



Craig S. Mullins

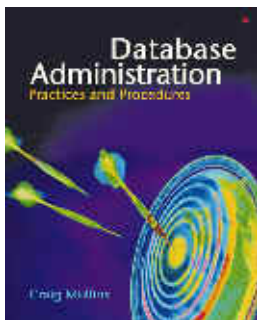
November 2006

[Return to Home Page](#)



The DBA Corner

by Craig S. Mullins



The Lifecycle of Data

Last month I discussed database archiving, 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. This month, I want to elaborate a bit more on the lifecycle of data – in other words – the **when** aspect of archiving.

The accompanying diagram helps to demonstrate this lifecycle. Basically, there are three major stages of “life” for any piece of data.

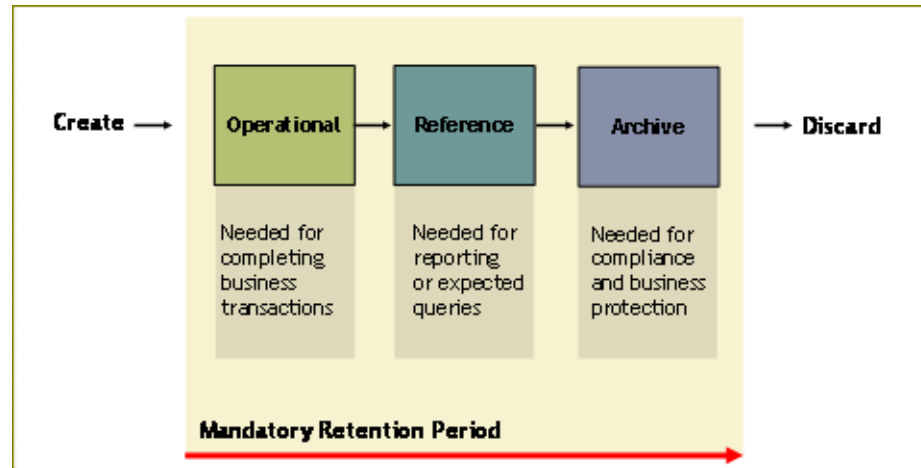


Figure 1. The Lifecycle of Data

Data is created at some point, usually by means of a transaction: a product is released, an order is processed, a deposit is made, etc. For a period of time after creation, the data enters its first state: it is ***operational***. That is, the data is needed to complete on-going business transactions. This is where it serves its primary business purpose. Transactions are enacted upon data in this state.

The operational state is followed by the ***reference*** state. This is the time during which the data is still needed for reporting and query purposes, but it is not necessarily driving business transactions. The data may be needed to produce internal reports, external statements, or simply exist in case a customer asks for it.

Then, after some additional period of time, the data moves into an area where it is no longer needed for completing business transactions and the chance of it being needed for querying and reporting is small to none. However, the data still needs to be saved for regulatory compliance and other legal purposes, particularly if it pertains to a financial transaction. This is the ***archive*** state.

Finally, after a designated period of time in the archive, the data is no longer needed at all and it can be discarded. This actually should be emphasized much stronger: the data **must** be discarded. In most cases the only reason older data is being kept at all is to comply with regulations, many of which help to enable lawsuits. When there is no legal requirement to maintain such data, it is only right and proper for organizations to demand that it be destroyed – why enable anyone to sue you if it is not a legal requirement to do so?

Perhaps a short example would help here. You are out shopping for clothing. You pick out a nice outfit and decide to charge the purchase to your credit card. As part of this transaction, the business captures your credit card data and the items you have purchased. In other words, the data is created and is in an operational state.

It remains operational until your monthly billing cycle is complete and you receive your statement in the mail. At some point after this happens the data moves from an operational state to a reference state. The data is not needed to conduct any further business, but it may be needed for

reporting purposes. Furthermore, the card processing company determines that there is a period of time – maybe 90 days – during which customers frequently call to get information on recent transactions. But after that time customer requests are rare.

At this point the data can pass into an archive state. It must be kept around until such time as all regulatory requirements have passed. After all need for the data, both for internal business purposes and external legal purposes, has expired it is purged from the system.

Don't think in terms of databases or technologies that you already know when considering these data states. The data could be in three separate databases, a single database, or any combination thereof. Furthermore, don't think about data warehousing in this context – here we are talking about the single, official store of data – and its production lifecycle.

The operational and reference states have been reasonably well implemented in organizations today, but not so for archived data. Think about how you archive data, if you archive anything today at all. Is it easily accessible? Or would it take weeks or months of work to get the archived data into any reasonable format for querying?

As you design your databases, be sure to consider the data lifecycle and plan for each stage accordingly. With increasing regulatory pressures the need to better plan for and implement database archiving will only become more pervasive over time.

From [Database Trends and Applications](#), November 2006.

© 2006 Craig S. Mullins, All rights reserved.

[Home](#).