

The Best Performance Practices DB2 for z/OS

*Akiko Hoshikawa, DB2 for z/OS Performance
Silicon Valley Lab*

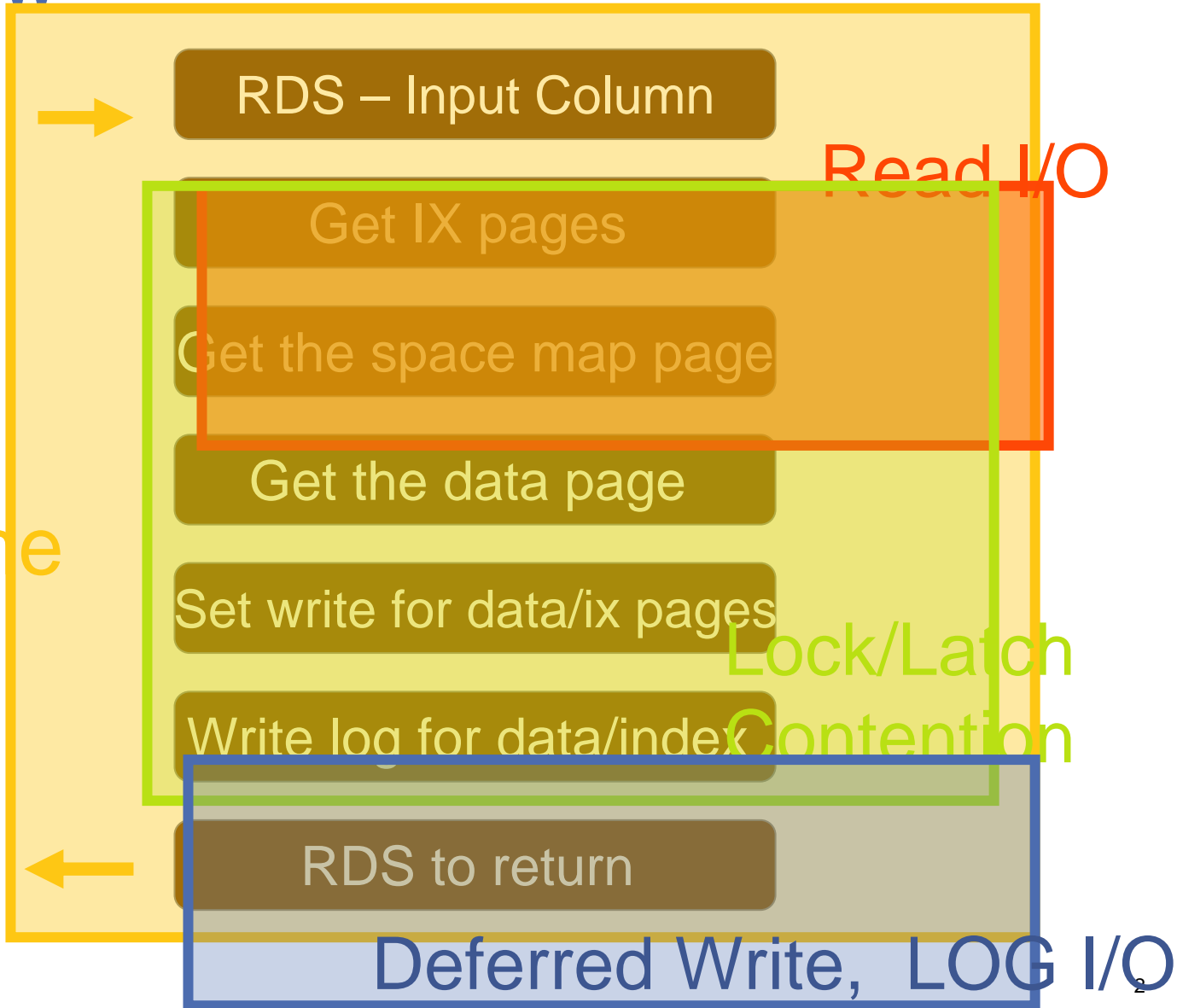
Agenda

- High Concurrent Insert
 - Bottlenecks and Tuning considerations
- General Performance Improvement in V9
 - Instrumentations
 - SORT
 - SQL Procedure
- Latest on Specialty Engines
- z10 Performance



Insert Flow

INSERT



CPU time

Read I/O

Lock/Latch Contention

Deferred Write, LOG I/O



High Concurrent Insert

| Class1 | Avg Time (ms) | |
|-------------|---------------|-----------|
| Elapsed | 1259 | |
| CPU (CP) | 10 | |
| CPU (zIIP) | 10 | |
| Class3 | Avg.Time (ms) | Avg.Event |
| Lock/Latch | 796 | 39 |
| DB I/O | 233 | 47 |
| Log I/O | 60 | 4.5 |
| Other write | 21 | 2.3 |
| Page Latch | 43 | 6.1 |
| Async CF | 3 | 11.3 |
| Locking | | Avg. |
| Lock sus | | 0.03 |
| IRLM sus | | 0.28 |

