DB2 Statistics

Craig Friske

DB2 Utilities Development

friske@us.ibm.com
Disclaimer

© Copyright IBM Corporation 2015. All rights reserved.

IBM’s statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM’s sole discretion.

Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including 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 results similar to those stated here.
Agenda

- DB2 RUNSTATS Statistics Concepts
- RUNSTATS Recommendations
- DB2 Real Time Statistics
- REORG, COPY, and RUNSTATS Scheduling
DB2 RUNSTATS Statistics Concepts
Useful Statistics Concepts

• Access Paths and Filter Factors
• High/Low bound, cardinalities
• Distribution Statistics
• Histogram Statistics
• Clustering, Cluster Count, Off Position
• LEAFNEAR/LEAFFAR for indexes
• Pointers to overflow rows (indirect references)
Access Paths Lite (very light)

• Performance involves Access Path and organization of data and keys. The Optimizer is cost based. Cost is minimized by decreasing I/Os or the number of rows accessed.

• The Optimizer decides things like whether to use an index for accessing data, whether an index alone can be used, or in which order tables are accessed when doing JOINs.

• Other statistics don’t affect access path selection, but they can be an indicator of performance degradation, and they may signal that action should be take for improved performance.
Simple Access Path Examples

• Indexes can be very useful and chosen, especially if the filter factor is good or clustering is good. Here are some examples for a table with NBA player info:
  
  • Example 1
    ```sql
    SELECT JERSEY_NUMBER FROM NBA_PLAYERS WHERE NAME='STEPHEN CURRY';
    ```
  
  • Example 2
    ```sql
    SELECT NAME FROM NBA_PLAYERS WHERE TEAM = 'WARRIORS';
    ```
  
  • Example 3
    ```sql
    SELECT NAME FROM NBA_PLAYERS WHERE YEARLY_COMPENSATION > $1M;
    ```
Tale of 3 access paths

• Filtering restricts access to a subset of the index entries or data rows
  – Can reduce index I/O
  – Generally results in reduction in data I/O
  – Can’t always filter

SELECT JERSEYNO WHERE NAME = name
SELECT NAME WHERE TEAM = team (cluster order by team)
SELECT NAME WHERE YEARLY > salary (no filtering)
Tale of 3 access paths

• Matching Index Probe
  – “Unique” index probe limited to 2 index pages and 1 data page
  – Access 1 data row

SELECT JERSEYNO WHERE NAME = name
Tale of 3 access paths

• Matching Index Scan
  – Index on Team with data ordered according to the index
  – Minimal leaf page and data pages access

SELECT NAME WHERE TEAM = team
(where cluster order is team)
Tale of 3 access paths

- If filtering cannot avoid any data pages
  - May as well to a table/table space scan

```sql
SELECT NAME WHERE YEARLY > salary (no filtering)
```
Oversimplified Optimizer Costing

- Optimizer assigns Filter Factors for each WHERE/ON predicate

- FFs are combined to determine the total filtering per object
  - Multiply “AND” predicate FFs
    - Available statistics determine “degree” of multiplication
  - Add “OR” predicate FFs

- FF accuracy and how to combine these is important for costing
  - Index matching
  - Total index filtering
  - Total table level filtering
Access Path Attributes

- Key attributes collected by RUNSTATS for use by the optimizer:
  - Size of the objects
    - NPAGESF, NLEAF, NLEVELS etc.
  - Range on records/keys
    - LOW2KEY, HIGH2KEY, LOWKEY, HIGHKEY
  - Selectivity or number of records/keys
    - CARDF, COLCARDF, FIRSTKEYCARDF, FULLKEYCARDF, FREQVAL etc.
  - Other important statistics
    - CLUSTERRATIOF, PCTROWCOMP etc.
Filter Factors and cardinality

SYSCOLDIST and SYSCOLDISSTATS contain frequency (or distribution) values.
If frequency stats do not exist, DB2 assumes that the data is uniformly distributed.
For example:

NBA players table (450 rows)

