IBM DB2® for z/OS: Data Sharing Technical Deep Dive

Mark Rader
mrader@us.ibm.com
IBM zGrowth Unit (ATS)

© Copyright IBM Corporation 2014
Acknowledgements and Disclaimers

Availability. References in this presentation to IBM products, programs, or services do not imply that they will be available in all countries in which IBM operates.

The workshops, sessions and materials have been prepared by IBM or the session speakers and reflect their own views. They are provided for informational purposes only, and are neither intended to, nor shall have the effect of being, legal or other guidance or advice to any participant. While efforts were made to verify the completeness and accuracy of the information contained in this presentation, it is provided AS-IS without warranty of any kind, express or implied. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this presentation or any other materials. Nothing contained in this presentation is intended to, nor shall have the effect of, creating any warranties or representations from IBM or its suppliers or licensors, or altering the terms and conditions of the applicable license agreement governing the use of IBM software.

All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer. Nothing contained in these materials is intended to, nor shall have the effect of, stating or implying that any activities undertaken by you will result in any specific sales, revenue growth or other results.

© Copyright IBM Corporation 2014. All rights reserved.

IBM, the IBM logo, ibm.com, DB2, Parallel Sysplex are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (@ or “®”), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml

Other company, product, or service names may be trademarks or service marks of others.
Assumptions

• This is a technical discussion of DB2 for z/OS Data Sharing topics

• The audience should be familiar with DB2 for z/OS Data Sharing concepts, behavior and benefits, based on
  • Experience with a data sharing environment
  • Recent data sharing education
  • DB2 for z/OS publications or Redbooks®
    – Flexible capacity
    – Scalability
    – High availability
    – Dynamic workload balancing
Acronyms

- CF – Coupling Facility LPAR
  - ICF – Integrated CF, aka Internal CF
- CFRM – CF Resource Management, definitions in CFRM policy
- CFCC – CF Control Code
- CF Links – connectivity between CF LPAR and ‘host’ CECs
  - ISC – fiber links, medium to long distance
  - ICB – copper links, very short distance
  - PSIFB - InfiniBand® links, short (12X IB) to long (1X IB) distance
  - IC – internal, microcode links for ICFs
- XCF – Cross-System Coupling Facility – communication between CECs
- XES – Cross-System Extended Services, z/OS component that manages CFs
Agenda

• DB2 Data Sharing
  – Configurations
  – Standard CF interaction
  – Performance monitoring
  – Auto Alter
• Workload growth
  – Lock structure
  – GBPs
  – Changes in configuration
    • CF considerations
• What’s New in DB2 10 and DB2 11
DB2 Data Sharing Starting Configuration

• Starting with DB2 V4
DB2 Data Sharing: Usual Configuration

- Introduction of ICF

- SCA and LOCK1 on external CF; isolated from DB2 and IRLM members
- Duplexed GBPs spread across CF01 and ICF2
DB2 Data Sharing: 2-ICF Configuration

- Reduced number of CEC footprints
- Risk of ‘double failure’: DB2 and SCA, IRLM and LOCK1
  - If structure and exploiter fail, other members fail, too.

- Duplexed SCA and LOCK1 strongly recommended in this configuration
  - DB2B remains active, even if CEC on left is lost
  - Additional cost: host CPU, CF CPU, and CF link busy
Data Sharing: Locking

- Global locking using Parallel Sysplex® coupling technology
  - Inter-system concurrency control

  ![Diagram showing data sharing and locking](image)

  - Cost of obtaining lock does not increase when adding 3rd through nth members
  - This example assumes no contention
Notes: Lock Structure (LOCK1)

- Used by IRLM to manage global locking

- Holds L-locks and P-locks
  - L-locks to track concurrency
    - Parent L-locks: e.g. table space intent locks
    - Child L-locks: page or row locks
    - Others...
  - P-locks to track coherency. Examples:
    - Page set P-locks: table space, partition, index, index partition
    - Page P-locks: data page (RLL), index leaf page, space map page
    - Others...

