

## **Kimball Design Tip #39: Bus Architecture Foundation For Analytic Applications**

By Bill Schmarzo

In our previous analytic application design tip, we explored the industry drivers behind the current analytic application hoopla. One of those drivers is the broad acceptance of dimensional modeling as a mature, business-centric data warehouse discipline. Dimensional models promote analytic exploration by providing a high performance and easy-to-use environment to identify performance exceptions and their causal factors.

There are two additional requirements placed on the data warehouse architecture to effectively support analytic applications.

First, the data warehouse bus architecture and its enabling conformed dimensions must serve as the cornerstone of the data warehouse architecture. Ralph and Margy's recent book, "The Data Warehouse Toolkit 2nd Edition," defines the data warehouse bus architecture as "the architecture for the data warehouse's presentation area based upon conformed dimensions and facts. Without adherence to the bus architecture, a data mart is a standalone stovepipe application." For more information about bus architecture and conformed dimensions, see Ralph's 1999 article from Intelligent Enterprise magazine.

Remember, we advocate implementing dimensional models focusing on business processes (e.g., orders, shipments, inventory, and payments), not on business departments (e.g., sales, marketing, manufacturing or finance). The bus architecture and conformed dimensions are a necessary prerequisite for analytic applications because the most interesting, high-value analytic applications require integrating metrics from multiple business processes. For example, a customer lifetime value analysis requires data from orders (to determine revenue and margin contributions), receivables (to determine collections costs), sales pipeline (to determine cost of sales), shipments (to determine manufacturing costs, particularly in build-to-order businesses), logistics (to determine distribution costs), and service calls (to determine support costs).

The key metrics from each of these business processes would be stored in separate dimensional models. The bus architecture allows us to traverse between the different business process models to pull together an integrated view of the customer. In a query tool, we use multi-pass SQL to perform the traversal. This cross-business process integration would be very challenging without a bus architecture using conformed dimensions.

Secondly, "Consolidated" data marts are often also required to support analytic applications. A consolidated data mart is an integrated superset of single business process-oriented dimensional models. It brings together the necessary data from multiple business processes into a single consolidated mart to support a suite of analytical applications. For example, a customer profitability consolidated mart might bring together data from orders, invoicing, solicitations and customer service in order to support customer attrition, customer promotional effectiveness, and customer cross-sell analytic applications.

Business user ease-of-use and performance are the key drivers behind the use of consolidated marts. Forcing users to become expert at multi-pass SQL, being dependent upon an IT professional, or "joining on the screen" on their local PC to integrate data from multiple business process marts is unrealistic. Unfortunately, many query and reporting tools do not provide adequate "drill across" support to navigate between multiple business process marts. The integrated consolidated mart provides a single analytic view (a single star schema) that is easier for users, provides high

performance, and is supported by most query and reporting tools. In a sense, a consolidated data mart is the next evolutionary step beyond separate data marts with conformed dimensions. The advantage of a consolidated data mart is that multi-pass SQL techniques are not required (because all the information has been consolidated into a single fact table), but the disadvantage is the cost and the delays associated with building the consolidated tables from the constituent separate data marts.

The move from today's data warehouse architecture to one that can support tomorrow's analytic applications is a natural evolution. It requires adherence to well-known and documented dimensional modeling concepts such as the data warehouse bus architecture, conformed dimensions, and consolidated data marts. Without the appropriate architecture, accessing and analyzing data from multiple business processes in support of these high-business-value analytic applications becomes foreboding.