#SHAREorg

SHARE
Technology · Connections · Results

DB2 10 for z/OS Performance and Scalability

Jeffrey Berger
IBM

August 5-10, 2012
Session 11296

SHARE in Anaheim 2012

#SHAREorg

SHARE
Technology · Connections · Results

Abstract and Agenda

- Abstract: Come learn why everybody is talking about DB2 10 for z/OS. DB2 10 delivers the most aggressive performance improvements of any DB2 release in the last 20 years. Gone are the virtual storage constraints and latching constraints that once limited the number of threads that one DB2 member could run. You will save considerable CPU resources and enjoy numerous I/O improvements with DB2 10. Most of these improvements are enabled even in CM mode, available “out of the box”. Furthermore, DB2 10 increases the synergy with new storage hardware enhancements that will save more system resources and makes it easier to manage your data.

- Agenda:
  - Conversion Mode, with and without Rebind
  - I/O performance, CPU performance, throughput
  - Scalability: virtual storage, latching, UTSERIAL, high concurrent inserts
  - New Function Mode enhancements
  - Beta feedback and summary

SHARE in Anaheim 2012

#SHAREorg

SHARE
Technology · Connections · Results

Disclaimer

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.

This information may contain examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious, and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Trademarks The following terms are trademarks or registered trademarks of other companies and have been used in at least one of the pages of the presentation:

The following terms are trademarks of International Business Machines Corporation in the United States, other countries, or both: AIX, AS/400, DataJoiner, DataPropagator, DB2, DB2 Connect, DB2 Extenders, DB2 OLAP Server, DB2 Universal Database, Distributed Relational Database Architecture, DRDA, eServer, IBM, IMS, iSeries, MVS, Net.Data, OS/390, OS/400, PowerPC, pSeries, RS/6000, SQL/400, SQL/DS, Tivoli, VisualAge, VM/ESA, VSE/ESA, WebSphere, z/OS, zSeries

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.
Intel and Pentium are trademarks of Intel Corporation in the United States, other countries, or both.
UNIX is a registered trademark of The Open Group in the United States and other countries.
Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
Other company, product, or service names may be trademarks or service marks of others.

SHARE in Anaheim 2012

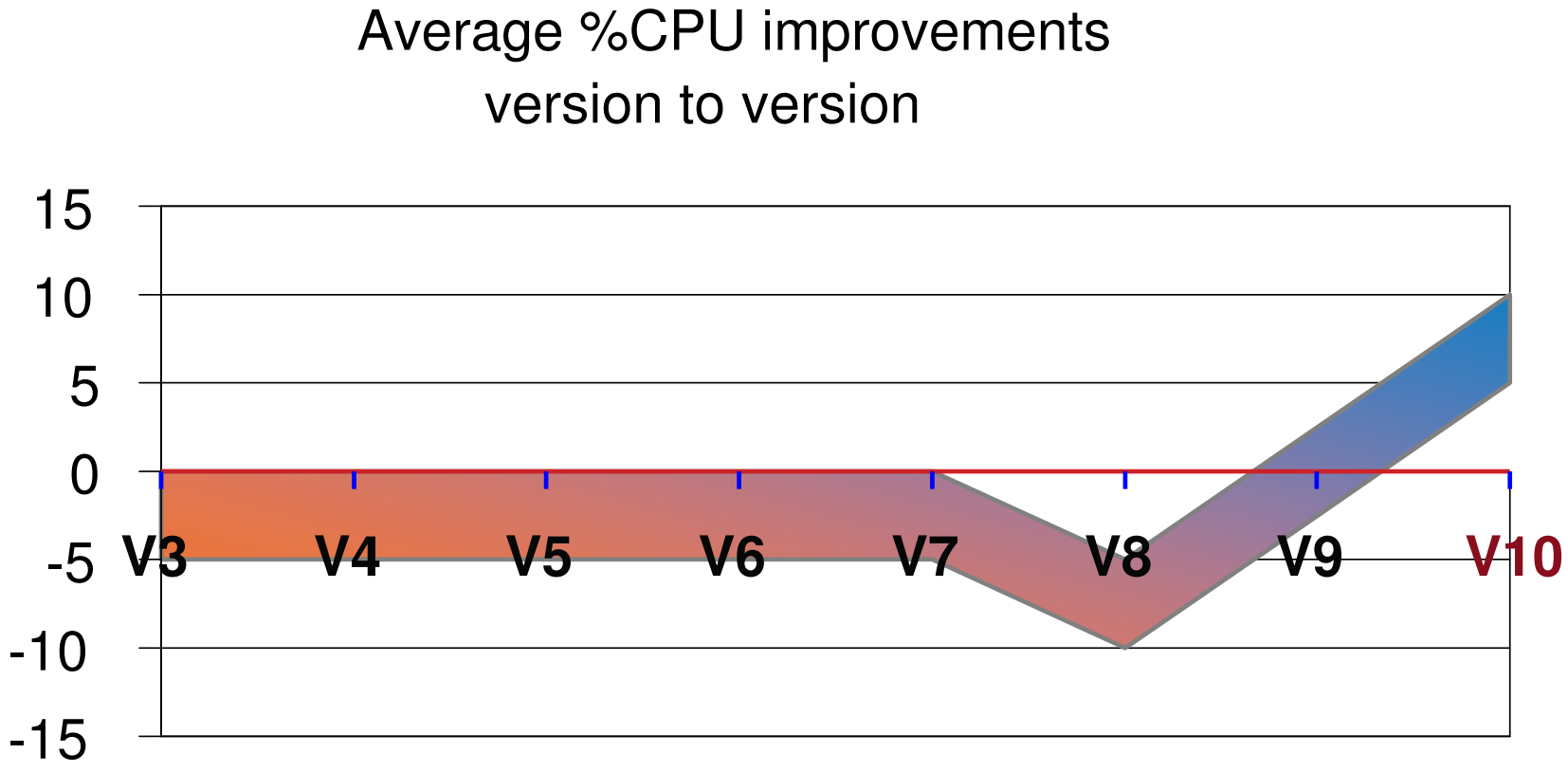
SHARE
Technology · Connections · Results

Performance Disclaimer

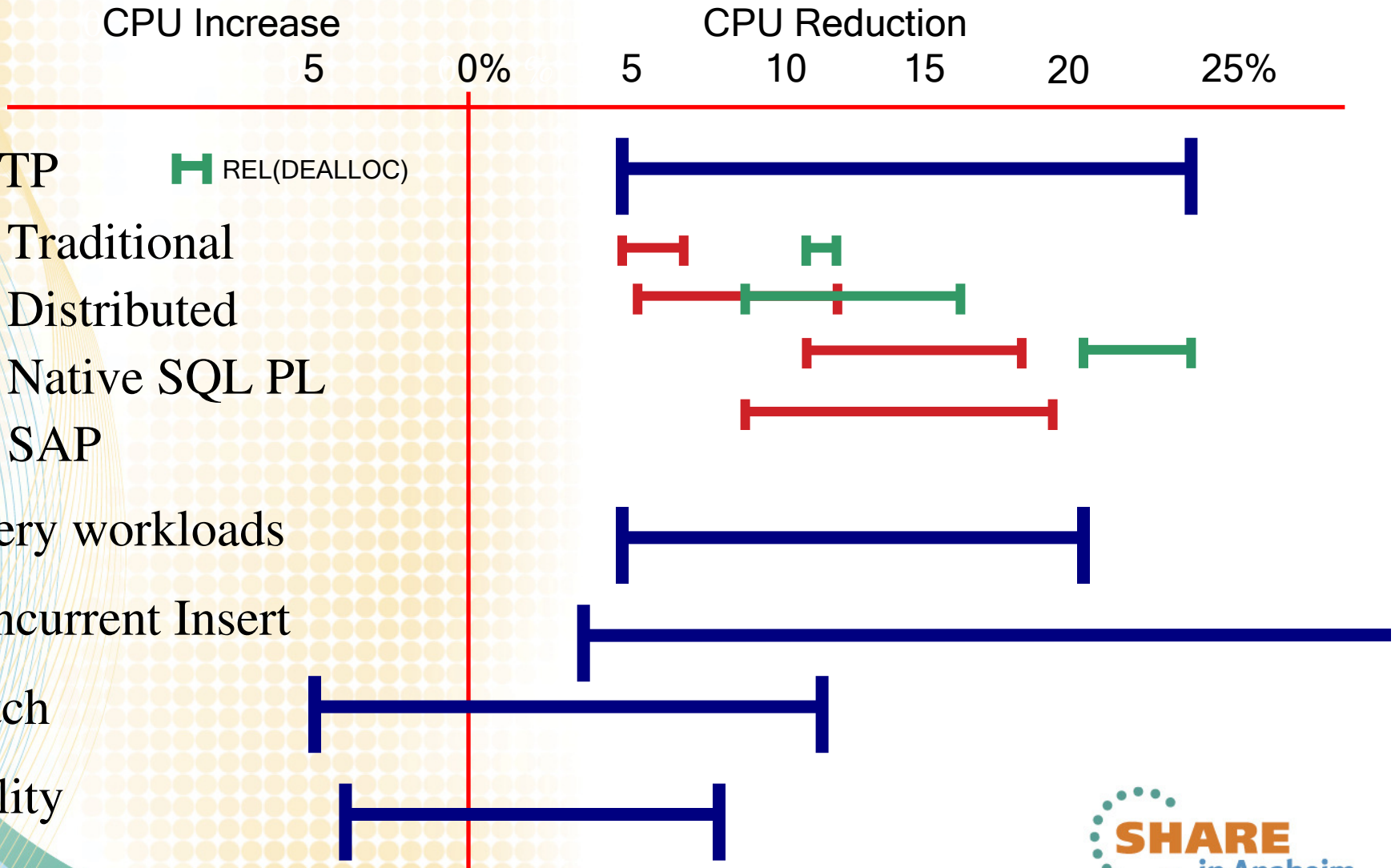
This document contains performance information based on measurements done in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput or performance improvements equivalent to the numbers stated here.

