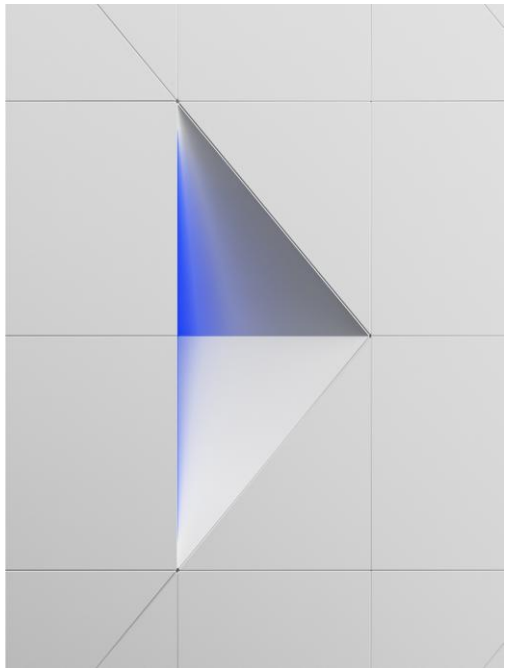
Db2 for z/OS data sharing updates

Mark Rader
mrader@us.ibm.com

Tori Felt
victoria.felt@ibm.com

Contents

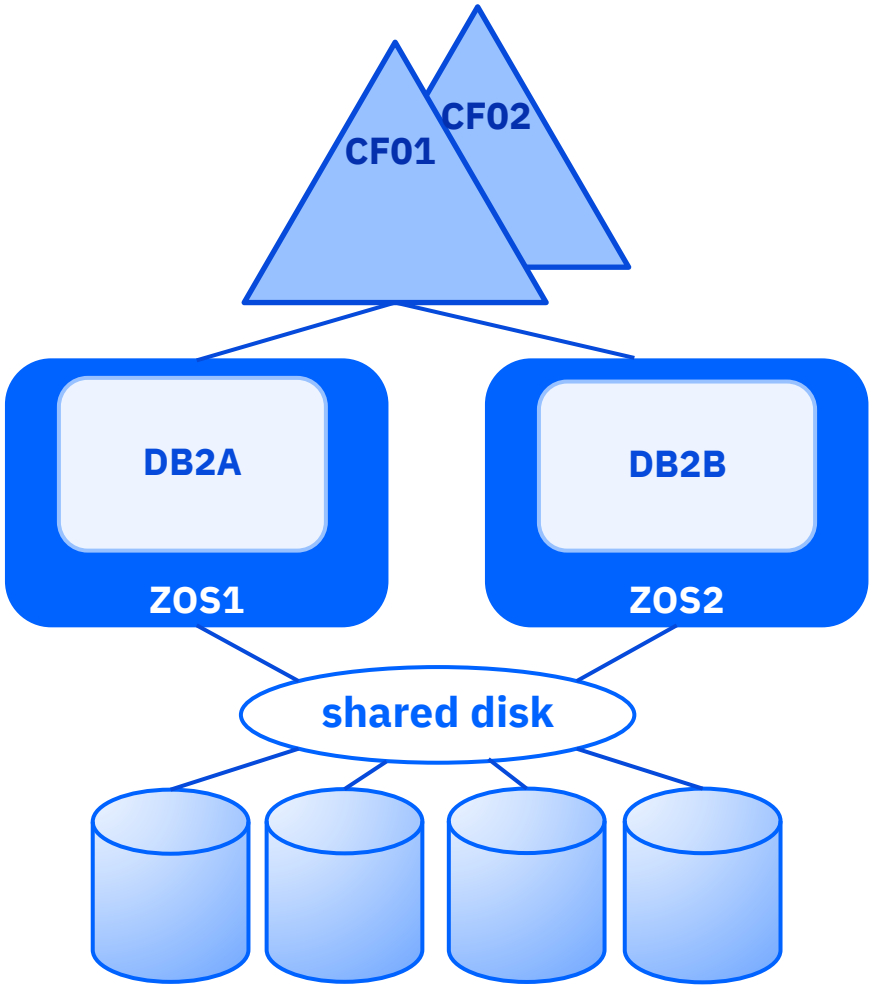


Overview of Db2 for z/OS data sharing

Why implement Db2 for z/OS data sharing?

