



DFSMShsm Best Practices

Glenn Wilcock wilcock@us.ibm.com

March 15, 2012 Session 10952



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 intellectually 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



SHARE Isthalogy - Consections - Results

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

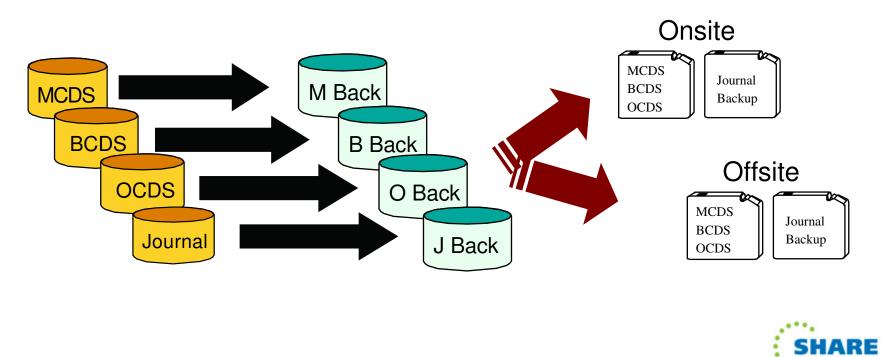


Control Data Sets Duplex CDS Backup Copies



in Atlanta

- 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



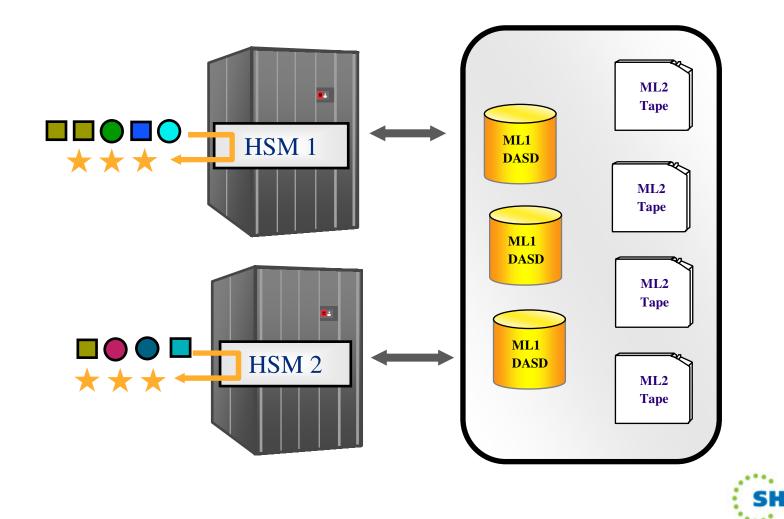
Recall Common Recall Queue



E in Atlanta

2012

• NonCRQ environment – Each host processes own requests

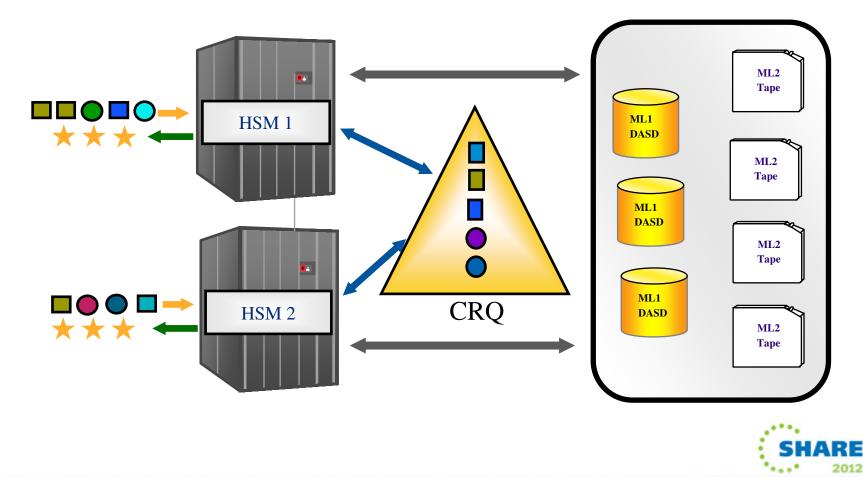