DB2 10 Performance Objective

Historical goal of <5 % version-to-version performance regression
Goal for DB2 : 5% -10% performance improvement for the most of OLTP workloads

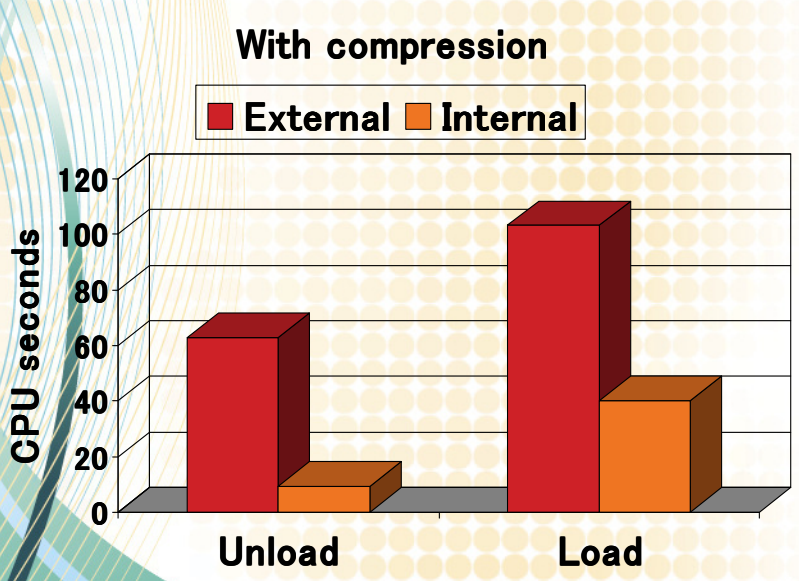


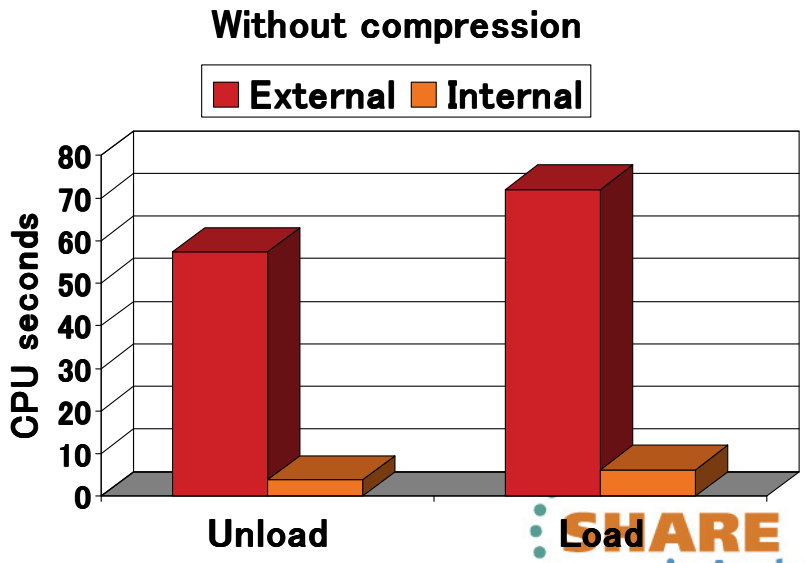
DB2 9 to DB 10 Migration CPU Reduction : Internal Studies



Load/Unload internal format

- Internal format avoids all of the CPU cost of column level processing
  - Unloaded data is always unpadded, no no CCSID conversion, no external format numerics
- Can be used only when data is to be reloaded into a table that has identical DDL as the original table
- PM19584, available for DB2 9 and 10, does not require NFM





Migration Story..

- DB2 V8 NFM or DB2 9 NFM REBIND with PLANMGMT
- Migrate to **DB2 10, CM** mode (CM8 or CM9) without REBIND
  - Ensure there are no issues
  - Enable 1MB page frames for key buffer pools
  - Virtual storage reduction for dynamic packages
- **REBIND** step by step under DB2 10 CM
  - Virtual storage reduction for static packages
  - CPU reduction from avoiding “conversion” of packages
  - If coming from DB2 9, take an advantage of APREUSE/APCOMPARE
- Rebind selective applications with **Release Deallocate**
  - Key online transactions frequently executed, DDF applications
- Enable **DB2 10 NFM**
  - Enable DB2 10 performance features

#SHAREorg

SHARE
Technology · Connections · Results

DB2 10
Conversion Mode

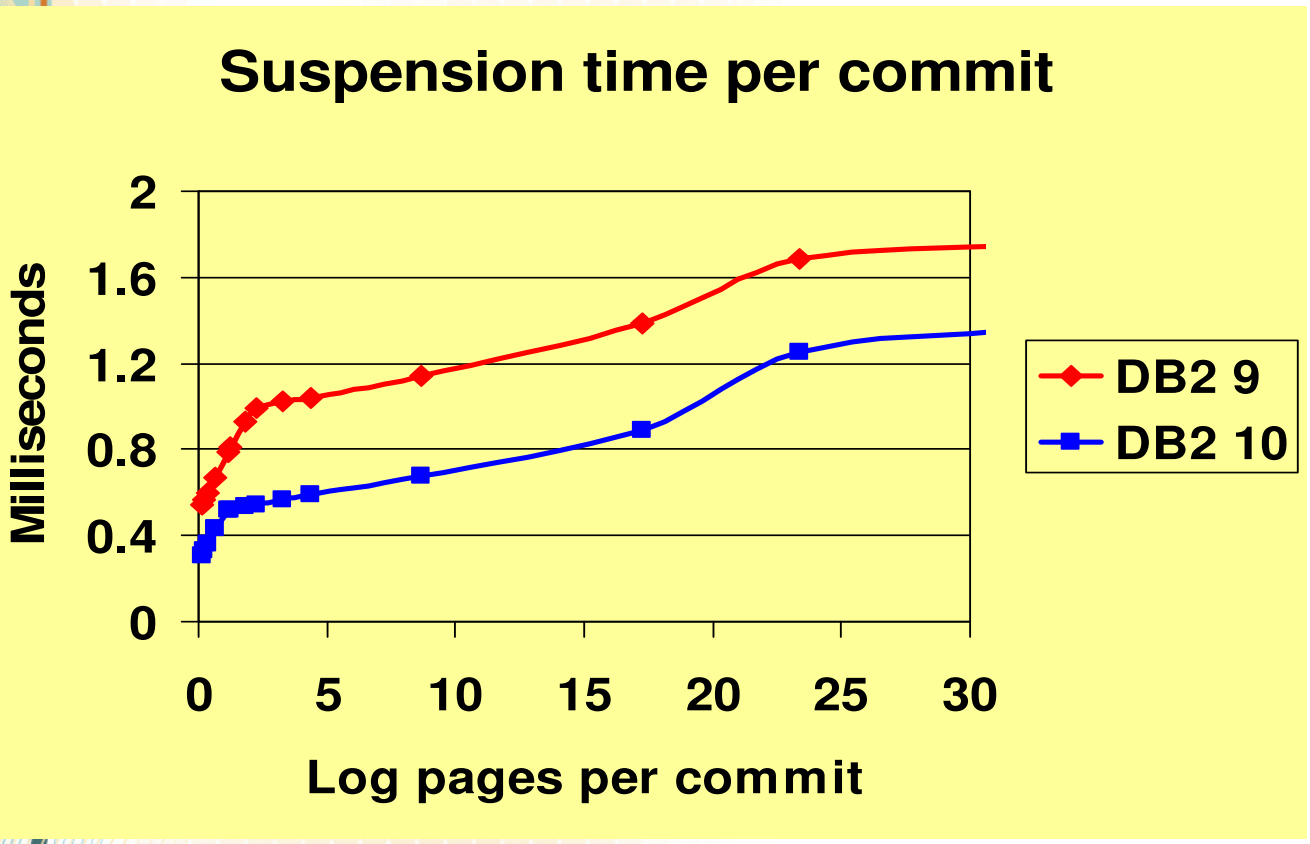
SHARE in Anaheim 2012

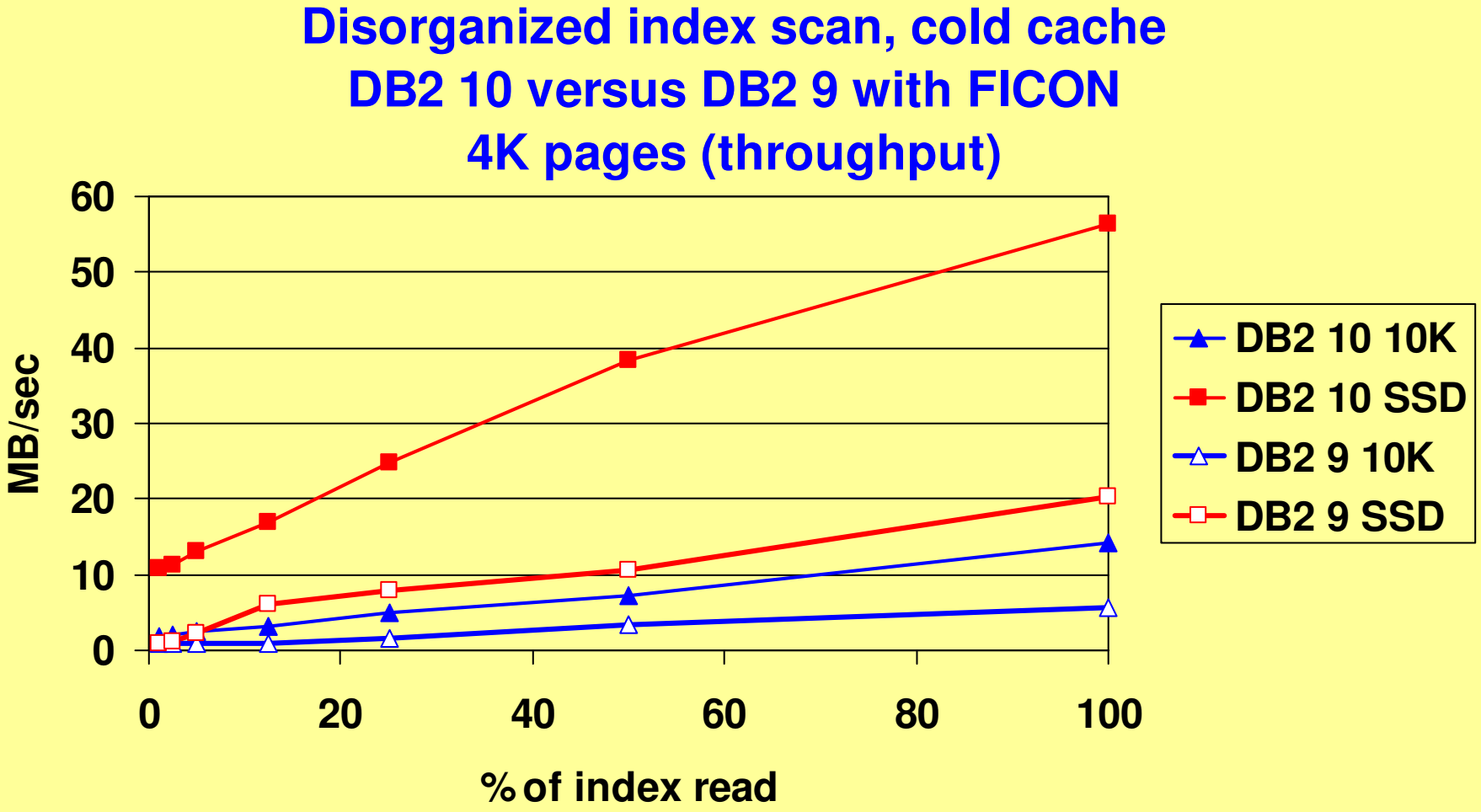
SHARE
Technology · Connections · Results

