IBM Systems

DFSMShsm best practices – Part 1

Legal Disclaimer

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

Any statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

References in this document to IBM products, programs, or services does not imply that IBM intends to make such such products, programs or services available in all countries in which IBM operates or does business. 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 NONINFRINGEMENT. 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.

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:

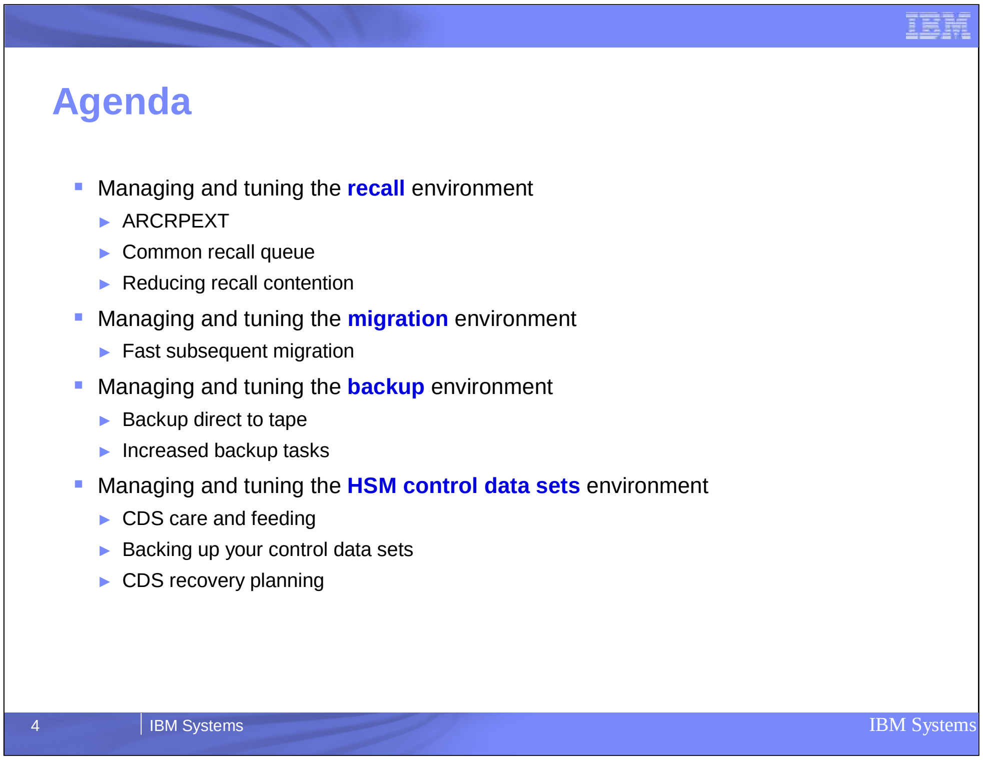
IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

Trademarks

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

DFDSS
DFHSM
DFSMSdfp
DFSMShsm
DFSMSdss

IBM®
System/390®
SNAP/SHOT®
Enterprise Storage Server®
FlashCopy®
OS/390®
z/OS®

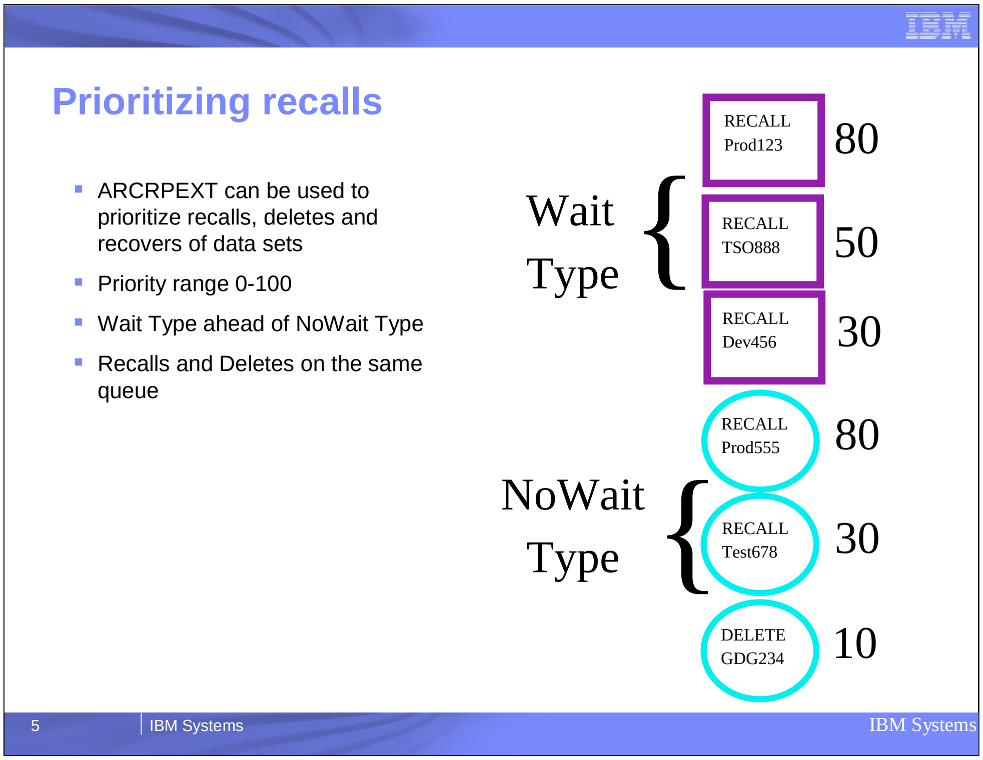


Agenda

- Managing and tuning the **recall** environment
  - ARCRPEXT
  - Common recall queue
  - Reducing recall contention
- Managing and tuning the **migration** environment
  - Fast subsequent migration
- Managing and tuning the **backup** environment
  - Backup direct to tape
  - Increased backup tasks
- Managing and tuning the **HSM control data sets** environment
  - CDS care and feeding
  - Backing up your control data sets
  - CDS recovery planning

This presentation is a loose collection of hints and tips collected through the years as well as highlights of certain functions that installations should consider taking advantage of.

To make it easier to follow, the presentation was broken down into the topics described in the agenda



Prioritizing recalls

- ARCRPEXT can be used to prioritize recalls, deletes and recovers of data sets
- Priority range 0-100
- Wait Type ahead of NoWait Type
- Recalls and Deletes on the same queue

