DB2 11 for z/OS
Availability Enhancements

More Goodies Than You May Think

Bart Steegmans
bart_steegmans@be.ibm.com

June 2014 - DB2 GSE user group meeting - Brussels

Disclaimer and Trademarks

Information contained in this material has not been submitted to any formal IBM review and is distributed on "as is" basis without any warranty either expressed or implied. Measurements data have been obtained in laboratory environment. 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. The use of this information is a customer responsibility.

IBM MAY HAVE PATENTS OR PENDING PATENT APPLICATIONS COVERING SUBJECT MATTER IN THIS DOCUMENT. THE FURNISHING OF THIS DOCUMENT DOES NOT IMPLY GIVING LICENSE TO THESE PATENTS.

TRADEMARKS: 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, e-business logo, Enterprise Storage Server, ESCON, FICON, OS/390, OS/400, ES/9000, MVS/ESA, Netfinity, RISC, RISC SYSTEM/6000, System i, System p, System x, System z, IBM, Lotus, NOTES, WebSphere, z/Architecture, z/OS, zSeries eserver

The FOLLOWING TERMS ARE TRADEMARKS OR REGISTERED TRADEMARKS OF THE MICROSOFT CORPORATION IN THE UNITED STATES AND/OR OTHER COUNTRIES: MICROSOFT, WINDOWS, WINDOWS NT, ODBC, WINDOWS 95

For additional information see ibm.com/legal/copytrade.phtml

Agenda

- Extended RBA/LRSN
- BIND / REBIND / DDL / Online REORG concurrency with persistent threads running packages bound with RELEASE(DEALLOCATE)
- More online schema changes
- More REORG avoidance
- Data sharing availability enhancements

Expanded (Long) RBA / LRSN

- Wrappig of the RBA range
  - Documented procedure (Admin Guide)
  - Very painful in non-data sharing environment
- Reaching end of LSRN range
  - No procedure to deal with that – DB2 11 is the answer
- DB2 11 can expands the RBA and LRSN to 10 bytes after reaching NFM
  - RBA addressing capacity of 1 yottabyte (2**80)
  - LRSN extended on the left by 1 byte, on the right by 3 bytes
    - >30,000 years and 16Mx more precision
  - 8 bytes is not sufficient to solve LRSN issues and may not give enough capacity for the longer term
  - DB2 11 in all modes operates internally with 10 byte RBA / LRSN
  - But externally DB2 continues to use 6 byte values in CM
  - Once in NFM, DB2 continues to use 6-byte values until you take action to convert

Expanded (Long) RBA / LRSN

- Two conversion tasks
  - Convert BSDSes to new format to enable logging with larger RBAs/LRSNs
  - Convert pagesets to new page format
  - These tasks are optional
    - If you do not care about larger RBAs/LRSNs then you do not have to convert
    - BSDSes can be converted without converting pagesets
    - Pagesets can be converted in a piecemeal fashion
    - Performance benefit accrued earlier if you convert BSDSes first
- All the gory details are explained in Timm’s presentation

Break-in support – Current Problem

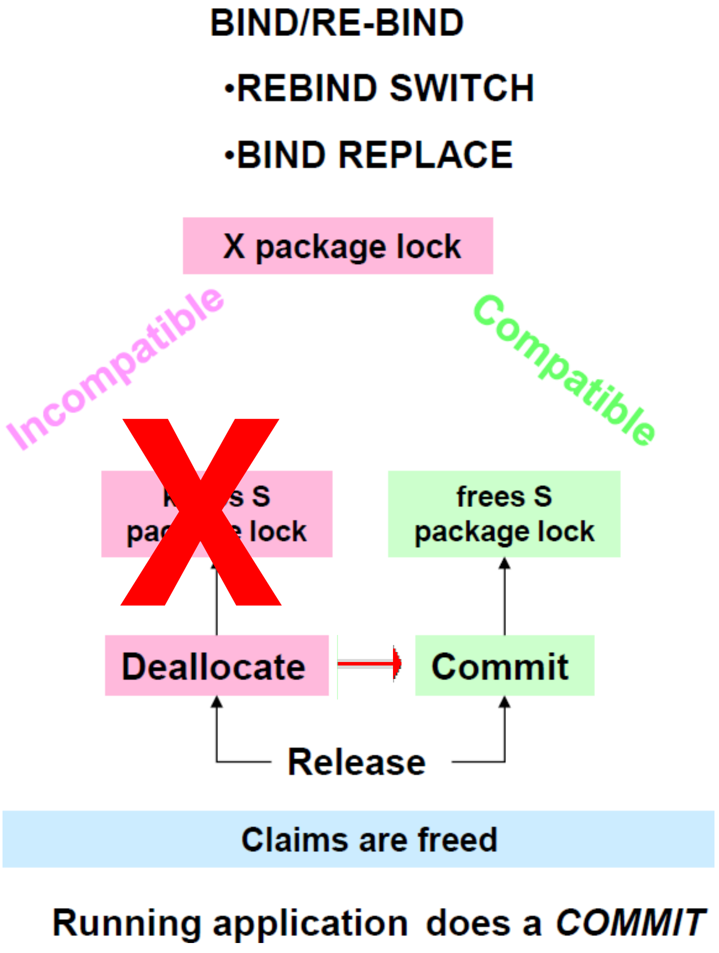
- Running BIND / REBIND / DDL / Online REORG concurrently with persistent threads running packages bound with RELEASE(DEALLOCATE) can lead to timeouts
- Problem became worse with increased use of persistent threads with DB2 10 after DBM1 ASID 31-bit VSCR
  - Examples: IMS Pseudo WFI, CICS protected ENTRY
- Currently having to shut down these applications to get such ((RE)BIND/DDL/OLR) operation through

