
CDB/Runstats

Introduction

Application performance and user response time are greatly affected by the runstats values used by the DB2 optimizer. The values gathered by the standard IBM Runstats utility are not sufficient in most cases, particularly in ad hoc queries common in web-based applications.

Function

DSTATS is an IBM-supplied program used to gather additional statistics relating to the skew of the data within the tablespace. According to the IDUG Solutions Journal (Volume 6, Number 1), “ By providing the optimizer with more information, we can enable it to choose more efficient access paths, more consistently, for a wider range of queries, leading to improved response times and lower resource usage.” CDB/Runstats gathers all of the statistics gathered by DSTATS that can be gathered from the index, but in a fraction of the time and with added function.

Speed

DSTATS scans the user tablespace once per column on which it is gathering statistics. CDB/Runstats gathers all statistics in one pass of the indexes of the tablespace. If distribution statistics are being gathered for 5 columns, DSTATS will scan the tablespace 5 times while CDB/Runstats will scan the indexes of the tablespace once. A huge reduction in elapsed time will be seen using CDB/Runstats, particularly when running on tablespaces with millions of rows.

Wildcards

CDB/Runstats provides the option of using wildcards in object names. Special symbols ‘_’ and ‘%’ are wildcards with special meanings that match those used by DB2: “%” - Replace with any string, and “_” - Replace with any character. This allows you to easily gather statistics for an entire application or system with one simple control card.

No-Sort Technique

CDB/Runstats uses a proprietary sorting mechanism to gather all statistics in one pass of the data. All of the problems traditionally associated with Sort are not a problem with CDB/Runstats.

Gathering Runstats

CDB/Runstats uses a DB2 table to control statistics-gathering options. IBM uses a lengthy and clumsy syntax that must be specified for each object, including each Reorg or Load. Instead, with CDB/Runstats, the statistics options are kept in a single place, in DB2 in a table that can be easily modified using standard SQL.

Statistics options include such variables as:

- Whether or not statistics should be gathered for this table,
- The NUMCOLS value for each index,
- The number of distribution values to be gathered by CDB/Runstats on each column of each index. (CDB/Auto-Reorg, CDB/Restore and CDB/Load, when gathering index statistics, report column distributions only on the first column of each index.)

CDB/Runstats Benchmarks

