



Craig S. Mullins

[Return to Home Page](#)

December 2006 / January 2007



zData Perspectives

by Craig S. Mullins

[Database Archiving for Long-term Data Retention](#)

The growing number of regulations, and the need for companies to be in compliance with them, is driving the need to retain data for longer durations. There are many regulations (such as the Sarbanes-Oxley Act and HIPAA) that govern how long data must be retained. Indeed, one analyst group estimates that there are over 150 federal, state and local laws that dictate how long data must be retained. As such, businesses of every size have begun to realize that they must formulate plans for archiving data from their operational databases.

Database archiving is component of the larger discipline of data archiving. Data exists in many formats and for many purposes, and only a small percentage of it is actually stored in databases. Yet even this small proportion comprises quite a lot of data.

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 may need to command its own 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, we need to define exactly what we mean by 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. Consider the lifecycle of a "piece" of data: upon creation the data is in an operational state. After a period of time the data is still needed for reference, but not to drive operational transactions. After another period of time it is no longer needed for reference, but is needed for legal purpose: this is the archive state. So when data is archived 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 or table 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 we must 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, as well as CUSTOMER and PRODUCT data. This data likely spans multiple constructs within the database.

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 to be useful. Archived data is separate and independent from any and all 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 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 query it later, right?

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

From [zJournal](#), Dec 2006 / Jan 2007

© 2007 Craig S. Mullins, All rights reserved.

[Home](#).