DB2 10 for z/OS Optimization and Query Performance Improvements

James Guo
DB2 for z/OS Performance
IBM Silicon Valley Lab

August 11, 2011 6 PM – 7 PM
Session Number 9524

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.

Agenda

- **Bind/Prepare**
  - Plan management
  - Hints/Bind options
  - Explain
  - Dynamic Statement Caching
  - REOPT
- **Optimizer costing**
- **Runtime query performance**
- **Indexing**
- **Complex queries**

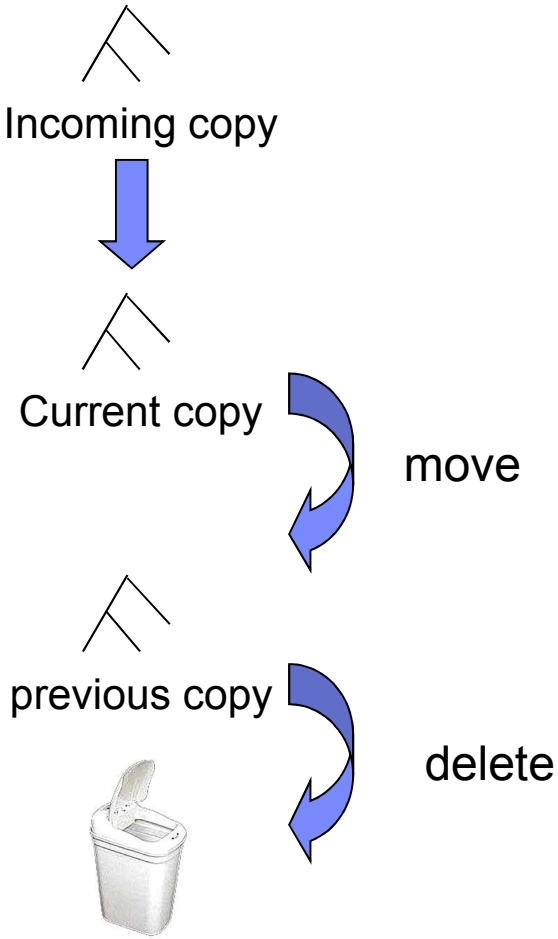
Plan Management Overview

- **Ability to backup your static SQL packages (DB2 9)**

- **At REBIND**
  - Save old copies of packages in Catalog/Directory
  - Switch back to previous or original version
- **Two flavors**
  - BASIC
    - 2 copies: Current and Previous
  - EXTENDED
    - 3 copies: Current, Previous, Original
  - Default controlled by a ZPARM
  - Also supported as REBIND options

Plan Management - BASIC support

REBIND … PLANMGMT(BASIC)



REBIND … SWITCH(PREVIOUS)

