

An Analytical Reporting Design Worksheet

Ross Morrissey - ross@tantiva.com

International Spectrum – November/December 2008

Last article presented the design steps around taking specific files in a MultiValue database and shaping them for analytical reporting. This time, the business process is the starting point of a design process that helps in identifying the appropriate MultiValue files to model the process and provides a concrete deliverable – an Analytical Reporting Design Worksheet. This worksheet can be used with any business intelligence tool, and follows a methodology fine-tuned with workshop participants from over 100 MultiValue sites over the course of several years. A PDF version of the worksheet can be downloaded from <http://www.tantiva.com/library/worksheet.pdf>.

Each of the worksheet entries will be described, and then examples from a report analyzing returns will be supplied. The target solution is a Tantiva Velocity matrix, but the same analysis would be required for any OLAP solution – so the word hypercube or the phrase dimensional model can be used where the word matrix occurs.

Housekeeping

The following entries will help identify this matrix.

ID: The filename associated with this design in the business intelligence environment. Example: *RETURNS*.

Designer: The party taking credit for this matrix design.

Date: The date of this design – helpful when understanding some design decisions. The sophistication of these designs will evolve over time.

Description: A brief description of the solution. Example: *Returns Analysis*.

High-level Vision

Business Process: The aspect of operations that will be exposed with this analysis. The flow of transactions as shipments, sales, or customer service calls, or snapshot of inventory, accounts payable, or accounts receivable. Example: *Returns*.

Grain: The smallest level of analysis for this process. A typical example is invoice line item (as opposed to order). A line item grain will support analysis of product sales, while an order grain will support customer level analysis only. Example: *return line item*.

Audience: The matrix designed for high-level executives will often be significantly different than the one designed for line managers. Understanding the goals of the intended audience is the key to satisfying their requirements. An audience that is too broad might be better served by two separate matrixes. Example: *Purchasing Manager*

Vision Statement: A vision statement has three important uses, it defines the scope and capability of the matrix so that completion can be defined and met, it guides minor design decisions without having to go back to project sponsors over and over, and provides the best tool for ensuring that the project sponsors understand and agree with what they are getting. Example: *Provide detailed information on returns to allow the purchasing manager to rank vendors on product quality.* Notice that returns aren't being analyzed from the customer satisfaction perspective, although the end goal of this analysis is certainly to improve customer satisfaction.

Key Questions: One or two key questions that should reflect or amplify the vision statement, which will be used to validate the design and help test. Certainly the matrix should answer many questions, but if it fails to answer these questions, it must be redesigned. Example: *What vendor had the largest dollar volume of returns last month? What product was returned most frequently?*

Choosing the Source Data

Source File matching Grain: There are three possibilities when selecting a primary source file: just one file matches or contains the grain of the business process, multiple possibilities exist, or no files are appropriate. The business process you are tracking may be represented in several different files, for example sales showing up in open orders, open order lines, order history, order history lines, shipments, and GL. Choosing the wrong file is very risky, so it's important to get this right up front.

First look at the dates associated with transactions in each file. Does order date, invoice date, posting date(s), shipping date(s), or payment date(s) make the most sense with the vision statement? Next, look at the volatility of the data in the file. Is the data constantly fluctuating as orders change based on actual inventory levels? Does data only reach the history file after a week or two? There is usually a trade-off between volatility and immediacy. Which is most appropriate for the vision statement?

How easy will it be to validate the resulting solution against the operational system? Are there existing production reports drawn directly from the file in question? If not, it may be difficult to isolate reporting logic errors from data extraction errors when you validate. Are you already using the source file in another matrix? This can speed validation, as well as offer the opportunity to leverage existing data extraction.

The source file matching the grain of analysis is not the only source for data. The analysis may be driven by an order line file, but information will be drawn from product, customer, and control files. Example: *RMA*

Source Account: To avoid ambiguity when there are multiple versions of the source file.

Validation Plan: The final result will not be trusted unless it can be successfully audited. The numbers either need to match or a rigorous explanation of why they don't match will need to be created – and sold to the end users. Either way, a solid validation strategy must be created. If there is no good way to validate the results based on the choice of source file – for example building a matrix based on a transaction journal that seems very detailed and complete could cause problems with phantom transactions that were created and backed out without ever hitting production files that drive other reporting. It might be a better idea to build from a more widely used file, even if it involves more work.

Example: *Match totals against the existing RMA Summary Report*

SELECT: If the source file contains more data – more history – or more types of transactions than required by the vision statement, a SELECT statement will be required.

Record Count: Understanding the approximate record count in both the entire file and the subset of required data can impact the frequency with which the data is refreshed.

Volatility: The ideal situation would be where complete records are added and never changed. This will open the possibility of streamlined extraction of data from the source file. The worst case is where records are potentially updated. Tracking changed records will require more than a SELECT against two files, usually refreshing the entire source file is in order. If records are deleted, it may be possible to sync up with a SELECT against two files.

Defining the Solution

Transaction Date matching Business Process and Grain: if the source file was chosen correctly, the date choice will usually be obvious. If there are several dates, choose the date that best matches the vision statement. As with the other design elements, the date is specified as a Source File DICT item. Example: *RETURN.DATE*

Dimension Attributes: the by-words of the matrix – by customer, by product, by warehouse. These will typically be “foreign keys” and have a supporting file containing full names or descriptions. Along with knowing how the supporting file is referenced, it will be helpful to know how many values there are in each dimension, and whether this is a small subset of the whole – for example an ENTITY file may contain both customers and vendors, and for most businesses, the customers will make up the bulk of the file while vendors will be a small subset. Provide an English name as well as the DICTIONARY name. Example: Customer – CUST.NO (CUSTOMER NAME), Part – PARTNO (PART DESC), Vendor – VENDOR (VENDOR NAME), Reason – REASON.CD (CODES DESC subset).

Measures: The additive numeric values – counts and amounts - surrounding this business process. Provide an English name as well as the DICTIONARY name. Example: Quantity – QTY, Extended Price – EXT.PRICE, Return Count – “ 1.”

Validating the Design

Key Questions Answered in design? Do the Dimension Attributes and Measures provide the data required to answer the key questions? How? Example: Yes, tracking vendor, price, and date answers the first question; tracking product and count answers the second.

Validation Possible? If the existing RMA report uses the same date definition and mix of measures and dimensions, validation should be possible.

If either of these two design validation steps fails, the design, and perhaps the vision statement, needs to be revisited. About one quarter of designs uncover some inconsistency during this validation stage, and this is the best time to uncover it – before any “real” work has been done.

Those experienced with the various OLAP tools will already know exactly what to do with this paper design, and will find their task significantly easier than if they worked through design decisions as they interacted with the software. For those not familiar with these tools, they will be covered in future articles.