Break-in support

- DB2 11 delivers a break-in mechanism for persistent RELEASE(DEALLOCATE) threads
- Will automatically detect operations that would like to break in and blocked by persistent threads running packages bound with RELEASE(DEALLOCATE)
  - Active and (local) idle threads are supported
    - Idle thread support requires PM95929, PM96001, and PM96004 (PE PI14705)
    - Idle thread support requires NFM
- If detected, then RELEASE(DEALLOCATE) packages will behave like RELEASE(COMMIT)
- Packages resume normal RELEASE(DEALLOCATE) behavior after the break-in operation completes
- New zparm PKGREL_COMMIT must be set to YES
  - Default is YES
  - Online changeable

Break-in support

- Break in mechanism only applies if
  - No CURSOR WITH HOLD is open at commit
  - Packages not bound with KEEPDYNAMIC(YES)
  - COMMIT does not occur inside a stored procedure
  - Note: These 3 restrictions already applied to packages originally bound with RELEASE(COMMIT)
- Performance impact
  - Should be minor but TS-locks and package lock and package need to to be re-acquired again
  - “Small price to pay” for increased availability



BIND/RE-BIND

•REBIND SWITCH

•BIND REPLACE

X package lock

Incompatible

Compatible

K s S pa lock

frees S package lock

Deallocate

Commit

Release

Claims are freed

Running application does a *COMMIT*

On-line Schema Change Enhancements

- Online ALTER Partition Limit Keys
- DROP COLUMN
- Point in time recovery support for deferred schema changes

- Alter Drop Pending Changes: AREOR status is now removed

Online ALTER Partition Limit Keys

- Currently:
  - Affected partitions are set to REORP
  - These partitions cannot be accessed
  - REORG is run to redistribute the data and remove the status
- In DB2 11 NFM
  - ALTER limit key is treated as a pending alter
  - The affected partitions are set to AREOR
  - Online REORG must be run to materialize the pending changes
  - PIT recovery prior to the ALTER limit keys is supported (RECOVER+REORG)
- Supported table spaces types are:
  - UTS – partitioned by range (PBR)
  - Classic partitioned table spaces (using table controlled partitioning)
- The new limit keys are materialized in SYSTABLEPART in the SWITCH phase
- Restrictions
  - MQT, field-procedure, RI, index on expression, trigger

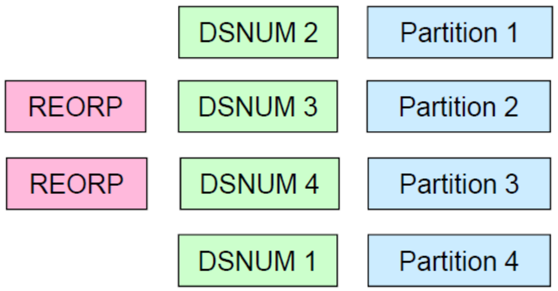
Online ALTER Partition Limit Keys

DB2 10

Alter Table ... Alter Partition 3 Ending At 350 Inclusive

SYSIBM.SYSTABLEPART

Part-ition	Logical _Part	Limitkey _Internal
2	1	200
3	2	~~300~~ 350
4	3	400
1	4	500