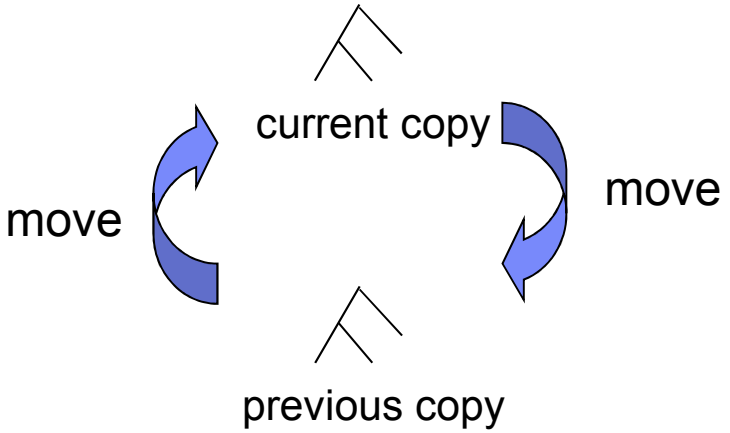
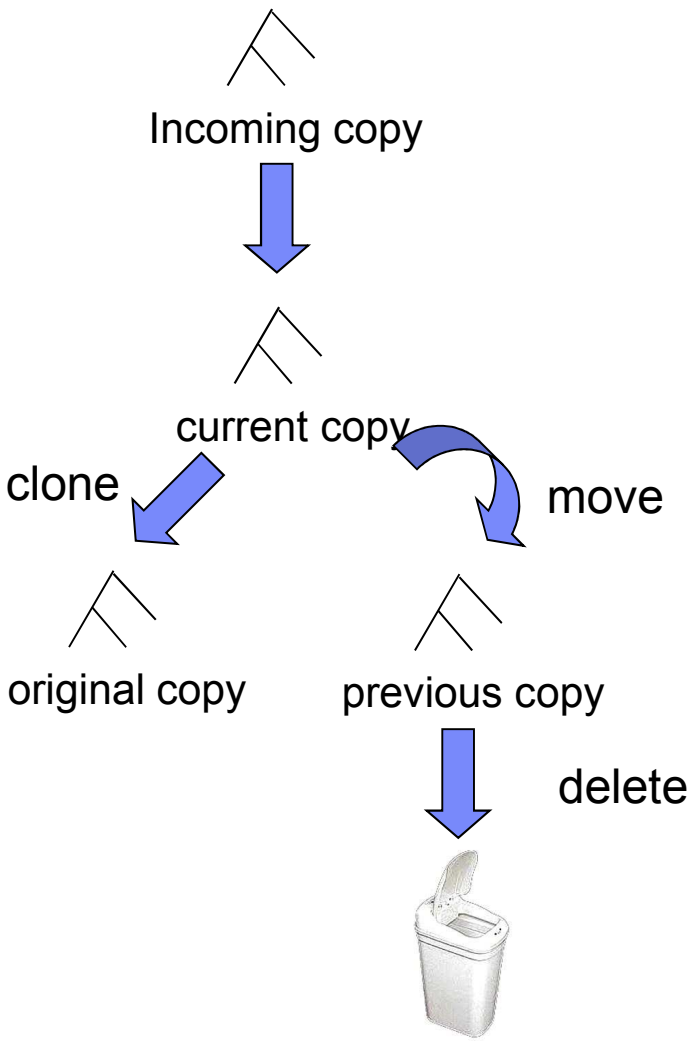


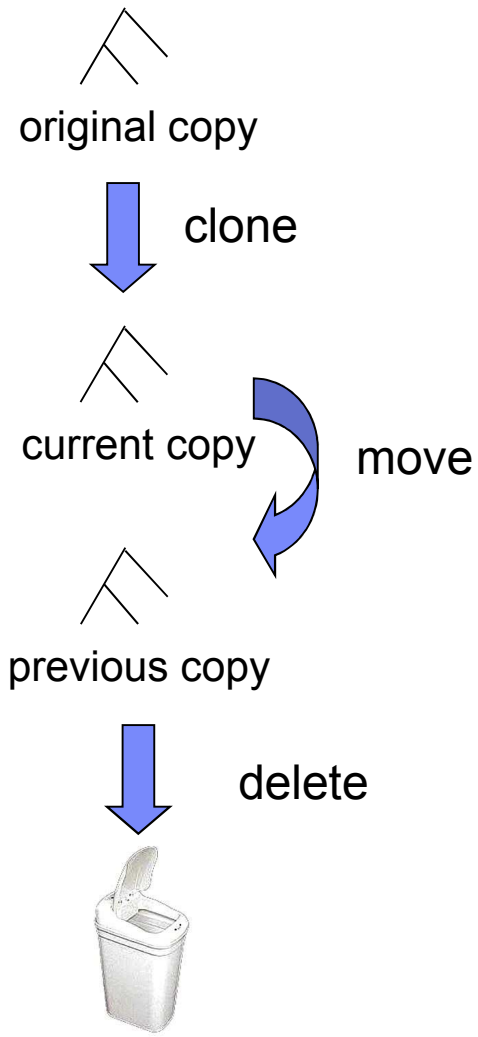
Chart is to be read from bottom to top

Plan Management - EXTENDED support

REBIND … PLANMGMT(EXTENDED)



REBIND … SWITCH(ORIGINAL)



DB2 10 Updates to Plan Management

- **SYSIBM.SYSPACKCOPY**
  - New catalog table
  - Hold SYSPACKAGE-style metadata for any previous or original package copies
  - No longer need to SWITCH to see information on inactive copies
    - Complaint from DB2 9