13



Atlanta

- 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



Advantages of CRQ

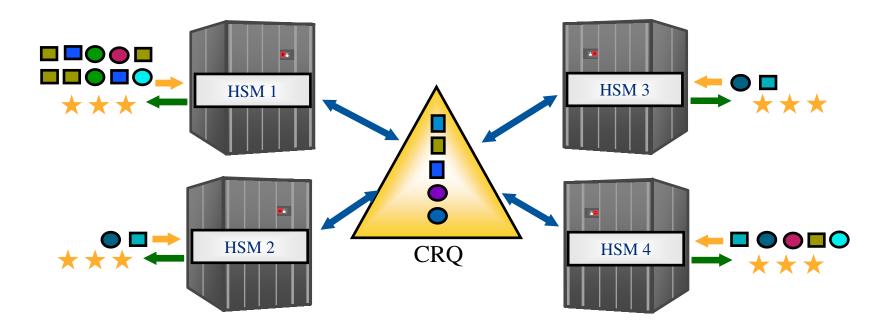
- Workload balancing
- Tape mount optimization
- Quiesce Activity w/o impacting Recall
- Priority optimization
- Flexible configurations
- Request persistence







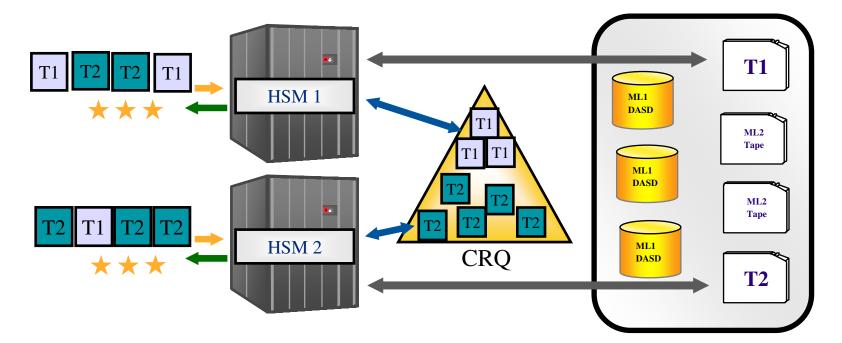
- Workload balancing
 - Requests are evenly distributed among hosts until the maximum tasking level has been reached







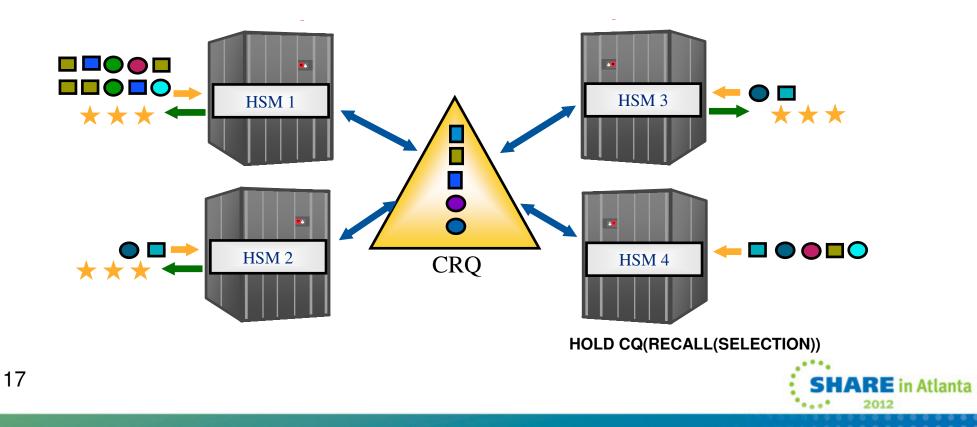
- Tape Mount Optimization
 - A recall task will process all requests in the CRQ that require the same tape
 - ★ Only a single tape mount is required