- Consists of a lock table (hash table) and a modify lock list
  - Lock table controls access to resources
    - One entry can record multiple readers and one updater (owner)
  - Modify lock list contains detailed information for update-type locks
    - Entries become retained locks in case of an IRLM or DB2 failure
Lock Structure (LOCK1)

- Simplified view

Modify Lock List

Lock Table

- 'S' lock: several readers
- 'X' lock owner: DB2A
- 'X' lock owner: DB2B

Modify Lock List entries (RLE) record info for update-type locks (U, X, IX, SIX)

* Examples

Lock Table Entries (LTE) correspond to hash classes
Data Sharing: Managing changed data

- Inter-system buffer coherency control
  - Example: DB2A has write interest in the table space, and page P1 is in DB2A’s buffer pool

  ![Diagram](image)

  - * Cross-invalidate (XI) to other member without interrupt
Notes: Group Buffer Pools (GBPs)

- DB2 uses GBPs to
  - Manage buffer coherency
  - Cache changed pages
    - Optionally cache read-only pages
- GBP consists of directory entries and data elements
  - Directory entries manage coherency by tracking interest in a data or index page by any DB2 member in the data sharing group
    - There is one directory entry for each page in the aggregate pool, no matter how many DB2 members have a copy of that page
  - Data elements are the cached pages that a DB2 member changed
  - In GBP duplexing, data elements exist in both the primary and secondary GBP
    - Directory entries in secondary GBP only exist for the changed pages
Group Buffer Pool (GBP)

• Simplified view
GBP Duplexing

- DSNDB20_GBP2 – Primary; “Old” on CF01

Data Elements

Page 1
Page 13
Page 27

Directory Entries

TS A Pg 1
1 0 1 1
x y u

TS A Pg 13
1 1 1 1
z w v s

TS B Pg 27
0 0 0 1
t

- Cache changed pages
- Register interest
- Search if local buffer miss
- Castout

- DSNDB20_GBP2 – Secondary; “New” on ICF2

Data Elements

Page 1
Page 13
Page 27

Directory Entries

TS A Pg 1
0 0 0 1
u

TS A Pg 13
0 0 0 1

TS B Pg 27
0 0 0 1
t

- Cache changed pages
Monitoring LOCK1

• RMF CF Activity Report
  – Structure Summary

| STRUCTURE TYPE | NAME         | STATUS | CHG | SIZE | STOR | REQ | CF | % OF CF | % OF ALL CF | AVG CF | REQ/SEC | TOT/CUR | TOT/CUR | TOT/CUR | TOT/CUR |
|----------------|--------------|--------|-----|------|------|-----|----|--------|------------|--------|---------|---------|---------|---------|---------|---------|
| LOCK           | DSNDB2R_LOCK1 | ACTIVE |     | 16M  |      | 0.1 |    | 0.0    | 0.0        | 0.00   | 24K     | 0       | 4194K   | N/A     | 13      | 5717    |
|                | DSNDB2P_LOCK1 | ACTIVE |     | 64M  |      | 0.5 |    | 10.5   | 0.0        | 1648.5 | 100K    | 0       | 1792    | N/A     | 2079    | 2579    |
|                | DSNDB2Q_LOCK1 | ACTIVE |     | 16M  |      | 0.1 |    | 0.0    | 0.0        | 0.00   | 24K     | 0       | 4194K   | N/A     | 272     | 48K     |
|                | DSNDB2R_LOCK1 | ACTIVE |     | 16M  |      | 0.1 |    | 0.0    | 0.0        | 0.00   | 24K     | 0       | 4194K   | N/A     | 13      | 5717    |
Notes: Key Points – LOCK1 Structure Summary

• Size can be an issue
  – Determines the number of Lock Table entries (LTE) and space for Modify Lock List entries (RLE)

• Requests per second is important
  – “Busy” is relative; < 5K/sec is not very busy
  – Observed: 166K/sec – very busy

• LIST/DIR ENTRIES = Modify Lock List entries (RLE)

• LOCK ENTRIES = 2-byte Lock Table entries (LTE)
  – May be 4- or 8-byte entries if > 7 members in the data sharing group
  – IRLM automatically rebuilds the lock structure when the 8th member (4-byte entries) or 23rd member (8-byte entries) joins the data sharing group
**RMF CF Activity Report**

