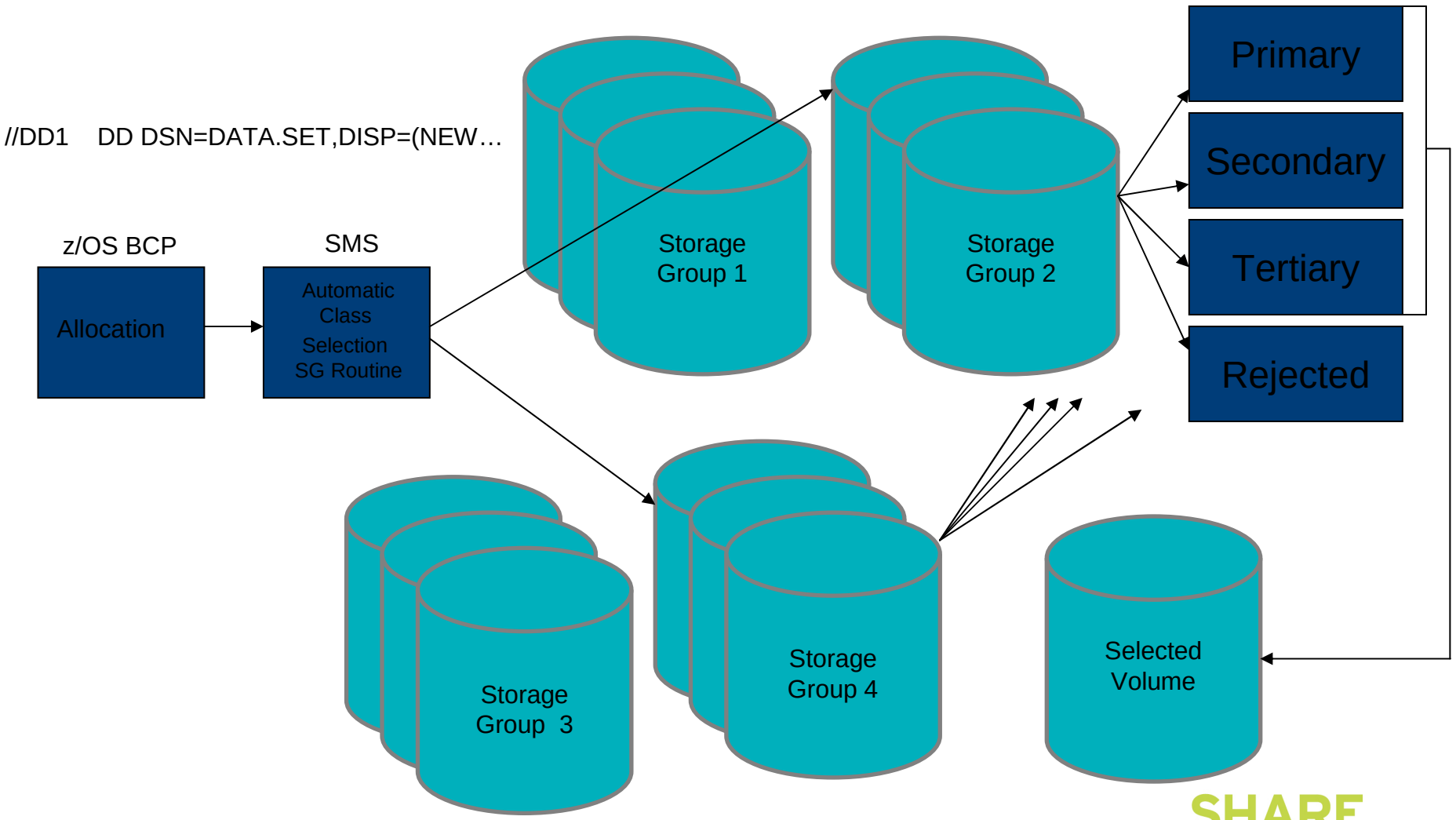
DFSMS Basics: How SMS Volume Selection Works

