May 6-10, 2007

San Jose Convention Center

San Jose, California, USA

Session: B01

Availability enhancements in DB2 for z/OS

IDUG® 2007 North America

Haakon Roberts
IBM Corp

May 07, 2007 09:50 a.m. – 10:50 a.m.

Platform: z/OS

GoFurther

Agenda

- Introduction
- Virtual storage management
- Online schema & structure enhancements in V9
- Locking & concurrency
- LOBs
- Utilities
- Restart
- Recovery
- Data Sharing
- Other
- Recent maintenance stream changes
- Future enhancements
- Conclusion

Introduction

- Putting enhancements into perspective
  - Exploitation of existing function
  - Effective application & product testing
  - Effective change management procedures
  - Service strategy
  - New functional availability enhancements

Virtual storage management

- DB2 64 bit evolution
- V8 areas moved above 2Gb bar
  - Buffer pools, BM control blocks
  - Castout buffers
  - RID pool
  - EDM DBD cache (OBDs)
  - Global dynamic stmt cache
  - Sort pool
  - Trace tables (Global, Lock, BB)
  - Accounting blocks
  - Compression dictionaries
  - IRLM locks
- Degree of relief in V8 varies

Virtual storage management

- DBM1, the following are moved above the bar in V9
  - Parse trees
    - Peak below-the-bar storage for full prepare reduced 10%
  - EDM fixed pools
    - V8 customer dumps show as much as 50Mb moved
    - Allows larger above the bar EDM pools
  - SKPTs / SKCTs (primarily static SQL)
    - Also part of the CTs/PTs
    - New EDM pool for skeletons
    - Savings in below the bar 10Mb to 300Mb
  - Pageset blocks, RTS blocks
    - Up to 10's of Mb savings
  - Local SQL statement cache
    - Est. 60% moves above bar
  - Thread-related storage:
    - Certain RTs, space block, DMTR
    - 10's of Mb or more in savings

Virtual storage management

- DDF address space runs in 64-bit addressing mode in V9
  - Shared 64-bit memory object avoids xmem moves between DBM1 and DDF and improves performance
  - Constraint relief
- Overall, degree of relief in V9 varies

Universal table spaces

- Improved space management
- Improved mass delete performance
- Max size for segmented 64Gb -> 128Tb
- Flexible partitioning benefits
  - Alleviate space limitations
  - Partition independence
- Segmented and partitioned
  - No conversion for existing table spaces
- Single table only

Universal table spaces

- Partition by range
- Partition by growth
  - Partitions added on demand
  - Up to MAXPARTITIONS
  - No partitioning key
  - Currently, once grown, will not shrink
- Always LARGE
- LOBs remain tightly coupled
  - All LOBs in a single LOB table space belong to a single base partition
- XML table spaces inherit structure from base
  - PBR base – PBR XML
  - PBG base – PBG XML

Cloned tables

- Continued schema evolution
- Eliminate outage for LOAD REPLACE
- Clone is identical to base table
- EXCHANGE DATA BETWEEN syntax
  - Drain & switch
- Can create indexes & before triggers
- ALTER TABLESPACE applies to both base and clone

Not logged tables

- LOGGED/NOT LOGGED option on CREATE/ALTER TABLESPACE
- LOBs will still log control information
- Indexes inherit logging attribute from base
- LOB & XML table spaces inherit logging attribute from base
  - LOB logging attribute can still be set separately
- Availability implications
  - UR still created even though only not logged data is updated
  - Alter logged to not logged sets a recoverable point
  - Alter not logged to logged sets COPYP
  - Rollback
    - Sets RECP,LPL or RBDP,LPL, no effect on LOBs
  - Recovery
    - Only SHRLEVEL REFERENCE imagecopies
    - RECOVER to current implies last recoverable point
  - Restart
    - Sets RECP,LPL or RBDP,LPL if pageset was open for update, no effect on LOBs

Schema evolution contd.