I/O Performance Highlights

- Log I/O improvement yields up to 2x improvement in commit time
- List prefetch exploitation for index scans of a disorganized index can produce 5x or more improvement
- RID pool to overflow to work files, enables list prefetch to be more viable
- Row level sequential detection: An algorithm to avoid synch I/Os for clustered pages when using index-to-data access
- Insert Index I/O parallelism yields up to 2x improvement for Inserts when using many indexes

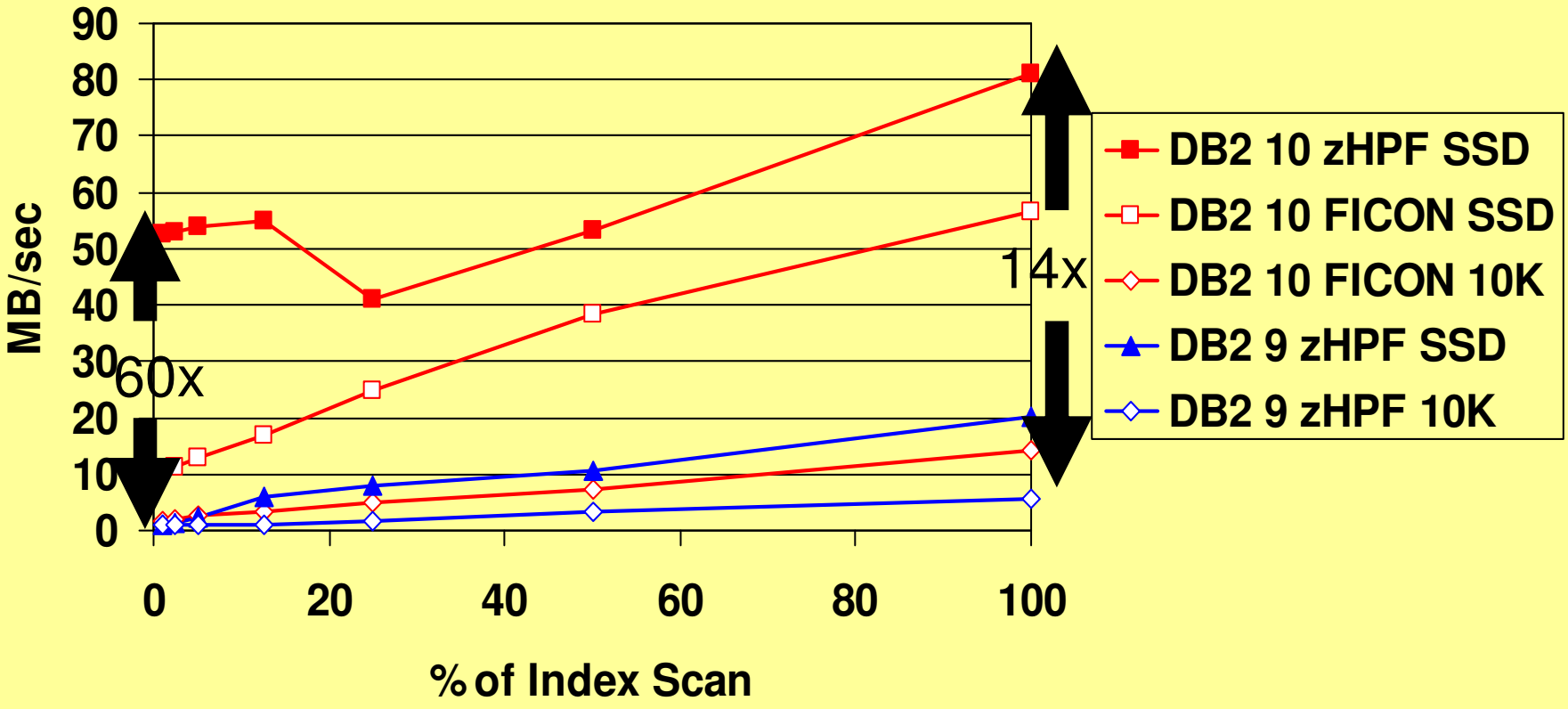
DB2 10 Logging suspension time





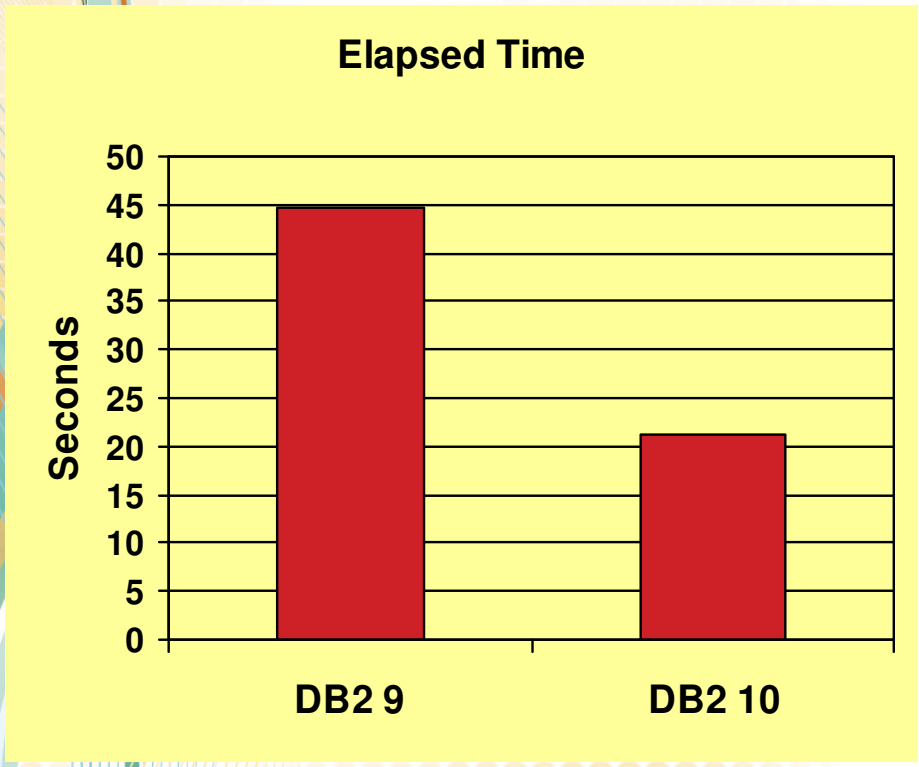
- DB2 10 is 2x to 10x faster than DB2 9 due to list prefetch

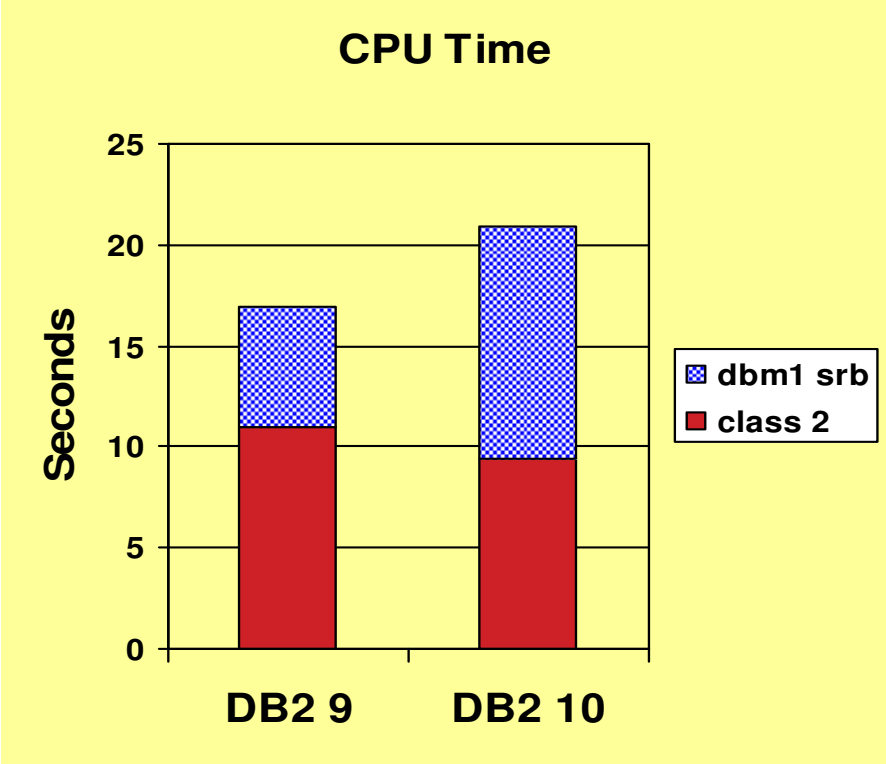
Disorganized index scans, cold cache, 4K pages
Throughput