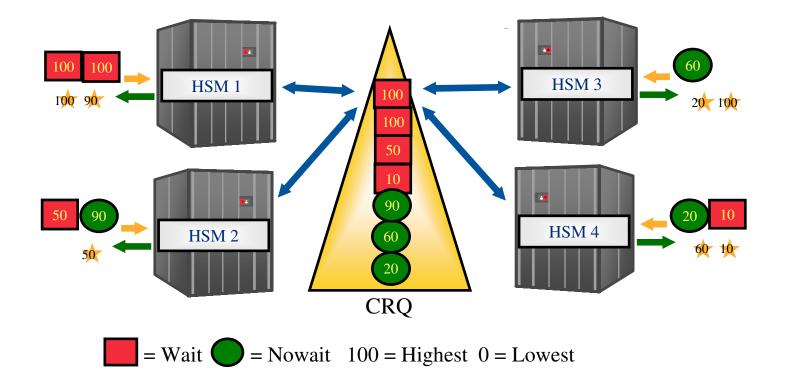


- Quiesce activity in preparation for a shutdown without holding Recall Activity
 - HOLD CQ(RECALL(SELECTION))
 - Places Recalls onto CRQ, but does not process any





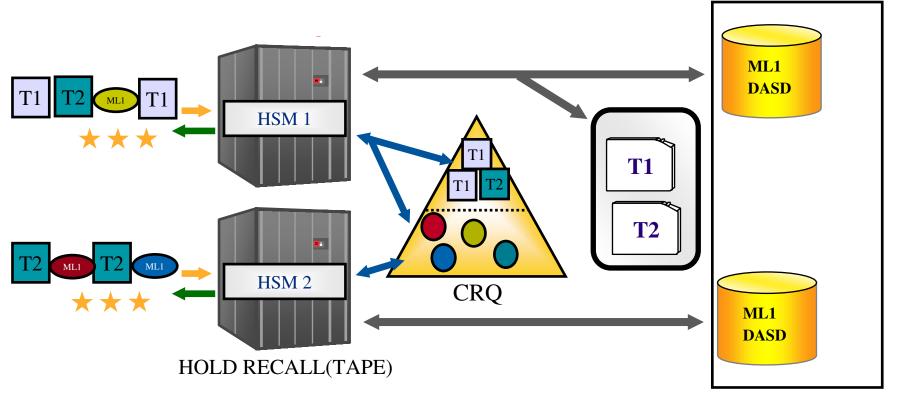
- Priority Optimization
 - Highest priority requests in the HSMplex are processed first







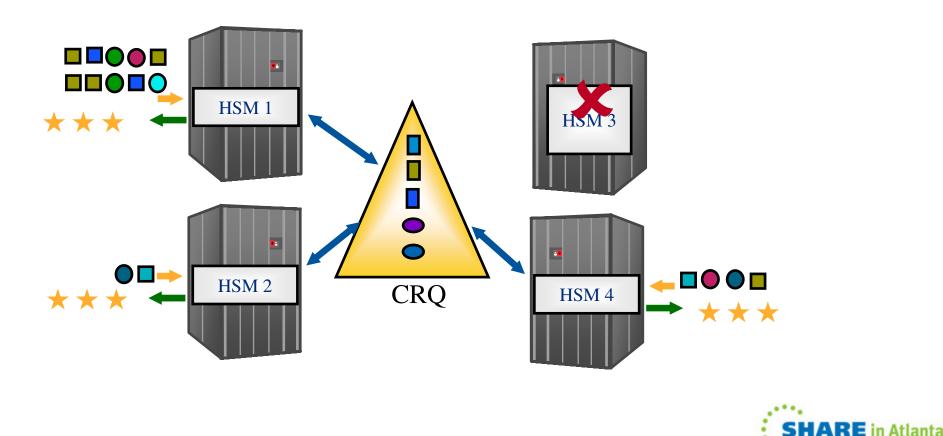
- Flexible Configurations
 - Hosts not connected to tape drives can be configured to only select non-tape requests