- **Structure Activity**

<table>
<thead>
<tr>
<th>Structure Name</th>
<th>Type</th>
<th>Status</th>
<th>Total Req</th>
<th>% of Total</th>
<th>Avg Serv Time (Msec)</th>
<th>Reason</th>
<th>% of Reason</th>
<th>Delayed Requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSNDB2P_LOCK1</td>
<td>LOCK</td>
<td>ACTIVE</td>
<td>784K</td>
<td>0</td>
<td>16.8</td>
<td>NO SCH</td>
<td>94.5</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1447</td>
<td>0</td>
<td>3.4</td>
<td>PR CMP</td>
<td>1.9</td>
<td>-CONT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-FALSE CONT</td>
</tr>
</tbody>
</table>

**Contention Summary**

- **SYSA**
  - 567K Sync: 35.3% of total, Avg Serv Time: 44.6
  - 630.1 Async: 3.0% of total, Avg Serv Time: 150.0
  - 0 Change: 0.0% Included in Async

- **SYSB**
  - 916K Sync: 57.6% of total, Avg Serv Time: 43.5
  - 1017 Async: 4.2% of total, Avg Serv Time: 147.3
  - 0 Change: 0.0% Included in Async

**Total Delayed Requests**

- 2040K
- 3457
- 1447

**Contention Details**

- Req Total: 784K
- Req Deferred: 4634
- Req Total: 1256K
- Req Deferred: 5437
- Req Total: 2040K
- Req Deferred: 10K
- Req Total: 8901
- Req Deferred: 1447
Notes: Key Points – LOCK1 Structure Activity

# REQ TOTAL
- These are requests on the subchannel
  - Compare with EXTERNAL REQUEST CONTENTIONS: REQ TOTAL, which reflects API requests to XES and should be the higher number

- SERV TIME(MIC) – service time in microseconds
  - SYNC is key metric – ‘good’ number is relative to CF configuration
    - If ASYNC is non-zero it could be ‘block unlock’, or some requests were converted, either due to subchannel busy or heuristic algorithm

- CONT and FALSE CONT
  - Contention - recommend: CONT/REQ TOTAL < 2%
  - False Contention - recommend: FALSE CONT/REQ TOTAL < 1%
    - If higher, adjust size of LOCK1 to double size of Hash Table
Heuristic Algorithm and LOCK1

• Most LOCK1 requests are synchronous CF requests
  – Synchronous CF request means host CP is busy for duration of request
  – Long synchronous service times = high host CPU overhead

• XES can convert synchronous request to asynchronous
  – Heuristic algorithm based on measured lock service times
  – Host CP can now do other work during CF request
  – There is some host CP cost to setting up asynchronous request
  – Also elapsed time impact on lock requests
Monitoring GBPs

- RMF CF Activity Report

- Structure Summary

<table>
<thead>
<tr>
<th>STRUCTURE</th>
<th>NAME</th>
<th>STATUS</th>
<th>ALLOC</th>
<th>STORAGE</th>
<th>% OF</th>
<th>% OF</th>
<th>% OF</th>
<th>AVG</th>
<th>LST/DIR</th>
<th>DATA</th>
<th>LOCK</th>
<th>DIR REC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CACHE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSNDB2P_GB0</td>
<td>ACTIVE</td>
<td>34M</td>
<td>0.3</td>
<td>529</td>
<td>0.0</td>
<td>0.0</td>
<td>0.59</td>
<td>29K</td>
<td>5732</td>
<td>N/A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DSNDB2P_GB1</td>
<td>ACTIVE</td>
<td>501M</td>
<td>4.1</td>
<td>18380</td>
<td>0.1</td>
<td>0.0</td>
<td>20.41</td>
<td>194K</td>
<td>823</td>
<td>N/A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DSNDB2P_GB16X0</td>
<td>ACTIVE</td>
<td>8M</td>
<td>0.1</td>
<td>120</td>
<td>0.0</td>
<td>0.0</td>
<td>0.13</td>
<td>127K</td>
<td>1876</td>
<td>N/A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DSNDB2P_GB16X1</td>
<td>ACTIVE</td>
<td>32M</td>
<td>0.3</td>
<td>42641</td>
<td>0.3</td>
<td>0.0</td>
<td>47.35</td>
<td>50K</td>
<td>2876</td>
<td>N/A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DSNDB2P_GB2</td>
<td>ACTIVE</td>
<td>2G</td>
<td>16.8</td>
<td>8681</td>
<td>0.1</td>
<td>0.0</td>
<td>9.64</td>
<td>12356</td>
<td>412K</td>
<td>N/A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DSNDB2P_GB3</td>
<td>ACTIVE</td>
<td>8M</td>
<td>0.1</td>
<td>94</td>
<td>0.0</td>
<td>0.0</td>
<td>0.10</td>
<td>844</td>
<td>844</td>
<td>N/A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DSNDB2P_GB32X0</td>
<td>ACTIVE</td>
<td>10M</td>
<td>0.1</td>
<td>132</td>
<td>0.0</td>
<td>0.0</td>
<td>0.15</td>
<td>840</td>
<td>1344</td>
<td>N/A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DSNDB2P_GB32X1</td>
<td>ACTIVE</td>
<td>16M</td>
<td>0.1</td>
<td>120</td>
<td>0.0</td>
<td>0.0</td>
<td>0.13</td>
<td>1438</td>
<td>2862</td>
<td>N/A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DSNDB2P_GB5</td>
<td>ACTIVE</td>
<td>256M</td>
<td>2.1</td>
<td>358</td>
<td>0.0</td>
<td>0.0</td>
<td>0.40</td>
<td>521K</td>
<td>13K</td>
<td>N/A</td>
<td>0</td>
<td>7</td>
</tr>
</tbody>
</table>
Notes: Key Points – GBPs Structure Summary