Db2 data sharing and Parallel Sysplex provide infrastructure for highest levels of:

- Availability – protection against planned and/or unplanned outages
- Non-disruptive scalability – growth to handle business, market, or regulatory changes
- Dynamic workload balancing – using available system capacity, handling spikes

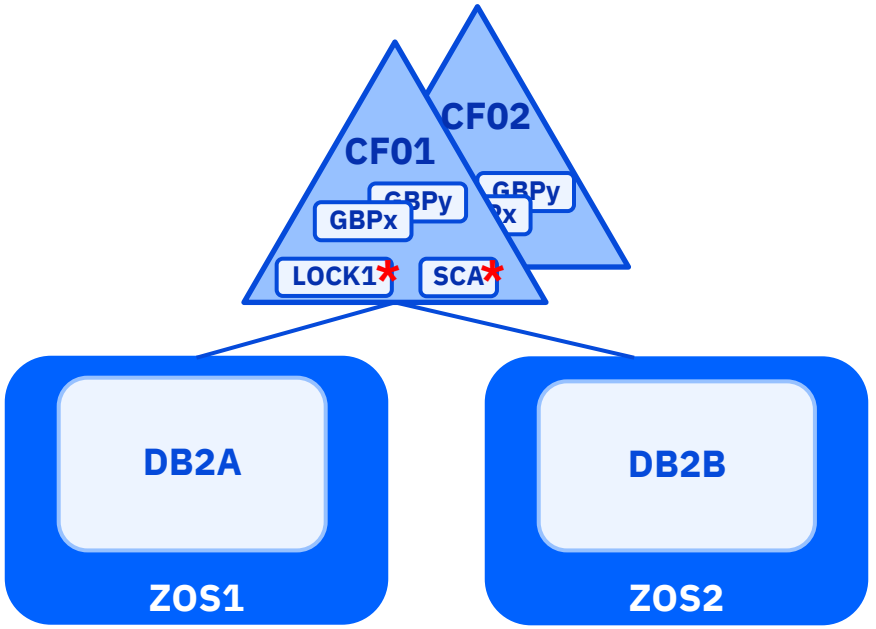


Db2 for z/OS data sharing concepts

Coupling facilities (CFs) contain:

- 1 lock structure (LOCK1) per data sharing group *
- 1 shared communications area (SCA) per group *
- Multiple group buffer pools (GBPs) per group
  - 1 GBP per local buffer pool (BP) containing shared data
  - Db2 allocates GBP0, GBP8K0, GBP16K0, GBP32K

* = Required structures
  - Db2 will not start if it cannot allocate LOCK1 or SCA
  - Db2 will fail if it loses access to LOCK1 or SCA and cannot rebuild it
- GBPs not 'required' strictly speaking
  - Duplexing highly recommended



More Db2 for z/OS data sharing concepts

LOCK1 contains record list entries (RLEs)

– Become retained locks if Db2 or IRLM lost

– If insufficient RLEs, IRLM does not accept more locks

GBPs track pages in local BPs (directory entries, data elements)

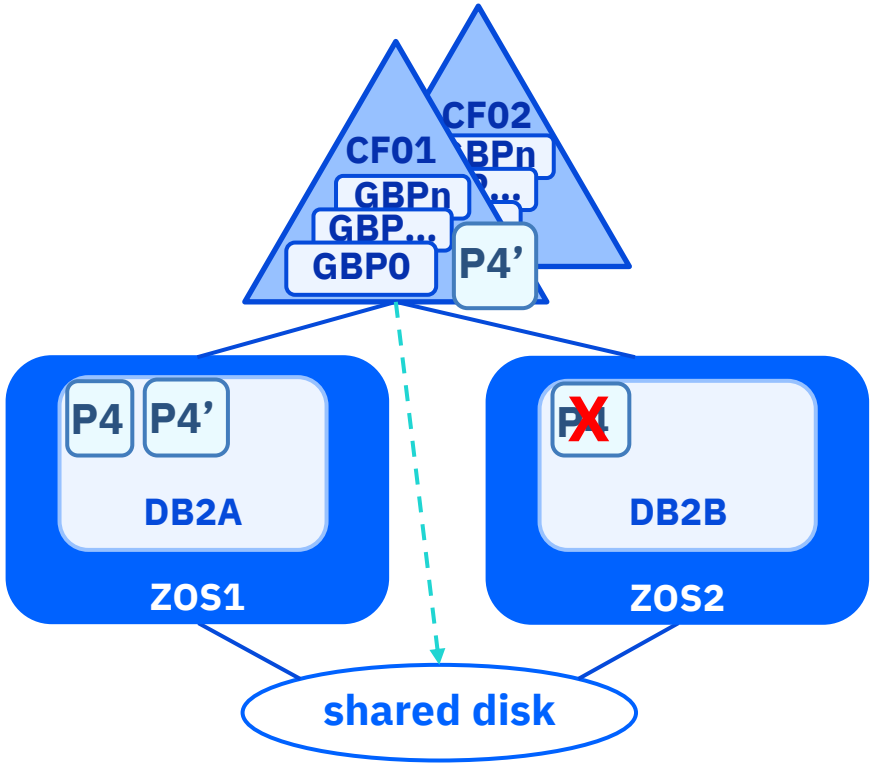
– Cross invalidated (XI) if changed on other member

- DB2A changes P4 to P4’, which is written to GBP
- DB2B copy of P4 invalidated [X]

GBP pages castout to disk (green arrow)

– Via special buffers in castout owner

– Allows room for more writes to GBP



A specific Parallel Sysplex configuration

2-CEC, 2-ICF configuration

– SCA and LOCK1 not isolated from Db2, IRLM

– Duplexed GBPs spread across CFs

What if one system is lost?