Steve Huber
IBM Corporation

March 3, 2011
Session 9012

Volume Classification



The Primary List

- Meet data set separation requirement
- SMS storage group and volume statuses are enabled
- MVS status is online
- IART requirement is met
- Number of volumes in storage group >= volume count
- Accessibility requested can be met
- Availability requested can be met
- Meets the guaranteed space requirement
- Can perform the allocation & stay below high threshold
- For MSR=999, volume is non-cached
- Data class extended format request can be met

The Secondary List

- ABEND X37 prevention - the most available space
- Meet data set separation requirement
- Meet volume count requirement
- Can perform the allocation without going more than 20% over high threshold
- SMS storage group and volume status
- Honors tiering of storage groups
- Spill/Overflow volumes
- Volume characteristics
  - Availability
  - Accessibility
  - Extended format
  - Guaranteed space
- Mount time performance

The Tertiary List

- Only used for:
  - Non-guaranteed space requests
  - Non-VSAM data sets
- Consists of volumes in storage groups that do not meet the volume count requested

Conventional Volume Selection

- Used for all non-striped data sets
- Used for all data sets with zero or blank SDR
- Uses a preference sequence to sort volumes in the candidate storage groups into:
  - Primary
  - Secondary
  - Tertiary
  - Rejected

Volume Selection Evaluation Process

Criteria	Preferences
PCU Separation	Volume not on same PCU as data set from which it is separated.
Extent Pool Separation	Volume not in same extent pool as data set from which it is separated.
Volume Separation	Volume does not contain a data set from which this data set should be separated.
Volume Count	Volume is in a storage group that can satisfy the volume count.
Primary Threshold	Volume has sufficient space in target addressing space without exceeding high threshold
Secondary Threshold	Volume has sufficient space without exceeding high threshold

Volume Selection Evaluation Process

Criteria	Preferences
SMS Status	Volume and its storage group are both enabled.
Multi-tiered Storage Group	Volume resides in a storage group that is elected in order of specification
EOV Extend	For EOV Extend, volume does not reside in extend storage group.
Non-Overflow	Volume resides in a non-overflow storage group.
IART	Volume is mountable and IART specified is non-zero.
Fast Replication	Volume is eligible for fast replication request
Extended Attribute	Volume has extended attributes (EAV)