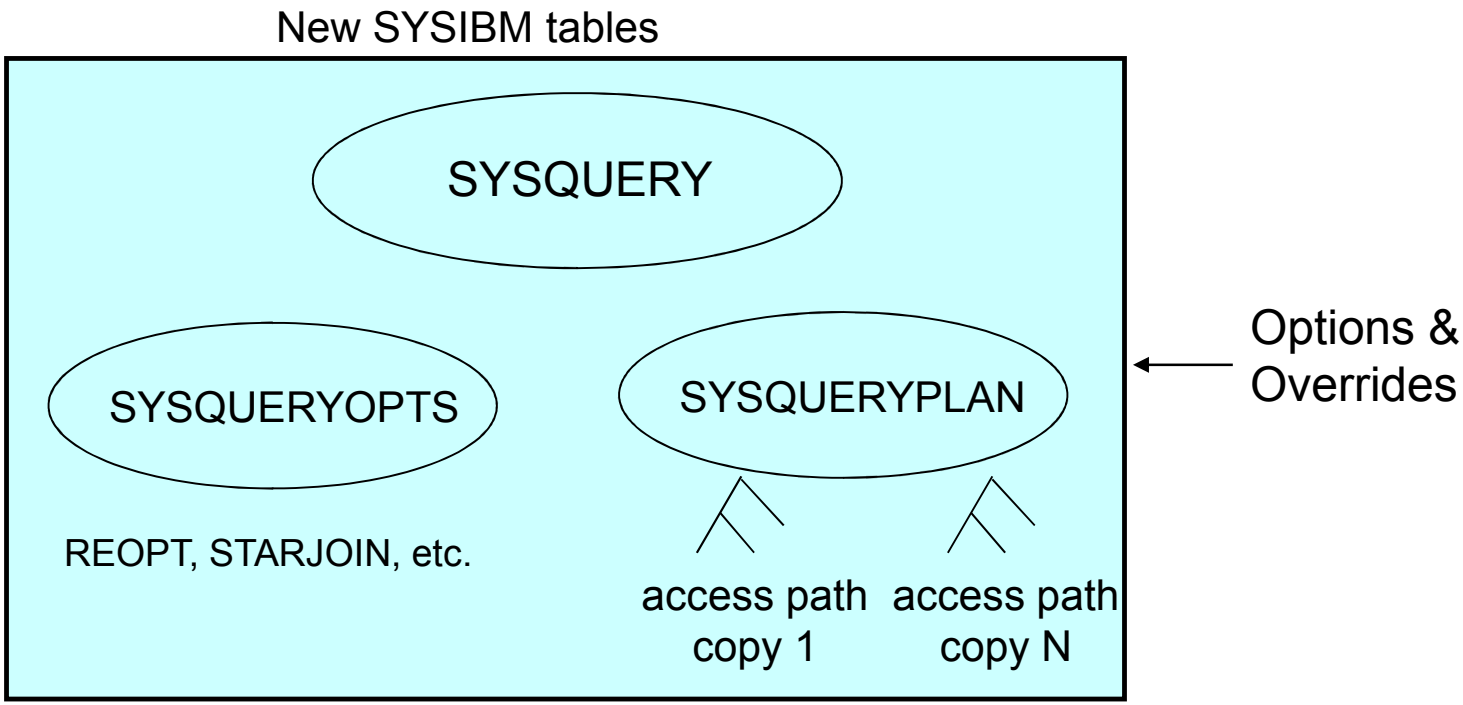
- **APRETAINDUP option of REBIND**
  - Default YES
    - Retain duplicate for BASIC or EXTENDED
  - Optional NO
    - Do not retain duplicate access path as PREVIOUS or ORIGINAL
      - PREVIOUS/ORIGINAL must be from DB2 9 or later

Access Path Stability with statement level hints

- **Current limitations in hint matching**
  - QUERYNO is used to link queries to their hints – a bit fragile
  - For dynamic SQL, require a change to apps – can be impractical
- **New mechanisms:**
  - Associate query text with its corresponding hint … more robust
  - Hints enforced for the entire DB2 subsystem
    - irrespective of static vs. dynamic, etc.
  - Hints integrated into the access path repository

- **PLAN_TABLE isn't going away**
- **Only the "hint lookup" mechanism is being improved.**

Access Path Repository – Hints/Statement level



Statement level hints (cont.)