V11

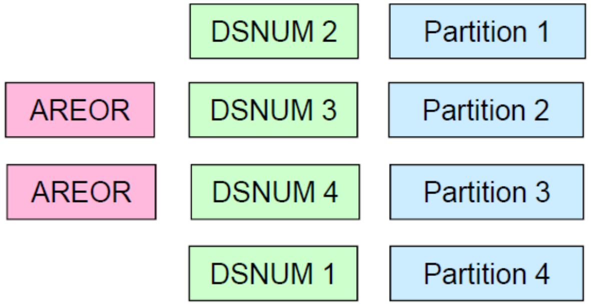
Alter Table ... Alter Partition 3 Ending At 350 Inclusive

SYSIBM.SYSTABLEPART

Part-ition	Logical _Part	Limitkey _Internal
3	2	300

SYSIBM.SYSPENDINGDDL

Option_Value	Part-ition
…350	3



- Behavior of ALTER limit key against index-controlled TS not affected - still REORP
  - PM89655 introduced PREVENT_ALTERTB_LIMITKEY and PREVENT_NEW_IXCTRL_PART to avoid ‘surprises’

DROP COLUMN Support

- Pending ALTER
- AREOR is set for the table space
- Materialization via REORG SHRLEVEL REFERENCE/CHANGE
  - Partitioned Table space: Materialization only if all partitions are addressed
- Invalidation of all packages and DSC that are dependent on the TB
- PIT recovery is not allowed (after materialization of the ALTER)
- SYSCOPY record with
  - ICTYPE=A (=alter)
  - STYPE=C (=column)
  - TTYPE=D (=drop)
- Restrictions
  - UTS only
  - Check SQL Ref for complete list

DROP COLUMN: Materialization of the Pending Change

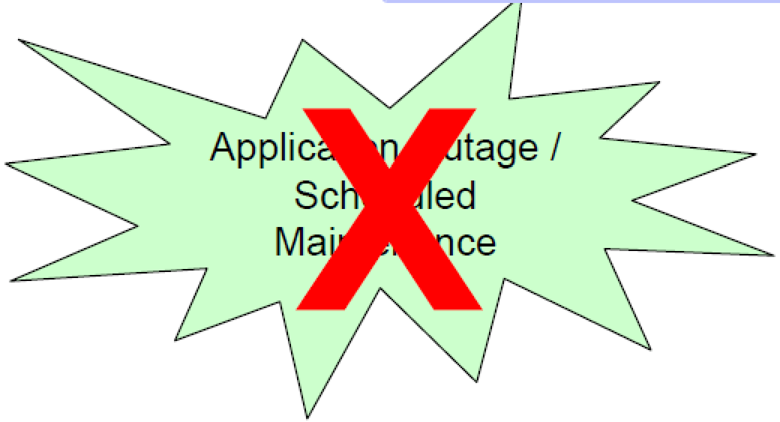
```
ALTER TABLE sc1.table1 DROP COLUMN Col_3 RESTRICT
REORG TABLESPACE DB1.TS1 SHRLEVEL REFERENCE
```

