IBM Information On Demand 2007

Session: 1265

IBM

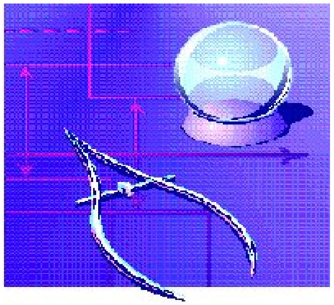
What's New in DB2 for z/OS Buffer Pool Management

Dr. Jim Teng
IBM Distinguished Engineer

Act.Right.Now.

IBM INFORMATION ON DEMAND 2007
October 14 – 19, 2007
Mandalay Bay
Las Vegas, Nevada

Disclaimers & Trademarks*



Information in this presentation about IBM's future plans reflect current thinking and is subject to change at IBM's business discretion. You should not rely on such information to make business plans. Any discussion of OEM products is based upon information which has been publicly available and is subject to change. The opinions expressed are those of the presenter at the time, not necessarily the current opinion and certainly not that of the company.

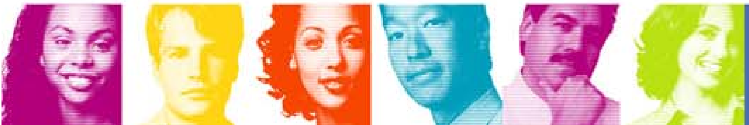
The following terms are trademarks or registered trademarks of the IBM Corporation in the United States and/or other countries: AIX, AS/400, DATABASE 2, DB2*, Enterprise Storage Server, ESCON*, IBM, iSeries, Lotus, NOTES, OS/400, pSeries, RISC, WebSphere, xSeries, z/Architecture, z/OS, zSeries, System p, System I, System z

The following terms are trademarks or registered trademarks of the Microsoft Corporation in the United States and/or other countries: MICROSOFT, WINDOWS, ODBC