– It depends…

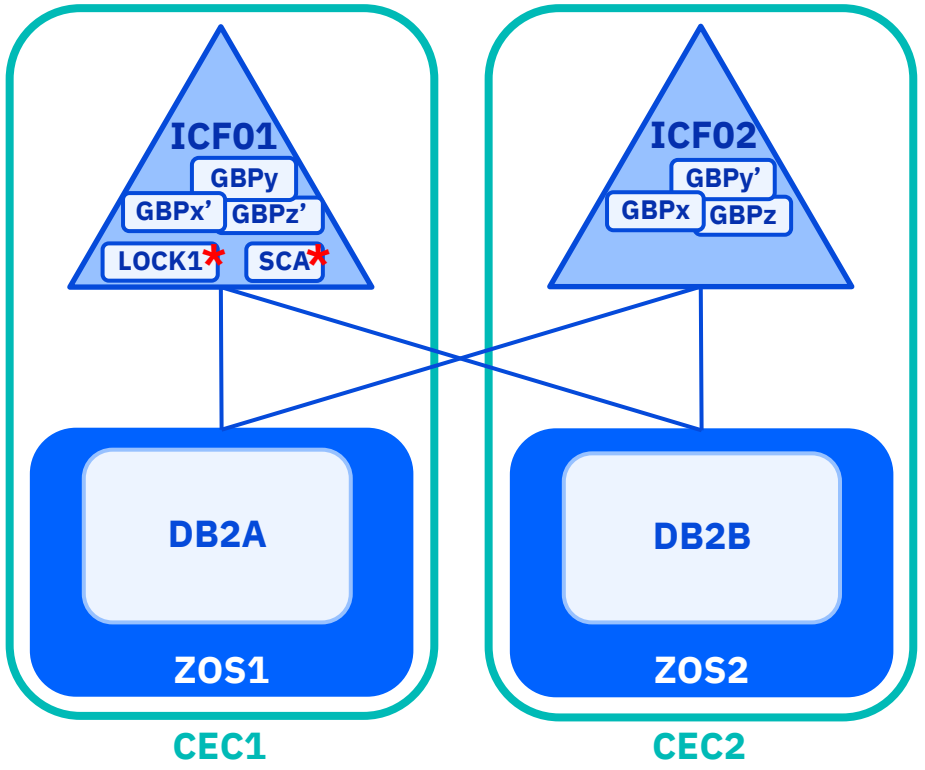
– If CEC2, would be loss of 1 Db2: DB2B

  • DB2A would continue running

– If CEC1, would be loss of entire data sharing group

  • DB2B alone would not be able to rebuild LOCK1 and SCA, so DB2B would fail, too

  • Here CEC1 represents a single point of failure (SPoF)



Db2 12 for z/OS data sharing

Db2 12 for z/OS base data sharing

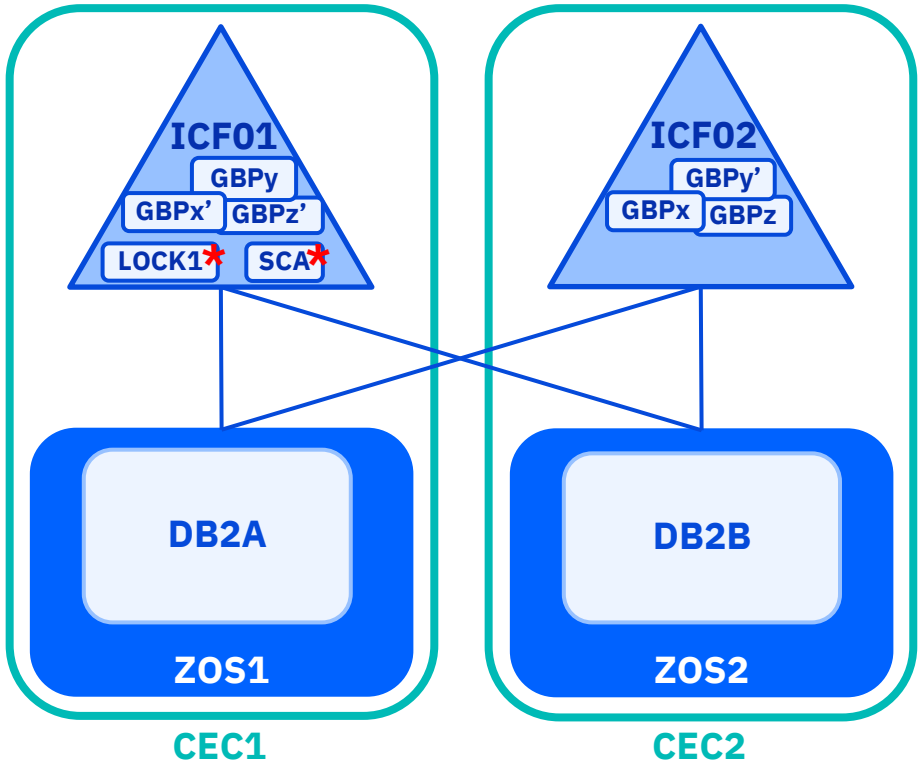
Asynchronous CF duplexing for LOCK1 (more to come on this)