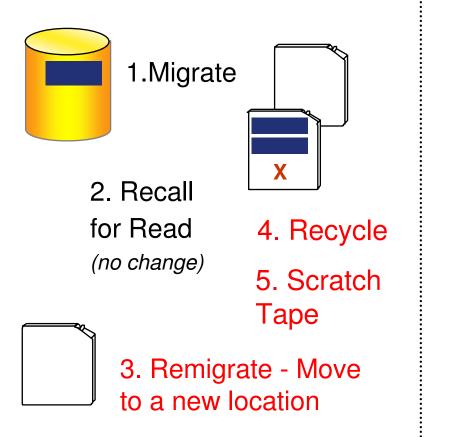
- Request Persistence
 - Outstanding Recall requests from unavailable hosts are processed by available hosts



Migration Fast Subsequent Migration

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

Without FSM



1.Migrate

With FSM

2. Recall for Read (Keep a pointer to the location on tape)

 Remigration -Reconnect to Original Tape





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(<u>NONE</u> | ALL | ML2DIRECT))

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

SETSYS MIGRATIONCLEANUPDAYS(recalldays statdays reconnectdays)

reconnectdays – Number of days to keep records for migrated data sets that
 are candidates for reconnection

Migration Duplex Tape Error Handling



SETSYS DUPLEX(MIGRATION(Y ERRORALTERNATE(CONT | <u>MARKFULL</u>)))

 For duplexing of migration tapes, both the original and the alternate will be marked full and two new tapes will be mounted

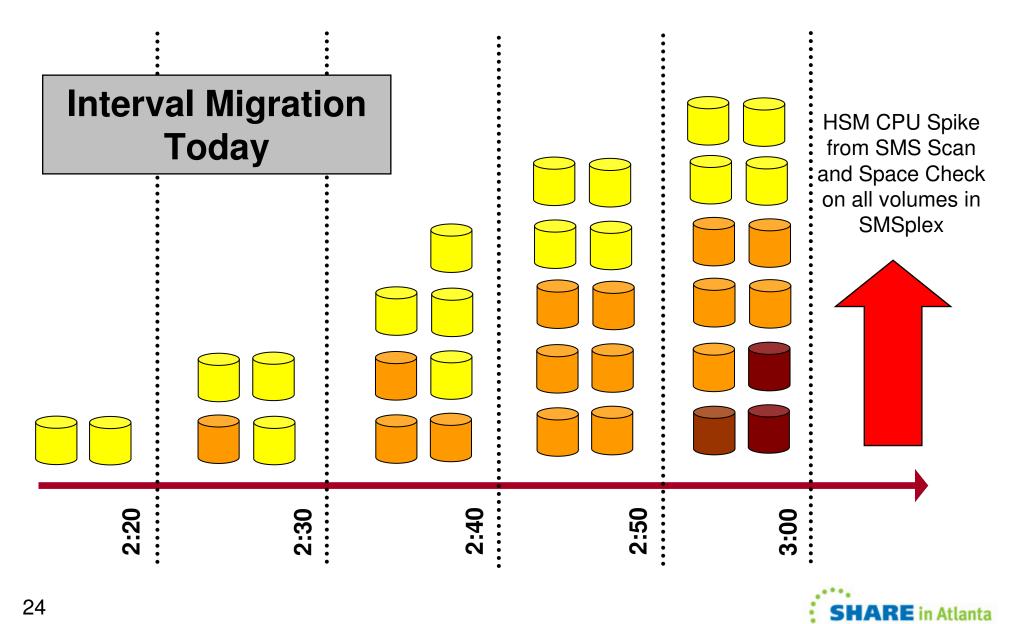
★Ensures that the original and alternate tapes are always identical

- Greatly reduces the need for Tape Copies
- No delay in creating the alternate copy
- Certain abends require a tape copy to be created



