



DB2 ATS (Advanced Technical Skills)

# DB2 and Storage Management, a Guide to Surviving a Perfect Marriage

John Iczkovits

[iczkovit@us.ibm.com](mailto:iczkovit@us.ibm.com)

JPMC



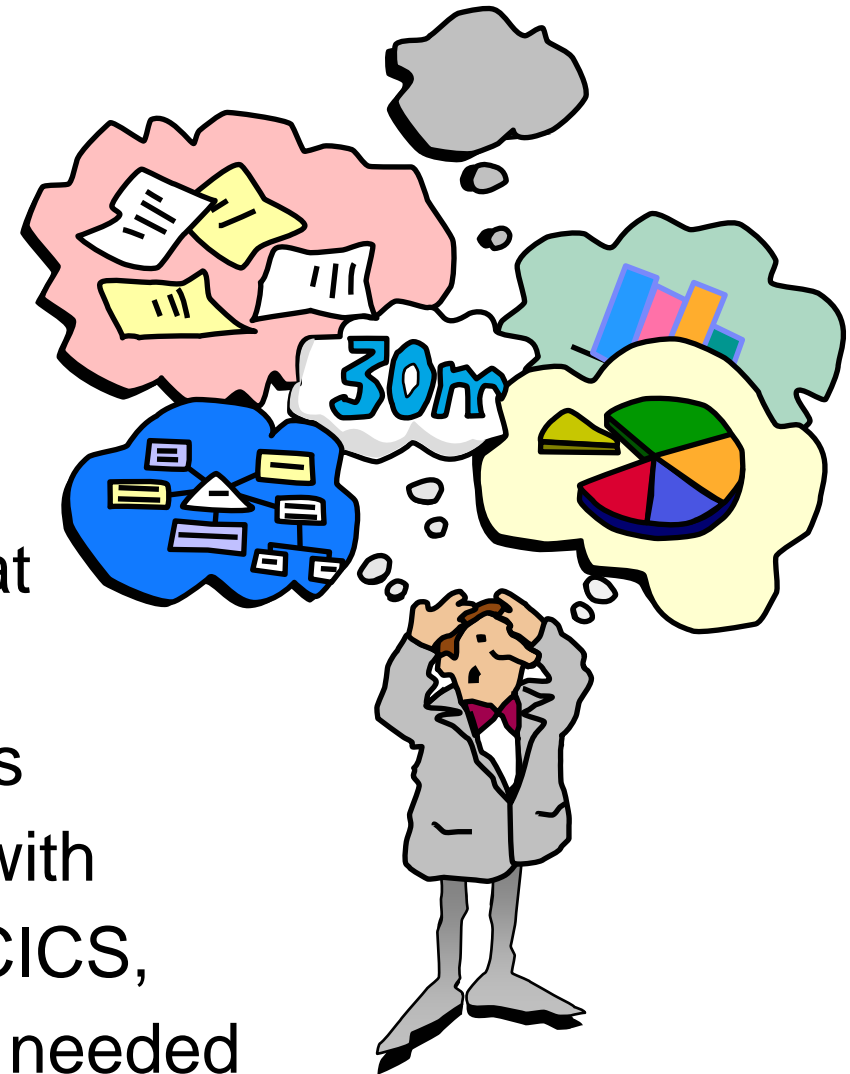
August 7, 2013

# Disclaimer

- **© Copyright IBM Corporation 2010. All rights reserved. U.S. Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.**
- **THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. WHILE EFFORTS WERE MADE TO VERIFY THE COMPLETENESS AND ACCURACY OF THE INFORMATION CONTAINED IN THIS PRESENTATION, IT IS PROVIDED “AS IS” WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. IN ADDITION, THIS INFORMATION IS BASED ON IBM’S CURRENT PRODUCT PLANS AND STRATEGY, WHICH ARE SUBJECT TO CHANGE BY IBM WITHOUT NOTICE. IBM SHALL NOT BE RESPONSIBLE FOR ANY DAMAGES ARISING OUT OF THE USE OF, OR OTHERWISE RELATED TO, THIS PRESENTATION OR ANY OTHER DOCUMENTATION. NOTHING CONTAINED IN THIS PRESENTATION IS INTENDED TO, NOR SHALL HAVE THE EFFECT OF, CREATING ANY WARRANTIES OR REPRESENTATIONS FROM IBM (OR ITS SUPPLIERS OR LICENSORS), OR ALTERING THE TERMS AND CONDITIONS OF ANY AGREEMENT OR LICENSE GOVERNING THE USE OF IBM PRODUCTS AND/OR SOFTWARE.**
- IBM, the IBM logo, ibm.com, DB2 are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at “Copyright and trademark information” at [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml)

## Intent of this Presentation

This presentation is not intended to make DB2 professionals into Storage Administrators, rather to educate DB2 professionals on major storage related items that impact them. Please keep in mind that your Storage Administrator probably does not know as much about DB2 as you do. They are typically also busy with other products, such as: MVS, IMS, CICS, open edition, etc. Common ground is needed to discuss how DB2 uses storage.






For more detailed information review

**Redpaper - Disk storage access with DB2 for z/OS**

<http://www.redbooks.ibm.com/redpieces/pdfs/redp4187.pdf>

**DB2 9 for z/OS and Storage Management SG24-7823**

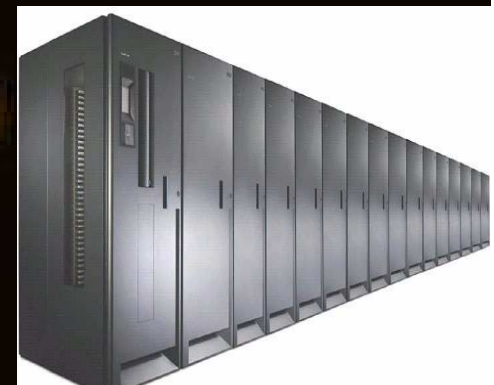
<http://www.redbooks.ibm.com/abstracts/sg247823.html?Open>



**Want to understand  
how DB2 works with tape?**

**Learn about DB2/tape  
best practices:**

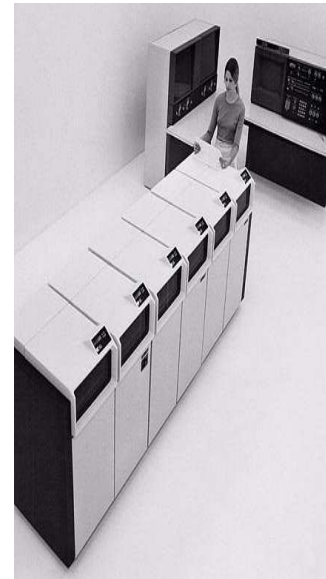
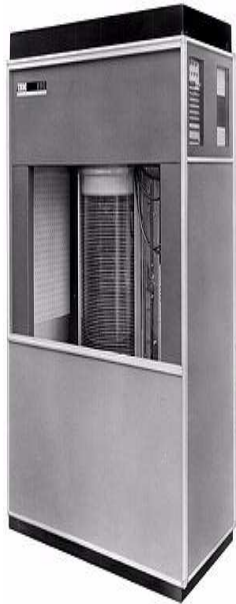
**<http://www.ibm.com/support/techdoc/s/atmastr.nsf/WebIndex/PRS2614>**



## Agenda

- Skipping through IBM Disk history -
  - Disk Architecture -
  - DFSMS Basics for DB2 Professionals –
  - ZPARMs Relating to Storage Management -
  - Utilities –
  - DB2 V10 SMS Required –
    - A few notes about tape -
  - z/OS 1.11, 1.12, and 1.13 enhancements useful for DB2 -
  - FAQ -

# Skipping through IBM Disk history – How did we get here?



First Disk  
350  
5-20 MB  
Total storage

2314  
Up to 10 GB  
Total storage

3330  
Up to 1.6 GB  
Storage  
per box

3350  
635 MB  
Per unit

3380  
10 GB  
Per string

3390  
22.7 GB  
Per unit

# RAMAC Array Direct Access Storage Device (DASD) and the RAMAC Array Subsystem - 1994

## Virtual Disk Architecture



IT'S DASD, BUT NOT AS YOU KNOW IT !



RVA = RAMAC Virtual Array (1997)

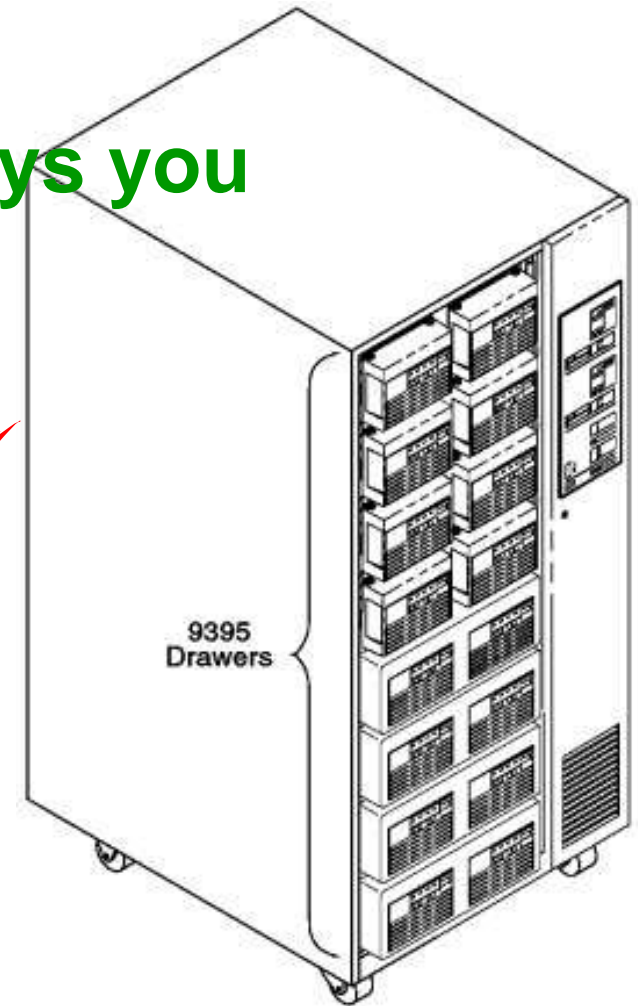


## RVA - 1997

Up to 840 GB of information storage

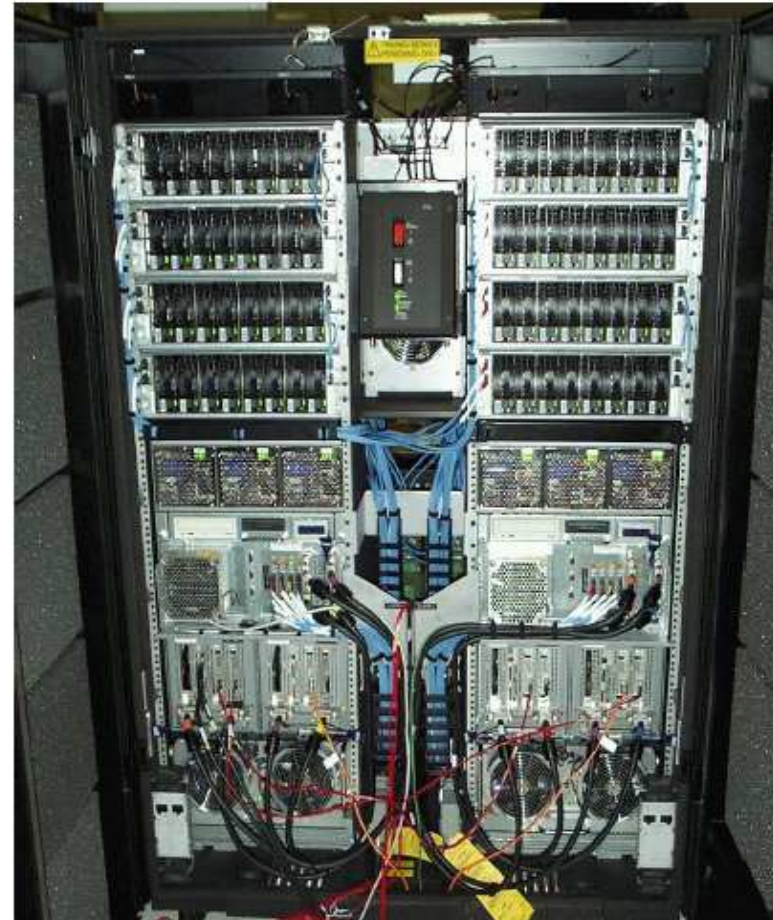
## Virtual Disk Architecture - What it buys you

- Enables compression (different on other IBM boxes) ✓
- Enables performance ✓
- Enables SnapShot (1997) ✓
- Reduces IOSQ ✓
- Simplifies operations and management ✓
- Hot spot avoidance ✓
- Dynamic configuration ✓
- RAID write penalty avoidance ✓
- 3380 and/or 3390 emulated in one box ✓



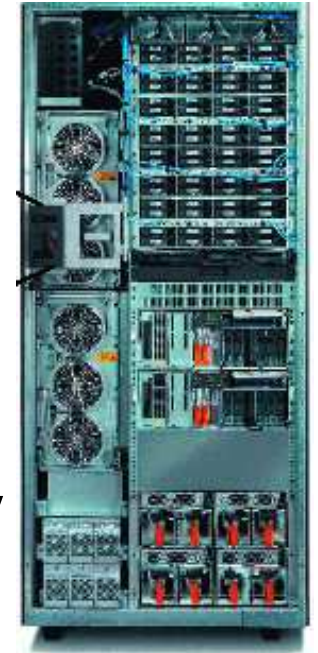
## IBM Enterprise Storage Server – code named "Shark" - 1999

- Scalable up to 55.9 TB
- FlashCopy
- Supports PAV and MA
- Large cache structures and sophisticated caching algorithms
- Different than RVA, but still using disk array concept.



## IBM TotalStorage DS8000 Series - 2004

- Different than RVA and ESS, but still using disk array concept.
- Capacity scales linearly up to 1,024 TB (more for raw data)
- FlashCopy
- Supports PAV and MA
- With the implementation of the POWER5 (POWER6+ with DS8800) Server Technology in the DS8000 it is possible to create storage system logical partitions (LPAR)s, that can be used for completely separate production, test, or other unique storage environments
- New caching algorithms (all variations still used):
  - 2004 - ARC (Adaptive Record Cache)
    - Dynamically partitions the read cache between random and sequential portions
  - 2007 – AMP (Adaptive Multi-Stream Pre-Fetch – R3.0)
    - Manages the sequential read cache and decides what, when and how much to prefetch
  - 2009 – IWC (Intelligent Write Cache – R4.2)
    - Manages the write cache and decides what order and rate to destage



# IBM System Storage DS8000 Turbo – Powerful Innovation (information below are not turbo specific)

## Innovation that extends DS8000 world class performance

- Storage Pool Striping** –new volume configuration option to maximize performance without special tuning
- AMP**-breakthrough caching technology can dramatically improve sequential read performance to reduce backup times, processing for BI/DW, streaming media, batch
- z/OS Global Mirror Multiple Reader**- IBM unique Innovation to improve throughput for z/OS remote mirroring

## Innovation to simplify and increase efficiency

- IBM FlashCopy SE** (space efficient snapshot capability) can lower costs by reducing capacity needed for copies.
- Dynamic Volume Expansion** - Easier, online, volume expansion to support growth
- Expansion frame warranty intermix** - Increased upgrade flexibility and investment protection through base and expansion frame machine type (warranty) Intermix
- SSL**-New secure connection protocol option for call home support and additional audit logging
- IBM System Storage Productivity Center** -Enhanced user interface with single pane control and management

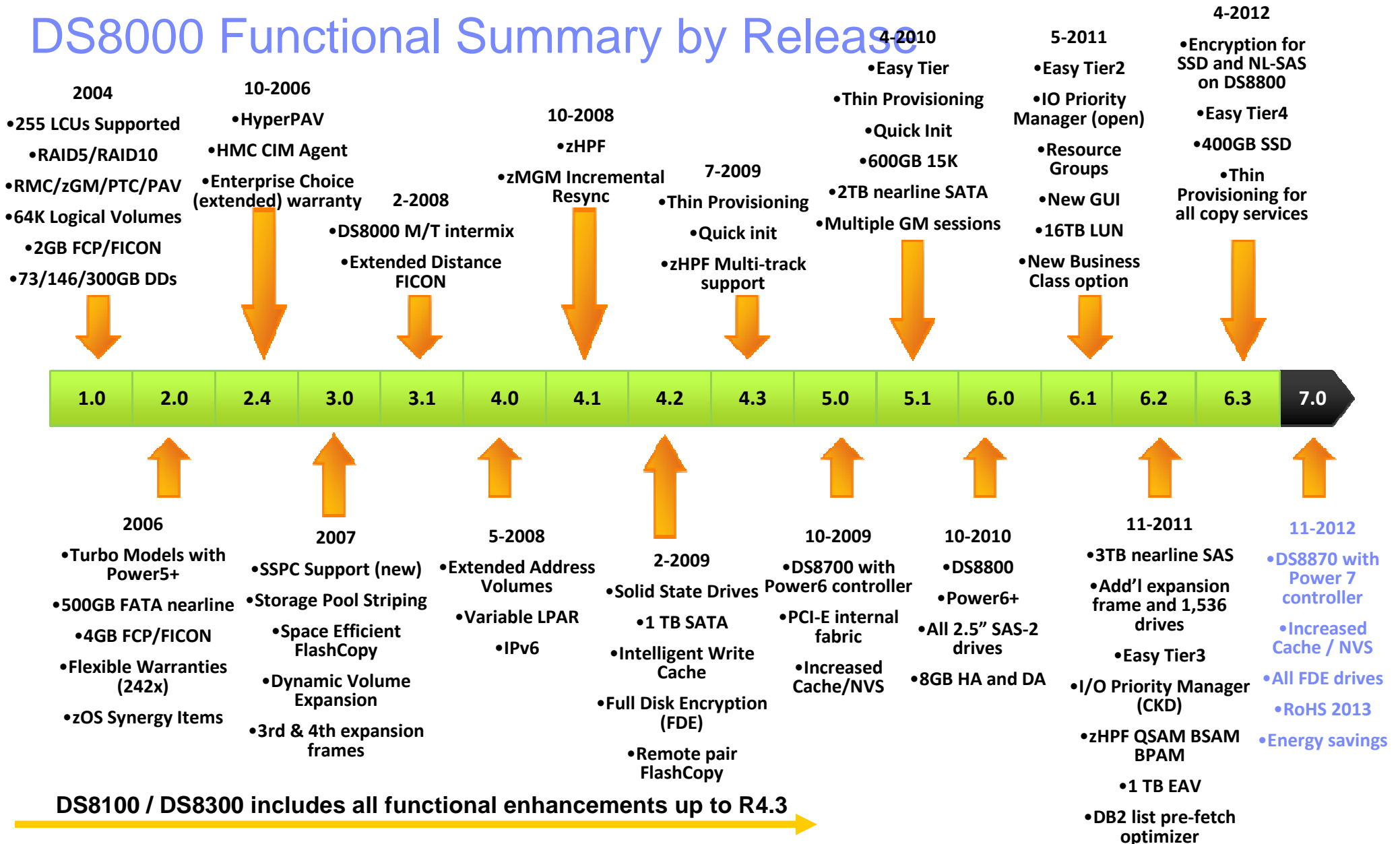


***Enabled through new release of microcode***

***New and existing customers can leverage the new capabilities!!***

**Note: not all enhancements are included**

# DS8000 Functional Summary by Release





# IBM System Storage DS8800 (Release 6.0)

*Up to over 40% increase in performance and almost 90% more drives in same single-frame footprint*

- **System-wide hardware upgrade and IBM POWER Server synergy**
  - New POWER6+ controller 5Ghz (2-way/4-way)
  - New 8Gb/s host and device adapters
  - New 2.5" 6Gb/s SAS (SAS-2) drives and drive enclosures
- **Features / Business Value**
  - New processors and adapters boost performance with faster access to host servers and disk drives
  - New small-form-factor drives offer faster performance, higher availability, lower energy consumption, and allow denser footprint for more effective scalability
- **Client Benefits**
  - System-wide upgrades enable faster performance and higher storage capacity within the same footprint
  - Smaller form factor drives offer faster performance, higher availability, smaller footprint, and lower energy consumption



# Disk Architecture

# RAID (Redundant Arrays of Inexpensive Disks) Technology- Disk Type Dependent

- RAID 0 - data striping without parity
- RAID 1 - dual copy
- RAID 2 - synchronized access with separate error correction disks
- RAID 3 - synchronized access with fixed parity disk
- RAID 4 - independent access with fixed parity disk
- **RAID 5 – data striping, independent access with floating parity (tolerates loss of 1 physical volume (DDM))**
- **RAID 6 – data striping, dual redundancy with floating parity (tolerates loss of 2 physical volumes (DDMs))**
- **RAID 10 (DS8000 and some ESS) - RAID 0 + RAID 1, mirrored, data striping but with no parity**

Parity is additional data, “internal” to the RAID subsystem, that enables a RAID device to regenerate complete data when a portion of the data is missing. Parity works on the principle that you can sum the individual bits that make up a data block or byte across separate disk drives to arrive at an odd or even sum



# Array

## ■ DS8000 – 8 DDM arrays

–1 array site

–RAID5

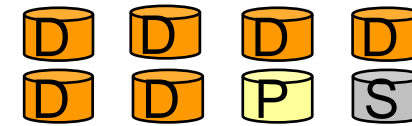
- 6+P
- 7+P
- Parity is striped across all disks in array but consumes capacity equivalent to one disk

–RAID 6

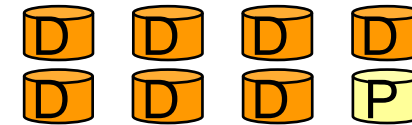
- 6+P+Q
- 5+P+Q+Spare

–RAID10

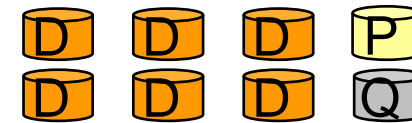
- 3+3
- 4+4



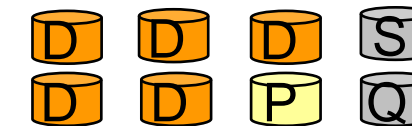
RAID5 6+P+S



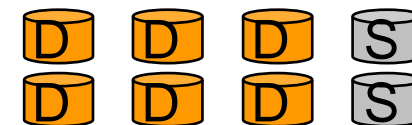
RAID5 7+P



RAID6 6+P+Q



RAID6 5+P+Q+S



RAID10 3+3+S+S

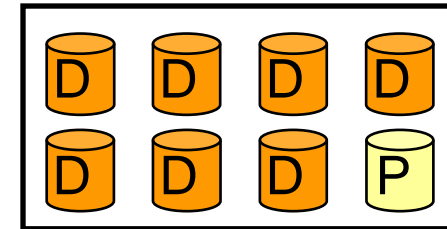


RAID10 4+4

## Rank

- **RAID array with storage type defined**
  - CKD or FB
  
- **One-to-one relationship between an array and a rank**
  - One RAID array becomes one rank
  - DS8000 – 8 DDMs

DS8000 CKD rank

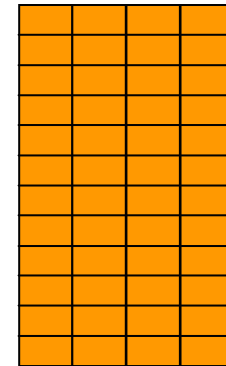


RAID5 7+P

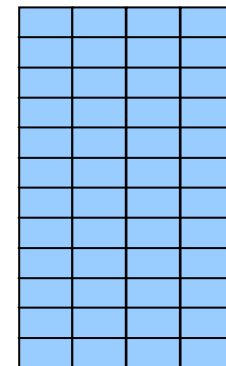
## Rank (continued)

- **Ranks have no pre-determined or fixed relation to:**
  - Server0 or Server1
  - Logical Subsystems (LSSs)
  
- **Ranks are divided into ‘extents’**
  - Units of space for volume creation
  - CKD rank
    - Extents equivalent to a 3390M1
    - 1113 cylinders or .94GB
  - FB rank
    - 1GB extents

CKD Rank



FB Rank



# Storage Resource Summary

## ■ Disk

- Individual DDMs

## ■ Array Sites

- Pre-determined grouping of DDMs of same speed and capacity (8 DDMs for DS8000; 4 DDMs for DS6000)

## ■ Arrays

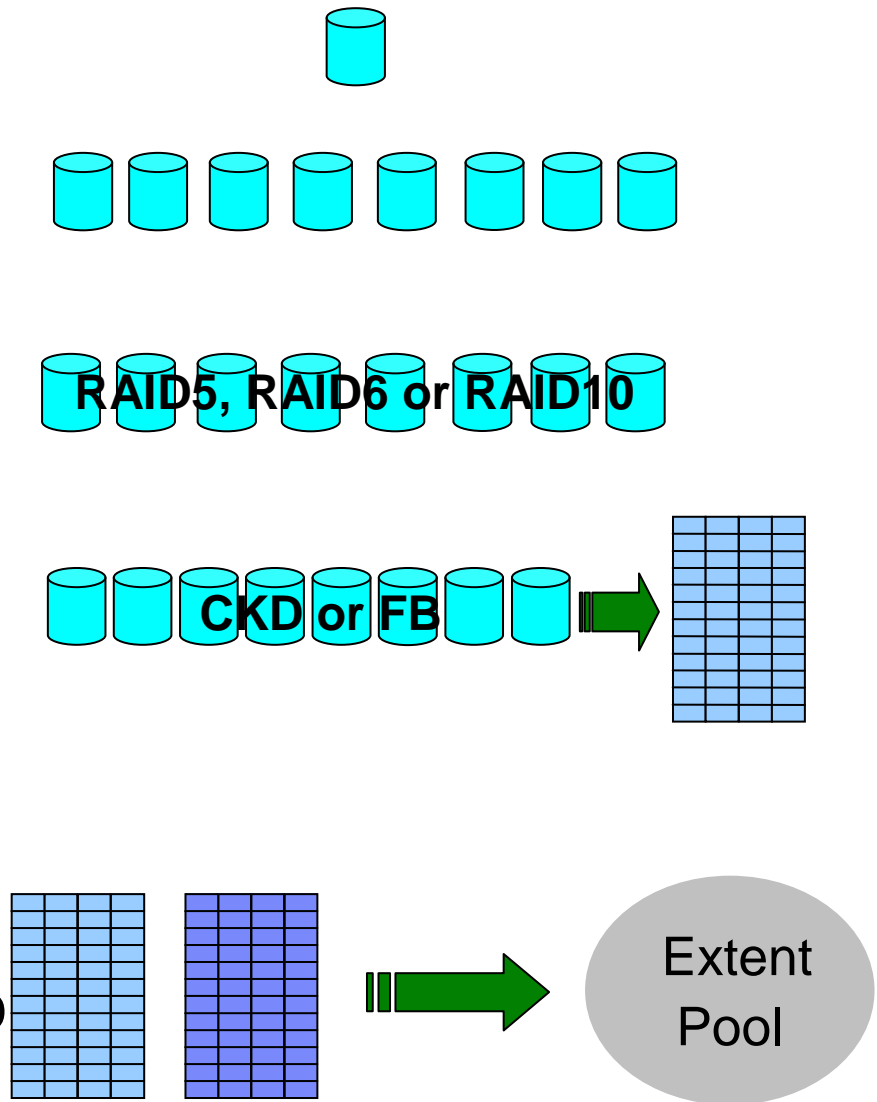
- One 8-DDM Array Site used to construct one RAID array (DS8000)

## ■ Ranks

- One Array forms one CKD or FB Rank (8 DDMs for DS8000; 4 or 8 DDMs for DS6000)
- No fixed, pre-determined relation to LSS

## ■ Extent Pools

- 1 or more ranks of a single storage type (CKD or FB)
- Assigned to Server0 or Server1



## Disk Space Numbers - 3390 Emulated Volumes

Model	Cylinders	Tracks	Bytes/Volume	Bytes/Track **
3390-1	1113	16695	946 MB	56664
3390-2	2226	33390	1.89 GB	56664
3390-3	3339	50085	2.83 GB	56664
3390-9 *	10017	150255	8.51 GB	56664
3390-27 *	32760	491400	27.84 GB	56664
3390-54 *	65520	982800	55.68 GB	56664
3390-A <b>EAV</b>	1,182,006	* * *	1TB	56664

-EAVs now supports 1TB volumes with z/OS 1.13 and DS8K LIC (Release 6.2).

-z/OS 1.12 supports 1TB volumes with required PTFs

\* Storage Administrators refer to 3390 mod 9, 27, and 54 as large-volume support 10017 cylinders, 32760 cylinders, and 65520 cylinders respectively. These large volumes are all considered mod 9s with varying number of cylinders.

\*\* Bytes per track refers to the device capacity only. The actual allocation for a DB2 LDS data set is 48 KB, not the total of 56 KB. This allows for 12 - 4K DB2 pages to be allocated per track.

Why larger disk sizes?

- reduces issues with MVS addresses (maximum number of devices met)
- simpler management of storage

# What is a DDM (Disk Drive Modules) physical drive?

- **Three different Fibre Channel (FC) DDM types (available in both non-encrypted and encrypted versions):**
  - 146 GB, 15K RPM drive
  - 300 GB, 15K RPM drive
  - 450 GB, 15K RPM drive
  - See next page regarding 2.5 inch SAS drives for the DS8800, which includes 600 GB, 10K RPM drives
- **One Serial Advanced Technology Attachment (SATA) DDM drive type:**
  - 1 TB, 7.2K RPM drive
  - 2 TB, 7.2K RPM drive
- **Two different Solid State Drive (SSD) types:**
  - 73 GB
  - 146 GB
  - See next page regarding for the DS8800, which includes the new 300 GB drives
- **Disks are 2.5 inch SAS (Serial Attached SCSI) for DS8800, otherwise 3.5 inches FC**
- **Each DDM (physical disk) can hold many logical volumes. For example, a mod 3 device is 2.83 GB. Many mod 3s can reside on one DDM. When you lose one DDM, how many logical volumes are you really losing? Think about storage pool striping in the next few slides as well.**
- **Three types of drives:**
  - HDD (Hard Disk Drive) Fibre Channel. Almost all mainframe disk is HDD.
  - SATA. Too slow for most mainframe uses. Some companies will use SATA drives for Data Warehousing if slower performance is acceptable, or for HSM Level 2, or backups, such as image copies. Much slower performance must be acceptable when putting data on SATA drives. Generally SATA drives are not used for DB2 environments.
  - SSD. VERY expensive and not generally used for DB2 environments. Very useful when access is random. Random access is better on SDD vs. HDD or SATA.