For more copyright & trademark information see ibm.com/legal/copytrade.phtml

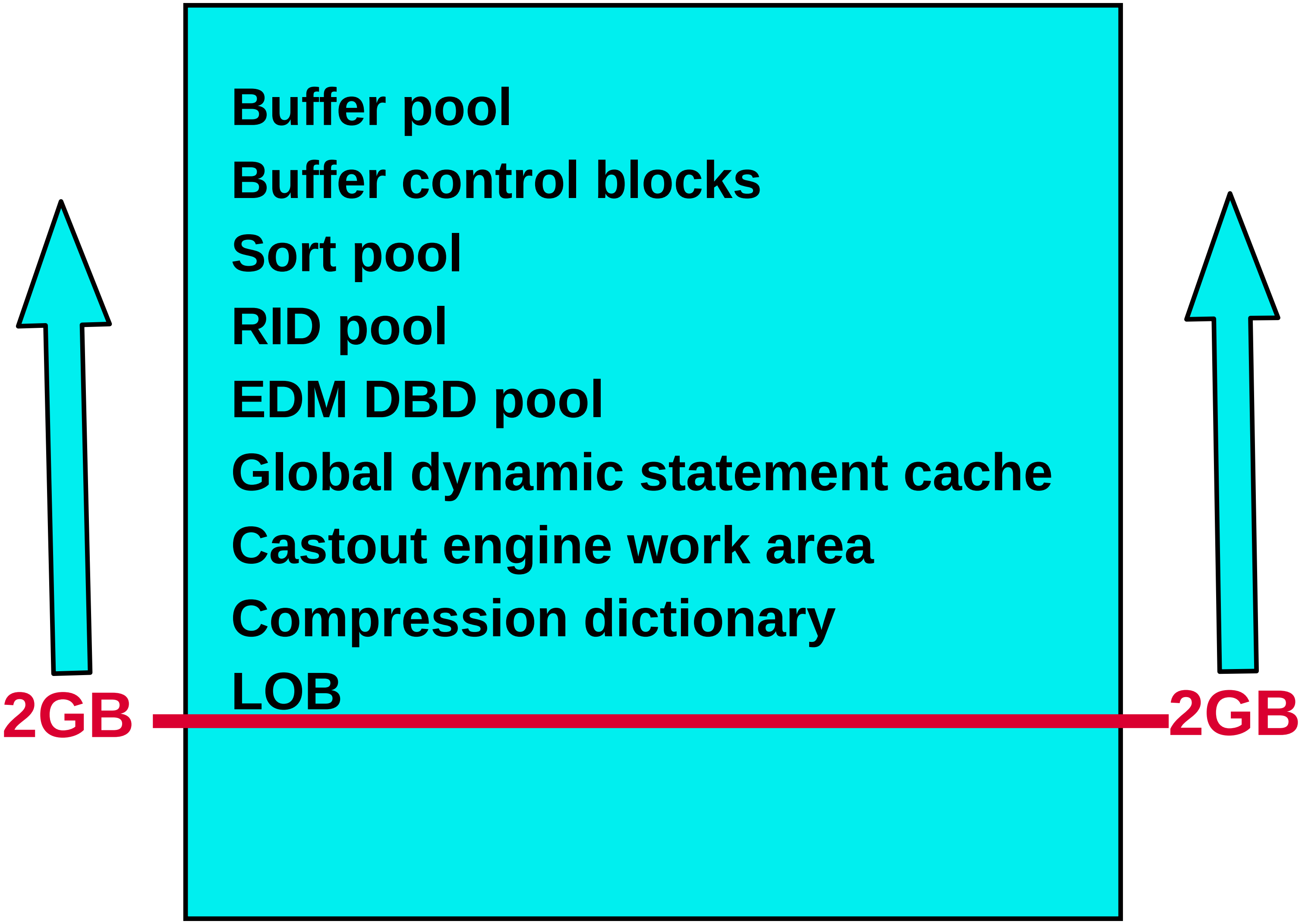
Agenda

- Buffer Pool Enhancements in V8
  - ►64-bit Buffer Pools
  - ►Batch Group Buffer Pool Writes
  - ►Miscellaneous Enhancements
- Buffer Pool Enhancements in V9
  - ►Automatic Buffer Pool Management
  - ►Workfile Buffer Pools
  - ►Commands to Open/Close Tablespace and Index
  - ►Miscellaneous Enhancements



Buffer Pool Enhancements in DB2 Version 8

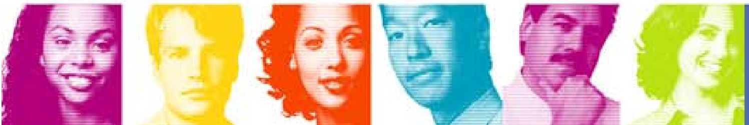
DBM1 Virtual Storage Constraint Relief



Buffer pool
Buffer control blocks
Sort pool
RID pool
EDM DBD pool
Global dynamic statement cache
Castout engine work area
Compression dictionary
LOB

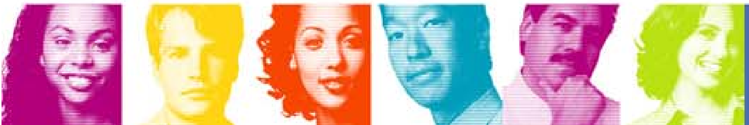
64-bit Buffer Pools

- Max BP size is lifted to 1TB
  - Max size of single or summation of all
  - The actual maximum = the REAL storage available
  - Always allocated above 2GB
  - Castout buffers and Buffer Control Blocks are also allocated above 2GB
- Data space pools and Hiperpools are eliminated
  - Simplifying DB2 system management
- Migration
  - V8_VPSIZE = V7_VPSIZE + HPSIZE
  - Fallback uses V7_VPSIZE and HPSIZE



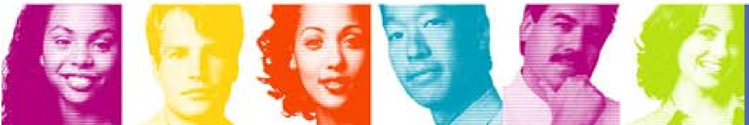
64-bit Buffer Pools ...

- PGFIX = YES option to long-term page fix buffers in real storage (i.e. virtual = real)
  - Use where I/O rate is high
  - Must have real storage available to back the pool
  - Up to 10% CPU saving
  - Independent setting by buffer pool
  - Issues DSNB541I and ignores PGFIX = YES if the total active BP storage > 80% of REAL storage
  - PGFIX option is enabled in V8 CM Mode
- PGFIX = NO (which is the default)
  - Needs to do page fix/free for each I/O or each GBP operation



64-bit Buffer Pools ...