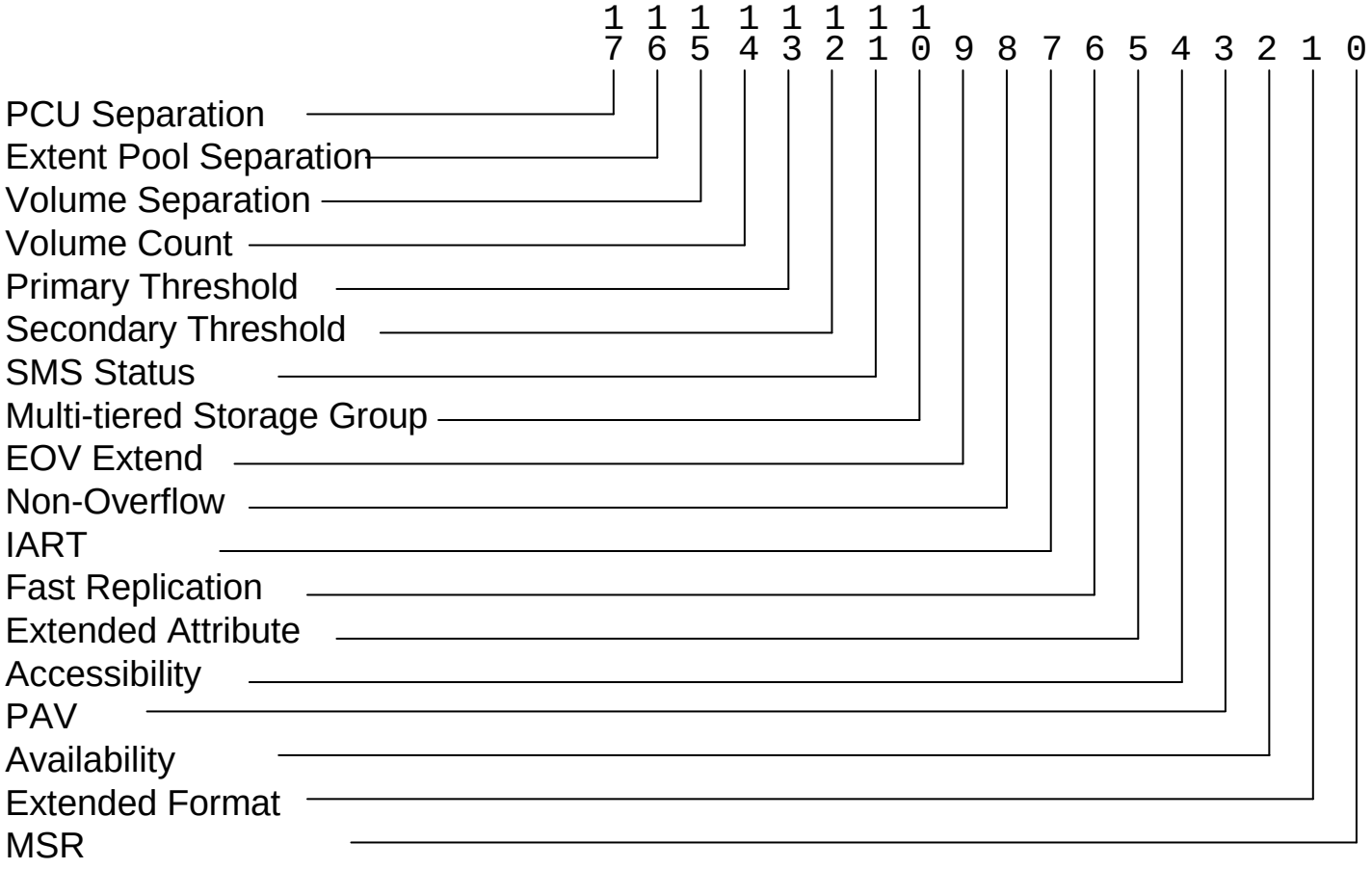
Volume Selection Evaluation Process

Criteria	Preferences
Accessibility	Controller for volume supports accessibility & value is PREF or controller does not support it & value is STANDARD
Parallel Access Volume	Volume supports PAV
Availability	Controller for volume supports availability & value is PREF or controller does not support it & value is STANDARD.

Volume Selection Evaluation Process

Criteria	Preferences
Extended Format	Volume is on a control unit that supports extended format and IF EXT is PREF.
Millisecond Response (MSR)	Volume provides the requested response time specified in direct/sequential MSR or Volume provides a faster response time than requested in the direct or sequential MSR.

Volume Selection Evaluation Process



Example

	Preference	Pos
Volume A	Volume and storage group enabled	11
	Non-overflow	8
	Volume PCU supports accessibility & PREF	4

Total Preference Value = 2320

Volume B	**Volume has sufficient space**	**12**
	Volume and storage group enabled	**11**

Total Preference Value = 6144

Data Set Separation

- Allows you to designate groups of data sets which are to be physically separated by PCU or volume
- SMS attempts to allocate the data sets behind different control units or volumes
- A data set separation profile must be provided
  - Indicates separation by PCU or volume
- The name of the data set containing the profile must be specified in the SMS base configuration
- Cannot be used with non-SMS-managed data sets or with full volume copy utilities such as PPRC

Recommended Use of Separation

Use only for a small set of mission-critical data sets

- Volume rejection because of separation may drastically reduce the number of eligible volumes
- Data set separation can affect system performance
  - The number of data sets should be small as SMS must scan them all
  - The use of wildcards characters in separation data set names
  - The number of eligible volumes
- Take care when using separation with striping
- May require constant updating if used with GDSs

Specifying Separation

SEPARATIONGROUP | SEP (PCU | {VOLUME | VOL})
TYPE ({REQUIRED | REQ | R} | {PERFERRED | PREF | P})
DSNLIST | DSNS | DSN (data set name[, data set name,...]);

Examples:

```
SEPARATIONGROUP(VOL) TYPE(PREF)-
      DSNLIST(SMS.PROD.SCDS, SMS.PROD.ACDS, SMS.PROD.COMMDS);

SEP(PCU) TYPE(REQ)-
  DSNS(SYS1.JESCKPT1,-          /*PRIMARY*/
       SYS1.JESCKPT2,-          /*SECONDARY*/
       SYS1.JESCKPT3);          /*TERTIARY*/

SEPARATIONGROUP(VOL) TYPE(PREF)-
      DSNLIST (DB2.DATA.LIB.V%%%%.**)
```

