7th ERCIM Database Research Group Workshop on Object Oriented Databases

Lisbon, 15-16 May 1995

Beyond Relations: Multi-dimensional Data Objects

Brian J Read

b.j.read@rl.ac.uk

CLRC Rutherford Appleton Laboratory
Chilton, Didcot, Oxon OX11 0QX, UK

Extended Abstract of Work in Progress

Only comparatively recently has the multi-dimensional data object received much attention. It arises when a quantity of interest is a function of parameters in two or more orthogonal dimensions. Examples taken from scientific and statistical data are discussed to illustrate the data modelling and operational issues for databases to handle this kind of object.

Starting with flat files and relations, scientists have extended this structure in a number of pragmatic ways. They store and manipulate physical measurements expressed as a multi-dimensional array to reflect the dependence on a number of independent variables. These might, for example, be wind velocities as a function of latitude, longitude and height. Widely used data formats from astronomy (FITS) and geophysics (CDF) are expressed here as complex object types.

The set of operations familiar from relational algebra can be generalised and extended. This allows for an understanding of the *slice* and *dice* operations on a multi-dimensional grid to be interpreted as kinds of *project, select* and *group by* or aggregation. New domain functions, which may have arguments and/or values which are not scalars, can be introduced. Examples are sorting or rotating a data grid.

Though normally outside the relational model, reports from database queries are often presented as cross tabulations. This can be achieved by a proprietary extension of SQL. Such an operation (and its inverse!) is now brought within an algebra for multi-dimensional objects. Indeed, data and meta-data become to some extent interchangeable.

Spreadsheets are the other paradigm for gridded data. For long untouched by database theory, from them has emerged the concepts of *pivot tables* and so-called On-line Analytical Processing (OLAP). The interest for this paper is to understand OLAP by relating it to variants of familiar database operations.

While the multi-dimensional spreadsheet is most often thought of in the context of manipulating financial data, it is interesting to apply it to the example of the statistical summary table. Such data have never fitted comfortably into conventional databases. Rather than invoke, say, the nested relations of the NFNF model, such

tables can often be better viewed as just multiple cross tabulations of an underlying multi-dimensional grid data object. An example presented is census data for Oxfordshire districts tabulating numbers of persons broken down by age and sex.

In summary, the aim is to gain an insight into multi-dimensional data objects by interpreting their structures and particularly the operations on them as a generalisation of the relational model and the relational algebra. The formalism so developed allows an abstraction from the physical storage driven structured files of the scientists for such data and from the *ad hoc* nature of the spreadsheet.

---

1 April 1995