Wait Type

RECALL Prod123 80

RECALL TSO888 50

RECALL Dev456 30

NoWait Type

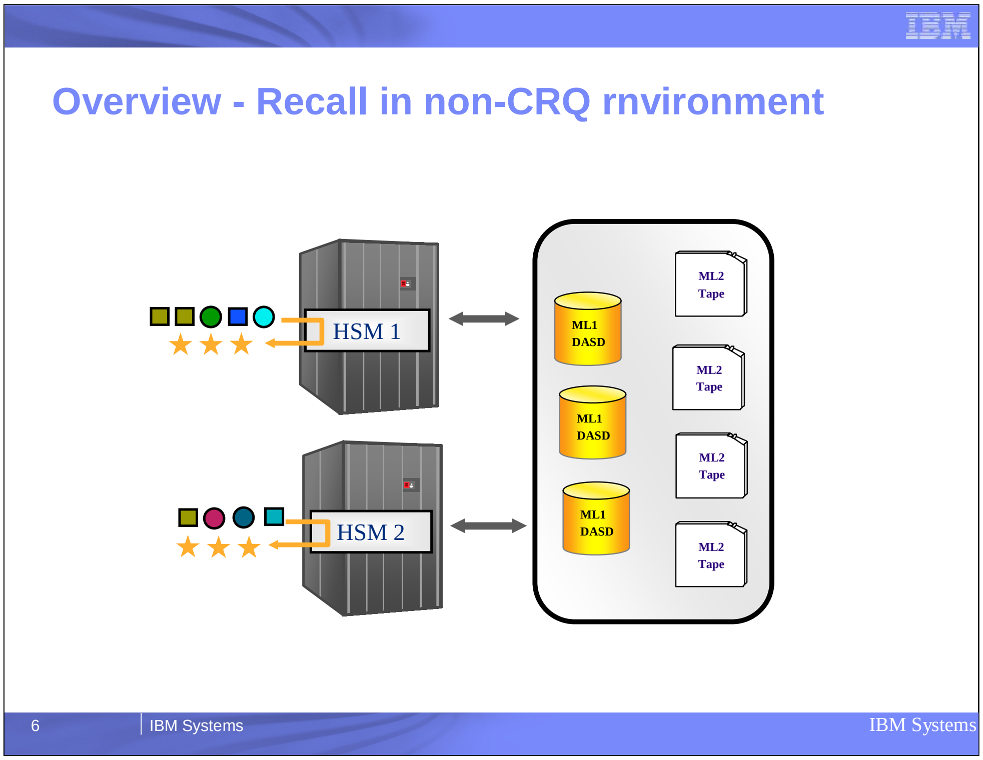
RECALL Prod555 80

RECALL Test678 30

DELETE GDG234 10

The ARCRPEXT, introduced via APAR OW07248 on Release 120, allows customers to assign relative priorities to each incoming recall, delete and recover request.

APAR OW29730 enhanced this support by allowing nonwait requests to also be prioritized. All wait type requests, however, are processed above nonwait type requests, regardless of assigned priority.
Priorities range from 1 (lowest) to 100 (highest) with 50 as default for any request not explicitly assigned by this exit.



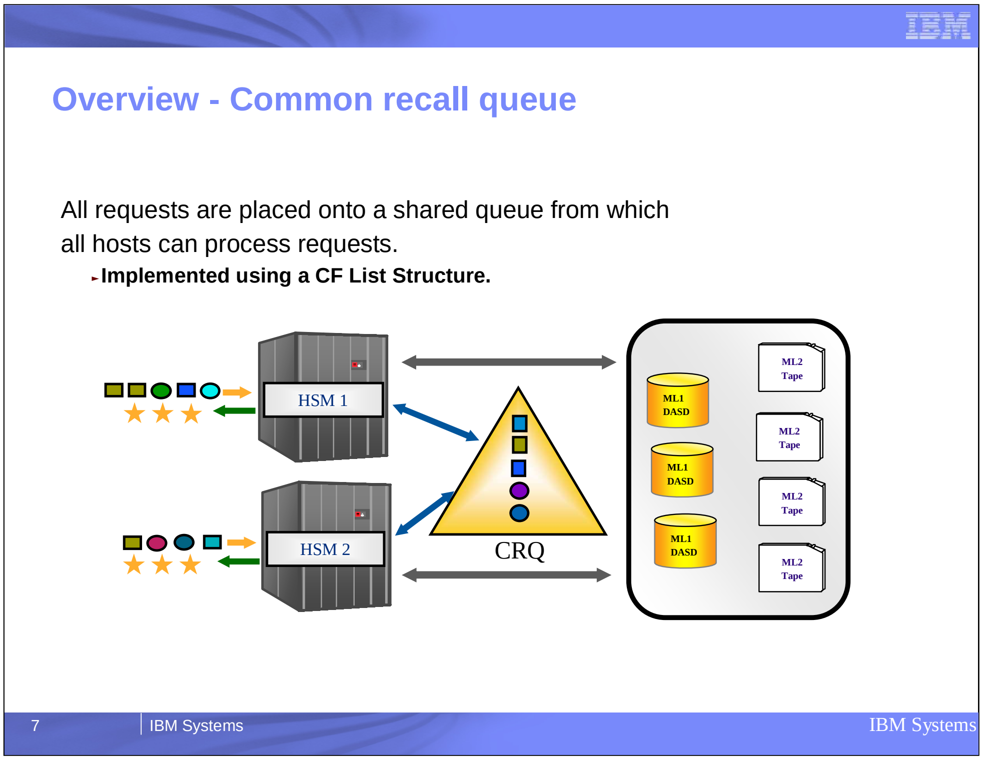
Overview - Recall in non-CRQ rnvironment

This slide shows how non-CRQ recall works. It highlights the fact the recall workload is not shared among the hosts even though tape and disk connection is shared.

Even if one host is overloaded and one lightly loaded there is no sharing

If one host mounts a tape for recall, the other host must wait for that tape if there are recall requests for data sets residing on the tape.

This can also create a situation called "recall tape takeaway" where a tape is taken away from one host because a higher priority recall request for a data set on that tape has been issued by another host.



Overview - Common recall queue

All requests are placed onto a shared queue from which all hosts can process requests.

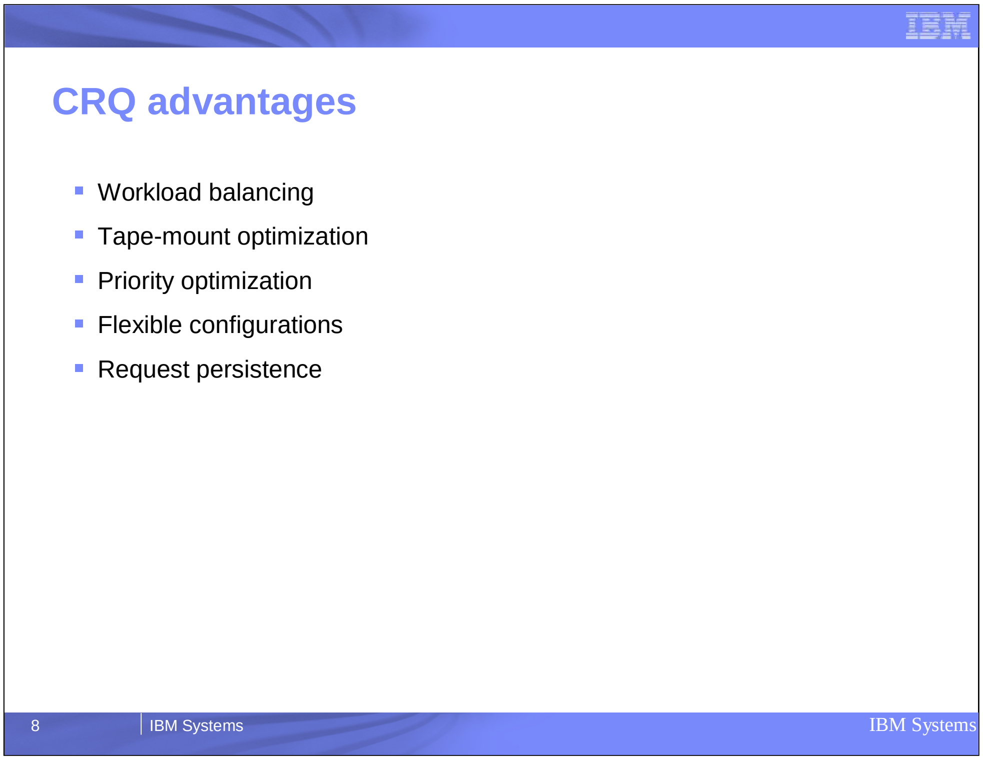
- Implemented using a CF List Structure.

