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

DB2 Information Management Software

RICDUG – June 14, 2005

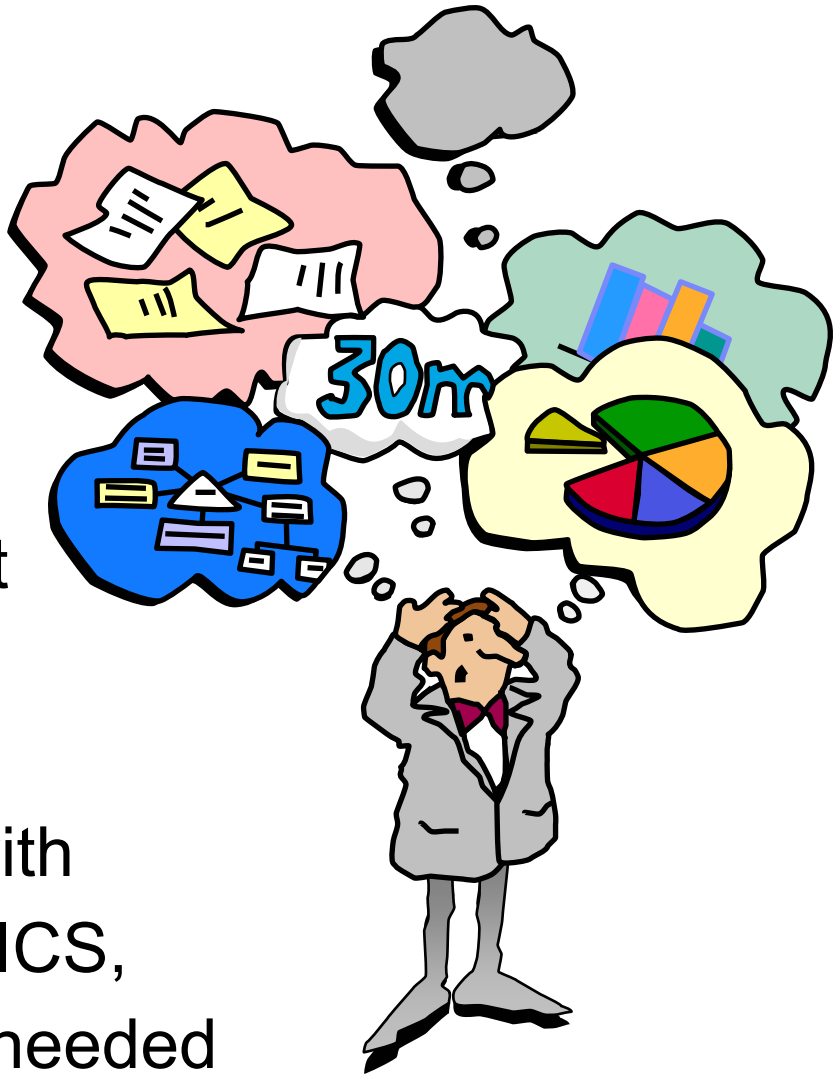
IBM storage

John Iczkovits
iczkovit@us.ibm.com

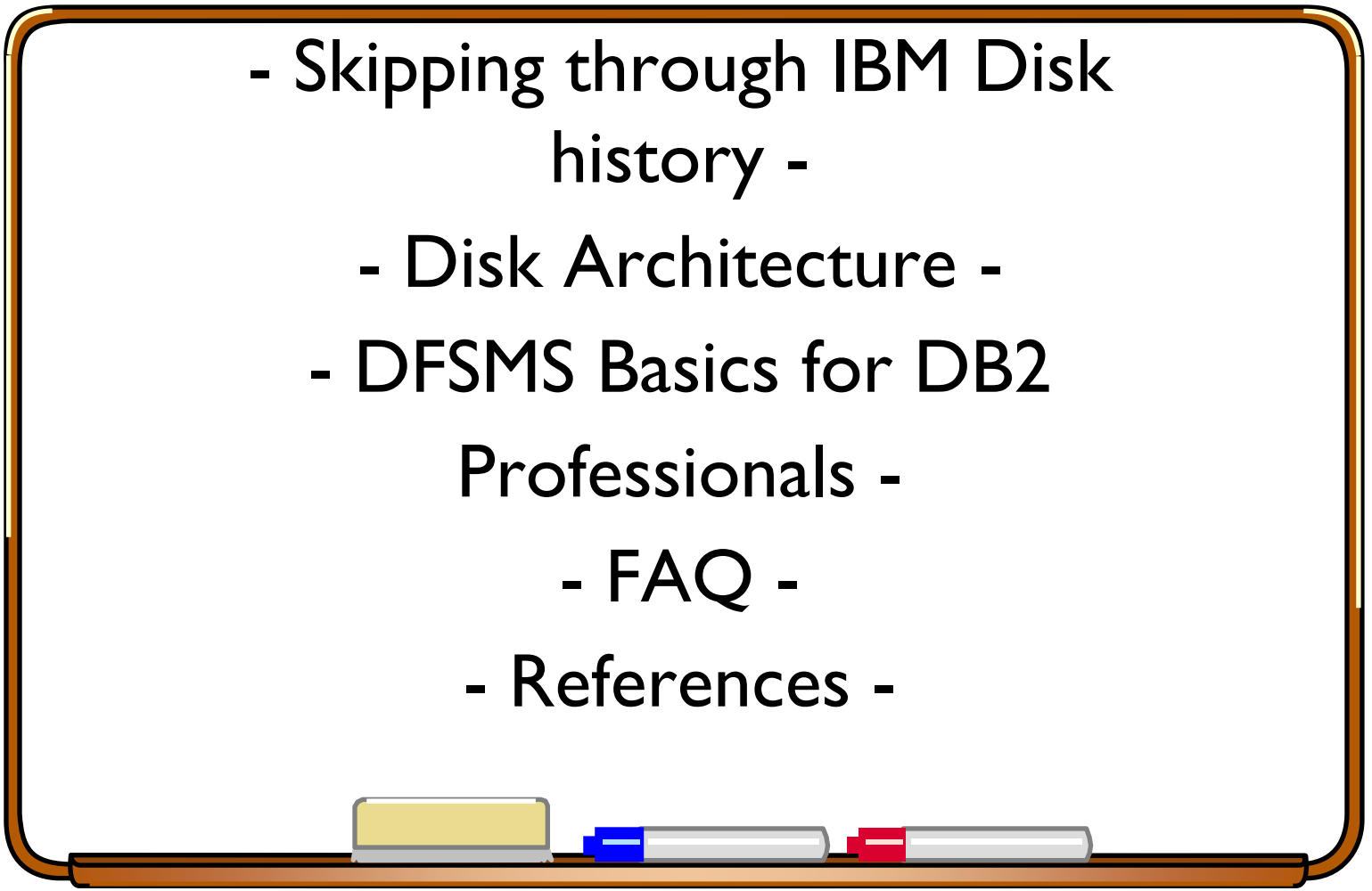
Advanced Technical Support
TECHNICAL SALES SUPPORT AMERICAS

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.