# DS8800 Hardware Changes

## ■ Drives

- Seagate Hornet 15K 146 GB non-FDE (Full Disk Encryption)
- Seagate Firestorm 10K RPM 450 GB and 600 GB both FDE and non-FDE.
- STEC Hikari SSD 300 GB non-FDE
  - SSD drive sets are not supported in RAID-6 or RAID-10 configurations
- **All vendors including IBM depend on other vendors for their physical disk. For example, for IBM devices, Seagate provides all vendors disk for HDD and SATA, while STEC provides all vendors for SSD. What is very different is how each vendor chooses to implement the software that drives the hardware.**

## SSD (Solid State Disk) Candidates

- **SSD has no moving parts. This also means that traditional latency times are reduced to primarily data transfer, since disk seek, scan, etc. are no longer performed for SSD.**
- **Today SSD is only practical if there is high access density (IO/sec/GB)**
- **Data sets with good cache hit ratios or low I/O rates should remain on spinning disks**
- **Sequential data sets should remain on spinning disks**
- **Data sets with high I/O rates and poor cache hit ratios are good candidates for SSD, but they are likely to be large data sets**



## Easy Tier - Overview

- **Easy Tier is a DS8700 and DS8800 feature that supports online dynamic relocation of data at the sub-volume/LUN level**
  - Workload learning algorithms collect I/O statistics and generate a heat map of data to be optimized on SSD
  - Data can be relocated to/from SSD and FC/SATA
    - Automatic Mode
    - Manual Mode
  - Storage Tier Advisor Tool (STAT) for I/O analysis and projected benefit
- **DS8700 Hardware feature**
  - Easy Tier is a new LIC feature available with Release 5.1
  - Supported by all server platforms with no additional software

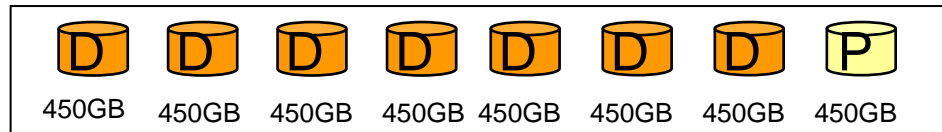
- **MYTH: My data always resides on the physical volser they were allocated on**
  
- **FACT: IBM's Easy Tier product can move "data" between different types of disk or the same disk**
  - Customer can have a mix of two or all three types of IBM disk. Easy Tier will move "data" which is a volume extent (not a data set extent) to the best fit disk for performance
  - Extents are moved dynamically without an outage
  - Customer example – owns HDD and SSD volumes. DB2 data is mostly random residing on HDD. Easy Tier can move the volume extent to an SSD volume.
    - No outage required!
    - VOLSER is the same, however the volume extent resides on different physical volumes and different type of disk
    - DB2 is not aware of the move
    - When the data set is REORGed and no longer flagged as random, placement will probably be on HDD and the process has to start over again.
  - IBM disk competitors have similar products for their disk
  
  - **Warning!** Even if you only have HDD, extents can be moved laterally to any rank in the disk box, from a hot rank to cool rank. Make sure that the BSDS and active log data sets are excluded from East Tier or equivalent product to avoid them residing in the same extent pool.

## Extents – what are they?

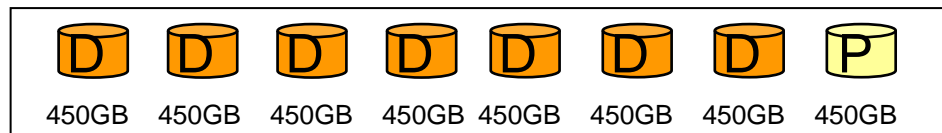
- **Lots of discussions and opinions regarding extents.**
  - Do they still matter?
  - How many extents are too many extents?
  - Based on how many extents should objects be REORGed?
- **Extents are more logical than physical**
  - Mostly people view an extent as a long stripe of data under one read/write head
  - In older technology, data was viewed from a vertical point of view. Tracks are placed in the same area on different read/write heads on different platters. Tracks refer to the read/write head. Depending on the technology, cylinder 1 track 1 would be on platter 1 using read/write head 1, cylinder 1 track 2 would be on platter 2 using read/write head 2, etc. Data is aligned vertically on different platters.
  - Newer technology does not use this approach. Tracks are spread out using a different approach. Tracks are no longer aligned with read/write heads on different platters.

## Storage Pool Striping

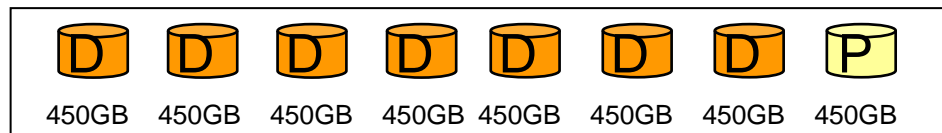
- Example of using storage pool striping with rotating extents across 3 ranks with RAID 5 (7+P) for a mod 3 device (2.83GB - 3,339 cylinders) – 450GB DDMs



Rank 1 – 1,113 cylinders  
Logical volume - .94GB



Rank 2 – 1,113 cylinders  
Logical volume - .94GB



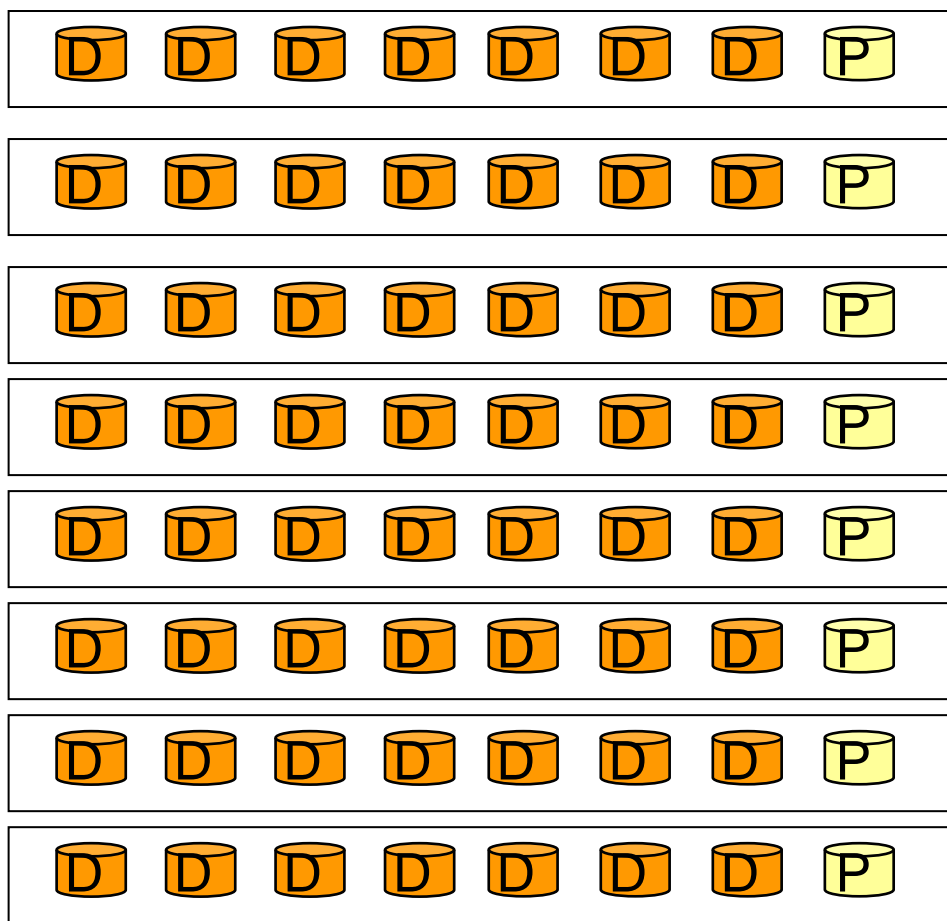
Rank 3 – 1,113 cylinders  
Logical volume - .94GB

**This one logical mod 3 volume with 3,339 cylinders really resides on 24 physical devices (DDMs)**

**In this scenario when you backup or Flash Copy one logical volume you are really accessing 24 physical volumes.**

- Parity is striped across all disks in array but consumes capacity equivalent to one disk

- **Example of using storage pool striping with rotating extents across 8 ranks with RAID 5 (7+P) for a mod 9 device (8.51GB - 10,017 cylinders) - 450GB DDMs**



Rank 1 – 1,113 cylinders

Logical volume - 1.88GB

Rank 2 – 1,113 cylinders

Logical volume - .94GB

Rank 3 – 1,113 cylinders

Logical volume - .94GB

Rank 4 – 1,113 cylinders

Rank 5 – 1,113 cylinders

Rank 6 – 1,113 cylinders

Rank 7 – 1,113 cylinders

Rank 8 – 1,113 cylinders

Logical volume - .94GB

This one logical mod 9 volume with 10,017 cylinders really resides on 64 physical devices (DDMs). Mod 9 devices require the equivalent of 9 ranks. Since only 8 ranks exist, the last 1,113 cylinders will be added back to rank 1. When adding to ranks, full ranks are skipped

## IBM System Storage DS8800 Architecture and Implementation SG24-8886

**“Storage Pool Striping can enhance performance significantly, but when you lose one rank (in the unlikely event that a whole RAID array failed due to a scenario with multiple failures at the same time), not only is the data of this rank lost, but also all data in this Extent Pool because data is striped across all ranks. To avoid data loss, mirror your data to a remote DS8000.”**

## Storage Pool Striping - Advantages

### ■ **Technical Advantages**

- Method to distribute I/O load across multiple Ranks
- DS8000 optimized for performance
- Far less special tuning required for high performance data placement.
  - Means less work for the storage administrator
- Reduces storage administrator work needed to optimize performance
- Less potential for hot spots

## How many DDMs does my single volume data set reside on?

- From a VTOC perspective an allocation on a single volume will display one (logical) volume. In reality the data set resides on several different physical volumes.
- For our examples we are using a RAID 5 7+P configuration with at least 3 ranks
- Data set is on a single logical volume with 2,500 cylinders allocated. Can be 1 or multiple extents, in this case it does not matter. In this scenario the data set is allocated on 24 DDMs (3 ranks \* 8 DDMs).
- Data set is on a single logical volume with 25 cylinders allocated in 1 extent. Allocation can be on 8 DDMs (1 rank), or 16 DDMs (2 ranks). Since the VTOC is not aware of ranks, it is possible there is a chunk of space that spans ranks that they are both used. For something as small as 25 cylinders, 3 ranks would not be used as they would not exceed 1 extent on 2 ranks.
- Data set has 2 extents on a single logical volume, each extent is 1 track. It depends. For example, extent 1 was allocated on 1 array, extent 2 is not allocated until a month later. Extent 2 could be on any rank with sufficient space. It is possible that extent 1 resides on an array in one rank, and that extent 2 resides on a different array in a different rank.



# Do extents really matter?

- **On what physical volumes do my extent reside? No idea!**
- **1 extent does not mean the data is necessarily contiguous, nor ever on the same array and rank**
- **From my very rudimentary tests:**
  - 1 data set with 1 extent allocated with 100-200 cylinders of data
  - Have another data set with 100 extents allocated with 100-200 cylinders of data. Data inserted is the same across both data sets
  - SELECT \* for both data sets took about the same time.
  - Open time for the 100 extent data set was roughly double the one for the 1 extent data set. MVS open is one of the most expensive operations.
    - VSAM data sets can now be 7,257 extents when overriding the Data Class attribute for your LDS. If 100 extents took 2x open over the 1 extent data set, how long would it take to open a 3,000 extent data set?
  - Tests were based on read (SELECT) and not insert, which would take additional time to create the extents, as well as a very small additional time for sliding secondary if on

After I REORG my data set in many extents, I always see performance improvements even on reads

- **Does the problem relate to extents, or rather disorganized data?**
- **Many times the problem is not extents, rather some common problems that reorganizing the data resolves as opposed to the number of extents:**
  - Pseudo deleted RIDs
  - Cluster ratio < 94% (fixed in DB2 V10)
  - High page splits

## LDS limit of 7,257 extents

- **LDS (Linear Data Sets – almost all of your DB2 VSAM data sets) can now reach 7,257 extents. dfp rules state that an LDS can reside on:**
  - 59 volumes \* 123 extents per volume = 7,257 extents
  - 7,257 extents is the architectural limit. You have 10 volumes in the Storage Group, 10 volumes \* 123 extents per volume = 1,230 extents. Another example, SG = 10 volumes, with no more than 40 extents per volume, 10 volumes \* 40 extents per volume = 400 extents.
  - By default the maximum is still 255 extents per LDS. The Storage Administrator must override the 255 extent rule in the Data Class.
    - From my experience, most customers do not have Data Class assigned for the DB2 LDSes. This means that 255 extents is used as it is the default.
  - To exceed the 255 extent limit, data set must be SMS managed

# Why is it hard to test performance on LDSes with more than 255 extents?

59 volumes \* 123 extents per volume = 7,257 extents (architectural limit)

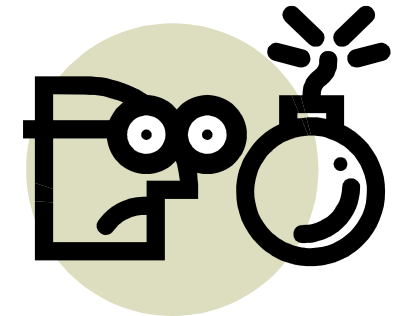
- **Only SMS managed data sets can exceed 255 extents**
- **SMS uses adjacent extent removal**
- **Having an SMS managed data set reach 123 extents on 1 volume is very difficult – the volume would need to be extremely fragmented. Having 59 volumes that are extremely fragmented allowing for 123 extents per volume for an LDS is even more difficult.**
- **To test 7,257 extents, because of adjacent extent removal:**
  - Turn off sliding secondary in ZPARMS (MGEXTSZ=NO). Do not run this test with other work for DB2 depending on sliding secondary.
  - Allocate a table space and insert an amount of data on 1 volume
  - Allocate a 1 track sequential data set residing right behind your LDS allocation so that the table space takes a new extent on next insert to avoid adjacent extent removal.
  - More inserts to fill up the next extent
  - Allocate a 1 track sequential data set residing right behind your next extent
  - Do this 123 times per volume on 59 volumes!
  - Want to try a smaller test, just 3,000 extents?
    - 123 extents per volume \* 25 volumes
      - You may only get 119 extent per volume, to test 3,000 extents with 119 extents per volume, you will need 26 volumes

## z/OS 1.12 Faster Data Set Open

- **See APARs:**
  - PM00068
  - PM17542
  - PM18557
  
- **I have not yet tested the effects of these APARs in regards to DB2 and number of extents**

## MTBF (Mean Time Between Failure)

- Depending on the manufacturer, the MTBF for disk is generally 1,000,000 (1M) + hours.
- MTBF also depends on type of disk, where SDD ranges between 1.5M and 2M hours.
- Even at the lower number of 1M, that is 114 years as mean time between failures!
- MTBF refers to the physical disk, it does not include problems such as:
  - Outside of the physical disk itself, such as a bad card in the controller
  - VTOC being ZAPed incorrectly causing the volume to become unusable



# Overview - DS8000 Support

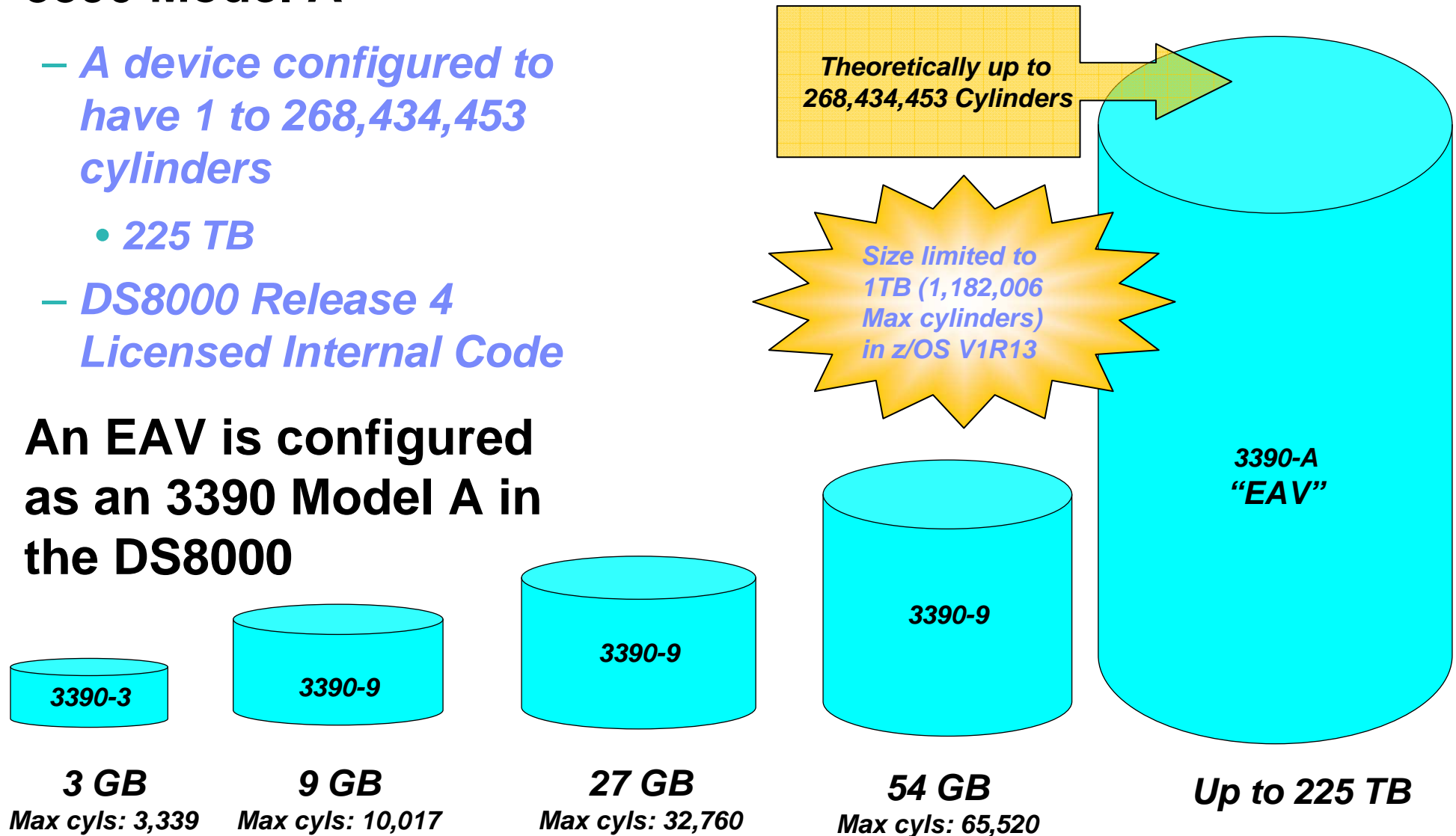
## ■ 3390 Model A

– A device configured to have 1 to 268,434,453 cylinders

- 225 TB

– DS8000 Release 4 Licensed Internal Code

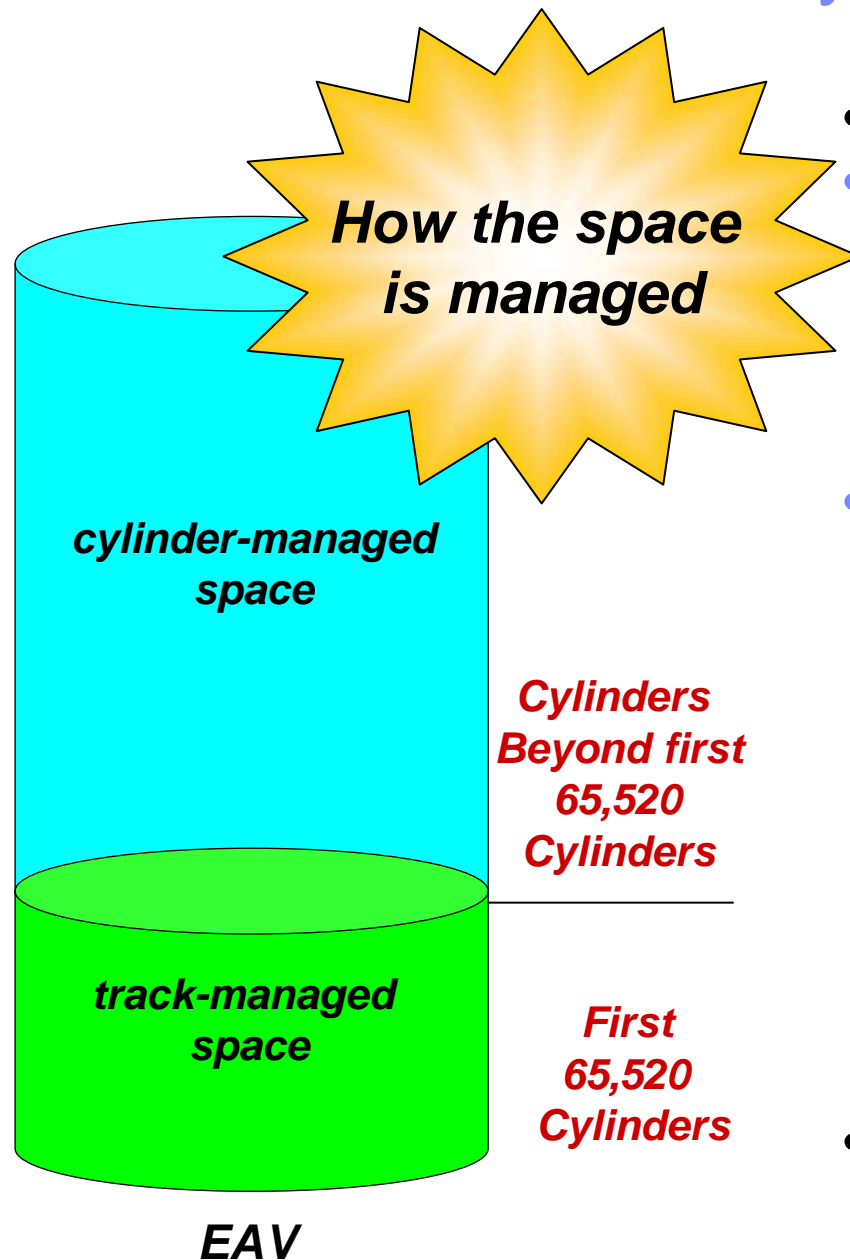
## ■ An EAV is configured as an 3390 Model A in the DS8000



–EAVs now supports 1TB volumes with z/OS 1.13 and DS8K LIC (Release 6.2).

–z/OS 1.12 supports 1TB volumes with required PTFs

## Overview – EAV Key Design Points



- Maintains 3390 track format
- **Track-managed space:** the area on an EAV located within the first 65,520 cylinders
  - Space is allocated in track or cylinder increments
  - Storage for “small” data sets
- **Cylinder-managed space:** the area on an EAV located above the first 65,520 cylinders
  - Space is allocated in **multicylinder units**
    - A fixed unit of disk space that is larger than a cylinder. Currently on an EAV it is 21 cylinders
    - System may round space requests up
  - Storage for “large” data sets
- Track-managed space comparable to same space on non-EAVs



**DB2 data sets may reside in EAS, by z/OS release with appropriate maintenance applied**



**\* eligible starting in DB2 V8**

**\*\* eligible starting in DB2 V9**

	z/OS 1.10	z/OS 1.11	z/OS 1.12
Tables and Indexes *	Yes	Yes	Yes
BSDS *	Yes	Yes	Yes
Active Logs *	Yes	Yes	Yes
Archive Logs **	No	Yes, if EF sequential	Yes
Utilities sequential input and output data sets *	No	Yes, if EF sequential	Yes
Utilities partitioned data sets and PDSEs	No	No	Yes
Sort Work data sets	No	No	Yes, if using DFSORT for DB2 utilities
DB2 Installation data sets (clists, panels, samples, macros, etc)	No	No	Yes
SDSNLINK, SDSNLOAD	No	No	Yes
HFS	No	No	No

# Using larger volume sizes

## ■ Benefits

- Fewer objects to define and manage
- Less processing for fewer I/O resources
  - CF CHPID, VARY PATH, VARY DEVICE
  - Channel path recover, link recovery, reset event processing
  - CC3 processing
  - ENF Signals
  - RMF, SMF
- Number of physical resources: CHPIDs, Switches, CU ports, fibers
- Each device consumes real storage:
  - 768 bytes of real storage for UCB and related control blocks
  - 256 bytes of HSA
  - 1024 bytes/device \* 64K devices = 64MB
  - 31 bit common storage constraints
- EOVS processing to switch to the next volume of a sequential data set significantly slows the access methods

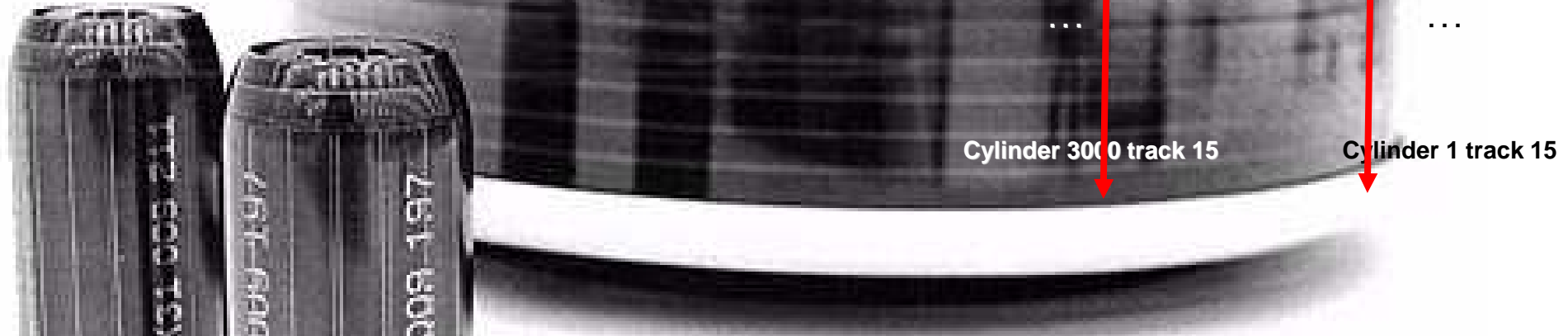
## ■ Considerations

- Data migration to larger devices may be challenging, time consuming

IBM 3850 data cartridge

and

IBM 3336 disk pack



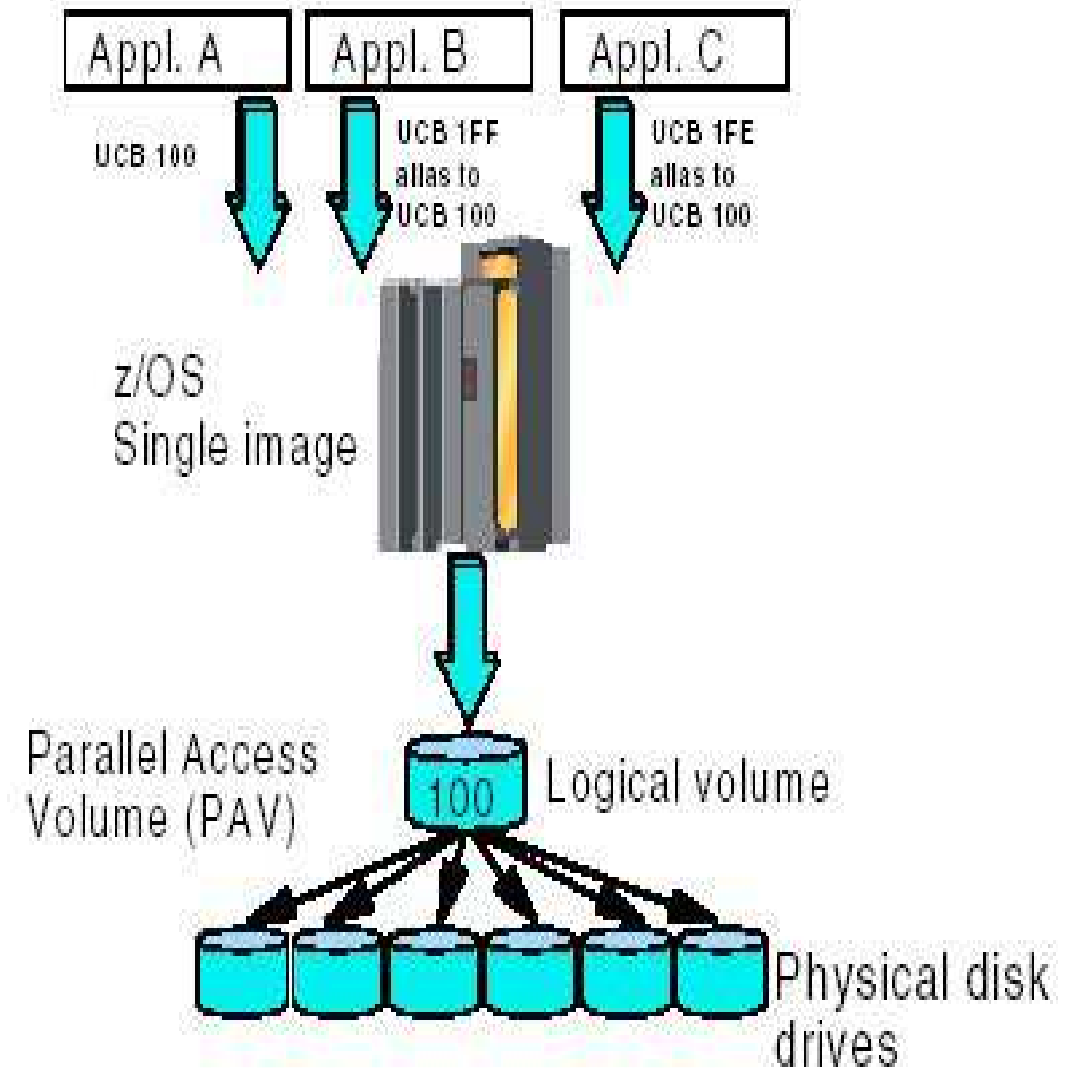
When reviewing CCCHHHH (Cylinder Head) information, MVS still views technology when multiple disk platters were used. Above is not a 3390 depiction, but will do for illustration.

- CCCC is the cylinder HHHH is the read/write head. CCCHHHH is the location of 1 track.
- Each track is either on its own platter or top and bottom of a platter. Each platter has either a dedicated or shared read write head.
- Cylinder and head information is off by 1 in LISTCAT and IEHLIST output
- CCCHHHH '0000000' translates to cylinder 1, head 1 (cylinder 1 track 1)
- CCCHHHH '0000001' translates to cylinder 1, head 2 (cylinder 1 track 2)
- CCCHHHH '000000E' translates to cylinder 1, head 15 (cylinder 1 track 15)
- When 1 cylinder is read, all read/write heads read data in unison

**•NOTE – this format is no longer reality with disk arrays. It has not been reality for two decades, however MVS still reports disk allocation with the old cylinder and head formats**

## Parallel Access Volumes (PAV) - ESS "Shark" and DS8000

- Multiple UCBs per logical volume
- PAVs allow simultaneous access to logical volumes by multiple users or jobs from one system.
- Reads are simultaneous
- Writes to different domains are simultaneous
- Writes to same domain are serialized
- Eliminates or sharply reduces IOSQ
- High I/O activity, particularly to large volumes (3390 mod 9, 27, and 54) greatly benefits from the use of PAV.
- WLM GOAL mode management of dynamic PAVs. Static PAVs do not require WLM. Dynamic PAVs and Priority I/O Queueing recommended. Hyper PAVs are gaining in popularity and is recommended over static and dynamic PAV.
- Multiple paths or channels for a volume is nothing new, but multiple UCBs (MVS addresses) for a volume is.