CRQ is implemented using coupling facility list structures

These list structures can be shared among multiple hosts in an HSMPlex

Introduces concept of CRQplex
HSMs connected via a single CRQ
CRQplex cannot span HSMplex

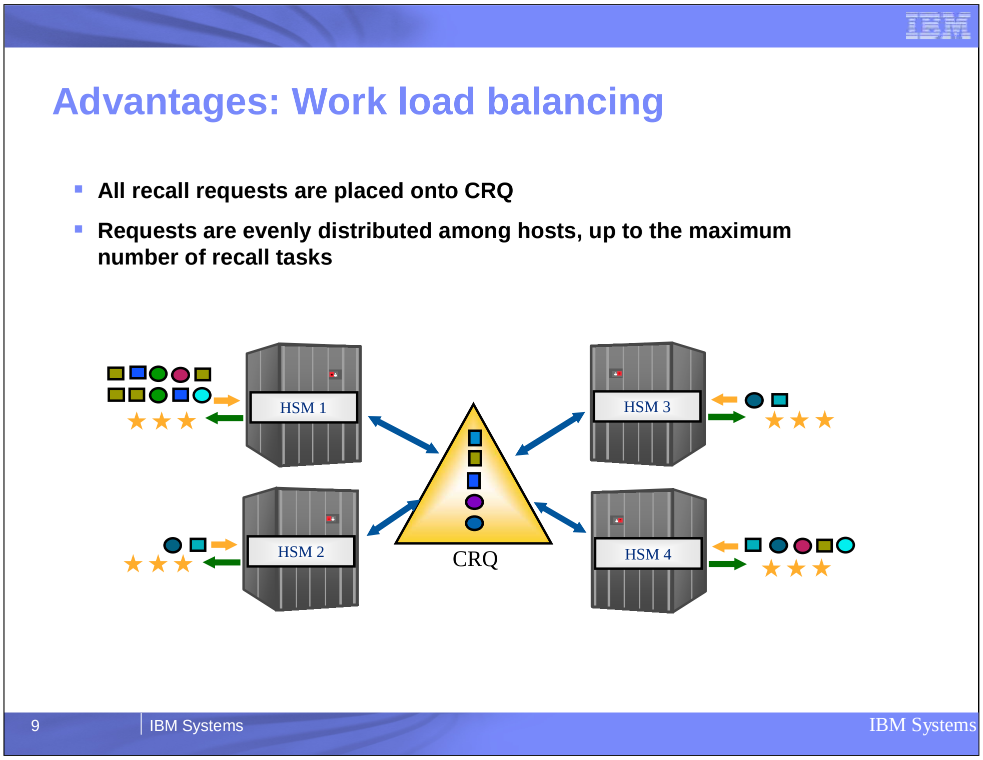
Can have CRQ and non-CRQ in a single HSMplex



CRQ advantages

- Workload balancing
- Tape-mount optimization
- Priority optimization
- Flexible configurations
- Request persistence

This slide lists the major advantages of CRQ. Each is highlighted in a follow-on slide



Advantages: Work load balancing

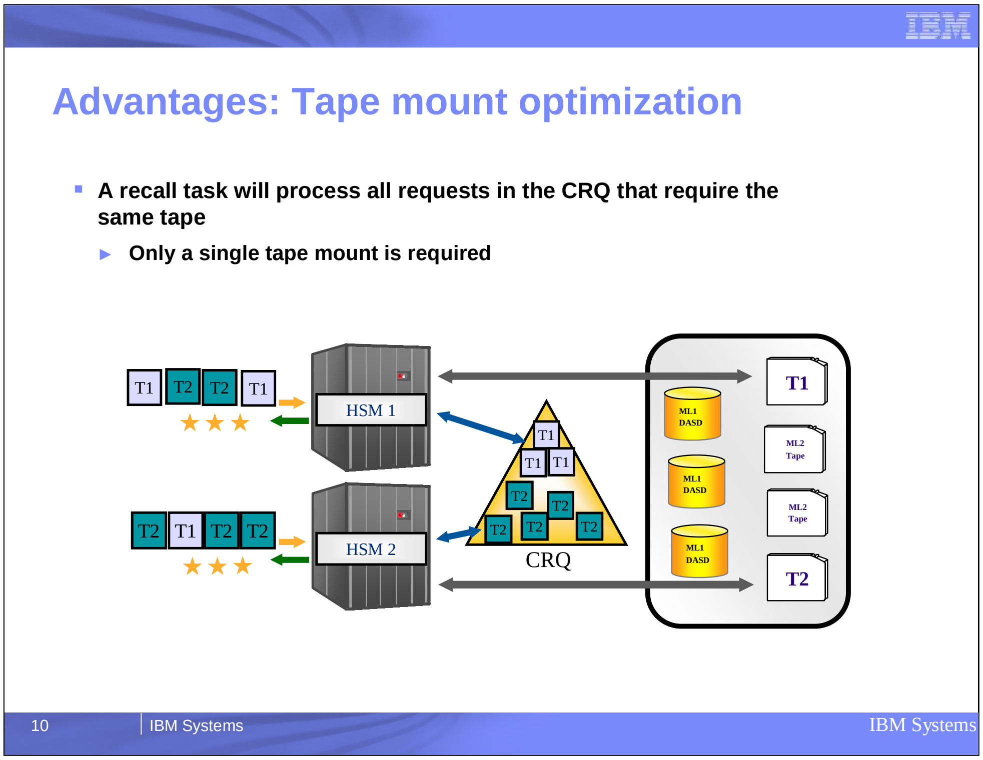
- All recall requests are placed onto CRQ
- Requests are evenly distributed among hosts, up to the maximum number of recall tasks

In a CRQplex the workload is balanced between hosts participating in the CRQ

Each host can set its own tasking level for recall. Up to 15 tasks per HSM

Recall requests will be shared up to the tasking levels on each host

HSMs will operate in a round-robin fashion selecting recall requests from the CRQ until all recall requests are exhausted or the particular instance of HSM has reached its recall task limit

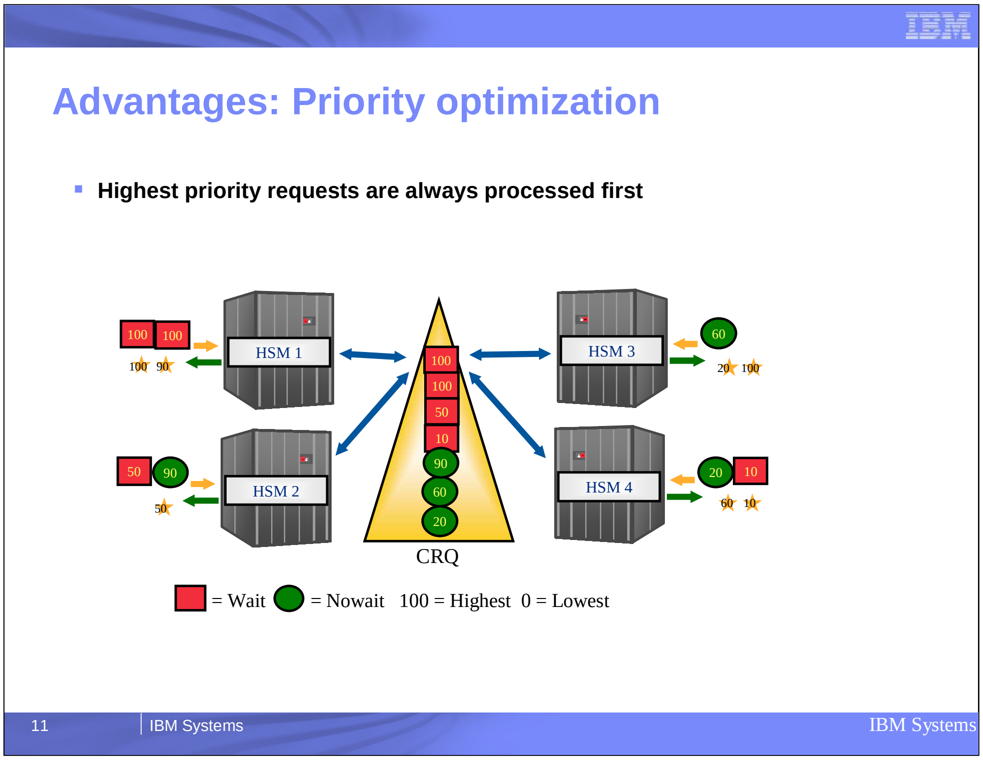


Advantages: Tape mount optimization

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

This slide shows that once a tape is mounted to a particular HSM host in the CRQplex, that HSM host will process all recall requests on the CRQ for that host.

