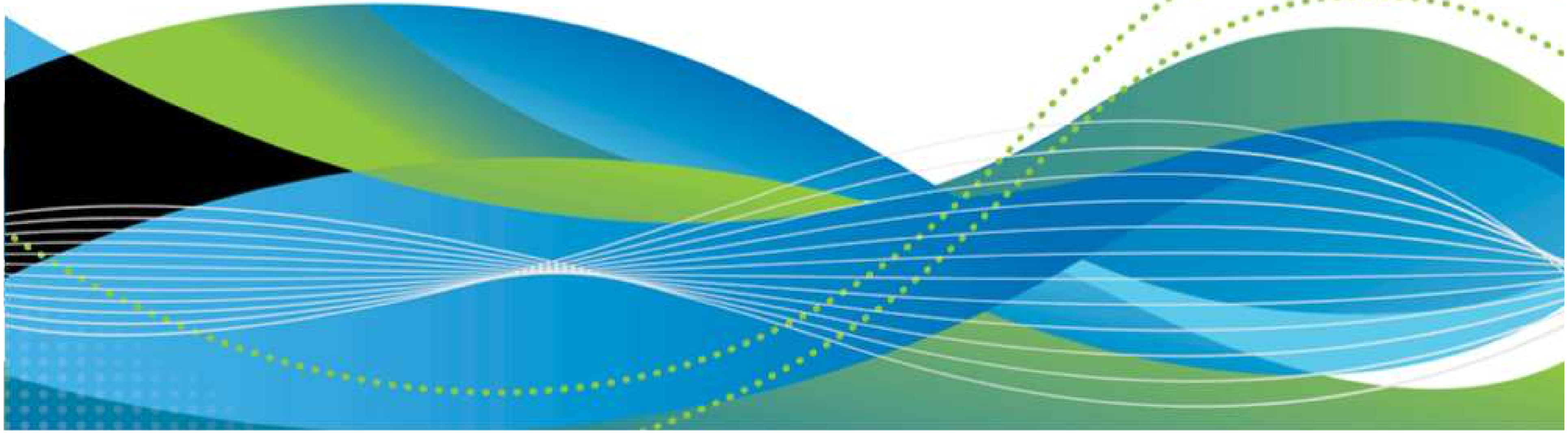
What Every Storage Administrator Should Do For YOUR DB2 Environment

John Iczkovits
IBM

March 15, 2012
Session Number 10475



- **10475: What Every Storage Administrator Should Do For YOUR DB2 Environment**

- Storage Administrators can make or break your DB2 environment. Storage Administrators have many different options when dealing with disk and tape for your DB2 environment. Which options should be turned on and which turned off for an optimal DB2 configuration? Come to this session and learn what your Storage Administrator should be doing for YOUR DB2 environment.

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/techdocs
/atsmastr.nsf/WebIndex/PRS2614

- MYTH: My Storage Administrator understands what DB2 requires, so there is no need for me to worry about disk or tape
- FACT: You can have the best Storage Administrator in the world, but the vast majority do not know how DB2 operates or what it requires.
  - Your Storage Administrator is working with dozens of products, many their own, such as dss, hsm, SMS, rmm, etc. Do not assume they understand database requirements.
  - DB2, IMS, CICS and other products have specific product related requirements that are not all the same.
  - Do not assume that your Storage Administrator knows what boot strap data sets, active and archive logs, image copy data sets, etc. are, nor their specific requirements. They generally do not.
    - Case in point – some customers have ALL of their DB2 data sets described above, plus the DB2 Catalog and Directory, user data sets, and DB2 sort data sets on the same set of volumes causing a single point of failure and performance problems.

- MYTH: As a DB2 professional, I have no influence over the disk and tape environment. I take what I get.

- FACT: Storage Administrators do their best to allocate DB2 data sets based on their understanding of requirements.
  - Be specific about DB2 requirements. Keep in mind, most Storage Administrators do not know the difference between an image copy data set and a DB2 user data set. Educate your Storage Administrator on DB2 and it's requirements.
  - At many sites, data sets are placed on less than the most efficient devices. At times your Storage Administrator has special technology, but has no idea DB2 can benefit from it.
    - Case in point – you have heavily random DB2 data sets that reside on HDD (regular drives), but your company has lots of new SSD (Solid State Devices) with little or nothing allocated on them). Your Storage Administrator does not know that these specific DB2 data sets can have 15 – 60x performance improvements by allocating the data sets on SSD devices instead.
    - Not all DB2 related tape data sets should reside on virtual tape. Some DB2 data sets may be much better off on manual devices.
  - Your company may have a much better technology to allocate your DB2 data sets, but it does not help if your Storage Administrator does not know your requirements.