- Migration from 10K HDD to SSD, DB2 9 to DB2 10, and FICON to zHPF is 14x to 60x faster

Insert Index I/O Parallelism





- Elapsed time decreases
- Class 2 CPU decreases
- Overall CPU increases, but the dbm1 srb time is zIIP eligible

Performance and Scalability …

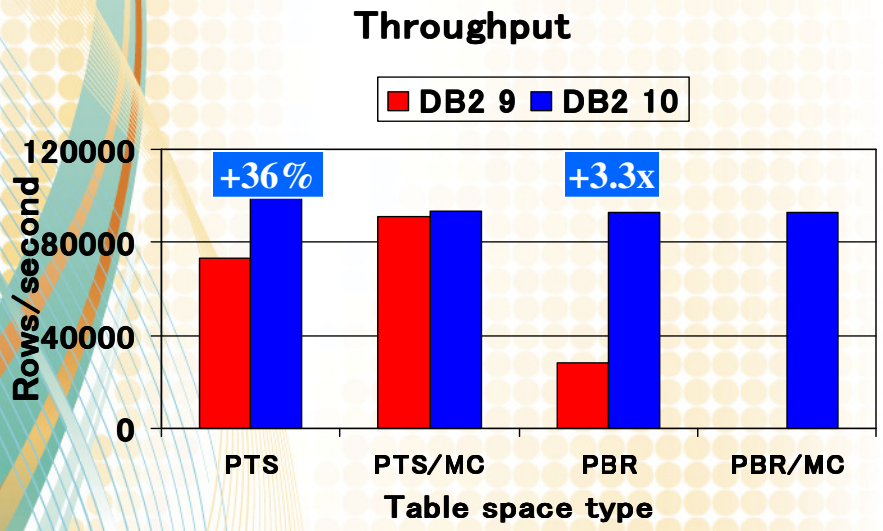
- High Concurrent INSERT performance …
  - Reduced LRSN spin for inserts to the same page
    - Works well for Multi-row Insert and INSERT within loop in a data sharing environment
- Optimization for 'pocket' sequential insert works well
  - Index manager picks the candidate RID during sequential insert (next lowest key rid)
  - Higher chance to find the space and avoiding a space search
- Parallel index IO works very well when activated for random key inserts
  - >= 3 indexes
  - Prefetch offload to zIIP to compensate for overall increase in CPU

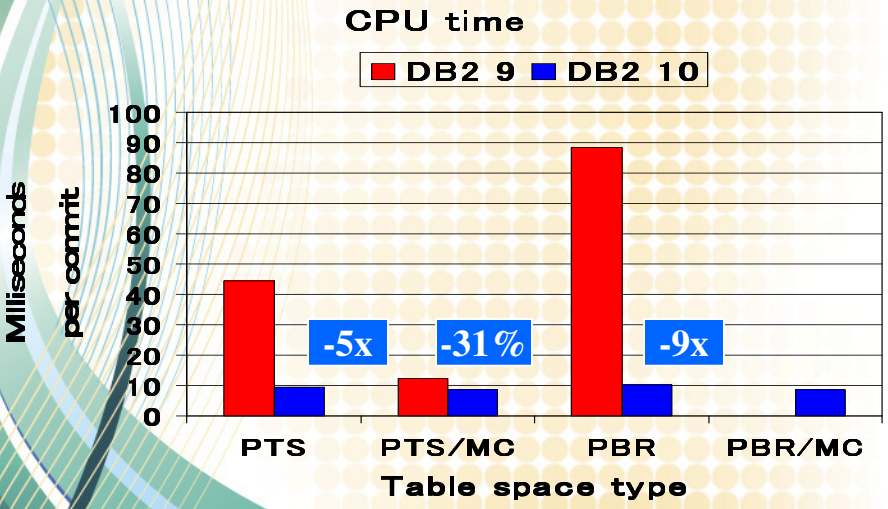
Notes: High concurrent Inserts

- Table space types
  - UTS - Universal Table Space
  - PTS - Partition table spaces
  - SEG – Segmented
  - PBR - Partition By Range (UTS)
  - PBG – Partition By Growth (UTS)
  - MC – Member cluster
- Two-way data sharing environment
  - 3 tables containing 5, 9 and 46 columns containing various data types
  - 6 indexes (4 unique, 2 non-unique indexes, 2 secondary indexes are a superset of the primary indexes)
  - Page level locking and row level locking
  - Java application
    - Multi-row (100), sequential inserts to empty table, 200 threads
    - Single row, random inserts, 240 threads

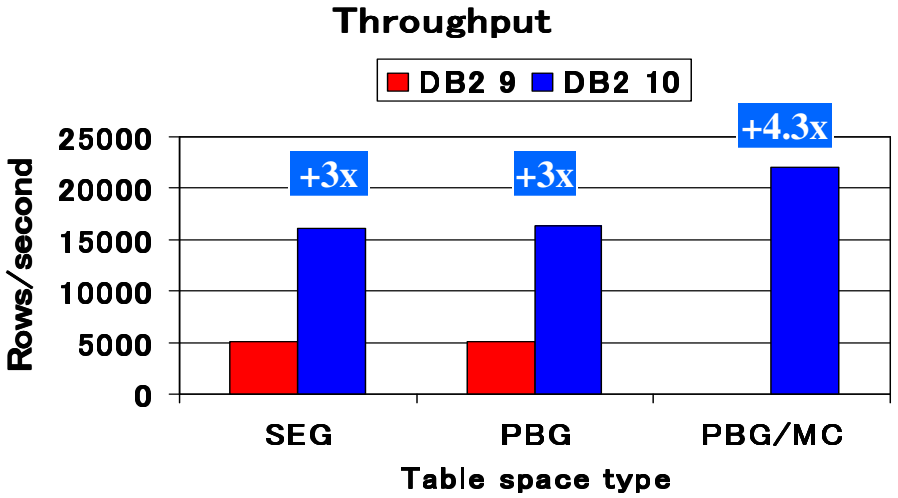
Highly concurrent sequential inserts (page level locking)