- **Steps to use new hints mechanism**
  - Populate a user table DSN_USERQUERY_TABLE with query text
  - Populate PLAN_TABLE with the corresponding hints
  - Run new command BIND QUERY
    - To integrate the hint into the repository.
  - FREE QUERY can be used to remove the hint.

Statement-level BIND options

- **Statement-level granularity may be required rather than:**
  - Subsystem level ZPARMs (STARJOIN, SJTABLES, MAX_PAR_DEGREE)
  - Package level BIND options (REOPT, DEF_CURR_DEGREE)
- **For example**
  - Only one statement in the package needs REOPT(ALWAYS)

- **New mechanism for statement-level bind options:**
  - Similar to mechanism used for hints
  - DSN_USERQUERY_TABLE can also hold per-statement options

Literal Replacement

- **Dynamic SQL with literals can now be re-used in the cache**
  - Literals replaced with &
    - Similar to parameter markers but not the same
- **To enable either you:-**
  - Put CONCENTRATE STATEMENTS WITH LITERALS in the PREPARE ATTRIBUTES clause
  - Or set LITERALREPLACEMENT in the ODBC initialization file
  - Or set the keyword enableLiteralReplacement='YES' in the JCC Driver
- **Lookup Sequence**
  - Original SQL with literals is looked up in the cache
  - If not found, literals are replaced and new SQL is looked up in the cache
    - Additional match on literal usability
    - Can only match with SQL stored with same attribute, not parameter marker
  - If not found, new SQL is prepared and stored in the cache

Literal Replacement …

- **Example:**

```
WHERE ACCOUNT_NUMBER = 123456
```

  – This would be replaced by

```
WHERE ACCOUNT_NUMBER = &
```

- **Performance Expectation**
  – Using parameter marker still provides best performance
  – Biggest performance gain for repeated SQL with different literals
  – NOTE: Access path is not optimized for literals
    - True for parameter markers/host variables today
    - Need to use REOPT for that purpose

Agenda

- **Bind/Prepare**
- **Optimizer costing**
  - RUNSTATS
  - Cost model enhancements
  - Subquery costing
- **Runtime query performance**
- **Indexing**
- **Complex queries**

Autonomic Statistics Solution Overview

- **Autonomic Statistics is implemented though a set of Stored Procedures**
  - *Stored procedures are provided to enable administration tools and packaged applications to automate statistics collection.*
    - ADMIN_UTL_MONITOR
    - ADMIN_UTL_EXECUTE
    - ADMIN_UTL_MODIFY
  - Working together, these SP's
    - Determine what stats to collect
    - Determine when stats need to be collected
    - Schedule and Perform the stats collection
    - Records activity for later review
  - *See Chapter 11 "Designing DB2 statistics for performance" in the DB2 10 for z/OS Performance Monitoring and Tuning Guide for details on how to configure autonomic monitoring directly within DB2.*

RUNSTATS Simplification/Performance Overview

- **RUNSTATS options to SET/UPDATE/USE a stats profile**
  - Integrate specialized statistics into generic RUNSTATS job
    - RUNSTATS … TABLE tbl COLUMN(C1)… **SET PROFILE**
      - Alternatively use **SET PROFILE FROM EXISTING STATS**
    - RUNSTATS … TABLE tbl COLUMN(C5)… **UPDATE PROFILE**
    - RUNSTATS … TABLE tbl **USE PROFILE**

- **New option for page-level sampling**
  - But what percentage of sampling to use?
    - RUNSTATS … TABLE tbl **TABLESAMPLE SYSTEM AUTO**

Optimizer Validation with Realtime Stats

- **Index Probing & RTS lookup**
  - Estimate # of rids within a given start/stop index key range at bind/prepare

- **Exploited when these two conditions are met.**
  - Query has matching index-access local predicate
  - Predicate contain literals, or REOPT(ALWAYS|ONCE|AUTO)
- **And 1 of the following is also true**
  - Predicate is estimated to qualify no rows
  - Stats indicate the table contains no rows
  - Table is defined as VOLATILE or qualifies for NPGTHRSH