<table>
<thead>
<tr>
<th>Player</th>
<th>J#</th>
<th>Team</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stephen Curry</td>
<td>30</td>
<td>Warriors</td>
<td>CA</td>
</tr>
<tr>
<td>Lebron James</td>
<td>23</td>
<td>Cavaliers</td>
<td>OH</td>
</tr>
<tr>
<td>James Harden</td>
<td>13</td>
<td>Rockets</td>
<td>TX</td>
</tr>
<tr>
<td>Anthony Davis</td>
<td>23</td>
<td>Pelicans</td>
<td>LA</td>
</tr>
<tr>
<td>Chris Paul</td>
<td>3</td>
<td>Clippers</td>
<td>CA</td>
</tr>
<tr>
<td>Russell Westbrook</td>
<td>0</td>
<td>Thunder</td>
<td>OK</td>
</tr>
</tbody>
</table>

Cardinality  447  50  30  25
#Rows/Card    1.01  9  15  17 (est rows per value)

Complete your session evaluations online at www.SHARE.org/Orlando-Eval
Filter Factors and skewed data

Distribution statistics can help produce accurate filtering for skewed data if there are 30 teams and 66 champions, shouldn’t I expect each team to have 2 entries (3.3%)?

### NBA Historical Stats

<table>
<thead>
<tr>
<th>Year</th>
<th>Team</th>
<th>Wins</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>Warriors</td>
<td>67</td>
</tr>
<tr>
<td>2015</td>
<td>Cavaliers</td>
<td>53</td>
</tr>
<tr>
<td>1996</td>
<td>Bulls</td>
<td>72</td>
</tr>
<tr>
<td>1998</td>
<td>Nuggets</td>
<td>11</td>
</tr>
<tr>
<td>1973</td>
<td>Sixers</td>
<td>9</td>
</tr>
</tbody>
</table>

More…

### NBA Champs

<table>
<thead>
<tr>
<th>Year</th>
<th>Winner</th>
<th>Loser</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>Warriors</td>
<td>Cavaliers</td>
</tr>
<tr>
<td>2014</td>
<td>Spurs</td>
<td>Heat</td>
</tr>
<tr>
<td>1997</td>
<td>Bulls</td>
<td>Jazz</td>
</tr>
<tr>
<td>1987</td>
<td>Lakers</td>
<td>Celtics</td>
</tr>
<tr>
<td>1950</td>
<td>Lakers</td>
<td>Sixers</td>
</tr>
</tbody>
</table>

More…

#rows=1650?, Card=33 65? Highkey 72 Lowkey 9 High2key 69 Low2key 11

Distributions Stats on Winner

Value = ‘Celtics’ Count = 17 Freq=25%
Value = ‘Lakers’ Count = 16 Freq=24%
Value = ‘Bucks’ Count = 1 Freq=1.5%
Value = ‘Hornets’ ??

Complete your session evaluations online at www.SHARE.org/Orlando-Eval
Histogram Statistics

Think of filter factor stats on a range (quantile) of data (helpful for range predicates)

LOWVALUE, HIGHVALUE, CARDF, and FREQUENCYF

RUNSTATS TABLESPACE ts TABLE(tb) COLGROUP(C4)
  FREQVAL COUNT 20 MOST HISTOGRAM NUMQUANTILES 100
RUNSTATS INDEX(IX FREQVAL NUMCOLS 15 COUNT 10 MOST
  HISTOGRAM NUMQUANTILES 100)

Catalog Table:
Kept in SYSCOLDIST and SYCOLDISTSTATS

Complete your session evaluations online at www.SHARE.org/Orlando-Eval
CLUSTERING INDEX

An index that determines how rows are physically ordered (clustered) in a table space. If a clustering index on a partitioned table is a DPSI, the rows are ordered in cluster sequence within each data partition instead of spanning the partitions.