Lock avoidance: improved performance, less need for short commit scopes

Db2 peer recovery: alternative to automated restart of failed Db2 member

Retry of automatic recovery for logical page list (LPL) and GBP recovery pending (GRECP) conditions

- Pages added to LPL if repeated inability to write to GBP
- Objects set to GRECP if GBP(s), or CF containing GBPs, lost and unable to rebuild
- Db2 has been automatically recovering LPL and GRECP for some time, now retries the automatic recovery

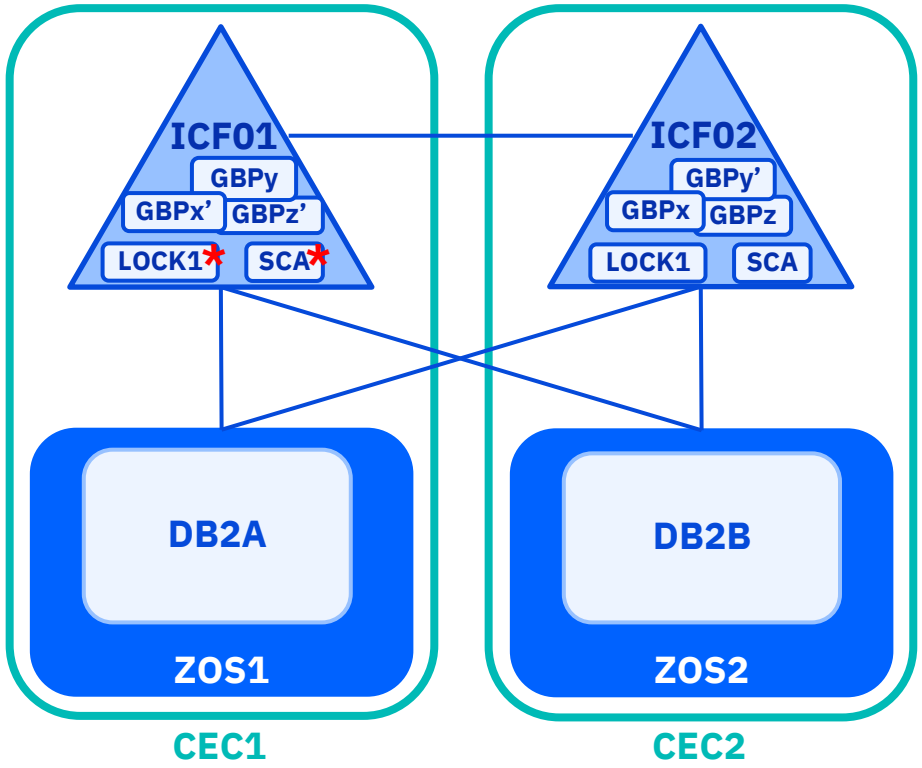


Asynchronous duplexing of LOCK1

Ideally for 2-CEC, 2-ICF configuration

Avoid CEC with IRLM as single point of failure by duplexing LOCK1 and SCA

- Option 1: synchronous duplexing of LOCK1 (since 2001)
  - Lock service time 3x - 4x > simplex lock service time
- Option 2: asynchronous duplexing of LOCK1 (Db2 12)
  - Lock service time 1.2 x simplex lock service time
- In either case, synchronous duplexing of SCA
  - SCA duplexing not a performance concern



Db2 12 for z/OS: post-GA data sharing updates

IRLM deadlock process improvements: suppress DSNT376I

– Still produce IFCID 196 for deadlock info

Asynchronous cross-invalidation of GBPs

– Greatest benefit with long CF link distance

-ALTER GROUPBUFFERPOOL (-ALT GBPOOL) ratio changes (ratio of directory entries to data elements)