Application Table

Col_1	Col_2	Col_3	Col_4
Data_1	Data_2	Data_3	Data_4
Data_1	Data_2	Data_3	Data_4

ALTER TABLE

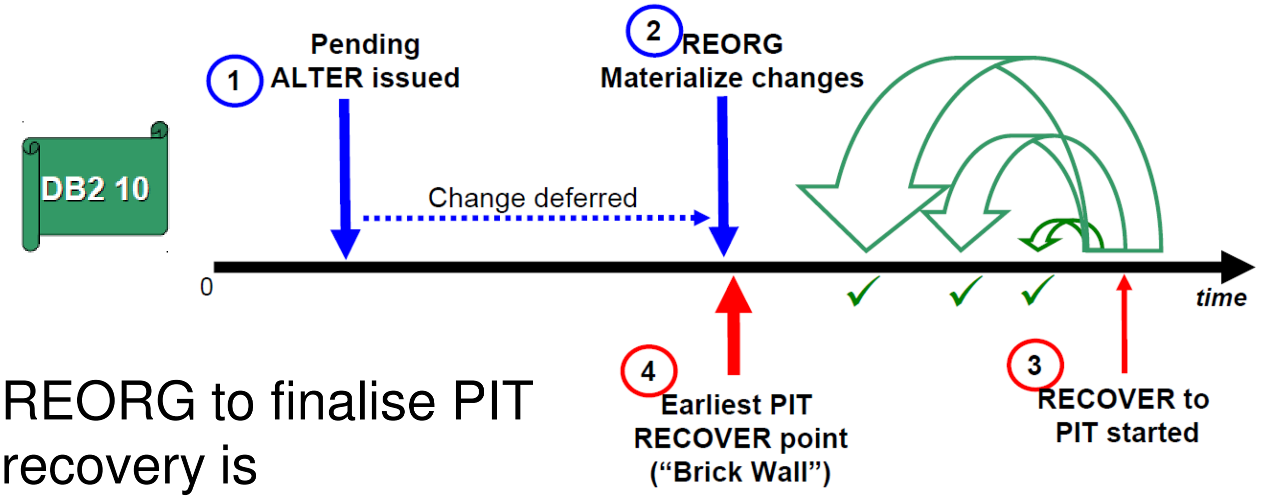
Online REORG



Application Table

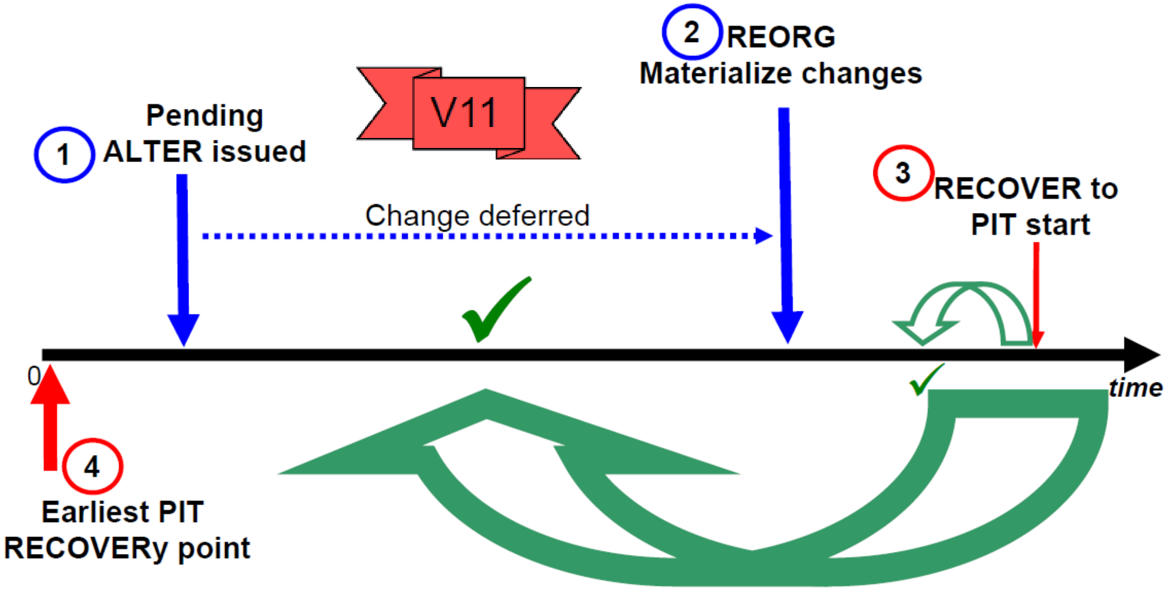
Col_1	Col_2	Col_4
Data_1	Data_2	Data_4
Data_1	Data_2	Data_4

PIT recovery for deferred schema changes (ALTERs)



Alter Segment Size
Alter Data Set Size
Alter BP T/S Page Size
Alter BP Index Page
Alter Member Cluster
Alter Limit key

- REORG to finalise PIT recovery is MANDATORY
- Recovery of **data only**, **changed schema remains**
- PIT recovery not supported for all pending alters



PIT recovery for deferred schema changes

- PIT recovery scenarios **supported**: table space attribute alters

(also with immediate ALTERs in the window between the materializing REORG and the PITR)

Pending ALTERs	PBR ts	LOB ts	XML ts
segsize	✓		✓
dssize	✓	✓	✓
Bufferpool	✓	✓	
member cluster	✓		

- PIT recovery scenarios **not supported**:
  - ✖ PBG table space (except when it is an XML tsp)
  - ✖ Indexes (but you can rebuild those)
  - ✖ All table space type conversions

🚫 RECOVER utility will issue existing message DSNU556I (recover cannot proceed), RC8 and terminate

