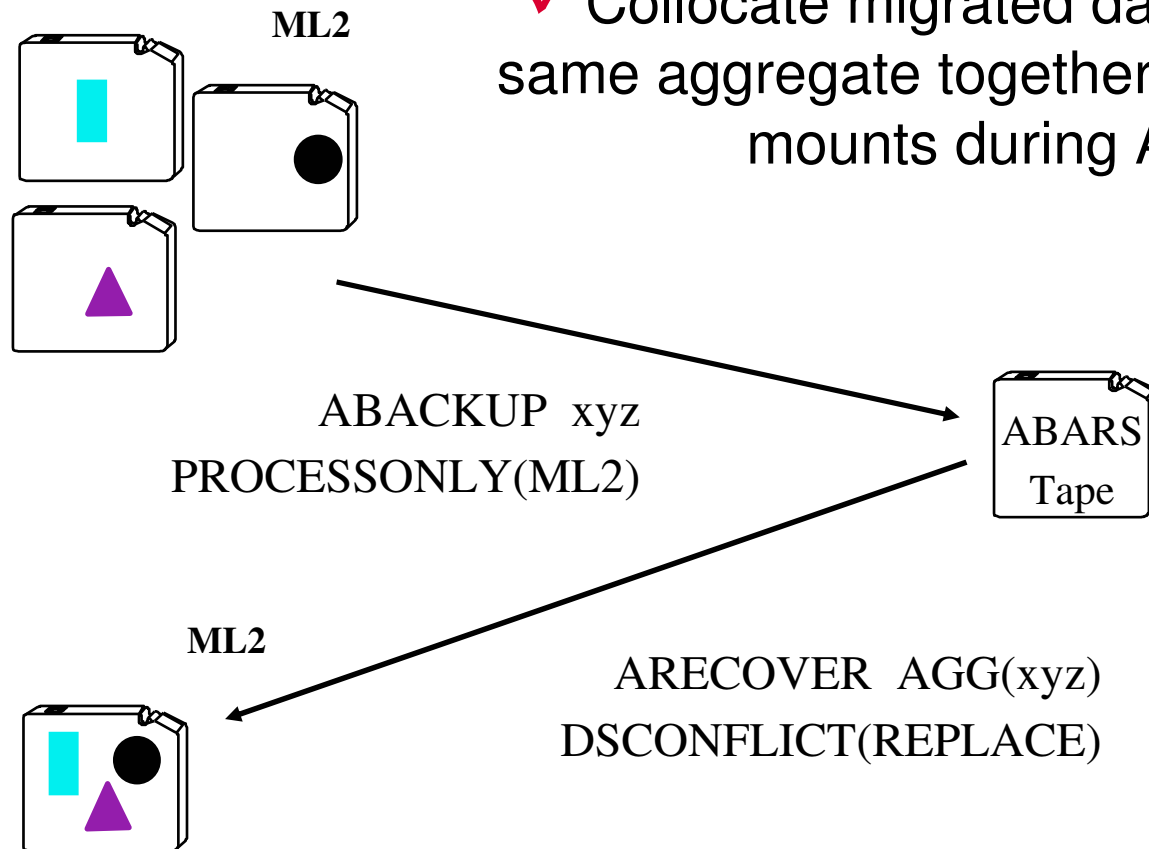
- You can determine if you have connected sets by issuing **LIST TTOC SELECT(CONNECTED)**
- Consider use of new **CHECKFIRST(N)** parameter on generic **RECYCLE** command if significant number of connected sets that meet PERCENTVALID criteria are not being recycled
- You can break a connected set by doing one of the following to the spanning data set
  - Recall a migrated data set
  - Deleting a data set backup using the BDELETE command



# Tape

## Collocate ML2 Data for ABARS

✓ Collocate migrated data sets for same aggregate together to reduce mounts during ABACKUP