- Need to have sufficient real storage to back BP
  - Paging I/O's will affect performance
  - Issue DSNB536I if the total active BP storage > REAL storage capacity
  - Issue DSNB610I if the total active BP storage > 2 x REAL storage
    - Adjust BP size downward or use the minimum size
  - Issue DSNB508I if the total BP size > 1 TB
- Default BP0 size raised from 2000 to 20000
- Default BP32K size raised from 24 to 250

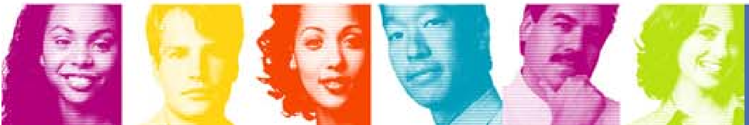


64-bit Buffer Pools ...

- ALTER BUFFERPOOL command parameters are NO LONGER supported:
  - ►VPTYPE, HPSIZE, HPSEQT, CASTOUT
- Parameters remaining unchanged:
  - ►VPSEQT, VPPSEQT, VPXPSEQT, DWQT, VDWQT, and PGSTEAL
- DISPLAY BPOOL LSTATS report removes references to hiperpool related counters

Batching of GBP Writes and Castout Reads

- Objectives
  - Write/Castout multiple pages in a single CF operation
  - Reduce traffic to and from CF
  - Improved data sharing performance for most workloads, especially for batch update
    - Workloads that updating large numbers of updated pages for GBP-dependent objects
    - ReduceDBM1's CPU time and CF link utilization due to less CF messages
- What are the prerequisites ?
  - New commands in z/OS 1.4
  - CF microcode shipped with CF LEVEL 12

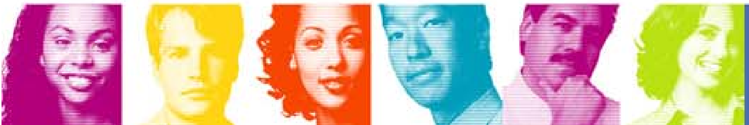


Batching of GBP Writes and Castout Reads ...

- z/OS 1.4 commands:
  - ►WARM - Write And Register Multiple command
    - –Registers and Writes Multiple Pages to a GBP
  - ►RFCOM - Read FOR Castout Multiple
    - –Read multiple pages from a GBP for CASTOUT processing
- Statistics and accounting records updated for measurement

Improved LPL Recovery

- Avoid global drain of entire object when LPL pages are being recovered
  - Only pages in LPL are unavailable
  - Other pages remain available
- Automatically trigger LPL recovery whenever possible
  - Skip triggering if: DASD I/O error, during DB2 restart, GBP structure failure, and GBP 100% loss of connectivity
- Improved diagnostics
  - To give better indication exactly why pages were added to LPL (DSNB250E message)



Pages Written to GBP at Phase 1 instead of Phase 2

- Some Tx managers spawn other transactions at syncpoint
- Spawned Tx could encounter "record not found" if it tries to read originating tx's update from another member (rare but a few customers have reported it)
- IMMEDWRITE NO will now write pages at commit phase 1
  - ZPARM IMMEDWRI(PH1) option removed
  - BIND IMMEDWRITE(PH1) option kept for compatibility
- Equivalent performance for Ph1 vs. Ph2 writes
- CPU cost to write pages to GBP being transferred from MSTR SRB to allied TCB
  - Included in the class 2 accounting CPU time

Greater Than 4K VSAM CI Support

- Problems for >4K page size table spaces in V7:
  - No VSAM striping
  - No Concurrent Copy
  - Exposure to create inconsistent pages during FlashCopy or GDPS/FREEZE
- Solution: V8 NFM
  - New CI size equals page size by default
    - e.g. 16K CI for 16K page
  - Use REORG to convert existing table spaces
  - Improve data rate for 8K, 16K, and 32K page
  - BACKUP SYSTEM Utility and the SET LOG SUSPEND will not suspend 32K page write I/Os

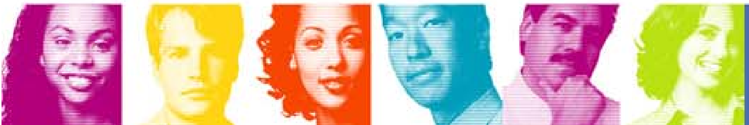
Miscellaneous Enhancements

- Identify long running reader without commit
  - Use read claim to identify long running readers
  - Enable by setting a non-zero ZPARM LRDRTHLD value (0 - 1439 minutes, default = 0)
  - Write IFCID 313 records to identify long running readers
- Remove the limit of 180 CIs per I/O for list prefetch and castout I/O