- ALTER TABLE tb1 RENAME COLUMN
- RENAME INDEX
- ALTER TABLE LONG VARCHAR to VARCHAR
- ALTER TABLE LONG VARGRAPHIC to GRAPHIC
- Update schema
  - New CATMAINT options
  - Switch schema name
  - Change from owner to role
  - Change VCAT

Locking & concurrency

- SKIP LOCKED
  - New keywords on SELECT, UPDATE, DELETE
  - CS or RS isolation level only
  - Page or row locking
  - Skips pages or rows for which incompatible locks are held
    - SELECT can access data with held S or U locks
    - SELECT with update clause skips data with held S, U or X locks
    - UPDATE & DELETE skips data with held S, U or X locks

Locking & concurrency

- Sequential index key insert improvement
  - Asymmetric index leaf page split
  - >4Kb page size allowed
  - Improved space utilization
  - Reduced index tree lotch contention due to fewer page splits

Locking & concurrency

- LOB locking algorithm redesign
  - Pre – V9
    - S LOB locks acquired temporarily for space search during LOB allocation
    - S LOB locks acquired for LOB deallocation and held until commit
    - S LOB locks acquired for LOB read and held until commit
      - Even for isolation UR readers!
    - X LOB locks acquired for LOB allocation and held until commit

Locking & concurrency

- LOB locking algorithm redesign
  - V9 enhancements
    - Lock escalation can no longer occur
    - No LOB locks acquired for space search
      - Wholly reliant on oldest read claim, more granular than in V8
    - No LOB lock acquired for LOB deallocation
    - X LOB lock acquired temporarily for LOB allocation
      - Released once allocation complete
      - Data sharing: Changed pages must be written prior to lock release – recommend GBPCACHE CHANGE rather than SYSTEM for performance
    - S LOB lock acquired momentarily for UR readers
      - Released immediately
      - Ensures LOB is complete
    - No LOB lock acquired for non-UR readers
      - Rely on base row locking
    - V9 NFM “protocol 3” only

LOBs

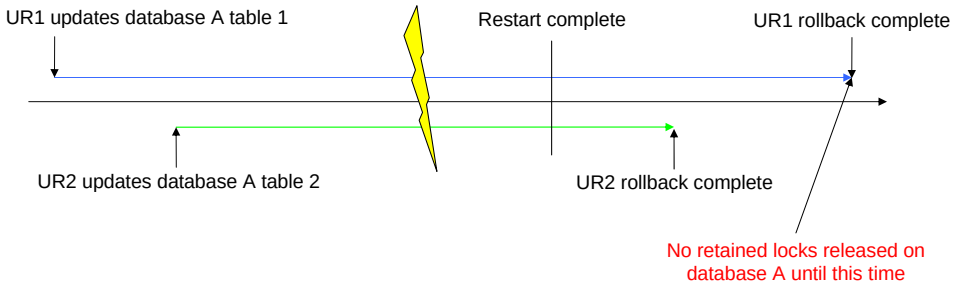
- LOB lock reduction/elimination
  - See earlier
- REORG SHRLEVEL REFERENCE
  - Allow read activity except during switch phase
  - Reclaim physical space
  - Complete REORG of LOB data
  - LOG NO only
  - Available in Compatibility Mode
  - SHRLEVEL NONE still supported
  - Maintains independence from base data

Utilities

- REORG BUILD2 phase eliminated
  - REORG TABLESPACE SHRLEVEL REFERENCE PART n will now have a LOG phase
- CHECK INDEX SHRLEVEL CHANGE
  - V7 & V8
- CHECK DATA / CHECK LOB SHRLEVEL CHANGE
  - V9
  - Restrictive states neither set nor reset
- REBUILD INDEX SHRLEVEL CHANGE
  - V9
  - Good for CREATE INDEX DEFER YES
    - Insert/update still restricted if unique index in RBDP/PSRBD
  - Not for NOT LOGGED (obviously)
- COPY CHECKPAGE will no longer set COPY pending when error detected