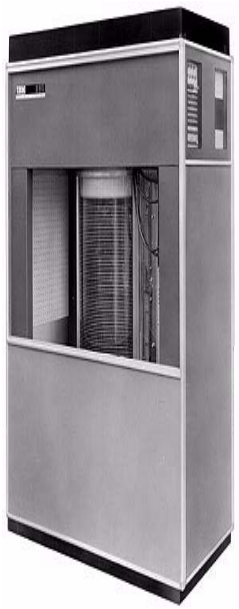


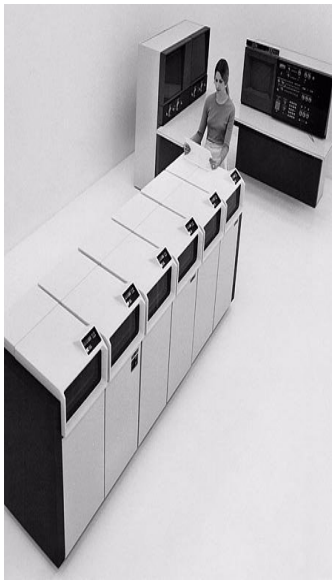
Agenda



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

1956 1965 1970 1975 1980 1989





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

IBM

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

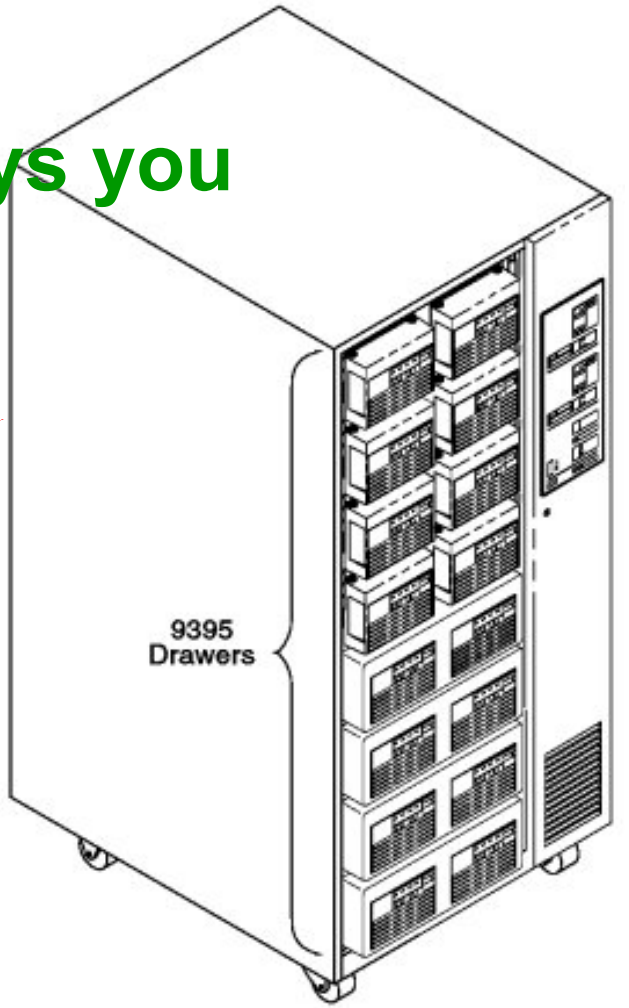
RVA = RAMAC Virtual Array (1997)

RVA - 1997

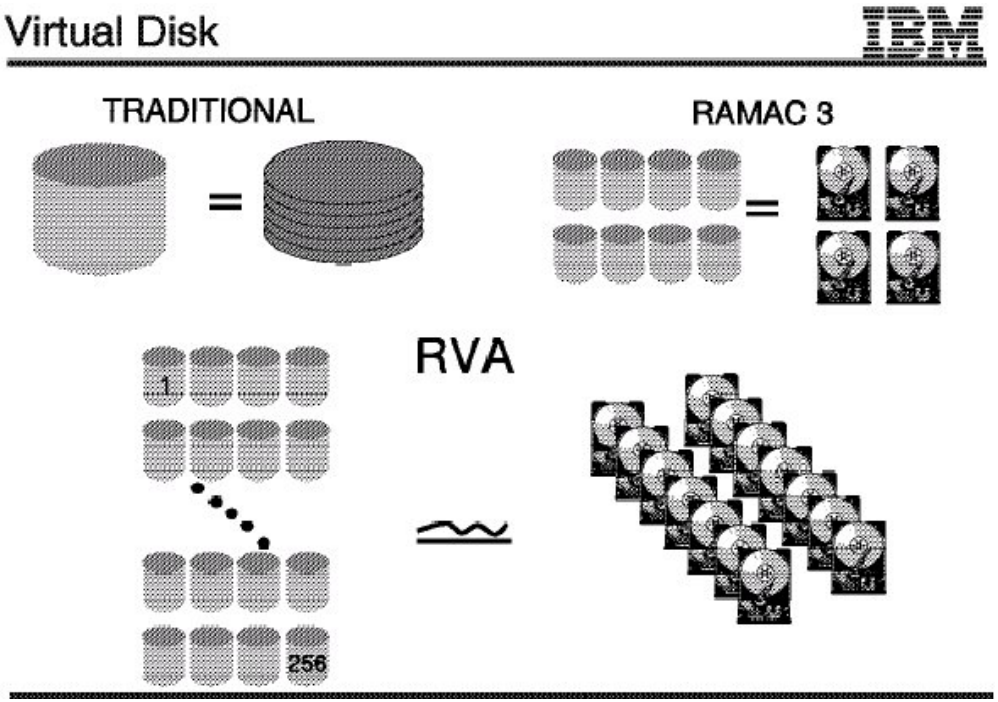
Up to 840 GB of information storage

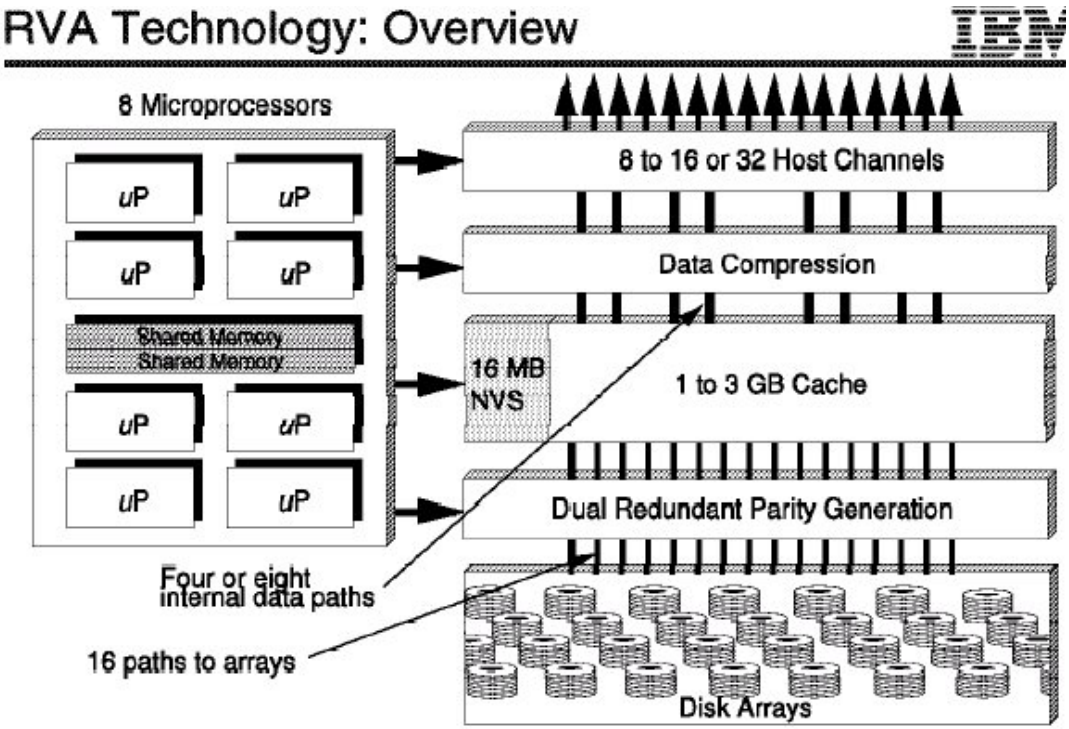
Virtual Disk Architecture - What it buys you