# Dynamic PAV (Parallel Access Volumes)

aka “WLM-managed PAV”

- ✓ **PAVs reduce or avoid IOSQ time. PAVs are essential for large volumes.**
- ✓ **Each PAV is assigned one of 64K UCB addresses. The number of UCB addresses limits the number of PAVs in a sysplex.**
- ✓ **PAVs are pooled by LCU (Logical Control Unit)**
- ✓ **PAVs are dynamically assigned by Workload Manager to a base volume**
- ✓ **The assignment must be coordinated throughout the sysplex**
  - Lots of WLM overhead
- ✓ **Different systems compete for PAVs if the systems share the same LCU**
- ✓ **The number of PAVs needed to avoid IOSQ time far exceeds the maximum number of concurrent I/Os, especially in a sysplex environment**

## Hyper PAV

- ✓ **PAVs are assigned to a base volume only for the duration of the I/O**
- ✓ **No coordination required across the sysplex**
- ✓ **Different systems won't compete against each other**
- ✓ **PAVs are still pooled by LCU**
- ✓ **Advantages:**
  - 64K UCB address constraint is relieved
  - No WLM overhead
  - The total number of PAVs needed for an LCU is equal to the maximum number of concurrent I/Os
  - The same alias address can be used by different systems for different volumes
- ✓ **System Requirements: z/OS 1.8, Release 2.4 ucode for the DS8000, upgradeable for existing DS8000 control units**
  - PTFs to be supplied for z/OS 1.6 and 1.7

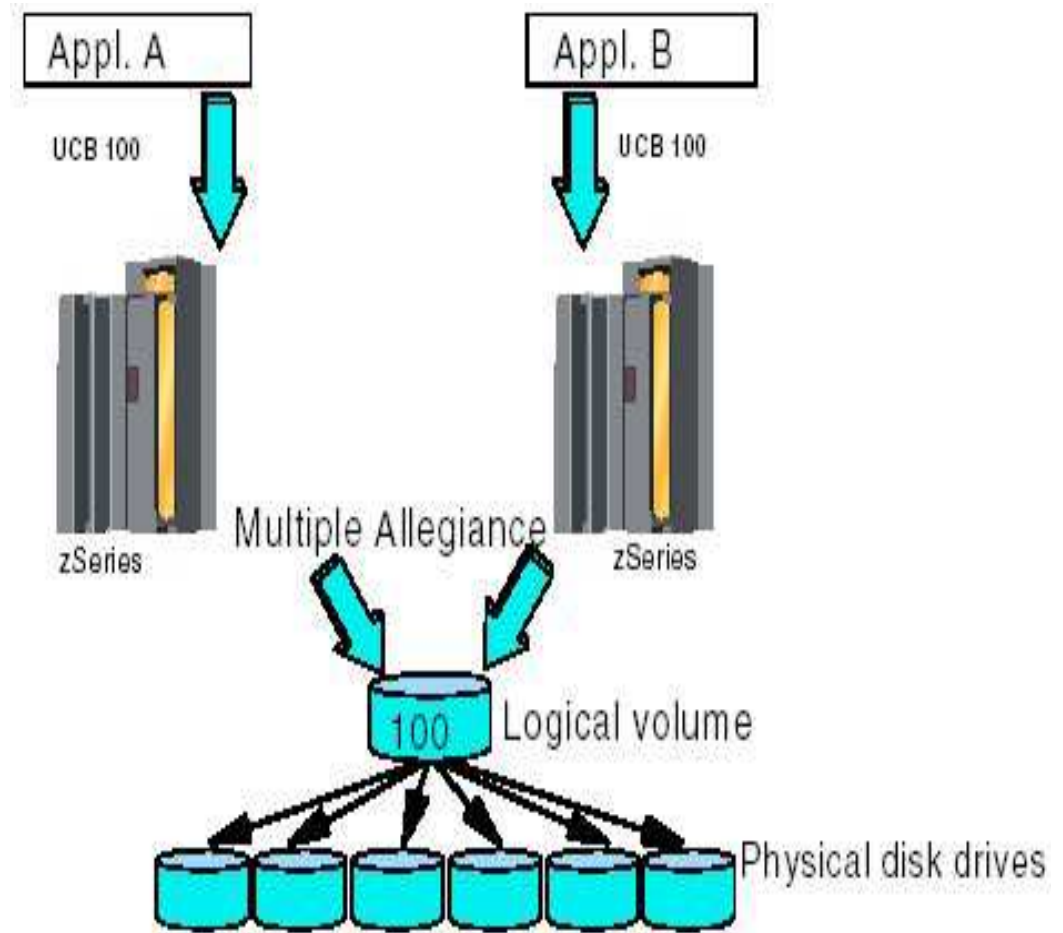
# Hyper PAVs

Workloads that benefit most from UCB constraint relief

- ✓ **PPRC (Synchronous Peer to Peer Remote Copy)**
  - Eliminating IOSQ time allows reads to be serviced during writes and also allows more writes to be service in parallel.
- ✓ **If the cache hit ratio is poor, PAVs may only shift queue time from IOSQ to disconnect time, because of DDM contention. However, shifting the queues into the control unit prevents cache hits from queueing behind cache misses.**
- ✓ **Different hosts access the same LCUs, but with different device skews.**

## Multiple Allegiance (MA) - ESS "Shark" and DS8000

- **Similar to PAV, however for more than one LPAR. Unlike PAV, MA is automatically turned on with your IBM disk.**
- **Incompatible I/Os are queued in the ESS/DS8000**
- **Compatible I/O (no extent conflict) can run in parallel**
- **ESS/DS8000 boxes guarantee data integrity**
- **No special host software required, however: Host software changes can improve global parallelism (limit extents)**
- **Improved system throughput**
  - Different workloads have less impact on each other
- **Reduces PEND time (device busy)**

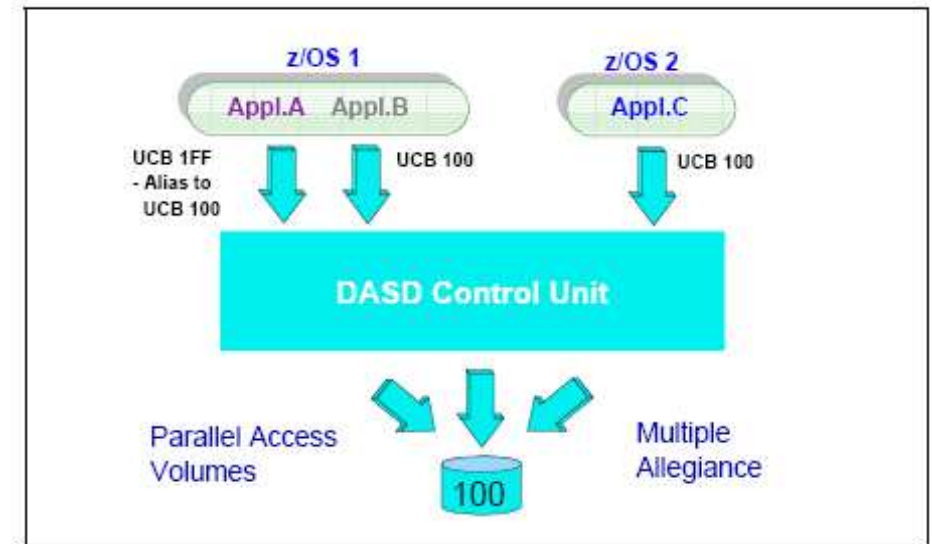
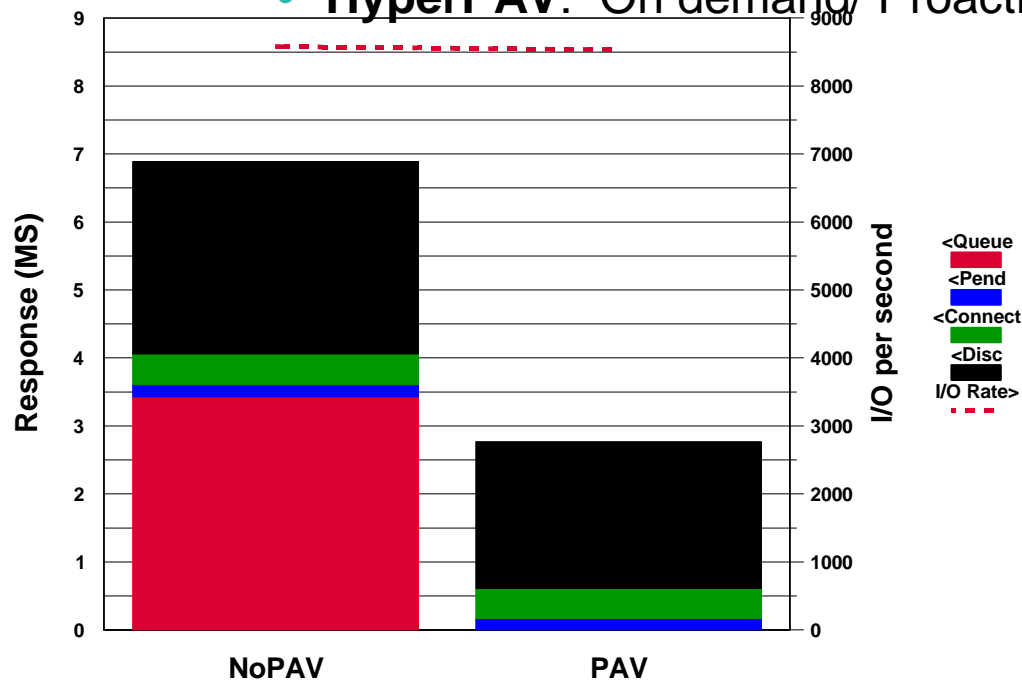


**Useful for Data Sharing!**

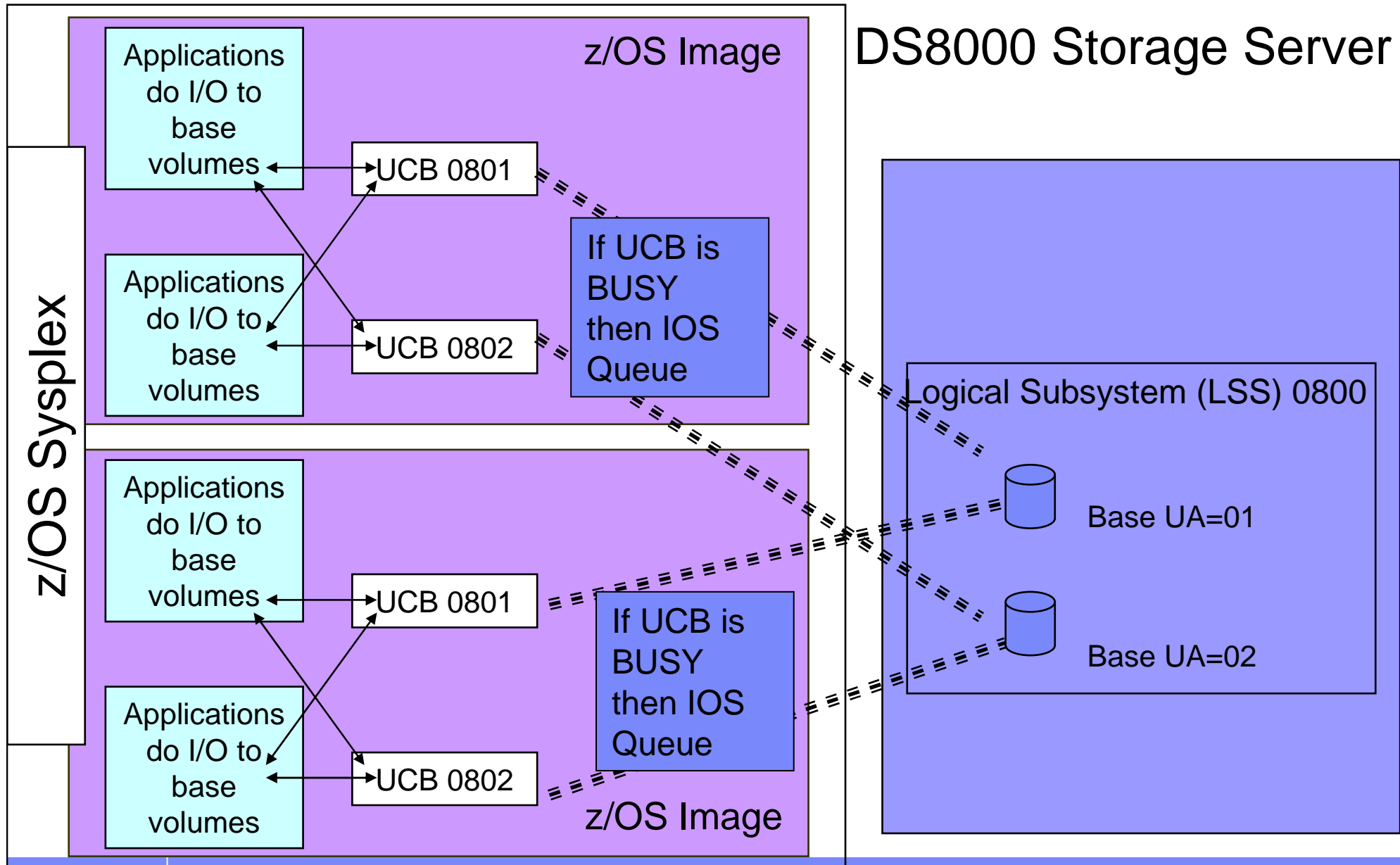


# PAV and Multiple Allegiance

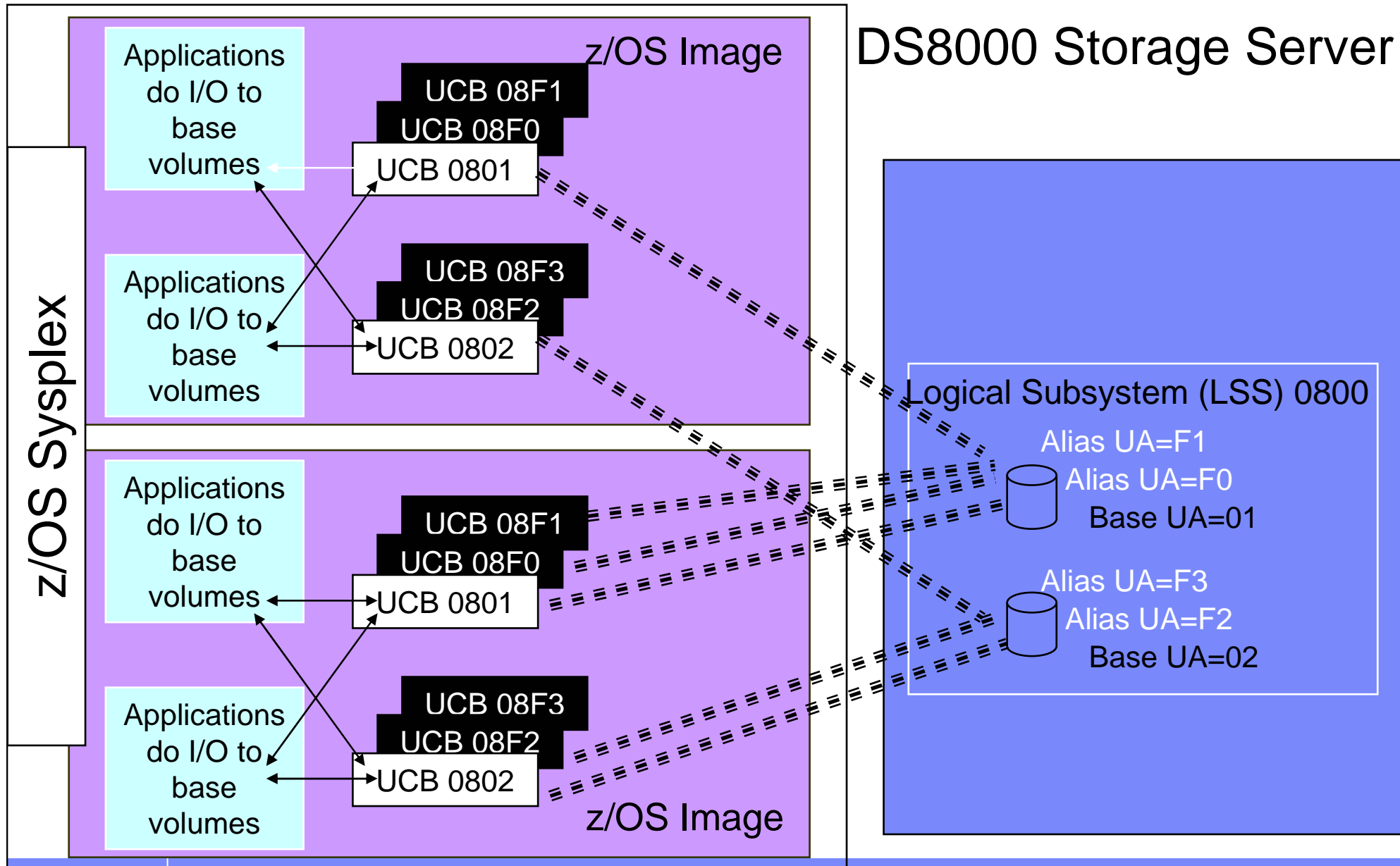
- Multiple Allegiance and PAV functions allow multiple I/Os to be executed concurrently against the same volume in a z/OS environment
  - With Multiple Allegiance, the I/O are coming from different LPAR of z/OS Sysplex
  - With Parallel Access Volumes, the I/O are coming from the same LPAR of z/OS systems
    - **Static PAV:** Aliases are always associated with the same Base Address
    - **Dynamic PAV:** Aliases are assigned up front but can be reassigned to any base address as need dictates: WLM function call Dynamic Alias Management - reactive alias assignment
    - **HyperPAV:** On demand/ Proactive alias assignment



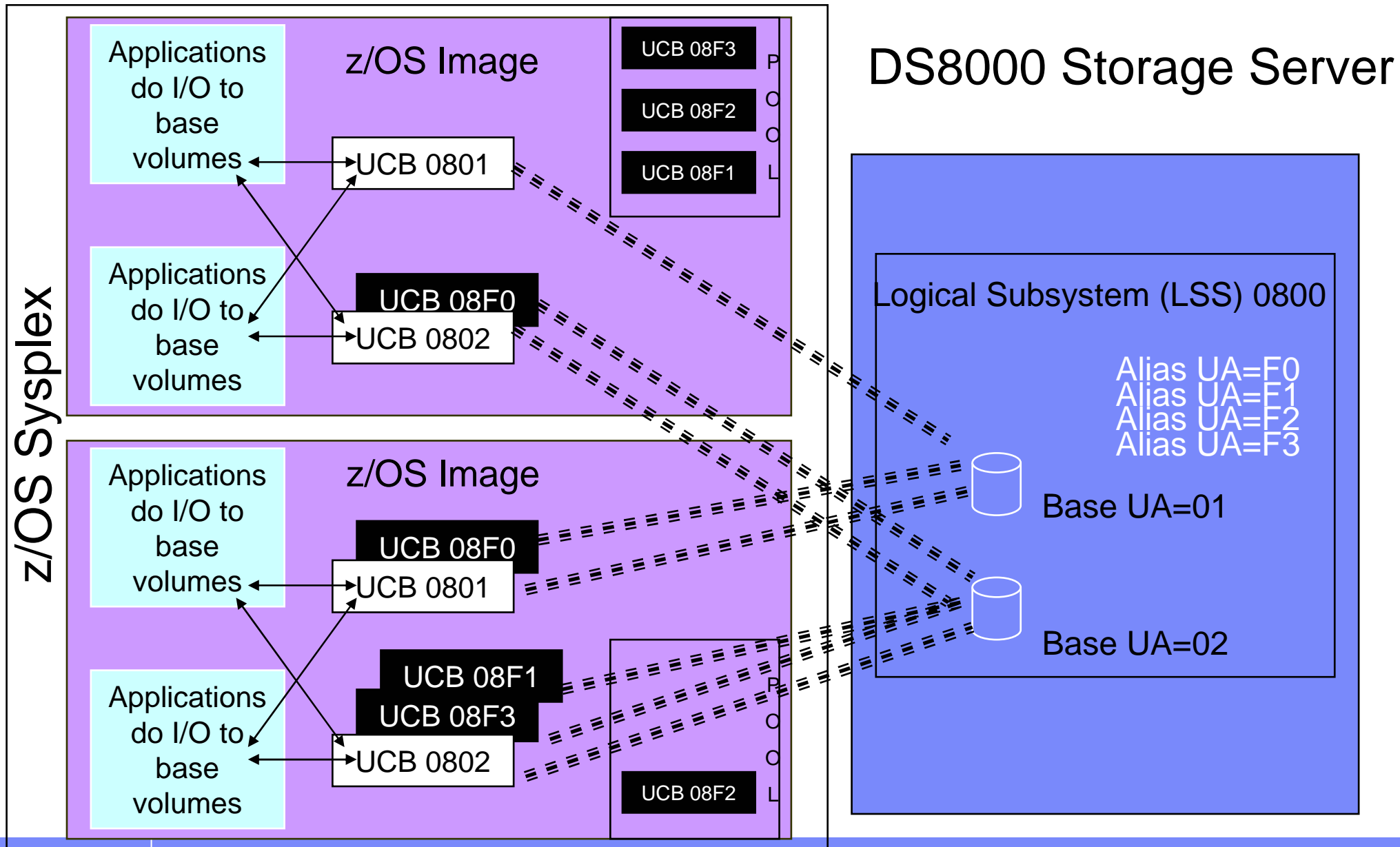
# IO without PAV



# Parallel Access Volumes - Dynamic



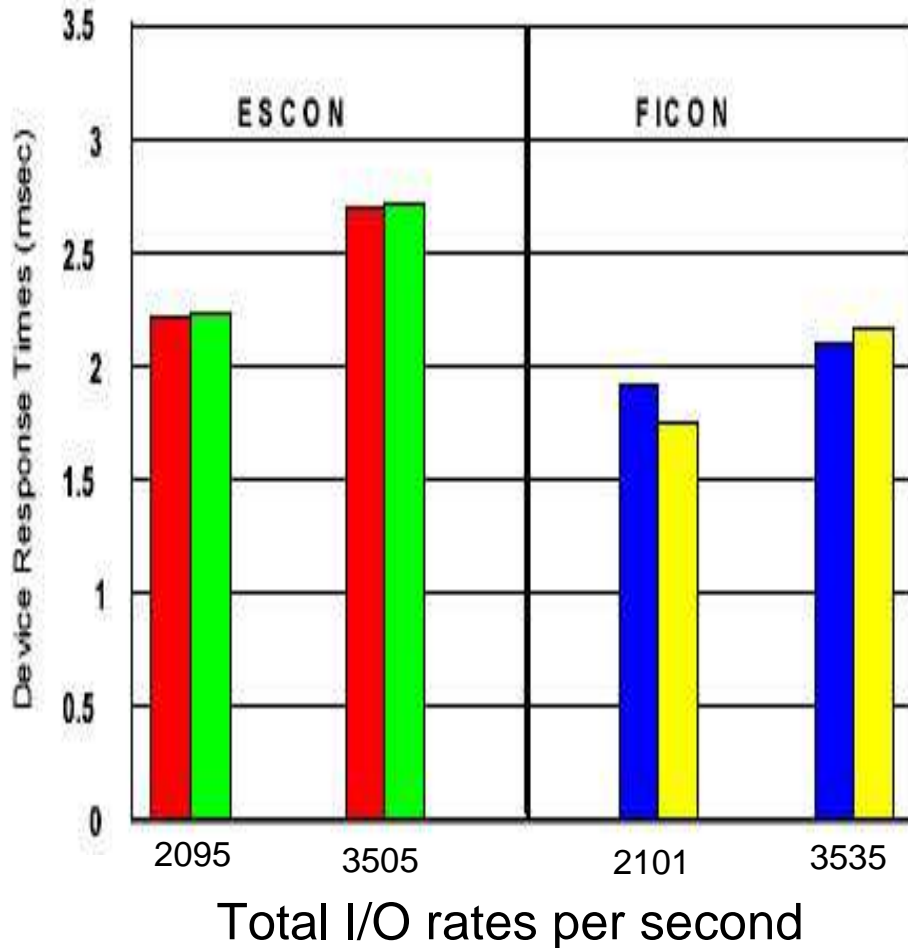
# HyperPAV



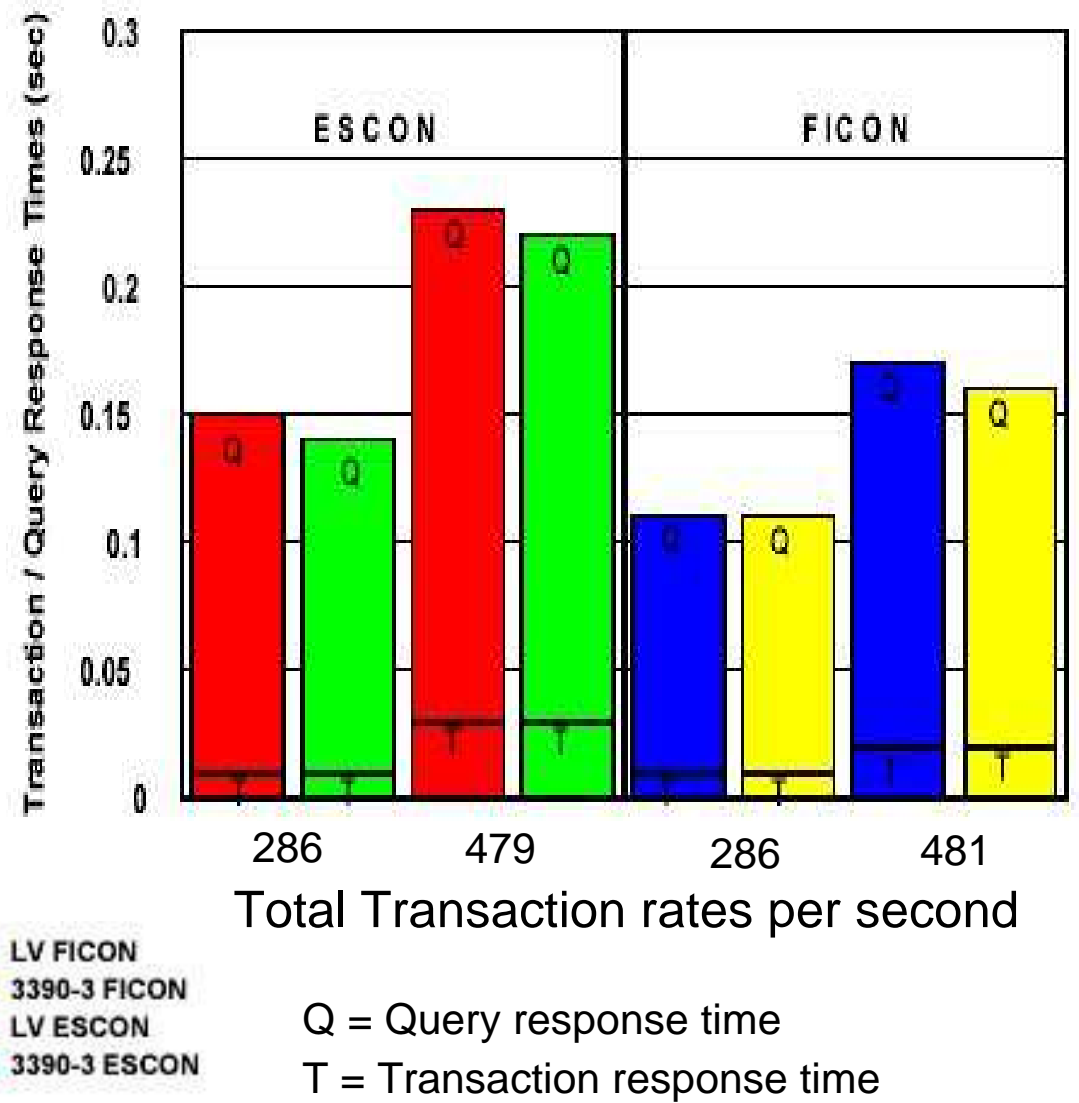
## Benefits of HyperPAV

- **Reduce number of required aliases**
  - Give back addressable device numbers
  - Use additional addresses to
    - support more base addresses
    - larger capacity devices.
- **z/OS can react more quickly to I/O loads**
  - React instantaneously to situations like ‘market open’ conditions
- **Overhead of managing alias exposures reduced**
  - WLM not involved in measuring and moving aliases
  - Alias moves not coordinated throughout Sysplex
- **Initialization doesn’t require “static” bindings**
  - Static bindings not required after swaps
- **IO reduction, no longer need to BIND/UNBIND to manage HyperPAV aliases in the DS8000**
- **Increases I/O Parallelism**

### DB2 Device Response Times 60 3390-3 vs. 6 Large Volumes (mod 27)



### DB2 Transaction/Query Response Times 60 3390-3 vs. 6 Large Volumes (mod 27)



Tests run using Dynamic PAV

- LV FICON
- 3390-3 FICON
- LV ESCON
- 3390-3 ESCON

Q = Query response time  
T = Transaction response time

## VSAM Data Striping - ESS, DS8000, and some RVA

- **Mainframes have RAID 5, 6, and/or 10 devices. All of these RAID types single stripe data. VSAM and sequential striping are used for multi striping data.**
  - **LISTC only displays stripe related information when multi-striped**
- **Spreads the data set among multiple stripes on multiple control units (this is the difference from hardware striping which is within the same disk array)**
- **An equal amount of space is allocated for each stripe**
- **Striped VSAM data sets are in extended format (EF) and internally organized so that control intervals (CIs) are distributed across a group of disk volumes or stripes.**
  - **DB2 V8 now allows striping for all page types, while V7 only allows striping for 4K pages. See APAR PQ53571 For issues with objects greater than 4K pages prior to DB2 V8.**
- **Greater rate for sequential I/O**
- **Recommended for DB2 active log data sets**
- **I/O striping is beneficial for those cases where parallel partition I/O is not possible. For example: segmented tablespace, work file, non partitioning indexes (NPIs, NPSIs), DPSIs, LOBs, etc.**

## Sequential Data Striping

### ■ Recommendations –

– Stripe all possible objects that are input or output for utilities when the associated table space is striped. For example (when the associated table space is striped):

- Stripe the image copy data sets
- Stripe input files for LOADs, etc.

### ■ **Only disk data sets can be striped.**

### ■ **You can now stripe your disk archive log data sets beginning in DB2 9.**

– You can now also compress archive log data sets providing a considerable savings in disk space.



## FAQ Checkpoint



Is this really a DB2 problem or a disk I/O problem?

**■ I am a DB2 professional, why should I care about RMF reports, PAVs, MAs, etc.? Doesn't my Storage Administrator deal with all of that?**

- You may be partially correct. However, your DB2 data resides on disk. If you do not receive data in a timely manner, then it becomes your problem. All the DB2 EXPLAINS in the world will not clue you into disk related problems.
- Periodically installations will review RMF data for disk performance. Discuss with your Storage Administrator and/or performance team (if one exists) how frequent periodic really is. You may find the answer to be daily, quarterly, or never. Never is not a good idea, quarterly may not be frequent enough.
- Does your Storage Administrator and/or performance team know your concerns? They may be tracking 5,000 disk volumes, in which case YOUR volumes may not be their top priority, although it is yours. Is your installation reviewing volumes only above specific disk thresholds?