– Ratio limit raised from 255 to 1024; default ratio raised from 5 to 10

Support for multi-factor authentication (MFA) and IBM RACF PassTickets

– Sysplex group authentication for IBM data server driver with Sysplex workload balancing (WLB) or seamless failover
  - AUTHEXIT_CACHEREFRESH=ALL required (DSNZPARM)

– MFA if no WLB or seamless failover: MFA_AUTHCACHE_UNUSED_TIME
  - Also requires AUTHEXIT_CACHEREFRESH=ALL

Db2 13 for z/OS and data sharing

Db2 13 for z/OS data sharing enhancements

IRLM dynamic increase of LOCK1 size, producing more RLEs

GBP residency time

GBP castout full conditions:

– Increased castout frequency

– More frequent GBP-write retries

Reduced P-lock contention for PBR table spaces with relative page numbers and row level locking

– UTS PBR RPN

Request Sysplex Workload Balancing via MODIFY DDF

APAR PH65562: reduce impact of changing GBP dependency

IRLM dynamic increase of LOCK1 size (1|2)

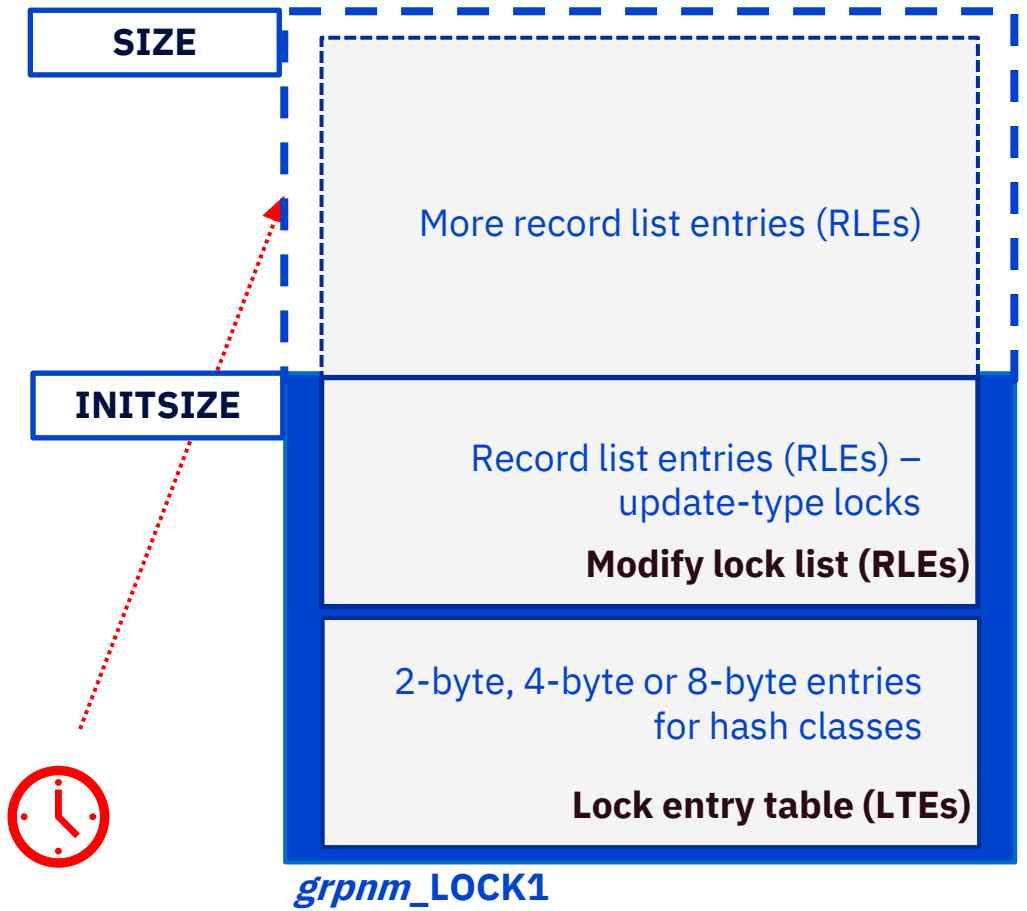
Db2 lock structure (grpnm_LOCK1) allocated based upon CFRM policy

- INITSIZE = initial allocation, SIZE = max size
- Lock entry table = $2^n \leq 0.5$ x INITSIZE
- Remainder holds RLEs; too few RLEs leads to IRLM slowdown and/or application timeouts

Increase structure allocation to get more RLEs

- 1. Operator or automation
  - SETXCF START,ALTER,STRNM=grpnm_LOCK1,SIZE=newsize
- 2. XES: CFRM ALLOWAUTOALT(YES)
  - FULLTHRESHOLD > 0