- Size and requests per second important
- LIST/DIR ENTRIES = directory entries
- DATA ELEMENTS = data pages
  - If current directory entries = current data pages, probably secondary GBP (GBP duplexing)
    - Could also be the effect of Auto Alter
- DIR REC/DIR REC XI’S = directory reclaims / cross-invalidations (XI's) due to directory reclaims
  - Should be zero! Investigate if non-zero, especially XI’s
    - If DIR REC XI’S non-zero, potential performance impact
  - CF report does not have directory reclaim details
    - Use –DIS GBPOOL GDETAIL
Monitoring GBPs, cont.

- RMF CF Activity Report
  - Structure Activity

<table>
<thead>
<tr>
<th>SYSTEM</th>
<th>AVG/SEC</th>
<th># REQ</th>
<th>ALL</th>
<th>AVG</th>
<th>STD_DEV</th>
<th>NO SCH</th>
<th>MRK</th>
<th>INCL</th>
<th>CHNGD</th>
<th>DUMP</th>
<th>XI'S</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYSA</td>
<td>12.4</td>
<td>19</td>
<td>0.1</td>
<td>12.0</td>
<td>6.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ASYNC</td>
<td>12K</td>
<td>67.6</td>
<td>73.2</td>
<td>297.3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CHNGD</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>INCL</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ASYNC</td>
<td>58K</td>
<td>32.0</td>
<td>54.7</td>
<td>110.9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CHNGD</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>INCL</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>SYSB</td>
<td>51</td>
<td>44</td>
<td>0.2</td>
<td>54.9</td>
<td>246.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

**Secondary GBP**

---

<table>
<thead>
<tr>
<th>SYSTEM</th>
<th>AVG/SEC</th>
<th># REQ</th>
<th>ALL</th>
<th>AVG</th>
<th>STD_DEV</th>
<th>NO SCH</th>
<th>MRK</th>
<th>INCL</th>
<th>CHNGD</th>
<th>DUMP</th>
<th>XI'S</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ASYNC</td>
<td>18K</td>
<td>100</td>
<td>67.3</td>
<td>253.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CHNGD</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DUMP</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

---

Secondary GBP
Monitoring GBPs, cont.

- **RMF CF Activity Report**
  - **Structure Activity**

<table>
<thead>
<tr>
<th>STRUCTURE NAME</th>
<th>TYPE</th>
<th>STATUS</th>
<th>TOTAL #</th>
<th># % OF</th>
<th>SERV TIME(MIC)</th>
<th># % OF</th>
<th>AVG</th>
<th>STD_DEV</th>
<th># % OF</th>
<th>AVG TIME(MIC)</th>
<th>STD_DEV</th>
<th>ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSNDB2P_GBP1</td>
<td>CACHE</td>
<td>ACTIVE</td>
<td>599K</td>
<td>90.9</td>
<td>16.6</td>
<td>66.7</td>
<td>943</td>
<td>0.2</td>
<td>1266</td>
<td>2665</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>665.9</td>
<td>8.9</td>
<td>146.0</td>
<td>358.5</td>
<td>0</td>
<td>0.0</td>
<td>16712</td>
<td>29186</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>942</td>
<td>0.2</td>
<td></td>
<td></td>
<td>0</td>
<td>0.0</td>
<td>12172</td>
<td>20401</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