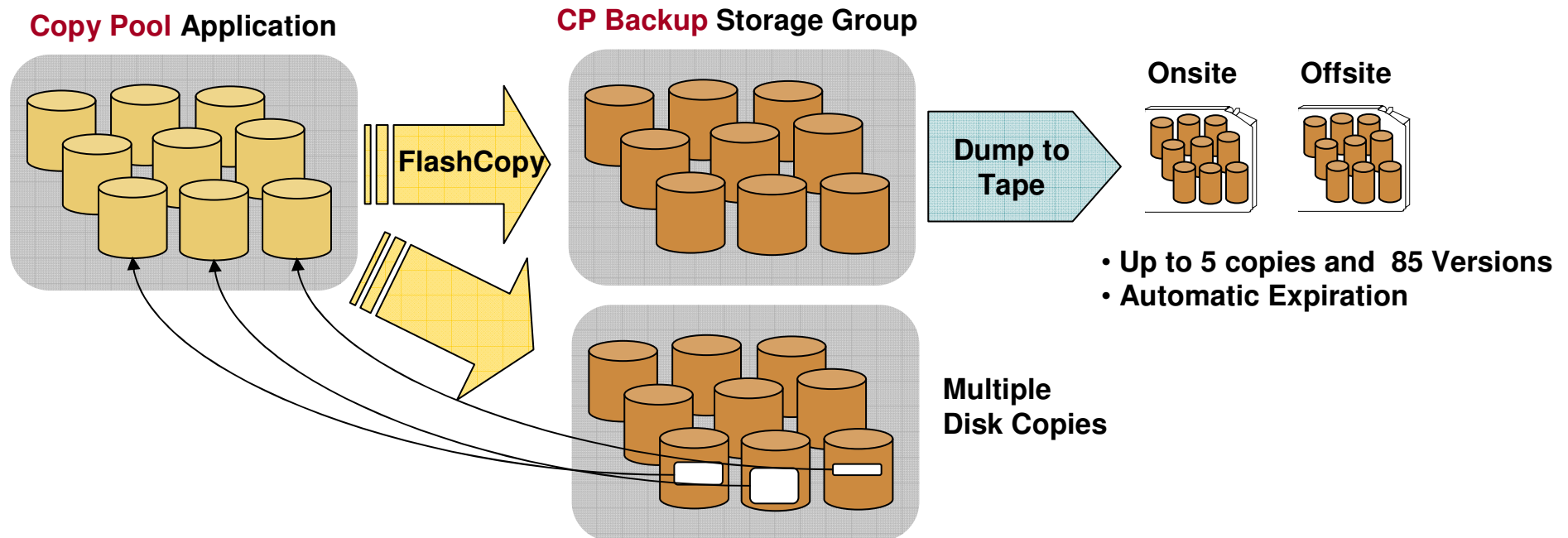




# Fast Replication

## HSM function that manages Point-in-Time copies

- Combined with DB2 BACKUP SYSTEM, provides non-disruptive backup and recovery to any point in time for DB2 databases and subsystems (SAP)

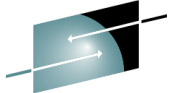


### ★ Recovery at all levels from either disk or tape!

- Entire copy pool, individual volumes and ...
- Individual data sets

# Fast Replication DFSMShsm Advantages

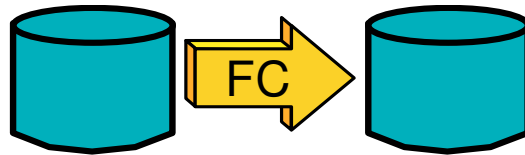
- ★ New Source Volumes always included in backup
- ★ Copy Pool Backup Storage Group disallows allocations on target volumes
- ★ Managed creation/expiration of tape copies
- ★ DFSMShsm ensures valid tape copies
- ★ Data set level recovery from physical backup copies
- ★ Catalog capture during FlashCopy enables deleted data sets to be recovered
- ★ Managed retry of failed recovery volumes