Under heavy workload, neither option is responsive enough to increase number of RLEs



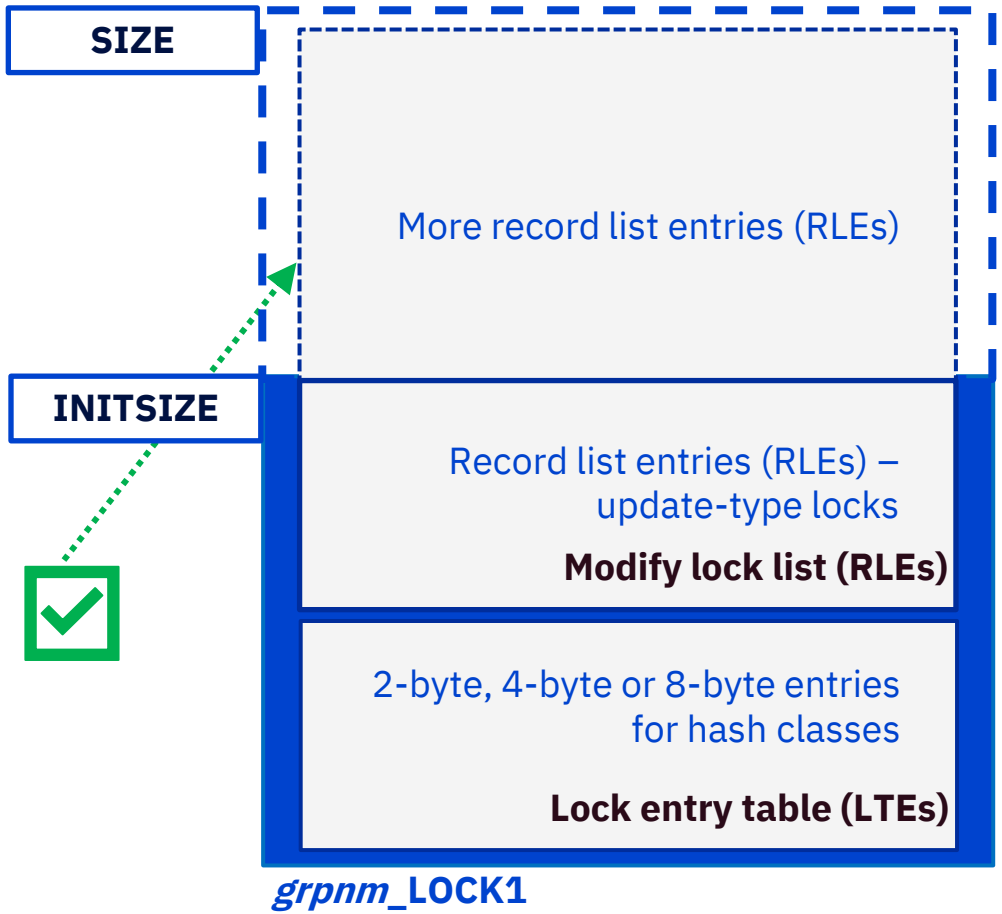
IRLM dynamic increase of LOCK1 size (2|2)

Db2 13: IRLM initiates dynamic alter of lock structure to avoid structure full conditions

- Result is equivalent of SETXCF START,ALTER, STRNM=grpnm_LOCK1,SIZE=newsize
- CFRM ALLOWAUTOALT (YES) should be specified
- Expansion of LOCK1 results in more RLEs

Under heavy workload IRLM request is more responsive

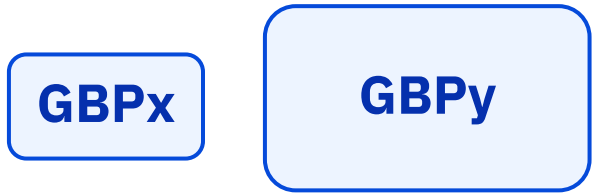
```
DXR189I  <irlmname> ALTERING LOCK STRUCTURE SIZE
DXR190I  <irlmname> ALTER LOCK STUCTURE COMPLETED

IXC530I START ALTER REQUEST FOR STRUCTURE
grpnm_LOCK1 ACCEPTED
  TARGET  SIZE:                  256 M
```



Group buffer pool (GBP) residency time

Challenge:

– More information required to tune GBPs effectively and to balance resources between GBPs



Solution:

– Collect residency time for directory entries and data elements

  • Record average in microseconds of residency time for directory entries and data elements in IFCIDs 230, 254