- Lock/latch wait
 - DB2 Lock
 - IRLM Latch
 - DB2 Latch
- Page latch
- Data base and log I/Os
- CPU time



DB2 Latch Counter

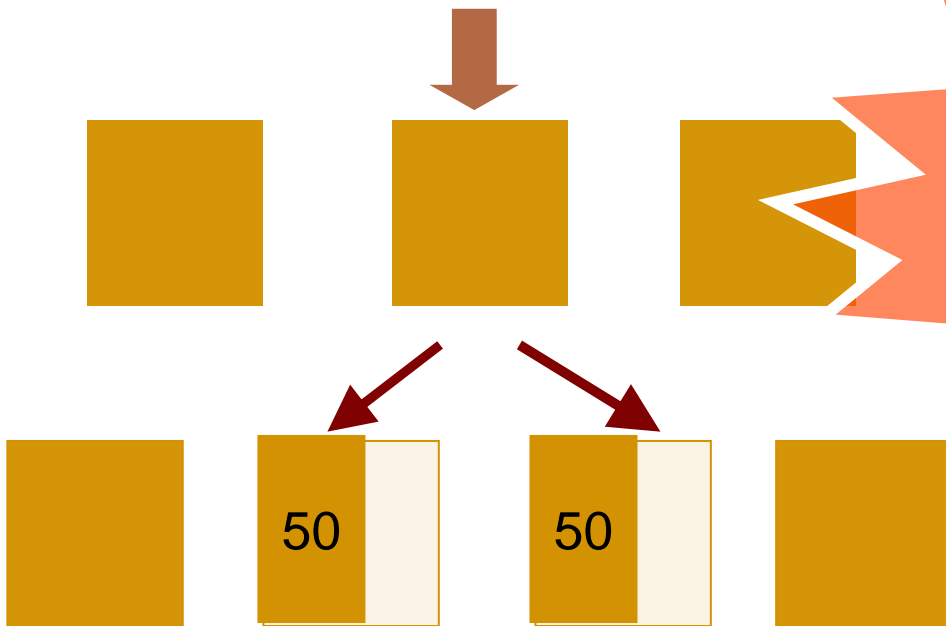
| LATCH CNT | /SECOND | /SECOND | /SECOND | /SECOND |
|-----------|---------|---------|----------|---------|
| ----- | ----- | ----- | ----- | ----- |
| LC01-LC04 | 0.00 | 0.00 | 8.05 | 0.00 |
| LC05-LC08 | 0.59 | 759.62 | 0.06 | 0.00 |
| LC09-LC12 | 0.00 | 0.00 | 0.00 | 27.07 |
| LC13-LC16 | 0.04 | 332.98 | 0.00 | 6.00 |
| LC17-LC20 | 0.00 | 0.00 | 20048.36 | 0.00 |
| LC21-LC24 | 0.03 | 0.00 | 1980.15 | 304.57 |
| LC25-LC28 | 30.59 | 0.09 | 55.32 | 2.27 |
| LC29-LC32 | 0.10 | 8.68 | 38.84 | 108.99 |

- Rule of Thumb: Try to keep below 10,000 per second
 - LC6 - Index Tree Latch (X'46' or 'FE' in IFCID 57)
 - LC14 - Buffer Manager
 - LC19 - Log Latch (X'13' in IFCID 57)
 - LC23 - Buffer Manager (BM timer : Indication of page latch)
 - LC24 - Prefetch Latch or EDM chain latch



Index Split and Latch

INSERT



Index Tree
LATCH

Log Force Write
in Data Sharing

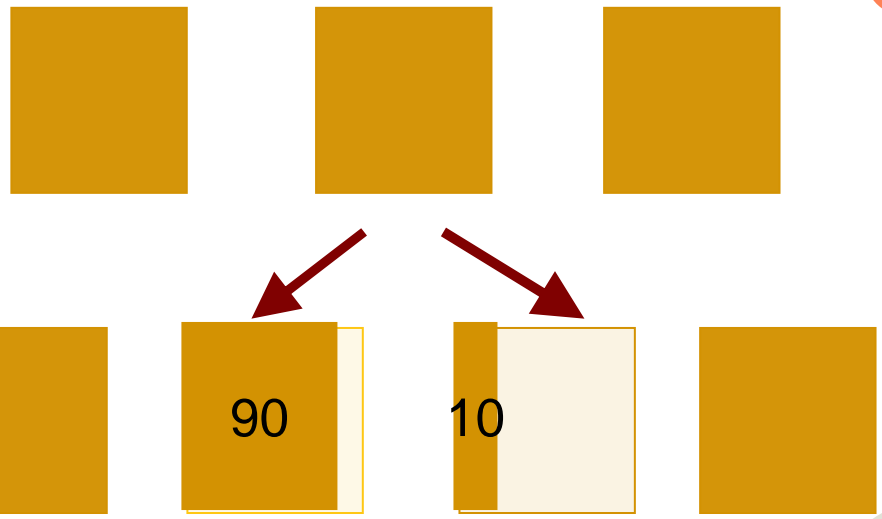
Freespace Tuning
Default PCTFREE 10%

Reduce Index Keysize
by NOT PADDED (V8)



Asymmetric Index Split in V9

INSERT



Monitor Insert Pattern
Seq or Random?