When data row obtained via index scan using “TEAM”. All rows are in optimal order (or clustering order, clustered) except when accessing row ‘L’akers in 2015.

```
SELECT SUM(YEARLY) FROM NBA_PLAYERS WHERE TEAM='WARRIORS' AND YEAR=2015;
```
CLUSTERRATIO

SYSIBM.SYSINDEXES.CLUSTERRATIO

• An access path statistic that can also help in determining when to REORG
• % of the rows that are in cluster order (adjusted formula) for clustering indexes (100% is ideal). There can only be 1 clustering index, but other indexes can be correlated with the clustering index.
• Rows are counted as being “clustered” if they are within the prefetch range for either a forward or backward reference.
• This is a statistic that describes the data in the table(space), even though it is reported in SYSINDEXES – REORG INDEX will never affect this statistic

Examples:  CREATE INDEX ICLUST on TABLE TB1(TEAM) CLUSTER;

CREATE INDEX ICLUST2 on TABLE TB1(TEAM, JERSEY#);

CREATE INDEX NOCLUST on TABLE TB1(TEAM) DESC;
LEAFNEAR/LEAFFAR

Measures the disorganization of physical leaf page. Pages are not in an optimal position due to index pages being deleted or index leaf page splits caused by an insert that cannot fit onto a full page. Affects performance during an index scan.

Logical and physical views of an index in which LEAFNEAR=1 and LEAFFAR=3

RUNSTATS: LEAFNEAR, LEAFFAR; RTS: LEAFNEAR, LEAFFAR
Action: REORG INDEX
(V10 list prefetch mitigates LEAFNEAR/FAR performance degradation)
INDEX key and data row referencing

RID “Part 33, Page 21, ID n-1”

Partition 33

Page ‘21’ Header

DB2 Control Area

Data Row 1

Data Row 2

Data Row 3

Data Rows

Data Row n-2

Data Row n-1

Data Row n

Free Space

DB2 Control Area

Table Space

Index

Key/rid
UPDATE to the record increases the row length, no room to fit (e.g. UPDATE ADDRESS = “…Oakland” WHERE NAME=‘KEVON LOONEY’)

Indirect reference occurred
NEARINDREF: <= “search interval” pages away
FARINDREF: > “search interval” pages away
Relief: REORG TABLESPACE or PCTFREE FOR UPDATE (in V11)
RUNSTATS stats gathering
Statistics gathered by RUNSTATS TABLESPACE

SYSIBM.SYSTABLESPACE
- NACTIVEF
- AVGROWLENG
- SPACEF

SYSIBM.SYSTABLES/HIST
- CARD/F
- NPAGESF
- PCTPAGES
- PCTROWCOMP
- AVGROWLENG
- SPACEF

SYSIBM.SYSTABSTATS
- CARD/F
- NPAGES
- PCTPAGES
- PCTROWCOMP

SYSIBM.SYSTABLEPART/HIST
- AVGROWLEN
- CARD/F
- DSNUM
- EXTENTS
- NEARINDREF
- FARINDREF
- PAGESAVE
- PERCACTIVE
- PERCDROP
- SPACE/F
- PQTY
- SQTY
- SECQTYI

SYSIBM.SYSCOLSTATS
- COLCARD
- HIGHKEY
- HIGH2KEY
- LOWKEY
- LOW2KEY
- COLCARDDATA

SYSIBM.SYSCOLDIST/HIST/STATS
- NUMCOLUMNS
- COLGROUPCOLNO
- COLVALUE
- CARDF
- TYPE (‘C’, ‘F’, ‘H’)
- FREQUENCY/F
- KEYCARDDATA
- QUANTILENO

- Covered in RUNSTATS concepts section previously
- Overlap with RUNSTATS INDEX

Complete your session evaluations online at www.SHARE.org/Orlando-Eval
# Statistics gathered by RUNSTATS INDEX