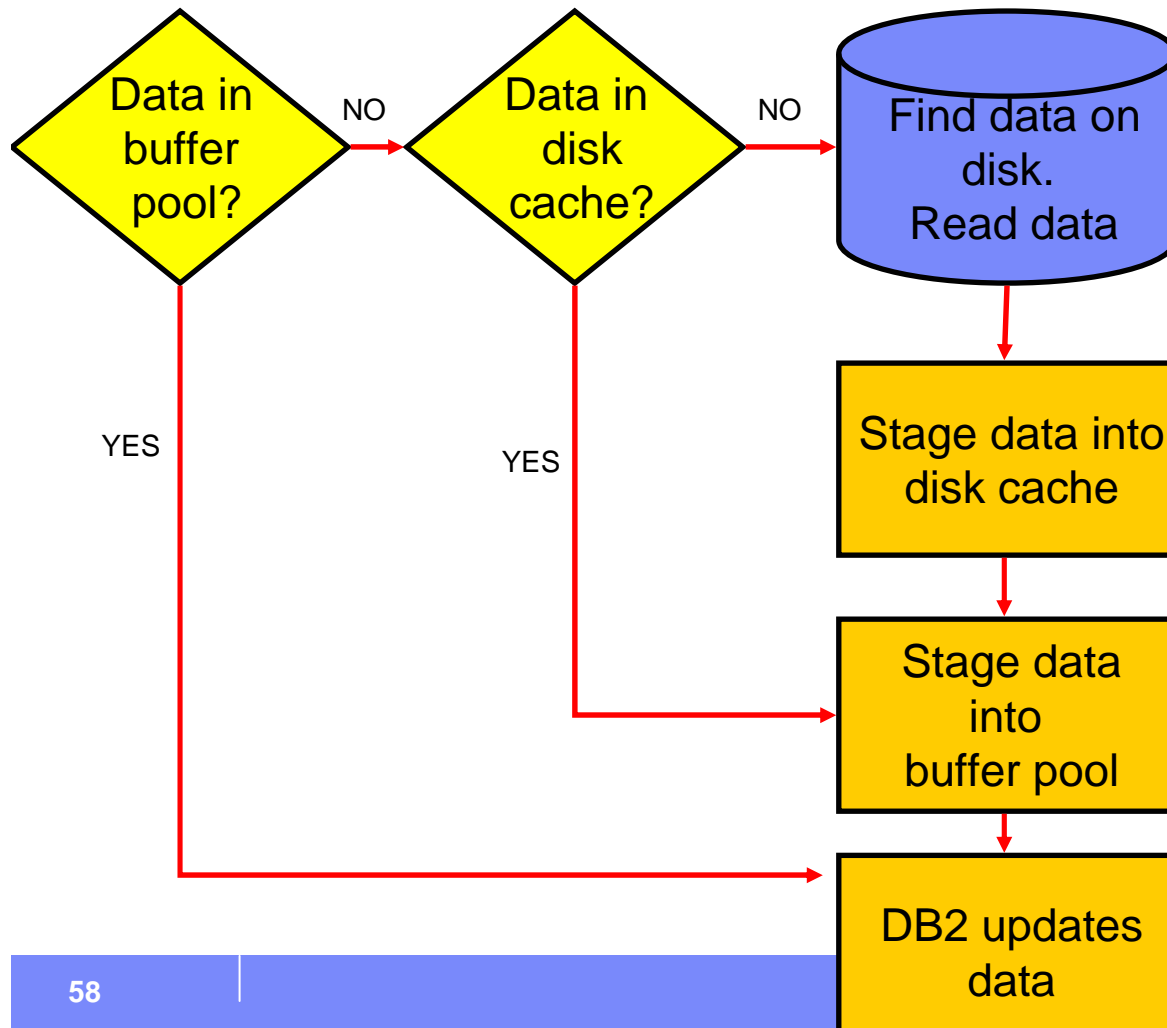
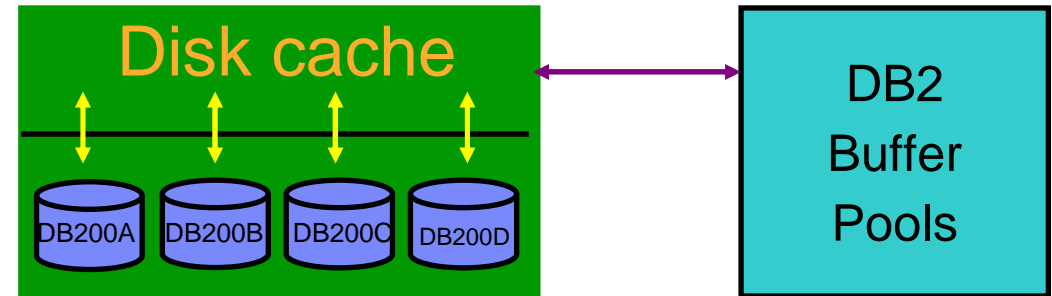


## Disk Cache - Why Storage Administrators think we are unfriendly! page 1 of 2

Illustration does not include Data Sharing with GBP dependent data, disk perspective is still the same.

- at a very high level -

```
UPDATE IBM.JOHN
SET SALARY = SALARY + 5000
WHERE NAME='JOHN';
```



- Based on DB2 checkpoint or buffer pool thresholds:

- DB2 destages updated data back down to cache.
- DB2 sees this as a disk write, even though the write is to cache.
- Data at this time may or may not remain in the buffer pool depending on why data was destaged.
- Data is also written to the NVS (Non Volatile Storage) part of the disk controller that is battery backed. If the controller crashes, data is not lost.

- Based on disk thresholds, cache destages data back down to disk.

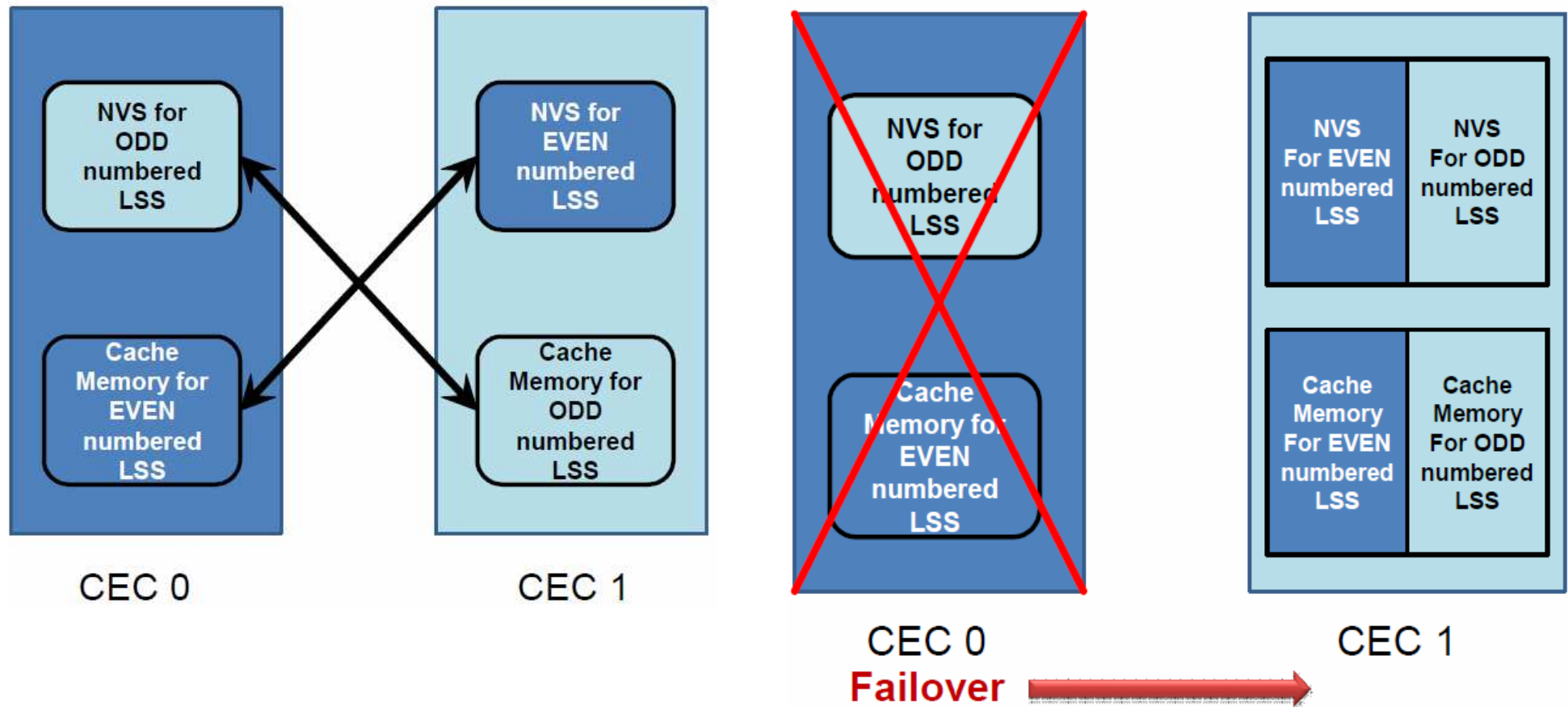
## Disk Cache - Why Storage Administrators think we are unfriendly! page 2 of 2

- **SELECT \* FROM IBM.JOHN;**
  - Read only operations do not require data to be destaged to cache.
  
- **From a conceptual point, the function of cache and buffer pools are similar.**
  - From a storage perspective, DB2 data is typically considered “unfriendly” because of the relatively low reuse of data in cache.
  - DB2 will use the data residing in the buffer pool when available. It may not require the data in disk cache at all.
  
- **Read from cache is exponentially faster than from disk.**
  - No need to go to disk, find the data, and bring it back through cache.
  
- **Just because your buffer pool casts out data, it does not mean that it is no longer retained in cache.**
  - Newer disk controllers have very large cache sizes and can retain data for longer periods.
  - With DB2 V8 and above, you can allocate very large buffer pools, as long as they are backed by real storage.

## Data IBM disk controllers bring into cache:

- **Data access is determined through the controllers adaptive cache.**
- **Random read** –
  - record or block staging (page)
  - partial track staging (from the page requested until the end of the track)
  - full track staging
  - Adaptive cache will determine which of the three will be used based on previous use of data whereby anything from as small as one page, or as large as from the page until the end of the cylinder plus the next cylinder will be read in.
- **Sequential prefetch** – full track staging – prefetch from the page up to the end of the cylinder plus the next cylinder (extent boundary). Actual operation is done on extent boundaries or stage group (an internal construct, depends on rank dimensions).
- **See ZPARM section for SEQCACH.**

# Write data when CECs are dual operational and failover



## CCHH (Cylinder Head) for newer disk

Messages will still show CCHH information for newer disk (RVA, ESS, DS8000). For example:



### LISTCAT output

EXTENTS:

LOW-CCHH-----X'079D0000'

HIGH-CCHH-----X'07AC000E'

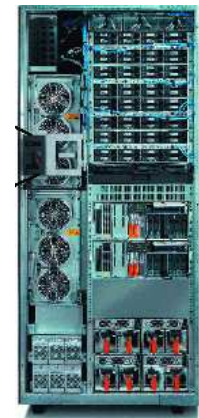


### IEHLIST utility with FORMAT option

EXTENTS	NO	LOW(C-H)	HIGH(C-H)	NO	LOW(C-H)	HIGH(C-H)
	0	0 1	0 1	1	0 12	0 14

DSNU538I RECOVER ERROR RANGE OF DSN=dataset name ON  
VOLUME=volser FROM CCHH=X'cccchhhh' TO CCHH=X'cccchhhh'  
CONTAINS PHYSICAL ERROR

- The disk controller itself will track allocations between the VTOC and where data really resides. Data sets no longer really reside at the CCHH reported.



## Volume Fragmentation - page 1 of 3

- **Because of the nature of allocation algorithms as well as the frequent creation, extension, and deletion of data sets, the free space on disk volumes become fragmented. This results in:**
  - Inefficient use of disk space
  - An increase in space-related abends
  - Performance degradation caused by excessive disk arm movement
  - An increase in the time required for functions that are related to direct access device space management (DADSM)
- **RVA, ESS, and DS8000 no longer require volume DEFRAgs!?**
  - “The box itself works differently; the box needs to be DEFRAged very infrequently (RVA). Individual volumes no longer require DEFRAgs.” This is not the case.
  - New disk technology is still tied to the old flavor of VTOC.
- **Periodic DEFRAgs on highly fragmented volumes are still recommended.**

# Volume Fragmentation - page 2 of 3

## ISMF Volume option

VOLUME	FREE	%	ALLOC	FRAG	LARGEST	FREE
SERIAL	SPACE	FREE	SPACE	INDEX	EXTENT	EXTENTS
-(2)--	---(3)---	(4)-	---(5)---	-(6)-	---(7)---	--(8)--
PTS005	370529	13	2400971	508	24182	180
PTS002	414522	15	2356978	537	23241	121

```
// DSN=TABLE.SPACE.IC,
  DISP=(NEW,CATLG),
  VOL=SER=PTS002,
  SPACE=(CYL,(250,25))
```

will this allocation work?  
Find out later in this presentation!

## ISPF 3.4 Display VTOC Information for volume PTS005

Volume . . : PTS005

Unit . . : 3390

Volume Data	VTOC Data	Free Space	Tracks	Cyls
Tracks . . : 50,085	Tracks . . :	75 Size . . :	6,696	393
%Used . . : 86	%Used . . :	22 Largest . . :	437	28
Trks/Cyls: 15	Free DSCBS: 2,962	Free		
		Extents . . :	180	

## ISPF 3.4 Display VTOC Information for volume PTS002

Volume . . : PTS002

Unit . . : 3390

Volume Data	VTOC Data	Free Space	Tracks	Cyls
Tracks . . : 50,085	Tracks . . :	75 Size . . :	7,491	462
%Used . . : 85	%Used . . :	30 Largest . . :	420	27
Trks/Cyls: 15	Free DSCBS: 2,641	Free		
		Extents . . :	121	

Note the low amount of largest free space. There are extent issues for large data sets!





## ▪ **Solution - run DEFRAG**

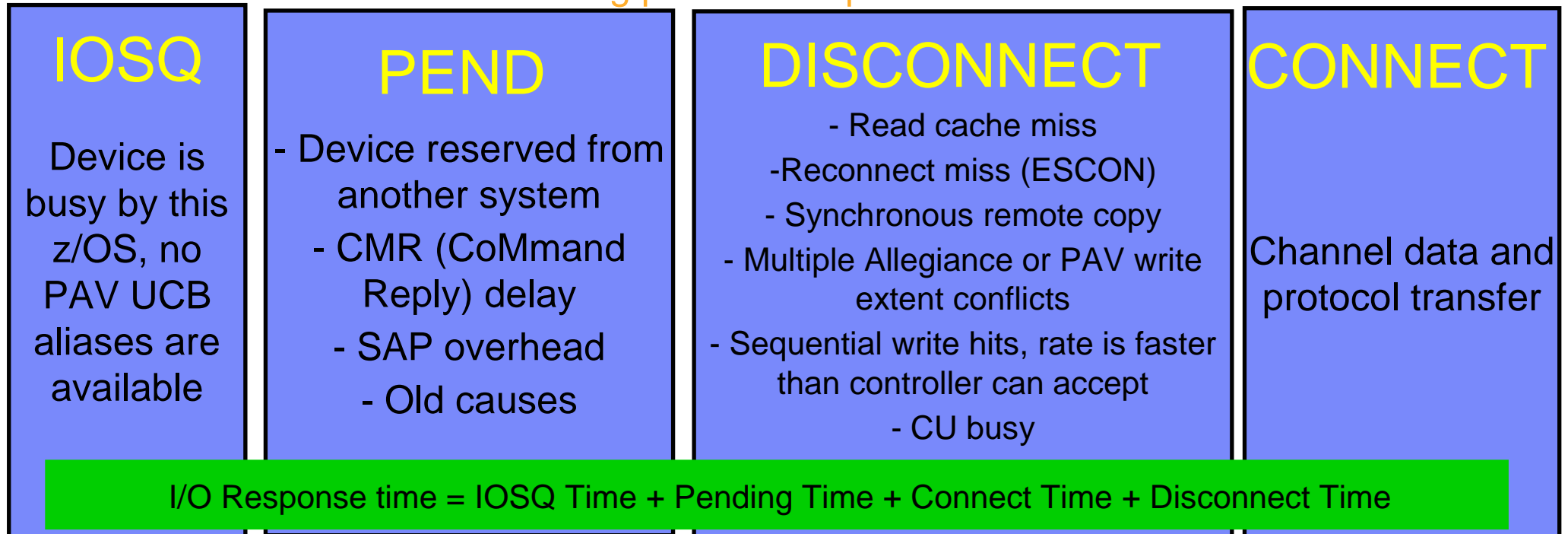
- Consolidates free space on volumes
- Relocates data set extents on a disk volume to reduce or eliminate free space fragmentation
- Storage Administrators can use data set FlashCopy V2 for ESS and DS8000 devices or SnapShot for RVAs to provide much faster DEFRAG operations.
- Recommendation - Your Storage Administrator may want to use the FRAGMENTATIONIndex keyword. Start the value for FRAGI at 250 and determine if it needs to go lower. FRAGI(250) will only DEFRAG the volume if the fragmentation index is 250 or above. Larger emulated disk may tolerate a greater fragmentation index.
  - Ignore volumes that have high fragmentation indexes, but are full, or close to full and have a small amount of free extents.

## ▪ **Drawback:**

- DEFRAG processing locks the VTOC (through the RESERVE macro) and the VVDS. The DEFRAG function also serializes on data sets through ENQ or dynamic allocation. What effect will this have on your DB2 data?
- BEWARE! In order to run a DEFRAG, the volume must be offline to all LPARs except for the one running the DEFRAG. The only alternative for a 7x24 installation is to not execute DEFRAGs and add volumes as needed.
- The CONSOLIDATE keyword attempts to consolidate data set extents and perform extent reduction for data sets that occupy multiple extents. Discuss this with your Storage Administrator! More information later in this presentation.
  - CONSOLIDATE can also be used as a separate command.

# What Storage Administrators look for in RMF DASD Reports - Four Stages of an I/O Operation

What is causing poor I/O response times?



**IOSQ:** Time waiting for the device availability in the z/OS operating system.

**Pending:** Time from the SSCH instruction (issued by z/OS) till the starting of the dialog between the channel and the I/O controller.

**Disconnect:** Time that the I/O operation already started but the channel and I/O controller are not in a dialog.

**Connect:** Time when the channel is transferring data from or to the controller cache or exchanging control information with the controller about one I/O operation.

# RMF Report - What Storage Administrators Look For

RMF - DEV Device Activity

14:48:42	I=35%	DEV				ACTV	RESP	IOSQ	-DELAY-	PEND	DISC	CONN	%D	%D	
STG	GRP	VOLSER	NUM	PAV	LCU	RATE	TIME	TIME	CMR	DB	TIME	TIME	TIME	UT	RV
SGDB2		DB200A	9034	4	0018	22.058	24.5	0.6	0.5	0.0	2.2	15.9	6.0	3.29	12.04
CB390		CBSM03	9035	1	0018	0.023	0.7	0.0	0.0	0.0	0.1	0.1	0.5	0	0
		WAS600	901F	4*	0018	17.12	14.2	0.0	0.0	0.0	0.2	0.0	14.0	24	0

•**Problem:** High disconnect time – disk **hot spots** (in this scenario)

•**Investigate:** Did your Storage Administrator front load the disk box? In other words, were your many disk volumes assigned in device order, all within an LSS? This is a relatively common technique.

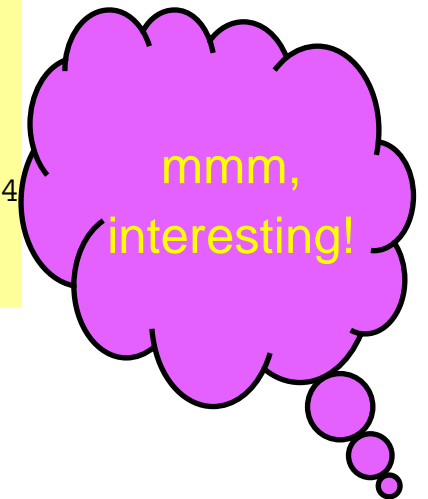
•**Solution:** Swap more active DB2 volumes with less active non DB2 volumes in another LSS. Other solutions are possible.

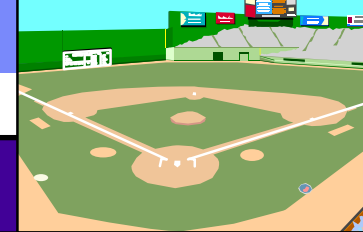
-This applies to newer DASD types as well, but less of an issue with the DS8000.

-Just one case where you can dramatically increase DB2 performance without doing anything with DB2.

LSS (Logical SubSystem) - controls a set of devices. Disk controllers contain one or more LSSes.

Front loading should be less of an issue when using storage pool striping as opposed to rotate Volumes or extents. Device Activity report and ESS (Enterprise Disk Systems report) are still Important for extent pools



**Ballpark I/O time per page - are you hitting home runs?**

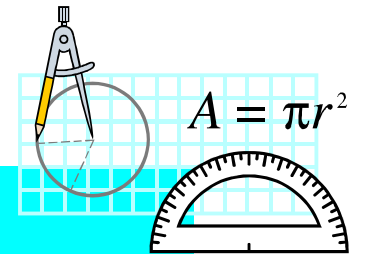
	Sequential Read or Write		Random Read	
	4K page	32K page	4K page	32K page
3390, RAMAC1, RAMAC2	1.6 to 2 ms	14 ms	20 ms	30 ms
RAMAC3, RVA2	0.6 to 0.9 ms	6 ms	20 ms	30 ms
ESS E20/ESCON	0.3 to 0.4 ms	3 ms	10 ms	15 ms
ESS F20/ESCON	0.25 to 0.35 ms	2 ms	10 ms	15 ms
ESS F20/FICON	0.13 to 0.2 ms	1.5 ms	10 ms	15 ms
ESS 800/FICON	0.09 to 0.12 ms	0.5 to 1 ms	5 to 10 ms	10 to 15 ms
DS8000	0.035 to 0.06 ms	0.3 to 0.5 ms	5 to 10 ms	5.5 to 11 ms

Sequential based on prefetch whereby one I/O can read in multiple pages. Pages are contiguous, no skip sequential.

Sequential numbers are also based on pages being in cache while random reads from disk

## Performance Notes

- For 8K and 16K page, interpolate from 4K and 32K page numbers.
- For skip sequential read or write, i.e. reading or writing of pages which are not contiguous, the time would be somewhere between sequential and random and depends on the distance between pages read or written.
- I/O time for sequential read or write is prorated on a per page basis, since multiple pages can be read or written by one Start I/O.
- Read I/O time tends to be faster than write I/O; eg use 1.6 to 1.8ms seq read and 1.8 to 2 ms seq write for 3390.
- Random read I/O time would go down if cache hit.



### Formula:

We obtained the time per page by dividing an average wait time for one prefetch or deferred write I/O from accounting class 3 by the number of pages read or written. "Sequential read or write" column represents the case in which all pages read or written are contiguous, not skip sequential.

## How fast is 0.035 or 0.3 ms (milliseconds)?

- **Order of time magnitude:**
  - 1 second  $10^0$
  - 1 millisecond (ms)  $10^{-3}$  is a thousandth (1/1,000) of a second
  - 1 microsecond  $10^{-6}$  ( $\mu$ s or usec) is 1/1000 millisecond, one millionth of a second
- **0.035ms = 35 usec, 0.3ms=300 usec**
- **Ballpark transfer rate for a 4k page based on prefetch from disk cache to a buffer pool is 35 usec.**

# Extent Consolidation

- **Consolidates adjacent extents for DB2 LDS - VSAM data sets when extending on the same volume.**
- **Automatic and requires no action on your part**
- **If the extents are adjacent, the new extent is incorporated into the previous extent.**
  - 1<sup>st</sup> extent tracks (seen on ISPF 3.4 Data set level listing) will show allocations more than the primary when extents are adjacent to the primary (see example in two pages).
- **LDS must be SMS managed.**
- **There is a dfp limit of 123 extents per volume. Extent consolidation continues even when an LDS exceeds what would have been 123 extents. For example, my allocation is PRIQTY 720 SECQTY 720 which is 1 cylinder primary and secondary and we have an open volumes so that we do not run into another data set, if you insert 190 cylinders worth of data, you will not stop at 123 cylinders for the volume, rather allocate the 190 cylinders and have 1 extent.**
- **When extent consolidation is in effect in z/OS V1.5 the secondary space allocation quantity can be smaller than specified or calculated by the sliding scale based on the physical extent number. PTF UQ89517 for APAR PQ88665 corrects this problem by using the logical extent number together with the sliding scale.**

# Extent Consolidation on Empty Volumes

Extent 1  
500 cylinders

Case: 3390 mod 3 with 3,330 cylinders of free space, all in one chunk of space. The VTOC, VTOC index and VVDS have been allocated and are now all at the beginning of the volume with no free space between them, leaving the rest of the volume totally empty. Allocation is PRIQTY 360000 SECQTY 7200, which is equivalent to CYL(500,10). How will extent consolidation work in this case based on inserting rows and increasing the space required?

Extent 1  
500 cylinders

Extent 2  
10 cylinders

Extent 1  
510 cylinders

Extent 1  
510 cylinders

Extent 2  
10 cylinders

Extent 1  
520 cylinders

much later ...

Extent 1  
2900 cylinders

Extent 2  
10 cylinders

Extent 1  
2910 cylinders

In our case this segmented table space has a DB2 limitation of 2GB, which is a little bit more than 2,900 cylinders. The end result of going to the 2GB limit is the data set is allocated at over 2,900 cylinders with only 1 extent.



Allocate 10 cylinder primary for VSAM object DSNB, due to volume

fragmentation largest cylinder chunks available are 4,3,3. What happens to extents:

### Prior to allocation:

Cylinder 15 head 0  
– cylinder 18 head  
14 (4 cylinders)  
**extent**  
*Free space*

Cylinder 19 head 0  
– cylinder 21 head  
14 (3 cylinders)  
**extent**  
*Free space*

Cylinder 22 head 0  
– cylinder 30 head  
14 (9 cylinders)  
**extent**  
DSNA

Cylinder 31 head 0  
– cylinder 33 head  
14 (3 cylinders)  
**extent**  
*Free space*

### Non SMS post allocation:

Cylinder 15 head 0  
– cylinder 18 head  
14 (4 cylinders)  
**extent**  
DSNB

Cylinder 19 head 0  
– cylinder 21 head  
14 (3 cylinders)  
**extent**  
DSNB

Cylinder 22 head 0  
– cylinder 30 head  
14 (9 cylinders)  
**extent**  
DSNA

Cylinder 31 head 0  
– cylinder 33 head  
14 (3 cylinders)  
**extent**  
DSNB

### SMS z/OS 1.5 and after, post allocation:

Cylinder 15 head 0  
– cylinder 18 head  
14 (4 cylinders)  
**extent 1 for DSNB**  
DSNB

Cylinder 19 head 0  
– cylinder 21 head  
14 (3 cylinders)  
DSNB

Cylinder 22 head 0  
– cylinder 30 head  
14 (9 cylinders)  
**extent**  
DSNA

Cylinder 31 head 0  
– cylinder 33 head  
14 (3 cylinders)  
**extent 2 for DSNB**  
DSNB

# Example – extend a data set until you must jump over another data sets extent

```
CREATE TABLESPACE JOHN
PRIQTY 48 SECQTY 48
1st extent tracks . : 1
Secondary tracks . : 0
Current Allocation
  Allocated tracks . : 1
  Allocated extents . : 1
```

```
CREATE TABLE JOHN
(ISPF 3.4 Data set listing)
1st extent tracks . : 4
Secondary tracks . : 0
Current Allocation
  Allocated tracks . : 4
  Allocated extents . : 1
```

```
Insert 1,500 rows into JOHN
(ISPF 3.4 Data set listing)
1st extent tracks . : 6
Secondary tracks . : 0
Current Allocation
  Allocated tracks . : 6
  Allocated extents . : 1
```

```
After many more inserts into JOHN
1st extent tracks . : 26
Secondary tracks . : 0
Current Allocation
  Allocated tracks . : 26
  Allocated extents . : 1
```

```
Many more inserts into JOHN and
extend onto new volume
1st extent tracks . : 28
Secondary tracks . : 0
Current Allocation
  Allocated tracks . : 42
  Allocated extents . : 2
```

**LISTC showed throughout:**

```
SPACE-TYPE-----TRACK
SPACE-PRI-----1
SPACE-SEC-----1
TRACKS/CA-----1
```

```
EXTENTS :
TRACKS-----28
TRACKS-----14
```

**28 tracks – 1 extent**

**14 tracks – 1 extent**

```
IEHLIST LISTVTOC FORMAT
```

EXTENTS	NO	LOW(C-H)	HIGH(C-H)	NO	LOW(C-H)	HIGH(C-H)
	0	17	2	18	14	
	1	519	0	519	13	

•PRIQTY and SECQTY are based on KB, so we can specify the following -

(allocations are slightly higher when DSVCI=YES, and you are allocating a 32K page):

- 1 track = 48 (see previous page)
- 1 cylinder =720 (15 tracks \* 48 KB per track) <=above 672 is allocated in cylinders due to CA
- If PRIQTY>1 cylinder and SECQTY<1 cylinder, secondary rounded up to 1 cylinder (CA 1 cyl)
- If PRIQTY<1 cylinder and SECQTY>1 cylinder, allocations in tracks, not cylinders (CA<1 cyl)
- See PK05644 – preformat up to 2 cylinders at a time regardless of track or cylinder allocation
- Preformat starting in DB2 9 is up to 16 cylinders at a time, up from 2 cylinders
- Conversions (4K tables, will vary when DSVCI=YES and you are allocating CI greater than 4K, see next page) :
  - PRIQTY to the number of cylinders=PRIQTY/720
  - PRIQTY to the number of pages = PRIQTY/4
  - Number of pages to cylinders = pages/180
  - Number of pages to PRIQTY=pages\*4

## Disk Space Allocations and Extents

Maximum disk volumes a data set can reside on - 59 (dfp limitation)

- Maximum number of volumes in a DB2 STOGROUP - 133 (DB2 limitation)
- Maximum size of a VSAM data set is 4GB unless it is defined with a data class that specifies extended format with extended addressability (dfp limitation). However:
  - Maximum simple or segmented data set size - 2 GB
  - Largest simple or segmented table space - 64 GB (32 data sets \* 2 GB size limit)
- Maximum extents for simple or segmented table space - 8160 (32 data sets \* 255 extents per data set) – see next page for z/OS 1.7 extent change
- Maximum size of a partition created with LARGE keyword - 4 GB
- Maximum size of a partition created with DSSIZE keyword - 64 GB (depending on page size)
- Largest table or table space - 128 TB (32K partitioned table \* (32 GB\*4096 parts or 64 GB\*2048 parts)). Maximum size of a UTS is 128 TB when using 32K tables.

## PM42175 – DB2 V10 – DSSIZE support for 128G and 256G

- **Total size limitations based on page size remains the same**
- **With higher sizes, less partitions can be allocated**

## ▪Extents:

- Non-VSAM (e.g.: image copies), non-extended format data sets: up to 16 extents on each volume
- Non-VSAM (e.g.: image copies), extended format data sets, up to 123 extents per volume
- PDS (e.g.: DB2 libraries) up to 16 extents - one volume max
- PDSE (see FAQ section) up to 123 extents - one volume max
- VSAM data sets, up to 255 extents prior to z/OS 1.7 per component, but only up to 123 extents per volume per component (123 extents per volume in z/OS 1.7 as well). **Starting with z/OS 1.7 architectural new maximum extents for SMS managed VSAM objects (Data Class 'Extent Constraint Removal' must be set to YES):**
  - 123 extents per volume \* 59 volumes (both dfp limitations per component) = 7,257 extents
- Striped VSAM data sets, up to 4080 extents per data component (16 stripes (volumes) max for VSAM\*255 extents per stripe). With z/OS 1.7 the 255 extents maximum is removed for SMS managed components.

▪**Small track allocations for large amounts of data (with the equivalent of many cylinders worth of data) can result in insert performance degradation. APAR PK05644 allows DB2 to preformat 2 cylinders worth of data regardless if an object was created in tracks or cylinders. Preformatting of data is done by extent, however it can not exceed 2 cylinders in DB2 V8. There is a difference between allocating 1,000 tracks of data in 100 track chunks or 5 track chunks. Providing the PTF for APAR PK05664 has been applied, the 100 track allocation can preformat the equivalent of 2 cylinders at a time, while the 5 track allocation can only preformat 5 tracks at a time, thereby causing performance degradation.**

## ▪With DSVCI=YES

- splits 32 KB CI over two blocks of 16 KB in order not to waste space due to track usage (48 KB)
- A 16 KB block is not used every 15 tracks (CA size) because the CI cannot span a CA boundary

DB2 page size	VSAM CI size V7/V8	VSAM physical block size V7/V8	blocks per track V7/V8	DB2 pages per tracks
4	4	4	12	12
8	4/8	4/8	12/6	6
16	4/16	4/16	12/3	3
32	4/32	4/16	12/3	1.5



## APAR - PQ53571 (BELOW DB2 V8!)

00C20111

Explanation: An attempt was made to access a data set with a page size of 8K, 16K or 32K that is defined with multiple stripes. DB2 does not permit 8K, 16K or 32K page size data sets to be striped because partial writes cannot always be detected.

System Action: A 'resource not available' code is returned to the user. This reason code and the data set name are made available to the user in the SQLCA.