**Primary GBP**
Notes: Key Points – GBPs Structure Activity

- SERV TIME (MIC)
  - SYNC is key metric – ‘good’ number is relative to CF configuration
    - If REQ/SEC < 100, variations in service time probably not significant
    - ASYNC requests are expected, especially in secondary GBPs
  - XI’s in lower right are not necessarily reclaims
    - Most likely business as usual
Monitoring GBPds: -DIS GBPOOL

- DIS GBPOOL(*) TYPE(GCONN) GDETAIL(*)
  - Contains status and definition information as well as statistics
  - Reports statistics since GBP allocation
- DIS GBPOOL(*) TYPE(GCONN) GDETAIL( INTERVAL )
  - To monitor an interval, execute this command before and after the desired interval.
  - Output messages from second command will show GBP statistics for the interval
- Typical problems due to incorrectly defined GBP
  - Directory entry reclaim
  - XIs due to directory entry reclaim
  - Writes failed due to lack of storage
-DIS GBPOOL(*) TYPE(GCONN) GDETAIL(*)

-DIS GBPOOL(*) TYPE(GCONN) GDETAIL(*)

07.57.32 STC34822 DSNB784I -DB2A GROUP DETAIL STATISTICS 362
362 READS
362 DATA RETURNED = 1842830

07.57.32 STC34822 DSNB785I -DB2A DATA NOT RETURNED 363
363 DIRECTORY ENTRY EXISTED = 1490516
363 DIRECTORY ENTRY CREATED = 9995482
363 DIRECTORY ENTRY NOT CREATED = 26712646, 0

07.57.32 STC34822 DSNB786I -DB2A WRITES 364
364 CHANGED PAGES = 50473770
364 CLEAN PAGES = 3408467
364 FAILED DUE TO LACK OF STORAGE = 48
364 CHANGED PAGES SNAPSHOT VALUE = 5568

07.57.32 STC34822 DSNB787I -DB2A RECLAIMS 365
365 FOR DIRECTORY ENTRIES = 80726
365 FOR DATA ENTRIES = 28878053
365 CASTOUTS = 28679918

07.57.32 STC34822 DSNB788I -DB2A CROSS INVALIDATIONS 366
366 DUE TO DIRECTORY RECLAIMS = 56680
366 DUE TO WRITES = 2666240
366 EXPLICIT = 0

07.57.32 STC34822 DSNB762I -DB2A DUPLEXING STATISTICS FOR GBP11-SEC 367
367 WRITES
367 CHANGED PAGES = 50072797
367 FAILED DUE TO LACK OF STORAGE = 48
367 CHANGED PAGES SNAPSHOT VALUE = 5568

27
Notes: Sizing CF Structures

  - CF Structure Sizer Tool
- *DB2 Version 9.1 for z/OS Installation Guide*, GC18-9846
- *DB2 10 for z/OS Installation and Migration Guide*, GC19-2974
- *DB2 11 for z/OS Installation and Migration Guide*, GC19-4056
  - Knowledge Center: [cf sizing for DB2 10 or 11](http://www.ibm.com/systems/support/z/cfsizer)
- Rule of thumb for GBPs
  - Start with CFSizer INITSIZE
  - Round up
  - Make that result INITSIZE; make SIZE up to twice that value
  - Use Auto Alter
Auto Alter – What is it?

- Autonomic effort by XES to avoid filling up any kind of structure. For GBPs:
  - If all data elements (pages) are changed, writes cannot occur
  - If all directory entries are marked changed, new pages cannot be registered
- Auto Alter has algorithms that
  - can increase or decrease number of entries and/or elements to avoid structure full conditions
  - can increase or decrease the size of the structure
- Can alter, dynamically, the precise directory to data ratio for GBPs
- Design point is for gradual growth, not spikes
Auto Alter and DB2