- **New EXPLAIN table to externalize runtime estimates**
  - User managed DSN_COLDIST_TABLE

DB2 10 - Minimizing Optimizer Challenges

- **Potential causes of sub-optimal plans**
  - Insufficient statistics
  - Unknown literal values used for host variables or parameter markers

- **DB2 10 Optimizer will evaluate the risk for each predicate**
  - For example: WHERE BIRTHDATE < ?
    - Could qualify 0-100% of data depending on literal value used
  - As part of access path selection
    - Compare access paths with close cost and choose lowest risk plan

Extending VOLATILE TABLE usage

- **VOLATILE TABLE support added in DB2 V8**
  - Targeted to SAP Cluster Tables
    - Use Index access whenever possible
    - **Avoids list prefetch**
      - Can be a problem for OR predicates or UPDATEs at risk of loop

- **DB2 10 provides VOLATILE to general cases**
  - Tables matching SAP cluster tables will maintain original limitations
    - Table with 1 unique index
  - Tables with > 1 index will follow NPGTHRSH rules
    - Use Index access whenever possible
    - **No limitation on list prefetch**
    - Less chance of getting r-scan when list-prefetch plan is only alternative

Agenda

- **Bind/Prepare**
- **Optimizer costing**
- **Runtime query performance**
  - Sort/sort avoidance
  - Sparse index
  - Predicate application
- **Indexing**
- **Complex queries**

Sort Performance Enhancements

- **FETCH FIRST n ROWS ONLY (FFnR) and Sort**
  - DB2 9 added in-memory replacement for FFnR to avoid sort
    - Provided (n * (sort key + data)) < 32K
  - DB2 10 extends this to 128K

- **Avoid workfile usage for small sorts**
  - DB2 9 avoided allocating WF for final sort only
    - If <= 255 rows and result < 32K (sort key + data)
  - DB2 10 extends this to intermediate sorts also
    - Except for parallelism or SET function

Improvements to predicate application

- **Major enhancements to OR and IN predicates**
  - Improved performance for AND/OR combinations and long IN-lists
    - General performance improvement to stage 1 predicate processing
  - IN-list matching
    - Matching on multiple IN-lists
    - Transitive closure support for IN-list predicates
    - List prefetch support
    - Trim IN-lists from matching when preceding equals are highly filtering
  - SQL pagination
    - Single index matching for complex OR conditions

- **Many stage 2 expressions to be executed at stage 1**
  - Stage 2 expressions eligible for index screening
    - Not applicable for list prefetch
  - Externalized in DSN_FILTER_TABLE column PUSHDOWN

IN-list Table - Table Type 'I' and Access Type 'IN'

- The IN-list predicate will be represented as an in-memory table if:
  - **List prefetch is chosen, OR**
  - **More than one IN-list is chosen as matching.**
  - **The EXPLAIN output associated with the in-memory table will have:**
    - **New Table Type: TBTYPE – 'I'**
    - **New Access Type: ACTYPE – 'IN'**

```
SELECT *
   FROM T1
   WHERE T1.C1 IN (?, ?, ?);
```

QBNO	PLANNO	METHOD	TNAME	ACTYPE	MC	ACNAME	QBTYPE	TBTYPE	PREFETCH
1	1	0	DSNIN001(01)	*IN*	0		SELECT	*I*	
1	2	1	T1	I	1	T1_IX_C1	SELECT	T	*L*

IN-list Predicate Transitive Closure (PTC)

```
SELECT *
FROM T1, T2
WHERE T1.C1 = T2.C1
  AND T1.C1 IN (?, ?, ?)

  AND T2.C1 IN (?, ?, ?) ←   Optimizer can generate
                             this predicate via PTC
```

- **Without IN-list PTC (DB2 9)**
  - Optimizer will be unlikely to consider T2 is the first table accessed
- **With IN-list PTC (DB2 10)**
  - Optimizer can choose to access T2 or T1 first.