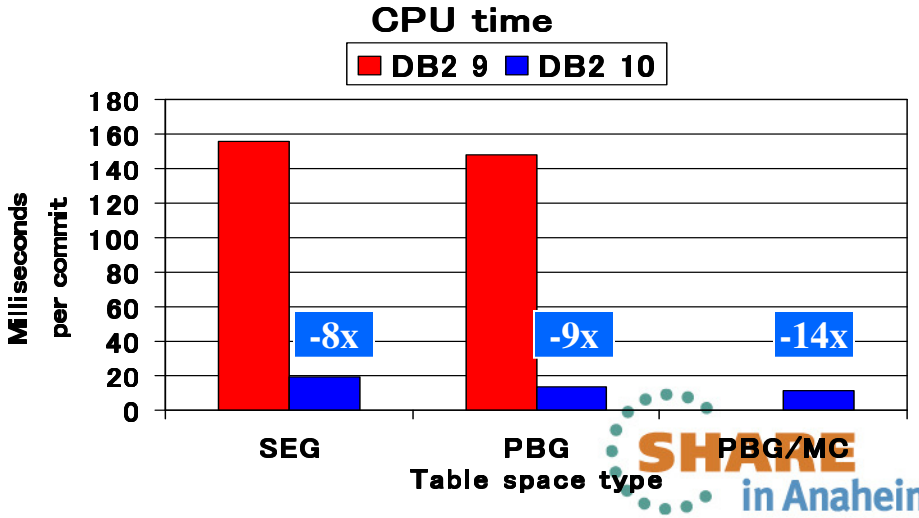
Range defined table spaces





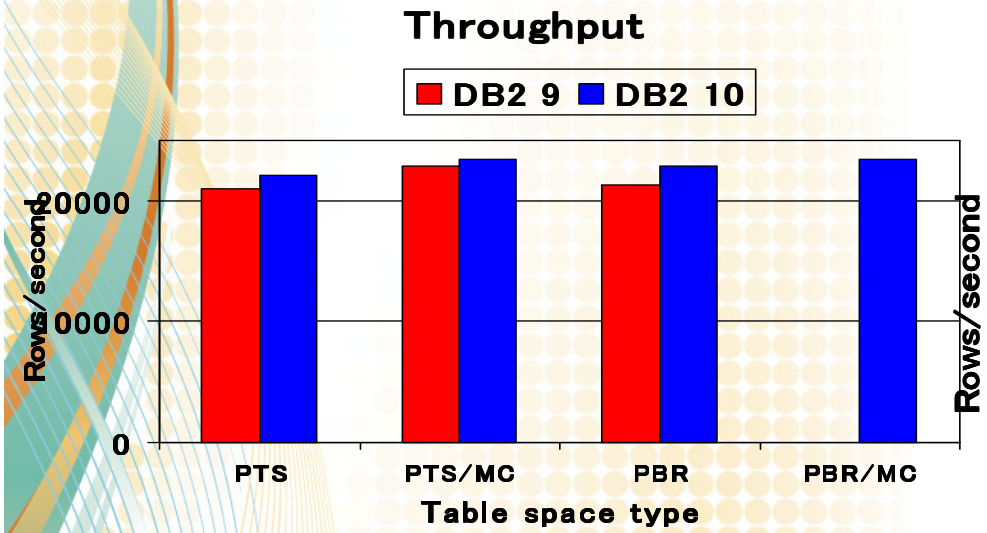
Non-range defined table spaces

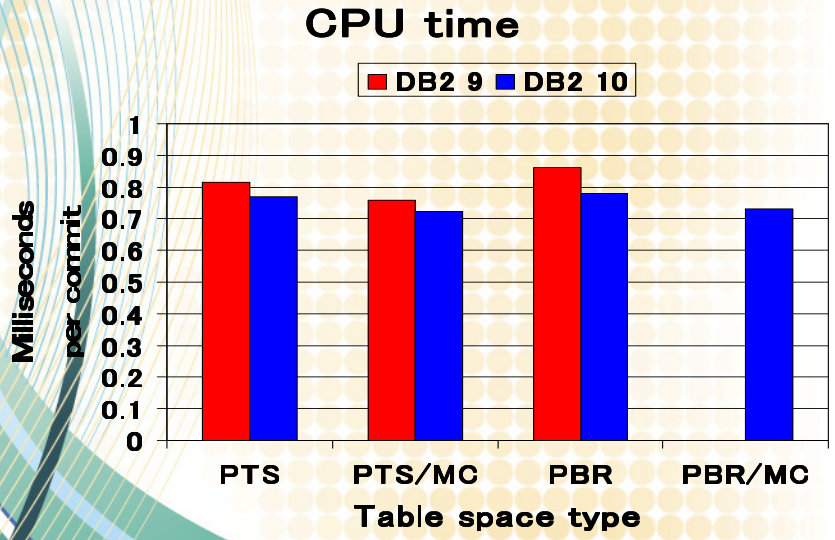




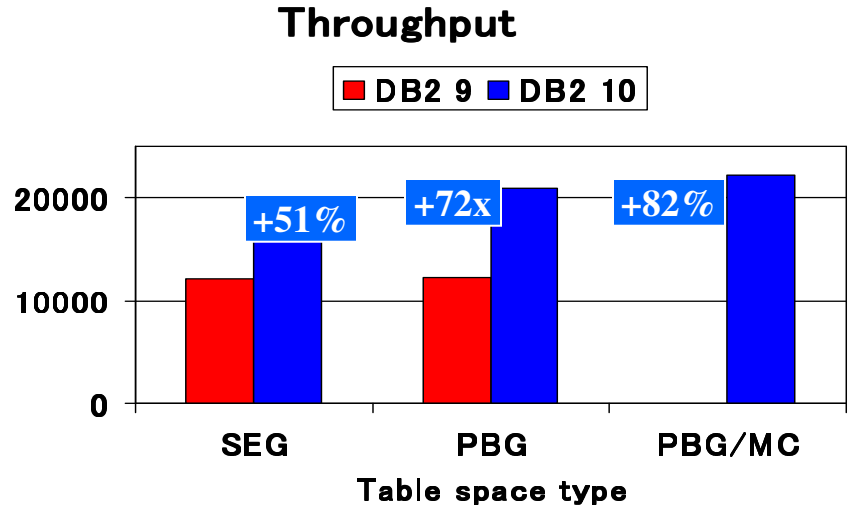
Highly concurrent random inserts (page level locking)

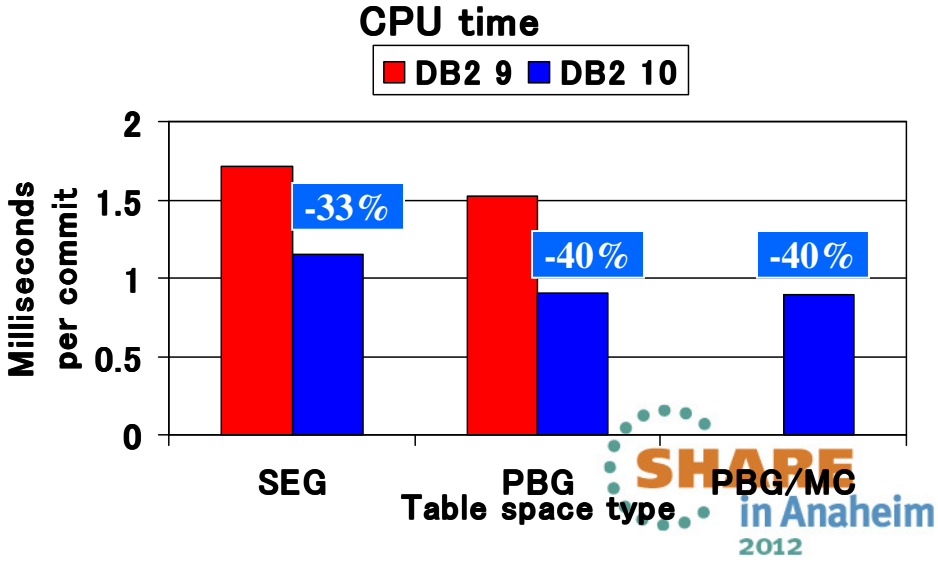
Range defined table spaces





Non-range defined table spaces

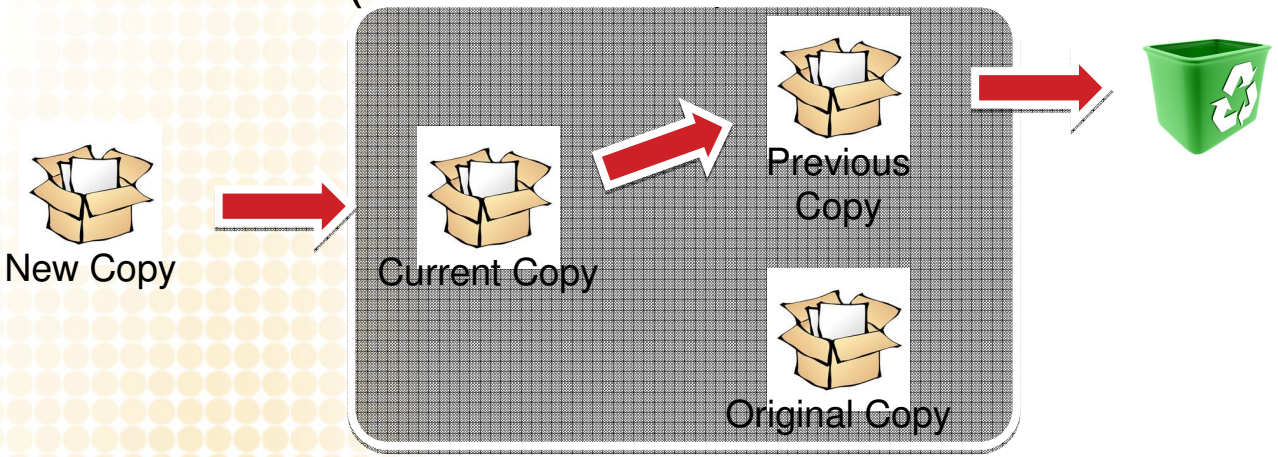




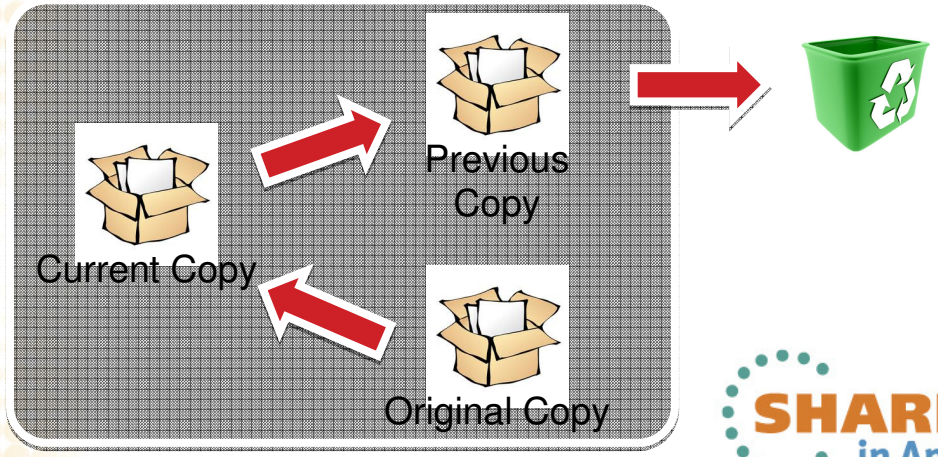
#SHAREorg

Plan Management - EXTENDED

- REBIND … PLANMGMT(EXTENDED)



- REBIND … SWITCH(ORIGINAL)



SHARE in Anaheim 2012

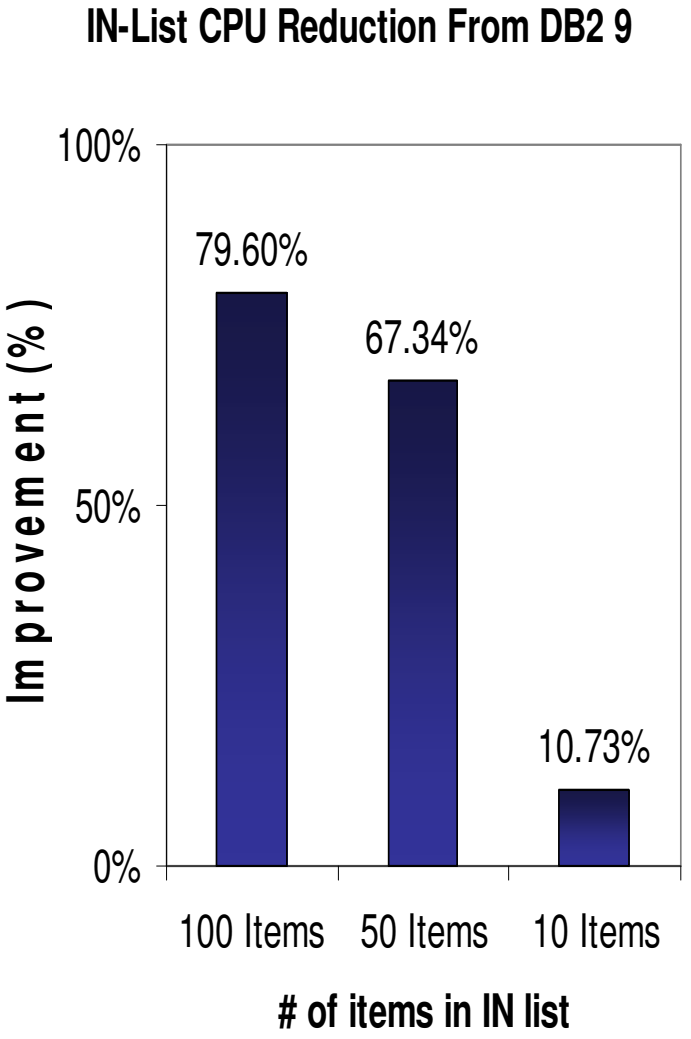
DB2 10 Plan Management

- EXTENDED is now the DEFAULT in DB2 10
  - May cause some CPU increase, but you can disable it
- No longer need to SWITCH to see information on inactive copies
- Access path enhancement
  - BIND option APREUSE to allow a package to reuse access paths
    - Useful to reduce CPU usage when migrating to DB2 10 if you don't want to risk changing the access path
  - BIND option APCOMPARE to specify what actions to take if access paths change