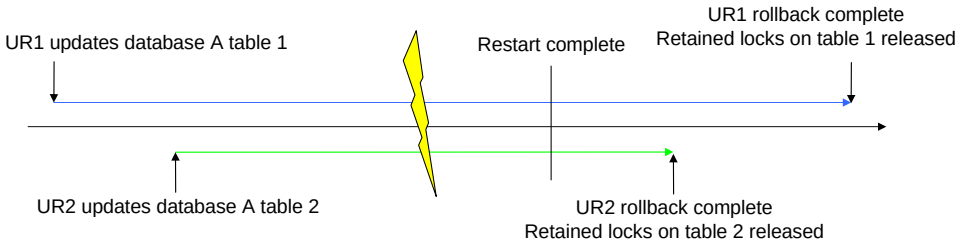
Restart

- -RECOVER POSTPONED: Improved availability for segmented tables
- Pre-V9:



Restart

- V9:



Restart

- Improved toleration of DBET inconsistencies
  - Pre-V9 may result in restart failures
  - V9: New restrictive DBETE state
    - Tablespace or indexspace level
    - Occurs when internal structure or logical inconsistency detected
    - New message DSNI045I indicates object involved
    - -STA DB() SP() ACCESS(FORCE) can reset DBETE
    - Allows restart & remedial action
- Allow log truncation by timestamp
  - New parameters on CRESTART CREATE
    - ENDTIME
    - SYSPITRT

Restart performance enhancements

- Initiate pageset open earlier in forward log phase
  - Asynch pageset open at log read rather than waiting until logapply
- Simplify pageset open processing during restart
  - Avoid acquiring GBP-conversion and extend locks
  - Also reduces risk of hang conditions during restart
- Defer closing SYSLGRNX ranges until after restart
  - Close at pseudo-close interval instead

Recovery

- RECOVER to point in time with consistency
  - Occurs for TOLOGPOINT & TORBA
    - Not TOCOPY
  - Data or index
  - New phases
    - LOGCSR
    - LOGUNDO
      - New pseudo-CLRs written
  - Inflight, inabort, postponed abort and indoubt URs will be rolled back
  - Reduce need for QUIESCE points
- New RESTOREBEFORE keyword
  - Allows choice of recovery base
- Object level recovery from system level backup
- -DISPLAY UTILITY to show RECOVER logapply progress

Recovery

- Archive log enhancements
  - Convert archive logs from BDAM to BSAM
    - V9 NFM
    - Support EF datasets
      - Striping
      - DFSMS compression
  - Larger internal read/write archive log buffers
    - Moved above 16Mb line
  - Dual buffering for archive log reads
  - Support >64K track archive logs
    - V9 NFM
    - z/OS 1.7 lifts limit with DSNTYPE=LARGE support
    - Allows 4Gb archive logs on DASD
      - Previously 4Gb archive logs had to be on tape

Data sharing

- Initiate automatic GRECP recovery at end restart
  - Previously only done when GBPs lost during mainline processing
  - Significant DR enhancement
  - Exceptions:
    - PITR mode
    - Tracker site
    - DEFER ALL

Other availability enhancements

- Workfile improvements
  - Converge WORKFILE and TEMP databases
  - Limit workfile consumption by thread
    - New ZPARM MAXTEMPS
  - Allow SEGSIZE to be specified for workfiles
  - Improved instrumentation for workfiles
    - New IFCID 2 statistics record information
- Automatic object creation
  - Implicit creation of
    - Database
    - Primary key index
    - Unique key index
    - ROWID index
    - LOB tablespace, table & aux index

Other availability enhancements

- Avoid re-IPL for EARLY code changes
  - Recycle of DB2 still required
- WLM-assisted bufferpool management
  - Manages bufferpool size
  - Range is +/- 25% of initial size
  - Tracks wait for read I/O for random getpage
  - -ALTER BUFFERPOOL () AUTOSIZE(YES)