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



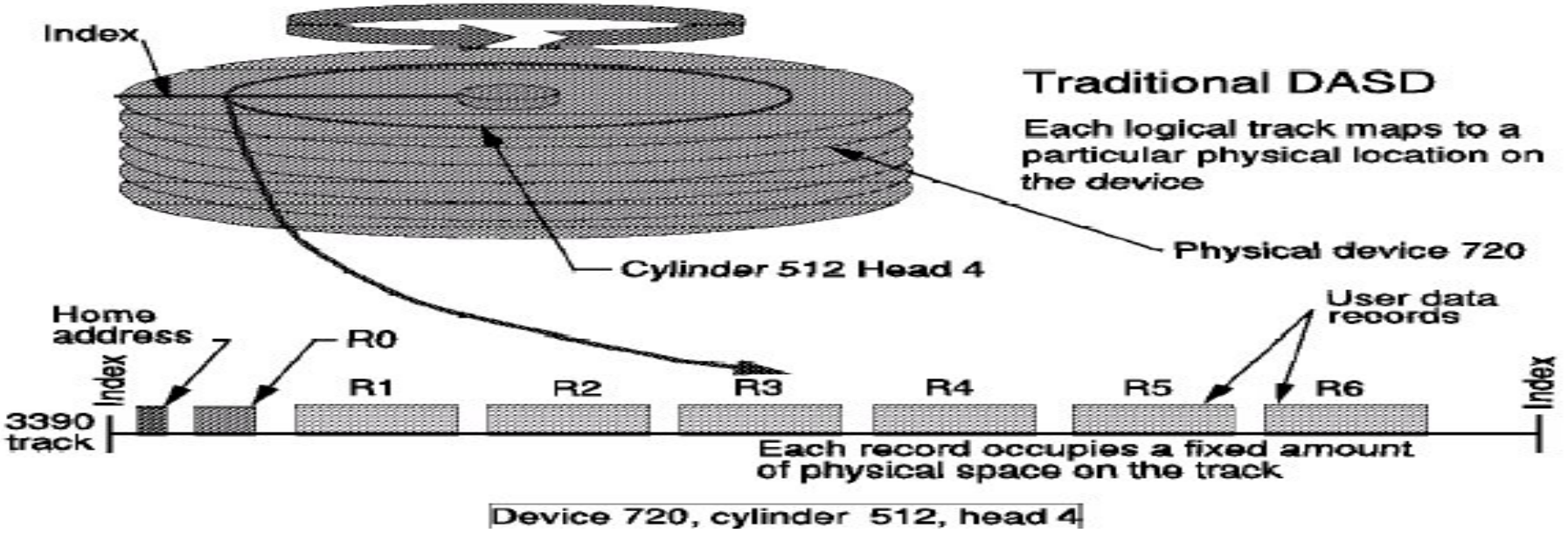
RVA - 1997



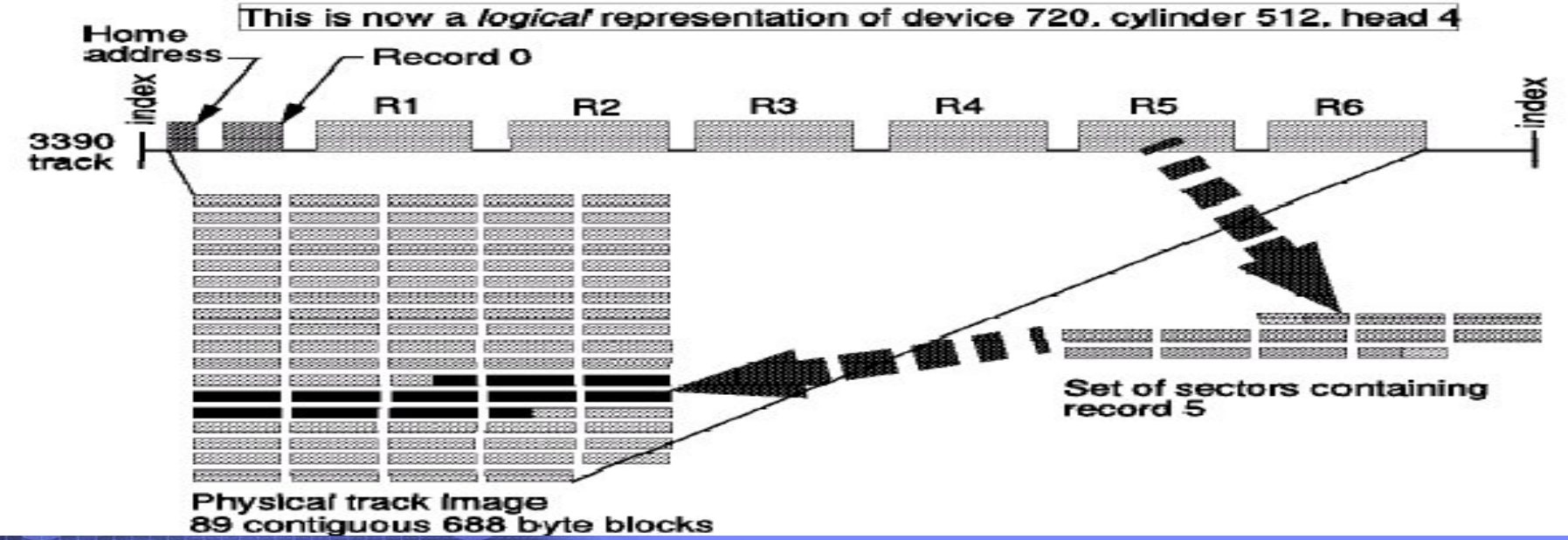


There is no fixed physical to logical mapping. The RVA dynamically maps functional volumes to physical drives. This mapping structure is contained in a series of tables stored in the RVA control unit.

Traditional DASD

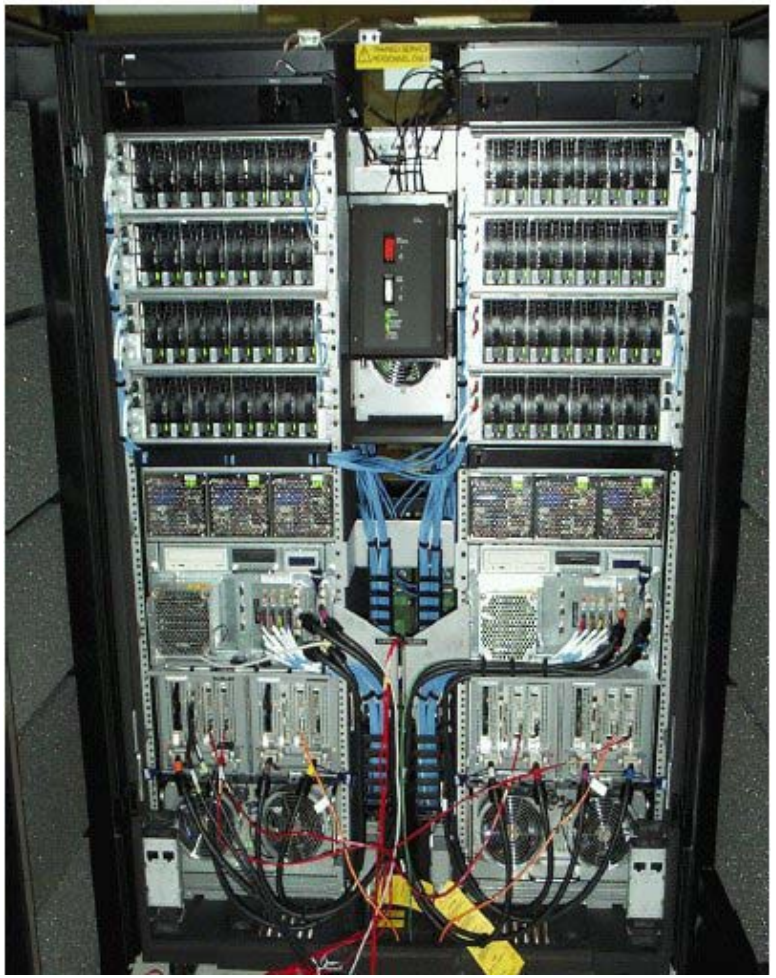


Array DASD



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

- Scalable from 420GB 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 from 1.1 TB up to 192 TB
- FlashCopy
- Supports PAV and MA
- With the implementation of the POWER5 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
- Sequential Prefetching in Adaptive Replacement Cache (SARC)
  - Sophisticated, patented algorithms to determine what data should be stored in cache based upon the recent access and frequency needs of the hosts
  - Prefetching, which anticipates data prior to a host request and loads it into cache
  - Self learning algorithms to adaptively and dynamically learn what data should be stored in cache based upon the frequency needs of the hosts
  - SARC provides up to a 25% improvement of the I/O response time in comparison to the LRU algorithm.