Multiple Separation Profiles

- You can create multiple separation profiles in different data sets or PDS members
- You can only specify one separation profile in the configuration base
- If you have multiple configurations, they can all share the same profile
- Or they can all have separate profiles
- Profile is read when SMS initializes or restarts and whenever a new configuration is activated

When Separation Does Not Work

- The allocation is not SMS-managed.
- The separation profile cannot be accessed.
- The separation profile is invalid.
- The allocation uses a temporary data set name.
- Two data sets are allocated on different systems.
- A volume is varied online during allocation.
- An IODF change occurs during allocation.
- A data set name not in the profile is specified during HSM recover.
- The profile was modified after configuration activation.
- SMS does not perform separation during:
  - Rename.
  - HSM migration to level 1 or 2.
  - Full volume image copy.

MSR and Bias

- MSR (Millisecond Response)
  - Uses only the stored MSR; cached if cache is active, native otherwise
  - Devices close to the requested MSR are placed on the primary list
  - Devices not close to the requested MSR are placed on the secondary list
- Bias
  - Determines which volumes MSR performance numbers (read, write, or both) to consider during volume selection.
- If you leave all MSR and bias fields blank (direct and sequential), SMS ignores device performance during volume selection

Sustained Data Rate (SDR)

- SDR > zero
  - Causes striping volume selection to be used
  - May cause MSR, availability, accessibility, and free space criteria to be ignored
  - Considers controllers over volume attributes

Initial Access Response Time (IART)

- Object access:
  - Specifies the desired response time (in seconds) required for locating, mounting, and preparing a piece of media for data transfer. OAM uses this value to interpret the storage level, that is, to place an object at an appropriate level in the object storage hierarchy. For objects, both the IART and the sustained data rate (SDR) are applicable.
    - OAM uses IART as follows:
      - If the IART value is 0, OAM writes to DASD.
      - If the IART value is 1-9999, OAM selects removable media, either optical or tape.
- DASD Data set access:
  - SMS allows the system resources manager (SRM) to select a DASD volume from the primary volume list if the IART value is 0 or unspecified. SRM volume selection is ideally suited for batch jobs.
- VTS (Virtual Tape Server) cache management:
  - An initial access response time (IART) of 100 or greater means the volume has least preference in the cache. 0-99 means the volume has most preference.

Availability

- Specifies whether data set access should continue in the event of a single device failure.
  - CONTINUOUS
    - Specify an availability of CONTINUOUS if you do not want a device failure to affect processing. Only duplexed and RAID volumes are eligible for this setting. If CONTINUOUS availability is specified, data is placed on a device that can guarantee that it can still access the data in the event of a single device failure. This option can be met by
      - *A dual copy volume*
      - *An array DASD*
  - PREFERRED
    - Array DASD volumes are preferred over nonduplexed volumes.
    - Dual Copy volumes are not candidates for selection.
  - STANDARD
    - If data sets do not require such a high level of availability, specify STANDARD availability, which represents normal storage needs. Specify an availability of STANDARD to cause processing of a data set to stop after a device failure. Simplex volumes are preferred over array DASD. SMS selects only volumes that are not dual copy. This attribute does not apply to objects. Array DASD are acceptable candidates for both STANDARD and CONTINUOUS availability requests.
  - NOPREF
    - Simplex and array DASD are equally considered for volume selection. NOPREF is the default. Dual copy volumes are not candidates for selection.