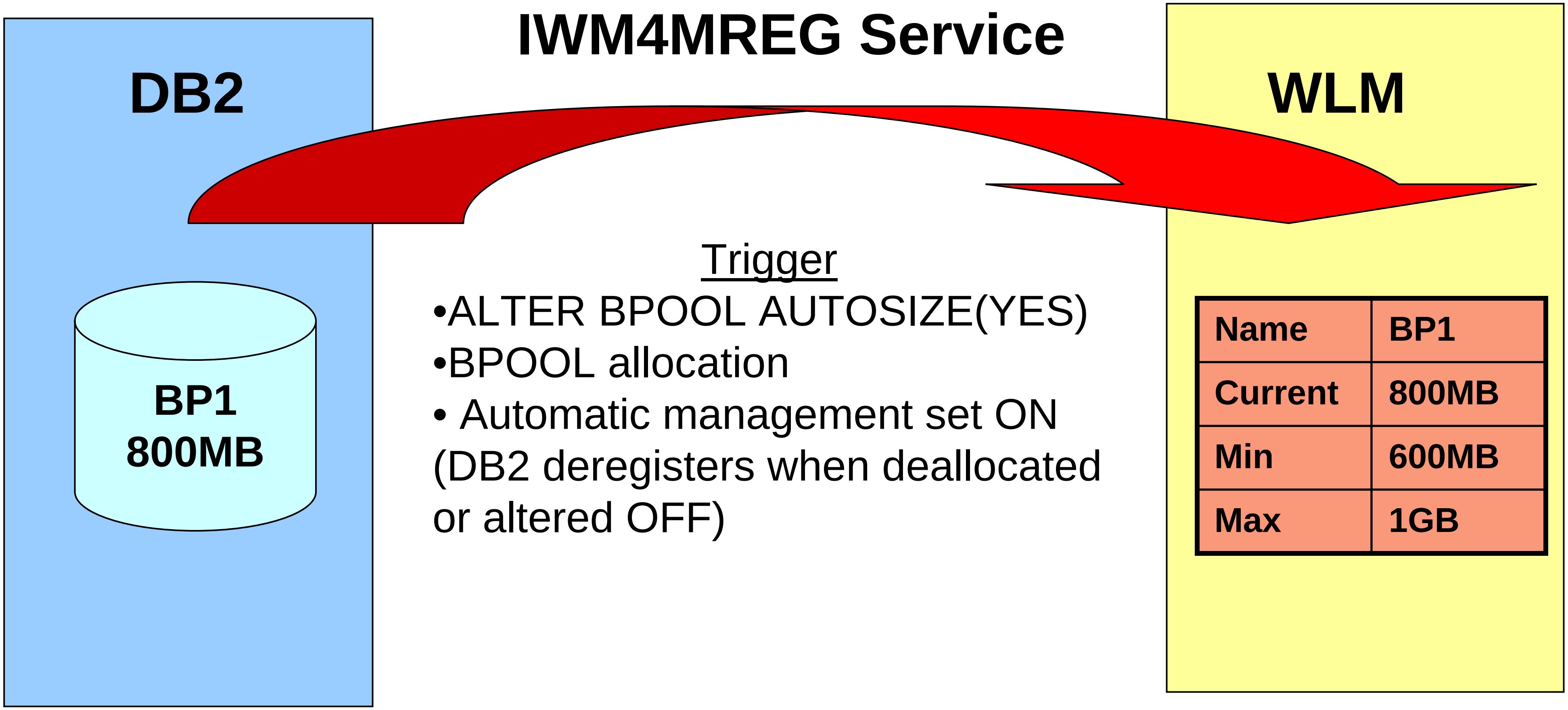
Buffer Pool Enhancements in DB2 9

Automatic buffer pool management

- Only the size attribute of the buffer pool.
- Can be enabled or disabled at the individual buffer pool level.
- Automatic management entails the following*:*
  - ►DB2 Registers the BPOOL with WLM
  - ►DB2 provides sizing information to WLM
  - ►DB2 communicates to WLM each time allied agents encounter delays
  - ►DB2 periodically reports BPOOL size and random read hit ratios to WLM