Stage 1 Predicate Evaluation (CM)

```
SELECT name FROM employee
WHERE  id > :H
AND  status IN ('a','b','c','d') ;
```

- Queries with stage 1 predicates
  - Both Rscan and Index access
  - Range (>, <, BETWEEN), IN, LIKE predicates
  - Dynamic optimization to evaluate the rows at runtime
  - Higher improvement with more rows evaluated and more predicates
  - Performance improvement
    - Average improvement 20% from generic 150 queries
    - Individual queries shows between 1 and 80% improvement



**Using single threads with table space scan

Buffer Pool Related Enhancements

- Large (1 MB) page frames, requires z10, PGFIX(YES)
  - Save 1-4% CPU
- In memory buffer pool: PGSTEAL=NONE
  - Similar to VPSEQT=0, but since Open preloads the object into the buffer pool, synch I/Os are avoided when DB2 is restarted
- Buffer pools allocated piecemeal as needed
  - Save storage if buffer pools had been over-allocated
- Table space buffer pools are no longer allocated when index-only access
- Avoid exhaustive local BP scan
  - Avoid CPU spike when an object's GBP dependency changes

SHARE
Technology · Connections · Results

More zIIP support

- Portion of Runstats
  - Redirection rate varies depends on the RUNSTATS options
- Parsing process of XML Schema validation
  - 100% of new validation parser is eligible
    - Can be zIIP, zAAP, or zAAP on zIIP
    - Retrofit into DB2 9 via PK90032 (preconditioning), PK90040 (enabling)
- Portion of DBM1 SRB processes
  - Especially useful with index compression and Insert Index I/O Parallelism

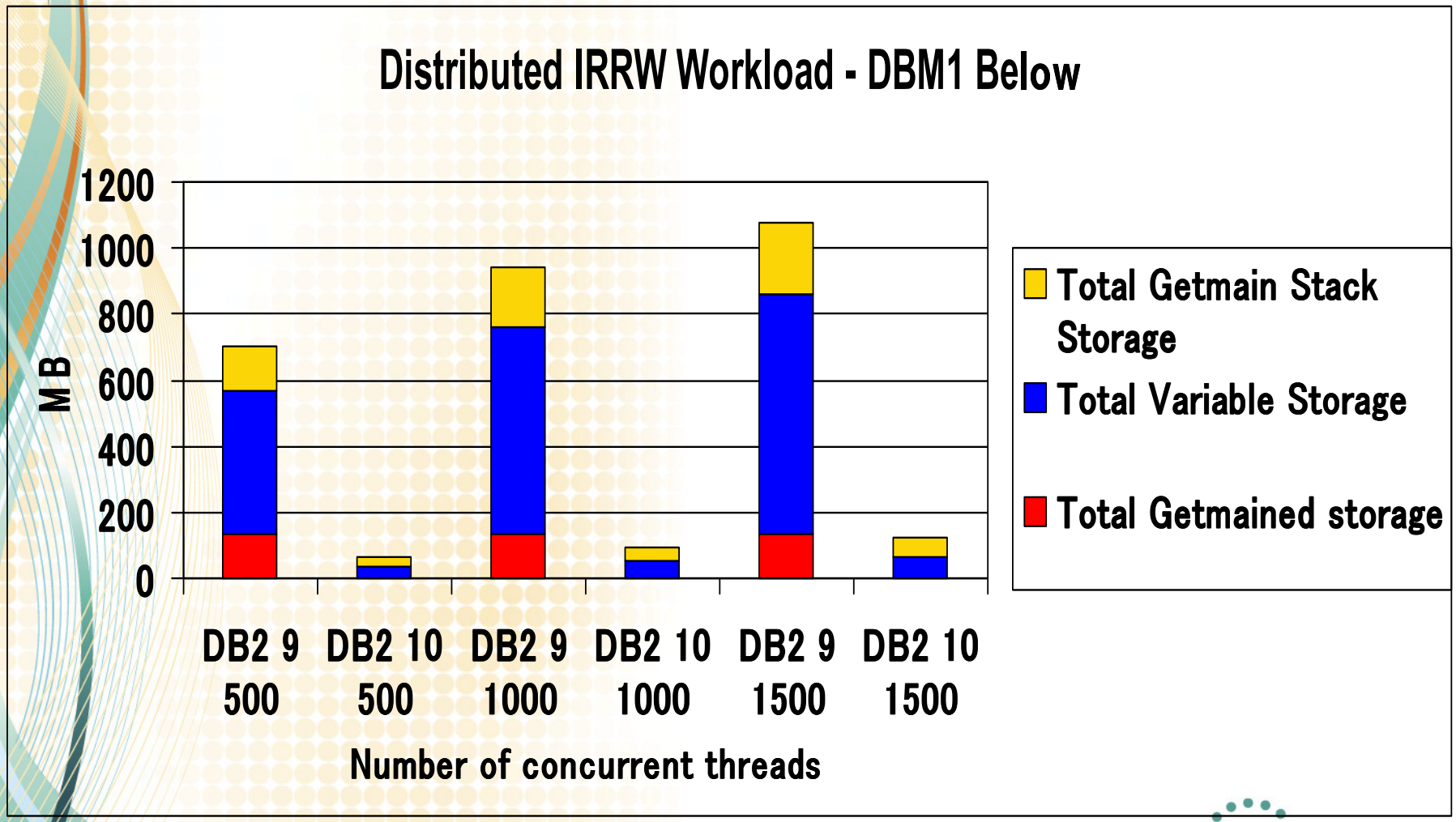
Scalability constraints that prevent a DB2 member from managing more threads

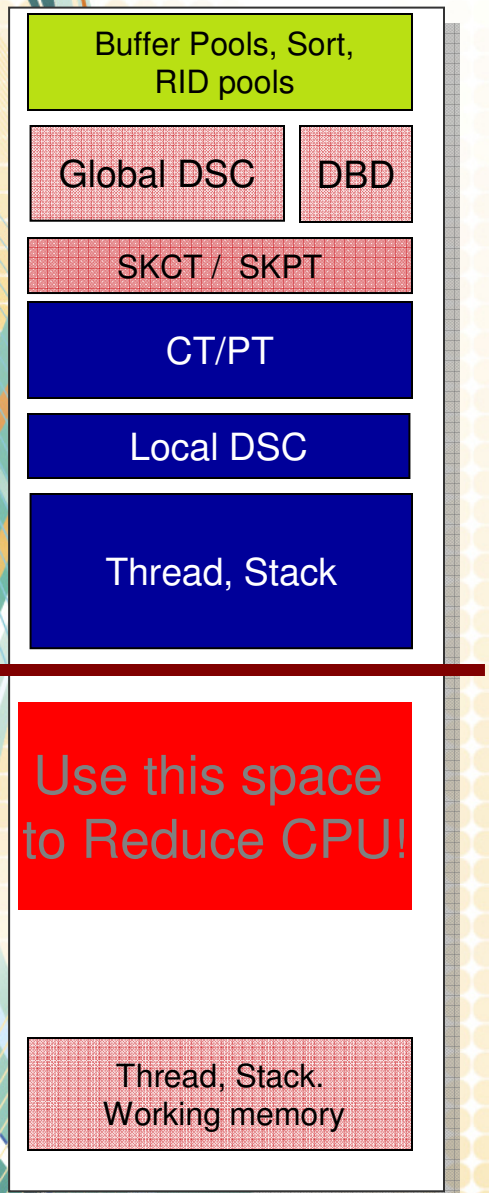
- Virtual storage constraints
  - Relieving such constraints helps reduce CPU time by using Release Deallocate, High Performance DBATs, etc.
- Latch constraints
  - Besides relieving the constraints themselves, DB2 10 eliminates most of the class 2 CPU time of suspend/resume when there is latch contention
- Catalog/directory constraints
  - UTSERIAL lock serializes DDL and utility operations
  - DB2 10 removes the UTSERIAL lock
  - Links in the directory are removed and replaced by relational structures, facilitating row level locking

CM9 or CM8 After REBIND – A Quick Look

- Possible Improvement because…
  - Re-enable SPROC
  - Optimizer enhancements by better access path selection
  - Virtual storage reduction with static applications
  - Tuned for CPU reduction instead of virtual storage constraint
    - Usage of thread reuse & RELEASE(DEALLOCATE) for local access
    - High Performance DBAT Release(Deallocate) for distributed access
    - Larger MAXKEEPD values for KEEPDYNAMIC users
  - Native SQL procedure optimization
- User Actions needed
  - REBIND
  - Monitor Common storage and Real. Possibly add more real storage or reduce BP sizes