Once all recall requests for data sets on that tape are exhausted, the tape will remain mounted for a period of time and if no more requests appear for that tape, then the tape will be demounted



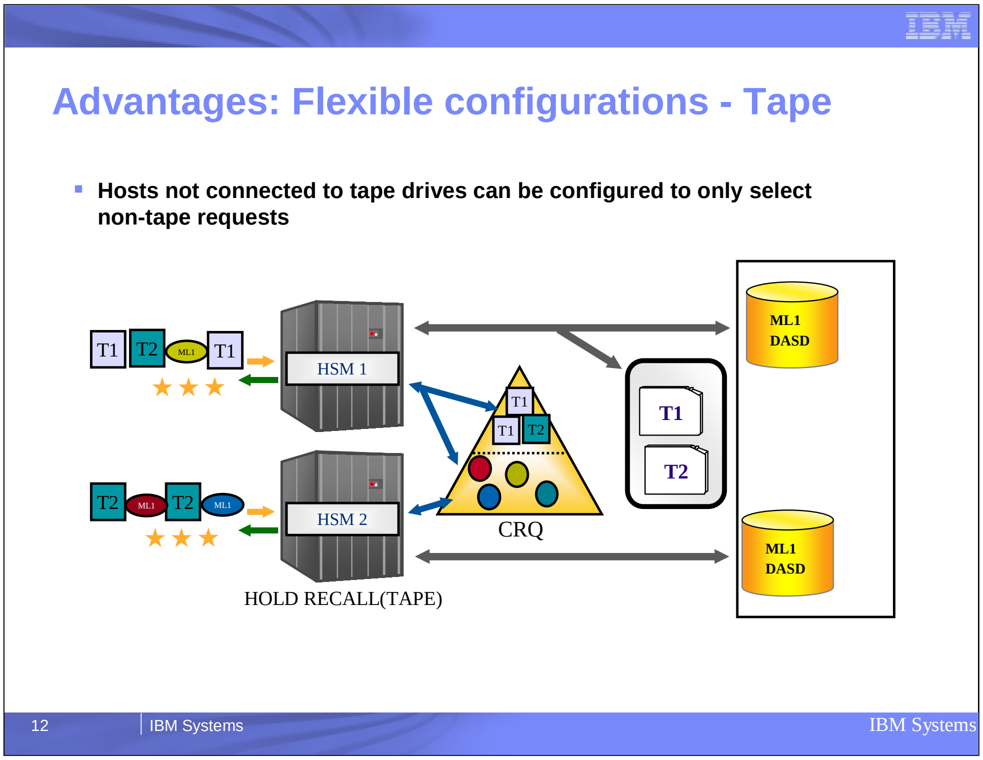
Advantages: Priority optimization

- **Highest priority requests are always processed first**

One advantage of CRQ is priority optimization

Highest priority requests are processed CRQplex wide because all hosts have access to the CRQ structures and will select the highest priority request available on the queue.

This is generally true except for recall requests from tape. In this case if a tape is mounted then the recall requests will be processing in FIFO order



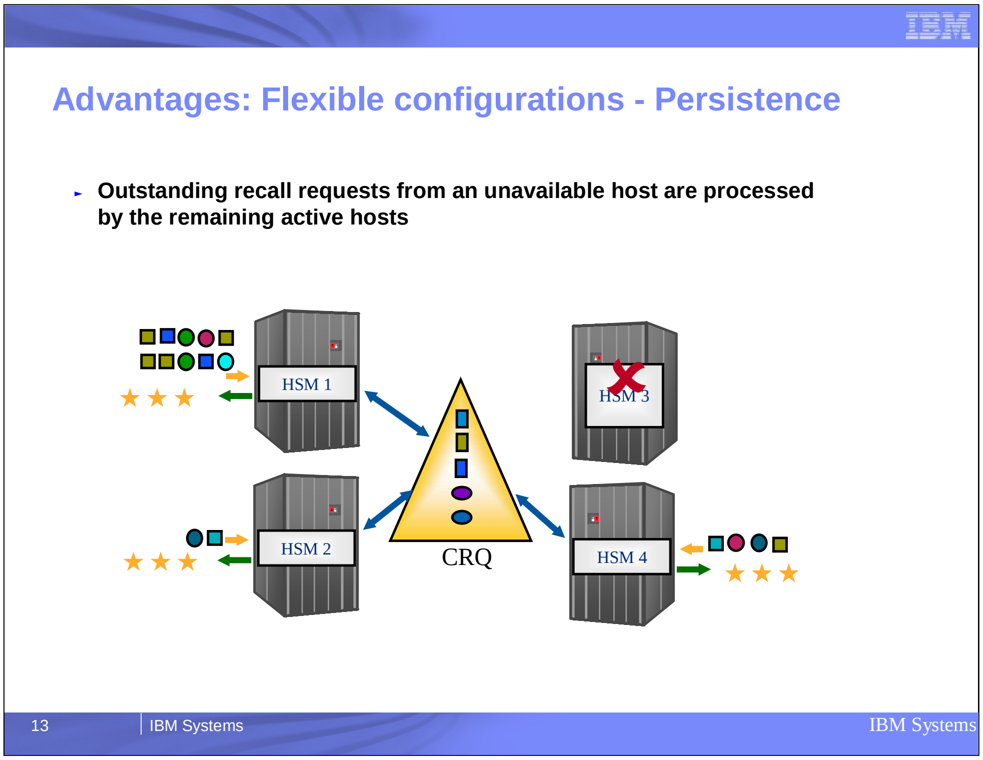
Advantages: Flexible configurations - Tape

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

You may have a case where one HSM in a CRQplex does not have access to drives.

In this case you can issue a HOLD RECALL(TAPE) to tell HSM not to select recall requests for data sets on tape.

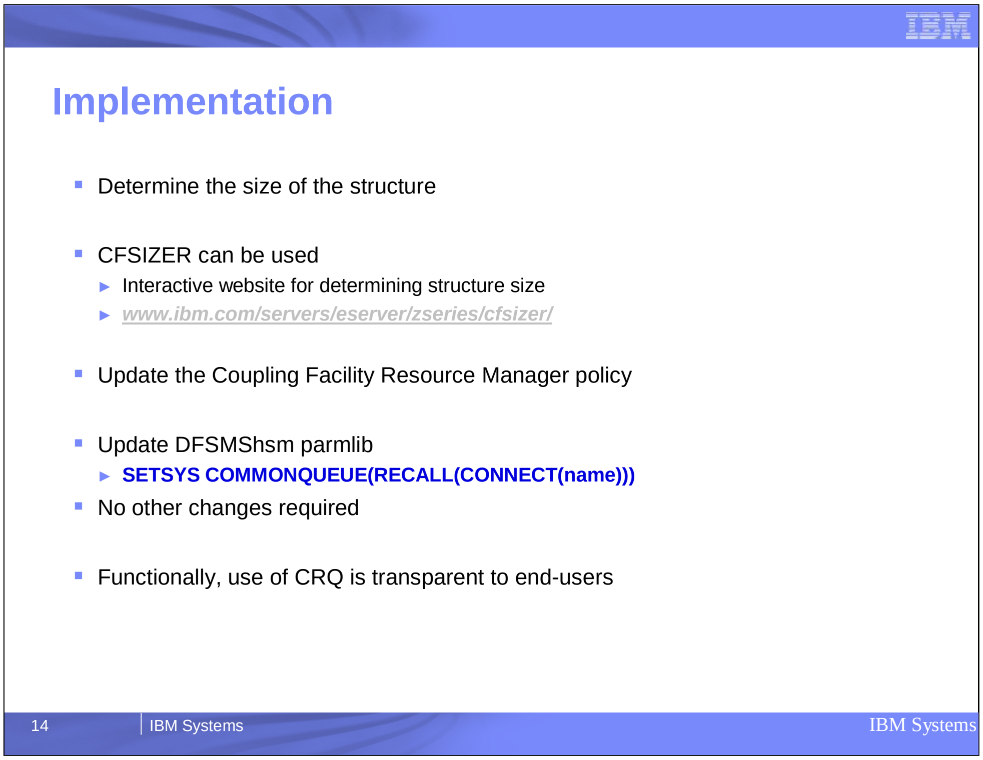
Tape recall requests on the HELD HSM can then be handled by other HSMs that are connected to the CRQ and tape drives



Advantages: Flexible configurations - Persistence

- Outstanding recall requests from an unavailable host are processed by the remaining active hosts

If for any reason an HSM host participating in a CRQplex becomes unavailable, recall requests originating from that host can be processed by the other hosts connected to the CRQ.