SQL Pagination

- **Targets 2 types of queries**
  - Cursor scrolling (pagination) SQL
    - Retrieve next n rows
      - Common in COBOL/CICS and any screen scrolling application
    - Not to be confused with “scrollable cursors”
  - Complex OR predicates against the same columns
    - Common in SAP

- **In both cases:**
  - The OR (disjunct) predicate refers to a single table only.
  - Each OR predicate can be mapped to the same index.
  - Each disjunct has at least one matching predicate.

Simple scrolling – Index matching and ORDER BY

- Scroll forward to obtain the next 20 rows
  - Assumes index is available on (LASTNAME, FIRSTNAME)
  - WHERE clause may appear as:

```
WHERE (LASTNAME='JONES' AND FIRSTNAME>'WENDY')
   OR (LASTNAME>'JONES')
ORDER BY LASTNAME, FIRSTNAME;
```

  - DB2 10 supports
    - Single matching index access with sort avoided
  - DB2 9 requires
    - Multi-index access, list prefetch and sort
    - OR, extra predicate (AND LASTNAME >= ‘JONES’) for matching single index access and sort avoidance

Complex OR predicates against same index

- Given WHERE clause
  - And index on one or both columns

```
WHERE (LASTNAME='SMITH' AND FIRSTNAME='JOHN')
   OR (LASTNAME='JONES');
```

QBlockno	Planno	Accessname	Access_Type	Matchcols	Mixopseq
1	1	IX1	NR	2	1
1	1	IX1	NR	1	2

Minimizing impact of RID failure

- **RID overflow can occur for**
  - Concurrent queries each consuming shared RID pool
  - Single query requesting > 25% of table or hitting RID pool limit

- **DB2 9 will fallback to tablespace scan***

- **DB2 10 will continue by writing new RIDs to workfile**
  - Work-file usage may increase
    - Mitigate by increasing RID pool size (default increased in DB2 10).
    - MAXTEMPS_RID zparm for maximum WF usage for each RID list

* Hybrid join can incrementally process. Dynamic Index ANDing will use WF for failover.

Agenda

- **Bind/Prepare**
- **Optimizer costing**
- **Runtime query performance**
- **Indexing**
  - Index on expression
  - Tracking index use
  - Sparse index
  - Include columns
- **Complex queries**

Index Include Columns

- **Index INCLUDE columns**
  - Create an Index as UNIQUE, and add additional columns
  - Ability to consolidate redundant indexes

```
INDEX1 UNIQUE (C1)
INDEX2 (C1,C2)
```