DB2 Registers BPOOL to WLM



Trigger

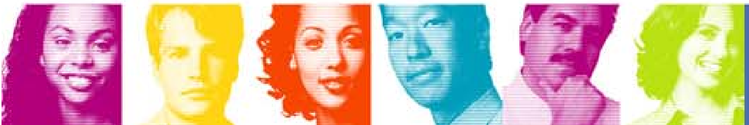
- •ALTER BPOOL AUTOSIZE(YES)
- •BPOOL allocation
- • Automatic management set ON (DB2 deregisters when deallocated or altered OFF)

Name	**BP1**
Current	**800MB**
Min	**600MB**
Max	**1GB**

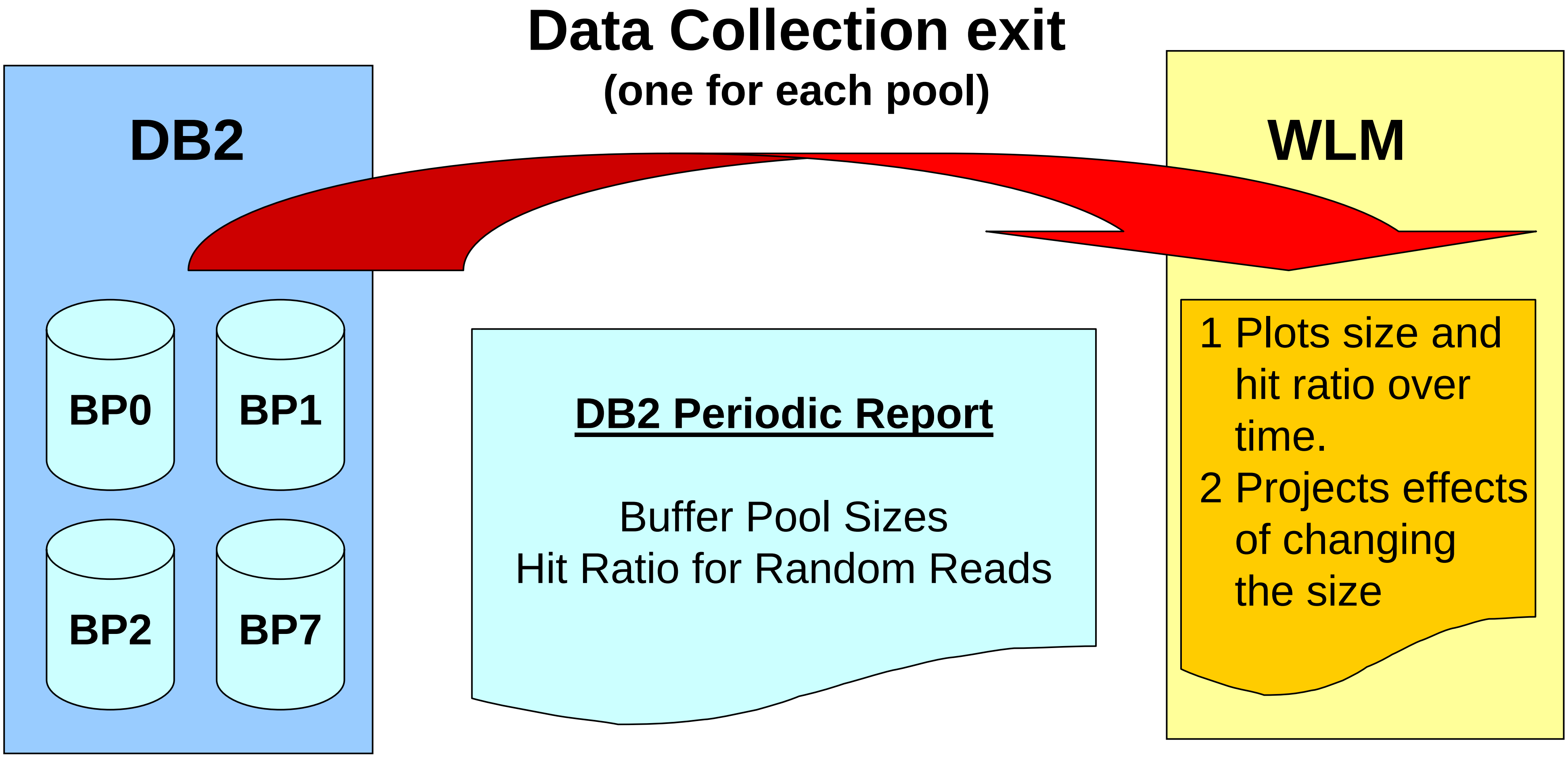
DB2 communication to WLM

Each time an allied agent encounters a delay caused by a random Get Page having to wait for read I/O.