Larger Index Page
(V9 NFM)

Asymmetric Index Split (V9 NFM)



Log Latch

- Hold Log latch (LC19) to serialize log operation
 - Hold latch while spin (data sharing V8, V9CM)
 - No longer holds log latch while spinning in data sharing (V9 NFM)
- LRSN spin and Log latch contention
 - In Data Sharing, DB2 may need to spin to generate unique LRSN
 - Impact in heavy insert operation with faster processor
- New no log option (V9 NFM)
 - Use with caution



DB2 Lock Consideration –(1)

- Take an advantage of lock avoidance
 - Use Cursor Stability and Currentdata No
 - Avoid locking in singleton select with CS CD YES (V8)
 - Currentdata NO Default in V9 (from YES)
 - LOB lock avoidance (V9 NFM)
 - Inserts in non data sharing generally do not wait for locks as conditional locks used
- Reorg to reduce locks on
 - Variable/Compressed Overflow record/pointer lock
 - Avoidance for overflow record in V8
 - Unique index insert with deleted record



DB2 Lock Consideration –(2)

- Lock waits by other concurrently running threads
 - V8 SKIPUNCI=Yes/No in DSN6SPRM macro to skip uncommitted inserts for CS and RS isolation with row level locking
- Skip locked data option in SELECT in V9
 - skips locked row or page, while Uncommitted Read isolation mode does not
 - For CS or RS, not UR or RR
- No more simple table space in V9
 - Newly defined table space becomes segmented



High Concurrent Insert

| Class1 | Avg Time (ms) | |
|-------------|---------------|-----------|
| Elapsed | 1259 | |
| CPU (CP) | 10 | |
| CPU (zIIP) | 10 | |
| Class3 | Avg.Time (ms) | Avg.Event |
| Lock/Latch | 796 | 39 |
| DB I/O | 233 | 47 |
| Log I/O | 60 | 4.5 |
| Other write | 21 | 2.3 |
| Page Latch | 43 | 6.1 |
| Async CF | 3 | 11.3 |
| Locking | | Avg. |
| Lock sus | | 0.03 |
| IRLM sus | | 0.28 |

- Lock/latch wait
 - DB2 Lock
 - DB2 Latch
 - IRLM Latch
- Page latch
- Data base and log I/Os
- CPU time



Page Latch

- Index Leaf page
 - With sequential key insert
 - Random option (V9 NFM)
- Space map and data page
 - Current last space map
 - Member Cluster option
 - MAXROWS
- Trim down the latch holding time
 - Deferred Write I/O tuning



High Concurrent Insert

| Class1 | Avg Time (ms) | |
|-------------|---------------|-----------|
| Elapsed | 1259 | |
| CPU (CP) | 10 | |
| CPU (zIIP) | 10 | |
| Class3 | Avg.Time (ms) | Avg.Event |
| Lock/Latch | 796 | 39 |
| DB I/O | 233 | 47 |
| Log I/O | 60 | 4.5 |
| Other write | 21 | 2.3 |
| Page Latch | 43 | 6.1 |
| Async CF | 3 | 11.3 |
| Locking | | Avg. |
| Lock sus | | 0.03 |
| IRLM sus | | 0.28 |

- Lock/latch wait
 - DB2 Lock
 - DB2 Latch
 - IRLM Latch
- Page latch
- Data base and log I/Os
- CPU time

I/O Wait Reduction

- Index read I/Os
 - Utilize large buffer pool to cache index if possible (V8)
- Larger VSAM CI size for larger data insert (V8)
- Larger Index page size (V9 NFM)
- Bigger preformat quantity and trigger ahead (V9 CM)
 - If >16CYL alloc, 16CYL per preformat (x'09 lock)
 - Less likely see the preformat wait
- Doubled prefetch and deferred write quantity (V9 CM)
- Active log stripe
- Archive log stripe (V9 NFM)



High Concurrent Insert

| Class1 | Avg Time (ms) | |
|-------------|---------------|-----------|
| Elapsed | 1259 | |
| CPU (CP) | 10 | |
| CPU (zIIP) | 10 | |
| Class3 | Avg.Time (ms) | Avg.Event |
| Lock/Latch | 796 | 39 |
| DB I/O | 233 | 47 |
| Log I/O | 60 | 4.5 |
| Other write | 21 | 2.3 |
| Page Latch | 43 | 6.1 |
| Async CF | 3 | 11.3 |
| Locking | | Avg. |
| Lock sus | | 0.03 |
| IRLM sus | | 0.28 |

- Lock/latch wait
 - DB2 Lock
 - DB2 Latch
 - IRLM Latch
- Page latch
- Data base and log I/Os
- CPU time