DSNB555I WLM RECOMMENDATION TO ADJUST SIZE FOR BUFFER POOL bpname HAS COMPLETED
OLD SIZE = csize BUFFERS
NEW SIZE = nsize BUFFERS

Other availability enhancements

- Internal DB2 health monitor
  - Internal –DIS THD(*) SERVICE(WAIT) driven at minute intervals
  - 3 monitors for failover protection
    - MSTR (main), DBM1, DIST
  - -DIS THD(*) TYPE(SYSTEM) displays monitor information
  - Automatic DBM1 31-bit storage constraint warning
  - Health of DB2 subsystem reported to WLM
    - If <100%, work may be routed to less-constrained members
  - Reduce/eliminate requirement for manual –DIS THD(*) SERVICE(WAIT) commands

V91A N * 0 002.VMON 01 SYSOPR 002E 0
V507-ACTIVE MONITOR, INTERVALS=31126, STG=73%, BOOSTS=0, HEALTH=100

DSNVMON) DB2 BELOW-THE-BAR STORAGE WARNING
93% CONSUMED
92% CONSUMED BY DB2

Other availability enhancements

- Improved Down Level Detection processing
  - Simplified
  - Improved availability due to reduction in false down level detection errors
- Allow cancel of database commands
  - -DISPLAY THREAD (*) TYPE(SYSTEM)
  - Displays system agent information
  - Then use –CANCEL THREAD
- Allow IFCID 306 reads of compressed log records from prior to REORG

Recent maintenance stream changes

- PQ83649
  - -DISPLAY THREAD SERVICE(WAIT)
- PK01230
  - -DISPLAY THREAD SERVICE(WAIT) boost enhancement
- PK34441
  - Only set UTRW during UTILTERM phase of online REORG
- PQ99524
  - Prevent DB2 failure when stored procedures are cancelled
- PQ98043
  - Compress stack storage when necessary
- PQ99159
  - Prevent DB2 failure on castout errors

Recent maintenance stream changes

- PK13458
  - Improved handling of DBET notify errors in data sharing
- PK19972
  - DSN1COPY CHECK support for LOBs
- PK22723
  - DSN1LOGP CHECK 66% performance improvement
- PK22926
  - Prevent DB2 failure on index LPL recovery errors
- PK10307 & PK14694
  - Quarantine damaged virtual storage
- PK19417
  - Toleration of DBET inconsistencies during restart

Recent maintenance stream changes

- PK19328
  - Prevent IRLM failure due to bad request input
- PK37402
  - Improved data capture for IRLM-detected errors
- PK01245
  - Add diagnostic messages for restart hang conditions
- PK01751
  - Prevent lock escalation by MODIFY utility
- PK27611
  - Prevent log RBA rollover
- PK28576
  - Allow COPY to reset page information & avoid DSN1COPY RESET

Recent maintenance stream changes

- PK18534
  - Increase storage cushion size to account for increase in stack storage in V8
    - Make sure CTHREAD & MAXDBAT not over-allocated!
- PK29791
  - New LIGHT(NOINDOUBTS) option for RESTART LIGHT to ignore indoubt URs

Future enhancements

- Continued focus on reduction of planned & unplanned outages
- Continued online schema enhancements
- Faster, more consistent & more robust restart
- Improved WLM synergy for system robustness at high utilisation
- Continued focus on VSCR
- Utility enhancements
- Improved synergy with storage subsystems

Summary

- Substantial improvements at release level and also between releases
- Investment required
- Understand & exploit capabilities
- Availability dependent upon database, application & subsystem design & practices
- Understand and avoid capacity issues

Session: B01

Availability Enhancements in DB2 for z/OS

Haakon Roberts

IBM Corp

haakon@us.ibm.com

GoFurther