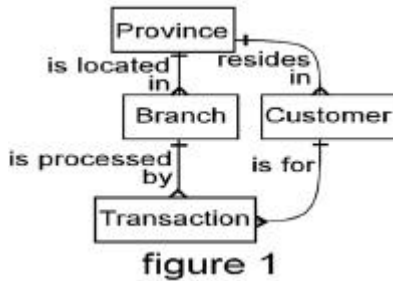


Thoughts about Data (2): Integrity of use of the Data Warehouse

By Mike King

(Read the left column completely before the right column - it seems easier to read in two columns.)



Location by Branch		Location by Customer	
Province	Transaction	Province	Transaction
North West	12 000.00	North West	10 000.00
	8 000.00		8 000.00
Gauteng	35 500.00	Gauteng	35 500.00
	45 450.00		22 450.00
	59 000.00		51 000.00
Western Cape	7 000.00	Western Cape	5 000.00
	3 500.00		3 500.00
Eastern Cape	12 330.00	Eastern Cape	31 330.00
	1 000.00		4 000.00

If the information retrieved from the data warehouse is often misunderstood, then who wants a data warehouse!

In the previous article on 'Thoughts about Data', it was suggested that Entity Models play a central role in setting up a Data Warehouse. However I have often observed the faulty use of these diagrams to try and understand the information of the business. In particular, one sees the entities on the data model named as if they were computer files. For instance, one entity is labelled 'Customer Master File', others are labelled 'Product Master File', 'Order List', 'Order Item File', and so on.

The fault is not just a technical detail, but has to do with clear business thinking. The entity model should be a semantic (focusing on meaning) model of the business. It would have been far more effective to label the entities as 'Customer', 'Product', 'Order' and 'Order Item'. These names describe the business objects themselves, and therefore enable sensible statements to be made about them.

Consider 'Each Order is placed by one Customer.' compared with 'Each record of an Order List is associated with one record of the Customer Master File', or 'Each Order Item orders one Product.' compared with 'Each record of the Order Item File must have one associated record of the Product Master File.' Clearly, one approach encourages clear thinking about the business, while the other interjects noise, which obscures clear thought.

In my experience this is not a trivial difference. The correct approach truly does lead to clear and precise understanding of the business.

The 'trick' is to choose entity names, which describe the business objects themselves, and to use a form of the name, which naturally describes one instance of the business object (i.e. use the singular).

In the above business statements about the orders, customers and products, a pattern is discernible. The statements all have the simple form

'Each ... one ,

where the boxes represent entities, and the three dots represent a sensible business phrase. In the statement 'Each Order is placed by one Customer', the boxes are the Order and the Customer, and the three dots become 'is placed by'.

This is the second technique to encourage clear business thinking. It is preferable to the statement 'There is a many to one relationship between the Order and Customer entities.'

The clarity of understanding of the business enabled by the disciplines touched on above, allows the businessman to perceive the misinterpretations he must guard against, when extracting information from the warehouse.

This is illustrated by an example in which branches (of a bank) and customers are grouped by province. Each bank transaction is originated by one customer who uses one branch for the purpose. The entity model represents this.

Suppose a simple report, showing transactions grouped by province, has been extracted from the warehouse. Depending on which relationships were utilized to produce it, the meaning and content would be completely different (not the column headings).

In one case it would mean 'The transactions processed by any branch in each province.' In the other case it would mean 'The transactions for customers residing in each province, regardless of processing branch.' Since customers often submit transactions when away from home, the difference in meaning and content is obvious, however without an Entity Model, the existence of different navigation paths linking the same entities may not have been so obvious!