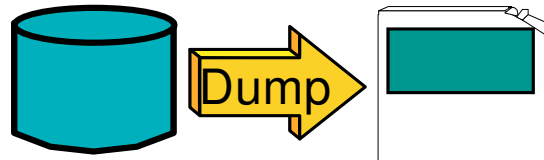
# Fast Replication Data Integrity

- Scenario 1: Dump is in progress

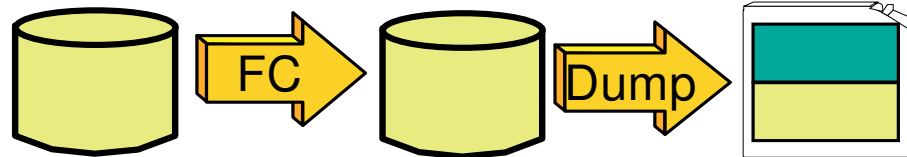
Time 1  
Create Copy



Time 2  
Start Dump

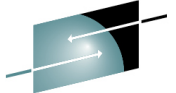


Time 3  
Create new copy  
before dump  
completes



Tape is corrupt.  
It represents a mix of  
two points-in-time.

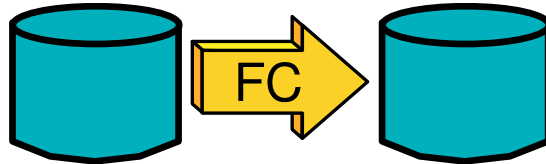
*DFSMShsm prevents this!*



# Fast Replication Data Integrity

- Scenario 2: Dump fails

Time 1  
Create Copy

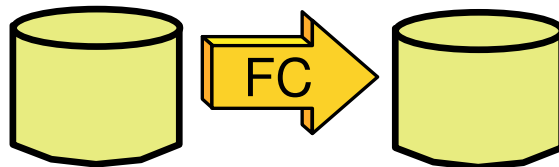


Time 2  
Start Dump



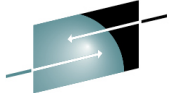
Dump fails.

Time 3  
Create new copy  
before dump  
retried.



Chance to retry failed  
dump is lost because  
PIT copy was overlaid.