<table>
<thead>
<tr>
<th>SYSIBM.SYSINDEXES/HIST</th>
<th>SYSIBM.SYSINDEXEXPART/HIST</th>
<th>SYSIBM.SYSCOLSTATS</th>
<th>SYSIBM.SYSCOLDIST/HIST/STATS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLUSTERRATIO/F</td>
<td>AVGKEYLEN</td>
<td>COLCARD</td>
<td>NUMCOLUMNS</td>
</tr>
<tr>
<td>CLUSTERED</td>
<td>CARDF</td>
<td>HIGHKEY</td>
<td>COLGROUPCOLNO</td>
</tr>
<tr>
<td>FIRSTKEYCARD/F</td>
<td>DSNUM</td>
<td>HIGH2KEY</td>
<td>COLVALUE</td>
</tr>
<tr>
<td>FULLKEYCARD/F</td>
<td>EXTENTS</td>
<td>LOWKEY</td>
<td>CARDF</td>
</tr>
<tr>
<td>NLEAF</td>
<td>FAROFFPOSF</td>
<td>LOW2KEY</td>
<td>TYPE ('C','F','H')</td>
</tr>
<tr>
<td>NLEVELS</td>
<td>LEAFNEAR</td>
<td>COLCARDDATA</td>
<td>FREQUENCY/F</td>
</tr>
<tr>
<td>AVGKEYLEN</td>
<td>LEAFFAR</td>
<td></td>
<td>KEYCARDDATA</td>
</tr>
<tr>
<td>SPACEF</td>
<td>NEAROFFPOS</td>
<td></td>
<td>QUANTILENO</td>
</tr>
<tr>
<td></td>
<td>LEAFDIST</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PSUEDO_DEL_ENTRIES</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SPACEF</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PQTY</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SECQTYI</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Covered in RUNSTATS concepts section previously
- Overlap with RUNSTATS TABLESPACE

Complete your session evaluations online at www.SHARE.org/Orlando-Eval
RUNSTATS Statistics Collection

- Basic statistics foundation
  ```sql
  RUNSTATS TABLE (ALL) TABLESAMPLE SYSTEM AUTO
  INDEX (ALL) KEYCARD
  SHRLEVEL CHANGE
  ```

- Supplement with more detailed statistics as needed
  - Distribution statistics
    - Frequencies
    - Histograms
  - Multi-column cardinality statistics

- Tradeoff with CPU savings and HISTOGRAM accuracy
How do I integrate supplemental statistics?

```
RUNSTATS LIST mylist
   TABLE (ALL) TABLESAMPLE SYSTEM AUTO
   INDEX (ALL) KEYCARD

RUNSTATS mydb.myts
   TABLE (NBA_PLAYERS)
   COLGROUP (TEAM, JERSEY#)
```

- Tradeoff with CPU savings and HISTOGRAM accuracy
Mixing Regular and “special” RUNSTATS

- If I run the following

1. RUNSTATS TABLE (ALL) TABLESAMPLE SYSTEM AUTO INDEX (ALL) KEYCARD
2. RUNSTATS TABLE (NBAChamps) COLGROUP (WINNER) FREQVAL COUNT 20
3. RUNSTATS TABLE (ALL) TABLESAMPLE SYSTEM AUTO INDEX (ALL) KEYCARD

- Won’t “Regular” RUNSTATS overwrite the “special”?
  - NO: RUNSTATS will only overwrite similar statistics
    - COLGROUP (STATUS) FREQVAL COUNT 20 is only overwritten if default statistics are collecting FREQVAL on this column
    - Is there an index leading with STATUS? Default is to collect top 10 (not top 20).

- Default since V10
DB2 V10 Simplifies Integration of Supplemental Stats

- Integrate specialized statistics into generic RUNSTATS job
  - RUNSTATS TABLE (mytb) COLGROUP(STATUS)… SET PROFILE
    - Or SET PROFILE FROM EXISTING STATS
  - RUNSTATS … TABLE (mytb) …. UPDATE PROFILE

- Next usage
  - RUNSTATS LIST mylist TABLE(ALL) USE PROFILE
    - RUNSTATS will execute as if all saved options were specified

- Caveats
  - Cannot specify USE PROFILE for a table without a defined profile (no defaults)
    - Restricts LISTDEF support
  - USE PROFILE not supported with inline stats
V11 Improvements

- Improved PROFILE usability with LISTDEF support
  - Gather default statistics if no profile exists for table
- More zIIP offload for RUNSTATS distribution statistics
  - Up to 80% zIIP-eligible
- Inline statistics RUNSTATS equivalence (avoid RUNSTATS)
  - Inline statistics collection on NPSIs during REORG with SORTNPSI
  - Inline histogram statistics
  - Inline DSTATS
  - zIIP offload up to an additional 30%
  - Still missing PROFILE support
- RUNSTATS RESET option deletes/clears all catalog stats for an object
DB2 V11 Optimizer externalization of missing stats