# Preformat – DB2 9 plus Examples



DB2 9 and later preformats **up to** 16 cylinders at a time (except HASH). Prior versions of DB2 preformatted up to 2 cylinders or 2 tracks depending on CA size and maintenance level. Preformat updates the VVDS.

```
CREATE TABLESPACE JOHNITS1
  USING STOGROUP SYSDEFLT
  PRIQTY          72000
  SECQTY          3600
  LOCKSIZE       ANY
  CLOSE          NO
  BUFFERPOOL     BP2
  FREEPAGE       0
  PCTFREE        0;
```

→ 100 cyl (48k per track \* 15 tracks per cyl)=720\*100  
→ 5 cyl (48k per track \* 15 tracks per cyl)=720\*5  
(keep in mind PRIQTY and SECQTY are specified in KB)

LISTCAT output:

```
ALLOCATION
SPACE-TYPE-----CYLINDER      HI-A-RBA-----73728000 → 100 cylinders=72000*1024(1kb)
SPACE-PRI-----100           HI-U-RBA-----11796480 → preformat 16 cylinders in kb
SPACE-SEC-----5
```

```
CREATE TABLESPACE JOHNITS1
  USING STOGROUP SYSDEFLT
  PRIQTY          3600
  SECQTY          720
  LOCKSIZE       ANY
  CLOSE          NO
  BUFFERPOOL     BP2
  FREEPAGE       0
  PCTFREE        0;
```

→ 5 cyl (48k per track \* 15 tracks per cyl)=720\*5  
→ 1 cyl (48k per track \* 15 tracks per cyl)=720\*1  
(keep in mind PRIQTY and SECQTY are specified in KB)

LISTCAT output:

```
ALLOCATION
SPACE-TYPE-----CYLINDER      HI-A-RBA-----3686400 → 5 cylinders=3600*1024(1kb)
SPACE-PRI-----5             HI-U-RBA-----3686400 → preformat 5 cylinders in kb
SPACE-SEC-----1
```



## DB2 9 – Indexes - Size and Compression

- **Starting with DB2 9, indexes can now be 4K, 8K, 16K, or 32K.**
- **Dictionaryless, software managed index compression at the page level.**
- **Indexes are compressed at write time, decompressed at read time. They are uncompressed in the buffer pools.**
- **Compression of indexes for BI workloads**
  - Indexes are often larger than tables in BI
- **Solution provides page-level compression**
  - Data is compressed to 4 KB pages on disk
  - 32/16/8 KB pages results in 8x/4x/2x disk savings
  - No compression dictionaries – compression on the fly
- **DSN1COMP can be used for indexes as well starting with DB2 9.**
- **Index compression is strictly for disk cost savings, not performance.**

## dfp Enhancements and DB2

- **As with DB2, dfp introduces enhancements on a periodic basis. These enhancements may effect data sets used by DB2, however, keep in mind that many times a DB2 APAR is required before DB2 can take advantage of the dfp enhancement.**
  - For example, the dfp enhancement to increase the VSAM data set maximum to 7,257 extents required some DB2 APARs in order to utilize the enhancement. Another example is when dfp introduced large sequential data sets in z/OS 1.7, DB2 could not take advantage of utilizing this enhancement for such things as archive log data sets and image copies until the introduction of DB2 9.
  - This is true for other dfp enhancements as well. For example, extended format sequential in z/OS 1.11 is not available for use for utilities with the current versions of DB2 (8 and 9).
- **Verify if any DB2 maintenance is required to take advantage of dfp enhancements.**
- **DB2 generally releases a new version every 3 years. MVS (z) on the other hand provides a release about once a year. There is 1 or several disk related updates once a year. Make sure you understand how these other newer releases relate to DB2.**

Simplified data allocation



Improved allocation control



Improved performance management



# DFSMS Basics for DB2 Professionals

Automated disk space management



Improved data availability management



Simplified data movement



# Managing Data With SMS

No SG info!



```
ISPF 3.4 Data Set listing:
General Data
Management class . . . : MCDB2
Storage class . . . . : SCDB2
Volume serial . . . . : DB2810
Device type . . . . . : 3390
Data class . . . . . : DCDB2
```

## ACS (Automatic Class Selection) Routines

What does it  
look like?

What are its  
requirements?

Where is  
it placed?

Data  
Class

Storage  
Class

Management  
Class

Storage  
Group

ACS routines invoked in order (DC,SC,MC,SG)

- Allows you to designate groups of data sets which are to be physically separated
- Volume separation is now available with z/OS 1.11
  - This new function may be useful when you have problems with hotspots. For example, you do not want partition 1 and 2 allocated on the same volume. **More important is to make sure dual software copied data sets such as the BSDS and active logs are allocated on different extent pools.**
  - SMS **attempts** to separate data sets listed in a separation group onto different extent pools (in the storage subsystem) and volumes.
    - **Separation Profile can be coded so that 2 data sets must never reside on the same logical volume, however it cannot guarantee they will reside in different extent pools.**
      - You can verify the extent pool a logical volume belongs to using the RMF Cache Subsystem Activity Report.
- For DB2, you may want to consider using this technique for the BSDS and active log data sets. This is an availability benefit for DB2.
  - My personal preference for the BSDS and active log data sets is to use the Storage Class Guaranteed Space attribute and hand place these data sets. The Separation Profile can be used along with Guaranteed Space to ensure data sets are allocated correctly.
- SMS attempts to allocate the data sets behind different physical control units and/or volumes. NOTE – control unit separation is different than the new volume separation profile.
- A separation profile data set must be provided

- A collection of allocation and space attributes - for new data sets only, e.g. - LRECL, BLKSIZE, space, etc.
  - Space specified by user can be overridden with Data Class space attributes.
- Can be used for SMS or non-SMS managed data sets. Some attributes such as EF or bypassing the 255 extent rule cannot be associated with non-SMS managed data sets. Attributes such as DCB and space can be associated with non-SMS managed data sets.
- For VSAM and sequential data sets
- Formats can be: EXTENDED Format (EF), Extended Addressable (EA), HFS, LIBRARY (PDSE), PDS, and LARGE format sequential
  - For DB2 partitions, LOBs, and XML greater than 4GB require EF/EA enabled.
  - When VSAM or sequential striping, EF must be enabled + Storage Class value for SDR
- Provides space constraint relief for VSAM and sequential data sets – avoids many X37 abends
  - SMS can be set up to allocate a percentage of the requested quantity if not enough space is available. This option is not available for data sets using Guaranteed Space.
  - It is possible to set up SMS whereby the 5 extent rule is bypassed for initial allocation, as well as when extending a data set to a new volume.
  - VSAM and non-VSAM multistriped data sets do not support space constraint relief. However, single-striped VSAM and non-VSAM data sets can use space constraint relief.
- Allows for data sets to be multi volumed - e.g. image copy data sets
- Allows certain VSAM data sets to exceed the 255 extent limit, see ‘Extent Constraint Removal’
- Compress (Compact) objects – archive log
- CA Reclaim (only useful for the ICF catalogs, not DB2 datasets as almost all are LDS, not KSDS)

# Data Class – VOLUME COUNT and DYNAMIC VC

- **Volume count and dynamic volume count are similar. They both add volumes at EOV. Volume count will add candidate volumes, dynamic volume count will not.**
  - VC shows candidate volume, DVC does not
- **Recommendation:**
  - For DB2 managed data sets, keep the default of 1 for volume count and do not have a value for dynamic volume count. This has to do with the way DB2 adds volumes at EOV.
  - For user managed data sets, such as the DB2 catalog and directory, keep the default of 1 for volume count and allow for a reasonable value for dynamic volume count. You may want to set VC=1, DVC=2 or 3. This would allow a DB2 catalog and directory data set to expand to the number of volumes set for DVC.
    - Using the VOLUMES parameter on DEFINE overrides VC and DVC
    - DEFINE using no Guaranteed Space, but with VOLUMES parameter specifying several volumes results in the allocation of the number of volumes requested, but only the first volume is allocated, the remaining volumes are candidate volumes with "\*" for the VOLSER. For example, GS=N, specify 6 volsers for VOLUMES, VC=3, DVC=5, allocation will result on 1 physical volume for the primary allocation, and 5 candidate volumes with "\*" as the volser.
    - DEFINE using Guaranteed Space with no VOLUMES parameter results in the allocation of volumes for VC and DVC. Volumes are allocated with no candidate volumes. For example, GS=Y, no VOLUMES, VC=3, DVC=5, allocation will result on 3 physical volumes, with the primary allocation value on all 3 volumes because of GS.
  - **NOTE** – when DB2 Catalog and Directory data sets are converted to DB2 managed in DB2 V10, VC and DVC are ignored even when specified. DB2 managed data sets use the normal EOV process as opposed to VC/DVC specifications.
    - Most of the time SMS rules apply, however there are certain circumstances such as above where DB2 code overrides SMS code.

## EF/EA Recommendation

- **EF (Extended Format) and EA (Extended Addressable) are primarily used in DB2 environments to be able to exceed the 4GB limit of LOBs, XML files, and partitions.**
- **Enabling EF/EA in the SMS Data Class is recommended for DB2 LDS with newer disk controllers and channels. This would provide DBAs the flexibility to assign objects requiring a large DSSIZE without the need to ask the Storage Administrators for a special change.**
- **If Storage Administrators do not want to assign EF/EA to all DB2 LDS, one Data Class can be assigned EF/EA and in DB2 9 DBAs can CREATE or ALTER a DB2 STOGROUP and specify the assigned Data Class. Objects requiring a larger DSSIZE can then be created using this special Data Class.**
- **NOTE - An EF data set requires a 32 byte suffix for every physical block. The number of physical blocks per CI depends on the CI size. If the CI size is such that the best use of DASD space is to create two physical blocks per CI (32K object), the 32 byte suffix is for every block. This does **NOT** add space for the data set – the additional bytes are generally hidden for DB2 objects.**



## Large Sequential Data Sets – see DB2 notes

- **Removes the size limit of 65535 tracks (4369 cylinders) per volume for sequential data sets**
  - BSAM, QSAM, and EXCP
    - Data sets do not have to be in SMS managed and in extended format
  - 16 extents is still the limit
  - Architectural limit is 16,777,215 tracks
  - JES2/JES3 spool can now be larger than 64K tracks
    - Must still be a single extent
- **Changed APIs supports all sequential and partitioned data sets**
- **DFSMSHsm support of migration/recall, backup/restore and ABACKUP/ARECOVER of large format data sets**
- **Addresses limits on capacity in customers' systems**
  - Direction is to support large capacity devices
  - Systems support up to 64,512 (63K) devices

# Specifying Large Sequential Data Sets

- **DSNTYPE DD statement, dynamic allocation text unit, TSO ALLOCATE keyword support 4 new values:**
  - LARGE
    - If the data set is sequential or DSORG is omitted, then large format data set
  - EXTREQ
    - If the data set is VSAM or sequential or DSORG is omitted, then extended format data set is required
  - EXTPREF
    - If the data set is VSAM or sequential or DSORG is omitted, then extended format data set is preferred
      - If not possible, then neither extended format, nor large format
  - BASIC
    - If the data set is VSAM or sequential or DSORG is omitted, then neither extended format nor large format
- **Data class can provide the new DSNTYPE**
  - New &DSNTYPE variable for ACS routines with values of either LARGE or BASIC

# DB2 9 and Large Sequential Data Sets

- **Can be non SMS managed.**
- **In DB2 9 NFM supported for:**
  - Utility input data sets.
  - Utility output data sets when coded in the SMS Data Class or DD includes DSNTYPE=LARGE. NOTE – at this time you cannot use DSNTYPE in your TEMPLATE statement, nor can you use a model DSCB that uses LARGE. You can however point to the TEMPLATE to a DATA CLASS with LARGE enabled.
  - Archive log data sets can be allocated on one disk as 4 GB -1 byte. Archive data sets can be created on disk only in NFM or ENFM modes, however, DB2 tolerates reading them in CM.
    - LARGE and EF are mutually exclusive. If you want to compress the archive log data sets you must specify EF which will exceed the 4,369 cylinder limit as well.
    - ZPARM values for the archive log Data Class do not exist. The Data Class ACS routine must associate the archive log with the correct Data Class.

## DB2 9 and Large Sequential Data Sets

- **Large sequential data sets are eligible for EAV devices. Do not confuse large sequential data sets with Extended format sequential data sets which are new types of data sets on the EAV introduced by z/OS 1.11.**

# Large Block Interface (LBI)



- **Large Block Interface allows block size > 32,760 bytes**
  - Up to 40% reduction elapsed time for COPY and RECOVER RESTORE phase
- **With LBI the maximum block size for most tapes increases from 32,760 to 256KB.**
- **LBI is supported on all tapes, DASD, dummy data sets, and z/OS UNIX files.**
  - Although LBI works with DASD, you cannot exceed a block size of 32,760
  - With LBI, IBM tape devices such as the VTS, TS7700, ATL, 3590, and 3592 can use a block size of 256KB.
  - Maximum block size limits are documented in the *DFSMS Using Data Sets* manual under “Optimum and Maximum Block Size Supported”
  - DB2 archive log data sets cannot exceed a block size of 28K, even of tape
- **No minimum level of z or tape microcode is required. The z/OS software asks the drive its maximum and optimum block size**
- **Block size can be set:**
  1. **BLKSZLIM** value in the DD statement or dynamic allocation.
  2. **ISMF Data Class** panel for option “Block Size Limit”
    - The Data Class ACS routine can automatically set allocation for a Data Class with LBI enabled, or can be specified by using the DATACLAS parameter on a DD statement if allowed by the Storage Administrator.
    - LBI enabled data sets can be SMS or non-SMS managed.
  3. **TAPEBLKSZLIM** value in the DEVSUP<sub>xx</sub> member of SYS1.PARMLIB. A system programmer sets this value, which is in the Data Facilities area (DFA) (see *z/OS DFSMSdfp Advanced Services*)
    - Only one option is required to use LBI. Any combination or all 3 options can be specified. JCL overrides the Data Class, and the Data Class overrides PARMLIB.
- **When migrating to DB2 V10:**
  - migrate from V8 -> LBI only allowed in V10 NFM, disallowed in CM8, ENFM8.
  - migrate from V9 -> LBI allowed in V9 NFM, CM9, ENFM9, and V10 NFM.

- **Separates data set performance objectives and availability from physical storage – it does not represent any physical storage, but rather provides the criteria that SMS uses in determining an appropriate location to place a data set.**
- **SMS determines at this time if a data set is SMS managed or not. Ones that are not do not continue down the ACS routines and stop here.**
- **Most Storage Class definitions are no longer generally required, such as MSR rate, accessibility, etc. Storage Class is used in DB2 environments for three general purposes:**
  - Guaranteed Space
  - SDR rate for VSAM and sequential striping
  - Enable multi-tiered Storage Groups
  - Direct allocations to SSD devices MSR=1, or HDD devices MSR=10 if both exist
- **Defining the Guaranteed Space Attribute (honor volser and space request)**
  - Not recommended for DB2 user data sets, especially when using PAV and MA (avoids **MOST** hot spots). DB2 BSDS and active log data sets can still be hand placed using guaranteed space for availability as opposed to performance, even when using PAV and MA. Also review the SMS data set separation profile with your Storage Administrator. Guaranteed Space and the Separation Profile can be used together.
  - When you allocate a data set with Guaranteed space and a data set becomes multi volumes, the extend to a new volume for the secondary extent will actually be what the primary for the initial extent, not the secondary. This is true for DB2 data sets that are user managed, however for DB2 managed data sets the extend to a new volume will be created on the secondary amount. Review APARS PK22108/PK50112 as well.

## Interesting Storage Class Information for DB2 Professionals - page 2 of 2

### ▪ **Other Specs**

- Allowance for multi-tiered SG (Storage Group) - see Storage Group section for additional information
- Determine if PAV volumes should be part of the Storage Group selection - see Storage Group section for additional information

### ▪ **Recommendations:**

- If your installation has a mix of all types, models, and configurable disk:
  - Determine if specific targeted response time rates are required, as well as such things as RAID, and guaranteed space.
  - For example: If your installation is all ESS and DS8000, you may not need to worry about these attributes for performance when using PAVs and MA. Place your BSDS and active log data sets on the fastest devices possible (DS8000), leaving your DB2 user data on the ESS.
  - Sit down with your Storage Administrator and determine the best use of data and technology. Do not assume that your Storage Administrator understands DB2 as well as you do.

## Interesting Management Class Information for DB2 Professionals - page 1 of 3

▪ **A management class is a list of data set migration, backup and retention attribute values, as well as expiration criteria which uses DFSMShsm and DFSMSdss. Some issues to consider are:**

- Requirements for releasing over-allocated space
- Migration requirements
- Retention criteria
- Treatment of expired data sets
- Frequency of backup
- Number of backup versions
- Retention of backup versions
- Number versions
- Retain only version
- Retain only version unit
- Retain extra versions
- Retain extra versions unit
- Copy serialization
- Generation data group (GDG) information

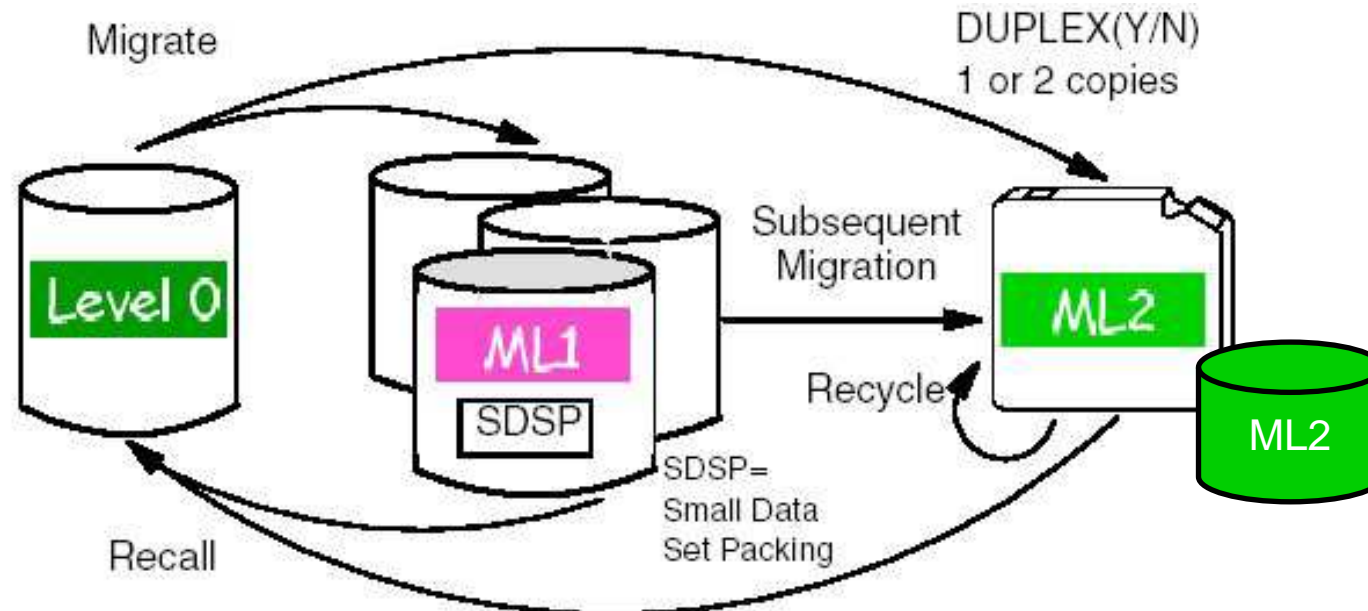


## Interesting Management Class Information for DB2 Professionals page 2 of 3

**Recommendations:**

- BEWARE! If your Storage Administrator has turned on the 'Partial Release Attribute', DFSMSHsm may compress any created VSAM data sets in extended format not using the guaranteed space attribute.
  - In a non-production environment releasing unused space in the DB2 LDSes can be of a great benefit.
- Discuss with your Storage Administrator the 'GDG Management Attributes' if you are creating GDGs for such things as image copies. Determine how many GDGs should be retained on disk.
- Keep in mind DR situations! ABARS may be used to backup your application's ML1 and ML2 data sets if required.
- Objects related to DB2 commonly migrated:
  - Archive log data sets
  - Image Copy data sets
  - Some customers migrate table spaces and indexes. Review ZPARM values for RECALL and RECALLD. Not recommended, especially in production.

# Interesting Management Class Information for DB2 Professionals - page 3 of 3



## Space Management Activities - DFSMSHsm

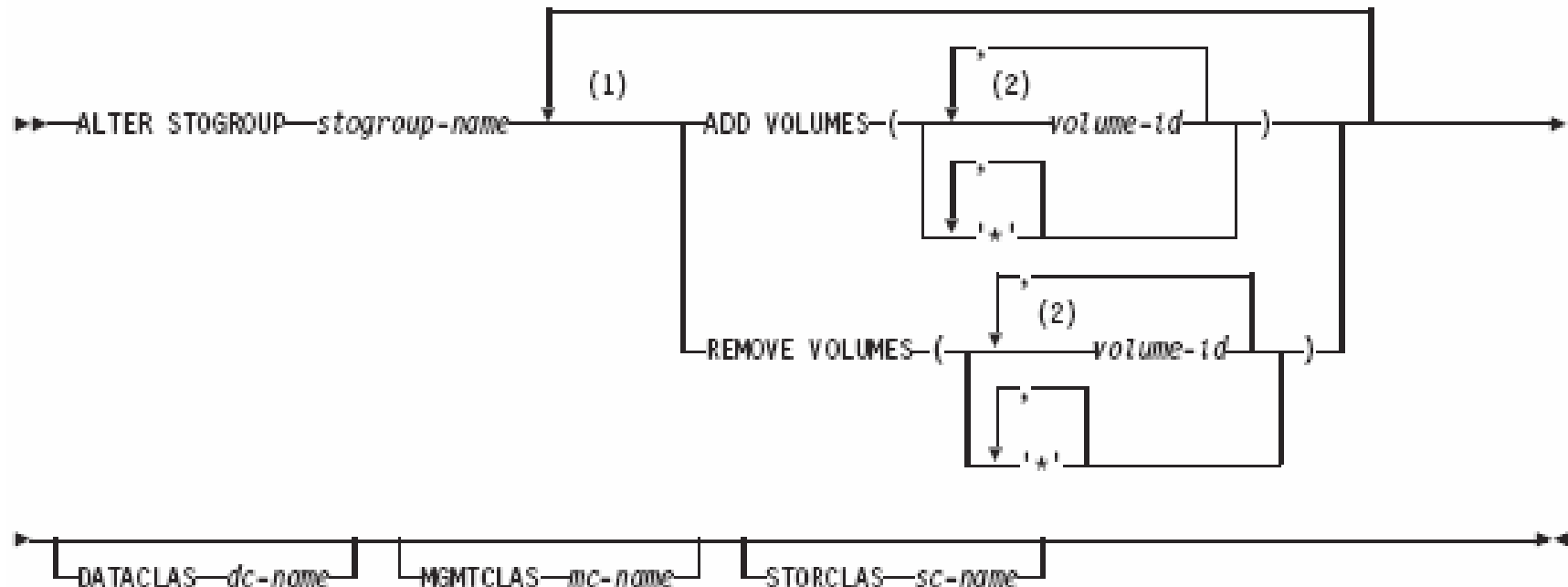
**Level 0** - user volume where data is migrated from

**ML1 (Migration Level 1)** - DFSMSHsm owned disk where data can be migrated. Data can then remain on this volume or further migrate to ML2.

**ML2 (Migration Level 2)** - DFSMSHsm owned tape or disk (traditionally tape) where data can be migrated to either from ML1 or directly from level 0. Discuss with your Storage Administrator which approach works best for you.

## DB2 9 and SMS Classes

- Starting in DB2 9, you can use SQL with **CREATE STOGROUP** or **ALTER STOGROUP** to specify a **DATACLAS**, **STORCLAS**, and/or **MGMTCLAS**. You cannot specify an SMS Storage Group.
- VOLUME** clause is now optional: it can be omitted if any of the DFSMS classes are specified.
  - NOTE! If you do not add VOLUMES, SMS is in control – by default the DC has VC=1, DVC=NULL. No multi volume data sets!**



Note – not all Storage Administrators will allow you to use the SMS classes. Verify this before using SMS Classes with STOGROUP.

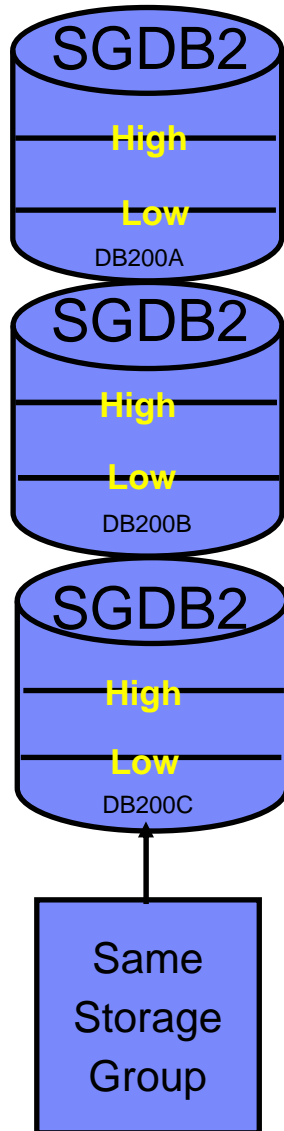
## What is an example of using STOGROUP with SMS Classes?

- **For example, my DB2 LDSes go through a Data Class that does not have EF/EA enabled for partitions, LOBS, or XML objects greater than 4GB.**
- **The Storage Administrator created a special Data Class with EF/EA enabled.**
- **I create a DB2 Stogroup pointing to the special newly created Data Class.**
- **I create new partitions, LOBS, or XML objects that require a DSSIZE > 4GB. My CREATE statement includes USING STOGROUP created specially for data sets > 4GB in the previous step.**

## Interesting Storage Group Information for DB2 Professionals - page 1 of 3

- **SMS uses the Storage Class attributes, volume and Storage Group SMS status, MVS volume status, and available free space to determine the volume selected for the allocation.**
- **DFSMSHsm functions to migrate data sets, backup (data sets incrementally), and dump (volume level) are decided at the Storage Group level.**
- Recommendations:
  - For production DB2 environments, do not allow DFSMSHsm to migrate or backup (data set level) table spaces or index spaces. However, depending on your installation's backup and recovery scenarios, you may want DFSMSHsm to dump (full volume) your volumes, even while DB2 is up. It is much better though if DB2 is down if at all possible.
  - My recommendation for non production environments is the same as for production environments. However, some installations will allow data to be migrated. If you decide to allow for migration, keep in mind ZPARM values for RECALL and RECALLD. Some considerations are:
    - How many objects need to be recalled at the same time?
    - How many objects reside on the same DFSMSHsm ML2 tape?
    - Will serial recalls complete in a timely manner?

## Interesting Storage Group Information for DB2 Professionals - page 2 of 3



## Migration and Allocation Thresholds - high and low values

High and low values are set for all volumes in a Storage Group

### HIGH - Used for disk volume allocation threshold

- If set to 75% for example, it will level all allocations to avoid volumes greater than 75% full
  - If set to 75%, then it allows for 25% growth in extents
  - Recommendation, start at 75% and then review
  - Also used to determine if volume migration will be executed

### LOW - Used for disk volume migration threshold

- Controls the low threshold by percentage of volume that eligible data sets can be migrated, thereby reducing stress on DFSMSHsm
- e.g., set low value to 50%, migrate after 100 days and now 90% of volume is eligible, only migrate down to 50%
- Recommendation, For image copy volumes that need to totally empty all data sets to allow for space for the next image copy run, set the low value 0. Otherwise, set the value higher to reduce the stress on DFSMSHsm for migration

- **It is not a one size fits all value.**

- For SMS Storage Groups housing active log and BSDS data sets – 99%. Active log data sets do not have secondary allocations and the BSDS has small secondaries. Data sets are not migrate, take LOW default.
- For SMS Storage Groups housing the DB2 catalog and directory – 99%. Typically the DB2 catalog and directory sits on one or two volumes. You may want to consider lowering this value if they reside on many more volumes.
- For SMS Storage Groups housing DB2 user data sets – IT DEPENDS! Between 75 and 85% may be a very good place to start as it would provide data sets the room to grow for secondaries.
  - If you have multiple Storage Groups based on size, you may want a higher value for your LARGE pool. You may need a much higher ceiling for large data sets.
  - Be careful when SGs have different flavors of emulated disk. 75% of a mod 3 is very different than 75% of a mod 54.
- For SMS Storage Groups housing archive log data sets – 99%. If you have specified the correct ZPARM values for the primary allocation there should be no secondary allocations. Some customers do not have uniform sizes for the active log data sets, in which case you may need to reduce the value.
  - If you migrate the archive log data sets, the HIGH must be set lower to trigger the archive. Determine what size LOW is required based on the amount of data migrated.
- For SMS Storage Groups housing image copy data sets – 99%. Typically allocations should fit into the primary application. Consider lowering the value If your utility or tool breaks down the allocation between primary and secondary.
  - If you migrate the archive log data sets, the HIGH must be set lower to trigger the archive. Determine what size LOW is required based on the amount of data migrated.

## Above HIGH Value Message

- **Storage group high allocation threshold exceeded message**  
**Messages IGD17380I and IGD17381I directed to the console so they will be recorded in SYSLOG and may be detected by automation, are issued. These messages alert storage administrators when a storage group's space usage has gone above its high threshold.**
- **One of the places you will see this message besides SYSLOG is on the DBM1 job log. As an example:**
- **IGD17380I STORAGE GROUP (*name*) IS ESTIMATED AT 86% OF CAPACITY,**  
**WHICH EXCEEDS ITS HIGH ALLOCATION THRESHOLD OF 85%**
- **Setting HIGH to 99% could defeat the purpose of receiving this message and taking action.**
- **You may not want to capture the SMS messages for all Storage Groups. For example, you may have a Storage Group for just archive log data sets with a very LOW value to force hourly migration, in which case you may have too many unnecessary messages. You may want to trigger the message only if the % is high on specific Storage Groups.**



## Interesting Storage Group Information for DB2 Professionals - page 3 of 3

- **Overflow Storage Group:**

- **An overflow Storage Group is used when non overflow Storage Groups are above their thresholds. An overflow storage group may also be specified as an extend storage group.**

- **Extend SG Name:**

- Data sets can be extended from one Storage Group to another, if there is insufficient amount of storage in the primary group.
- An extend SG name can also be an overflow Storage Group.
- Discuss with your Storage Administrator if using extend and/or overflow Storage Groups will help you.
- **Be aware of where your DB2 data sets are allocated. Review periodically for space related problems and inclusion for SnapShot and FlashCopy operations!**
- **COPY POOL BACKUP STORAGE GROUP - used starting in DB2 V8 to allow new BACKUP function.**

# Multi-Tiered Storage Groups

- **Specify Multi-Tiered SG Y in the Storage Class**

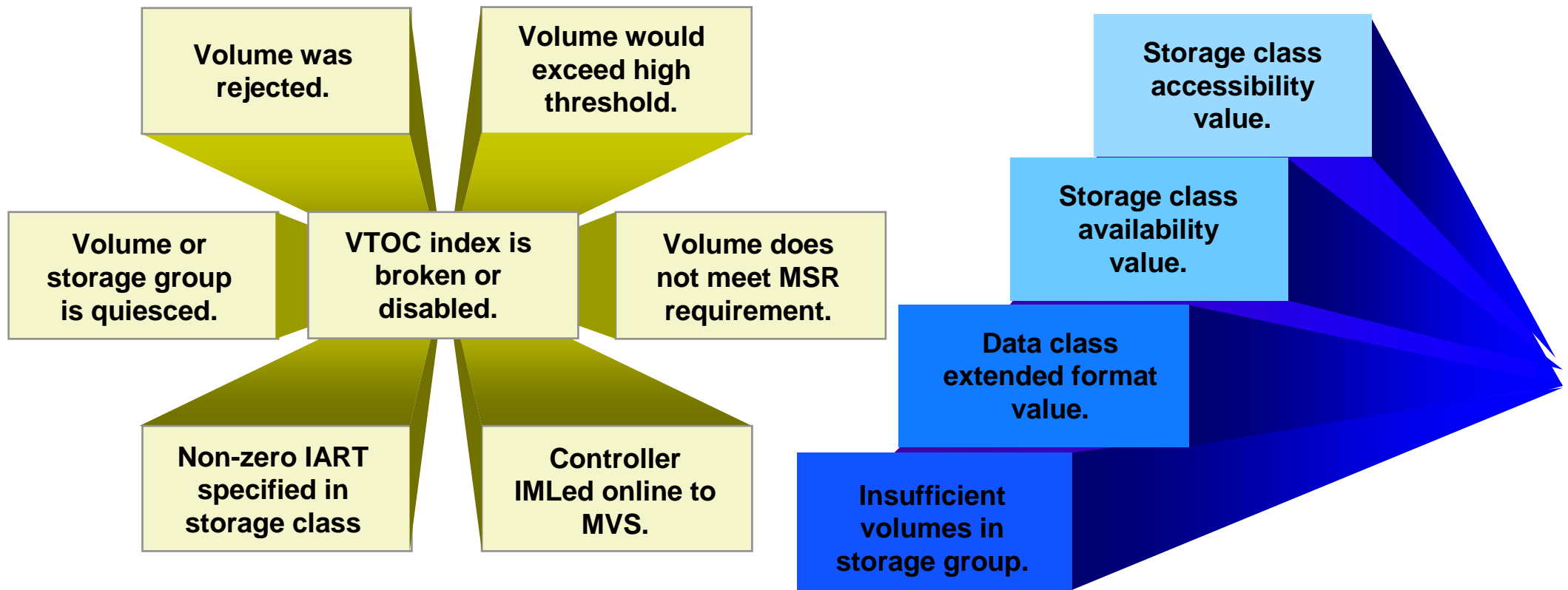
- **Example:**

- 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.