CPU Reduction in INSERT/DELETE

- V8
 - Multi row insert (NFM)
 - Long Term Page Fix (PGFIX=YES) with large number of BP or GBP read/write
 - Append-like (CM)
 - Member Cluster, PCTFREE/FREEPAGE 0
- V9
 - Shared memory between DDF and DBM1 (CM)
 - **More Index look aside (CM)**
 - Append YES (NFM)

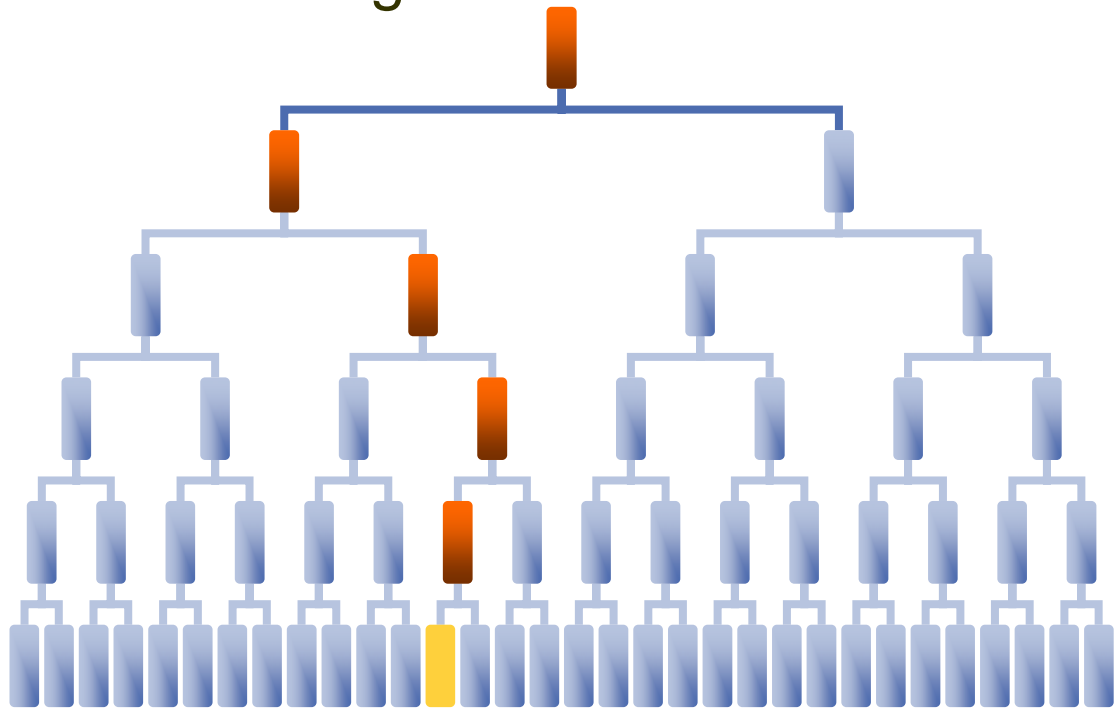


CPU: Index Lookaside

- V8
 - Insert : Look aside for Clustering Index
 - Delete : None

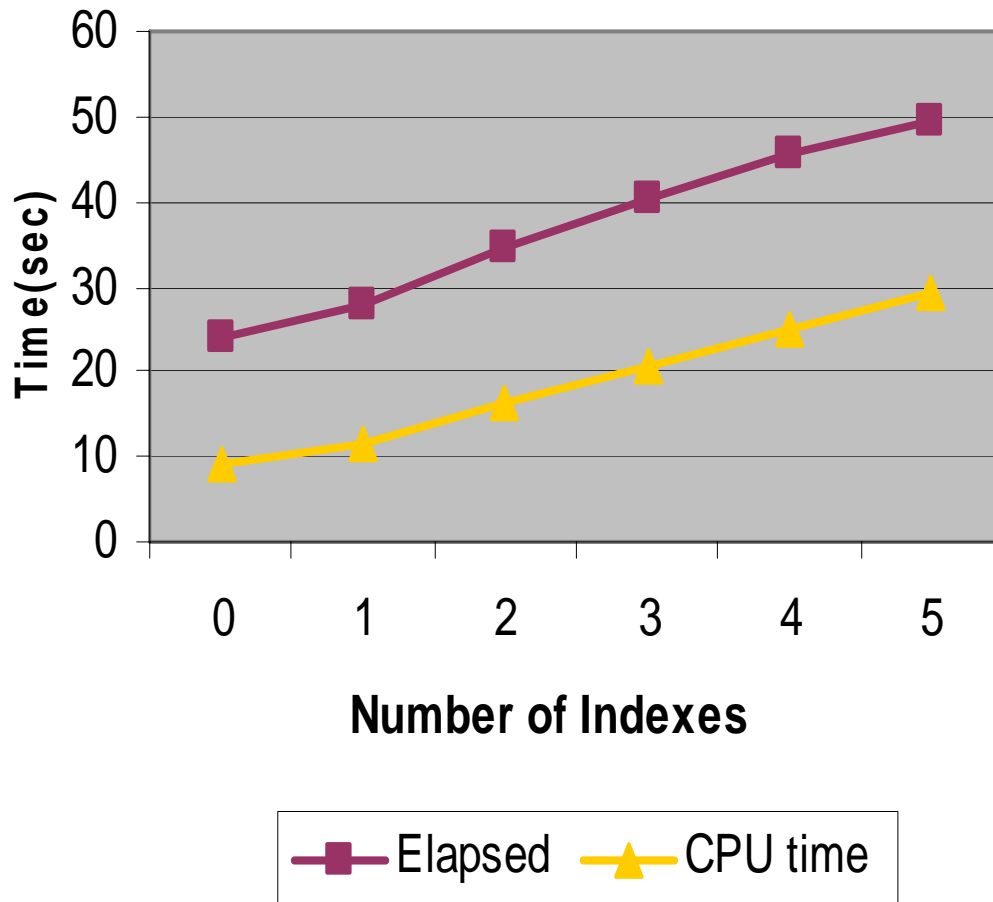
- V9 More Index Look aside

- Insert/Delete
- 3 index with all ascending key seq, 6 nlevel
 - Avg. 12 getpages per insert in V8
 - Avg. 2 getpages per insert in V9



CPU time (and I/O time)

1M Insert with Index



- Remove Indexes not used
- Backward Index scan (V8 CM)
 - No need for both ASC/DSC
- Real Time Stats
SYSINDEXSPACESTATS.LASTUSED (V9 NFM)
 - SELECT/FETCH, searched UPDATE/DELETE
 - Not INSERT, LOAD, etc.
 - Support RI, RID list, set functions, and XML index (PK44579 8/07)



Summary of Concurrent Insert

- Common issues in insert
 - Lock/Latch Wait
 - I/O wait
 - CPU
- Significant improvement in concurrent Insert/Update/Delete in DB2 9
 - Asymmetric index split, Large index pages
 - LOG latch contention reduction
 - More Index Lookaside
 - Larger preformat and deferred write quantity
 - Shared memory between DDF and DBM1



Performance Topics

Instrumentation updates

Open Datasets



Instrumentation : CPU Time Reduction

- Minimize phantom or orphaned trace records (V9 CM)
 - Example from customer's DB2 V8 statistics report in IFC records per commit
 - (1) **Phantom or orphaned trace** because monitoring (eg vendor tool) stopped but not DB2 trace. The same CPU overhead as real trace.
 - V9 tries to eliminate orphaned trace records