Disk Architecture

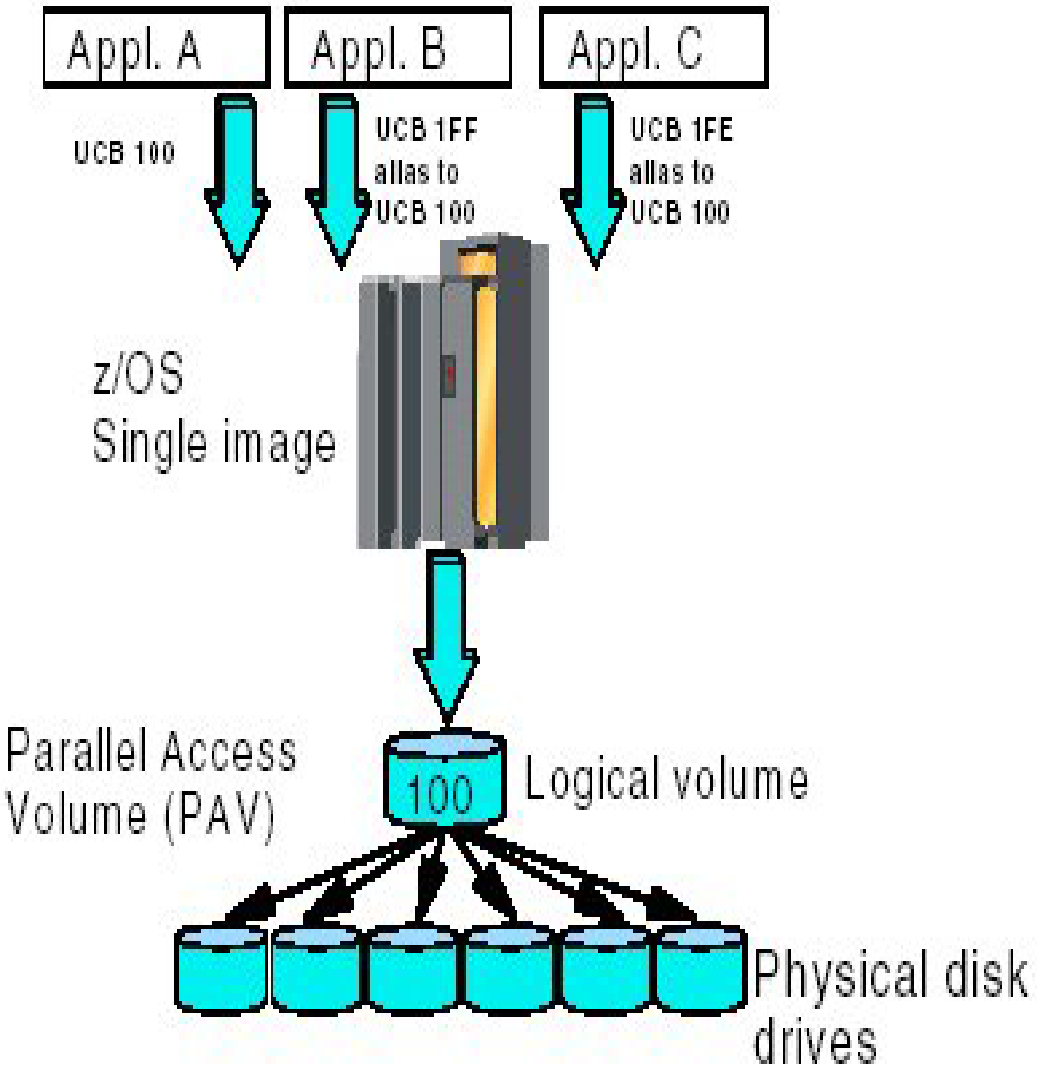
RAID (Redundant Arrays of Inexpensive Disks) Technology- Disk Type Dependent

- **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 - independent access with floating parity**
- **RAID 6 - dual redundancy with floating parity**
- **RAID 10 (DS8000 and some ESS) - RAID 0 + RAID 1, 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.

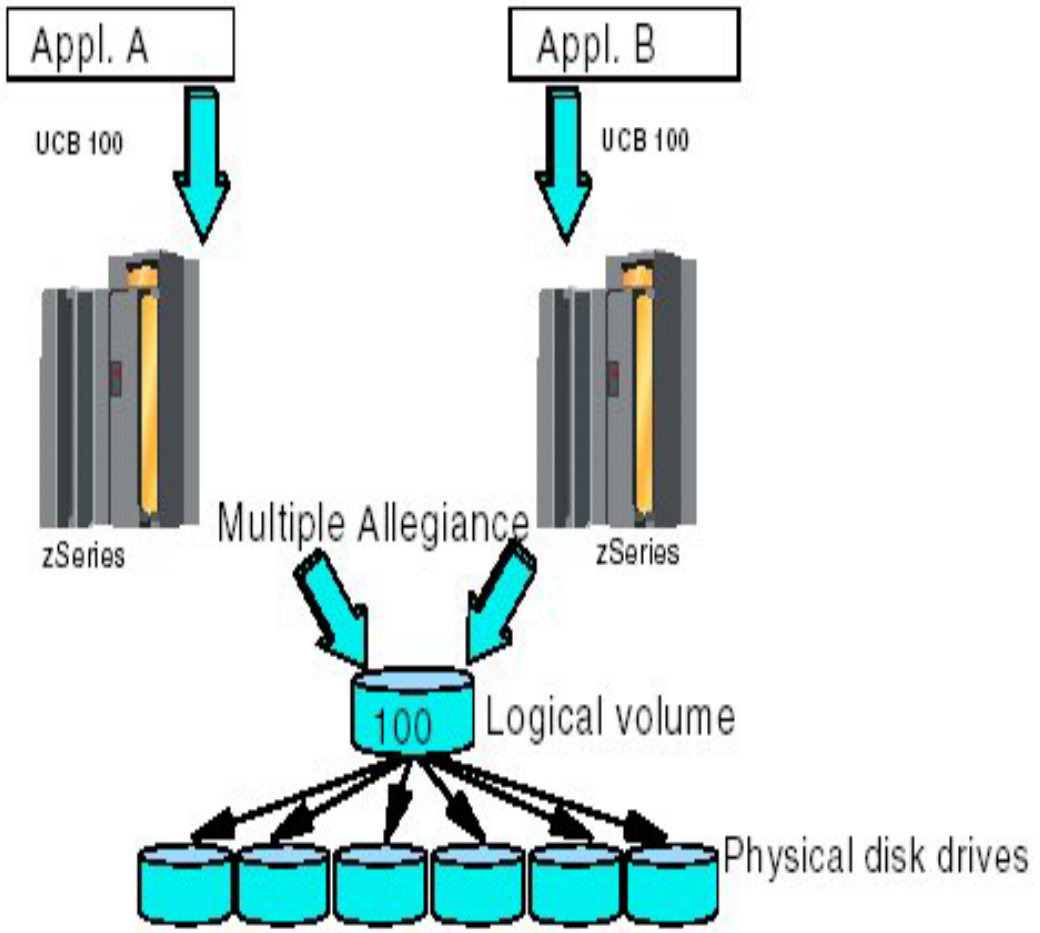
Parallel Access Volumes (PAV) - ESS "Shark" and DS8000

- **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.**



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

- **Similar to PAV, however for more than one LPAR**
- **Incompatible I/Os are queued in the ESS/DS8000**
- **Compatible I/O (no extent conflict) can run in parallel**
- **ESS/DS8000 boxes guaranty 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!

VSAM Data Striping - ESS, DS8000, and some RVA

- **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.**
- **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, etc.**

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.

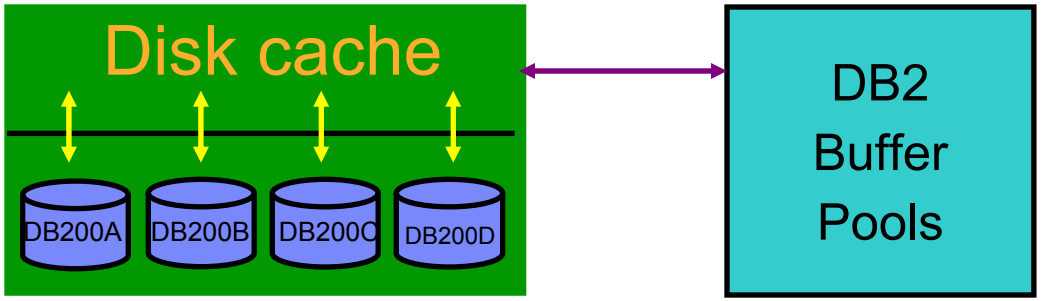
  - Periodically installations will review RMF data for disk performance. Discuss with your Storage Administrator and/or performance team (if one exists) how frequent periodic really is. You may find the answer to be daily, quarterly, or never. Never is not a good idea, quarterly may not be frequent enough.
  - 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?