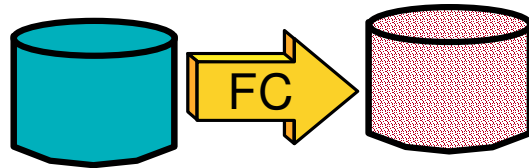
*DFSMShsm prevents this!*



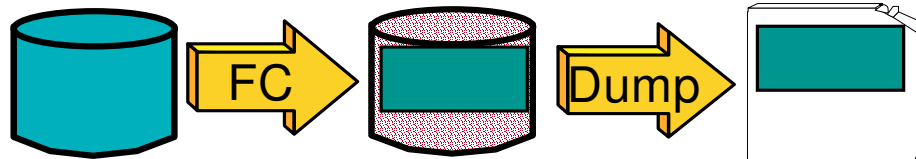
# Fast Replication Data Integrity

- Scenario 3: Relationship is Withdrawn

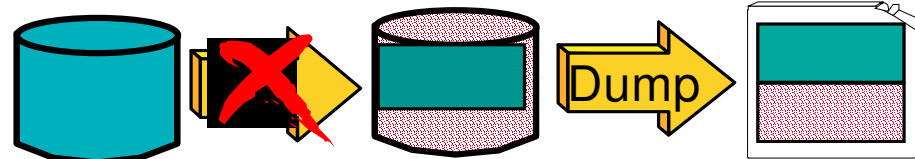
Time 1  
Initiate FC



Time 2  
Start Dump



Time 3  
Withdraw  
Relationship



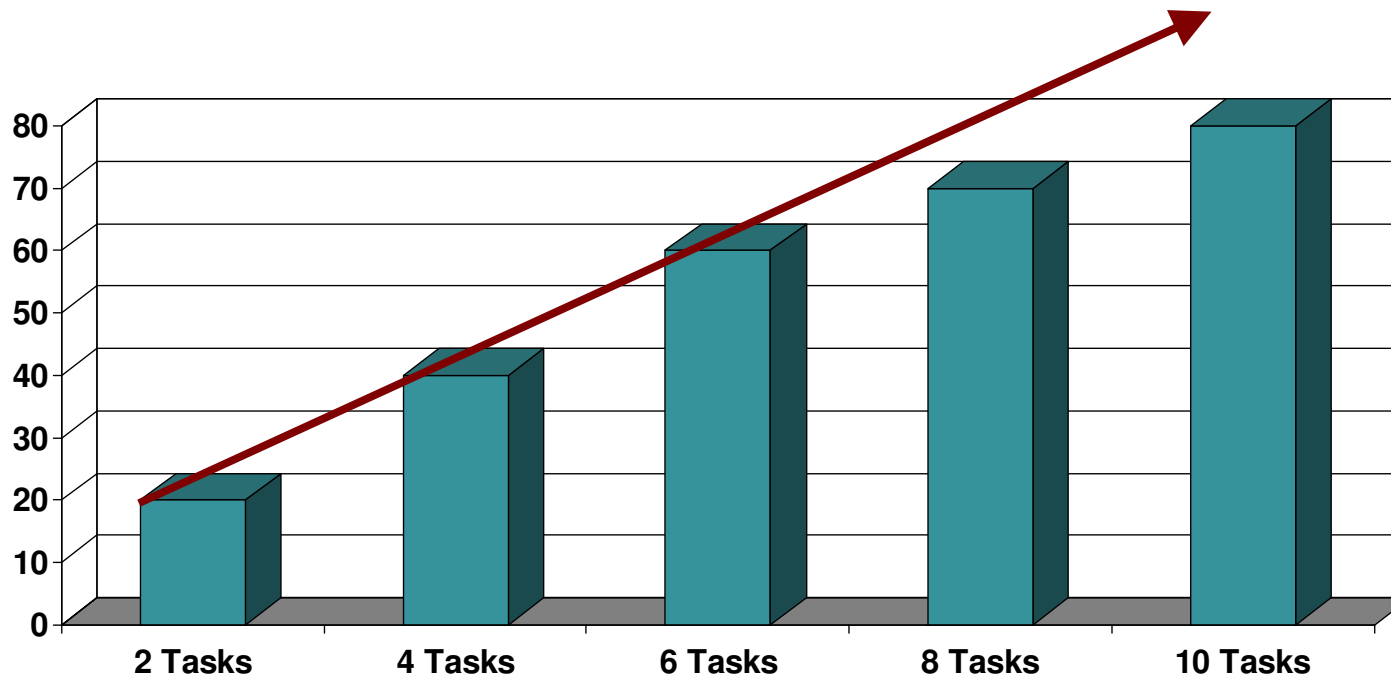
Tape is corrupt.  
Data copied after the  
withdraw is residual.

*DFSMSHsm prevents this!*

*(When Withdraw done with DFSMSHsm)*

# Throughput MASH

- In general, there is a performance ‘knee’ for DFSMShsm functions
- i.e. – the average throughput decreases per task after a certain number of tasks have been started
  - The knee for most functions is at 7-8 tasks
  - For Fast Replication the knee is at 24 of the possible 64 tasks
- Contention for the SYSZTIOT can be one cause