Migration V1R13 On Demand Migration

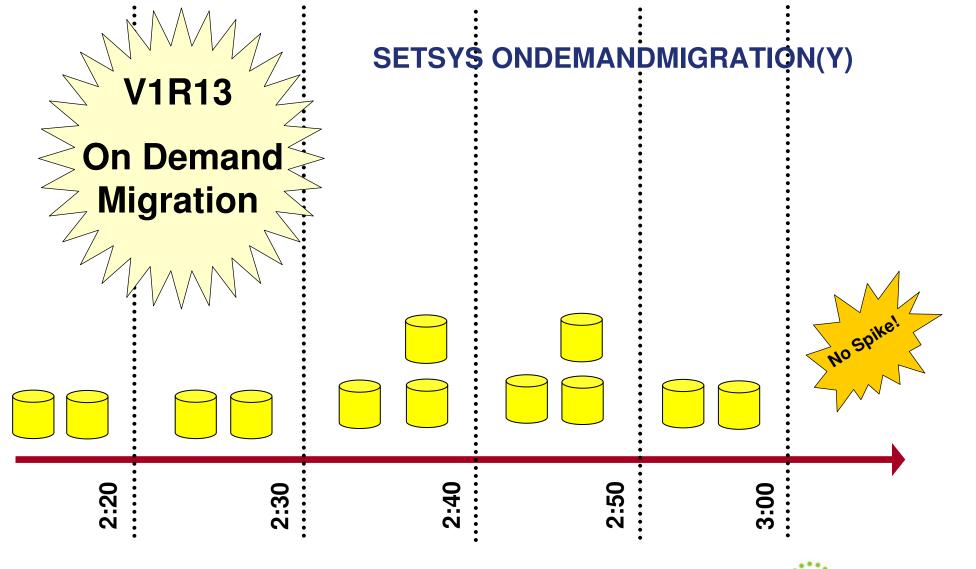




Migration V1R13 On Demand Migration



in Atlanta



Migration Management Class Values



- In the Management class, pay attention to "Primary Days Nonusage" 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'



Migration Small Data Set Packing



- The benefits of utilizing Small Data Set Packing
 - More efficient use of ML1 space
 - A migrated data set can take as little of 2K of space in an SDSP
 - As much as 24:1 compaction
 - More efficient migration processing of data to ML1
 - No need to perform a data set allocation, open & close for each migration data set
 - Only performed once for the SDSP



Migration Small Data Set Packing



- Recent enhancements that make SDSPs more efficient
 - V1R12
 - CA Reclaim Reduces the need to re-org SDSPs
 - V1R13
 - Updated SDSP selection algorithm
 - DFSMShsm now selects SDSPs based on highest freespace
 - Prior release select SDSPs based on ADDVOL order, so certain SDSPs were overused
 - Updated serialization logic
 - SDSPs can now have multiple concurrent 'readers'
 - Updated location of MM exit invocation
 - MM exit can be used to skip a data set for ML1 -> ML2 processing
 - Exit is now invoked up front, before serialization / queuing of data set



Migration General



- Do not make significant SMS configuration changes all at one time:
 - Expire after Days Non-usage
 - Expire after Date/Days
 - Level 1 Days Date/Days

Example

- You need to decrease Level 1 Days Date/Days from 100 to 70
- If that change is made all at once, there will be a significant spike in ML1 -> ML2 workload
- Instead, make the change gradually
 - 100 -> 90
 - 90 -> 80
 - 80 -> 70



Audit Mediacontrols



- Audit Mediacontrols can resume processing of a 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



Tape VTS



- 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)



- Other benefits of a Disk-Only environment
 - Eliminate ML1
 - Eliminates HSM Compaction CPU overhead
 - Eliminates Secondary Space Management
 - Faster Recalls
 - Data Deduplication for Backup Data
 - Storage Controller features of Replication, etc



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
 - Spanning data sets cannot be reconnected during migration
- You can minimize the occurrence of connected sets with the judicious use of SETSYS TAPEUTILIZATION(PERCENTFULL) and TAPESPANSIZE parameters
 - Never use TAPEUTILIZATION(NOLIMIT)





Tape Connected Sets (cont)

