IBM DATABASE MAGAZINE » Powering business information

Quarter 1, 2004 Vol. 9, Issue 1

http://www.dbmag.intelligententerprise.com/story/showArticle.jhtml?articleID=17602324

Turbo-Charged Data Warehousing

By David Beulke

IDUG Insight

The improvements in DB2 Universal Database (UDB) for z/OS version 8 have enhanced DB2's data warehousing capabilities on the z/OS platform. Like a high-performance car, DB2 has been turbo-charged with SQL enhancements, new table-design options, and new indexing choices. These features — combined with star join improvements, a dedicated workfile virtual memory pool, and other changes — enhance the appeal of DB2 for large-scale data warehousing and business intelligence projects.

There are more SQL enhancements in v.8 than can be highlighted in this brief column. For more details regarding the SQL changes, see Version 1.1 of *The SQL Reference for Cross Platform Development* and the *What's New?* publication for version 8 (both are listed in the resources at the end of this column). I'll limit my discussion to the ways IBM is leveraging and integrating its DB2 Linux, Unix, and Windows SQL functionality in the new z/OS release.

Two new features — first introduced on Linux, Unix, and Windows platforms — that will likely have a tremendous performance impact on mainframe data warehousing environments are materialized query tables (MQTs) and common table expressions (CTEs).

MQTs are new table types that can automatically precalculate or summarize data from other detailed data or fact tables. The DB2 v.8 optimizer can dynamically determine whether to rewrite a submitted SQL query to access the MQT or the detailed fact table. MQTs boost data warehouse performance by precalculating, aggregating, or summarizing data once instead of each time a frequently repeated SQL query is executed. MQT options provide flexibility regarding when data should be refreshed and materialized, giving designers and administrators complete control over the MQTs in their environments.

CTEs, another new feature, can be combined with recursive SQL (in other words, SQL that references itself) to provide a powerful data warehousing design solution for extracting and working on a distinct set of information. CTEs provide a way to extract a

result set from the database based on desired criteria and then to extract more tables from the data. This unique result set can be referenced in the SQL, further refining the answer for the user. This new feature is great for situations when you need to retrieve data from data already retrieved in an SQL statement. CTEs avoid the catalog overhead of views, provide the ability to use host variables, and avoid data changes from other `INSERT`, `UPDATE`, or `DELETE` SQL operations. With recursive SQL, you can apply the power of SQL repeatedly to distinct result sets to quickly and efficiently derive the answers.

In a bill of material application, a CTE would be ideal for getting part details from the result set of part information already extracted. The CTE avoids going back to DB2 for additional information, while allowing the application to extract only specific parts, based on a structure denoted by the program's host-variable values. Combining these techniques gives data warehouse users answers to their unique criteria while improving performance, avoiding conflicts with other users, maintaining data security, and easily leveraging SQL result set data.

New expansion and rotation capabilities in v.8 table partitioning provide powerful data warehouse design alternatives. V.8 offers an expanded maximum number of partitions (4,096 — up from 254). The capability to define a separate table partition for every day for more than 10 years is unique to DB2 and provides remarkable design flexibility for data warehouses. The ability to rotate or create additional partitions on already implemented partitioned table designs also enhances existing partitioned databases. This ability allows existing databases to move data around quickly and easily, which can be a great help in disaster recovery, expanding the database for changing business conditions, and getting rid of old data quickly.

Resources

What's New? Version 8

The SQL Reference for Cross-Platform Development, Version 1.1

Star join features have been improved in several ways. The most important features are the new sparse indexing capability, enhancements to the DB2 optimizer materialization, and a dedicated virtual memory pool. The sparse index capability reduces I/O associated with unqualified rows that might have been considered and then eliminated through a large result set. The new capability builds an index of only qualified rows. So unqualified data is eliminated earlier in the access path, which improves query response time and CPU performance. Improvements in the DB2 optimizer algorithms determine the cost effectiveness of materializing the snowflake or raw data warehouse dimensions. These improvements can have a dramatic performance impact on complex front-end OLAP tool queries for data warehousing.

V.8 also introduces the ability to create a dedicated virtual pool for star join workfiles that are usually created for materialized dimensions or composite data. When a dedicated data warehousing workfile virtual pool exists, repeated scanning or access to this work data is done in memory, improving performance for sort operations and queries. The star join workfile pool exists in addition to other buffer pools and can also help parallel workfile operations.

With all the changes in v.8, the International DB2 Users Group (IDUG) North America conference (Orlando, Fla., May 9-13) should be an educational priority. Getting the latest technical information from IBM developers and actual users who participated in the IBM Quality Partnership Program (QPP) will help you leverage this state-of-the-art technology. IDUG has also published v.8 white papers for z/OS (www.idug.org/idug/member/journal/may03/article02.cfm) and Linux, Unix, and Windows (www.idug.org/idug/member/journal/aug02/unveiling8.0.cfm) that detail many more v.8 features.

DB2 UDB for z/OS v.8 is a tremendous release that extends IBM's reputation as the leader in cutting-edge DBMS technology.

David Beulke specializes in design reviews of high-performance systems, data warehouses, and Internet architectures. He is a certified DBA for z/OS, Linux, Unix, and Windows platforms, a certified Business Intelligence Solutions Expert, an IBM Gold Consultant, and the president of IDUG. You can reach him through Pragmatic Solutions Inc. at 703-798-3283 or DBeulke@compuserve.com.

Return to Article