Logical vs. Physical Disk Shadowing

Goetz Graefe  
University of Colorado  
Boulder, CO 80309-0430

Leonard D. Shapiro  
Portland State University  
Portland, OR 97207-0751

For many applications, down-times are unacceptable, and database management systems for such applications offer facilities to maintain redundant copies of the database. According to Bitton and Gray [1], "Disk shadowing is a technique for maintaining a set of two or more identical disk images on separate disk drives. Its primary purpose is to enhance reliability and availability of secondary storage by providing multiple paths to redundant data. However, shadowing can also boost I/O performance." They provide insights on how disk shadowing can reduce seek distances for read operations, in principle by using the "shadow set" of "mirrored" disks as one disk with multiple access arms, at the expense that each new page image has to be written to all disks. Bitton and Gray show that disk shadowing improves I/O performance if read operations constitute about 55% or more of all I/O operations. In this talk, we demonstrate how disk shadowing can be further improved if not physical page images but logical data sets are mirrored with different clustering strategies on different disks.

For example, in a simple university database with the relations student (student-id, name), course (course-no, title), and enrollment (student-id, course-no, grade), one would index and cluster the student and course relations on their key attributes. For the enrollment relation, however, one must choose which key attribute to index and cluster on. If one chooses student-id’s, joins with students will be very fast, e.g., in a query to list a student’s current classes. On the other hand, a join of courses and enrollment will require sorting or hashing. If one chooses to cluster enrollment on the course-no, joins with the course relation will be fast but joins with the student relation will be slow.

As a second situation in which the limitation to only one clustering choice creates a dilemma, consider the customer information of a telephone company. The billing department might want accounts in ZIP order for better postage rates, while the marketing department might want to target customers with high monthly long-distance bills. The problem in both examples is that one can cluster data according to one criterion only. We contend, however, that multiple copies as assumed in disk shadowing are no less expensive than various copies in using different clusterings to serve all or almost all users optimally.

There are two disadvantages of maintaining different clusterings. First, maintenance is more expensive since more indices need to be updated for each data update. This is a normal tradeoff between retrieval and update performance. Second, recovery of a single disk is more complex because it also requires sorting into the clustering order that was lost in the disk failure. We offer three possible viewpoints. First, if the remaining disk(s) include secondary indices on the clustering attribute of the failed disk, the entire system should still perform as well as a system without shadowing or a system using physical shadowing with a failed disk. Second, very fast sorting and clustering methods are available. Third, it is possible to keep multiple copies of each ordering, i.e., to combine physical shadowing in addition to logical shadowing.

Many relational database management systems already offer sufficient physical design options to exploit logical shadowing. For example, to shadow a relation R(A,B,C) clustered on A on disk 1, one creates a composite (multi-attribute) index on (B,A,C) on disk 2. This index is sorted and clustered on B, and it is maintained automatically when R is updated. Most database vendors have implemented their query optimizers such that they consider scanning the index only. For availability and recovery, however, two pieces are missing. First, if disk 1 with R crashes, most database systems will not support recovery of R from the index but will recover from backups and logs. Second, during the disk repair, R is deemed unavailable and the database system does not use the index to satisfy queries against R. Thus, these two special cases should be included in database software to exploit logical shadowing to its full potential.

— This research has been partially supported by the Oregon Advanced Computing Institute (OACIS) and by NSF awards IRI-8805200, IRI-8912618, and IRI-9006348.