| IFC DEST | Written | Others (1) |
|----------|---------|------------|
| SMF | 2 | 0 |
| OP5 | 0 | 4 |
| OP6 | 0 | 4 |
| OP7 | 0 | 4 |
| OP8 | 2 | 0 |
| Others | 0 | 0 |



Instrumentation – Display Traces

-D91D DIS TRACE

| TNO | TYPE | CLASS | DEST | QUAL | IFCID |
|-----|-------|--------------|------|------|-------|
| 01 | STAT | 01,03,04,05, | SMF | NO | |
| 01 | | 06 | | | |
| 02 | ACCTG | 01 | OP1 | NO | |
| 03 | AUDIT | 01 | OP2 | NO | |
| 04 | MON | 30 | OP2 | NO | 031 |
| 05 | STAT | 04 | OP2 | NO | |
| 06 | STAT | 03 | OP2 | NO | |
| 07 | PERFM | 30 | OP2 | NO | 090 |

*****END OF DISPLAY TRACE SUMMARY DATA*****

- Know your trace by Display Trace
- Regular Accounting Class 1,2,3,7,8
 - Detail accounting Class 10 (IFCID 239)
- Dataset Statistics (Stats class 8) for I/O tuning



Instrumentation – Statistics Traces

- Set to 1 minute interval
 - Only 1440 intervals per day
 - Essential for slowdown, virtual storage monitoring
- Copy SMF 100/102 (statistics) to keep separate from accounting (101)



Instrumentation

- Dataset Statistics for Data base I/O tuning
 - Statistics class 8 (IFCID 199)

| BPOOL | DATABASE SPACENAM PART | TYPE GBP | SYNCH I/O AVG ASYNC I/O AVG ASY I/O PGS AVG | SYN I/O AVG SYN I/O MAX | AVG DELAY MAX DELAY | CURRENT PAGES (VP) CHANGED PAGES (VP) CURRENT PAGES (HP) NUMBER OF GETPAGES |
|-------|------------------------------|-------------|---|----------------------------|------------------------|--|
| BP10 | KAGURA24 | TSP | 23.35 | | 8 | 3433 |
| | TETHTS | N | 0.01 | | 78 | 0 |
| | 30 | | 32.00 | | | N/A |
| BP11 | KAGURA24 | IDX | 102.59 | | 1 | 18991 |
| | TETHIX1 | N | 4.04 | | 35 | 74 |
| | 36 | | 5.98 | | | N/A |
| | | | | | | 245586 |