Accessibility

- Defines the function of the hardware supporting the point-in-time copy using
  - Using either concurrent copy
  - Virtual concurrent copy
  - FlashCopy
- Options are;
  - CONTINUOUS (C)
    - Only point-in-time copy volumes are selected.
  - CONTINUOUS PREFERRED (P)
    - Point-in-time copy volumes are preferred over non-point-in-time copy volumes.
  - STANDARD (S)
    - Non-point-in-time copy volumes are preferred over point-in-time copy volumes.
  - NOPREF (N)
    - Point-in-time copy capability is ignored during volume selection (default)

Multi-Tiered Storage Groups

- Specify Multi-Tiered SG Y in the storage class
- Example:
  - SET &STORGRP = 'SG1', 'SG2', 'SG3'
- Result:
  - SMS selects volumes from SG1 before SG2 or SG3
  - If all enabled volumes in SG1 are over threshold, then SMS selects from SG2
  - If all enabled volumes in SG2 are over threshold, then SMS selects from SG3
  - If all volumes are over threshold, then SMS selects from the quiesced volumes in the same order

Parallel Access Volumes

- Feature of the Enterprise Storage Server (ESS)
- Available only when the PAV option is enabled
- Use the Parallel Access Volume Storage Class attribute:
  - Required: Only volumes with the PAV feature are selected
  - Preferred: Only volumes with the PAV feature are primary
  - Standard: Only volumes without the PAV feature are primary
  - Nopreference: All volumes, PAV and non-PAV are treated equally

Striping Volume Selection

- Used only for:
  - Initial allocation of extended format preferred or required data sets with SDR > 0
  - Recall/Recover of multi-stripe data sets
- Not used for Recall/Recover of single-striped multivolume data sets
- Similar to conventional volume selection
  - Eligible volumes classified as primary and secondary
    - For each controller, primary meets all requested preferences and is selected randomly
    - Secondary is all other volumes on the same controller
  - Assigned a volume preference weight based on preference table

Striping Volume Selection (cont)

- SMS calculates an average preference based on all volumes in an SG
- SMS selects the SG that has enough primary volumes to meet the stripe count and the highest average weight
- If no SGs meet this criteria, the one with the largest amount of primary volumes is selected
- If the largest amount of primary volumes is the same for multiple SGs, the SG with the highest weight is chosen
- If there are multiple SGs that meet the criteria, one is chosen at random

Striping Volume Selection (cont)

- No SGs with mixed device types
- Number of volumes computed from SDR
- Temporary data sets with volume count > 1 treated as non-striped
- Volume must be able to satisfy primary space requested

Striped Data Sets (general information)

- Maximum stripes for non-VSAM = 59
- Maximum stripes for VSAM= 16
- Maximum extents/space alloc = 5
- Non-VSAM max extents/stripe = 123
- For VSAM max extents/stripe:
  - Per volume = 123
  - Per stripe = 255
  - Per data component = 4080
  - VSAM extent constraint removal in data class, if set to Y, 59x123
  - VSAM stripes can extend to new volumes
- Minimum alloc = 1 track/stripe

Extents Per Volume

- Non-VSAM, non-extended format: Up to 16 on the volume
- Non-VSAM, extended format: Up to 123
- PDSE and HFS: Up to 123 on the volume
- VSAM: Up to 255 per component, but only 123 per volume per component
- Striped VSAM: Up to 4080 per data component
- VSAM extent constraint removal in data class, if set to Y, 59x123 or 7257

Extending Striped Data Sets

- All stripes must be able to satisfy secondary space/number of stripes
- Secondary space is divided by number of stripes and rounded up for each volume
- Non-VSAM striped data sets **cannot** extend to additional volumes
- VSAM striped data sets **can** extend to additional volumes
- Volume fragmentation may result in striping volume reselection

Requirements for Striping

- Volumes behind one of the following controllers:
  - ESCON-attached controller that supports concurrent copy
  - 3990-6 controllers
  - 3990-3 controllers with Extended Platform that are ESCON-attached
  - 3990-3 controllers with RAMAC support-level microcode
  - 9394 controllers
  - 9343 controllers with cache
  - IBM RAMAC Virtual Array
  - IBM Enterprise Storage Server
- Volumes must be ENABLED or QUIESCED and varied ONLINE

Why Isn't My Volume Primary?

- Volume was rejected
- Allocation would exceed high threshold
- Volume/SG quiesced
- VTOC IX disabled
- MSR not met
- Non-zero IART
- Controller IML'd while online to MVS
- Accessibility value
- Availability value
- Extended format value
- Insufficient number of volumes in SG

Why Was My Volume Rejected?