Relief in DBM1 Storage Below the 2GB Bar



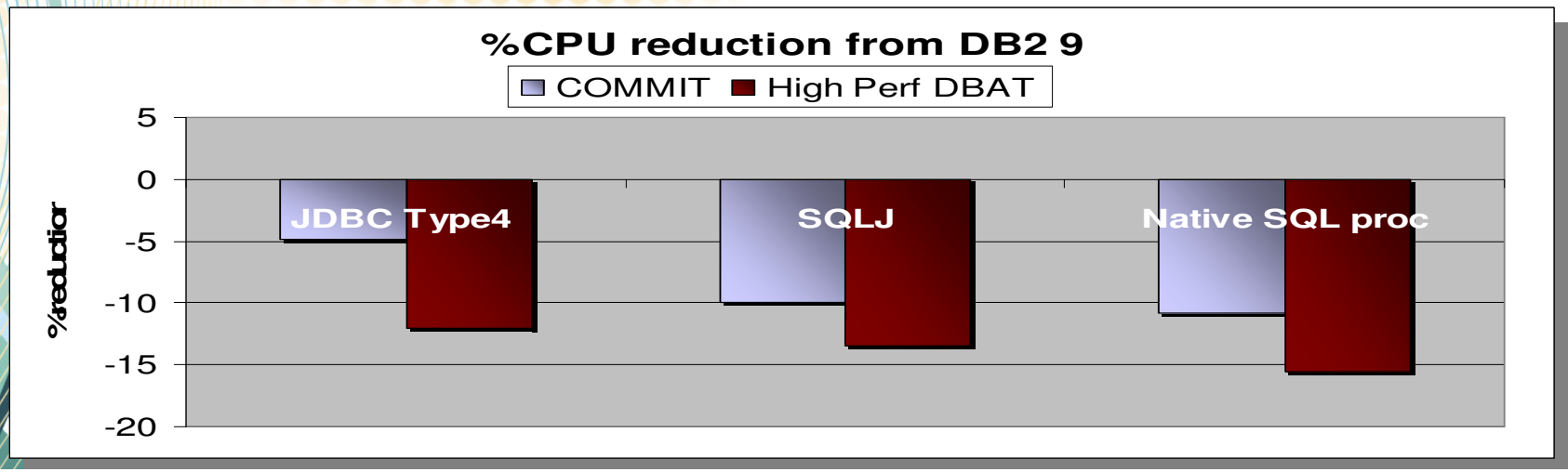


Relief in DBM1 Below The 2G Bar

- DBM1 below 2GB
  - EDM storage - All above
  - Thread + Stack - 70-90% less usage in DB2 10 compared to DB2 9
    - xPROC (SPROC, IPROC, UPROC, etc) loaded in below the 2GB bar
    - Built at BIND time, shared at runtime
- More number of threads
- Reduce CPU time at the expense of virtual storage
  - More thread reuse to avoid allocate/deallocate
  - Wider usage for bind option RELEASE(DEALLOCATE)
  - High Performance DBATs
  - Larger MAXKEEPD values for KEEPDYNAMIC=YES users to avoid short prepare

High Performance DBATs

- Re-introducing RELEASE(DEALLOCATE) in distributed packages
- High Performance DBATs reduce CPU consumption by
  - RELEASE(DEALLOCATE) to avoid repeated package allocation/deallocation
  - Avoids processing to go inactive and then back to active
- Using High Performance DBATs
  - Stay active if there is at least one RELEASE(DEALLOCATE) package exists
  - Connections will turn inactive after 200 commits to free up resources
  - Higher benefit with simple DRDA transactions that commit frequently using same packages



#SHAREorg

Notes: High Performance DBATs

- History:
  - V6 introduced Type-2 INACTIVE connection support (DBATs are pooled)
    - Initially, packages were allocated based on their RELEASE bind option
    - Pooled DBATS could linger (until POOLINAC expired) and/or get created with RELEASE(DEALLOCATE) package sections and Tablespace intent locks
  - Distributed usage of RELEASE(DEALLOCATE) magnified data maintenance restrictions
    - Customers could not break in to do DDL, BIND
    - STOP DDF MODE(SUSPEND) was provided to purge pooled DBATs but was not sufficient
    - PQ63185 was implemented to ALWAYS allocate packages on DRDA DBATs with RELEASE(COMMIT)
- DB2 10 High Performance DBAT (HP DBAT) behavior
  - DBAT will stay associated with connection at UOW boundaries if there is at least one RELEASE(DEALLOCATE) package allocated.
  - DBAT will be terminated after 200 uses (non-configurable).
  - Normal idle thread time-out (IDTHTOIN) detection will be applied to these DBATs.
  - If DBAT is in completed unit-of-work status, and POOLINAC is reached, connection will turn inactive and DBAT will be terminated.
  - No benefit and no support for CMTSTAT ACTIVE
  - No benefit for KEEPDYNAMIC YES users - such as SAP workloads

Enabling High Performance DBATs

- To Enable High Performance DBATs (two steps)
  1. REBIND with RELEASE(DEALLOCATE)
     - Default BIND option in DB2 LUW 9.7 FP3a is RELEASE DEALLOCATE for DB2 Connect and JDBC packages
  2. Then issue command –MODIFY DDF PKGREL (BNDOPT)
     - -DISPLAY DDF shows the option currently used
- To disable
  - -MODIFY DDF PKGREL (COMMIT) to go back to DB2 9 behavior
- To monitor
  - Statistics GLOBAL DDF activity report

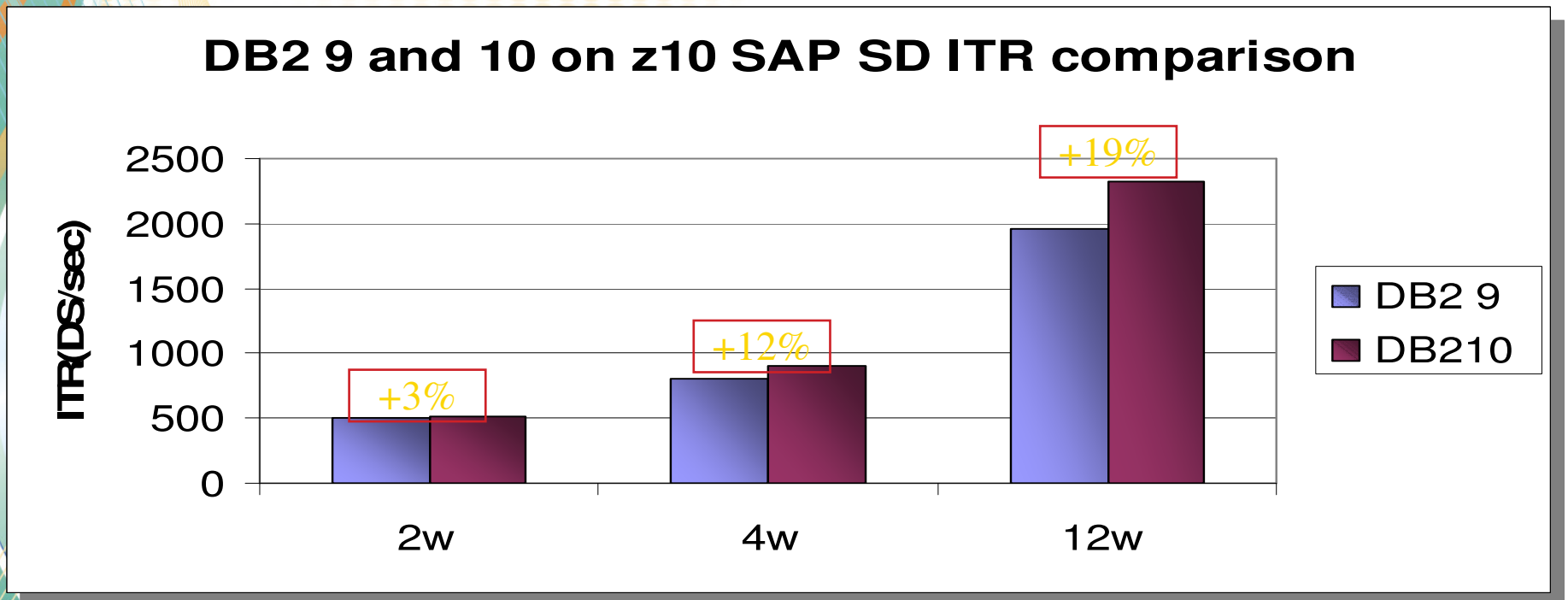
```
GLOBAL DDF ACTIVITY                QUANTITY
-------------------------          --------
CUR ACTIVE DBATS-BND DEALLC           5.39
HWM ACTIVE DBATS-BND DEALLC          10.00
```

Performance Scalability - DB2 Latches

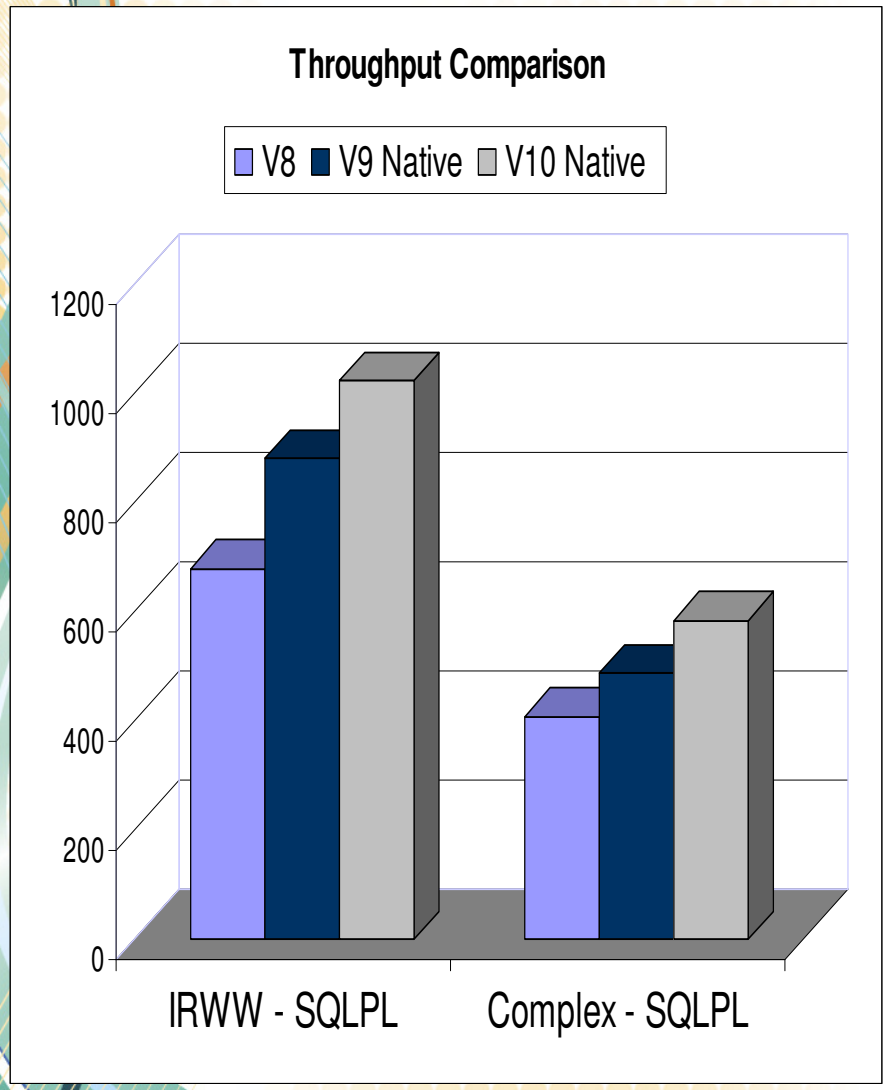
- Faster process on latch suspend/resume
  - No longer necessary to get z/OS Cross Memory Local lock
- Most of known DB2 latches are addressed in DB2 10
  - LC12 : Global Transaction ID serialization
  - LC14 : Buffer Manager serialization
  - LC19 : Log write in both data sharing and non data sharing
  - LC24 : EDM thread storage serialization (Latch 24)
  - LC24 : Buffer Manager serialization (Latch 56)
  - LC27 : WLM serialization latch for stored proc/UDF
  - LC32 : Storage Manager serialization
- Internal contention relief
  - IRLM : IRLM latch contention on hash table
  - CML : z/OS Cross Memory Local suspend lock