Count of Sync I/O per second

Sync I/O (ms)



Instrumentation – Package Traces V8

DML in Package Level

| MARKET#3 | AVERAGE |
|----------|---------|
| ----- | ----- |
| SELECT | 0.00 |
| INSERT | 3.97 |
| UPDATE | 13.97 |
| DELETE | 3.97 |
| DESCRIBE | 0.00 |
| PREPARE | 0.00 |
| OPEN | 13.00 |
| FETCH | 46.97 |
| CLOSE | 13.00 |

Buffer pool in Package Level

| MARKET#3 | AVERAGE |
|---------------------|---------|
| ----- | ----- |
| BPOOL HIT RATIO (%) | 98.83 |
| GETPAGES | 146.89 |
| BUFFER UPDATES | 38.74 |
| SYNCHRONOUS WRITE | 0.00 |
| SYNCHRONOUS READ | 1.70 |
| SEQ. PREFETCH REQS | 0.00 |
| LIST PREFETCH REQS | 0.00 |
| DYN. PREFETCH REQS | 0.13 |
| PAGES READ ASYNCHR. | 0.01 |



Dataset Open/Close

| Open/Close Activity | Total |
|-----------------------------|-------|
| OPEN DATASETS – HWM | 49806 |
| OPEN DATASETS | 49687 |
| DSETS CLOSED-THRESH.REACHED | 2394 |

| Buffer Pool | Total |
|------------------------|-------|
| NUMBER OF DATASET OPEN | 5163 |

- More and more DB2 datasets
 - high DBM1 TCB and high accounting class 3 wait
- DSMAX tuning
 - Study the impact on DBM1 storage below 2GB
 - DGTT Indexes are not governed by DSMAX



CLOSE (YES) and CLOSE (NO)

- CLOSE (YES) and CLOSE (NO)
 - When DSMAX is hit, DB2 will close in LRU, first CLOSE (YES).
- Data Sharing GBP dependent objects
 - PCLOSET and PCLOSEN hit
 1. Pseudo close
 2. Then physical close to remove GBP dependent for CLOSE (YES) object
 3. CLOSE NO datasets remain open
 - Remain as GBP dependents
 4. Physical close for CLOSE(NO) objects once DSMAX is hit and all CLOSE(YES) objects have been closed
 - Use CLOSE(YES) in general except the objects should be remained open.



V9 Performance Topics

Select

Utility

SQL Procedures



Select performance – Sort in V9

- Fetch First N Rows Only (CM)
 - 2x improvement measured
 - Fetch First N Rows Only in Subselect (NFM)
- Group by and Distinct Sort improvement (CM)
- In memory workfile for small sort (CM)
- Use for 32K workfile for larger records sort (CM)
 - **Assign bigger 32K work file BP**
 - Allocate bigger/more 32K work file datasets



Select Performance - continued

- Index on expression (NFM)

```
Create Index IX1 on  
Employee(salary+bonus,bonus*100/salary)
```

– Orders of magnitude improvement if a predicate using such an index

- Extra cost in Load, Insert, Update on key value, Rebuild Index, Check Index, and Reorg tablespace but not Reorg index as expressions are evaluated in Insert or index rebuild
- Not eligible for zIIP offload as index expression evaluation done at load or unload rather than build index phase



DGTT Performance

- Temp database are stored in workfile DB (V9 CM)
 - PK43765 6/07 to reduce LC24 DGTT prefetch latch contention
 - Increased prefetch quantity from 8 to segsize with 16, maximum 64
- 30 to 60% faster for **SELECT COUNT**
 - Bigger prefetch quantity, 32K workfile, dynamic pref
- 5 to 15% faster and less CPU for **INSERT**
 - Bigger preformat quantity and asynchronous preformat in V9 but not V8
- **Mass Delete/Insert Performance Apars**
 - Performance study in progress
 - PK62009 (04/08) to reclaim mass deleted space at Commit
 - PK67301 (08/08) to reduce the log activity at mass insert without commit



Access Path related improvement

- V8
 - Read multiple rows via index backward to avoid sort
 - More Index usage for unlike type or length
 - Materialized Query Table
 - Distribution Statistics
 - Star Join Sparse index and in memory workfile usage
- V9 CM
 - Optimization across, rather than within, query blocks
 - Pair-wise join in star schema queries
 - Optimizer cost model update
 - Access path stability for static SQL (PK52522/52523 12/07)
 - Histogram statistics over a range of column values



Multi Level Security

- Row level security via MLS in V8 NFM
 - RACF for authorization
 - Cost of SECLABEL column
 - Lower (<5%) for online transaction
 - Higher for cpu-bound sequential scan
- Index Only in V9 with column functions on SECLABEL
 - `SELECT SUM(ACCT_BALANCE) FROM TABLE_A WHERE CUSTNO = '78724' GROUP BY ...`