RAS Improvements

- Cancel DDF Threads – new FORCE option
  - Prior command without FORCE must be issued first
  - Only DDF threads
  - z/OS 1.13 APAR OA39392 required (CALLRTM TYPE=SRBTERM)
- DRDA SQLCancel() improvements
  - when DB2 receives a DRDA SQL Interrupt from a remote client, it closes the connection and terminate the thread under which the statement is running, instead of interrupting just the statement
  - Interrupt even when waiting on locks, executing SPs, or statement forwarded to another DB2
- Open data set limit raised to 200K
- Restrict hybrid join to 80% of the total RID pool
- Query parallelism dynamic adjustment to available system resources (incl. new instrumentation counters)

RAS Improvements

- Workfile space shortage warning new system parameters, instrumentation and messages
  - WFSTGUSE_AGENT_THRESHOLD subsystem parameter
    WFSTGUSE_SYSTEM_THRESHOLD subsystem parameter
  - Systems programmer response to DSNI052I/DSNI053I
- Compression dictionary availability for CDC tables
  - Replication products problems when old dictionary was no longer available after a REORG
  - Dictionary is written to the log now
- DEFER DEFINE improved concurrency
  - Lock on DBD released as soon as the table or index space is physically defined (not wait for UR to commit)

Autonomics improvements – REORG avoidance

- Automatic index pseudo-delete cleanup
- Reduction of overflow rows and indirect references

Auto Cleanup of Pseudo-deleted Index Entries

- Pseudo-deleted index entries (introduced with Type 2 IX in V4)
  - Index entries typically not actually deleted from the index but marked pseudo deleted
  - Increases getpages, lock requests, CPU cost
  - (Applications may encounter deadlocks and timeouts during update processing)
- Prior to DB2 11
  - Inline code to do cleanup of PD index entries and PD empty pages
  - REORG INDEX required in most cases to remove pseudo-deleted index entries

Auto Cleanup of Pseudo-deleted Index Entries

- DB2 11 automatically cleans up pseudo-deleted entries
  - zIIP-eligible processing runs in the background
  - Designed to have minimal disruption to applications
  - New zparm (INDEX_CLEANUP_THREADS) to control number of concurrent cleanup tasks, default=10
  - New SYSIBM.SYSINDEXCLEANUP catalog table to control auto cleanup at index level
    - Day of week/month, start/end time
    - By default cleanup is enabled for all indexes
- Benefits of automatic pseudo-delete cleanup
  - Reduce size of some indexes, fewer getpages
  - Improve SQL performance in terms of lower CPU and lower elapsed time
  - **Reduce the need to run REORG INDEX**

Reduction of Overflow Rows and Indirect References

- Row updates to variable length and/or compressed rows can increase the length of the row
  - If not enough space on the data page DB2 moves the row to another data page and replaces the original row with a pointer record
  - Index entries continue to refer to the original row (RID)
  - RTS indicators REORGNEARINDREF and REORGFARINDREF
- Good reasons to avoid indirect references
  - Often causes additional I/O to read the extra data page into a buffer pool
  - REORG TABLESPACE required to remove indirect references

Reduction of Overflow Rows and Indirect References

- DB2 11 solution to reduce indirect references
  - New **PCTFREE .. FOR UPDATE ..** attribute to reserve free space for updates
    - Default is zero, or current behavior
    - Value you specify (1-99)
    - **-1 is the autonomic option** – DB2 figures out optimal setting using RTS
  - E.g. PCTFREE *x* FOR UPDATE *y*
    - *x* = % of free space to leave in each data page by LOAD or REORG
    - *y* = % of free space to leave in each data page by INSERT, LOAD or REORG. INSERT will preserve y% while REORG will preserve (x+y) %
  - Zparm PCTFREE_UPD (PERCENT FREE FOR UPDATE)
    – System default for FOR UPDATE value when it is not specified in DDL
  - Default setting = zero (0) - V10 behavior
  - **Reduce the need to run REORG TABLESPACE**
  - Available in NFM

Data Sharing Improvements

- Group buffer pool write-around to avoid CF cache structure flooding issues
  - Autonomic tracking whether the GBP is filling up
    - No external knobs
  - Intelligently writing certain pages to disk instead of GBP to avoid GBP full
    - Enabled/disabled based on GBP-level and CLASS-level internal thresholds
    - Only for pages async written to CF
    - When GBP duplexing, CF req. for pages in write-around written are serially
  - Data integrity: cross invalidation notice sent to all members when a page is written to DASD with write around.
    - Uses new IXLCACHE LOCALREGCNTL=YES option
  - Available in CM
  - Not retrofitted to V10 (PM70575-CAN)
    - Requires z/OS maintenance and a certain CF level