- During access path calculation, optimizer will identify missing or conflicting statistics
  - On every BIND, REBIND or PREPARE
    - Asynchronously writes recommendations to SYSIBM.SYSSTATFEEDBACK
      - From DB2 11 NFM
  - DB2 also provides statistics recommendations on EXPLAIN
    - Populates DSN_STAT_FEEDBACK synchronously
      - Beginning in DB2 11 CM – provided explain table exists

- Contents of SYSSTATFEEDBACK or DSN_STAT_FEEDBACK can be used to generate input to RUNSTATS
  - Contents not directly consumable by RUNSTATS
Object, Type, and Reason for statistics recommendations

- Object is identified as table, index, or column
- TYPE specifies the statistics to collect

<table>
<thead>
<tr>
<th>TYPE</th>
<th>CHAR(1)</th>
<th>The type of statistic to collect:</th>
</tr>
</thead>
<tbody>
<tr>
<td>'C'</td>
<td>Cardinality.</td>
<td></td>
</tr>
<tr>
<td>'F'</td>
<td>Frequency.</td>
<td></td>
</tr>
<tr>
<td>'H'</td>
<td>Histogram.</td>
<td></td>
</tr>
<tr>
<td>'T'</td>
<td>Index.</td>
<td></td>
</tr>
<tr>
<td>'T'</td>
<td>Table.</td>
<td></td>
</tr>
</tbody>
</table>

- REASON identifies why statistics were recommended
Real Time Statistics - RTS
Real-Time Statistics (RTS) Objective

- Older DBA procedures and some home-grown monitor tools had no accurate data to identify objects that need maintenance.
- Spending time performing maintenance on static and unchanged objects inefficiently uses DBA's time, wastes batch window time, and CPU.
  - The “best” utility is the one not needed.
- DB2 systems becoming larger and more complex.
  - A single DB2 for z/OS may have large amounts of tables/indexes. For ERP-packaged applications, it can be 80K+.
  - Requires skilled DBAs (and lots of time) to identify unused / static objects.
- Goal is to self-managed or automate the maintenance process.
- DB2 Stored Procedures, DSNACCOX, IBM DB2 Automation Tool and the new DB2 Management Console exploit RTS.
 RTS Overview

• Runs in the background - automatically updates statistics, as the data rows and indexes for DB2 table spaces are modified
• RTS manager runs under a system task in DBM1 address space
  – CPU time is included in DBM1's SRB time
  – The system task is created during START DB2
• Statistics collected in memory, and periodically externalized
  – ACCESS DATABASE … MODE(STATS)
• Contains space and as well as some access path statistics
• Externalized into DB2 Catalog – SYSIBM.DSNRTSTS:
  SYSIBM.SYSTABLESPACESTATS
  SYSIBM.SYSINDEXSPACESTATS
• Helps eliminate scheduling RUNSTATS (but can’t replace RUNSTATS)
## V11 RTS Tables – SYSTABLESPACESTATS

<table>
<thead>
<tr>
<th>Global Statistics</th>
<th>Incremental Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>REORG-</strong></td>
</tr>
<tr>
<td>DBID</td>
<td>LASTTIME</td>
</tr>
<tr>
<td>PSID</td>
<td>INSERTS</td>
</tr>
<tr>
<td>PARTITION</td>
<td>UPDATES</td>
</tr>
<tr>
<td>INSTANCE</td>
<td>DELETES</td>
</tr>
<tr>
<td>DBNAME</td>
<td>DISORGLOB</td>
</tr>
<tr>
<td>NAME</td>
<td>UNCLUSTINS</td>
</tr>
<tr>
<td>NACTIVE</td>
<td>MASSDELETE</td>
</tr>
<tr>
<td>NPAGES</td>
<td>NEARINDREF</td>
</tr>
<tr>
<td>EXTENTS</td>
<td>FARINDREF</td>
</tr>
<tr>
<td>SPACE</td>
<td>CLUSTERSENS</td>
</tr>
<tr>
<td>TOTALROWS</td>
<td>SCANACCESS</td>
</tr>
<tr>
<td>DATASIZE</td>
<td>HASHACCESS</td>
</tr>
<tr>
<td>UNCOMPRESSEDDATASIZE</td>
<td></td>
</tr>
<tr>
<td>UPDATESTATSTIME</td>
<td></td>
</tr>
<tr>
<td>HASHLASTUSED</td>
<td></td>
</tr>
<tr>
<td>DRIVER TYPE</td>
<td></td>
</tr>
<tr>
<td>LPFFACILITY</td>
<td></td>
</tr>
<tr>
<td>UPDATESIZE</td>
<td></td>
</tr>
<tr>
<td>LASTDATACHANGE</td>
<td></td>
</tr>
<tr>
<td>GETPAGES</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LOAD-</strong></td>
<td></td>
</tr>
<tr>
<td>RLASTTIME</td>
<td></td>
</tr>
</tbody>
</table>