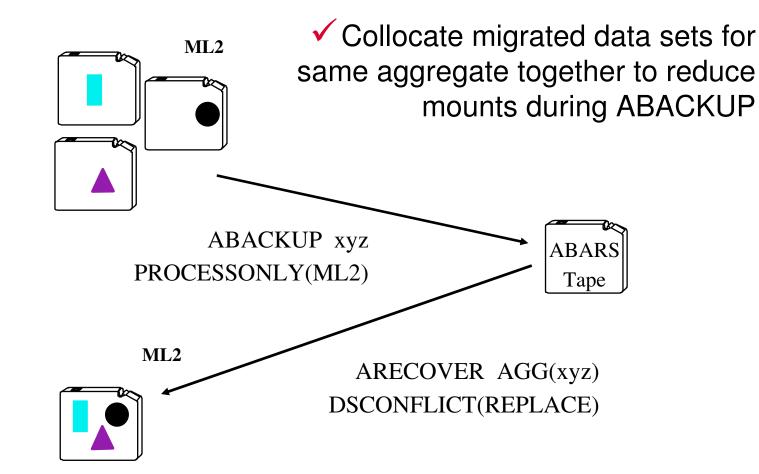


- You can determine if you have connected set by issuing LIST TTOC SELECT(CONNECTED)
- Periodically use the CHECKFIRST(N) parameter on generic RECYCLE commands if a significant number of connected sets that meet the PERCENTVALID criteria are not being recycled
- You can break a connected set by doing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



Tape Collocate ML2 Data for ABARS





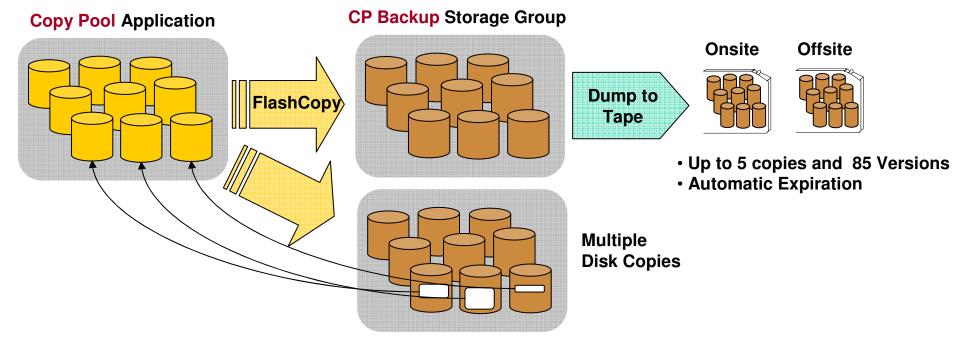


Fast Replication



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 volume recoveries



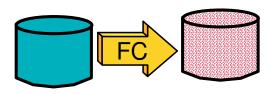
Fast Replication

Data Integrity

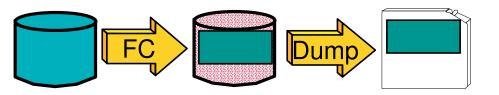
SHARE Istraligy - Canadians - Pasuta

Scenario: FlashCopy Relationship is Withdrawn

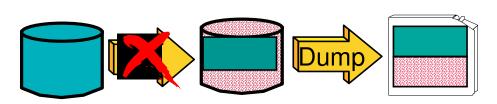
Time 1 Initiate FC



Time 2 Start Dump



Time 3 Withdraw Relationship



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

DFSMShsm prevents this!

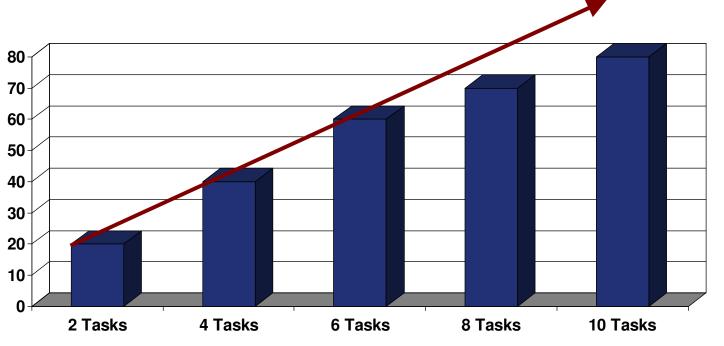
(When Withdraw done with DFSMShsm)