Utility Performance

- V7-> V8 Utility CPU time -5% to +5%
 - DPSI for Parallel Load/Reorg/Rebuild (V8 NFM)
- V9 CPU reduction in Index processing
 - 5 to 20% in Recover index, Rebuild Index, Reorg Tablespace/Partition
 - 5 to 30% in Load
 - 20 to 60% in Check Index
 - 35% in Load Partition
 - 30 to 50% in Runstats Index
 - 40 to 50% in Reorg Index
 - Up to 70% in Load Replace Partition with NPIs and dummy input

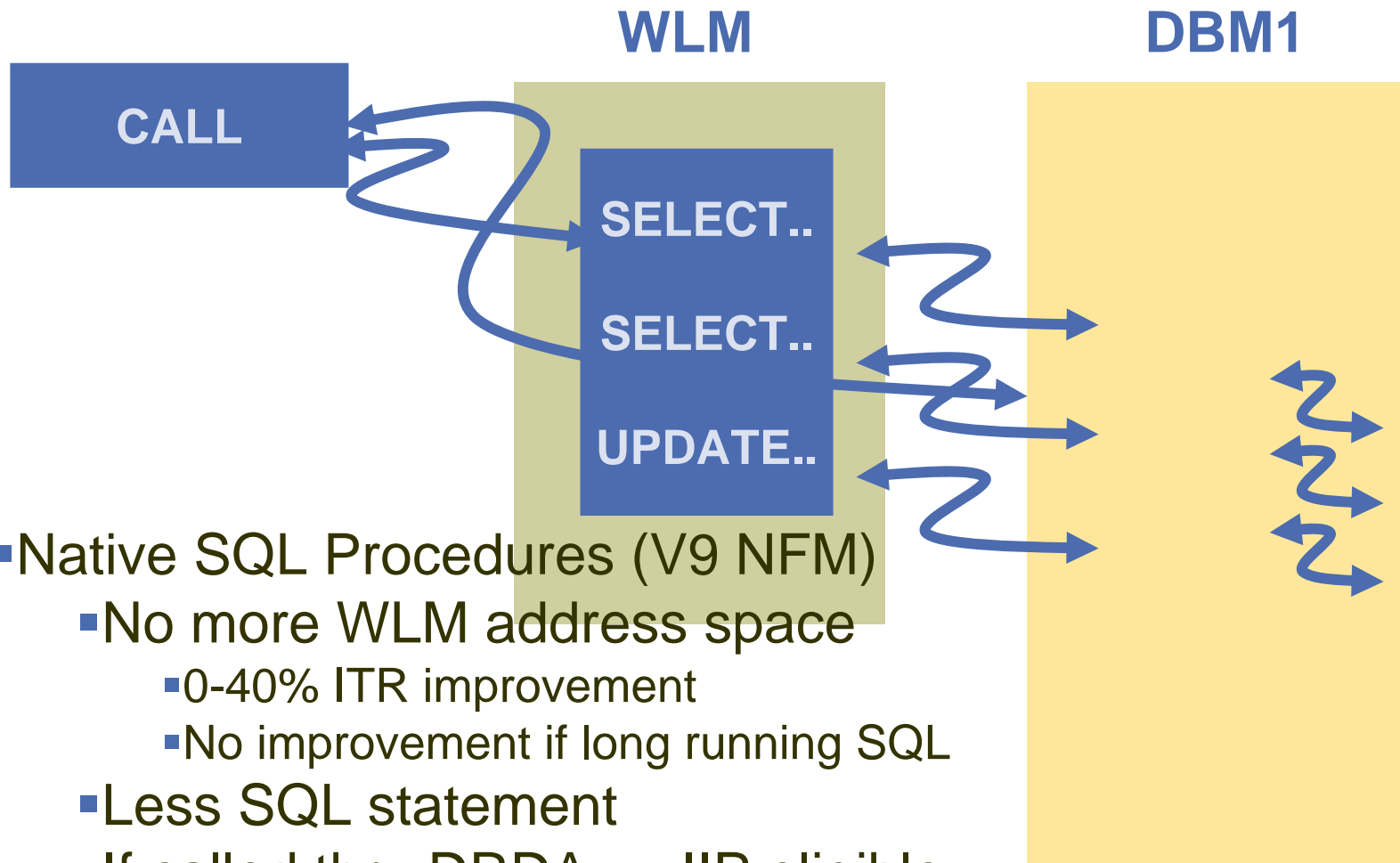


Utility Performance – More

- V8/V9 PK61759 5/08 Load/Reorg CPU reduction
 - 6 to 8% in DFSORT interface
- V9 : Build2 phase elimination in Online Reorg Partition with NPI for better availability :
 - Higher CPU time and elapsed time when few out of many partitions, especially with more NPIs, are Reorg'd as entire NPIs copied to shadow dataset
 - Additional temporary DASD space needed
 - NPIs are automatically Reorg'd
- Active log read buffers per Start IO increased from 15 to 120 for up to +70% recovery throughput