Implementation

- Determine the size of the structure
- CFSIZER can be used
  - Interactive website for determining structure size
  - *www.ibm.com/servers/eserver/zseries/cfsizer/*
- Update the Coupling Facility Resource Manager policy
- Update DFSMShsm parmlib
  - **SETSYS COMMONQUEUE(RECALL(CONNECT(name)))**
- No other changes required
- Functionally, use of CRQ is transparent to end-users

This slide depicts how easy it is to implement CRQ

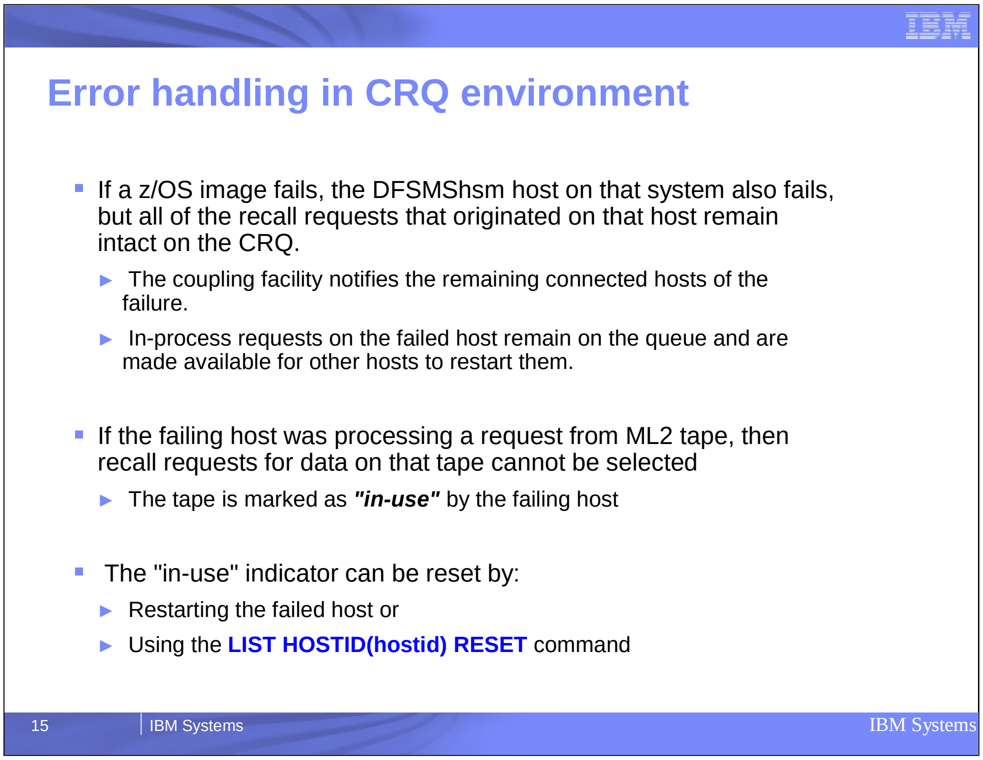
If an installation requires help determining the size of a structure they can look in the DFSMShsm I&C Guide or use an interactive tool called the CFSIZER

The size of the CRQ structure is dependent on the highest number of projected concurrent recall requests that are expected to be on the queue at any one time

Installations then need to define the CRQ structure to the coupling facility resource manager

Once the structure is defined to the CFRM, it can be defined to HSM via the SETSYS COMMONQUEUE command

End-users should see no external changes. Just better performance of their recall requests



Error handling in CRQ environment

- If a z/OS image fails, the DFSMShsm host on that system also fails, but all of the recall requests that originated on that host remain intact on the CRQ.
  - The coupling facility notifies the remaining connected hosts of the failure.
  - In-process requests on the failed host remain on the queue and are made available for other hosts to restart them.
- If the failing host was processing a request from ML2 tape, then recall requests for data on that tape cannot be selected
  - The tape is marked as ***"in-use"*** by the failing host
- The "in-use" indicator can be reset by:
  - Restarting the failed host or
  - Using the **LIST HOSTID(hostid) RESET** command

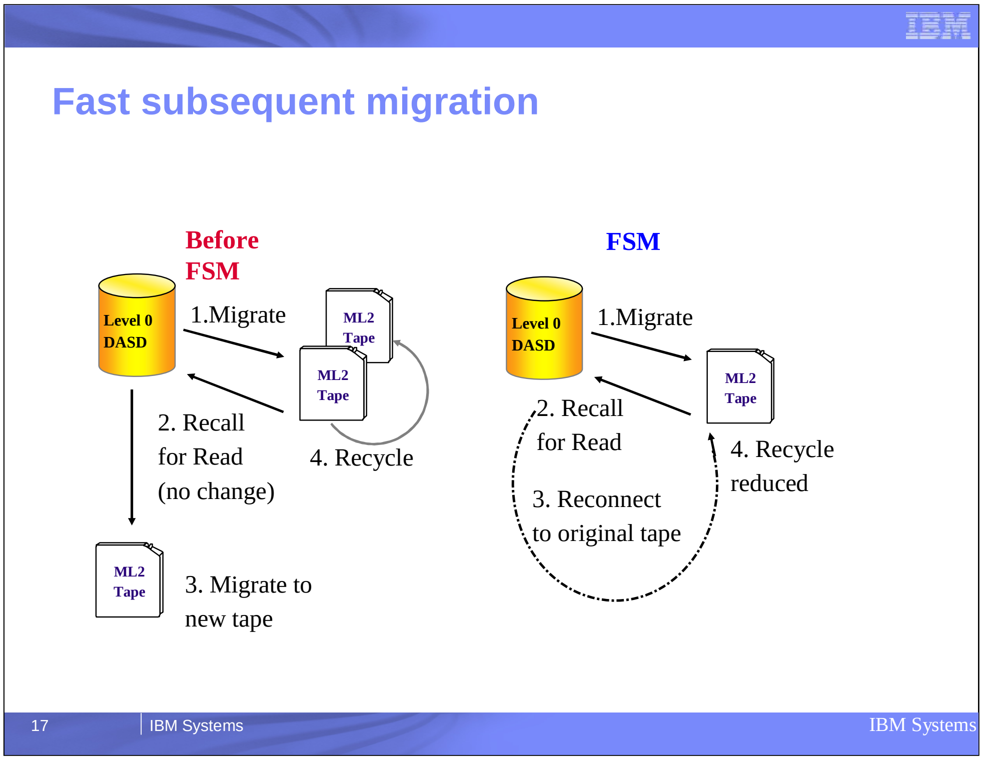
The slide explains a situation that can occur if an HSM host in the CRQplex is processing a tape recall request fails.

In this situation the tape is marked "in-use" by the failing host and other hosts are not able to mount this tape for new recall requests from that tape.