- The following cases are not communicated to WLM:
  - Prefetch I/O
  - Wait for I/O on a sequential GetPage
  - Group buffer pool reads



Periodic reporting



Data Collection exit

(one for each pool)

DB2

BP0 **BP1**

BP2 **BP7**

WLM

<u>DB2 Periodic Report</u>

Buffer Pool Sizes
Hit Ratio for Random Reads

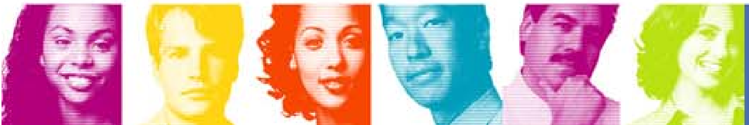
1. Plots size and hit ratio over time.
2. Projects effects of changing the size

Buffer Pool adjusting

- If the buffer pool is adjusted, the result will be just as though an ALTER BUFFERPOOL VPSIZE command had been issued
  - The new size is stored by DB2 in the BSDS

- If the buffer pool is deallocated (e.g. because DB2 is being stopped) it will subsequently be reallocated at its most recently allocated size.

  Example
  - If BPOOL is adjusted from 800 MB to 900 MB
  - Then DB2 is stopped and restarted
  - BPOOL will be subsequently allocated at 900 MB



What if the BPOOL is manually altered?

- If a buffer pool's size is manually altered (via the ALTER BUFFERPOOL VPSIZE command), it is deregistered and then registered at the new size.
- Example
  - –BPOOL registered at 800 MB
  - –Altered to a size of 1000 MB
  - –Then after the alteration has completed, DB2 deregisters and re-registers the buffer pool at 1000 MB with a new min of 750 MB and a new max of 1250 MB.

AUTOSIZE option

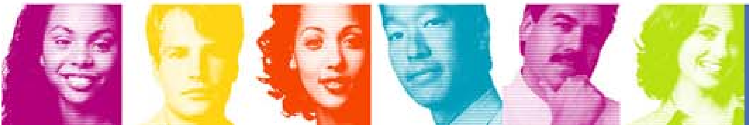
- DB2 will increase or decrease the size of a given buffer pool by up to 25% of the originally allocated size.
- By default, automatic buffer pool adjustment is turned off.
- It can be activated via a new AUTOSIZE(YES) option on the ALTER BUFFERPOOL command.
- Once activated, it can be deactivated by ALTER BUFFERPOOL(bpname) AUTOSIZE(NO).
- The AUTOSIZE attribute is added to the DISPLAY BUFFERPOOL output.

New messages

- **DSNB544I AUTOSIZE FOR bpname HAS BEEN SET TO nasize**
  - Message issued in response to an ALTER BUFFERPOOL command to indicate that the requested change to the AUTOSIZE attribute has been accepted.
- **DSNB555I WLM RECOMMENDATION TO ADJUST SIZE FOR BUFFER POOL bpname HAS COMPLETED**

  OLD SIZE = csize BUFFERS

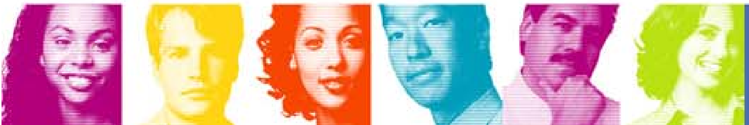
  NEW SIZE = nsize BUFFERS
  - Message issued when WLM notifies DB2 to adjust the size of a buffer pool.
  - Recommendation is made based on:
    - WLM's dynamic monitoring of the effects of buffer pool I/O on the achievement of workload goals
    - Amount of available storage on the system.