- MYTH: Knowing my disk channel speed does not help me understand DB2 performance

- FACT: Customers have different disk channel speeds:
  - ESCON (some customers with old EMC devices still run ESCON)
  - FICON Express 2 (FEx2)
  - FICON Express 4 (FEx4)
  - FICON Express 8 (FEx8)
  - FICON Express 8S (FEx8S)
  - zHPF (High Performance FICON)

- Channel speed and whether zHPF is used influence data transfer time and therefore relate to the DB2 performance

- MYTH: The amount of disk cache on the disk boxes has nothing to do with the performance of my DB2 applications

- FACT: Think of disk cache as DB2 buffer pools
  - There is an unofficial hand shaking between disk cache and DB2 buffer pools. In concept they work similarly to accomplish the same task
  - Disk cache comes in different sizes.
    - When disk boxes are purchased, it is not uncommon to purchase less cache than required.
    - The price is cheaper and the performance consequences not yet known
  - Does increasing the size of your buffer pools help your application because the data is re-referenced?
    - If not, increasing the size of disk cache will probably not help
    - Otherwise, would adding disk cache increase performance for your DB2 application?
    - For IBM disk, Storage ATS can help determine if purchasing more cache would benefit DB2 performance.

- MYTH: I do not need to know the type of RAID my DB2 data resides on

- FACT: RAID (Redundant Arrays of Inexpensive Disks) allows for different options for recovery of physical devices
  - The most common mainframe options are RAID 5, 6, and 10
  - Different RAID options provide different levels of performance and protection from device failures
    - Performance – Disk can be allocated based on different RAID options. For example, you may have RAID 5 for your DB2 environment, but RAID 10 for your non DB2 data. Discuss with your Storage Administrator which RAID option is most suitable for performance. Discuss RAID performance penalties as trade offs in relation to the total cost of your disk.
    - Protection from failure – Discuss with your Storage Administrator your companies threshold for failure. Most customers implement RAID 5 technology. Does your current RAID implementation adequately protect you from multiple physical failures?

- MYTH: It does not matter if my DB2 data sets reside on HDD, SATA, or SSD disk devices

- FACT: Physical disk devices come in three flavors:
  - HDD (Hard Disk Drives) are the most common.
    - Spins at either 10K or 15K RPM
    - Each physical device holds 146GB to over 900GB of data
  - SATA (Serial Advanced Technology Attachment)
    - Spins at 7.2K RPM (1/3 to ½ slower than HDD)
    - Each physical device holds 1 or 2 TB of data
    - Cheapest devices as they spin much slower
    - Used in mainframe DB2 environments as a cheaper alternative to house image copy data sets, or Data Warehouse data sets if slower performance at cheaper costs are acceptable.
  - SSD (Solid State Devices)
    - Does not spin – no moving parts. No read/write arm. Think of it as memory
    - Each physical device holds 73GB – 300GB of data
    - Best for random data. Can be 15 – 60x faster for random vs. HDD
    - At times, data transfer is almost on par with cache reads
    - Most expensive
- Is your data on the right type of device?

- MYTH: I do not need to know if my DB2 data resides on physical volumes that use rotate extents vs. rotate volumes

- FACT: Rotate volumes single stripes data from one logical volume on one rank (set of eight physical volumes). Rotate extents single stripes data from one logical volume on two or more ranks for every 1,113 cylinders - .94 GB.
  - Rotate volumes only deals with one rank
  - Rotate extents places data on two or more ranks to avoid disk hot spots and therefore improve performance.
    - Newer disk implementations generally choose rotate extents
  - Why does it matter which implementation is used?
    - Performance. If the RMF Volume Detail report shows your volume is having problems, is your data on one rank or spread across two or more ranks?
      - *When using Rotate extents, which rank is causing the problem?*
    - Recovery. When a physical failure occurs that causes a rank to fail, which logical volsers were lost?