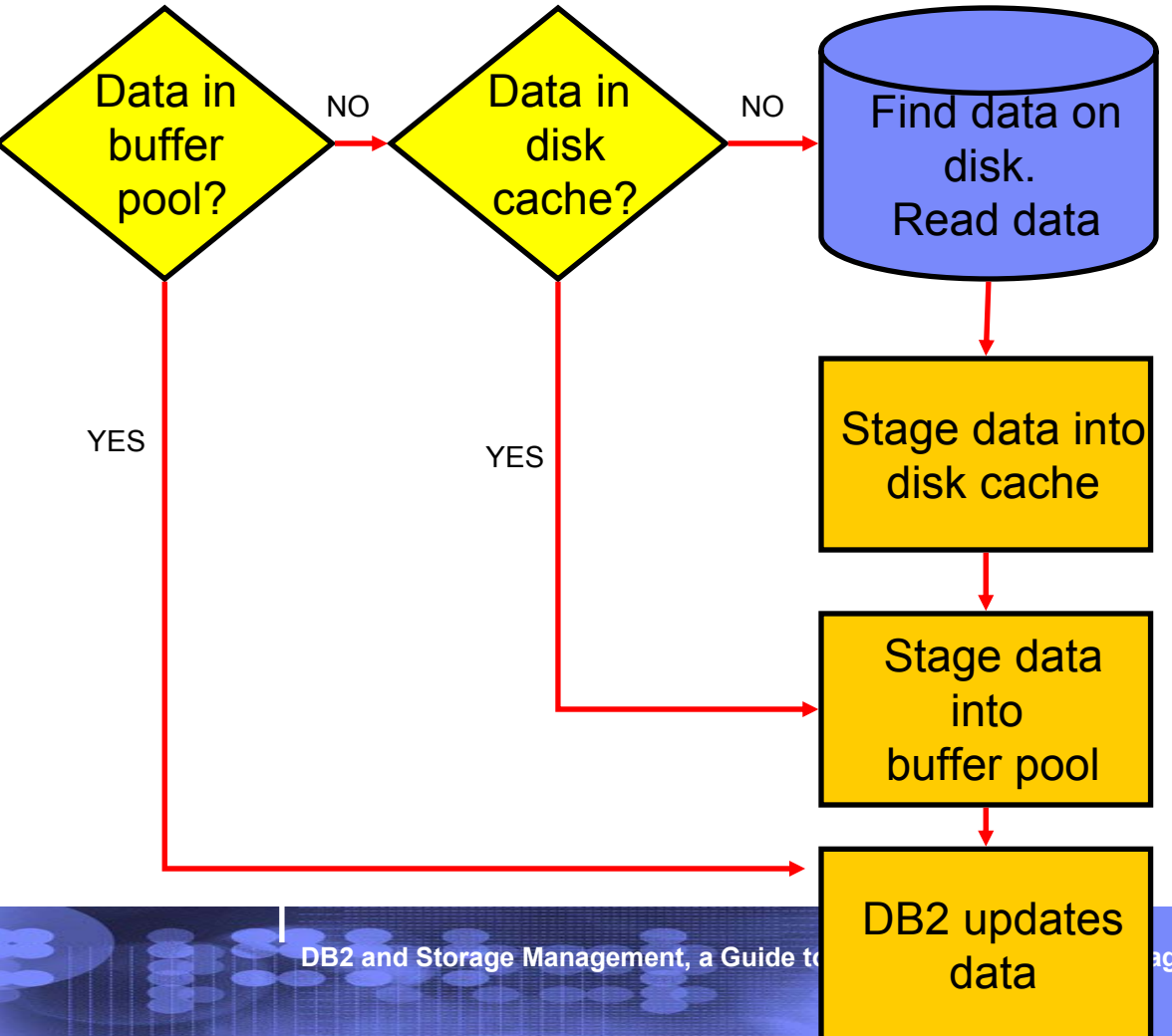
Disk Cache - Why Storage Administrators think we are unfriendly! page 1 of 2

- at a very high level –

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

```
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.
  - In DB2 V8. You can allocate very large buffer pools, as long as they are backed by real storage.

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 DUMP option