    – The higher the better, unless zero reclaims

– New message DSNB820I added to DISPLAY GROUPBUFFER POOL GDETAIL option

```
DSNB820I - AVERAGE RESIDENCY TIME
             FOR DIRECTORY ENTRIES         = directory-entry-reside-time
             FOR DATA ENTRIES              = data-area-reside-time
```

– Requires z/OS 2.4 or 2.5 and GBP in CF on z16+ with CF control code (CFCC) 25 or higher

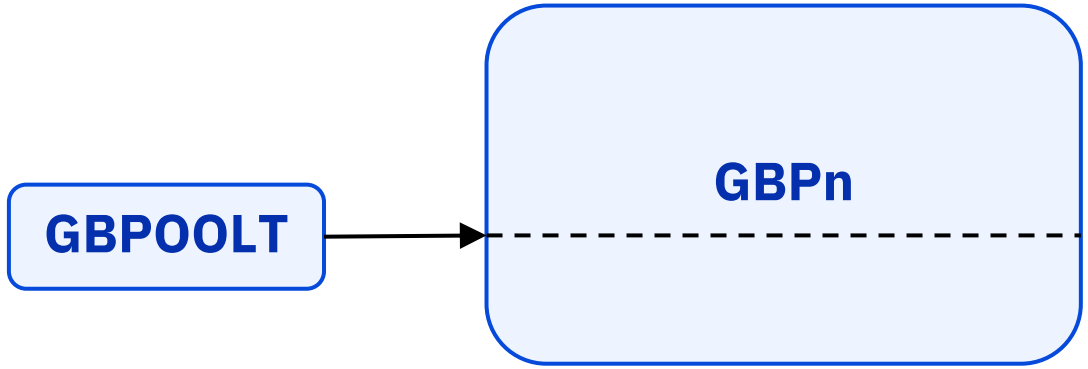
GBP full conditions (1|2)

GBP castout is critical for busy data sharing systems

- Castout delays can lead to GBP full, application delays, and coupling facility (CF) message overhead
- GBP full can lead to adding pages to the logical page list (LPL)
  - Pages on LPL are not available for any process
- Group buffer pool threshold (GBPOOLT) monitoring drives group level castout
  - GBPOOLT (30) can be adjusted with ALTER GBPOOL command
  - Threshold monitoring frequency values are static

Db2 13 behavior

- GBPOOLT threshold monitoring
  - Frequency increased
  - Castout triggered more quickly
- Overall reduced likelihood of delays due to GBP full conditions



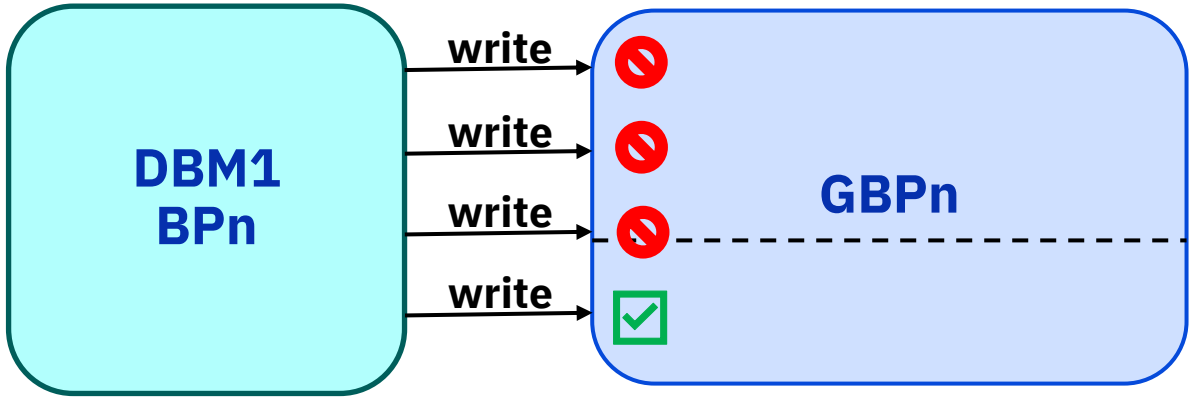
GBP full conditions (2|2)