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) Backup Single PSM HSM SSM **Address** Recycle MCDS Single Space Recall OCDS Recovery **Address** BCDS Journal Backup Space Single PSM HSM IM **Address** Recovery Space Autodump Recall Backup Autodump Backup Recovery IM PSM Recycle Auxiliary Auxiliary Auxiliary Main HSM HSM HSM HSM **Address** Address Address Address Space **Multiple** MCDS Space Space Space OCDS **Address** SSM Backup BCDS Recall Autodump Recycle Backup Recoverv Journal **Spaces** Main Auxiliary Auxiliary Auxiliary **HSM** HSM HSM HSM Address

Address

Space

Space

Address

Space

Address

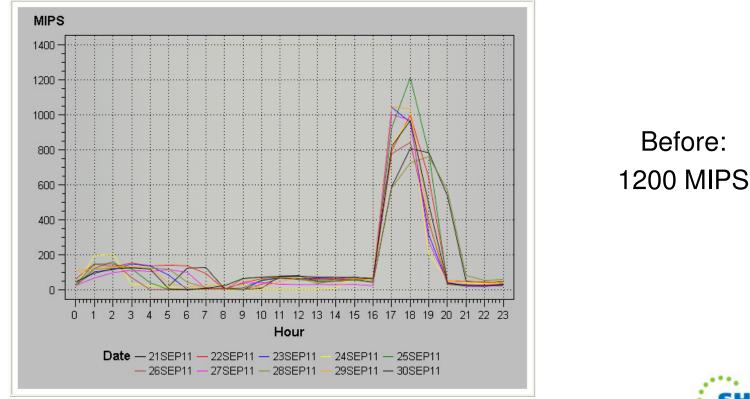
Space



Throughput Tune Maximum Tasking Levels



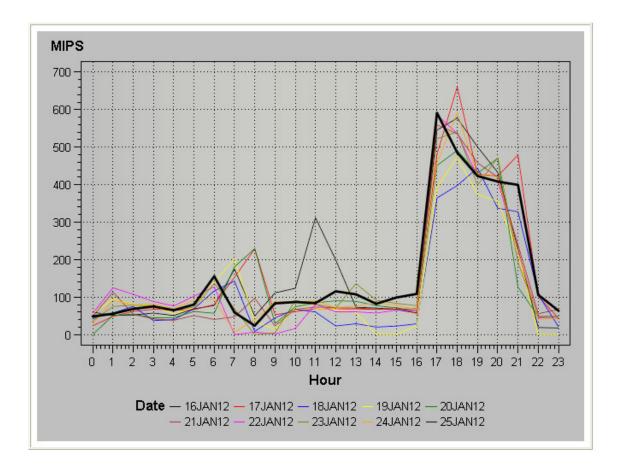
- Tune the maximum number of concurrent DFSMShsm tasks to minimize DFSMShsm's peak CPU consumption
 - Charts from Customer C showing peak DFSMShsm CPU before and after they
 minimized the number of DFSMShsm concurrent tasks to complete the workload
 within the allotted timeframe





Throughput Tune Maximum Tasking Levels





After: 600-700 MIPS



Availability Secondary Host Promotion

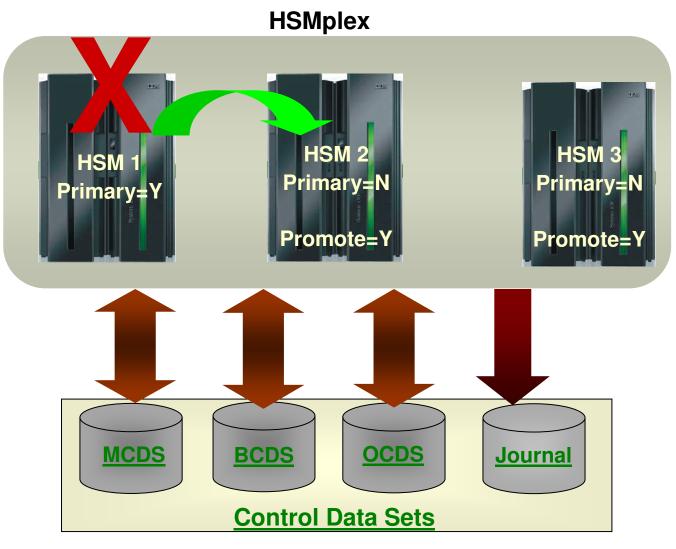


- In an HSMplex, Secondary Host Promotion enables secondary DFSMShsm hosts to take over the *unique* functions being performed by a disabled Primary and/or Secondary Space Management DFSMShsm host
- A Primary DFSMShsm host is the only host in an HSMplex that performs:
 - Hourly space checks
 - Automatic 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)