```
DSCB ADDR (CCHHR)
0001000004
```

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 - z/OS 1.6 DFSMSdss - 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 - z/OS 1.6 DFSMSdss - page 2 of 3

ISMF Volume option

VOLUME SERIAL	FREE SPACE	% FREE	ALLOC SPACE	FRAG INDEX	LARGEST EXTENT	FREE 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!

Volume Fragmentation - z/OS 1.6 DFSMSdss - page 3 of 3

- **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.
- **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.

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	PEND	DISCONNECT	CONNECT
Device is busy by this z/OS, no PAV UCB aliases are available	- Device reserved from another system - CMR (CoMmand Reply) delay - SAP overhead - Old causes	- Read cache miss -Reconnect miss (ESCON) - Synchronous remote copy - Multiple Allegiance or PAV write extent conflicts - Sequential write hits, rate is faster than controller can accept - CU busy	Channel data and protocol transfer

I/O Response time = IOSQ Time + Pending Time + Connect Time + Disconnect Time

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.

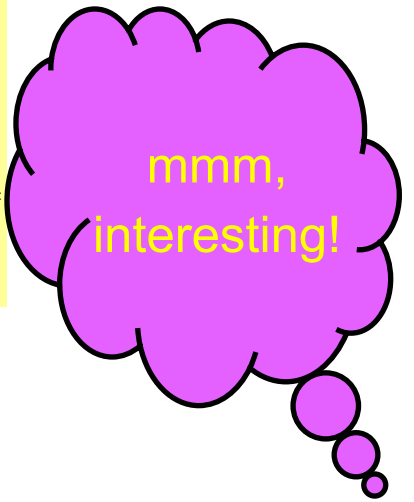
Connect: Time when the channel is transferring data from or to the controller cache or exchanging control information with the controller about one I/O operation.

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.

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 2ms	14ms	20ms	30ms
Ramac3, RVA2	0.6 to 0.9ms	6ms	20ms	30ms
ESS E20/ESCON	0.3 to 0.4ms	3ms	10ms	15ms
ESS F20/ESCON	0.25 to 0.35ms	2ms	10ms	15ms
ESS F20/FICON	0.13 to 0.2ms	1.5ms	10ms	15ms
ESS 800/FICON	0.09 to 0.12ms	1ms	5ms	10ms

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

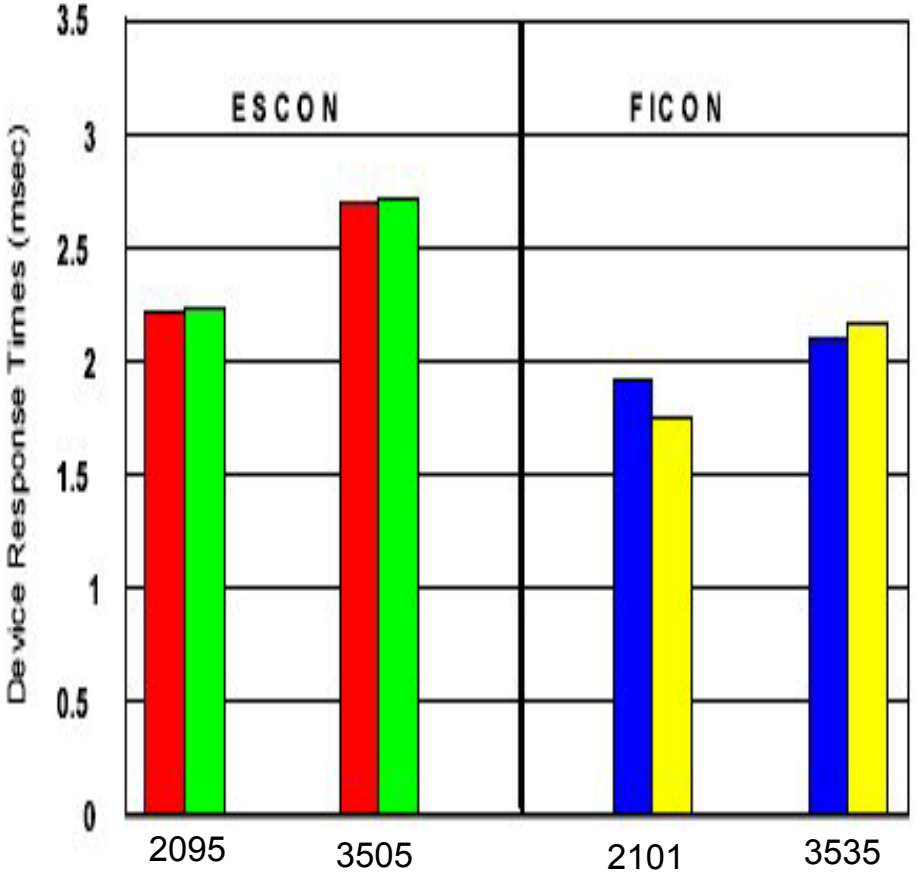
* Storage Administrators refer to 3390 mod 9, 27, and 54 as large-volume support 10017 cylinders, 32760 cylinders, and 65520 cylinders respectively. Mod 54 support can be found in the 2107device PSP bucket.

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

Why larger disk sizes?

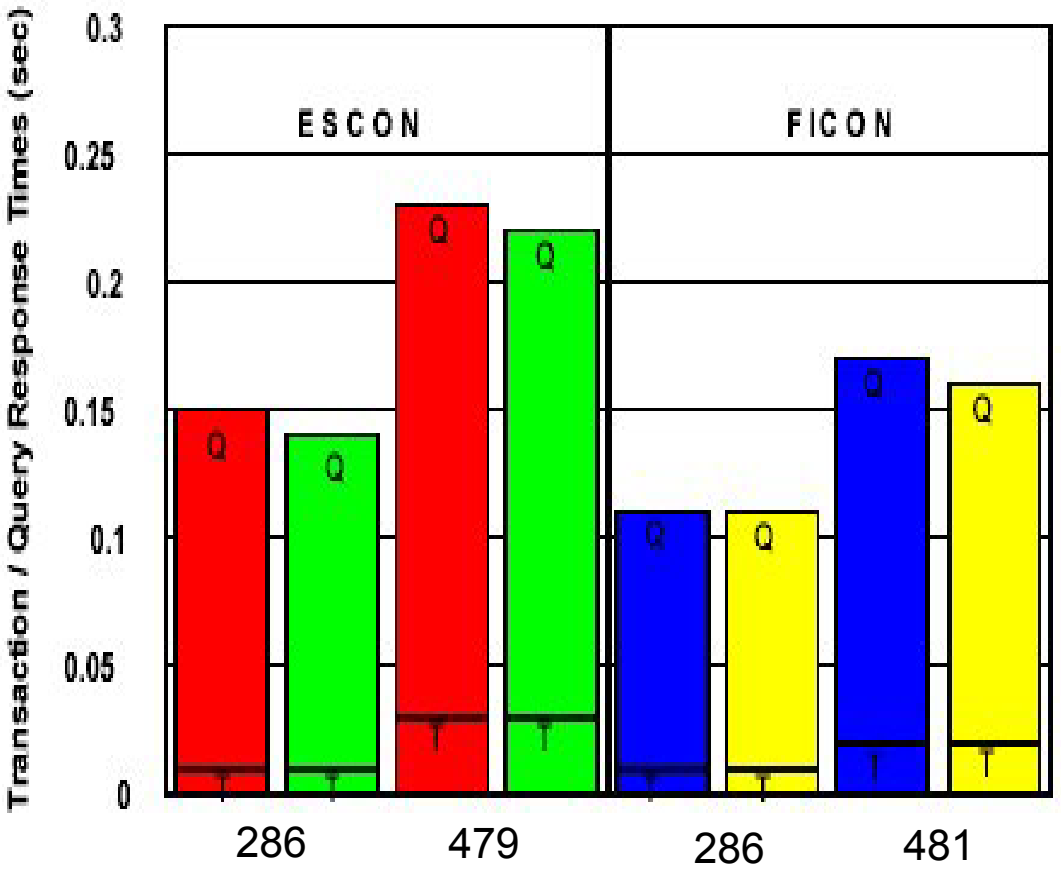
- reduces issues with MVS addresses (maximum number of devices met)
- simpler management of storage

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



Total I/O rates per second

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



Total Transaction rates per second

Q = Query response time
T = Transaction response time

Tests run using Dynamic PAV

Disk Space Allocations and Extents

- PRIQTY and SECQTY are based on KB, so we can specify the following - see next page (allocations are slightly higher when DSVCI=YES, and you are allocating a 32K CI):
  - 1 track = 48 (see previous page)
  - 1 cylinder =720 (15 tracks * 48 KB per track)
- 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
- 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 data set size - 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)
- 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))
- CREATE INDEX with the PIECESIZE keyword can allocate smaller data sets

Disk Space Recommendations (even with newest technology)

- **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 per component, but only up to 123 extents per volume per component
  - Striped VSAM data sets, up to 4080 extents per data component (16 stripes (volumes) max for VSAM*255 extents per stripe)
- **Allocate objects on cylinder boundary**
  - Improved SQL SELECT and INSERT performance
  - 2x improvement relative to track allocation
  - You will lose 16K for each track when a 32K table space is created in tracks and DSVCI=YES (DB2 V8). This is because the CA size in this case is one track and a CI can not exceed the size of a CA.
- **Extents still matter:**
  - For performance of objects with small secondary allocations (e.g. 1 track), increase the secondary quantity for small objects to avoid many multiple extents, which will cause significant performance penalties.
  - Avoid running out of extents for all size data sets
- **With DSVCI=YES**
  - splits 32 KB CI over two blocks (except for track allocation) of 16 KB in order not to waste space due to track usage (48 KB)
  - A 16 KB block is wasted 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

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)

Data Set Separation - Part of the SMS Base Configuration

- **Allows you to designate groups of data sets which are to be physically separated**
- **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.**
- **SMS attempts to allocate the data sets behind different control units**
- **A data set separation profile must be provided**
- **Cannot be used with non-SMS-managed data sets or with full volume copy utilities such as PPRC. Alternatives:**
  - Create the data set before PPRC is established
  - Break the pair after it is established, create the data set, then re-establish the pair

Interesting Data Class (z/OS 1.6) Information for DB2 Professionals

- **A collection of allocation and space attributes - for new data sets only, e.g. - LRECL, BLKSIZE, space, etc.**
- **Can be used for SMS or non SMS managed data sets**
- **For VSAM and sequential data sets**
- **Used for allocating extended addressable (EA) - allocate and access VSAM data sets more than 4 GB in size, extended format (EF - used in conjunction with EA), striped data sets, and PDSEs**
  - Recommendations - Have the Storage Administrator set only the required DB2 LDSes (table spaces and index spaces) to extended format, this will avoid any EF overhead if not required.
  - Sequential data sets do not have the same performance penalty, so you may want your Storage Administrator to set all allowable sequential DB2 data sets to EF. At this time. archive log data sets can not be EF managed, unless you use a tool, such as IBM's Archive Log Accelerator.
