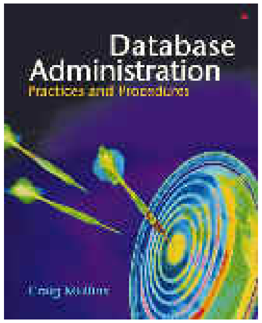
Craig S. Mullins

Return to Home Page

November 2004

DATABASE
TRENDS AND APPLICATIONS



The DBA Corner

by Craig S. Mullins

Ensuring Data Integrity is a Tricky Business

The term "data integrity" can mean different things. At the top level, there are two aspects of integrity with respect to databases: database structure integrity and semantic data integrity. Keeping track of database objects and ensuring that each object is created, formatted and maintained properly is the goal of database structure integrity. Each DBMS uses its own internal format and structure to support the databases, table spaces, tables, and indexes under its control. System and application errors at times can cause faults within these internal structures. The DBA must identify and correct such faults before insurmountable problems occur.

Semantic data integrity refers to the meaning of data and relationships that need to be maintained between different types of data. The DBMS provides options, controls and procedures to define and assure the semantic integrity of the data stored within its databases. Structural database integrity and consistency is critical in the ongoing administration of databases. If the structural integrity of the database is not sound, everything else will be suspect, too. There are multiple types of structural problems that can occur. Indexing problems are one. Certain types of database maintenance can cause such problems and DBAs need to be able to recognize the problem, and rebuild the indexes to correct their structural integrity. Indexes are not the only database objects that utilize pointers. Many DBMSs use pointers to store very large objects containing text and image data. These can become corrupted.

The more difficult and more pervasive problem is the semantic integrity of the data. Getting that right requires proper design, processes that match your business requirements, good communication skills, and constant vigilance.

Perhaps the number one cause of data integrity problems is improperly designed databases. Just getting the data type and length correct for each column can go a long way to making sure the right data is stored. Think about it. If you need to store dates but the column is defined as CHAR(8) how can you enforce that only valid dates are stored? You need to code some program logic to accomplish that. If the column were defined as DATE then the DBMS would take care of it -- and more of the data would be likely to be correct. And the DBA must also set up data relationships properly in the database. This is done using referential integrity (RI), a method for ensuring the "correctness" of data within a DBMS. People tend to over-simplify RI stating that it is merely the identification of relationships between relational tables. It is actually much more than this. Of course, the identification of the primary and foreign keys that constitutes a relationship between tables is a component of defining referential integrity. Basically, RI guarantees that an acceptable value is always in the foreign key column. Acceptable is defined in terms of an appropriate value as housed in the corresponding primary key (or perhaps null).

The combination of the relationship and the rules attached to that relationship is referred to as a referential constraint. The rules that accompany the RI definition are just as important as the relationship. These rules define how data is to be properly added to the databases and what happens when it is removed. There are other mechanisms in the DBMS that DBAs can use to enforce semantic data integrity. Check constraints and rules can be applied to columns that dictate valid values. The DBMS will reject invalid data that does not conform to the constraints. More complex data relationships can be set up using database triggers.

Every DBA should take advantage of the mechanisms provided by the DBMS to ensure data integrity. When DBMS-provided methods are used, fewer data integrity problems are likely to be found. Fewer data integrity problems mean higher quality databases and more proficient end users. You have to know what integrity rules are proper for the DBMS to enforce. But once defined, many of those rules can be enforced by the DBMS. And that is good!

From Database Trends and Applications, November 2004.

Home.