The Matrix: Revisited

By Margy
Ross

With the current industry buzz focused on master data management (MDM), it's time to revisit one of the most critical elements of the Kimball method. Back in 1999, Ralph Kimball wrote an Intelligent Enterprise column entitled *The Matrix*. The 1999 movie of the same name spawned two sequels, but we haven't devoted a column to our matrix in more than six years.

Dimensional modelers strive to deliver information in a way that's easily understood. The same objectives apply when representing an organization's breadth of performance information and associated descriptive reference data. That sounds like a formidable task, but the tabular row-and-column orientation of a matrix lends itself beautifully to the challenge. The enterprise data warehouse bus matrix is akin to a Swiss Army pocket knife for data warehouse and business intelligence professionals; it's one tool that serves multiple purposes, including architecture planning, data integration coordination and organizational communication.

Matrix Columns for Reference Data

Delving into bus matrix fundamentals, let's start with the matrix columns, which address the demands of master data management and data integration head-on. Each column of the bus matrix corresponds to a natural grouping of standardized, descriptive reference data. In Kimball parlance, the matrix's columns are conformed dimensions containing the textual attributes used for filtering, constraining, grouping or labeling. Each attribute has an agreed-upon name, definition and domain values to ensure consistent data presentation, interpretation and content. The bus matrix includes separate columns to identify the "who, what, where, when, why and how," such as date, customer, product and employee associated with each business event or transactional activity.

Much is being said and written lately about the importance and value of master data management and data integration. We wholeheartedly agree; the Kimball Group has been talking about conformed dimensions since our first use of the terminology in 1984. We're thrilled others are jumping on the soapbox and embracing the concepts. It's simply unacceptable to build separate data stores (warehouses, marts, cubes and so on) without a framework to tie the data together. Reusable conformed dimensions supply potent integration glue, letting businesses drill across core processes with consistent, unified views.

Data Stewardship Required

Unfortunately, you can't purchase a wonder product to create conformed dimensions and miraculously solve your organization's master data management issues. Defining master conformed dimensions to be used across the enterprise is a cultural and geopolitical challenge. Technology can facilitate and enable data integration, but it doesn't fix the problem. Data stewardship must be a key component of your solution.

In our experience, the most effective data stewards come from the business community. As with technology, the data warehouse team facilitates and enables stewardship by identifying problems and opportunities and then implementing the agreed-upon decisions to create, maintain and distribute "gold standard" dimensions. But the subject-matter experts in the business are the ones rationalizing the diverse business perspectives and driving to common reference data. To reach a consensus, senior business and IT executives must openly promote and support the stewardship process and its outcomes, including the inevitable compromises.

Over the years many have criticized the concept of conformed dimensions as being "too hard." Yes, it's difficult to get people in different corners of the business to agree on common attribute names, definitions and values, but that's the crux of unified, integrated data. If everyone demands their own labels and business rules, then there's no chance of delivering the single version of the truth DW promises.

Process-Centric Rows

While the matrix columns refer to the business's nouns, the matrix rows are typically expressed as verbs. Each row of the bus matrix corresponds to a business process within the organization. A business process is an activity that the business performs, such as taking orders, shipping, invoicing, receiving payments and handling service calls. In most cases, measurements or metrics are generated each time one of these actions or events occurs. When the order is taken, the source system captures the order quantities and amounts. At shipping and customer invoicing, you again deal with quantities and amounts, albeit different from the order metrics. Each customer payment has an amount associated with it. Finally, as the business receives service calls from customers, metrics such as call duration are captured.

Each business process is typically supported by an operational system, which can introduce complexity. Some of you may deal with large, monolithic source systems that support a handful of business processes; conversely, others may have several order source systems in their environments. Insert a row in the matrix for each business process that collects or generates unique performance metrics with unique dimensionality.

After listing the core business process rows, you might also identify more complex cross-process or consolidated rows. These consolidated rows can be extremely beneficial analytically, but they are typically much more difficult to implement given the need to combine and potentially allocate performance metrics from multiple source systems; they should be tackled after the underlying processes have been built.

Associate Matrix Columns and Rows

Once you've determined several dozen core processes and dimensions, shade the matrix cells to indicate which columns are related to each row. Presto! You see the logical relationships and complex interplay between the organization's conformed reference dimensions and key business processes. By looking across a row, you quickly understand its dimensionality. Looking down a column gives immediate feedback about conformed dimension opportunities and obstacles, visually highlighting dimensions that warrant special attention given their participation in multiple matrix rows.