- **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.
  - 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**

Interesting Storage Class (z/OS 1.6) Information for DB2 Professionals - page 1 of 2

- **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.**
- **Defining Performance Objectives:**
  - MilliSecond Response (MSR) times
    - Each device type and model has a predetermined MSR capability
    - Use or non use of disk cache, and performance requirements if used
    - Sustained Data Rate (SDR) - extended format data sets using striping, for VSAM as well as physical sequential – the algorithm is for the number of stripes.
- **Defining Availability**
  - Determine if the data set is eligible for RAID and/or dual copy volumes
- **Defining Accessibility**
  - Point-in-time copy, using either concurrent copy, virtual concurrent copy, or FlashCopy (for ESS and DS8000)
- **Defining the Guaranteed Space Attribute (honor volser 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.

Interesting Storage Class (z/OS 1.6) 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.

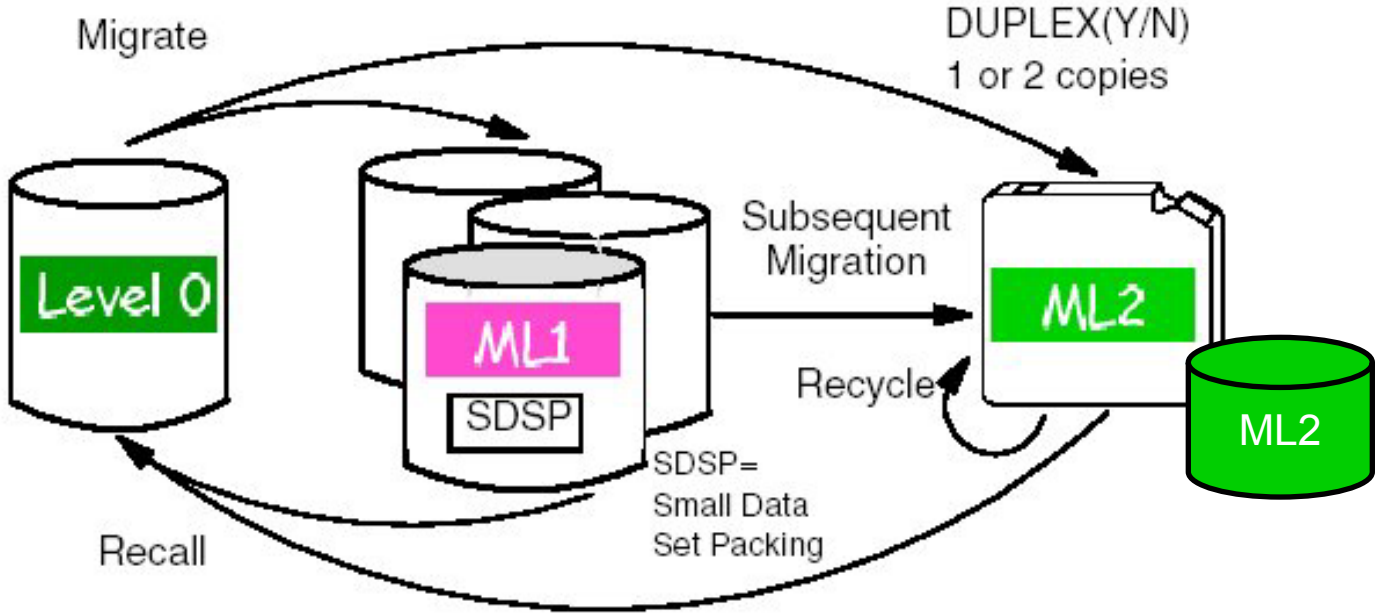
Interesting Management Class (z/OS 1.6) Information for DB2 Professionals - page 1 of 3

- **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 (z/OS 1.6) 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.
  - 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.
  - Match the SMS Management Class expiration date for DB2 archive logs with the ZPARM value for ARCRETN. DB2 does not automatically expire archive logs. Follow this recommendation only if another product, such as RMM is not already expiring the data for you.
  - Match the SMS Management Class expiration date for image copy data sets with the DB2 utility MODIFY DELETE date. DB2 does not automatically expire image copy data sets. Follow this recommendation only if another product, such as RMM is not already expiring the data for you.
  - Keep in mind DR situations! ABARS may be used to backup your application's ML1 and ML2 data sets if required.

Interesting Management Class (z/OS 1.6) 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.

Interesting Storage Group (z/OS 1.6) 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 (z/OS 1.6) 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 primary 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

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

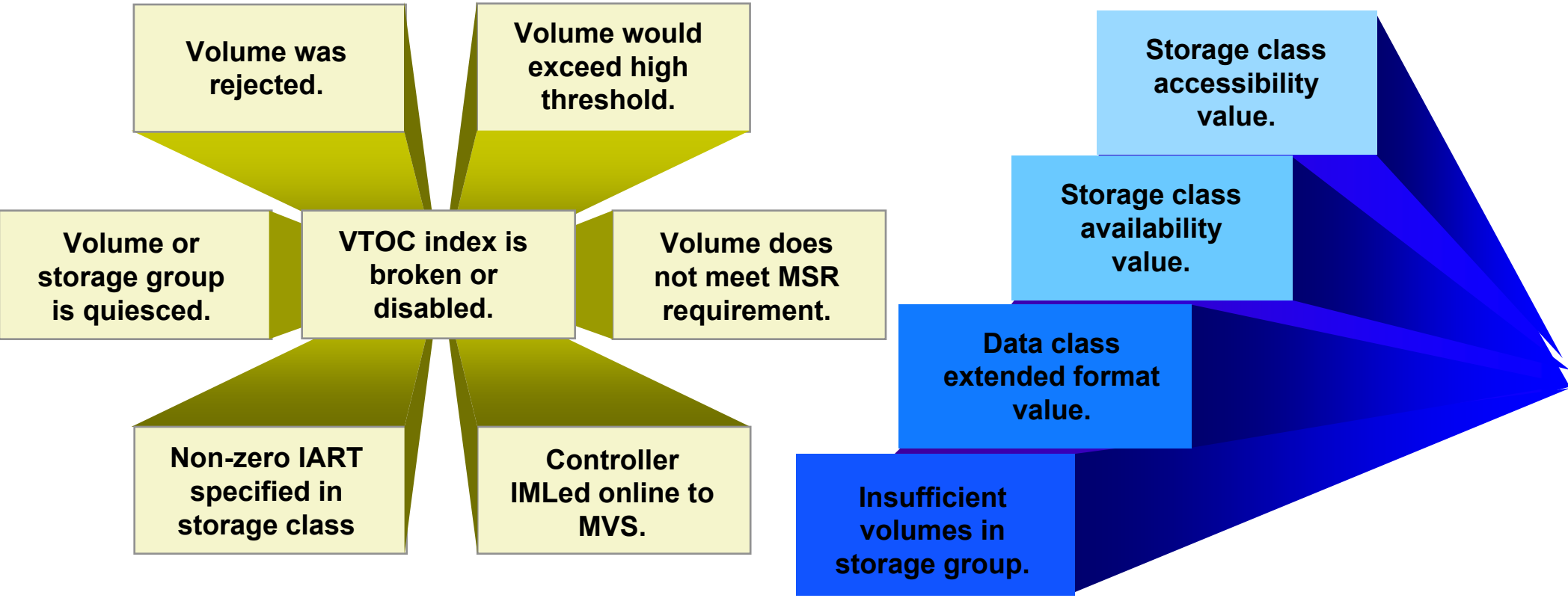
Interesting Storage Group (z/OS 1.6) Information for DB2 Professionals - page 3 of 3

- **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 extended SG name can also be an overflow Storage Group.
- **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 extended storage group.
- Discuss with your Storage Administrator if using extended 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

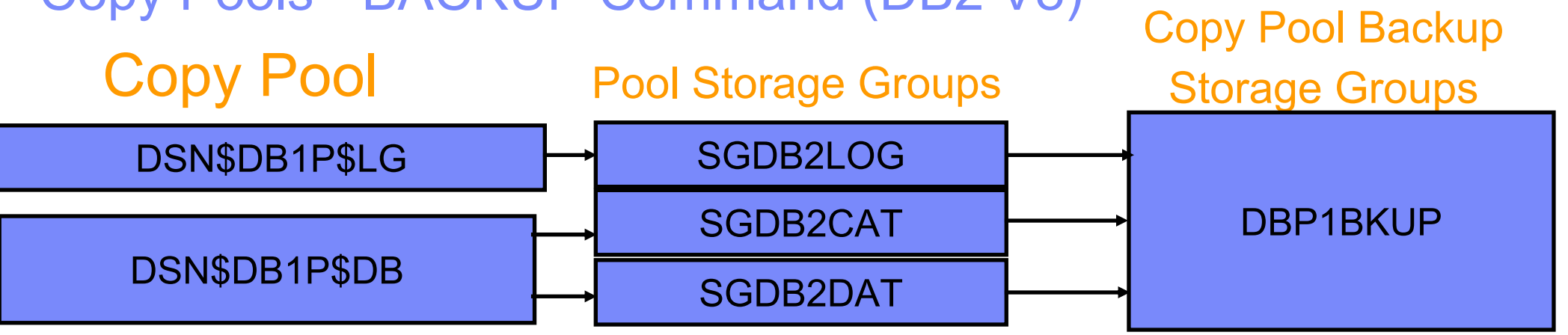


SMS can be set up to allow for volumes to be selected as primary, secondary, tertiary, or rejected. Discuss further with the Storage Administrator.

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

Copy Pools - BACKUP Command (DB2 V8)



- 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

MGEXTSZ - Sliding Secondary Allocation Size (DB2 V8) - page 1 of 3

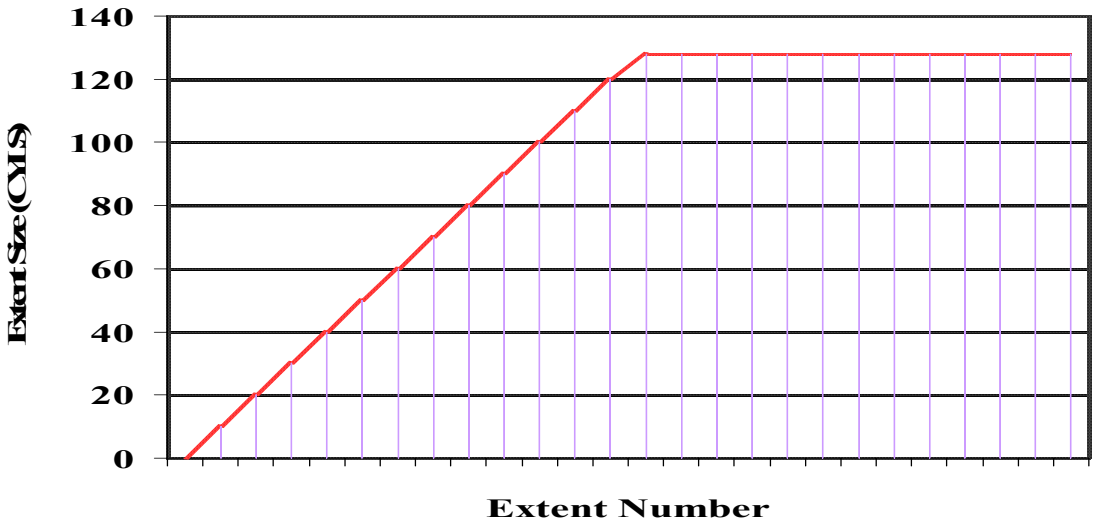
- **Applies to DB2 managed pagesets**
- **Tries to avoid VSAM maximum extent limit errors**
  - Can reach maximum dataset size before running out of extents. Beware of heavily fragmented volumes, which may impede this feature.
- **Requires ZPARM value MGEXTSZ in DSN6SYSP be set to YES, since the default is set to NO. It is set during the DB2 install in panel DSNTIP7 under OPTIMIZE EXTENT SIZING.**
- **Uses cylinder allocation**
  - Default PRIQTY
    - 1 cylinder for non-LOB tablespaces and indexes
    - 10 cylinders for LOB tablespaces
  - Improved SQL SELECT and INSERT performance when allocating on cylinder boundary
    - 2x improvement relative to track allocation
- **Can be used for:**
  - New pagesets: No need for PRIQTY/SECQTY values
  - Existing pagesets: Execute SQL to ALTER PRIQTY/SECQTY values to -1 plus schedule a REORG

MGEXTSZ - Sliding Secondary Allocation Size (DB2 V8) - page 2 of 3

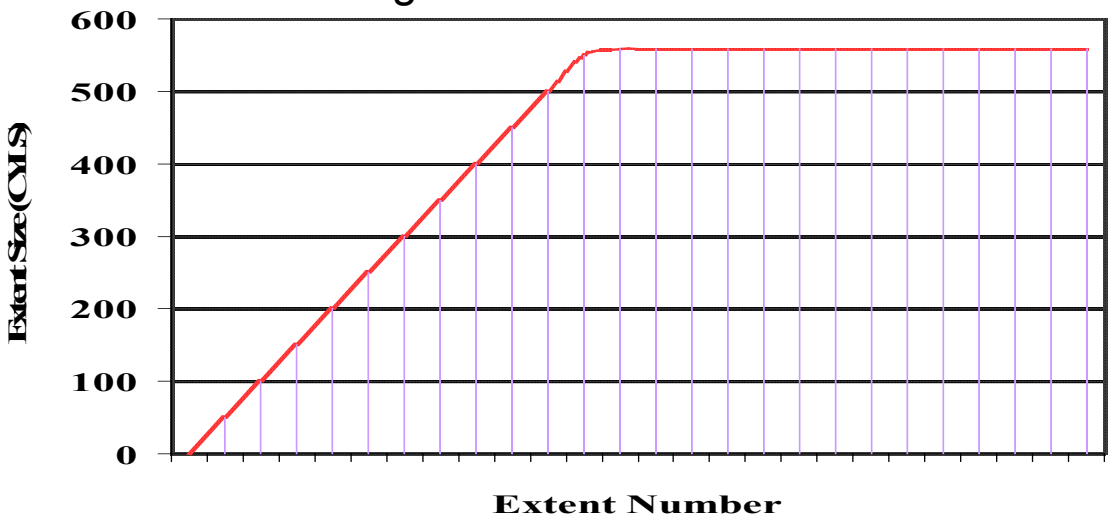
- **Two sliding scales will be used depending on the maximum dataset size. The first 127 extents are allocated in increasing size, and the remaining extents are allocated based on the initial size of the data set:**
  - For 32 GB and 64 GB data sets, each extent is allocated with a size of 559 cylinders.
  - 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.**
- **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

MGEXTSZ - Sliding Secondary Allocation Size (DB2 V8) - page 3 of 3

Sliding Scale for less than 1 GB to 16 GB



Sliding Scale for 32 GB and 64 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

TSQTY and IXQTY (DB2 V8) 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:**
  - 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.**

DSVCI - Support for VSAM Control Interval (CI) Greater than 4K (DB2 V8) - page 1 of 2

- **Support for CI sizes of 8, 16, and 32 K table spaces. For indexes, only 4K pages are supported.**
- **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 page sizes of table spaces, which will not take effect until after a LOAD or REORG of the table space.**
- **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
  - I/O bound executions benefit with using larger VSAM CI sizes, including COPY and REORG.

DSVCI - Support for VSAM Control Interval (CI) Greater than 4K (DB2 V8) - page 2 of 2

- **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:
    PRIQTY 72000

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

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
```