Consolidate to

```
INDEX1 UNIQUE (C1) INCLUDE (C2)
```

Agenda

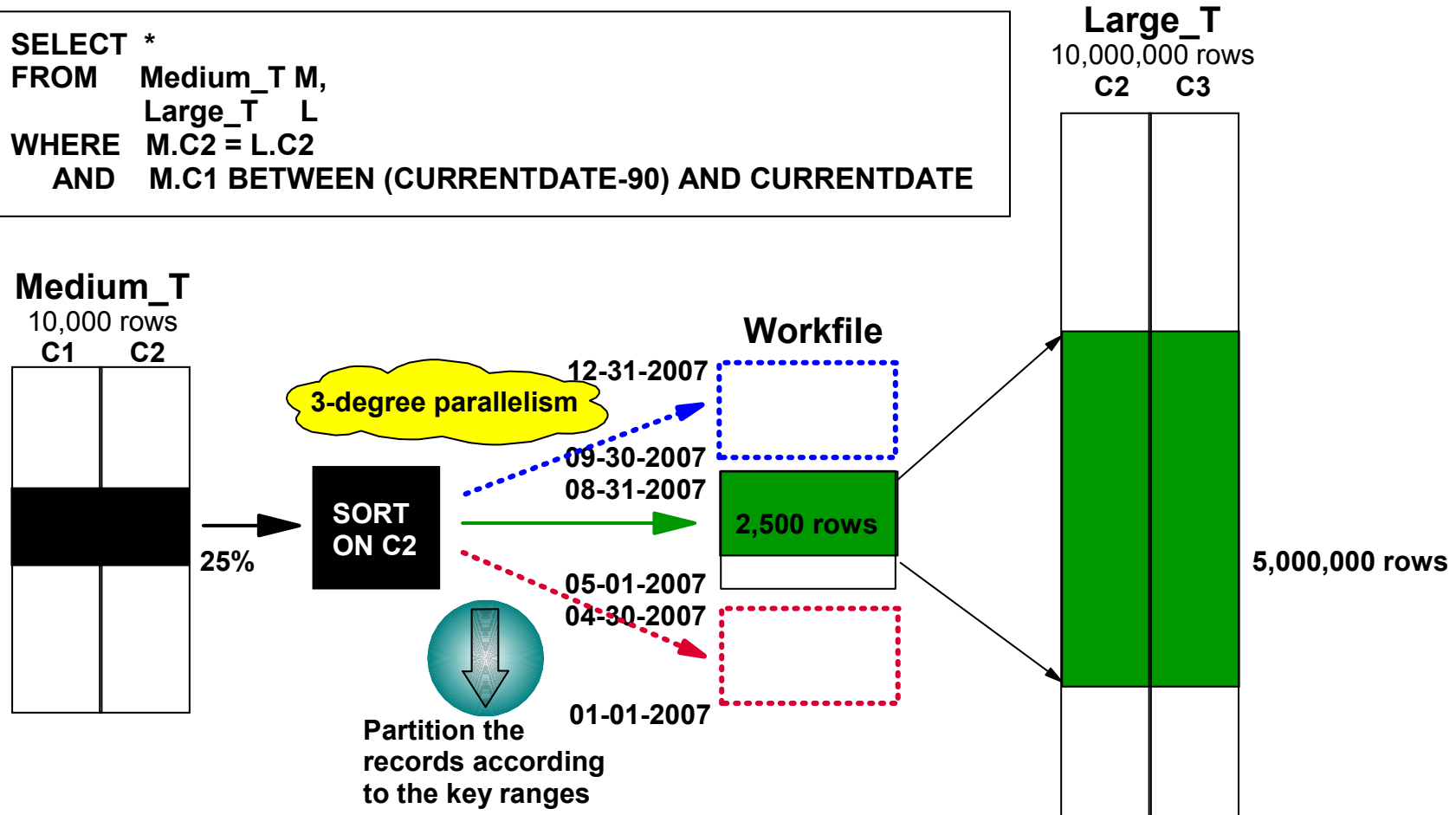
- **Bind/Prepare**
- **Optimizer costing**
- **Runtime query performance**
- **Indexing**
- **Complex queries**
  - Parallelism
  - BI/DW

Parallelism Enhancements - Effectiveness

- **Previous Releases of DB2 may use Key Range Partitioning**
  - Key Ranges Decided at Bind Time
  - Based on Statistics (low2key, high2key, column cardinality)
    - Assumes uniform data distribution
    - Histograms can help
      - But rarely collected
  - If Statistics are outdated or data is not uniformly distributed what happens to performance?

Key range partition - Today

```
SELECT  *
FROM    Medium_T M,
        Large_T  L
WHERE   M.C2 = L.C2
  AND   M.C1 BETWEEN (CURRENTDATE-90) AND CURRENTDATE
```