To alleviate this situation, installations should issue the LIST HOSTID(hostid) RESET command with the hostid of the failing host to reset the "in-use" indicator and allow other HSM hosts to be able to select this tape

Agenda: Managing and tuning the migration environment

- Managing and tuning the recall environment
- **Managing and tuning the migration environment**
- Managing and tuning the **HSM control data sets** environment



Fast subsequent migration

Before FSM

Level 0 DASD

1.Migrate

ML2 Tape

ML2 Tape

2. Recall for Read (no change)

4. Recycle

ML2 Tape

3. Migrate to new tape

FSM

Level 0 DASD

1.Migrate

ML2 Tape

2. Recall for Read

3. Reconnect to original tape

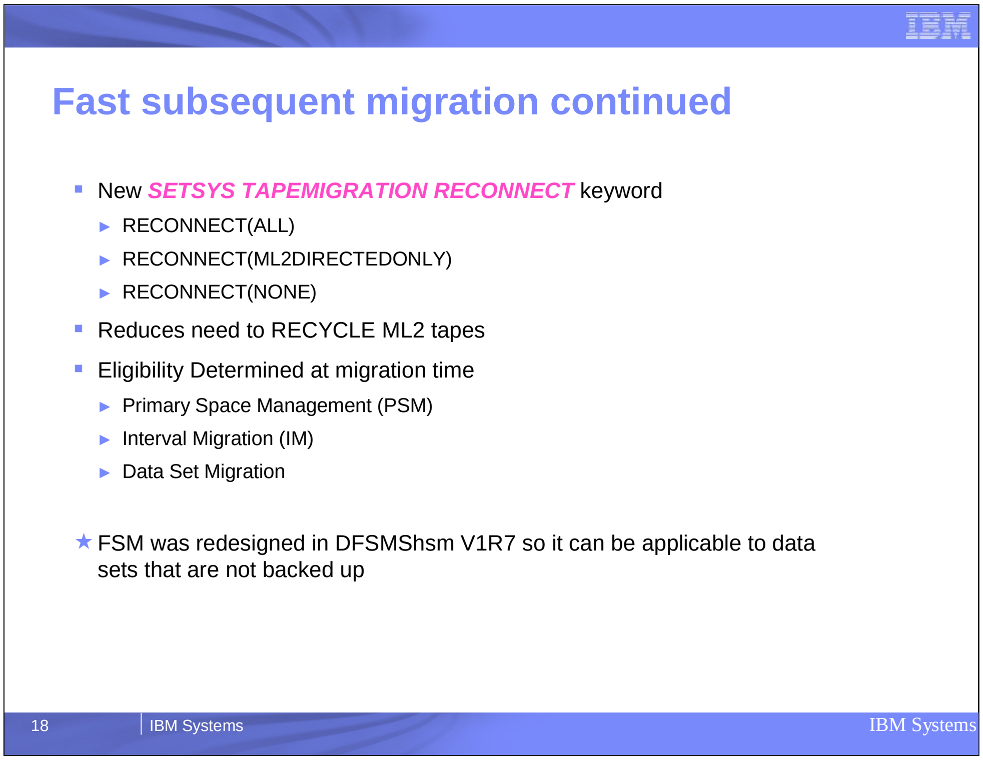
4. Recycle reduced

This slide shows HSM migration processing before and after the implementation of Fast Subsequent Migration (FSM).

FSM was implemented in DFSMS R10

Prior to FSM if a data set was migrated to ML2 tape, recalled for a browse request (data set not updated), when the data set became eligible for migration a new migration copy is created on ML1 disk or ML2 tape even though a valid copy still existed on the original ML2 tape

With FSM if the data set is recalled for browse only and a valid copy still exists on ML2 tape when that data set again is eligible for migration, that data set is reconnected to the tape from which it was recalled rather than a new migration copy being created on ML1 disk or ML2 tape.



Fast subsequent migration continued

- New ***SETSYS TAPEMIGRATION RECONNECT*** keyword
  - RECONNECT(ALL)
  - RECONNECT(ML2DIRECTEDONLY)
  - RECONNECT(NONE)
- Reduces need to RECYCLE ML2 tapes
- Eligibility Determined at migration time
  - Primary Space Management (PSM)
  - Interval Migration (IM)
  - Data Set Migration

★ FSM was redesigned in DFSMShsm V1R7 so it can be applicable to data sets that are not backed up

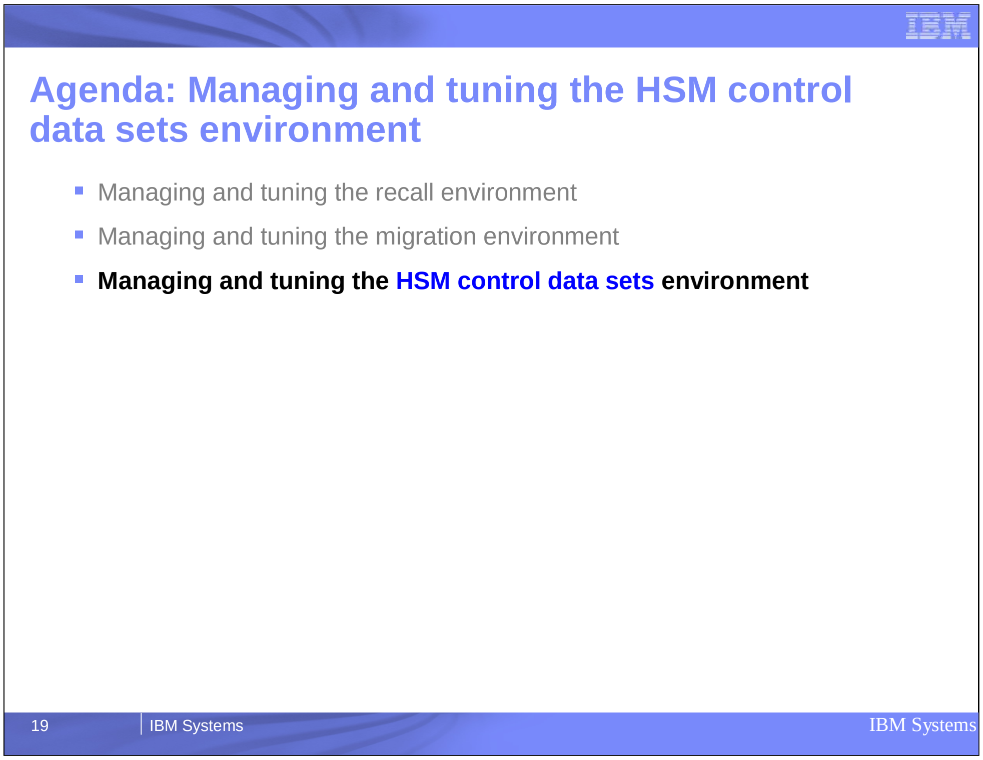
This slide shows the DFSMShsm commands needed to indicate whether or not FSM is to be used. This is controlled via keywords on the SETSYS TAPEMIGRATION RECONNECT command

ALL - indicates to reconnect when the data set becomes eligible for either ML1 or ML2 migration.

ML2DIRECTEDONLY - indicates to reconnect only when the data set becomes eligible to be migrated to ML2 tape, either by direct migration to ML2 tape or if the data set is first migrated to ML1 disk and then to ML2 tape

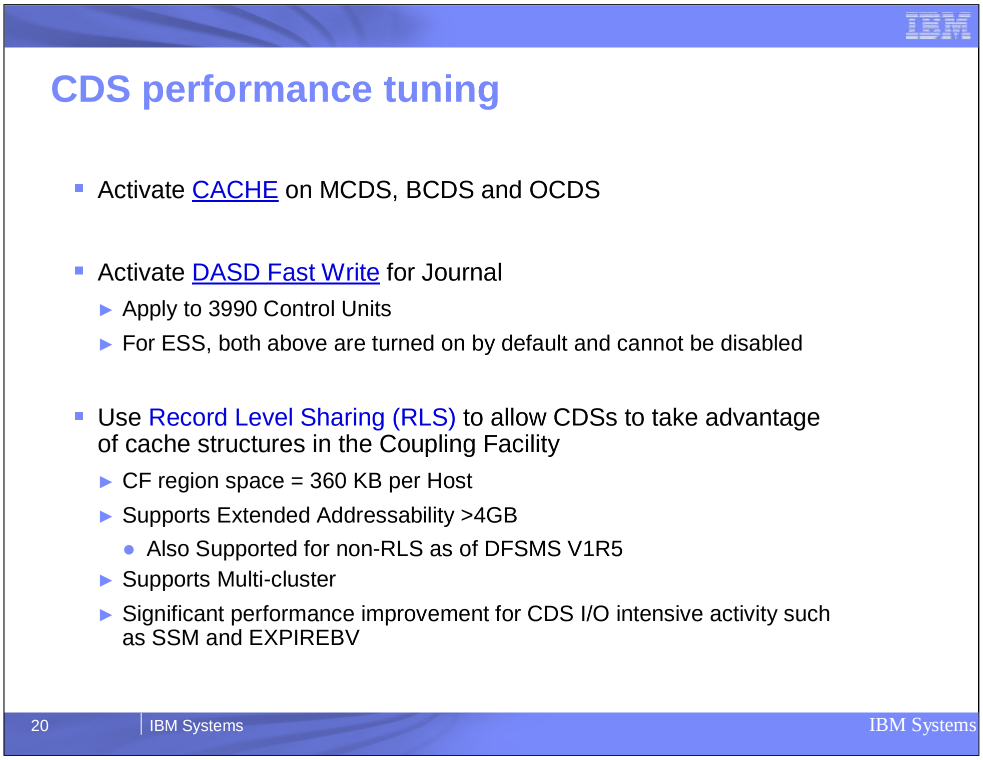
NONE - indicates to not use FSM and always create a new migration copy on ML1 disk or ML2 tape