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,

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.

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 SERIAL	FREE SPACE	% FREE	ALLOC SPACE	FRAG INDEX	LARGEST EXTENT	FREE 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. DEFRAGing 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.

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.

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 V8:**
  - ADSNLOAD
  - ADSNLOD2
  - SDSNLOAD
  - SDSNLOD2

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

- **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.

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                          Tracks   XT   Device
-------------------------------------------------------------------
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).

I understood the last chart, but how did I actually get LESS space then 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.**

I am migrating my DB2 to V8, is there anything I need to tell my Storage Administrator?

- **It depends on how your Storage Administrator set up the ACS routines for the DB2 data set names. If they are looking at the low level qualifier of your DB2 data set name and you plan on using partitions above the V7 limit of 254, then the answer is yes. The LLQ in DB2 V8 will have 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
- Your Storage Administrator may need to change some reporting programs as well.

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 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 STOSPACE, 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.

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
- **Using Such functions as DFSMSdss COPY may be much faster than running REORGs.**
- **Do you have SQL that uses the DB2 catalog or RTS that 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.**
- **Do you use high extents as a tool to review issues with clustering indexes? This can potentially be a problem. Review CLUSTERRATIOF 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.**
- **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.
- **So, do I care if my data sets are multi volume?**
  - Yes, based on the information above. Lots of multi volume data sets can cause you lots of additional restores if a piece of your data resided on the crashed volume.

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:
    - (PSEUDO_DEL_ENTRIES/CARDF)*100
  - Recommendation - REORG INDEX, if the percentage of pseudo deleted entries is greater than 10%.

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 – 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.
- **TimeFinder - EMC Hardware**
  - similar in concept to FlashCopy for EMC
  - for more information see, http://www.emc.com/products/software/timefinder.jsp

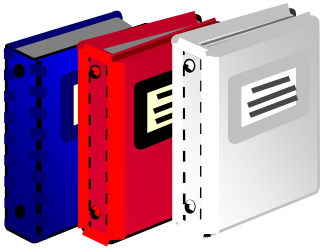
I have FlashCopy and/or SnapShot Technology, can it help if I want to clone my DB2s?

- **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.**

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

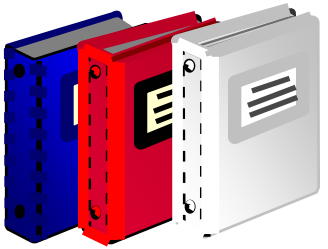
- **Ask your Storage Administrator for a copy of their procedures as it pertains to DB2. Review the document with your Storage Administrator to understand the concepts.**
- **Discuss with your Storage Administrator your installation's mix of hardware and software. How do they work and how can they work 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.**

References - page 1 of 2



- DB2 V8 Administration Guide SC18-7413
- DB2 V8 Installation Guide GC18-7418
- DB2 V8 SQL Reference SC18-7426
- DB2 V8 Utility Guide and Reference SC18-7427
- DB2PE Report Reference SC18-7978
- DB2 UDB for z/OS Version 8: Everything You Ever Wanted thing to Know, ... and More SG24-6079
- DB2 UDB for z/OS Version 8 Performance Topics SG24-6465
- DB2 for z/OS and OS/390 Version 7 Performance Topics SG24-6129
- DB2 for z/OS and OS/390 Version 7 Selected Performance Topics SG24-6894
- Effective zSeries Performance Monitoring using Resource Measurement Facility (RMF) SG24-6645
- Storage Management with DB2 for OS/390 SG24-5462
- z/OS 1.6 DFSMS: Implementing System-Managed Storage SC26-7407
- z/OS 1.6 DFSMSdfp Storage Administration Reference SC26-7402
- IBM RAMAC Array Introduction GC26-7012
- Disaster Recovery with DB2 UDB for z/OS SG24-6370
- z/OS V1R6.0 DFSMS Advanced Copy Services SC35-0428

References - page 2 of 2



- z/OS V1R6.0 DFSMS: Using DFSMSdfp in the z/OS V1R6.0 Environment SC26-7473
- z/OS V1R6.0 DFSMS Access Method Services for Catalogs SC26-7394
- z/OS V1R6.0 DFSMShsm Storage Administration Reference SC35-0422
- z/OS V1R6.0 DFSMSdss SAG SC35-0423
- DFSMS/MVS Software Support for IBM Enterprise Storage Server SC26-7318
- IBM Enterprise Storage Server SG24-5465
- IBM RAMAC Virtual Array SG24-4951
- The IBM TotalStorage DS8000 Series: Concepts and Architecture SG24-6452
- z/OS V1R6.0 RMF Performance Management Guide SC33-7992z/OS V1R6.0
- z/OS V1R6.0 RMF Report Analysis SC33-7991
- z/OS V1R6.0 RMF User's Guide SC33-7990
- z/OS V1R6.0 DFSMS: Using Data Sets SC26-7410
- IBM® TotalStorage Enterprise Storage Server Large Volume Support Performance Evaluation and Analysis - White Paper
- http://www.ibm.com/ibm/history/exhibits/storage/storage_photo.html
- http://www.ibm.com/ibm/history/exhibits/storage/storage_profile.html
- http://www.ibm.com/ibm/history/exhibits/storage/storage_basic.html