**Craig S. Mullins**
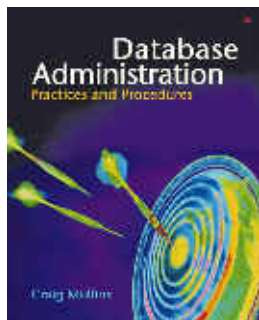
October 2006

## The DBA Corner
*by Craig S. Mullins*

### Databases Must be Archived for Long-Term Data Retention

The growing number of regulations, and the need for organizations to be in compliance with them, is driving data retention requirements to expand over longer and longer periods of time. Regulations such as the Sarbanes-Oxley Act and HIPAA govern how long data must be retained. Indeed, industry analysts have estimated that there are over 150 federal and state

laws that dictate how long data must be retained. As such, organizations are beginning to realize that they have to formulate plans for archiving data from the operational databases.

Database Archiving is part of the larger topic of Data Archiving. Data exists in many formats and for many purposes; only a small percentage of it is actually in a database. Physical documents, electronic documents, computer files and data sets, e-mail, and multimedia files are all examples of data that may reasonably need to be archived at some point. Each of these "things" needs to be archived to fulfill regulatory, legal, and business requirements.

But each type of data requires different archival processing requirements due to its form and nature. What works to archive e-mail is not sufficient for archiving database data, and so on. In other words, each type of data probably requires its own archive technology. This is most certainly true for database data. Why?

Well, data stored in a database is different than other types of data in many ways. The main advantage of using a DBMS is to impose a logical, structured organization on the data. A DBMS provides a layer of independence between the data and the applications that use the data. In other words, applications are insulated from how data is structured and stored. The interface to the data is through the DBMS data language, whether it is SQL for relational databases, DL/1 for IMS, or even XQuery for XML databases. So the archival of data from a database requires

knowledge of, and operation in conjunction with, the mechanisms and interfaces of the DBMS.

OK, if we now accept that database archiving is a subset of data archiving, let's create a clear definition of the term. **Database Archiving** is the process of removing selected data records from operational databases that are not expected to be referenced again and storing them in an archive data store where they can be retrieved if needed.

Let's examine each of the major components of that last sentence. We say **removing** because the data is deleted from the operational database when it is moved to the data archive. If the data is still required for operational requirements it is not ready to be archived. When data moves into the archive state, query and access is no longer anticipated to be required, so it can be removed from the operational database.

Next, we say **selected records**. This is important because we do not want to archive database data at the file level. We need only those specific pieces of data that are no longer needed for operational and reference purposes by the business. This means that the archive needs to be able to selectively choose particular pieces of related data for archival... not the whole database, not an entire table or segment, and not even a specific row. Instead, all of the data that represents a business object is archived at the same time. For example, if we choose to archive order data, we would also want to archive the specifics about each item on that order. This data

likely spans multiple constructs within the database (tables for DB2 or Oracle; segments and/or databases for IMS).

The next interesting piece of the definition is this: ***and storing them (the data) in an archive data store***. This implies that the data is stored separately from the operational database and does not require either the DBMS or the operational applications any longer. Archived data is separate and independent from the production systems from which it was moved.

The final component of the definition that warrants clarification is... ***where they can be retrieved if needed***. The whole purpose of archiving is to maintain the data in case it is required for some purpose. The purpose may be external, in the form of a lawsuit or to support a governmental regulation; or the purpose may be internal, in the form of a new business practice or requirement. At any rate, the data needs to be readily accessible in a reasonable timeframe without requiring a lot of manual manipulation. I mean, let's face it, anyone can archive data if they don't have to worry about how to meaningfully access it later.

So, what do you think? Do you have the technology and resources at your disposal to archive your database data in accordance with legal requirements?

From Database Trends and Applications, October 2006.

Home.