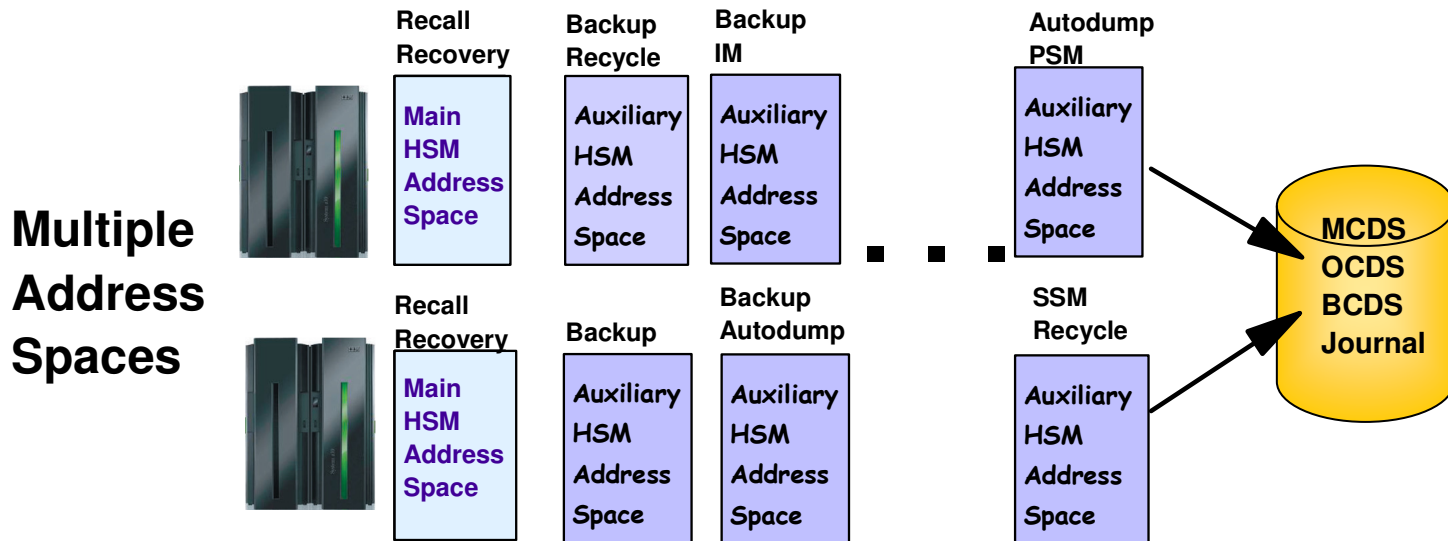
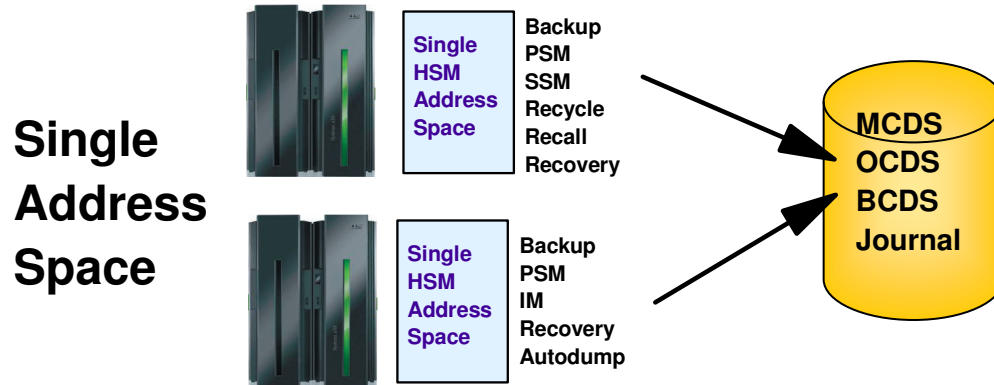


# Throughput

## MASH *(cont)*

- Multiple Address Space HSM (MASH)
  - Each LPAR can have multiple active DFSMSHsm address spaces
  - Up to 39 active DFSMSHsms in an HSMplex
    - HSMplex – All DFSMSHsm's sharing the same control data sets
- Potential benefits of spreading out the DFSMSHsm workload to more hosts
  - Maintain tasks at optimal level
  - Increase overall tasking level
  - Hosts can be assigned different WLM Velocity Goals
  - Recall hosts via Common Recall Queue
    - Start hosts just to process Recall requests during high recall activity
  - Reduces SYSZTIOT contention for disk/tape allocations
  - Increased availability