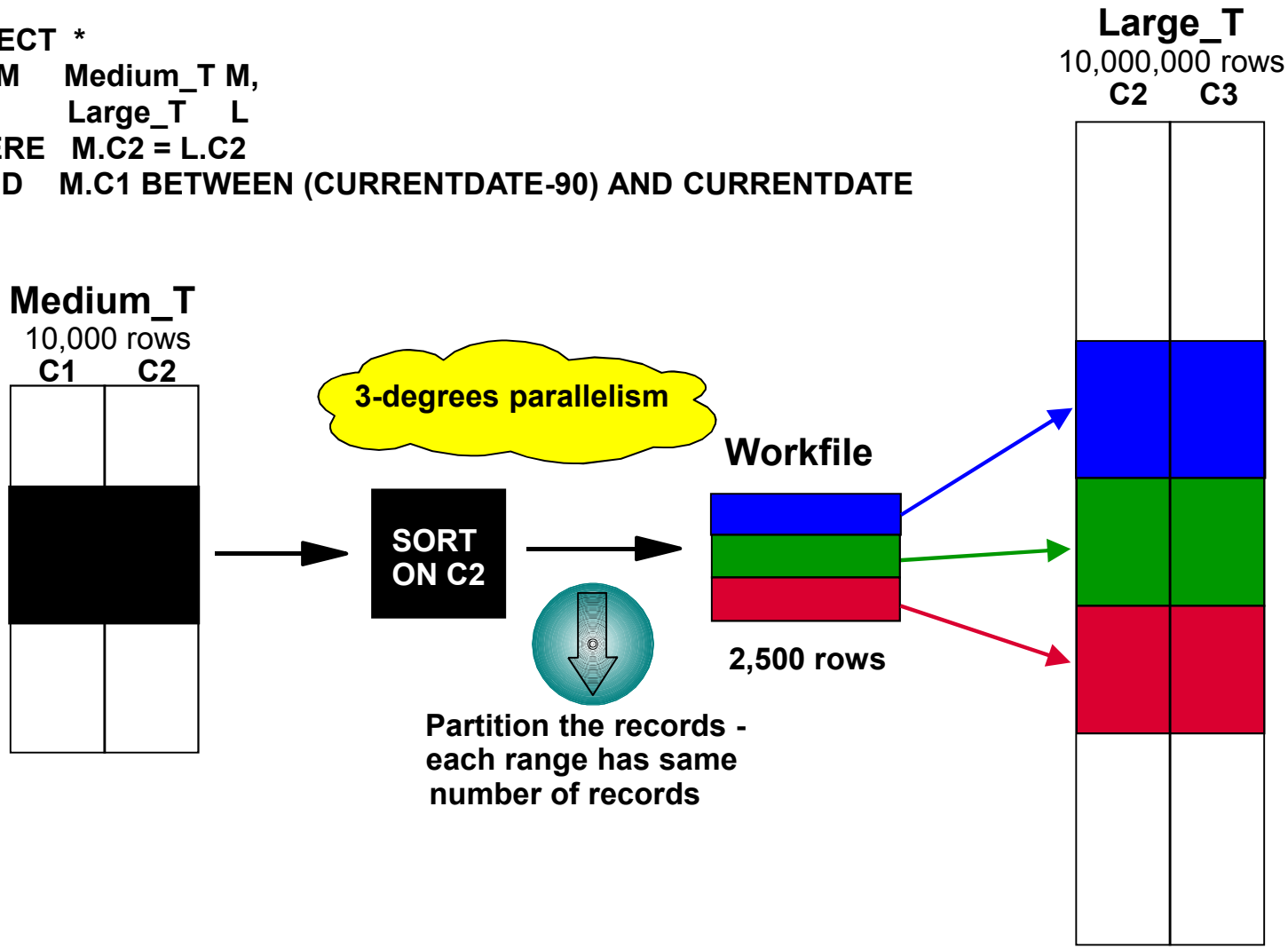
M.C1 is date column, assume currentdate is 8-31-2007, after the between predicate is applied, only rows with date between 06-03-2007 and 8-31-2007 survived, but optimizer chops up the key ranges within the whole year after the records are sorted :-(

Parallelism Effectiveness – Record range

- **DB2 10 can use** Dynamic record range partitioning
  - Materialize the intermediate result in a sequence of join processes
  - Results divided into ranges with equal number of records
  - Division doesn't have to be on the key boundary
    - Unless required for group by or distinct function
  - Record range partitioning is dynamic
    - no longer based on the key ranges decided at bind time
  - Now based on number of composite records and parallel degree
    - Data skew, out of date statistics etc. will not have any effect on performance

Dynamic record range partition

```
SELECT  *
FROM    Medium_T M,
        Large_T  L
WHERE   M.C2 = L.C2
  AND   M.C1 BETWEEN (CURRENTDATE-90) AND CURRENTDATE
```



Parallelism Effectiveness - Straw Model

- **Previous releases of DB2 divide the number of keys or pages by the number representing the parallel degree**
  - One task is allocated per degree of parallelism
  - The range is processed and the task ends
  - Tasks may take different times to process

- **DB2 10 can use the Straw Model workload distribution method**
  - More key or page ranges will be allocated than the number of parallel degrees
  - The same number of tasks as before are allocated (same as degree)
  - Once a task finishes it's smaller range it will process another range
  - Even if data is skewed this new process should make processing faster