BUSINESS PROCESSES	Date	Product	Warehouse	Store	Promotion	Customer	Employee
Issue Purchase Orders	X	X	X				
Receive Warehouse Deliveries	X	X	X				X
Warehouse Inventory	X	X	X				
Receive Store Deliveries	X	X	X	X			X
Store Inventory	X	X		X			
Retail Sales	X	X		X	X	X	X
Retail Sales Forecast	X	X		X			
Retail Promotion Tracking	X	X		X	X		
Customer Returns	X	X		X	X	X	X
Returns to Vendor	X	X		X			X
Frequent Shopper Sign-Ups	X			X		X	X

COMMON DIMENSIONS

The number of data warehouse bus matrix rows and columns varies by organization. For many, the matrix is surprisingly square, with approximately 25 to 40 rows and a comparable number of columns. However, there are some interesting industry-specific anomalies, such as in insurance and health care, where matrices typically have more columns than rows.

It's relatively straightforward to lay out the matrix's rows and columns, and you're essentially defining the overall data architecture in the process. The matrix delivers the big-picture perspective, regardless of database or technology platform preferences, while also identifying reasonably manageable development efforts. Separate development teams can work on components of the matrix fairly independently, with confidence that the puzzle pieces will fit together.

The matrix is a succinct and effective communication tool. It lets you visually convey the entire plan within and across development teams, as well as upward and outward throughout the organization, including directly with senior IT and business management. The matrix isn't intended to be a static document that's created and then gathers dust. It will evolve as you delve more deeply into the requirements of the business and realities of your operational source systems.

Common Matrix Mishaps

When drafting a bus matrix, people sometimes struggle with the level of detail expressed by each column or row. Row mishaps commonly fall into the following two categories:

- *Departmental or overly encompassing rows*. The matrix rows shouldn't correspond to the boxes on your corporate organization chart, which represent functional groups, not business processes. Sure, some departments may be responsible or acutely interested in a single business process, but the matrix rows shouldn't look like a list of direct reports to your CEO.

- *Report-centric or too narrowly defined rows*. At the opposite extreme, your bus matrix shouldn't resemble a laundry list of requested reports. A single business process, such as shipping orders, often supports numerous analyses such as customer ranking, sales rep performance and product movement analysis. The matrix row should reference the business process, not the derivative reports or analytics.

When defining the matrix columns, architects naturally fall into the similar traps of defining columns that are either too broad or narrow:

- *Overly generalized columns.* A "person" column in a bus matrix typically refers to a wide variety of people, from internal employees to external suppliers and customer contacts. Since there's virtually zero overlap between these populations, it adds confusion to lump them into a single, generic dimension. Similarly, it's not beneficial to put internal and external addresses referring to corporate facilities, employee addresses and customer sites into a generic location column in the matrix.

- *Separate columns for each level of a hierarchy*. The columns of the bus matrix should refer to dimensions at their most detailed level. Some business process rows may require an aggregated version of the detailed dimension, such as sales forecast metrics at the brand level. But rather than creating separate matrix columns, such as product, brand, category and department, for each level of the product hierarchy, we advocate a single column for product. As the cell is shaded to show participation with a business process row, you can denote the level of detail in the cell (if it's not at the most granular level). An even more extreme example of this mishap is to list each individual descriptive attribute as a separate column; this defeats the concept of dimensions and results in a completely unruly matrix.

Matrix Extensions

One of the beauties of the bus matrix is its simplicity. You can reuse the familiar tabular format to convey other DW/BI "relationships." These extensions aren't a substitute for the enterprise DW bus matrix, but are intended as complementary opportunities to reuse the framework.

- *Opportunity matrix*. Once the bus matrix rows have stabilized, replace the dimension columns with business stakeholders, such as marketing, sales and finance. Based on the business requirements gathering activities, shade the cells to indicate which stakeholders (columns) are interested in which business process rows. This is a useful tool to assist with the prioritization of matrix rows, plus it establishes which business groups should be sending representatives to the collaborative dimensional modeling sessions.

- *Detailed implementation bus matrix*. A single business process matrix row sometimes spawns multiple fact tables or OLAP cubes. For example, this occurs when there's a need to view the metrics at both atomic and summarized levels of detail or with both transactional and snapshot perspectives. In this scenario, the matrix rows are expanded to list individual fact tables or OLAP cubes, along with their specific granularity and captured or derived metrics. The standard bus matrix dimension columns are retained.