# Throughput MASH (cont)





# Availability

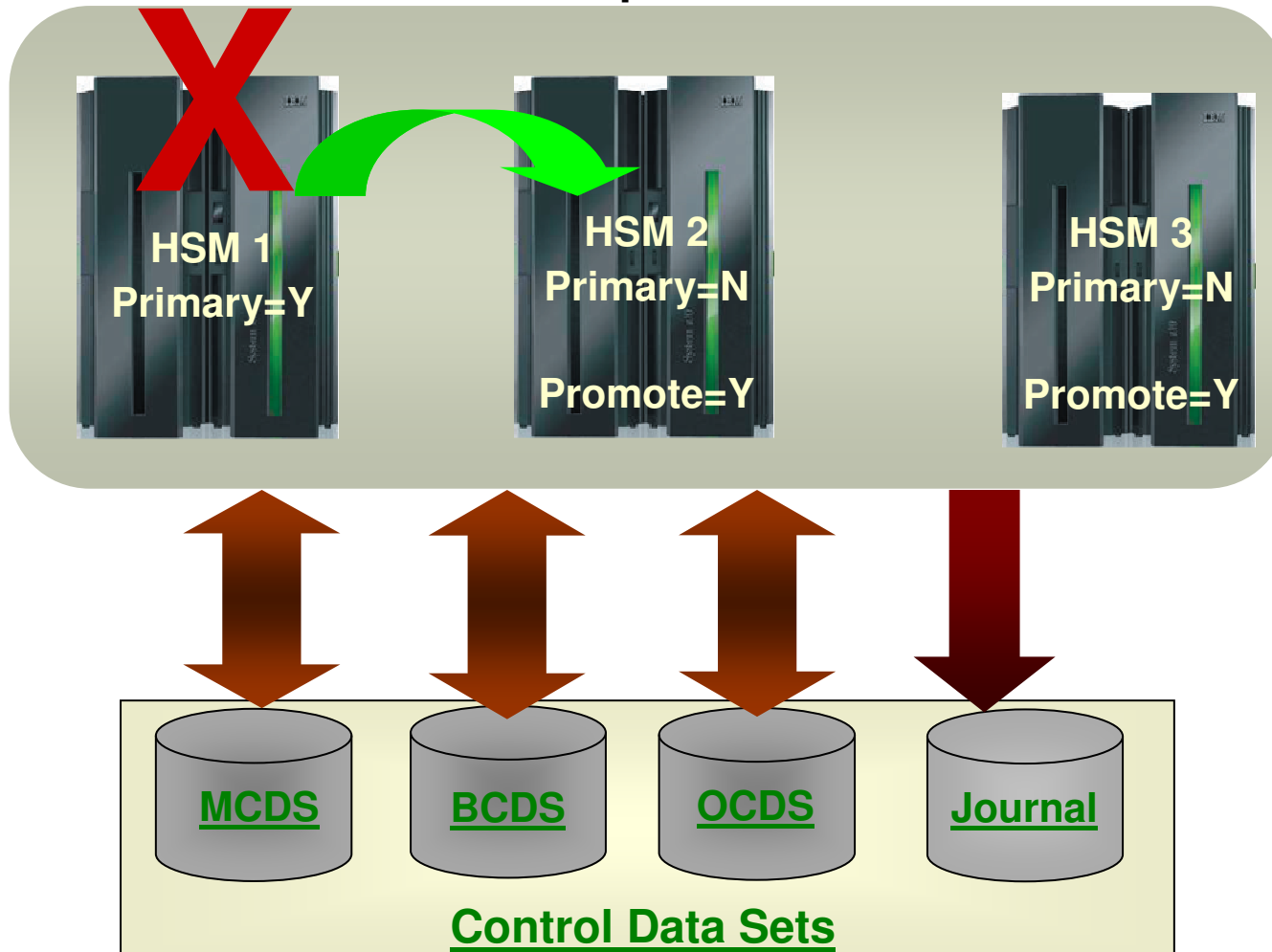
## Secondary Host Promotion

- In an HSMplex, **Secondary Host Promotion** enables secondary DFSMSHsm hosts to take over the *unique* functions being performed by a disabled Primary and/or Secondary Space Management DFSMSHsm host
- A Primary DFSMSHsm host is the only host in an HSMplex that performs:
  - Hourly space checks
  - Automatic CDS backup
  - Automatic movement of backup versions from ML1 to tape
  - Automatic backup of migrated data sets
  - Expiration of dump copies
  - Deletion of excess dump VTOC copy data sets
- There is generally only a single DFSMSHsm host that performs SSM
- ! Without SHP, when either the Primary or SSM host is disabled, the above functions are not performed

# Availability

## Secondary Host Promotion (cont)

### HSMplex



# Availability

## Secondary Host Promotion *(cont)*

DFSMSHsm host must be on a system within a HSMplex that has XCF active and configured in multisystem mode

- SETSYS PLEXNAME(HSMplex\_name\_suffix)
  - Default: ARCPLEX0
  - Must be specified if more than one HSMplex within a sysplex. Must be specified on all hosts in that HSMplex.
  - Must be specified in ARCCMDxx member
- SETSYS PROMOTE(PRIMARYHOST(Y|N) SSM(Y|N))
  - Default: No
  - PRIMARYHOST(Y) is ignored for Primary host
  - A SSM host cannot be promoted for another SSM host. ARC1521I issued if SSM(Y) specified on a SSM host

# Availability

## Cancel Active Tasks

- Function enables hung tasks to be cancelled
  - Reduces need to restart DFSMSHsm due to a hung task