Changed message DSNB402I

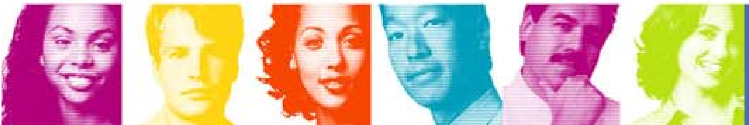
```
BUFFERPOOL SIZE = vpsize         BUFFERS AUTOSIZE = auto
ALLOCATED           = vpalc          TO BE DELETED        = vptbd
IN-USE/UPDATED    = vpcba          BUFFERS ACTIVE       = vpact
```

- Message displayed by the **DISPLAY BUFFERPOOL** command:
  - vpsize - The user-specified buffer pool size
  - auto - the buffer pool AUTOSIZE attribute that is applicable to the current allocation of the buffer pool.
    - YES - Buffer pool uses WLM to automatically adjust the size of the buffer pool
    - NO - Buffer pool does not use WLM services for automatic buffer pool sizing
  - vpalc - Number of allocated buffers in an active buffer pool.
  - vptbd - Number of buffers to be deleted in an active buffer pool because of pool contraction.
  - vpcba - Number of currently active (not stealable) buffers in the buffer pool.
  - vpact - Number of allocated buffers which contain data.



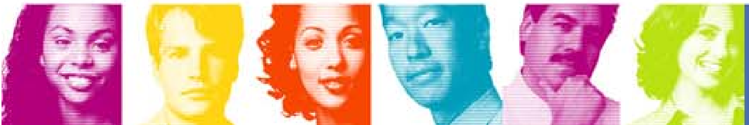
Migration considerations

- Functionality is available in DB2 9 CM or NFM mode.
- There are no fallback or coexistence considerations.
  - If a buffer pool is defined with AUTOSIZE(YES) while in V9, then the user falls back to V8, the AUTOSIZE(YES) option will be honoured upon remigration to V9.
- The AUTOSIZE option is ignored while running in V8.
- IFCID 0201 ("alter buffer pool") additions / changes in support of function



Prefetch and Deferred Write Quantity

- **Bigger prefetch and deferred write quantity for bigger buffer pool**
  - Max of 128KB V8 ->256KB V9 in SQL table scan
  - 256KB V8 ->512KB V9 in utility
  - +36% MB/sec in non striped prefetch
  - +47% in 2-striped prefetch -> more effective striping
- **"Bigger buffer pool"**
  - For sequential prefetch, if VPSEQT*VPSIZE> 160MB for SQL, 320MB for utility
  - For deferred write, if VPSIZE> 160MB for SQL, 320MB for utility



Dynamic Prefetch & Preformat

- **Replace all sequential prefetch, except in tablespace scan, with dynamic prefetch in SQL calls**
  - Up to 50% faster
  - Dynamic prefetch is more intelligent and robust

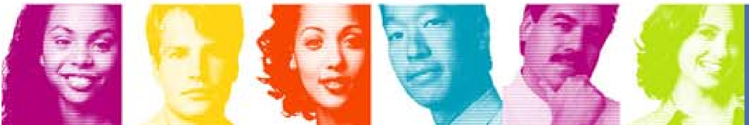
- **Bigger preformatting quantity and trigger ahead**
  - From 2 (V8) to 16 (V9) cylinders if >16cyl allocation
  - 27% faster Insert in one measurement

Workfile Buffer Pools

- **Heavier use of 32K workfile BP instead of 4K BP**
  - V9 tries to use 32K BP for bigger record size to gain improved performance, especially I/O time
    - Less workfile space and faster I/O
      - Example: 15 2050byte records on one 32K page vs 8 records on 8 4K pages
  - Recommendation
    - Assign bigger 32K workfile BP
    - Allocate more 32K workfile datasets
    - If 4K workfile BP activity is significantly less, corresponding BP size and workfile datasets can be reduced.

Miscellaneous Enhancements

- Command to remove GBP-dependency at object level
  - ACCESS DB MODE(NGBPDEP)
  - Typical usage would be before batch run
  - Issue on the member on which you plan to run batch
- Command to “prime” open dataset
  - START DB MODE(OPEN) [PART]
- Improved performance for GBP writes
  - Avoid copying pages for batched writes
- Auto-recover GRECP/LPL objects on group restart
- Long-term page fix castout buffers and index compression buffers



Session Title: What's New in DB2 for z/OS Buffer Pool Management

Session: 1265

Dr. Jim Teng

IBM Distinguished Engineer

IBM Silicon Valley laboratory

jteng@us.ibm.com