PEBBLE PILES & INDEX STRUCTURES

Imagine trying to convince ancient sheepherders (who calculated and compared their wealth by equating piles of pebbles to the size of herds of sheep) that a piece of parchment with numbers on it would give the same results as a pile of pebbles, but that would be more accurate, easier to use, faster in processing, less resource intensive, and more portable? Now imagine trying to convince modern database researchers that the mathematical identity of records can be used to replace the physical indexing of records.

MODELING RECORDS Since the advent of computers, *records* have been used to capture and relate data. Manipulating records is what computers do. Explaining to a computer exactly what should be done with and to records, requires some form of model for describing record processing. Any such model requires some definition of what a *record* is and what the allowable *operations* are that can be applied to them.

PHYSICAL MODELS One such model defines records to be strings of bytes acted upon by operations: create, put, find, get, modify, and destroy. This is a physical model. The definition of a record is in terms of physical entities, bytes. The operations of this model are physical manipulations of records. This model, like the piling of pebbles, is conceptually unchallenging. It is simple to understand, simple to implement, and simple to use. This model is the present standard of the database community.

MATHEMATICAL MODELS Physical models model physical objects. Mathematical models model mathematical objects. Physical objects, like pebbles, have substance. They are tangible, have form and location. Mathematical objects, like numbers, are abstract. They are not tangible, have no form and no location. Physical models adhere to the laws of physics. Faithful implementations still have to be verified through testing. Mathematical models adhere to the laws of logic. Faithful implementations need no testing since the laws of logic are not subservient to the laws of physics. (In both cases above, *faithful* implies that the intent of the model has been faithfully transferred to the implementation.)

DATABASE RESEARCH Unlike the ancient sheepherders' unawareness of number, the advantages of mathematical modeling were known to database researchers long before the advent of modern computers. Given this statement to be true, one might ask why the computer industry is still stuck with only a physical model of records. The answer is quite simple: there is no obvious *number* equivalent for the concept of record. Database researchers tried very hard during the 1960s and 1970s to use the concept of *n-tuple* and classical set theory to develop a comprehensive mathematical model for the application, processing, and storage of data. Only the modeling of applications was successful. No formal mathematical model of data processing and data storage has yet to be adopted by the database research community. Today, the concept of the *mathematical identity* of a record is no better understood by database researchers than the concept of number was understood in the past by the ancient sheepherders.

EXTENDED SET-PROCESSING The problem is not that the mathematical identity for records does not exist. It does, but it is not widely understood and therefore not readily accepted. Like the concept of number, an abstract concept is hard to grasp. What is not hard to grasp is the dramatic performance improvement that implementations using mathematical identities have over systems using index structures. Given the mathematical identity of a record as a building block and operations provided by *extended set-processing*, systems can now be built that yield the same results as do index structure implementations, but that are more accurate, easier to use, faster in processing, less resource intensive, and more portable. Just as number replaced the ancient use of pebble piles, so can extended set-processing replace the current use of index structures.

Copyright © 2005 INTEGRATED INFORMATION SYSTEMS . ANN ARBOR MI . iis@umich.edu