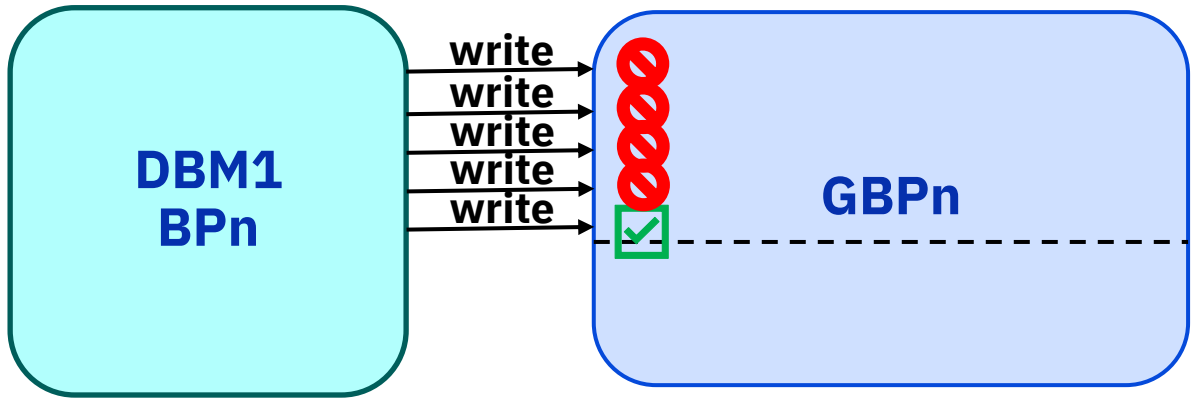
GBP castout is critical for busy data sharing systems

– If Db2 cannot write a page to the GBP, then Db2 will retry the write
  - Retry occurs until the write is successful or until page added to the LPL
  - “Transaction pacing”: interval between retries

Db2 13 behavior

– “Transaction pacing”
  - Frequency increased dramatically
  - Faster response to GBP full relief
– Overall reduced likelihood of delays due to GBP full conditions





Reduced P-lock contention for UTS PBRs with RPN

Db2 12 introduced relative page numbering (RPN) for universal PBR table spaces

– RPN advantages versus absolute page numbering:
  - Much greater data capacity
  - Maximum number of partitions not affected by choice of page size or DSSIZE
  - DSSIZE can be specified at partition level (and DSSIZE increase is immediate change)

Performance tip for RPN table spaces defined ***with LOCKSIZE ROW in a data sharing system***: after activating function level V13R1M500, online REORG those table spaces

– Why?
  - Online REORG will modify header pages of table space's partitions so that Db2 will use a new lock hashing algorithm for the data page P-lock requests associated with row-level L-locks
  - New hashing algorithm will boost CPU efficiency by reducing page P-lock contention
– Note that this online REORG can be done at the partition level – full-table space REORG not required
– For RPN table spaces created after activation of function level V13R1M500, Db2 will use the new hashing algorithm for data page P-locks – no need to REORG those table spaces to get that performance benefit

Db2-controlled Sysplex Workload Balancing

Sysplex workload balancing (transaction-level workload balancing)

- High availability for DDF client applications
- In some situations, Sysplex workload balancing is not being exploited by DDF clients as often as desired
  - Now you can initiate Sysplex workload balancing
- -MODIFY DDF (Db2) command changes (PH48253; October 2022)
  - RQSTWLB – request client driver to enable Sysplex workload balancing; Db2 LUW 11.5 or later, JDBC 4.26.14 or later
    - * enabled for entire data sharing group (default)
    - *location-name*
    - *alias-name*
  - DFLTWLB – honor Sysplex workload balancing option requested by client
    - * - enabled for entire data sharin group (default)
    - *location-name*
    - *alias-name*

Db2 13 for z/OS: APAR PH65562 (April 2025)

Refresh cross-invalidated BP pages

– Pseudo-close required for recovery points in SYSLGRNX

– Pseudo-close removes GBP dependency – drives cross-invalidation

– Pseudo-close controlled by PCLOSET subsystem parameter (DSNZPARM)

Issue: spiky, intermittent workloads

– Performance depends on pages always being in BP, especially indexes

– Cross-invalidation results in synchronous reads, which impacts performance

Solution in 2 parts

– Refresh prefetch: refresh pages after cross invalidation, avoid sync I/O

  • ALTER BUFFERPOOL REFPF (NO, YES)

– BPOOL level pseudo-close interval: overrides PCLOSET

  • ALTER BUFFERPOOL VPCLOSET (integer; initial default 0)

Questions

Summary

Db2 12 and Db2 13 data sharing features

– Availability

– Performance

– Operations

– Security

And more with continuous delivery

Thank you

This document is current as of the initial date of publication and may be changed by IBM at any time. Statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

THIS DOCUMENT IS DISTRIBUTED "AS IS" WITHOUT ANY WARRANTY, EITHER EXPRESS OR IMPLIED. IN NO EVENT, SHALL IBM BE LIABLE FOR ANY DAMAGE ARISING FROM THE USE OF THIS INFORMATION, INCLUDING BUT NOT LIMITED TO, LOSS OF DATA, BUSINESS INTERRUPTION, LOSS OF PROFIT OR LOSS OF OPPORTUNITY.

Client examples are presented as illustrations of how those clients have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary.

Not all offerings are available in every country in which IBM operates.

It is the user's responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs.

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.