- New in V10
- New in V11

Complete your session evaluations online at www.SHARE.org/Orlando-Eval
<table>
<thead>
<tr>
<th>Global Statistics</th>
<th>Incremental Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>REORG</strong>-</td>
<td><strong>COPY</strong>-</td>
</tr>
<tr>
<td>DBID</td>
<td>REBUILDLASTTIME</td>
</tr>
<tr>
<td>ISOBID</td>
<td>LASTTIME</td>
</tr>
<tr>
<td>PSID</td>
<td>INSERTS</td>
</tr>
<tr>
<td>PARTITION</td>
<td>UPDATES</td>
</tr>
<tr>
<td>INSTANCE</td>
<td>DELETES</td>
</tr>
<tr>
<td>NACTIVE</td>
<td>APPENDINSERT</td>
</tr>
<tr>
<td>NLEVELS</td>
<td>PSEUDODELETES</td>
</tr>
<tr>
<td>NPAGES</td>
<td>MASSDELETE</td>
</tr>
<tr>
<td>NLEAF</td>
<td>LEAFNEAR</td>
</tr>
<tr>
<td>EXTENTS</td>
<td>LEAFFAR</td>
</tr>
<tr>
<td>SPACE</td>
<td>NUMLEVEL</td>
</tr>
<tr>
<td>TOTALENTRIES</td>
<td>INDEXACCESS</td>
</tr>
<tr>
<td>LASTUSED</td>
<td></td>
</tr>
<tr>
<td>UPDATESTATSTIME</td>
<td></td>
</tr>
<tr>
<td>DBNAME</td>
<td></td>
</tr>
<tr>
<td>NAME</td>
<td></td>
</tr>
<tr>
<td>INDEXSPACE</td>
<td>LOAD-</td>
</tr>
<tr>
<td>DRIVETYPE</td>
<td>RLASTTIME</td>
</tr>
<tr>
<td>GETPAGES</td>
<td></td>
</tr>
</tbody>
</table>
RTS Usage – History and trending

• There is currently no historical capability in RTS in DB2 itself
• Create a history table manually
  
  ```sql
  CREATE
  SYSIBM.(TABLE | INDEX)SPACESTATS_HIST
  LIKE
  SYSIBM.SYS(TABLE | INDEX)SPACESTATS
  
  then add
  
  CAPTURE_TIME AS TIMESTAMP NOT NULL WITH
  DEFAULT column
  ```

• Periodically insert into RTS history tables with a sub select from the RTS tables those rows that aren’t already in the history tables; and delete old information.
  Some customers do this weekly, others monthly – depending on needs
Object activity

- How active are my DB2 objects?
- What activity has taken place for a specific time for TS’ and IX’s
- Use `UPDATESTATSTIME`

```sql
SELECT DBNAME, NAME, PARTITION, UPDATESTATSTIME
FROM SYSIBM.TABLESPACESTATS
WHERE (JULIAN_DAY(CURRENT DATE) – JULIAN_DAY(UPDATESTATSTIME)) <= 14
   AND NAME = xxx;
```

- Use DB2 Administration Tool – DB2 Performance Queries
Unused or (in)activity of INDEXES

–LASTUSED column in SYSINDEXSPACESTATS
  • Is a **date** field
  • Consider using for identifying which IXs to drop
  • The date indicates the index is last used for SELECT, FETCH, searched UPDATE, searched DELETE, or used to enforce referential integrity constraints.
  • The default value is 01/01/0001.