- **QUERY ACTIVE (TCBADDRESS)**
  - Requests only messages directly related to specific data movement activities be listed along with extra information allowing cancellation of the active task
- **CANCEL TCBADDRESS() or SASINDEX()**
  - **TCBADDRESS** - Address of TCB returned from Q ACTIVE
  - **SASINDEX** - Index to be used when canceling an ABARS task
    - Can be done instead of canceling the ABARS started task

# Performance

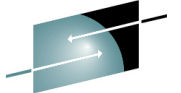
## SMF Consolidation Processing

- Specify **DDCONS(NO)** on SMF parameters to avoid DD name consolidation during shutdown
  - DDCONS is specified in SMFPRMnn parmlib member
  - See *MVS Initialization and Tuning Reference* for more information
- DFSMShsm is a started task with thousands of DD name allocations
- DFSMShsm shutdown may be delayed up to 45 minutes if consolidation is performed
- SMF Type 30 records are a bit longer

# Performance

## Avoid LOG Overhead

- Use **HOLD LOG** to avoid DFSMSHsm logging overhead
    - Command can be added to PARMLIB
  - Turns off writing to the LOGX/LOGY data sets
  - Information available elsewhere, such as FSR records in SMF, Activity Logs, PDA trace data
  - Reduces DFSMSHsm overhead activity
- ✓ *Some ISV products require the LOGX/LOGY data sets as input*



# Performance

## Miscellaneous

- Asynchronous scratch of migrated nonVSAM data sets
- RECYCLE SYNCHDEV at intervals
- Multitask Secondary Space Management
- Overlap Phase 1 and Phase 2 of Volume Migration
- No Recall on an IEFBR14 Delete and DELETE GDG FORCE