• DB2 Structures support Auto Alter
• LOCK1 – effective on Modify Lock List entries (RLEs)
  – Lock Table entries (LTE) cannot be changed without a rebuild
• SCA – can be increased
• Main value is for Group Buffer Pools (GBPs). Why?
  – People tend not to tune GBPs
    • Organizational division of labor
      – DB2 DBAs responsible for local BPs – may forget about GBPs
      – z/OS responsible for GBPs – and they own the CFRM Policy
  – DB2 needs ?? more directory entries than data page elements
  – Each –ALTER to change directory entries means manual GBP rebuild
• Works for duplexed GBPs
Auto Alter – When not to use it

- CF available storage is <10%
  - Auto Alter reduces the size of “alterable” structures below INITSIZE (to MINSIZE), attempting to get 10% available storage in the CF
- Not enough storage for size of structure, especially in Test environments
  - XES reaches SIZE quickly
  - Reclaim avoidance results in constant XES attempts to increase directory entries and reduce data pages
    - Reclaim avoidance alone does not allow structure size increase
    - Attempts usually fruitless - produce alarming console messages
  - Hint: test one structure, correctly sized, instead of all
Workload Growth

- Increased transaction, batch and/or query volumes
- New applications
- Mergers
- New business opportunities
- Regulatory compliance
- Technology advances
Workload Changes and LOCK1

• Increased lock requests may lead to
  – Higher CF CPU busy
  – Higher synchronous service time, and host CPU cost
  – Higher transaction or query elapsed time, higher job run time

• New applications may not follow standards
  – Less lock avoidance by new applications may mean more locking for existing applications
  – Long commit scopes hold Modify Lock List entries (RLEs) longer
  – Row level locking increases demand for RLEs

• False contention could increase, requiring more Lock Table Entries (LTEs)
Workload Changes and LOCK1

• Possible solutions
  – Increase CF CPU capacity
    • More CPs and/or faster CPs
  – Increase the number of RLEs
    • SETXCF START,ALTER,strnm=&,SIZE=&
      to increase the size of LOCK1
      – Assumes allocation < SIZE in CFRM policy
      – Else change CFRM policy definition, rebuild structure
  – Increase the number of LTEs
    • Requires a structure rebuild with larger allocation
      – CFRM policy change required if allocation already = SIZE
    – CF storage increase may be necessary
### LOCK1 Example

- **RMF CF Activity Report**
  - **Structure Activity**

```plaintext
<table>
<thead>
<tr>
<th>SYSTEM</th>
<th>NAME</th>
<th>REQ</th>
<th>AVG/SEC</th>
<th>ALL</th>
<th>AVG</th>
<th>STD_DEV</th>
<th>REQUESTS</th>
<th>DELAYED REQUESTS</th>
<th>EXTERNAL REQUEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>S***</td>
<td>233M</td>
<td>323M</td>
<td>323.8</td>
<td>11.0</td>
<td>5.8</td>
<td>NO SCR 89K</td>
<td>0.0 30.8 236.2</td>
<td>0.0</td>
<td>REQ TOTAL 133K</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td></td>
<td>-CONT 233K</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td></td>
<td>-FALSE CONT 122K</td>
</tr>
<tr>
<td>S***</td>
<td>17W</td>
<td>17W</td>
<td>31.2</td>
<td>5.5</td>
<td>NO SCR 14K</td>
<td>0.0 15.0 47.2</td>
<td>0.0</td>
<td>REQ TOTAL 99K</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td></td>
<td>-CONT 218K</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td></td>
<td>-FALSE CONT 111K</td>
</tr>
<tr>
<td>S***</td>
<td>17W</td>
<td>17W</td>
<td>30.0</td>
<td>5.9</td>
<td>NO SCR 50K</td>
<td>0.0 19.7 37.5</td>
<td>0.0</td>
<td>REQ TOTAL 102K</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td></td>
<td>-CONT 333K</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td></td>
<td>-FALSE CONT 242K</td>
</tr>
<tr>
<td>S***</td>
<td>28H</td>
<td>28H</td>
<td>10.2</td>
<td>5.1</td>
<td>NO SCR 11K</td>
<td>0.0 29.2 241.5</td>
<td>0.0</td>
<td>REQ TOTAL 27H</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td></td>
<td>-CONT 79K</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td></td>
<td>-FALSE CONT 512K</td>
</tr>
</tbody>
</table>
```

**Note:** The diagram shows a visual representation of the structure activity, with colors indicating different statuses and request types. The numbers represent specific metrics such as request counts, average times, and status flags.
Workload Changes and GBPs