Use of FSM can reduce the need for recycle since the data on the tape is being reused and is not marked invalid



Agenda: Managing and tuning the HSM control data sets environment

- Managing and tuning the recall environment
- Managing and tuning the migration environment
- **Managing and tuning the HSM control data sets environment**



CDS performance tuning

- Activate CACHE on MCDS, BCDS and OCDS
- Activate DASD Fast Write for Journal
  - Apply to 3990 Control Units
  - For ESS, both above are turned on by default and cannot be disabled
- Use Record Level Sharing (RLS) to allow CDSs to take advantage of cache structures in the Coupling Facility
  - CF region space = 360 KB per Host
  - Supports Extended Addressability >4GB
    - Also Supported for non-RLS as of DFSMS V1R5
  - Supports Multi-cluster
  - Significant performance improvement for CDS I/O intensive activity such as SSM and EXPIREBV

In order to get the best CDS I/O performance on a particular disk control unit, installations should turn on CACHE and DASD fast write

With newer disk subsystems CACHE and DASD Fast Write are always enabled

Installations should strongly consider the use of VSAM Record Level Sharing or RLS.
RLS allows sharing of a VSAM KSDS at the record level rather than at the data set level, thus reducing contention for the data set by DFSMShsm functions and address spaces

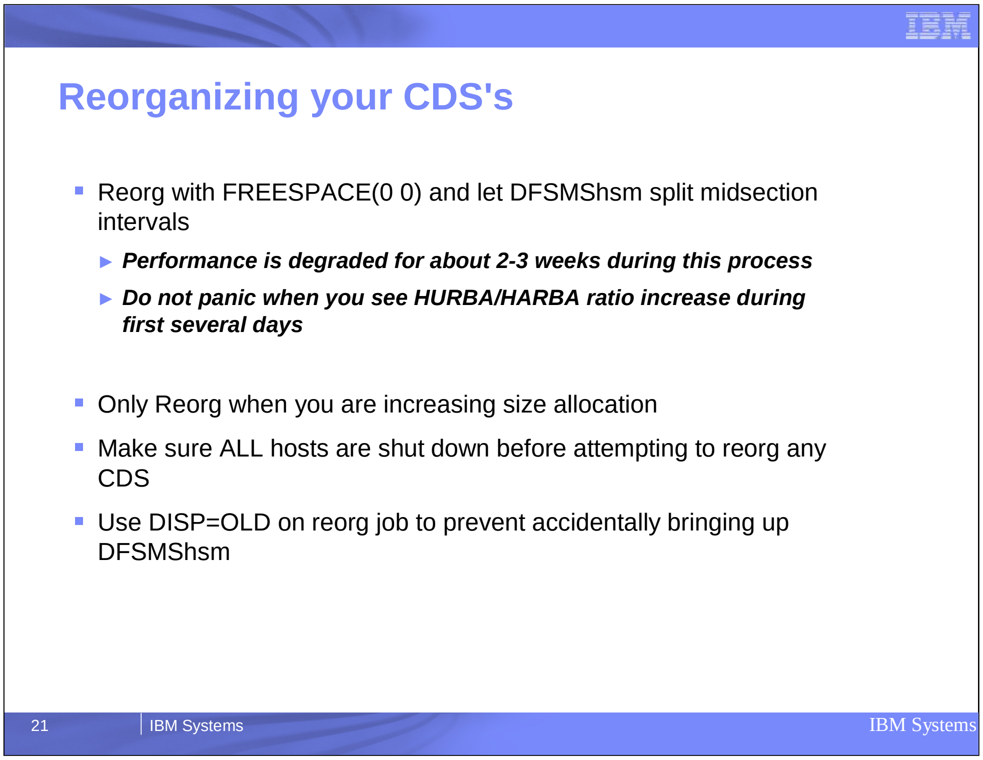
RLS has shown to provide significant performance improvement for any CDS activity that is heavily I/O intensive such as Secondary Space Management and EXPIREBV processing

RLS requires cache structures in the Coupling Facility and requires approximately 360 KB per host sharing the CDSs

RLS also supports Extended Addressability

Extended Addressability for the CDSs was extended to non-RLS in DFSMShsm V1R5

EA allows each cluster to be larger than 4GB



Reorganizing your CDS's

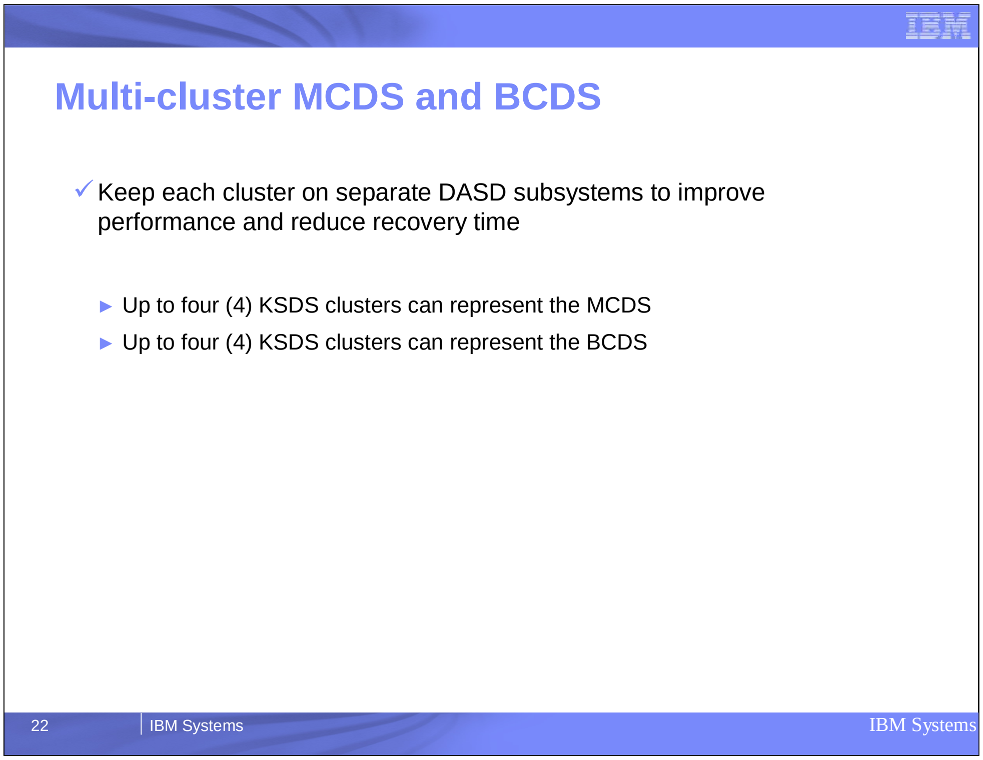
- Reorg with FREESPACE(0 0) and let DFSMShsm split midsection intervals
  - ***Performance is degraded for about 2-3 weeks during this process***
  - ***Do not panic when you see HURBA/HARBA ratio increase during first several days***
- Only Reorg when you are increasing size allocation
- Make sure ALL hosts are shut down before attempting to reorg any CDS
- Use DISP=OLD on reorg job to prevent accidentally bringing up DFSMShsm

Many customers reorg their CDSs on a regular basis thinking that by doing so, they make the most efficient use of their CDSs

IBM recommends that installations only reorg their CDSs when they are increasing the size of a CDS or changing attributes of a CDS

Studies have shown that when CDSs are REORGed, VSAM performs a large number of CI/CA splits to create space required for inserts. It can take as many as 2 weeks for the CDSs to reach a steady state after a REORG.