## Why Isn't My Volume Primary? - One example of why my data was allocated to a volume I did not expect



SMS can be set up to allow for volumes to be selected as primary, secondary, tertiary, or rejected. Discuss further with the Storage Administrator.

## Volume selection, is primary always what you want?

- **SMS allows the System Resources Manager (SRM) to select a DASD volume from the primary candidate list. SRM has a very short history span, typically seconds, generally not more than 2 minutes.**
- **When SRM is used, all volumes are in the primary list first and SMS uses an elaborate weighing algorithm, with one of the factors being the most amount of free space. This technique does not pick volumes at random.**
  - With SRM, when many data sets are created in a short period, SMS will at times place the data sets on the same volume and not randomly allocate data sets on other volumes.
- **If random volume selection is desired over SRM, set the Storage Class value for IART>0. Volumes are picked from the secondary candidate list and randomized.**
  - May not be a good idea with data sets that have Guaranteed Space
- **TEST, TEST, and TEST more! Make sure random volume selection works properly for you, especially when using overflow or spill volumes.**

## Why not have just one large SMS Storage Group?

- **Most Storage Administrators allocate the least amount of Storage Groups possible. This helps with overall workload for the Storage Administrator and management of the devices. Then when should multiple Storage Groups be used?**
  - Availability - When using Flash Copy or any other full volume dump/restore mechanism, Storage Groups should contain data sets from only one DB2 Data Sharing group, or subsystem. If multiple ones exist on the same volumes, flashing back or recovering a volume will effect not just the intended member/subsystem, but rather all others as well.
    - Specific data sets, such as the BSDS and active log data sets that are duplexed should be on allocated at a minimal on different extent pools
  - Performance – although volumes are logical and many logical volumes can be placed on the same physical DDMS there are still times when applications do not coexist well from an I/O perspective and cause hot spots.
  - Space requirements – often architecting a large and small, or small, medium, and large pools based on space will avoid many space related problems. This is commonly a solution for times as when DB2 allocates a shadow data set for REORGs.

- **Since volumes are logical, mod 1, 2, 3, 9, 27, 54, and EAV are recommended types. Some customers find a better fit by allocating for example mod 18 and 45 devices for their DB2 data sets, even though they are not device types commonly used. So long as space is used efficiently and within proper bounds, volumes can be allocated to fit your DB2 needs.**
- **With the DS8000, volumes can be dynamically expanded. This means that so long as the VTOC, VTOC index, and VVDS are large enough, a mod 3 for example can be dynamically expanded to a larger size emulated disk.**
- **How many data sets are or will be used for DB2?**
- **How large are the data sets? Most customers have a natural break between large and small data sets. The value between large and small is arguable. Most customers find the line between 200 and 400 cylinders. Typically 10% of the data sets are above 200-400 cylinders, 90% are below.**
- **Many customers still use 3390 mod 3 emulated disk. Allocating just 1 DSSIZE 64GB data set would require at least 32 mod 3 devices. If this same customer had 10 – 64GB data sets, at least 320 devices are required.**
  - An EAV device can house not just 1 - 64GB data set, rather about 2-3. 2 mod 54 devices can house a 64GB data set.
  - Management of devices can be greatly reduced by allocating data sets to device types meant to hold a specific amount of data

## CREATE STOGROUP VOLUMES

- **For allocations with non-Guaranteed Space, specify VOLUMES(“\*”) with one asterisk only. Using a real volser without GS results in unpredictable results under certain conditions.**
- **Using a non-SMS volser for VOLUMES does not make the volume non-SMS managed.**

Thinking about SMS managing all of your DB2 volumes (good idea!) and for STOGROUP specifying VOLUMES("\*") (another good idea!)?

- **Creating a STOGROUP and specifying VOLUMES("\*") allows SMS to choose a volume for allocation.**
- **Here is the issue at hand – at times a Storage Administrator sets up the SMS ACS routines to check for specific volume types for allocation. Much of the time an \* is not provided for the test. Without the ACS routine having an \* as part of a valid test, your allocation will fail.**
- **If you will be creating STOGROUPs with VOLUMES("\*") make sure your Storage Administrator adds an \* to their test of device allocations in the ACS routines.**



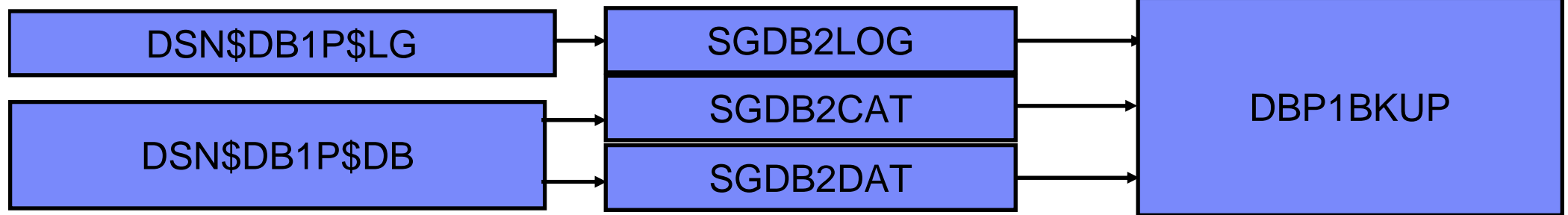
## Differences Between DB2 STOGROUP and SMS Storage Group

DB2 STOGROUP	SMS Storage Group
Different STOGROUPs can share the same disk volume(s)	One disk volume can only belong to one SMS Storage Group
VOLSERs are specific	Recommendation - code VOLUMES("*") to allow for SMS management. Avoid Guaranteed Space and specific VOLSERs where possible since this defeats the purpose of SMS.
SYSIBM.SYSVOLUMES has a row for each volume in column VOLID	When created with VOLUMES("*") SYSIBM.SYSVOLUMES has an * for column VOLID
Limited to management of 133 volumes	No volume limit
Volume selection based on free space	Volume selection based on SMS algorithms

# Resetting SMS Classes and Storage Group

- **You can change a data sets Storage Class and/or Management Class by using:**
  - IDCAMS ALTER command
  - dss COPY or DUMP/RESTORE commands
  - **NOTE!** Be careful when changing any classes as it can be used by a following ACS routine and change where data is placed or how it is used. For example, changing the Storage Class may affect the placement of a data set in the Storage Group if the ACS routine is driven and uses the Storage Class to determine the allocation for the Storage Group.
- **Data Class and Storage Group cannot be changed online**
  - dss COPY or DUMP/RESTORE will not drive the Data Class. In order to drive the Data Class, you must create a new data set and then use a copy mechanism, such as REPRO for the new data set.
    - REORG without the REUSE parameter will also drive the ACS routines, including Data Class
  - When using IDCAMS ALTER, or dss COPY, DUMP/RESTORE driving the Storage Class and/or Management Class are considered when the STORCLAS and/or MGMTCLAS parameters are used, which may again drive the Storage Group (see NOTE above). In this case the Data Class is not driven. Much of what is decided on what will be driven is based on if the dss parameters BYPASSACS or NULLSTORCLAS are specified.

# Copy Pools - BACKUP Command



- A copy pool is a defined set of pool storage groups that contain data that DFSMSHsm can backup and recover collectively, using fast replication.
- DFSMSHsm manages the use of volume-level fast replication functions, such as FlashCopy and SnapShot.
- Provides point-in-time copy and recovery services.
- `DSN$locn-name$cp-type`, `DSN` and `$` are required, `locn-name` is the DB2 location name, `cp-type` is the copy pool type. `DB` is for database. `LG` is for logs. For example: DB2 DB1P would have copy pools named `DSN$DB1P$DB` for the database copy pool and `DSN$DB1P$LG` for the log copy pool.
- **BACKUP** command records entry in BSDS, as well as the DFSMSHsm BCDS.

# ZPARMs Relating to Storage Management

## DSVCI - Support for VSAM Control Interval (CI) Greater than 4K - page 1 of 2

- **Support for CI sizes of 8, 16, and 32 K table spaces. For indexes, only 4K pages are supported. Starting with DB2 9, indexes can be 8, 16, or 32 K as well.**
- **Requires ZPARM value DSVCI in DSN6SYSP be set to YES, which is the default. It is set during the DB2 install in panel DSNTIP7 under VARY DS CONTROL INTERVAL.**
- **Supported for DB2 managed as well as user managed DB2 table spaces. DB2 install procedure provides JCL that will convert user defined DB2 catalog table spaces to proper CI sizes during ENFM. User managed table spaces require manual IDCAMS CI change.**
- **Activated in NFM for corresponding non LOB page sizes of table spaces, which will not take effect until after a LOAD or REORG of the table space. Note – this issue is resolved in DB2 9. Some alternatives prior to DB2 9 for LOBs:**
  - **If it is a LOB with LOG YES, one way to switch CI size is to use the RECOVER utility. For LOG NO LOBs, you can still use COPY and RECOVER after switching the LOB to access r/o and quiescing updaters.**
  - **STOP the LOB, IDCAMS EXPORT, then IMPORT with NEWNAME, change the names, then START the LOB.**
  - **DSN1COPY into a new object.**
- **New CI sizes:**
  - Reduce integrity exposures
  - Relieves some restrictions on concurrent copy (of 32K objects) and the use of striping (of objects with a page size > 4K).
  - Potentially reduce elapsed time for table space scans

- **Striped VSAM data sets are in extended format (EF) and internally organized so that control intervals (CIs) are distributed across a group of disk volumes or stripes. A CI is contained within a stripe.**
- **Increase your non 4K buffer pool sizes to accommodate new CI sizes.**
- **Some test results:**
  - 16K page measurement with 16K instead of 4K CI
    - +40% for non EF (Extended Format) datasets
    - +70% for EF datasets
    - EF getting nearly equivalent to non EF in data rate performance
  - Some table spaces may have negative results - test and verify
- **LISTCAT results (SPACE does not change for 8K and 16K table spaces, just 32K):**

–after CREATE TABLESPACE to a 4K buffer pool with PRIQTY 72000:

```

•CISIZE-----4096
•PHYREC-SIZE-----4096
•PHYRECS/TRK-----12
•SPACE-TYPE-----CYLINDER
•SPACE-PRI-----100

```

–after CREATE TABLESPACE to a 32K buffer pool with DSVCI=YES:

```

•CISIZE-----32768
•PHYREC-SIZE-----16384
•PHYRECS/TRK-----3
•SPACE-TYPE-----CYLINDER
•SPACE-PRI-----103

```

PRIQTY 72000

Notice, DB2 adds  
the 2.2%  
overhead for you!

```

DB2 V7 – 32K,
PRIQTY 72000
CISIZE-----4096
PHYREC-SIZE-----4096
PHYRECS/TRK-----12
SPACE-TYPE-----CYLINDER
SPACE-PRI-----100

```

## MGEXTSZ - Sliding Secondary Allocation Size - page 1 of 7

- **Applies to DB2 managed pagesets**
- **Tries to avoid VSAM maximum extent limit errors**
  - Can reach maximum dataset size before running out of extents. Beware of heavily fragmented volumes, which may impede this feature. This is less of an issue in z/OS 1.7 where an LDS can have 7,257 extents. The sliding secondary rule can exceed 255 extents when 'Extent Constraint Removal' is turned on in z/OS 1.7.
  - Calculation for next extent is based on total space and not total extents because of such factors as extent consolidation.
- **Uses cylinder allocation**
  - Default PRIQTY
    - 1 cylinder for non-LOB table spaces and indexes
    - 10 cylinders for LOB table spaces
- **Can be used for:**
  - New pagesets: No need for PRIQTY/SECQTY values
  - Existing pagesets: Execute SQL to ALTER PRIQTY/SECQTY values to -1

## MGEXTSZ - Sliding Secondary Allocation Size - page 2 of 7

- **Two sliding scales will be used depending on the maximum dataset size. The first 127 extents are allocated in increasing size, and the remaining extents are allocated based on the initial size of the data set:**
  - For 32 GB to 256 GB data sets, each extent is allocated with a size of 559 cylinders.
  - For data sets that range in size from less than 1 GB to 16 GB, each extent is allocated with a size of 127 cylinders.
- **Maximum dataset size determined based on DSSIZE, LARGE and PIECESIZE and defaults.**
- **BEWARE! If your Storage Administrator has set up a “LARGE” DB2 Storage Group, using this technique will probably not work well, unless you explicitly specify a large enough PRIQTY instead of DB2 implicitly specifying one.**
- **Advantages:**
  - Minimizes the potential for wasted space by increasing the size of secondary extents slowly at first
  - It prevents very large allocations for the remaining extents, which would likely cause fragmentation.
  - It does not require users to specify SECQTY values when creating and altering table spaces and index spaces.
  - It is theoretically possible to always reach maximum data set size without running out of secondary extents.
  - Particularly helpful for users of ERP/CRM vendor applications, which have many small data sets that can grow rapidly



## 4 Rules for Sliding Secondary

- 1. If PRIQTY is specified by the user, the PRIQTY value will be honored, otherwise the new default value as determined by either TSQTY, TSQTY\*10 or IXQTY will be applied: 1 cylinder for non-LOB table spaces and indexes, and 10 cylinders for LOB table spaces.**
- 2. If no SECQTY is specified by the user, the actual secondary quantity allocation will be determined by maximum of 10% of PRIQTY, and the minimum of calculated secondary allocation quantity size using the slide scale methodology and 559 (or 127) cylinders depending on maximum DB2 dataset size. When a pageset spills onto a secondary dataset, the actual secondary allocation quantity will be determined and applied to the primary allocation. The progression will then continue. Prior to DB2 Version 8, the PRIQTY would have been used.**
- 3. If SECQTY is specified by the user as 0 to indicate do not extend, This will always be honored. This condition will apply to DSNDB07 work files where many users set SECQTY to 0 to prevent work files growing out of proportion.**
- 4. If SECQTY specified by the user is greater than 0, the actual secondary allocation quantity will be the maximum of the minimum of calculated secondary allocation quantity size using the slide scale methodology and 559 (or 127) cylinders depending on maximum DB2 dataset size, and the SECQTY value specified by the user. When a pageset spills onto a secondary dataset, the actual secondary allocation quantity will be determined and applied as the primary allocation. The progression will then continue. Prior to DB2 Version 8, the PRIQTY would have been used.**

***Rules 1, 2 and 3 above apply regardless of the ZPARM MGEXTSZ setting.***

Rules 1, 2 and 3 apply to data sets created prior to DB2 Version 8.

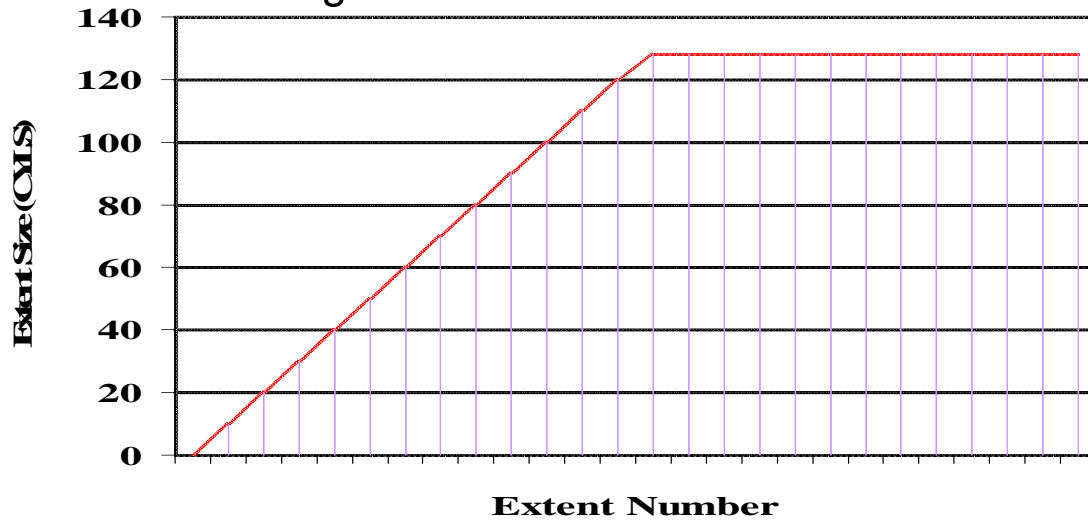
Rule 4 only applies when ZPARM MGEXTSZ is set to YES. The actual secondary allocation quantity applied will not be reflected in the Catalog. The primary allocation quantity and the actual secondary allocation quantities will never exceed DSSIZE and PIECESIZE.

# MGEXTSZ - Sliding Secondary Allocation Size

## - page 4 of 7

- **If PRIQTY and/or SECQTY is not specified by the user then -1 will be recorded in the associated Catalog columns:**
  - PQTY and SQTY in SYSTABLEPART
  - PQTY and SQTY in SYSINDEXPART.
- **With MGEXTSZ NO/YES (see APAR PK50113 for change to this design):**
  - **YES** – when allocating a new data set for a new piece (after A001) the primary and secondary allocation will be last size calculated by the sliding secondary of the previous piece. The maximum allocation will not exceed 127 cylinders for objects 16GB or below, and 559 cylinders for objects 32 or 64GB.
    - *For example, A001 hits the 2GB limit and DB2 now needs to create the A002 data set. A001's last allocation based on sliding scale was 79 cylinders, A002 will be created with a primary and secondary allocation of 79 cylinders.*
  - **NO** – When specifying a value for PRIQTY and SECQTY the new data set (created after A001) will have the same allocation as A001 for primary and secondary and therefore take on the characteristics of A001's PRIQTY and SECQTY. If PRIQTY and SECQTY were not specified then the allocation works as documented in YES above.

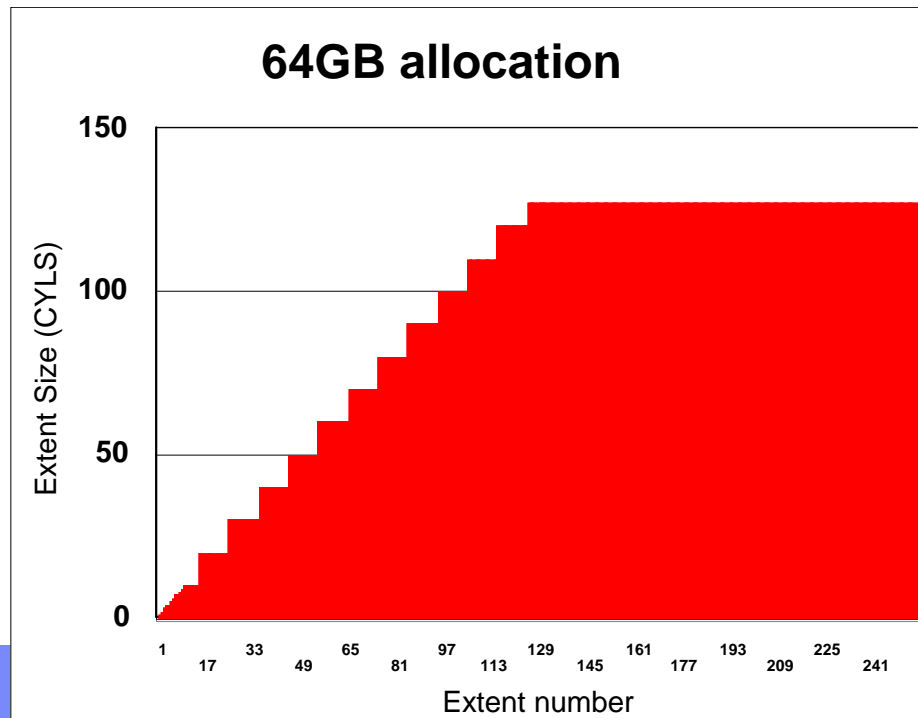
Sliding Scale for less than 1 GB to 16 GB



Maximum allocation of secondary extents

Max DS size in GB	Max Alloc in Cylinders	Extents to reach full size
1	127	54
2	127	75
4	127	107
8	127	154
16	127	246
32	559	172
64	559	255
128	559	414
256	559	740

Sliding Scale for 32 GB to 256 GB



Requires DB2 V10 – see PMR PM42175  
Set Data Class to exceed 255 extent limit

# Sample Allocations for A001 Segmented (assuming TSQTY=0 and IXQTY=0)

## With MGEXTSZ=NO

```
PRIQTY          48
SECQTY          48
```

```
1st extent tracks . : 1
Secondary tracks   . : 1
```

## With MGEXTSZ=YES

```
PRIQTY          48
SECQTY          48
```

```
1st extent tracks . : 1
Secondary tracks   . : 15
```

## With MGEXTSZ=YES and DSVCI=YES. For 32K pages only - we add 2.2% to the allocation:

```
PRIQTY          48
SECQTY          48
```

```
1st extent tracks . : 2
Secondary tracks   . : 16
```

## With MGEXTSZ=NO or YES

```
PRIQTY          48
no secondary
```

```
1st extent tracks . : 1
Secondary tracks   . : 15
```

```
no primary
SECQTY          48
```

```
1st extent cylinders: 1
Secondary cylinders : 1
```

## no primary and no secondary

```
1st extent cylinders: 1
Secondary cylinders : 1
```

```
PRIQTY          72000
no secondary
```

```
1st extent cylinders: 100
Secondary cylinders : 10
```

## no primary

```
SECQTY          72000
```

```
1st extent cylinders: 1
Secondary cylinders : 100
```

```
PRIQTY          72000
SECQTY          0
```

```
1st extent cylinders: 100
Secondary cylinders : 0
```

# Sample Allocations for A001 Segmented (assuming TSQTY=3600 (5 cylinders) and IXQTY=3600)

## With MGEXTSZ=NO

```
PRIQTY          48
SECQTY          48
```

```
1st extent tracks . : 1
Secondary tracks   . : 1
```

## With MGEXTSZ=YES

```
PRIQTY          48
SECQTY          48
```

```
1st extent tracks . : 1
Secondary tracks   . : 15
```

With MGEXTSZ=YES and DSVCI=YES. For 32K pages only - we add 2.2% to the allocation:

```
PRIQTY          48
SECQTY          48
```

```
1st extent tracks . : 2
Secondary tracks   . : 16
```

## With MGEXTSZ=NO or YES

```
PRIQTY          48
no secondary
```

```
1st extent tracks . : 1
Secondary tracks   . : 15
```

no primary and no secondary

```
1st extent cylinders: 5
Secondary cylinders : 1
```

```
PRIQTY          72000
no secondary
```

```
1st extent cylinders: 100
Secondary cylinders : 10
```

```
no primary
SECQTY          72000
```

```
1st extent cylinders: 5
Secondary cylinders : 100
```

## With MGEXTSZ=NO

```
no primary
SECQTY          48
```

```
1st extent tracks . : 75
Secondary tracks   . : 1
TRACKS/CA-----1
```

See PK05644 - preformat up to 2 cylinders even though allocation is in tracks. .

## With MGEXTSZ=YES

```
no primary
SECQTY          48
```

```
1st extent cylinders: 5
Secondary cylinders : 1
TRACKS/CA-----15
```

## Sliding Secondary FAQ

- **When using sliding secondary, after REORG without REUSE, do allocations start as new on the sliding scale as it did during allocation prior to the REORG, or is allocation based on current size and extents?**
  - Answer: After REORG without REUSE, allocations start from the beginning on the sliding scale. Size and extents are not factored in. If the goal is to reduce extents or size, then you must execute an ALTER to space prior to the REORG, which is the same operation were you not using sliding secondary.
  - Using ALTER to resize data sets that you know will grow using sliding secondary avoids the minimal performance penalty of using sliding secondary.

## TSQTY and IXQTY DB2 V6 and V7 by the PTF for APAR PQ53067

- **Specifies the amount of space in KB for the primary space allocation quantity for DB2-managed table spaces (TSQTY) and indexes (IXQTY) which are created without the USING clause.**
- **It uses cylinder allocation. The default values are set to 0, which means in DB2 V8:**
  - Default PRIQTY and SECQTY
    - 1 cylinder for non-LOB tablespaces and indexes
    - 10 cylinders for LOB tablespaces
- **Autonomic selection of data set extent sizes with a goal of preventing extent errors before reaching maximum data set size - use with sliding secondary allocation.**
- **Prevents heavy over-allocation and waste of excessive space.**
- **Can also result in better performance of mass inserts, prefetch operations, as well as LOAD, REORG and RECOVER utilities.**
- **PRIQTY honored if used.**

## SEQCACH

- **Determine for prefetch if `BYPASS` or `SEQ(quential)` is used for cache. Although `BYPASS` is an acceptable `ZPARM` value, we no longer bypass cache for the `DS8000`, `ESS`, or `RVA`. `3990` was the last controller to honor `BYPASS`.**
- **`BYPASS` for `DS8000`, `ESS`, or `RVA` will use sequential detect while `SEQ` will use explicit command.**
  - `SEQ` (explicit) puts tracks on the accelerated list and starts prestaging for next I/O. Operations are done on extent boundaries or stage group (an internal construct, depending on rank dimensions). Explicit reacts faster and can end sooner than using the detect mechanism (`BYPASS`).
  - `BYPASS` (sequential detect) stages data to the end of a stage group (an internal construct, depending on rank dimensions).
- **Recommendation – change `SEQCACH` to `SEQ`. NOTE – `BYPASS` is the default.**



## SEQPRES

- **Utility cache option.**
- **Recommendation – set to YES. Default is NO.**

## SVOLARC - Single VOLUME ARChive - PQ49360

- **DB2 allocates up to 15 volumes for archive log data sets to allow for space extension onto other volumes.**
  - Problem: For some SMS users who use guaranteed space for archive log data sets, DB2 may request primary allocation of up to 15 volumes (which it may not need) thereby causing space related problems.
  - Symptoms include: Message DSNJ103I with ERROR STATUS=970C0000, SMS REASON CODE=00004336 or REASON CODE=00004379
  - Solution: Consider changing to SVOLARC=YES if you want SMS to only allocate one volume to avoid this situation when using guaranteed space.

## UNIT and UNIT2 - for DB2 Archive Log Data Sets

▪ **It can be set to the same UNIT type. However, beware of the following:**

- Problem: many installations set UNIT and UNIT2 to a disk unit, VTS or use TMM.
- Storage Administrator sets ACS routines whereby:
  - ARCHLOG1 is allocated to a Storage Group that DFSMSHsm “sweeps” hourly to ML1 or ML2.
  - ARCHLOG2 is allocated to a Storage Group that DFSMSHsm migrates once every 24 hours.
- If they stay on ML1 or eventually migrate to ML2, their final residence can land on the same volume which can be a single point of failure. When ML2 tapes are set to MOD (typical scenario):
  - Tape can tear and become unusable
  - Tape can be misfiled by Operator or Tape Librarian
  - Tape is maliciously lost or destroyed
- Solutions:
  - Allow DFSMSHsm to backup the archive data sets before migrating
  - Duplex your ML2 tapes – may be an issue after tapes are recycled, then may result in a single point of failure again down the line
  - Transmit a copy of at least one set to another site
  - Split UNIT and UNIT2 between 2 device types, e.g. one to disk, one to tape
  - Use ABARS to copy one of the archives from ML1 or ML2
- NOTE: This can happen to any dual copied data set, including image copies. Dual copy data sets residing on VTS or when using TMM can also have a single point of failure. Discuss your requirements with your Storage Administrator,

# Utilities

## CHECK INDEX and FlashCopy

- **When using SHRLEVEL CHANGE in DB2 V8:**
  - Drains all writers and forces the buffers to disk for the specified object and all of its indexes
  - Invokes DFSMSdss™ to copy the specified object and all of its indexes to shadow data sets
  - Enables read-write access for the specified object and all of its indexes
  - Runs CHECK INDEX on the shadow data sets
- **Note: DFSMSdss uses FlashCopy Version 2 if available. Otherwise, DFSMSdss might take a long time to copy the object, and the time during which the data and indexes have read-only access might increase.**

## Other Utilities using FlashCopy

- **BACKUP and RESTORE utilities.**

## Utility Template Switching

```
TEMPLATE LESS_80 DSN &DB.&TS..IC..D&DA..T&TI. UNIT(SYSDA) LIMIT(80 CYL,GREAT_80)  
TEMPLATE GREAT_80 DSN &DB.&TS..IC..D&DA..T&TI. UNIT=TAPE  
COPY TABLESPACE DSN8S91E.DSN8D91A COPYDDN(LESS_80)  
COPY TABLESPACE AUXD501.AUXSITH COPYDDN(LESS_80)
```

Starting with DB2 9, TEMPLATES can switch based on size. In this example objects < 80 cylinders are allocated on disk, objects > 80 cylinders are Allocated to tape. Tapes can be stacked for multiple objects.

# DB2 V10 SMS

**Required!**



## DB2 10 requires that the DB2 catalog and directory be SMS managed

- **New Catalog and Directory data sets created during CM require SMS, with EF (Extended Format) and EA (Extended Addressability) enabled.**
- **During ENFM, some of the DB2 catalog and directory objects are converted to PBG with a DSSIZE 64G.**
- **VSAM Objects > 4GB are required to be SMS managed.**
  - Data Class in the ISMF panel - EF (Extended Format) and EA (Extended Addressability) must be turned on.
  - Data Class ACS routine must set the DB2 catalog and directory objects to this Data Class with EF/EA enabled. Setting the value is easily done as we know that the 3<sup>rd</sup> qualifier is DSNDB01 or DSNDB06.
- **DSNTIJSS (in SDSNSAMP) provides SMS classes for customers without SMS in use.**
  - The environment created by DSNTIJSS is ONLY for DB2 Catalog and Directory data sets.
    - Other DB2 data sets such as logs and BSDS not covered.
    - Parts of DSNTIJSS use NaviQuest procedures

# Installation process changes

Introducing job DSNTIJSS: Creates a sample SMS environment

- **Background:**

- In V10, data sets being defined for the catalog and directory are managed by DB2 – not the user
- These data sets also must be associated with an SMS data class for allocating them in extended format and using extended addressability.
- You are not required to convert existing DB2 catalog and directory data sets to the SMS environment before migrating. These can remain non-SMS managed indefinitely, but will be converted to SMS management the next time the related table space is reorganized

- **DSNTIJSS shows how to create a sample SMS environment for DB2 catalog and directory data sets that consists of:**

- An SMS Source Control Data Set (SCDS) for storing the base configuration
- An SMS storage Class
- An SMS storage Group with 3 enabled volumes
- An SMS data class for allocating data sets in extended format and to use extended addressability
- SMS Automatic Class Selection (ACS) routines for managing the DB2 catalog and directory data sets in the storage group, storage class, and data class

- **The sample environment is to help you get started – it is not the ideal or recommended SMS environment for your system**

- **After running the job, use the following commands to make the SCDS active**

```
SETSMS SCDS(scds-name)
```

- Serviceability:

- **Common problems:**

- Symptom: After installing and starting DB2 V10, job DSNTIJTC fails with abend S04E, reason code 00C200EF
- Cause: The catalog and directory data sets are not defined in a valid SMS environment
- Response: Provide an SMS environment with an SMS data class that allocates the data sets in extended format and for extended addressability.