–REORGINDEXACCESS column in SYSINDEXSPACESTATS
  • # of times the IX was **accessed** (read and updates) since last reorg or since creation
  • NULL denotes never used
RTS Usage – Track Utility execution

- When was the last time a utility was run against my objects?
- When was COPY, REORG, LOAD REPLACE, and RUNSTATS last executed against objects ..

```sql
SELECT DBNAME, NAME, PARTITION, TOTALROWS, NACTIVE, SPACE, EXTENTS, UPDATESTATSTIME, STATS_LASTTIME, LOADR_LASTTIME, REORG_LASTTIME, COPY_LASTTIME
FROM SYSIBM.TABLESPACESTATS
ORDER BY DBNAME, NAME, PARTITION
```

- Or use DB2 Administration Tool for reporting
- For object maintenance queries/info consider DB2 Automation Tool, free stored procedure DSNACCOX and DB2 Management Console
Object Maintenance (aka utilities scheduling)
What is DSNACCOX?

A DB2 stored procedure that accesses the RTS tables and looks at DBET states to give recommendations for when schedule table spaces or indexes maintenance for reorganization, taking image copies, or updating statistics:

- REORG TABLESPACE, REORG INDEX
- RUNSTATS TABLESPACE, RUNSTATS INDEX
- COPY TABLESPACE, COPY INDEX
Reorg table space recommendations

- #Inserts since last REORG > 25% of total rows and #Inserts > 0
- #Deletes since last REORG > 25% of total rows and #Deletes > 0
- #Cluster Accesses since last REORG > 0 and #Unclustered Inserts since last REORG > 10%
- #Overflow Rows since last REORG > 10%
- #Mass deletes since last REORG > 0
- #Extents > 254
- #Disorganized LOBS > 50%
- #Hash Index Entries since last REORG > 15%
- Object is in REORG Pending (Alter Limit Key, Add Identity Column blocks access)
- Object in Advisory REORG Pending AREOR (Pending Alter materialization)
- Object in Advisory REORG AREO* (Immediate Alter materialization)
Reorg index recommendations

- #Inserts since last REORG > 30% and #Inserts > 0
- #Deletes since last REORG > 30% and #Deletes > 0
- #Inserts appended since last REORG > 20%
- #Pseudodeletes since last REORG > 10%
- #Mass Deletes since last REORG > 0
- #LEAFFAR since last REORG > 10%
- #Levels > 0
- #Extents > 254
- #Extra formatted pages > 10%
- Object in Advisory REORG Pending AREOR (Pending Alter materialization)
Copy scheduling recommendations

Full Image Copy on a Table Space
• Table space has never had a full image copy
• Last image copy is older than 7 days
• #Updated pages since the last copy > 10% and # pages changed > 0
• The object is in Copy Pending

Incremental Copy on a Table Space
• Table space has never had an incremental image copy
• Last image copy is older than 7 days
• #Updated pages since the last copy > 1% of the total pages, and #updated pages > 0
• #Updated rows since the last copy > 1% of the total rows.
RUNSTATS scheduling recommendations

RUNSTATS on a Table Space
- If RUNSTATS has never been run
- #Inserts, #Deletes, and #Updates > 20%, and #changes > 0
- #Mass Deletes > 0

RUNSTATS on an INDEX
- If RUNSTATS has never been run
- #Inserts and #Deletes > 20%, and #changes > 0
- #Mass Deletes > 0

Remember, RUNSTATS followed by REBIND may alter the access path. Consider using plan stability (e.g. PLANMGMNT and REBIND SWITCH) to avoid surprises.
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.

Special thanks to Bryan Smith, Terry Purcell, Henni Mynhardt for their help.

© Copyright IBM Corporation 2013. All rights reserved.

- U.S. Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

IBM, the IBM logo, ibm.com, DB2, and z/OS 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](http://www.ibm.com/legal/copytrade.shtml)

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

Complete your session evaluations online at [www.SHARE.org/Orlando-Eval](http://www.SHARE.org/Orlando-Eval)