- **MYTH: So long as my data is separated onto different logical volumes, job complete. I am protected.**

- **FACT: Tens or hundreds of logical volsers can reside on the same rank – same set of physical disks**
  - Availability – you have done your due diligence, BSDS1 and the LOGCOPY1 data sets reside on DB2001, and BSDS2 and LOGCOPY2 data sets reside on DB2002. If DB2001 and DB2002 are on the same rank or extent pool and a failure beyond the RAID implementation occurs, you have lost both copies of the BSDS and active logs. Best practices dictate that the pairs be split at a minimum on different ranks or extent pools.
    - Placing the BSDS and active log data sets in the SMS Separation Profile data set assures that the data sets do not reside on the same logical volume, but they do not protect against the data sets residing on the same physical rank or extent pool.
  - Performance – if your DB2 data resides on 100 logical volumes, best practices dictates that the data is spread across different ranks or extent pools. This would help avoid some potential hot spot issues as well.
    - Having data sets on different logical volsers, but on the same rank may eliminate VTOC and VVDS reserve issues, but it does not mitigate the availability issue
    - Hot spots can still occur, but are much less frequent. Even with rotate extents, if by chance two heavily competing data sets reside on the same DDMs, hot spots can still occur.

SMS Data Class Best Practices for DB2

- **By default VSAM data sets can have a maximum of 255 extents.**
  - For availability, change the Data Class to allow for the maximum of 7,257 extents. **Extent Constraint Removal=YES**
- Specify CA RECLAIM for the ICF catalogs (not the DB2 LDSes) **– IDCAMS ALTER/CREATE parameter RECLAIMCA**
- **EF/EA all DB2 VSAM and sequential (not temporary) data sets to allow for:**
  - Data sets > 4GB (EF and EA)
  - Sequential files > 4,369 cylinders per volume (EF)
  - Striping (plus Storage Class SDR) (EF)
- **To bypass the 5 extent rule, specify any value (even 0) for Space Constraint Relief**
  - Specifying a value other than 0 will have the same result, with the added benefit of reducing allocations by the % specified when disk is out of space (not for Guaranteed Space allocations)

SMS Data Class Best Practices for DB2

- **COMPACT (compression) of archive log data sets**
  - Do not compress when further moving the archive log data sets to compressed tape
  - Requires DB2 9 NFM
- **Volume Count/Dynamic Volume Count – for DB2 Catalog and Directory data sets only – prior to DB2 V10**
  - Do not specify a volser as it will override VC/DVC
  - Do not specify for DB2 managed data sets as DB2 overrides VC/DVC
    - DB2 10 DB2 manages the Catalog and Directory data sets after REORG or when new data sets/pieces are created
- **Use Large Block Interface (LBI) - especially with tape**
  - BLKSIZE can go up tp 256K on tape (archive log data sets cannot exceed 28K)
  - Potentially large elapsed time reduction for COPY, and RECOVER RESTORE phase

SMS Storage Class Best Practices for DB2

- **Guaranteed space**
  - – Only specify for the active log and BSDS for hand placement
  - – Use along with the SMS Separation Profile for added insurance
- **Specify SDR rate for VSAM and sequential striping**
  - – Consider VSAM striping heavily sequential data sets. Start with striping the active log data sets
  - – Consider striping sequential data sets
    - Fully test all variations of striping for performance gains
- **Enable multi-tiered Storage Groups when multiple SMS Storage Groups are used**
- **Direct allocations to SSD devices MSR=1, or HDD devices MSR=10 if both exist**
  - – Review use of Easy Tier
  - – If Easy Tier is not used, place DB2 data that is random on SSD

SMS Management Class Best Practices for DB2

- **For DB2 production environments, generally specify for DB2 VSAM data sets:**
  - No backup
  - No migrate
- **For production data sets, determine migration strategies for:**
  - Image copies
  - Archive log data sets
  - BSDS data sets
  - When dealing with GDGs, consider GDG strategy. Migrate all but the 0 generation?
  - Decide if hsm L1, L2, or a combination is used
    - ZPARM values for RECALL & RECALLD must be specified with a reasonable value
      - Specify higher values for manual tape data sets
      - Consider higher values for ATL/VTS when recalling large number of data sets

SMS Management Class Best Practices for DB2

- **For non production environments:**
  - Consider migrating data sets to save on space
    - Decide if hsm L1, L2, or a combination is used
    - ZPARM values for RECALL & RECALLD must be specified with a reasonable value
      - Specify higher values for manual tape data sets
      - Consider higher values for ATL/VTS when recalling large number of data sets
  - Consider releasing over allocated space for VSAM data sets to save on space
    - Data sets must have been allocated EF with no Guaranteed Space
    - For space savings only, can cause slight performance degradation
    - For production VSAM data sets that are read only and will never be updated