## DB2 catalog and directory are now DB2 managed

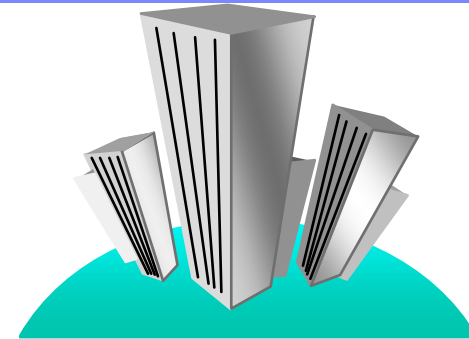
- **Not only does the DB2 catalog and directory require SMS, the data sets will also be DB2 managed instead of user managed. No more IDCAMS DEFINE requirement after ENFM for DB2 Catalog and Directory objects.**
- **Starting with CM, new catalog and directory data sets are DB2 managed. Existing data sets are DB2 managed from space and data perspective**
  - For customers that do not SMS manage the catalog and directory data sets, moving your data sets to SMS managed volumes is not required.
    - New catalog and directory objects must be SMS managed with EF/EA enabled
    - Catalog and directory objects that are REORGed will automatically become DB2 managed, and SMS managed with EF/EA required for the newly allocated data sets.
    - During REORG of catalog and directory objects, an internal STOGROUP that cannot be viewed is used. The STOGROUP was created with VOLUMES(“\*”).
    - Data sets that are not REORGed can remain on the non-SMS volumes until REORG.
    - This requirement is for the DB2 catalog and directory objects only and do not require the BSDS or logs to be SMS managed.
    - User defined data sets for the Catalog and Directory remain the same and do not require any changes.

## DB2 managed data sets - Benefits

- **Minimize user's effort to maintain data sets.**
  - No need to allocate data sets and extended pieces (A002, A003 etc...)
  - No need to allocate data sets as part of the migration (next next release)
    - We still have DSNTIJIN in this release because there are new catalog indexes to be created during migration and we cannot make them DB2 managed until leaving CM.
  - No need to allocate data sets for the shadow for online REORG.
  - No need to estimate the size for the data sets (had to provide space allocations in the DEFINES)
    - DB2 will use a default for the primary and a sliding scale for secondary.
      - ALTER PRIQTY, SECQTY, and STOGROUP are not permitted for the DB2 Catalog and Directory data sets after they are DB2 managed.
    - Minimize the chance of running out of extends before reaching the maximum data set size.
  - SMS will determine which dataset goes to which volume.
  - Minimize outage due to improper data set allocations.
  
- **New fields in installation/migration panels (CLIST)**
  - SMS information (data class, storage class, management class) stored in ZPARM.

# DB2 Catalog Evolution

*The DB2 catalog continues to grow with every DB2 release.*



DB2 Version	Table Spaces	Tables	Indexes	Columns	Table Check Constraints
<b>V1</b>	<b>11</b>	<b>25</b>	<b>27</b>	<b>269</b>	<b>N/A</b>
<b>V3</b>	<b>11</b>	<b>43</b>	<b>44</b>	<b>584</b>	<b>N/A</b>
<b>V5</b>	<b>12</b>	<b>54</b>	<b>62</b>	<b>731</b>	<b>46</b>
<b>V6</b>	<b>15</b>	<b>65</b>	<b>93</b>	<b>987</b>	<b>59</b>
<b>V7</b>	<b>20</b>	<b>84</b>	<b>117</b>	<b>1212</b>	<b>105</b>
<b>V8</b>	<b>21</b>	<b>87</b>	<b>128</b>	<b>1286</b>	<b>105</b>
<b>V9</b>	<b>28</b>	<b>106</b>	<b>151</b>	<b>1668</b>	<b>119</b>
<b>DB2 10</b>	<b>81 (88-7)</b>	<b>124</b>	<b>195</b>	<b>1701</b>	<b>119</b>

**Does not include objects for XML Schema Repository!**

## Catalog Restructure enhancement summary

- DB2 (SMS) managed catalog and directory data sets
- New CLOB/BLOB columns to the catalog
  - Merge records that store SQL statements' text
- Reduce catalog contention
  - Removal of links
  - Change to row-level locking
- Convert some catalog and directory table spaces to partition-by-growth (PBG)
  - With MAXPARTS = 1
- Combine SYSUTIL and SYSUTILX into a single table SYSUTILX



# Eliminate 64Gb limit on Catalog & Directory – e.g. SPT01

- V10 will relieve SPT01 space problems
- **SPT01 can grow beyond 64G, when system in DB2 10 NFM.**
- **Hence Application Developer creates packages using:**
  - Creates Packages using BIND PACKAGE.
  - Creates Packages using CREATE TRIGGER.
  - Creates Packages using SQL procedures.
  - Binds a package, using plan stability.
- **DB2 extends SPT01 beyond 64G when needed.**



**Binds and rebinds never fail for lack of SPT01 space.**

- **DB2 9 changed the work file allocations from user managed to DB2 managed, however did not require that the data sets be DB2 managed.**
- **DB2 managed work file data sets are not required for V10. Customers can continue to use user defined work file table spaces in V10.**
- **With V10 NFM work files can be allocated as PBG data sets allowing keywords MAXPARTITIONS and DSSIZE to be used. This is very useful when having problems with large DGTTs.**
  - Using keywords MAXPARTITIONS and DSSIZE can limit the size of the data sets and thereby limit runaway transactions
    - For example, to limit the space used to 3 GB, you could set MAXPARTITIONS 3 DSSIZE 1G. With DB2 managed classic segmented table spaces, this was not possible. In this case you could only limit the growth at 2 GB or less (via PRIQTY nK SECQTY 0)
  - With PBGs, a DGTT can span multiple 64GB data sets, however it is limited to one table
  - PBGs must be DB2 managed data sets
  - Limiting the size can be done today using IDCAMS for user managed work files, but in this scenario PBGs are not used
- **For work file data sets that exceed 4GB, Data Class EF/EA must be enabled**
- **If ZPARM WFDBSEP=NO (the default), DB2 will try to use work file PBG table spaces for DGTT only, however if there is no other table space available DB2 will also use it for workfiles (e.g. sorts, CGTTs). With WFDBSEP=YES, DB2 will only use work file PBG table spaces for DGTT and if there is no other table space available, a work file application (e.g. sorts, CGTTs) will get a 'resource unavailable' message.**



# **A few notes about tape**

# Terminology

- The **ATL (Automated Tape Library)** is a device consisting of robotic components, cartridge storage areas, tape subsystems, and controlling hardware and software, together with the set of tape volumes that reside in the library and can be mounted on the library tape drives. **ATLs do not provide tape virtualization or automatic tape stacking.**
- The **VTS (Virtual Tape Server)** is a hardware-based solution that addresses not only tape cartridge utilization but also tape device utilization and hence overall tape subsystem costs. The VTS was introduced and made available in 1996 as a new breed of storage solution that **combines a high-speed disk cache with tape automation, tape drives, and intelligent storage management software running on a server.**
- The **IBM System Storage Virtualization Engine TS7700** is the newest member of the IBM TS7000 Virtualization Family. It represents the fourth generation of IBM Tape Virtualization for mainframe systems and **replaces the highly successful IBM TotalStorage Virtual Tape Server (VTS).**
  - Most of this presentation is based on IBM's newest VE TS7700.

# IBM Backup & Archive Portfolio



**TS1050  
(LTO5)**

**TS1140  
(Jaguar)**

## Tape Drives

- LTO5 tape drive
  - Encryption capable
  - 1.5TB Native capacity cartridge
  - Up to 140 MB/sec throughput
  - LTFS support
  - Dual ported drive
- TS1140 tape drive/controller
  - Fourth generation tape drive
  - Controller supports FICON & ESCON
  - Tape drive data encryption
  - Up to 4TB cartridge capacity
  - Up to 1200 MB/sec throughput
  - Auto Virtual Backhitch

**TS3200  
(3573)**



**TS3100  
(3573)**



**TS3310  
(3576)**



**TS3500  
(3584)**

## Tape Libraries

- TS3100 tape library (up to 19.2TB)
- TS3200 tape library (up to 38.4TB)
- TS3310 tape library (up to 316.8TB)
  - Stackable modular design
  - LTO Tape drives
- TS3500 tape library (up to 30PB with LTO5 or up to 180PB with TS1140)
  - Linear, scalable, balanced design
  - High Availability
  - High Density
  - Fastest robotics in industry
  - LTO and TS1100 tape drive
  - Connect up to 15 libraries for over 300,000 slots using shuttle complex



**TS7720  
(tapeless)**

**TS7740  
(Hydra)**

**TS7680  
(DeDup)**

**TS7650 App  
(DeDup)**

**TS7650G  
(DeDup)**

**TS7610  
(DeDup)**

## Virtualization

### Mainframe Virtual Tape

- TS7720 (Virtual Tape)
  - Tapeless
  - Up to 1000 MB/s throughput
  - Up to 440 TB native cache
  - Standalone or GRID (PtP)
  - Up to 6-way Grid
  - Hybrid Grid support
- TS7740 (Virtual Tape)
  - Up to 1000 MB/s throughput
  - Up to 28 TB native cache
  - Standalone or GRID (PtP)
  - “Touchless” with Export options:
  - Up to 6-way Grid
  - Hybrid Grid support
- TS7680G (Dedup)
  - 600Mb/s INLINE
  - Up to 1PB Repository
  - 100% Data Integrity
  - Data / Disk Agnostic
  - Native Replication

### Open Systems Virtual Tape

- TS7610 App Express (Dedup)
  - 80Mb/s INLINE
  - 4TB & 5.4TB Useable capacity
  - 100% Data Integrity
  - Data Agnostic
  - Native replication
  - Many to one replication support
- TS7650 Appliance (Dedup)
  - 500Mb/sec INLINE
  - 7TB to 36TB Useable capacity
  - 100% Data Integrity
  - Data Agnostic
  - Many to one replication support
  - High Availability (36TB option)
- TS7650G (Dedup)
  - 1GB/s (Cluster) INLINE
  - Up to 1PB Useable capacity
  - 100% Data Integrity
  - Data / Disk Agnostic
  - Native replication
  - Many to one replication support
  - High Availability

# Virtual Tape Concepts

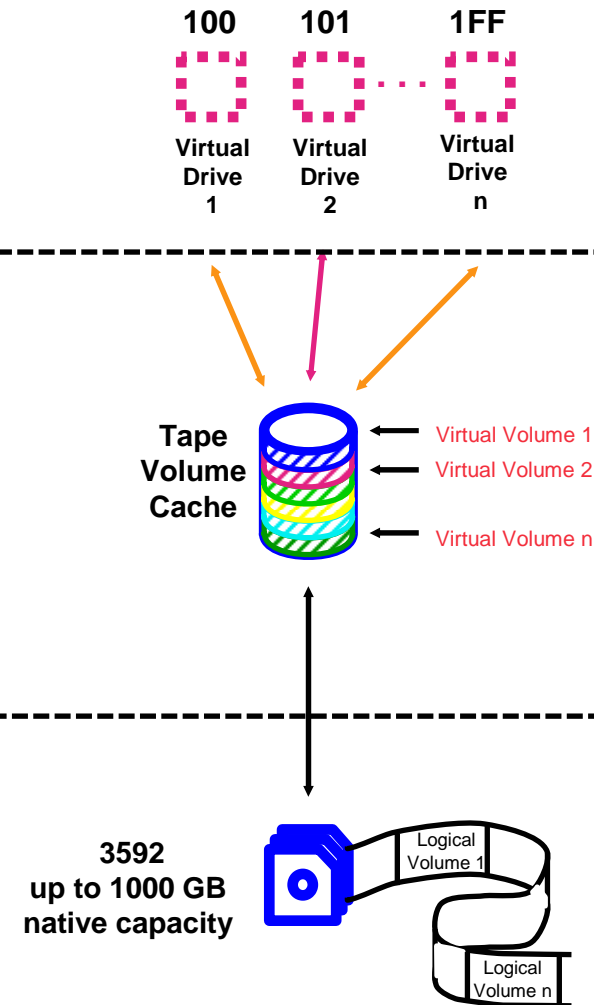
- Virtual tape drives
  - Appear as multiple 3490E tape drives
  - Shared / partitioned like real tape drives
  - Designed to provide enhanced job parallelism
  - Requires fewer real tape drives
  - TS7700 offers 256 virtual drives per cluster

---

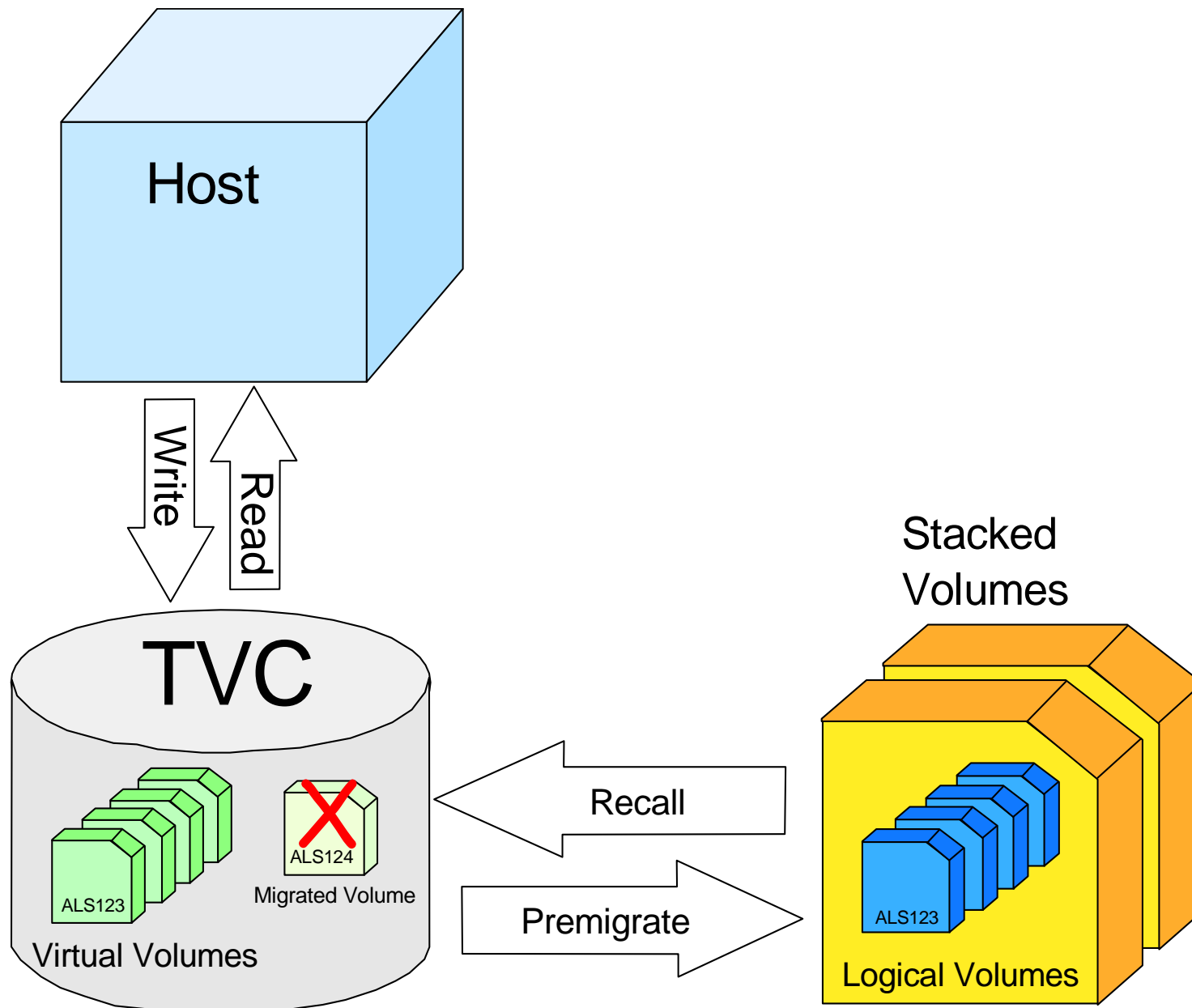
- Tape volume caching
  - All data access is to disk cache
  - Removes common tape physical delays
  - Fast mount, positioning, load, demount
  - Up to 28 TB / 440 TB of cache (uncompressed)

---

- Volume stacking (TS7740)
  - Designed to fully utilize cartridge capacity
  - Helps reduces cartridge requirement
  - Helps reduces footprint requirement



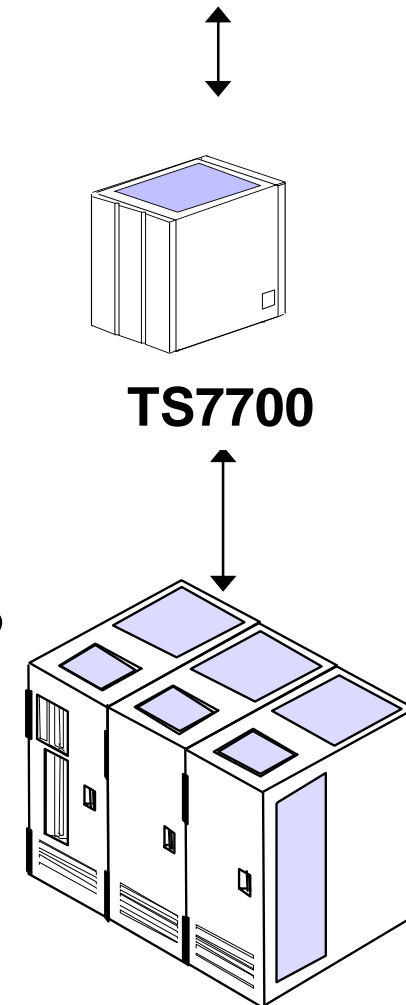
# Logical Volumes and Stacked Physical Volumes



# TS7700 - Capabilities

- Tape Volume Cache
  - TS7740
    - RAID 5
    - 1 to 28.17 TB cache (3 – 84.5 TB @3:1 compression)
  - TS7720
    - RAID 6
    - 20TB to 440TB cache (60 – 1320 TB @3:1 compression)
- 256 Virtual Tape Devices
- 2 Million Logical Volumes
- Advanced Policy Management
- 4 to 16 3592 Physical Drives (TS7740)
- 3584 Physical Library Support (TS7740)

## Mainframe Attachment



## Where do your tape data sets reside? Some things to think about

### ■ Virtual tapes

- Do your data sets use the right size virtual tape?
  - 400, 800, 1000, 2000, 4000, or 6000 MB – set via SMS Data Class
  - 4GB data set can reside on one virtual tape volume or 10
- Do you stack/append your data sets or waste valuable logical volumes?
- Are your tapes part of a grid/PtP implementation, or duplexed another way?
- Same serialization problem as manual and ATL tapes
  - This is true as well, even if the “tape” data set resides in the TVC (disk) portion of the tape unit. MVS tells DB2 the data set resides on tape, even though it resides on disk and therefore DB2 must serialize all “tape” requests.

## Media for DB2 related data (archive logs, image copies, large sort work data sets, some other assorted data sets)

- **Disk – if parallelism is one of the key factors, create your data sets on disk. There are some alternatives:**
  - create data sets on tape, then use an MVS utility to recreate the data on disk when required.
  - Use hsm to migrate the data to tape, then recall the data. This is a more automated approach than the previous alternative. Both alternatives require additional time to bring your data sets back to disk.
  - Disk is also a good alternative if you need to recover hundreds or thousands of objects. For example, tape might not be the best alternative for an SAP customer requiring hundreds or thousands of tape image copy data sets for recovery.
- **ATL**
  - If you are sending tapes offsite use the ATL instead of virtualized tape.
  - Extremely large data sets. Because of data transfer time, ATLs may be a better alternative than virtualized tape. Benchmark testing should be executed.
- **VTS/TS7700**
  - When it is imperative to have a second copy of a tape data set automatically created at an alternate site.
  - When stacking methods are not used or you do not have many files to create and you want an automated way of stacking your data instead of making them a multi file tape.



- **MYTH: It does not matter if my archive log data sets reside on tape or disk**
  
- **FACT: Archive logs on disk only take up one slot in the BSDS as the must be cataloged. Tape allocations can be cataloged or uncataloged, therefore each volser takes up one BSDS slot for multi-volume tapes if not cataloged and one slot total if cataloged – as with disk.**
  - An archive log data set residing on three disk volumes consumes one BSDS slot – only the first volser is recorded
  - An archive log data set residing on three tape volumes that is not cataloged consumes three BSDS slot – one for each volser
  - Review ZPARM MAXARCH, ARCRETN, and CATALOG (default NO) settings
  
- **MVS serializes access to all DB2 data sets on tape. Archive log data sets residing on tape can only be serialized, even when residing on the disk portion of the tape unit.**
  - Archive logs residing on disk can be parallelized

- **MYTH: When I see a tape mount, a physical tape is always mounted**
  
- **FACT: Seeing a take mount, unmount, request for additional tape volumes, etc. does not mean tapes are real**
  - Virtual tapes still issue the mount/unmount, etc. commands as you would see on a manual or ATL tape. You cannot tell strictly by the mounts you see in the JESLOG what type of tape you are dealing with without knowing UCB addresses, etc.
  - When the hosts requests a volume that is still in cache, the volume will be virtually mounted, no physical mount is required.
  - Access to data is at disk speed. Tape commands such as space, locate, rewind, and unload are mapped into disk commands that are completed in tens of milliseconds rather than the tens of seconds required for traditional tape commands.
  - Mounting of scratch tapes is also virtual and does not require a physical mount.
  - The TS7740 Node manages the physical tape drives and physical volumes in the tape library and controls the movement of data between physical and virtual volumes.
  - Multiple, different, emulated 3490E volumes can be accessed in parallel because they physically reside in the tape volume cache. **(A single virtual volume cannot be shared by different jobs or systems at the same time, even though the media is really disk.)**

- **MYTH: For virtual tapes, my Storage Administrator does not need to know how long my data set should reside on the disk vs. tape portion.**
  
- **FACT: The Storage Administrator can influence the length of time your data set resides in the TVC vs. physical tape**
  - Discuss your requirements with your Storage Administrator
  - Do your image copy data sets really need to reside in the TVC for a long time or can they quickly be moved to physical tape?
  - How about your archive log data sets, etc.?
  - TVC space is limited. Most data sets do not require long term placement of data in the TVC.

- **MYTH: My Storage Administrator knows how to separate my data on physical tape volumes**
  
- **FACT: Most Storage Administrators do not know how archive logs and image copies are used**
  - Even for the Storage Administrators that know what archive logs and image copies are used for, they do not typically know the naming convention you are using for dual pair data sets
  - If you have dual pair of archive logs and/or image copies on virtual tape, let the Storage Administrator know that they must ultimately reside on different physical tapes.
  - The Storage Administrator can pool specific data sets onto different physical tape volumes.
  - This recommendation goes along with duplicating tapes in case one snaps or has other media failure

- **MYTH: It is a good idea to compress a data set that will reside on disk, and then to compress it again on tape**
  
- **FACT: Although disk subsystems no longer supports compression, you can CPU compress certain data sets residing on disk**
  - For example, starting with DB2 9 NFM, you can compress the archive log data sets, potentially saving a significant amount of disk space
  - Consider not compressing the data set if it moves from disk to hardware compressed tape
    - Customers have run into problems compressing on disk and then trying to re-compress on tape
  - CPU starved customers need to weigh compression vs. CPU costs
  - Generally, it is inefficient to compress very small data sets

# Tape Issues

- **Tape is an excellent medium to place your DB2 related data on. Consider tape for your:**
  - Archive logs
  - Image copies
  - Very large Sortwork data sets
  - Other assorted data sets
- **There are some challenges using tape in a DB2 environment:**
  - VSAM data sets, PDSs and PDSEs must reside on disk
  - No concurrency or sharing within a data set or volume, therefore parallelism does not exist.
  - Tapes perform best using pure sequential access. Other access types may result in some performance degradation.
  - Data sets residing on tape cannot be striped
- **Discuss your tape and data requirements with your Storage Administrator. Understand your company's tape solutions and how to best manage your DB2 related data.**

**z/OS 1.11, 1.12,  
and 1.13  
Enhancements  
useful for DB2**

## z/OS 1.11 enhancements useful for DB2

- IDCAMS DELETE command is enhanced to include a new function called DELETE MASK. It allows users to specify the data set name selection criteria desired with a mask-entry-name and a keyword “MASK”.
  - A mask-entry-name (also called as filter key) can have two consecutive asterisks (\*\*) or one or more percentage signs (%)
  - MASK keyword is the keyword to turn on the new feature. For example,
    - DELETE A.B.\*\* MASK
    - DELETE A.BC.M%%K MASK
  - AMS deletes only the first 100 data sets that matched the selection criteria
- SMS expands the data set separation function to the volume level to reduce I/O contention and risk of single points of failure
  - SMS provides new syntax for data set separation function
  - SMS attempts to separate data sets listed in volume separation groups onto different extent pools and volumes
- SMS striping restrictions lifted
- EAV can be used for more data types.
- IEFBR14 DELETE will not recall a migrated data set before it is deleted



# z/OS 1.11 enhancements useful for DB2

## ▪ HSM

- Provide the capability to specify a retention period on the data set backup command, which allows users to keep an individual backup copy for either a shorter or longer than normal period of time
- Enable individual backup copies to be retained for an indefinite amount of time
- Maintain more than 100 backup copies
- Larger data sets (>58 K tracks) will now be DFSMSHsm managed large format sequential data sets when migrated or backed up.
- OVERFLOW volumes can now be used for migration and backup.
- An installation can adjust the values that direct data sets to OVERFLOW versus NOOVERFLOW volumes and the threshold of the OVERFLOW volume pool.

## ▪ FlashCopy

- Fast Replication data set recovery will use saved catalog information to recover deleted and moved data sets
- The ALLOWPPRCP keywords from the FRBACKUP and FRRECOV commands can now be set in the SMS copy pool definition rather than specifying them on the command line
- User catalogs will automatically be unallocated prior to a FRRECOV of a copy pool version
- The ARC1803E message will include DFSMSdss messages and return codes for each failing volume
- Fast replication copy pool backup storage group volumes will be automatically reinitialized when DFSMSHsm has determined that the VTOC may become corrupt

## z/OS 1.12 enhancements useful for DB2

- **Ability to open more data sets concurrently**
- **Enhance VSAM and VSAM RLS to reclaim empty space (control areas) in KSDSs. Applies to SMS and non-SMS data sets.**
  - Will resolve the main reasons for KSDS reorganizations:
  - Reclaim space
  - Improve both sequential and direct performance
  - Can be used to reclaim space in an ICF catalog if running out of space or extents without down time
- **EAV supports almost all types of data sets**
- **ICF catalogs can be allocated greater than 4GB when Extended Addressability (EA) is enabled. ICF Catalogs can grow beyond the 4 GB size to a maximum size of 4 GB times 32K depending on CI size.**
- **IDCAMS command DELETE GDG FORCE will no longer require that migrated GDSs be recalled before being deleted. GDSs having a volser of “MIGRAT” invoke HDELETE instead and will not recall the data sets first.**

## z/OS 1.12 enhancements useful for DB2

- IDCAMS command DELETE is enhanced to allow users to specify a single asterisk (\*) as the member name of a PDS or PDSE. This will delete all the members in that PDS or PDSE.
  - For example, DELETE A.B(\*) will delete all the members residing in partitioned data set A.B
- MVS will call the PDSE validation tool to check the integrity of a PDSE in LNKLST when the system is coming up and when LNKLST is being changed. With this change, users will learn of PDSE structure errors earlier.
- Partial release can be used for Extended format (EF) data sets which are over allocated and span volumes (e.g. best fit data sets). Space on the last volume containing data and all unused empty volumes will be scratched and marked as candidates.

# z/OS 1.12 enhancements useful for DB2

## ▪ hsm

- Using Fast Reverse Restore, you can:
  - Recover a copy pool without waiting for physical background copy to complete. This is a significant benefit as the background copy may take a very long time to complete.
  - Recover a copy pool from a disk version in either COPY or NOCOPY environment
  - Recover a copy pool from space efficient volumes in NOCOPY environment. Space efficient volumes are beneficial for certain DB2 related functions, such as cloning.
- Restore from dump tape:
  - Multitasking volume recovery from dump tape can concurrently recover up to 64 volumes
  - Recovery of a copy pool version from dump can be done with single command. Before z/OS 1,12, A FRR (Fast Reverse Recovery) volume recovery command is needed for each volume when recovering a copy pool version from tape
  - Copy pools can be recovered from partially dumped versions
  - The Fast Replication Recover function will resume an interrupted recovery of a copy pool version unless RESUME(NO) is specified
  - Recover processing will reduce mounts and demounts by automatically scanning the Recover queue for additional recovery requests that need the currently mounted tape before demounting a tape.
- hsm space management:
  - Perform Primary Space Management (PSM), Interval Migration (IM) and Command Volume Migration in multiple overlapping phases.
  - Begin pre-processing the next eligible volume to be space managed while data movement occurs on the previous volume. Allows a separate task to begin working on the next volume before data is fully processed on the original volume.

## z/OS 1.13 enhancements useful for DB2

- **hsm**

Specify that space management be done when any volume in a storage group for which auto migration is enabled exceeds the utilization threshold, rather than waiting for Interval Migration processing. This can avoid the hourly spike for auto migration.

- A new subparameter for the RELEASE RECALL command you can use to specify that DFSMSHsm avoid recalling data sets from missing or faulty tapes while releasing the hold on recalls from DASD.

### **PARMLIB**

- A new parmlib member, IGGCATxx, allows users to specify a number of Catalog system parameters. Default is IGGCAT00
  - VVDS space defaults
  - Catalog utilization warning message threshold
  - Limit on CAS service tasks (overrides any specification in SYSCATxx)
  - Whether to enable extension records for user catalog aliases
  - A number of other things you also specify using MODIFY CATALOG
- Customers can now create their own Catalog parmlib member(s) to customize their Catalog environment; the parameters can be changed by doing an IPL or a simple restart of the Catalog address space.

## ■ PDSE

- New command to discard all buffered data for a PDSE when there is a need to discard cached directory pages in order to recovery from possible incorrect pages in the directory cache.
  - (V SMS,PDSE(1),REFRESH)
- New command to display the connections for a PDSE. Used to help determine scope of a problem with a particular PDSE.

## ■ Tape

- Allow different systems in the sysplex to concurrently read multivolume tape files in a way similar to the Deq at Demount Facility.
- A new JCL keyword (FREEVOL=EOV) will allow a tape for part of a multivolume data set to be available at end of volume rather than end of step.
- Limitations:
  - Does not require APF authorization, and since it is implemented in the JCL, no changes to the application are required.
  - Honored only for input processing.
  - EOV and CLOSE volume disposition processing will unload the volume when the disposition would otherwise be REWIND.

# FAQ

## DB2 and SMS

- **I hear that DB2 does not work with SMS, is that true?**

- When SMS was first introduced many years ago there were some recommendations not to use SMS for DB2 data sets. This recommendation was rescinded about a year or two after SMS was first announced. So, yes, you can fully use DB2 with SMS.

- **I hear that my DB2 user data can be SMS managed only and not the system data, is that true?**

- The short answer is no. All DB2 data can be SMS managed. As a matter of fact, if you want to use DB2 V8 system level backup and restore, SMS is required for user as well as system data sets.

- DB2 V10 requires that the DB2 Catalog and Directory data sets be SMS managed with the Data Class attributes EF/EA enabled.



Now that my volumes are SMS managed, do I still need to worry about space issues?