SAP Sales and Distribution Benchmark

- SAP SD Benchmark
  - z10 with 2, 4 and 12 processors per LPAR
  - Enabled 1 MB large frames for DB2 10
  - Higher throughput improvement as # of processors increase – Better scalability with DB2 10



Native SQL Procedure Performance Improvement



- DB2 9 Native SQL Procedure
  - No WLM address space overhead
  - zIIP eligible if called from DRDA
- DB2 10 Native SQL Procedures
  - Internal optimization to improve performance especially after DROP/RECREATE
  - Optimization for SET statements
- Measurements
  - DB2 V8-> 9 5-30% by converting external to native SQLPL
  - DB2 9 ->10 up to 20%
  - More if called from DRDA and use High Perf DBATs

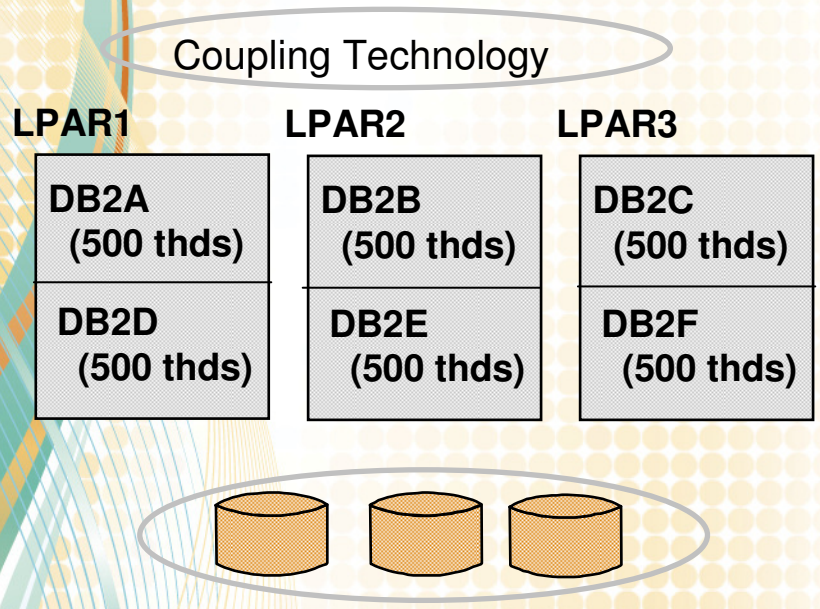
CPU Parallelism Enhancements

- More queries are eligible for parallelism
- More effective parallelism
  - Work is divided up into smaller, more even, chunks of work that are distributed among the sub-tasks
- Using a “straw model”, each sub-task selects the next chunk of work to processes as it completes the work for the previous chunk
  - If some chunks of work are slower than others, a straw model tends to keep all of the sub-tasks busy, until there is no more work left to do

#SHAREorg

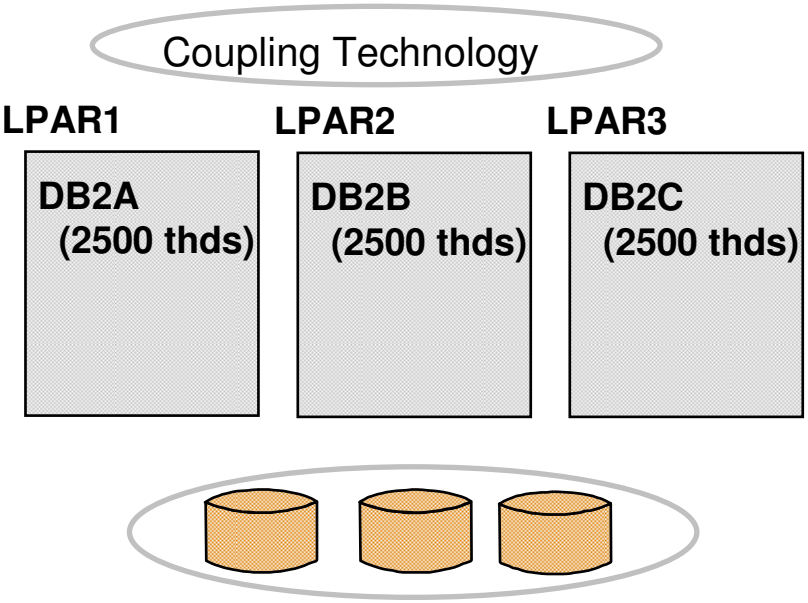
DB2 10: Massively Scalable DBMS

Today



- Data sharing and sysplex allows for efficient scale-out of DB2 images
- Sometimes multiple DB2s per LPAR

DB2 10



- More threads per DB2 image
- More efficient use of large n-ways
- Easier growth, lower costs, easier management
- Data sharing and Parallel Sysplex still required for very high availability and scale
- Rule of thumb: save ½% CPU for each member reduced, more on memory

Real Storage

- DB2 9 monitored its use of virtual storage
  - When a constraint is detected, DB2 employs storage contraction (at a CPU cost) to avoid losing *the DB2 member*
- DB2 10 likely won't have any such *private* virtual storage constraints, but it may encounter real storage constraints.
- DB2 10 monitors its use of real storage
  - When a constraint is detected, DB2 employs storage contraction (at a CPU cost) to avoid losing *the whole LPAR*.
  - A new message identifies when real storage is constrained
  - ZPARM REALSTORAGE_MAX can be used to restrict the amount of real storage used by an individual DB2 member

#SHAREorg

SHARE
Technology · Connections · Results

DB2 10
New Function Mode

SHARE in Anaheim 2012

New index option: INCLUDE indexes

- Addition of non-key columns don't affect the uniqueness contraint
- Performance improvement, CPU and Elapsed time reduction
  - 2 Indexes vs 1 index with INCLUDE columns shows 30% cpu reduction in insert
- Easy way to get index-only access without impacting insert performance
- DASD space saving
- Less choice for Access Path Selection

New Universal Table Space options

- Member cluster for both PBG and PBR
  - Avoids insert hot spots
  - Ideal for highly concurrent inserts in data sharing environment
- Hash Access
  - Avoids hot spots
  - Avoids the CPU and I/O overhead associated with the cluster index
- Inline LOBs
  - VARCHAR-like performance for small LOBs
  - 70-90% CPU and I/O reduction for LOBs that can be fully inlined
  - DASD space savings
- PBG Workfiles
  - Easier way to manage workfile DASD space usage
- ALTER/REORG can be use to convert to UTS, or to change phsyical attributes such as page size

SHARE
Technology · Connections · Results

Other LOB Enhancements

- LOAD/UNLOAD use of RECFM=VBS can replace LOB File Reference Variables when loading and unload LOBs
  - 5x improvement over USS file systems
  - Works with tape
- Two times faster LOAD LOB
- LOB materialization avoidance (in DBM1)
  - Applies to LOBs bigger than 2 MB
  - Less real storage used
  - 30% reduced elapsed time
  - 15% reduced CPU time
- REORG LOB with SHRLEVEL CHANGE
- REORG with AUX=YES allows rows containing LOBs to be moved from one partition to another

#SHAREorg

SHARE
Technology · Connections · Results

Beta Performance Feedback and Performance Summary

SHARE in Anaheim 2012

Some of Beta Feedback in Workload level

Workload	Results
CICS online transactions	Approx. 7% CPU reduction in DB2 10 CM after REBIND, additional reduction when 1MB page frames are used for selective buffer pools
CICS online transactions	Approx 10% CPU reduction from DB2 9
CICS online transactions	Approx 5% CPU reduction from DB2 V8, 15-20% reduction with protected thread and release deallocate
CICS online transactions	10% CPU increase from some of very simple transactions -> addressing with PM31614
Distributed Concurrent Insert	50% DB2 elapsed time reduction, 15% chargeable CPU reduction after enabling high perf DBAT
Data sharing concurrent insert	38% CPU reduction
Queries	Average CPU reduction 28% from V8 to DB2 10 NFM
Batch	Overall 20-25% CPU reduction after rebind packages

Other performance enhancements

- Improved monitoring
- SMF Compression and improved ROLLUP accounting
- Insert space search
- Insert auto compression
- Query
- Workfile
- Referential integrity checking
- XML
- Data set level flashcopy for creating VSAM image copies
- Auto stats using Runstats profiles
- Dynamic statement with literals
- Limited block fetch extended to JCC Type 2 drivers
- Return to client results sets

References

- DB2 10 for z/OS Performance Topics

  http://www.redbooks.ibm.com/abstracts/sg247942.html

- DB2 10 for z/OS Technical Overview

  http://www.redbooks.ibm.com/abstracts/sg247892.html

#SHAREorg

SHARE
Technology · Connections · Results

Questions?

Email : bergerja@us.ibm.com