Data Sharing Improvements

- Castout enhancements:
  - Reduced wait time for I/O completion
  - Reduced notify message size sent to castout owner
  - More granular class castout threshold for large GBP size (#pages in addition to %pages)
- CF DELETE_NAME utilizes a new CF request option to suppress XI signals when deleting directory entries
  - Improves efficiency of DELETE_NAME especially for sysplex over extended distance
  - Retrofitted to DB2 10 and V9 with APAR PM67544
    - Make sure to check maint (DB2 and z/OS) – went PE a couple of times (incl. command option to disable CFLEVEL17 functionality)
  - Adds a safety net to detect unexpected errors

Data Sharing Improvements

- New LIGHT(CASTOUT) option on Restart Light
  - – Causes **all** retained locks to be removed – except in-doubt or postponed abort URs
    - LIGHT(YES) Page set P-locks in IX or SIX mode are not freed
    - If cannot resolve all indoubt and postponed-abort units of recovery, DB2 11 does not release the associated page set P-locks that are in IX or SIX mode.
      - – Eg. Possible reason can be that the LBACKOUT subsystem parameter is set to LIGHT or LIGHTAUTO
  - – Accomplished by initiating castout at end of Restart Light
  - – After castout, pagesets become non GBP-dependent and retained page set P-locks can be safely released
  - – Utilities can now be run after Restart Light completes

Data Sharing Improvements

- Index split performance and other index availability improvements
  - Avoid placing indexes in RBDP during group restart in rare cases
    - In DB2 10 REBUILD IX is needed get passed this
    - DB2 11 NFM uses a 2- step LPL/GRECP recovery process and issue DSNI051I
    - Should be rare
    - Reduce DB2 outage time
- Improve index split performance
    - Reduce multiple log force write I/Os in data sharing for index split operation
    - Reduce multiple log force write I/Os for pseudo-delete operation
    - Improve index split rollback performance
      - Reduce backout time by reducing several log force write I/Os on rollback of deleted pages

Data Sharing Improvements

- Auto LPL recovery improvements
  - Prior to DB2 11, when pages were added to the LPL by an active member while one of the members was down and holding retained locks, **no automatic LPL recovery performed when the failed member restarted**
    - Eg. Pages added to LPL after several retries with GBP full
    - Resolve the LPL by manually issuing a –START DB(xx) SPACE(yy) command
  - DB2 11 initiates **automatic LPL recovery** of objects at the end of normal restart and restart light
    - At the end of auto-LPL recovery, each member issues a DSNI049I
    - Still cases where cannot be done (eg. When PA objects involved)
- Avoid child 'U' lock propagation for single-member read-only
  - Suppresses any update U state child lock propagations until there is global contention on the parent page set P-lock
    - Better performance of SELECT FOR UPDATE statements in data
    - S-mode pageset P-lock sent as X to XES while no update interest (PM85053 IRLM)
- Full LRSN spin avoidance

Application Availability

- DB2 10 introduced BIF_COMPATIBILITY and DDF_COMPATIBILITY
  - To deal with certain release incompatibilities
    - To buy more time to allow applications to change
- DB2 11 introduces APPLCOMPAT (available in NFM)
  - Provide a 'fence' to better control
    - When new DML DB2 functionality is available
    - Release incompatibilities
  - Migrate an application at a time
    - In the past switch all applications on day#1 of new version to new behavior
    - At package level
      - APPLCOMPAT bind option (cannot bind with V11R1 until NFM)
      - Dynamic SQL is governed by the CURRENT APPLICATION COMPATIBILITY special register

Application Availability

- Detect the use of incompatible changes via traces
  - IFCID 239
    - Indicates Packages using a function that changes in DB2 11
    - Field QPACINCOMPAT
    - See SDSNMACS(DSNDQPAC) for mapping
  - IFCID 366/376
    - Records indicate SQL using the V10 code path which is different from the V11 code path
    - Use these in CM to identify programs needing review
    - IFCID376 is new in V11 and is a roll up of activity reported in 366
      - Attempts once per dynamic and static statement (bound V10 or later)
      - Once per Plan, Package, Statement # bound prior to V10
    - See SDSNMACS(DSNDQW05) for detailed description

Questions?

Thank
You