- **Yes. Your likelihood of getting that 2 a.m. call or visit to your executive's office for a chat regarding an outage is greatly reduced. HOWEVER, there are some issues to deal with:**
  - Is the space you were provided sufficient?
    - Are there any unexpected increases that now makes it insufficient?
    - How do I track this now?
    - Who deals with these issues, the Storage Administrator or me?
  - OK, I am happy with the space I have, but am I getting the disk performance that will make my customer happy?
- **These issues are highly dependent on your relationship with your Storage Administration group.**

Now that my volumes are SMS managed, how can I tell what my volume names are, how much space I have left and how much space I have used?

First, find out what your Storage Group names are. Then, here are some things you can use to find out what the volume names are:

ISMF, option 6 - Storage Group, enter the Storage Group, and enter LISTVOL for a command. You must have Storage Administrator rights in ISMF to do this!

VOLUME	FREE	%	ALLOC	FRAG	LARGEST	FREE
SERIAL	SPACE	FREE	SPACE	INDEX	EXTENT	EXTENTS
-(2)--	---(3)---	(4)-	---(5)---	-(6)-	---(7)---	--(8)--
CBSM01	2598244	94	173256	71	2112449	17
CBSM02	2408773	87	362727	40	2223675	24
CBSM03	2566481	93	205019	39	2327430	16

If you know the volser, you can use ISPF 3.4 with option V. You will not see Storage Group information, just volume information:

```
Volume . : CBSM02

Unit . . : 3390

Volume Data          VTOC Data          Free Space   Tracks   Cyls
Tracks . : 50,085    Tracks . :          12   Size . . : 43,530  2,898
%Used . :          13    %Used . . :          12   Largest . : 40,185  2,679
Trks/Cyls:          15    Free DSCBS:    531   Free
                               Extents . :          24
```

You can also use DCOLLECT through ISMF or execute IDCAMS. A sample REXX for the output is available in ISMF under “Enhanced ACS Management” then under “SMS Report Generation”.

# Any other way I can figure out what volsers are in my Storage Group and what LPARs they are attached to?

If you have SDSF authority to issue MVS commands:

```
D SMS,SG(CB390),LISTVOL
```

```
IGD002I 10:54:27 DISPLAY SMS 404
```

```
STORGRP  TYPE      SYSTEM= 1 2 3 4 5 6 7 8
CB390     POOL          + + + + + + + +
```

```
VOLUME   UNIT      SYSTEM= 1 2 3 4 5 6 7 8      STORGRP NAME
CBSM01   9025          + + + + + + + +      CB390
CBSM02   9034          + + + + + + + +      CB390
CBSM03   9035          + + + + + + + +      CB390
```

```
***** LEGEND *****
```

```
. THE STORAGE GROUP OR VOLUME IS NOT DEFINED TO THE SYSTEM
```

```
+ THE STORAGE GROUP OR VOLUME IS ENABLED
```

```
- THE STORAGE GROUP OR VOLUME IS DISABLED
```

```
* THE STORAGE GROUP OR VOLUME IS QUIESCED
```

```
D THE STORAGE GROUP OR VOLUME IS DISABLED FOR NEW ALLOCATIONS ONLY
```

```
Q THE STORAGE GROUP OR VOLUME IS QUIESCED FOR NEW ALLOCATIONS ONLY
```

```
> THE VOLSER IN UCB IS DIFFERENT FROM THE VOLSER IN CONFIGURATION
```

```
SYSTEM 1 = SYSA      SYSTEM 2 = SYSB      SYSTEM 3 = SYSC
```

```
SYSTEM 4 = SYSD      SYSTEM 5 = SYSE      SYSTEM 6 = SYSF
```

```
SYSTEM 7 = SYSG      SYSTEM 8 = SYSPLEX1
```

How many Storage Groups should I have in my production environment? The answer of course is: It depends! Here are my recommendations - page 1 of 2

- **Production should have its own storage group for DB2 table spaces and indexes separate from all other environments.**
  - Start with at least 3 separate Storage Groups - if you will use the DB2 V8 BACKUP and RESTORE SYSTEM (requires z/OS 1.5) you are required to have separate Storage Groups for the BSDS and active log data sets from all other data. The three categories are:
    - DB2 Catalog and Directory objects
    - BSDS and active log data sets
    - DB2 user data
  - Discuss ICF catalog placement strategies with your Storage Administrator in regards to different recovery scenarios.
  - Place the sort (DSNDB07) data sets on a separate volume from the above if not using PAV and/or MA.

## How many Storage Groups should I have in my production environment? The answer of course is: It depends! Here are my recommendations - page 2 of 2

- Newer disk devices that use PAV, MA., etc. typically do not require data and indexes on separate volumes. Older versions on the other hand, may benefit by the separation (still watch for hot spots). For separate Storage Groups:
  - Use a unique data set naming convention separating data and indexes which will be resolved by the Storage Group ACS routine for correct placement.
  - Use ZPARM values for SMSDCFL and SMSDCIX, in macro DSN6SPRM. These hidden ZPARM values provide SMS with one Data Class for data (SMSDCFL) and another for indexes (SMSDCIX) which will be resolved by the Storage Group ACS routine for correct placement.
- Provide a Storage Group for your archive log data sets.
  - Recovery using archive log from disk is much faster than tape for parallel recoveries where several logs reside on the same tape. The exception to this is when archive logs are stored on DFSMSHsm tapes (aside from recall time). In this case, make sure the Storage Group can handle additional logs.
- Provide a Storage Group for image copy data sets. Recoveries from image copy data sets residing on disk will be much faster for parallel operations because there is no need to serialize image copy data sets if stacked on the same tape.
- Determine realistically how many archive log and image copy data sets are required for recovery situations and size the volumes in your Storage Groups accordingly.

How many Storage Groups should I have in my non-production environment? The answer of course is: It depends! Here are my recommendations:

- **For non production, it depends on the requirements for:**
  - **Performance**
  - **Backup and recovery**
  - **Types of environments, i.e. – sandbox, development, test, ERP and non-ERP, etc.**
    - **Amount of data in each environment**
- **Separation of environments will depend on business requirements. You may want to stay with the production strategy or you may want to combine some or all of the Storage Groups for the above environments.**

## Consolidating to large volumes - mod 27 or 54

- **My Storage Administrator just told me that they are putting in mod 54s, and instead of my current 15 mod 3 volumes, they are trading me up to one volume with the capacity of a little bit more than 19 mod 3s. That sounds like a great deal, is there anything I need to consider?**

- Although VSAM data sets can have up to 255 extents (increased in z/OS 1.7), there is a limitation of 123 extents per volume. This means that extending past 123 extents will not be possible with just one volume, so you will never grow beyond this point. This is true for other object types as well, such as extended format sequential data sets, which similar to VSAM will allow 123 extents per volume.
- Are you currently backing up (full volume dump) your volumes? There will be a lack of parallelism. There will be one very large dump instead of up to 20 dumps in parallel. How long will this function take? This is true for other operations, such as DEFRAg, etc.
- Consider the amount of time it takes to FlashCopy a volume. Again, similar to the dump issue, your copies will not be in parallel.

## Did that DEFrag allocation example work or fail?

### ▪ Back in the DEFrag section, we had an example:

```
--// DSN=TABLE.SPACE.IC,  
DISP=(NEW,CATLG),VOL=SER=PTS002,SPACE=(CYL,(250,25))
```

–Volume PTS002, has 462 free cylinders. However, the largest free extent is only 27 cylinders, and the remaining free extents are the same size or smaller.

### ▪ Results:

–Non SMS allocation: It failed! Dealing with the 5 extents for primary allocation rule, even if there are 5 extents of 27 cylinders, the result would be:

– (5 extents \* 27 cylinders = 135 cylinders) for the primary allocation, almost half the space required. DEFragging this volume would probably combine enough free extents to accommodate the space request.

–SMS allocation - it depends (Space relief is not valid for multi striped data sets):

- Failure again if the guaranteed space attribute is set to YES with no space constraint relief. Same scenario as above.
- Success when the guaranteed space attribute is set to YES with space constraint relief on. However, you may run out of extents trying to get the 250 cylinders if the allocation is for a non EF data set.
- Success when the guaranteed space attribute is set to NO and there are other volumes in the Storage group that can accommodate the allocation, otherwise failure.



## How did these space allocations happen?

- **I created an image copy data set as SPACE=(CYL,(1000,100)), I expected 1400 cylinders and 5 extents when multi volume, and I got 2300 cylinders with 5 extents, what happened?**
  - Did your data set use the guaranteed space attribute? Yes, well SMS is working as designed, since in this scenario the primary allocation is propagated to the next volume when it went multi volume. Extents are as follows - VOL1=1000,100 VOL2=1000,100,100
- **I image copied a 800 track table space, by mistake the output was TRK(1,1), but the allocation worked, how did that happen?**
  - Did your Storage Administrator assign EF for your image copy data sets? So long as there are at least 7 volumes with enough free space, the data set can spread up to 123 extents (as opposed to 16 extents non EF) on each volume. The end result will be a sequential data set with 800 extents. This is not a great way of doing business, but there is no outage.
- **\*\*\*\* NOTE – Guaranteed space will propagate the primary allocation to all multiple volumes for sequential and user managed DB2 LDSes. DB2 managed LDSes will propagate the secondary value to candidate volumes and not the primary.**

My Storage Administrator is seeing a lot of disk write bursts. Is there anything I can do to help?

- **Your Storage Administrator will see this in an RMF Cache Volume Detail report as “DFW Bypass”. For newer DASD, this is actually a DASD Fast Write retry and no longer a bypass.**
- **What this means is that NVS (Non Volatile Storage) is being flooded with too many writes. The controller will retry the write in anticipation that some data in NVS has been offloaded.**
- **For RAMAC devices the solution was to lower VDWQT to 0 or 1.**
  - This will cause high DBM1 SRM CPU time
  - May no longer be needed for ESS and DS8000 devices. Test and verify settings.
- **For very large buffer pools with many data sets, consider lowering VDWQT to 0 or 1.**
  - May work well for ESS and DS8000 as well. The tradeoff is still higher DBM1 SRM CPU time.
  - Test and retest! Validate such things as Class 3 times.

## REORG of LOBs

- **I have a LOB that is in many extents, I REORGed it and there was no space reduction. Is there a problem with z/OS or DB2?**
  - REORG of LOBs does not redefine your LOBs
  - There are only 3 phases that get executed - UTILINIT, REORGLOB, and UTILTERM.
  - REORG is done in place. It does not unload and reload any data. What does happen is that REORG removes imbedded free space, and attempts to make LOB pages contiguous.
  - REORG of LOBs help increase the effectiveness of prefetch.
  - Recommendation - After the REORG is complete, stop the object and use a dfp utility for extent reduction. E.g.: Use DFSMSHsm to migrate and recall the LOB. Verify the number of extents, then start the object.
  - \*\*\* NOTE – Resolved in DB2 9.

## I have heard about PDSEs, do I need them and what are they? Do they need to be SMS managed?

- **PDSEs are SMS managed data sets (non SMS possible, partial APAR list - OW39951)**
- **Before DB2 V8, there were no requirements for PDSEs. The following are required starting with V9:**
  - ADSNLOAD
  - ADSNLOD2
  - SDSNLOAD
  - SDSNLOD2
- **Recommendation: Use PDSEs for load libraries that contain stored procedures. This reduces the risk of out of space conditions due to adding or updating members. This is true for DB2 V7 as well.**
- **PDSEs are like a PDS, but much better!:**
  - Up to 123 extents (instead of 16 for PDSes). Cannot extend beyond one volume.
  - Number of directory blocks are unlimited.
  - Does not require periodic compression to consolidate fragmented space for reuse.
  - There is no need to recreate PDSEs when the number of members expands beyond the PDS's available directory blocks.

```
ISPF 3.4 data set listing for DB2 V8
Data Set Name . . . . : DSN810.SDSNLOAD
Organization . . . . : PO
Data set name type : LIBRARY
```

How did this happen? I have 90 cylinders primary, 12 cylinders secondary, 63 extents, but 25530 tracks allocated?

- **CREATE TABLESPACE PRIQTY 64800  
SECQTY 8640**
- **After some time and space ... ALTER  
TABLESPACE SECQTY 20000**
- **Add more rows, trip extents**
- **No REORG afterwards**
- **DB2 information correct**
- **MVS information is not!**
- **CYL(90,12) with 63 extents**
  - should be 12510 tracks
  - exception - fragmentation
  - BEWARE Storage Administrators! What you see is not always what you get!

```

ISPF 3.4 listing
-----
RI1.DSNDBD.A020X7KR.CKMLPR.I0001.A001  25530  63  3390

LISTCAT output:
ALLOCATION
      SPACE-TYPE-----CYLINDER      HI-A-RBA-----1254850560
      SPACE-PRI-----90      HI-U-RBA-----1113292800
      SPACE-SEC-----12

primary+(secondary*(extents-1))=space
90+(12*(63-1))=834 cylinders, 12510 tracks, not 25530!

```

This case is not a disk fragmentation issue. After the ALTER, DB2 knows the allocation converted to CYL(90,28). However, MVS still thinks it is CYL(90,12) until redefined by a process such as REORG without a REUSE. PRIQTY and SECQTY is actually what DB2 uses, not CYL(90,12).

Space related information not make sense? How did this happen?

Primary 25 cylinders, secondary 1 cyl, total 70 cyl, extents=3

- **Some reasons this can happen:**
  - Part of the data set is allocated on an EAV in the cylinder managed area.
  - SMS adjacent extent reduction
  - DB2 sliding secondary
  - DB2 ALTER executed increasing the space without running REORG or LOAD REPLACE to recreate the data set
  - DEFRAg was run with the CONSOLIDATE keyword
  - dss was run with the CONSOLIDATE function
- **What really caused this? There is no easy way to tell. It can be a combination of factors above.**
  - 32K objects will throw off numbers as well, but will not cause the above issue.

# Want to see what the secondary really was?

## ■ Using OMPE:

```
DB2PM GLOBAL(TIMEZONE(+7:00)
INCLUDE (IFCID(258)))
RECTRACE TRACE LEVEL(LONG)
EXEC
```

```
START TRACE (PERFM)
CLASS(30)
IFCID(258)
```

IFCID 258 is already on with statistics class 3, if class 3 is not on you can limit the tracing by specifying AUTHID, PLANID, etc.

PRIQTY=720 (1 cylinder) 180\*4. SECQTY=1440 (2 cylinders) 360\*4

High allocated before – 1 cylinder (180\*4=720), after - 3 cylinders (4\*(180+360)=2160)

2GB data set.  $524288 * 4096 = 2147483648$

DATA SET NAME	: DB9AU.DSNDBD.DSNDB04.JOHNEXT.I0001.A001	TIMESTAMP	: 06/30/10 19:21:10.736508
DATABASE NAME	: DSNDB04	DBID	: 4
TABLESPACE NAME	: JOHNEXT	PSID	: 562
PRIMARY QUANTITY	: 180	SEC. QUANTITY	: 360
HIGH ALLOC BEFORE	: 180	HIGH ALLOC AFTER	: 540
EXTENTS BEFORE	: 1	EXTENTS AFTER	: 1
VOLUMES BEFORE	: 1	VOLUMES AFTER	: 1
		MAX DS SIZE	: 524288
		MAX EXTENTS	: 251
		MAX VOLUMES	: 59

SMS managed data set – adjacent extent reduction

DFP max, not current data set

Sliding secondary data set with adjacent extent reduction. Started out with a 1 cylinder data set with 1 extent, took an extent (equivalent to extent 2 – 2 cylinders) and the total became 3 cylinders (logical extent 2).

QUANTITY values must be multiplied my 4

I understood the last chart, but how did I actually get LESS space than I requested?

- **When your Storage Administrator has set up a Data Class with the following attributes:**
  - **The space constraint relief attribute on, and with the request for a percentage for space reduction, your data set allocated can actually be less than requested . This can happen if your volume does not have enough space.**
  - E.g., 4K object, created with PRIQTY 72000 (100 cylinders), the Data Class space constraint was set up to allow 10% reduction, you had one volume with 92 cylinders remaining.  
Results:
    - The DB2 catalog will still show the equivalent of PRIQTY 72000.
    - The actual MVS allocation will be 90 cylinders or the equivalent of PRIQTY 64800.



When Storage Administrators talk about catalogs, are they talking about the DB2 catalog?

- **Generally, the answer here is no.**
- **Storage Administrators view the term catalog as the ICF catalog which they typically maintain.**
- **Make sure that when you or your Storage Administrator use the term catalog it is specifically stated which one, this will avoid needless confusion, errors, and arguments.**

## Storage Administrator should know about

- **With DB2 V8 the number of allowable partitions grew from 254 to 4096. The LLQ has the following pattern:**
  - A001-A999 for partitions 1 through 999
  - B000-B999 for partitions 1000 through 1999
  - C000-C999 for partitions 2000 through 2999
  - D000-D999 for partitions 3000 through 3999
  - E000-E096 for partitions 4000 through 4096
- DB2 9 allows cloned tables. This means that the fifth qualifier has added a new number. The fifth qualifier can now be I0001, I0002, J0001, or J0002. Note – REORG fast switches are not allowed for objects with cloned tables.
- With DB2 V9 NFM, implicit databases can be created with the range of DSN00001 to DSN60000. Depending on the CREATE statement either DSNDB04 will still be used, or the range for DSN00001 to DSN60000. This will be the third qualifier in the data set name.

## New RTS info

- **If you are using the RTS for your utilities or reporting, keep in mind that in DB2 9:**
  - The RTS has now moved into the DB2 catalog (DSNDB06.SYSRTSTS)
  - SYSIBM.TABLESPACESTATS is now SYSIBM.**SY**STABLESPACESTATS
  - SYSIBM.INDEXSPACESTATS is now SYSIBM.**SY**SINDEXSPACESTATS
  - Although Stored Procedure DSNACCOR can still be used, new SP DSNACCOX should be used instead.

My Storage Administrator tells me they are going to move disk volumes around during the day while DB2 is up. Don't I need to do online REORGs to do that?

- **Using online REORGs to accomplish this task can still be used, however, this can be very disruptive and time consuming for a DBA.**
- **There are some products on the market, such as Piper from IBM, that will actually do volume migration while DB2 is still up.**
- **Using Piper may be a less disruptive, time consuming, and error prone way of accomplishing the task that would have required additional Storage management changes followed by online REORGs. Using this type of product is a win - win for both the DBA and Storage Administrator.**
  - Some other products include TDMF from Amdahl (now IBM) and FDR/PAS

Since I am a DB2 professional, I get all of my space related information from the DB2 catalog. Should I consider something different?

▪ **It depends on what you are looking for:**

- If you are not executing frequent (perhaps daily) RUNSTATS or STOSPSPACE, then you could be looking at some very outdated information that you can not depend on.
- If you need more current information, consider using something like DCOLLECT and an interpretive report for the information as a replacement for what you are currently using. Samples are available for the DCOLLECT and report in ISMF. Use of processes outside of DB2 will reduce the stress on DB2, even when using shadow DB2 catalog information. It will also provide more up to date information.

– Review NaviQuest information as well. Refer to:

DFSMSdfp Storage Administration SC26-7402

Chapter 21. Using NaviQuest

DB2 RTS (Real Time Statistics) has relatively current information as opposed to stale information housed in the DB2 Catalog

My Storage Administrator told me they recreated my DB2 data sets with high extents. They are now as low as 1 extent per data set. Are there any issues?

- **Storage Administrators have a number of ways of causing extent reduction, thereby potentially bringing back a data set in 1 extent, among them:**
  - DFSMShsm MIGRATE and RECALL functions
  - DFSMSdss COPY, or DUMP and RESTORE functions
  - DEFRAG with the CONSOLIDATE keyword
  - dss with CONSOLIDATE operation
- **Using Such functions as DFSMSdss COPY may be much faster than running REORGS.**
- **Do you have SQL that uses the DB2 catalog to report on extents? This can potentially be a problem. You may be redoing REORGS unnecessarily since the move was done outside of DB2 and DB2 does not know about it. \*\*\* NOTE – the EXTENTS column in the RTS will only be updated once an update or applicable utility is run for the object. A simple start after extent reduction or a read based on SELECT will not update the EXTENTS column (same issue as the catalog).**
- **Do you use high extents as a tool to review issues with clustering indexes? This can potentially be a problem. Review CLUSTERING more closely.**

## Is it possible to lose just one or a few volumes with newer disk?

- **Yes, although it is extremely rare to lose just one or a few volumes instead of an entire disk controller's worth.**
  - **Although mean time between failure for disk devices is in the 1 million + hour range, there may be other factors causing disk failures, such as a bad disk controller card.**
- **Recommendation: Because of this (there are other reasons), I run a daily disk report by volume for my DB2 objects. Some things to think about if something does go wrong:**
  - What was on the volume I lost?
    - All indexes? Maybe I can just rebuild them
    - Part of one application or part of a partition data set, it MIGHT not be too bad then
    - My new mod 54 with ALL of my data? Find out what your alternatives are. There hopefully are some based on the architecture you have built in for this type of event.
    - Etc. - This gets into a much bigger discussion which we do not have time for now.

My index keeps on growing and tripping extents, even after deletes. What's wrong with DB2? How can I control the extent growth?

- **This one is actually a DB2 issue concerning pseudo deleted entries in your index, not a Storage management issue:**

- For an index, deleted keys are marked as pseudo deleted.
- Actual cleaning up will not occur except during certain processes. An example would be before a page split.
- High CPU cost of an index scan - Every time a SQL statement makes a scan of an index, it has to scan all entries in the index. This includes pseudo deleted entries that have not yet been removed.
- You can calculate the percentage of RIDs that are pseudo deleted based on the values of PSEUDO\_DEL\_ENTRIES and CARDF in SYSINDEXPART:
  - $(\text{PSEUDO\_DEL\_ENTRIES}/\text{CARDF}) * 100$
- Recommendation - REORG INDEX, if the percentage of pseudo deleted entries is greater than 10% for non Data Sharing and 5% for Data Sharing.

**\*\*\* NOTE** – Using procedures outside of DB2 for consolidation of space does not remove pseudo deleted data. You must use DB2 utilities in order to remove pseudo deleted data.



I hear that DB2 compression will reduce my disk space, and I will have better I/O and buffer pool hits. Do I now compress everything?

- **Do not compress small table spaces.**
- **Are you really tight on CPU? Keep in mind that compression will add a small amount of extra CPU cycles.**
- **Run DSN1COMP. Find out what you will be saving. If it is below 40%, it is probably not worth it.**
- **I found 5,000 table spaces I can compress, can I start now?**
  - DB2 V7 – No. You will hit a VSTOR issue. It is just too much to compress. Try to start out with a much smaller number and review your DBM1 VSTOR usage. The compression dictionary is still below the 2 GB bar.
  - DB2 V8 and above – Yes, only if you can fully back the compression dictionary with real storage. The compression dictionary is now above the 2 GB bar.

I hear about SnapShot, FlashCopy versions 1 and 2, and TimeFinder, what's the difference between all of these? (at a very high level)

- **SnapShot (RVA only)**

- SnapShot can quickly move data from the source device to the target device.
- Data is “snapped” (quickly copied) directly from the source location to the target location.

- **FlashCopy (ESS and DS8000) - versions 1 and 2**

- FlashCopy V1 requires the entire source volume and target volume to be involved in a FlashCopy relationship. FlashCopy V1 relationships do not allow any other FlashCopy relationships to exist on either the source or target volume.
- FlashCopy Version 2 enhances the FlashCopy function by providing an alternative method to copying an entire source volume to a target volume:
  - Multiple FlashCopy relationships are allowed on a volume.
  - Track relocation is possible because when copying tracks the target tracks do not need to be in the same location on the target volume as on the source volume.
  - The restriction that a FlashCopy target and source volume must be in the same logical subsystem (LSS) in an ESS is removed. However, FlashCopy must still be processed in the same ESS.
  - 10x faster than FlashCopy V1.

- **TimeFinder - EMC Hardware**

- similar in concept to FlashCopy for EMC
- for more information see, <http://www.emc.com/products/software/timefinder.jsp>

I have FlashCopy and/or SnapShot Technology, can it help if I want to clone my DB2s?

- **YES! Check out redbook: SAP on DB2 for z/OS and OS/390: DB2 System Cloning SG24-6287**
- **There are some products that will allow you to clone systems, such as Mainstar Volume Conflict Resolution (VCR). VCR allows you to clone data within the LPAR you are cloning from.**

What should I  
discuss with my  
Storage Administrator?

## Issues I should discuss with my Storage Administrator about my DB2 environment:

- **Who is the disk manufacturer for the volumes I use?**
- **Do I use ESCON or FICON and at what data transfer speed?**
  
- **What models of disk do I use?**
  
- **What type of 3390 do I emulate?**
  
- **How much disk do I have for each environment (# of volumes, and total GB)?**
- **How much cache does each disk box contain?**
  
- **If we buy more disk cache, how much performance will I gain?**
  
- **What type of RAID is my disk and what does it mean to me?**
  
- **What features are turned on for my disk?**
  - PAV (if available, dynamic or static)
  - FlashCopy (version 1 or 2)

## Issues I should discuss with my Storage Administrator about my DB2 environment:

- **Do you VSAM stripe my heavily sequential DB2 data sets (e.g. active logs)?**
- **Do you sequential stripe my applicable disk data sets that support my VSAM striped data sets?**
- **How often do you DEFRAg my volumes and based on what criteria?**
- **How well are my disks doing when reviewing RMF data during peak periods?**
- **For what DB2 LDSes do I have EF turned on?**
- **In the SMS Data Class, is 'Extent Constraint Removal' set to YES to exceed the 255 extent limit? Does it make sense to increase the limit?**
- **Is space constraint relief being used for my data sets? Include the advantage of 5 extent rule for allocations.**

## Issues I should discuss with my Storage Administrator about my DB2 environment:

- **Are my data sets created with the Guaranteed Space attribute or are the entries in the Separation Profile? If so, why? If you do have the attribute on, mention to your Storage Administrator that although on multi volume data sets the primary allocation is propagated to additional volumes, for DB2 managed data sets the secondary is propagated not the primary.**
- **Do you migrate any of my DB2 or related data sets, and if you do what are the Management Class attributes?**
- **Do I use the SMS Management Class to expire any of my data, such as my archive logs or image copies?**
- **If I archive my DB2 LDSes, what is my ZPARM value for RECALL and RECALLD?**
- **Do you full volume backup my volumes, and if yes, why?**
- **Do you incrementally backup my volumes, and if yes, why?**
- **In the SMS Storage Group, what are the values for HIGH and LOW and why were they set to those numbers?**
- **Do my DB2 data sets use Extend or overflow Storage Groups? If so, what are the VOLSERS? If we FlashCopy our volumes, we need to make sure these are included as well.**

## Issues I should discuss with my Storage Administrator about my DB2 environment:

- **If I have more than one disk controller, are my active log data sets segregated onto separate controllers for availability? Do you use the SMS Separation profile to accomplish this?**
  
- **For my archive log data sets, what units do I write to?**
- **How long do I keep them for? Is this a realistic number?**
- **What unit types do I use for my archive data sets?**
- **If I write to tape, what type of tape do I use? What does this tape type really mean to me? Also, what is the tape capacity?**
- **Is there a better technology to write my archive log data set to?**
  
- **How do you guarantee that my dual copied BSDS and archive log data sets do not wind up on the same volumes? Same question if you dual software copy your image copy data sets.**
  
- **What Storage Groups do I allocate my data sets on?**
- **What are their names?**
- **How many volumes do I have in each Storage Group?**



## Issues I should discuss with my Storage Administrator about my DB2 environment:

- **Are the specs different for my different Storage Groups?**
- **How can I tell how much space I have available in each of my Storage Groups?**
- **Do you make sure I have enough space in my Storage Groups or do I?**
- **Are you seeing write bursts in the RMF Cache Volume Detail report? If so, I can probably relieve this situation.**
- **Are you using some type of disk migration tool such as Piper, TDMF, or FDR/PAS?**
- **How are we handling excessive extents? Will you run processes to reduce them or do I?**

## Issues I should discuss with my Storage Administrator about my DB2 environment:

- **If my data sets are migrated, does it get migrated to disk or tape? If tape, how long will it take to retrieve a data set? Does the tape contain multi data sets? Is the tape manually mounted or automatic (robotic)?**
- **How often do the Storage Administrators and/or performance team review RMF data for DB2 disk? Is it often enough? Can we sit down and review the information together and have the reviewer explain what may be a potential problem?**
- **Be specific when discussing ‘the catalog’. Your Storage Administrator is thinking about the ICF catalog, not the DB2 catalog.**
- **For your DB2 LDSes, make sure your Storage Administrator does not set a value for "Volume Count" or "Dynamic Volume Count" in the Data Class panel of ISMF. We want DB2 to be able to allocate LDSes to DFP limitations, not artificial limitations set in SMS. Consider VC and/or DVC for user managed objects, such as the DB2 Catalog and directory.**

## Things that influence my DB2 allocation

- **DB2 or user managed data sets**
- **Volume fragmentation**
- **z/OS 1.5 and above – extent consolidation**
- **z/OS 1.7 and above – ability to exceed 255 extents for LDS**
- **Size of CA**
- **Use of parameters such as PIECESIZE, LARGE, and DSSIZE**
- **Use of ZPARM for:**
  - DSVCI
  - MGEXTSZ (sliding secondary)
  - Use of TXQTY/IXQTY

# Things that influence my DB2 allocation

- **Use of Extended Format (EF) data sets**
- **Use of Extended Adressable (EA) data sets**
- **Use of SMS Data Set Separation Profile**
- **Use of SMS Data Class for:**
  - DCB and space Attributes
  - Space constraint relief
  - Multi volume allocations
  - Exceed 255 extent rule
  - Stripe data sets
  - Compress data sets
- **Use of SMS Storage Class for:**
  - Guaranteed Space attribute
  - Multi-tiered SGs
  - VSAM and/or sequential striping
- **Use of SMS Management Class for:**
  - Expiration date for data sets
  - Migration and backup of data sets
  - Release of unused space

## Things that influence my DB2 allocation

- **Use of SMS Storage Group for:**
  - HIGH and LOW values
  - Use of extend or overflow Storage Groups
  - Reasons for allocation on volume – primary, secondary, tertiary, or rejected

## Storage related items that effect my DB2 performance

- **Type of disk**
- **Use of FICON or ESCON as well as channel transfer rate**
- **PAV (this also depends on if static or dynamic)**
- **Priority I/O Queueing**
- **Size of disk cache**
- **Number of paths for devices**
- **Disk volume/address architecture. Not “front loading the box”**
- **VSAM data striping**
- **Sequential data striping**
- **Allocation in tracks instead of cylinders (track penalty)**
- **MIDAW**
- **CI size (ZPARM DSVCI)**
- **Migration/recall of data**
- **Use of data in cache (ZPARMs SEQCACH, SEQPRES)**
- **FlashCopy V2 for CHECK INDEX utility**
- **Write bursts and use of buffer pools**

## Now it is time to sit down with your Storage Administrator for an exchange of ideas

- **Ask your Storage Administrator for a copy of their procedures as it pertains to DB2. Review the document **with** your Storage Administrator to understand the concepts.**
- **Discuss with your Storage Administrator your installation's mix of hardware and software. How do they work and how can they work best for DB2? Keep in mind such things as DR requirements which, for the most part, are not discussed in this presentation.**
- **Discuss with your Storage Administrator the concepts you have seen in this presentation. Are there any options you can take advantage of that are not currently being used?**
- **Exchange information with your Storage Administrator about how DB2 works and find out more about your storage. The more you know about each other's technology, the better DB2 and MVS in general will perform and continue their happy marriage together.**



**Dank u**

Dutch

**Merci**

French

**Спасибо**

Russian

**Gracias**

Spanish

شكراً

Arabic

감사합니다

Korean

Tack så mycket

Swedish

धन्यवाद

Hindi

תודה רבה

Hebrew

**Obrigado**Brazilian  
Portuguese**Dankon**

Esperanto

谢谢

Chinese

**Thank You**

ありがとうございます

Japanese

**Trugarez**

Breton

**Danke**

German

**Tak**

Danish

**Grazie**

Italian

நன்றி

Tamil

děkuji

Czech

ขอบคุณ

Thai

go raibh maith agat

Gaelic