SMS Storage Group Best Practices for DB2

- **Place BSDS and active log data sets on fastest devices**
- **Set HIGH value reasonably to avoid multi volume data sets**
  - Multi volume data sets cause performance degradation when spanning to additional volumes
  - Theoretically HIGH can be higher when emulating larger devices
    - Avoid mixing different emulated types when possible. 70% of a mod 3 device is considerably smaller than 70% of a mod 54
  - Volumes housing the active log and BSDS data sets can max out at 99% for high when new data sets are not anticipated
  - Consider setting HIGH at 99% for archive log, BSDS, and image copy data sets. These data sets generally do not require additional space
  - Capture SMS message IGD17380I and send out e-mails to investigate lack of space
- **If migrating data sets, set LOW value reasonably**
  - Avoid hsm doing too much work
  - Depending on factors, HIGH value will be used to trigger LOW for migrations.
    - Consider a small HIGH and LOW value for such pools as for the archive log, BSDS, and image copy data sets

SMS Storage Group Best Practices for DB2

- **Use overflow volumes**
  - Provides additional protection when primary volumes are out of space
    - Review allocations on overflow volumes daily. Resolve space problems and online REORG data sets allocated to the overflow volumes when time allows.
  - Use Extend Storage Group
    - Overflow volumes are used for new allocations
    - Extend volumes are used for data sets already created requiring an extend function
    - Overflow volumes can be extend volumes
  - Backup the overflow/extend volumes when doing full volume backups, such as FlashCopy
- **Use multi-tiered volumes when more than one SMS Storage Group is specified in the ACS routine that require allocations in a specific Storage Group order**

SMS Storage Group Best Practices for DB2

- **When random volume selection is required and SRM (System Resources Manager) is not used as part of the weighing algorithm**
  - SRM only keeps seconds of data, generally not more than two minutes
  - SRM does not have the proper picture of long term usage of volumes
  - Use Storage Class IART>0 when SRM is not to be used as part of the allocation algorithm

SMS Storage Group Best Practices for DB2

- **Do not allocate all DB2 data sets to one large Storage Group**
  - At a minimal, BSDS and active log data sets should be placed on a separate SMS Storage Group with its own alias and ICF catalog. Consider adding the alias for the archive log data sets in the same ICF catalog
    - **BSDS (1 & 2) and active log data sets (LOGCOPY1 and 2) need to be separated by rank or extent pool**
  - Consider splitting the DB2 Catalog and Directory data sets from the user data sets
    - Splitting the data will avoid space from one type inadvertently using the space from another
  - Split DSNDB07 or equivalent data sets onto their own volumes when concerned about runaway transactions overusing sort files
  - Image copy and archive log data sets should be allocated on their own Storage Groups
  - Review alias and ICF catalog placement. When using full volume restores, avoid incorrectly overwriting ICF catalogs

What 3390 type should I emulate for my DB2 objects?

- **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

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.

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.

- **MYTH: I do not need to know if my DB2 data resides on PAV devices**

- **FACT: PAV (Parallel Access Volumes) allows for multiple reads and writes for a volser so long as those sets of tracks are not being updated at that time**
  - With older technology IOSQ (I/O Sequential Queuing) was a major performance inhibiter
    - DB2 thread 1 reads the first 100 cylinders from volume DB200A
    - While the first 100 cylinders are still being read in, thread 2 asks to read 5 tracks at cylinder 3000 from the same volume – DB200A
    - While the first 100 cylinders are still being read in, and thread 2 is still waiting, thread 3 asks to read 30 tracks at cylinder 1500 from the same volume – DB200A
    - Thread 2 and 3 wait (I/O Sequential Queuing) until the first 100 cylinders were read for thread 1.
      - All read/write arms are reading in the 100 cylinders and we do not do an I/O interrupt for thread 2 or 3.
      - Once thread 1 is complete, then thread 2 will process, once that is complete, thread 3 will process
  - Newer technology does not work with data on just one volume with several platters, and the access arms do not need to move very far as the entire disk is only 2.5 to 3.5 inches wide.
    - Several physical disks are read from at once
    - In the above example, 1 base UCB is used for DB200A and 2 aliases, 3 UCBs in total and the reads are done simultaneously and not queued
  - Same volser, several different MVS addresses (UCBs)
    - Logical volsers can have several different MVS addresses and several channel/paths
  - A heavily used DB2 logical volume can have dozens of MVS addresses
  - Three different types of PAV:
    - Static – original PAV. Specific number of PAVs for logical volumes
    - Dynamic (most common) – Use of WLM to manage logical volsers across a SYSPLEX. Logical volumes only acquire additional UCBs when required
    - Hyper (becoming more common) - Logical volumes only acquire additional UCBs when required, however does not have the WLM overhead.
  - With newer disk, IOSQ is totally eliminated or greatly reduced
  - Some customers do not have PAVs or enough PAVs. Validate the setup to ensure DB2 does not queue and wait for a device.