SQL Procedures



- Native SQL Procedures (V9 NFM)
 - No more WLM address space
 - 0-40% ITR improvement
 - No improvement if long running SQL
 - Less SQL statement
 - If called thru DRDA, zIIP eligible





Latest on zIIP/zAAP



zIIP and zAAP

Original zIIP support : V8(CM), z/OS 1.6, z9, SRB mode

- DRDA over TCP/IP including V9 SQLPL
- Index process in LOAD/REORG/REBUILD
- Parallel Query

zAAP:

- XML System Service on z/OS 1.7, z990, DB2 pureXML(V9 NFM)

zIIP:

- XML System Service on z/OS 1.8, z9, DB2 pureXML(V9 NFM)
- z/OS Communication Server
 - IPsec processing (z/OS 1.8, z9)
- z/OS Communication Server
 - HiperSockets for large outbound messages (z/OS 1.10, z10)



zIIP/zAAP Monitoring

- RMF Workload Activity Report
- DB2 Accounting : Combined zIIP/zAAP counter (PK50575 5/08)
- OMPE Accounting: PK51045

| AVERAGE | APPL (CL.1) | DB2 (CL.2) |
|-------------|-------------|------------|
| ----- | ----- | ----- |
| CP CPU TIME | 0.006263 | 0.005252 |
| AGENT | 0.006263 | 0.005252 |
| NONNESTED | 0.001553 | 0.000560 |
| STORED PRC | 0.004660 | 0.004660 |
| UDF | 0.000049 | 0.000032 |
| TRIGGER | 0.000000 | 0.000000 |
| PAR.TASKS | 0.000000 | 0.000000 |
| SECP CPU | 0.001739 | N/A |
| SE CPU TIME | 0.003546 | 0.002318 |
| NONNESTED | 0.001580 | 0.000351 |
| STORED PROC | 0.001967 | 0.001967 |

Eligible for zIIP but did not run on zIIP (same as IICP)

Time spent in Specialty Engines (zIIP + zAAP)



zIIP Tuning Practice

- Avoid zIIP Constraint situation
 - ZIIPAWMT (ZAAPAWMT for zAAP) in IEAOPTxx
 - Default 12000 (12 ms)
 - Impact when the specialty engine will ask for help from the general CPs.
 - If response time is growing and you want the general CPs to help with more work, lower to 1000 or 500 (z10)
 - If response time is fine, use default 12000
 - Add more zIIP processors
- Utility with zIIP : V8/V9 PK60956 (4/08)
 - Objects with small primary/secondary allocation



DB2 and z10



DB2 and z10 - z9 to z10

- Wide variation is observed
 - z9 to z10 Ratio : observed **1.2x to 2.1x** in LSPR
 - depends on the characteristics of workload
 - DB2 Workload 1.4 to 1.6x
 - Higher improvement if CPU intensive workload
 - Lower improvement if memory intensive workload
- DB2 OLTP workload - 1.4x to 1.6 x
 - Dataset Open/Close - 1.2x to 1.4x
 - DSMAX tuning
- DB2 Query - up to 2x
- DB2 Utility - 1.3 to 2.1x



DB2 and z10 –z9 to z10

- DB2 Batch insert/delete/update in non data sharing
 - 1.5x to 2.0x
- DB2 Batch insert/delete/update in data sharing
 - V8, V9CM : DB2 spins for TOD clocks to generate unique LRSN
 - May see little z10 improvement due to LRSN spin
 - Higher impact with MRI
 - V9 NFM : Relief on DB2 spin for LRSN



DB2 and z10 HiperDispatch

- HiperDispatch
 - MIPS improvement with large system
 - z10, z/OS 1.7 or above
- DB2 REORG Partition
 - 13% CPU reduction with HiperDispatch
- DB2 Query CPU Parallelism
 - 1-3% CPU reduction with HiperDispatch



Summary

- V9 has significant improvement in concurrent Insert/Update/Delete
- Use instrumentations wisely
- Other improvements in V9/V8
 - Open datasets
 - Sort
 - SQL Procedure
 - Utility CPU time reduction
- More Specialty Engine support and considerations
- z10 processor improvement varies by workload





Thank You.

email: akiko@us.ibm.com



Disclaimer

© Copyright IBM Corporation [current year]. All rights reserved.
U.S. Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. 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. IN ADDITION, THIS INFORMATION IS BASED ON IBM’S CURRENT PRODUCT PLANS AND STRATEGY, WHICH ARE SUBJECT TO CHANGE BY IBM WITHOUT NOTICE. IBM SHALL NOT BE RESPONSIBLE FOR ANY DAMAGES ARISING OUT OF THE USE OF, OR OTHERWISE RELATED TO, THIS PRESENTATION OR ANY OTHER DOCUMENTATION. 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 ANY AGREEMENT OR LICENSE GOVERNING THE USE OF IBM PRODUCTS AND/OR SOFTWARE.

IBM, the IBM logo, ibm.com and DB2 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