IBM also recommends that installations define their CDSs with FREESPACE(0) so VSAM can figure out where space is needed and perform the necessary CI/CA splits when doing inserts
. This will create space where needed and eliminate the need for the extra space that would be required if the CDSs were defined with FREESPACE(50 50)
To make sure an instance of DFSMShsm is not started while a REORG of the CDSs is occurring, installations should specify DISP=OLD in their REORG jobs
One of the most common ways customers break a CDS is by bringing up DFSMShsm on one system while reorging a CDS on another system

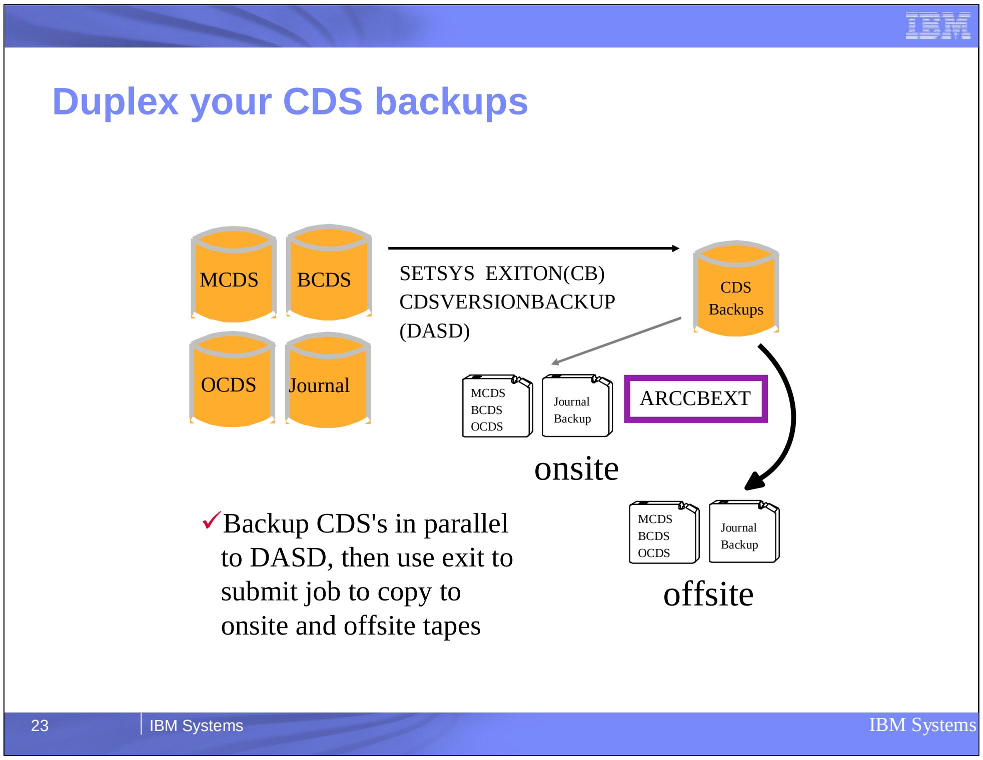


Multi-cluster MCDS and BCDS

- ✓ Keep each cluster on separate DASD subsystems to improve performance and reduce recovery time
  - ► Up to four (4) KSDS clusters can represent the MCDS
  - ► Up to four (4) KSDS clusters can represent the BCDS

An individual MCDS or BCDS can be defined to up to 4 different clusters. Splitting the MCDS and BCDS in 4 different clusters allows for more concurrent I/O.

Also having 4 smaller clusters rather than 1 monolithic cluster allows the clusters to be backed up in parallel to disk, thus reducing the backup time.



Duplex your CDS backups

MCDS

BCDS

OCDS

Journal

SETSYS EXITON(CB)
CDSVERSIONBACKUP
(DASD)

CDS Backups

MCDS
BCDS
OCDS

Journal Backup

ARCCBEXT

onsite

MCDS
BCDS
OCDS

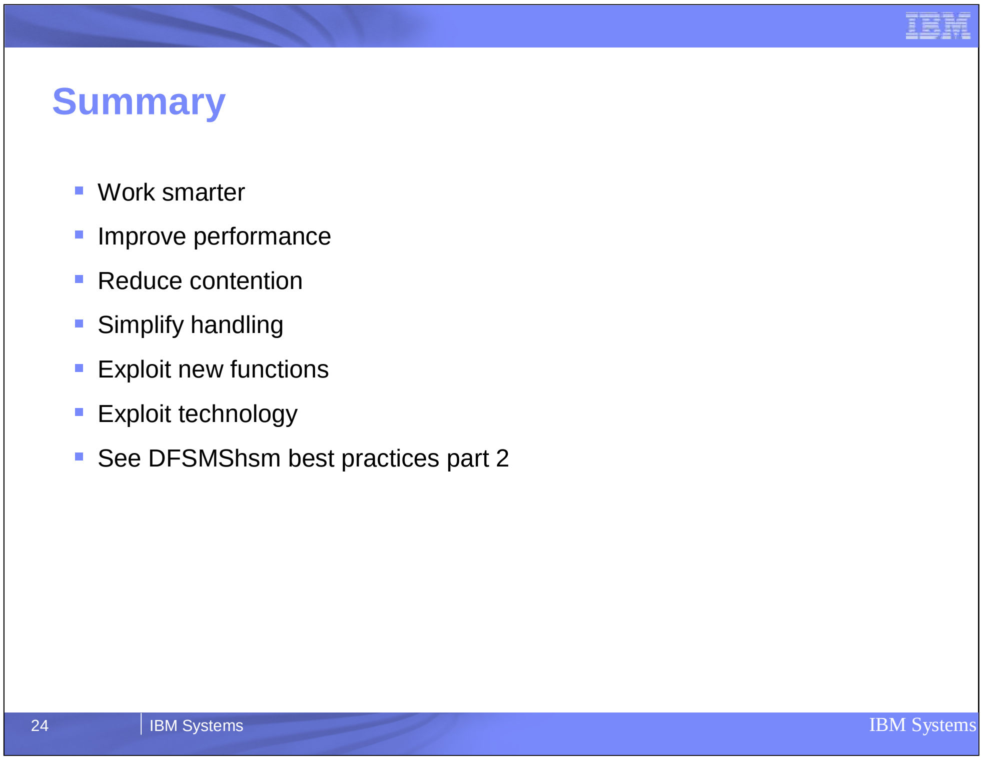
Journal Backup

offsite

✓Backup CDS's in parallel to DASD, then use exit to submit job to copy to onsite and offsite tapes

Because CDSs backup copies are so vital to ensuring DFSMShsm data availability it is suggested that installations keep multiple backup copies of each version of their CDS backup.

The fastest way to accomplish this is to first backup the CDS to disk and then use the ARCCBEXT to schedule a DSS dump job to dump multiple copies of the disk backup to tape



Summary

- Work smarter
- Improve performance
- Reduce contention
- Simplify handling
- Exploit new functions
- Exploit technology
- See DFSMShsm best practices part 2

This presentation covered a number of hints and tips to help you get the most out of your DFSMShsm environment

Please take time to read some of the DFSMShsm publications such as the DFSMShsm Storage Admin Ref and Storage Admin Guide so that you can get more detailed information on how to use DFSMShsm

If you are new to DFSMShsm , the DFSMShsm Primer Redbook (SG24-5272-01) provides an excellent overview of the functionality of DFSMShsm