# DFSMSHsm Reporting Report Generator

## Generate reports of DFSMSHsm functions and inventory using DFSMSrmm Report Generator

- DFSMSrmm Report Generator is an easy-to-use ISPF application
  - Create and customize reports specific to your needs
  - *Available without a DFSMSrmm license*
    - New option on ISMF panel to create 'Storage Management' reports
  - Sample Reports shipped in SYS1.SAMPLIB

## DFSMSHsm reporting based on

- DFSMSHsm Function Statistics Record (FSR)
- DFSMSHsm ABACKUP/ARECOVER Function Statistics Record (WWFSR)
- DFSMSHsm Inventory (control data set) data via DCOLLECT



# DFSMSHsm Reporting Report Generator



Migration Age of zero when data set is recalled

DFSMSHsm Thrashing Report

- 1 -

2008/02/04

15:06:18

DSN	AGE	SIZE KB	MC NAME
<b>HSMATH0.SMS.VBGPS1</b>	<b>0000</b>	<b>36830</b>	<b>MCLASS1</b>
HSMATH0.SMS.VSMALNA	0000	159	MCLASS1
HSMATH0.SMS.VSMALNB	0000	159	MCLASS1
HSMATH0.SMS.VSMALNC	0000	159	MCLASS1
HSMATH0.SMS.VSMALND	0000	159	MCLASS1
HSMATH0.SMS.VSMALNE	0000	159	MCLASS1
HSMATH0.SMS.VSMALNF	0000	159	MCLASS1
HSMATH0.SMS.VSMALNG	0000	159	MCLASS1
HSMATH0.SMS.VSMALNH	0000	159	MCLASS1
HSMATH0.SMS.VSMALNI	0000	159	MCLASS1
HSMATH0.SMS1.PS.TEST0	0000	3	MCLASS1
HSMATH0.SMS1.PS.TEST1	0000	3	MCLASS1
HSMATH0.SMS2.PS.TEST2	0000	3	MCLASS1

Other fields included in the sample report:

- Date; Elapsed time
- Target volume
- Return Code / Reason Code

# DFSMSHsm Reporting

## FSRSTAT

- **FSRSTAT** is a REXX sample program that reads DFSMSHsm FSR records, and generates a statistical summary report
- Shipped with DFSMSHsm
  - SYS1.SAMPLIB(ARCTOOLS)
- Since it is written in REXX:
  - Does not require any special programs or languages (SAS, MICS, etc.)
  - It can be easily modified and customized to meet your needs
  - It can be slow, consider running in batch using PGM=IKJEFT01
  - Requires input data to be converted to RECFM=VB format

# DFSMShsm Reporting

## FSRSTAT

### FSR records by Size (KB)

0 ->	49	84320	45.9%	45.9%	<b>Nearly half!</b>
50 ->	149	16568	9.0%	54.9%	
150 ->	749	26290	14.3%	69.2%	
750 ->	29MB	43066	23.4%	92.7%	
30MB ->	7GB	13464	7.3%	100.0%	

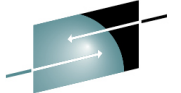
Average 12343.72 KB ← **Misleading**

### By rate (KB/sec)

0 ->	499	137052	74.6%	74.6%	<b>Smaller Data Sets</b>
500 ->	999	10318	5.6%	80.2%	
1000 ->	1499	6073	3.3%	83.5%	
1500 ->	1999	4550	2.5%	86.0%	
2000 ->	2499	6443	3.5%	89.5%	<b>Larger Data Sets</b>
2500 ->	2999	3219	1.8%	91.3%	
3000 ->	9999	16053	8.7%	100.0%	

Average 808.55 KB/sec

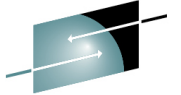
# Just For Fun



**SHARE**  
Technology • Connections • Results

- Penalize a user who continuously Recalls hundreds/thousands of data sets on a frequent basis by periodically moving all their requests to the bottom of the queues:

**ALTERPRI USERID(*anyuser*) LOW**



# Summary

- Improve performance
- Work smarter
- Exploit new functions
- Exploit technology