- Volume not in selected SG or SGs
- Not online to MVS
- Bad volume of SG status
- Insufficient free space
- Insufficient space in VTOC or IX
- Not init'd as SMS volume
- On exclude list
- Does not support extended format
- Not on include list

Why Was My Volume Rejected (cont)?

- Volume does not meet …
  - Availability
  - Accessibility
  - IART
- UCB type unusable
- Allocation attempted, but failed
- Too fragmented
- SG has insufficient volumes
- Geometry incorrect
- SG contains mixed SDRs
- Does not support Flashcopy

Wrong Volume Selected?

- Check construct assignments
- Check channel path utilization
- Check storage group/volume utilization
- No volumes in primary volume list
- Expected volume was on tertiary list
- Expected volume was rejected
- Products which hook into system code (such as SRM) can create unexpected results

If All Else Fails....

Data class contains two values which can be used to influence volume selection:

- Space Constraint Relief
- Reduce Space Up To (%)

If you specify the second, the first must be Y

If You Use Space Constraint Relief...

- Very large allocations may succeed with large enough volume count.
- Existing data sets may end up with less space than requested on extents.
- New data sets may be smaller than requested.
- Fewer extents may be available when the data set extends.
- May result in more than 5 primary extents
- X37 abends should occur less frequently.

The Retry Process...

- If the volume count is 1:
  - SMS retries the allocation after reducing the requested space as indicated
  - SMS removes the 5 extent limit
- If the volume count is greater than 1:
  - First, SMS uses best-fit volume selection
  - If this fails, SMS reduces the space quantity and removes the 5 extent limit

Requesting Assistance

- Use VOLSELMSG with JOBNAME, STEPNAME, ASID and/or DSNAME to get detailed volume selection analysis messages
- Turn SMS tracing on:
  - SETSMS TRACE(ON),TYPE(ALL),SIZE(100M),DESELECT(ALL), SELECT(MSG,VTOCC,VTOCA,MSG,MODULE),JOBNAME(jobname)
- Run the job
- Turn SMS tracing off
  - SETSMS TRACE(OFF)
- Make note of the dump data set name
- Take a dump of the SMS address space
  - DUMP COMM=(any dump title you desire)
  - R #,JOBNAME=SMS,CONT
  - R #,SDATA=(LPA,CSA,ALLNUC,GRSQ,LSQA,SWA,PSA,SQA,TRT,RGN,SUM)

Requesting Assistance....

- Activate IPCS from a TSO session.
- Set the defaults (dump data set name) using option 0
- Go to the IPCS COMMAND option (IPCS option 6)
- Issue: VERBX SMSDATA 'TRACE'
- If possible, use IPCS PRINT to create a hard copy of the trace

References: APARs

- II07464 - reasons for volume selection failure
- II08004 - reasons why wrong volume selected
- II08442 - volume selection and DCME settings
- II08618 - striping volume selection information
- II08987 - continuation of II08004

References: Publications

- z/OS DFSMSdfp Storage Administration Reference (SC26-7402)
- z/OS DFSMS: Implementing System-Managed Storage (SC26-7407)
- MVS/ESA SML: Managing Storage Groups (SC26-3125)
- z/OS DFSMShsm Storage Administration Reference (SC35-0422)
- z/OS DFSMShsm Storage Administration Guide (SC35-0421)
- z/OS DFSMSdss Storage Administration Reference (SC35-0424)
- z/OS DFSMSdss Storage Administration Guide (SC35-0423)
- z/OS DFSMS Advanced Copy Services (SC35-0428)

Legal Disclaimer

NOTICES AND DISCLAIMERS

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 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.

Legal Disclaimer

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, 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.

Those trademarks followed by ® are registered trademarks of IBM in the United States; all others are trademarks or common law marks of IBM in the United States.

For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml:

*, AS/400®, e business(logo)®, DBE, ESCO, eServer, FICON, IBM®, IBM (logo)®, iSeries®, MVS, OS/390®, pSeries®, RS/6000®, S/30, VM/ESA®, VSE/ESA, WebSphere®, xSeries®, z/OS®, zSeries®, z/VM®, System i, System i5, System p, System p5, System x, System z, System z9®, BladeCenter®

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.