Craig S. Mullins

January/February 1995

Relational Database Journal

Data Set Closing - From Deferred to Pseudo
by Craig S. Mullins

The CLOSE option can be specified on DB2 indexes and table spaces. It indicates whether or not DB2 should close the underlying VSAM data set when it is not in use. The valid options are YES or NO.

However, the ramifications of the CLOSE options are not quite so clear cut. The concept of what constitutes "in use" is completely different depending upon which DB2 version and release you are using. Likewise, the actual technique used to close the underlying data sets varies considerably—to the point of changing which CLOSE option is more efficient. Read on for the details!

The Original DB2 CLOSE

The default CLOSE option is, and always has been, CLOSE YES. Prior to DB2 V2.3, this caused numerous headaches because CLOSE YES meant that the underlying VSAM data set was to be opened and closed every time it was used. This was highly inefficient and caused most shops to enforce coding CLOSE NO on all DB2 objects.

The DB2 V2.3 CLOSE

DB2 V2.3 implemented a deferred close strategy for DB2 page sets. When a table space or index defined using CLOSE YES is no longer being accessed DB2 performs a "logical" close. The logical close places the page set on a "drain" queue and updates SYSIBM.SYSLGRNG, but does not physically close the data set. The physical close does not happen until the number of page sets on the "drain" queue is 99% of DSMAX. DSMAX is set in the DSNZPARMs and for DB2 V2.3 can be up to 10,000. When the maximum is met, page sets are removed from the "drain" queue, and physically closed, such that least recently used page sets are closed first.   Physical data set closing minimizes the amount of virtual storage required since each opened data set requires 2.2K of storage. As log ranges are recorded for "logically" closed table spaces, the SYSIBM.SYSLGRNG table is updated and hence, performance degradation tended to occur.

In addition to the SYSLGRNG updating, there was also some small amount of overhead encountered with CLOSE YES due to the additional code that DB2 must execute to perform the logical close. This additional overhead should be minute. However, because of the SYSLGRNG updating, few shops used CLOSE YES for DB2 V2.3 table spaces.

CLOSE NO under DB2 V2.3 caused DB2 to leave unused data sets open until a STOP DATABASE command is entered or the system terminates. This results in improved performance since the open and close is avoided every time the data set is used. Opening a data set takes approximately 0.7 seconds of CPU. A single complex query may require that many data sets be opened. Keep in mind that each table space and index space involved with a query has a data set which needs to be opened at a CPU cost of approximately 0.7 seconds each!

The DB2 V3 CLOSE

As of V3, DB2 has changed the process for opening and closing the underlying VSAM data sets. The end result essentially eliminates the differences between CLOSE YES and CLOSE NO.

The DB2 V3 deferred close process is fundamentally equivalent for both CLOSE YES and CLOSE NO data sets. Only one difference remains: the order in which the underlying data sets are physically closed when the open data set threshold is reached. This threshold is still 99% of the value specified for DSMAX (or the MVS limit for the number of open data sets). If either threshold is met, DB2 will physically close some of the open data sets (3% of DSMAX). This is accomplished closing by data sets until the 3% threshold is met as follows:

1.   First, DB2 will physically close data sets coded as CLOSE YES.

2.   When no remaining CLOSE YES candidates remain, DB2 will physically close data sets marked as CLOSE NO.

This is significant because data sets coded as CLOSE NO *can be* physically closed by DB2.

Another new feature of data set closing under DB2 V3, is when the VSAM open-for-update timestamp is set. When opening a data set, DB2 V3 will not modify this timestamp until data in the page set is updated. In all prior releases, it was updated when the data set was opened. This change effectively eliminates logging for read-only data sets.

A data set is considered logically closed when an application is deallocated from the data set. This occurs at commit time or deallocation time, depending on the RELEASE parameter specified at BIND time (i.e. COMMIT or DEALLOCATE). DB2 maintains an *in-use count* for each user of each data set. When a data set is logically closed, the in-use count is reduced by 1. When the data set in-use count reaches zero, the data set is considered to be not in use. Furthermore, DB2 V3 will not update SYSLGRNG when the data set is logically closed or when the in-use count for a page set reaches zero.

Additionally, two DSNZPARMs control a new type of DB2 V3 close processing known as "pseudo" close. Refer to Table 1. If an updated page set is not modified during the period defined by

PCLOSEN and PCLOSET, the page set will be set to a read-only state. This is called a pseudo-close and when it occurs SYSLGRNG is updated. When a table space data set is pseudo-closed, the SYSLGRNG entry is closed and any updated pages are externalized to DASD.

Table 1: Pseudo Close DSNZPARMs

Parameter	Definition	Default Value
PCLOSEN	Number of checkpoints between page set modification.	5 checkpoints
PCLOSET	Amount of time since page set modification.	60 minutes

Synopsis

As a general rule of thumb, consider using CLOSE YES for most objects in DB2 V3. For those objects that are performance-critical or hyper-sensitive (the CEO accesses them at whim), CLOSE NO can be specified so that they always will be the last objects physically closed.

In short, for DB2 V3 shops, there is little remaining difference between the CLOSE YES and CLOSE NO parameters.

From Relational Database Journal, 1995.

Home.