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. ARC15211 issued if SSM(Y) specified on a SSM host



Performance SMF Consolidation Processing



- Specify DDCONS(NO) on SMF parameters to avoid DD name consolidation during DFSMShsm shutdown
 - DFSMShsm shutdown has been reported to take up 45 minutes due to DD consolidation
 - DDCONS is specified in SMFPRMnn parmlib member
 - See MVS Initialization and Tuning Reference for more information
- Pros of DDCONS(NO):
 - Faster HSM shutdown
 - Less likelyhood of periodic slowdowns
- Cons of DDCONS(NO):
 - Lots more SMF type 30 records
 - Higher SMF filling/swapping rates



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



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
 - 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
HSMATHO.SMS.VBGPS1		0000	36830	MCLASS1
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 rec 0-> 50-> 150-> 750-> 30MB ->	49 149 749	y Size (KE 84320 16568 26290 43066 13464	3) 45.9% 9.0% 14.3% 23.4% 7.3%	45.9% 54.9% 69.2% 92.7% 100.0%	Nearly half!
Aver	age	12343.72	KB - Misl	eading	
By rate 0-> 500-> 1000-> 1500-> 2000-> 2500-> 3000->	2999	ec) 137052 10318 6073 4550 6443 3219 16053	74.6% 5.6% 3.3% 2.5% 3.5% 1.8% 8.7%	74.6% 80.2% 83.5% 86.0% 89.5% 91.3% 100.0%	Smaller Data Sets Larger Data Sets

Average

808.55 KB/sec



Just For Fun



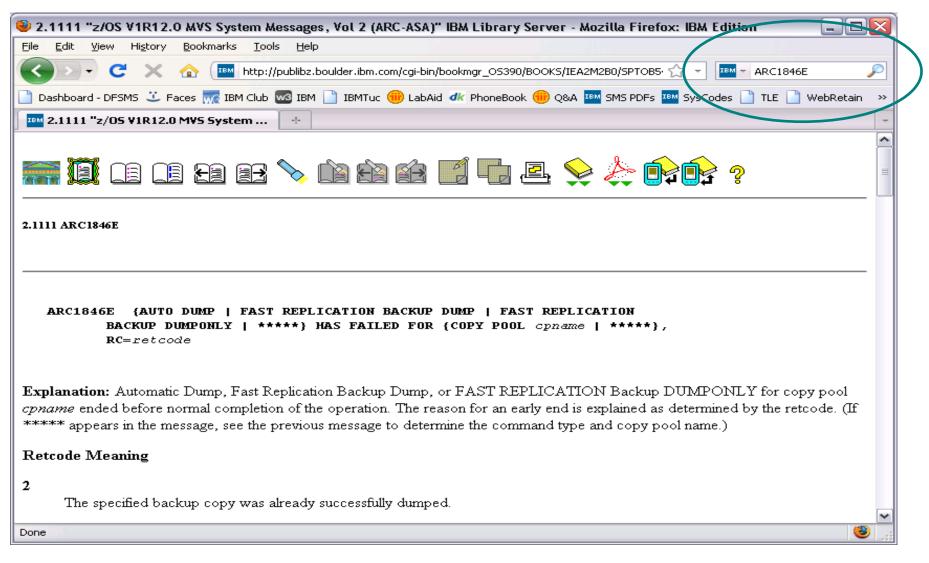
 Penalize a user who continuously Recalls hundreds/thousands of data sets on a frequent basis by periodically moving all their requests to the bottom of the queues:

ALTERPRI USERID(anyuser) LOW



Just For Fun





http://www-03.ibm.com/systems/z/os/zos/bkserv/lookat/lookatalerts.html





Summary

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