- Increased GBP requests may lead to
  - Higher CF CPU busy
  - Higher synchronous service time, and host CPU cost
  - Higher transaction or query elapsed time, higher job run time

- New applications may
  - Change access patterns of existing tables or indexes
  - Add tables and indexes to existing buffer pools

- Local buffer pool allocations may increase
  - GBPs might be forgotten
Workload Changes and GBPs

• Possible solutions
  – Increase CF CPU capacity
    • More CPs and/or faster CPs
  – Increase the size of the GBPs
  – Tune local buffer pool thresholds and GBP thresholds
  – CF storage increase may be necessary
GBPs and Impact of CF Busy

- ICF has two CPs
- CF (external) has three CPs
• GBP1 on ICF was very busy over a 10 hour interval
  – S1: 178 M synchronous requests
  – S3: 144.5 M synchronous requests
  – If 10 µsec saved from each request, over 300 CPU seconds per hour of ‘host effect’ could be saved from GBP1 alone

• How could 10 µsec be saved?
  – Increase number of CPs on ICF to reduce CF busy and improve service time
  – Upgrade CEC with ICF to reduce CF service times
Workload Changes and SCA

- New applications or new workloads may add tables and indexes
- New clients may require additional databases
- Auto Alter may be able to handle most of the increase
- Use CF Structure Sizer Tool to validate CFRM policy definition
When New Members Join the Data Sharing Group

- GBP – review sizes
  - Increased demand for directory entries and data elements
  - Auto Alter may not be sufficient to handle multiple new members
- LOCK1
  - 4-byte LTEs required when 8th member joins the group
    - Automatic rebuild will normally result in half as many LTEs, so false contention will increase
    - Prepare for larger LTEs before adding 8th member
  - 8-byte LTEs required when 23rd member joins the group
    - Automatic rebuild has same considerations
Configuration Changes

• CF Considerations
  – Balanced performance: CF technology = CEC technology
  – Unbalanced configuration examples:
    • zEC12 CF and z196 CEC – good for the CEC
    • z196 CF and zEC12 CEC – more Host Effect cost to CEC
    • z10 CF and zEC12 CEC – ‘heuristic algorithm’ likely to convert many synchronous requests to asynchronous
      – Algorithm represents tradeoff of host effect versus cost of conversion
      – Elapsed times, contention, and time outs likely to increase
  – Increase in distance between CF and CEC can have similar effect
    • Asynchronous conversion frequently observed as distance between CEC and CF increases
Creative Use of CF Storage

- As more DB2 members join the group
  - Consider GBPCACHE ALL
    - Each page is read into GBP on first access
    - Only one member incurs I/O cost for each page
    - Local buffers can be smaller – GBP acts as very fast cache
- If large objects with very random access and minimal page re-reference
  - Consider GBPCACHE NONE
    - Saves GBP access on local page miss
    - Enforces 'store through cache': synchronous writes to disk at commit
      - Modern cache controllers minimize negative impact
DB2 10 for z/OS
Data Sharing Highlights
DB2 10 for z/OS and Data Sharing

- Deleting member of data sharing group
  - Offline utility
- Deleting structures during group restart
  - DEL_CFSTRUCTS_ON_RESTART - DSNZPARM for DR
- Sub-group attach
- DDF Restart Light – handle indoubts
- MEMBER CLUSTER for UTS
- -MODIFY DDF – online changes for LOCATION ALIAS

- LRSN spin avoidance
- IFCID 359 – index split
- GBP DELETE_NAME processing
- BP scan avoidance
DB2 11 for z/OS
Data Sharing Highlights
DB2 11 for z/OS Data Sharing Enhancements

- Castout enhancements: New CLASST setting – similar to VDWQT
- RESTART LIGHT Enhancements
- Buffer pool enhancements

- GBP Write-around
  - If GBP / CF busy, write new pages to directly to disk
  - Reduce impact of flood of new pages on rest of GBP
- Automatic LPL or GRECP recovery
- CF DELETE_NAME
- Locking enhancements
- Index split performance
- LRSN spin avoidance – extended LRSN
Additional Resources

- Data Sharing: Planning and Administration
  - DB2 9 for z/OS: SC18-9845
  - DB2 10 for z/OS: SC19-2973
  - DB2 11 for z/OS: SC19-4055
  - KC db2 data sharing planning
Questions?
Thank you!