

## Design Tip #101 Slowly Changing Vocabulary

By Margy Ross

Ralph's first article on data warehousing appeared in 1995. During the subsequent 13 years, we've written hundreds of articles and Design Tips, as well as published seven books. Remarkably, the concepts that Ralph introduced in the 1990s have withstood the test of time and remain relevant today. However, some of our vocabulary has evolved slightly over the years. This became readily apparent when we were working on the 2<sup>nd</sup> edition of *The Data Warehouse Lifecycle Toolkit* which was released in January 2008.

### Data Warehouse vs. Business Intelligence

Traditionally, the Kimball Group has referred to the overall process of providing information to support business decision making as *data warehousing*. Delivering the end-to-end solution, from the source extracts to the queries and applications that the business users interact with, has always been one of our fundamental principles; we wouldn't consider building data warehouse databases without delivering the presentation and access capabilities. This terminology is strongly tied to our written legacy; nearly all our *Toolkit* books include references to the data warehouse in their titles.

The term *business intelligence* emerged in the 1990s to refer to the reporting and analysis of data stored in the warehouse. Some misguided organizations had built data warehouses as archival repositories without regard to getting the data out and usefully delivered to the business. Not surprisingly, these data warehouses had failed and people were excited about BI to deliver on the promise of business value.

Some folks continue to refer to *data warehousing* as the overall umbrella term, with the data warehouse databases and BI layers as subset deliverables within that context. Others refer to *business intelligence* as the overarching term, with the data warehouse as the central data store foundation of the overall business intelligence environment.

Because the industry cannot reach agreement, we have been using the phrase *data warehouse/business intelligence (DW/BI)* to mean the complete end-to-end system. Though some would argue that you can theoretically deliver BI without a data warehouse, and vice versa, we believe that is ill-advised. Linking the two in the *DW/BI* acronym reinforces their dependency.

Independently, we refer to the queryable data in your DW/BI system as the *enterprise's data warehouse*, and value-add analytics as *BI applications*. In other words, the *data warehouse is the foundation for business intelligence*.

### Data Staging → ETL System

We often refer to the extract, transformation, and load (ETL) system as the back room kitchen of the DW/BI environment. In a commercial restaurant's kitchen, raw materials are dropped off at the back door and transformed into a delectable meal for the restaurant patrons by talented chefs. Much the same holds true for the DW/BI kitchen: raw data is extracted from the operational source systems and dumped into the kitchen where it is transformed into meaningful information for the business. Skilled ETL architects and developers wield the tools of their trade in the DW/BI kitchen; once the data is verified and ready for business consumption, it is

appropriately arranged “on the plate” and brought through the door into the DW/BI front room.

In the past, we’ve referred to the ETL system as *data staging*, but we’ve moved away from this terminology. We used data staging to refer to all the cleansing and data preparation that occurred between the source extraction and loading into target databases, however others used the term to merely mean the initial dumping of raw source data into a work zone.

#### Data Mart → Business Process Dimensional Model

The DW/BI system’s front room must be designed and managed with the business users’ needs front and center. Dimensional models are a fundamental front room deliverable; fact tables contain the metrics resulting from a business process or measurement event, while dimension tables contain the descriptive attributes and characteristics associated with measurement events. *Conformed dimensions* are the master data of the DW/BI environment, managed once in the kitchen and then shared by multiple dimensional models for enterprise integration and consistency.

Historically, we’ve heavily used the term *data mart* to refer to these architected *business process dimensional models*. While data mart is short-and-sweet, the term has been marginalized by others to mean summarized departmental, independent non-architected dataset; unfortunately, this hijacking has rendered the data mart terminology virtually meaningless.

#### End User Applications → Business Intelligence Applications

It’s not enough to just deliver dimensional data to the DW/BI system’s front room. Some business users are interested in and capable of formulating ad hoc queries, but most will be more satisfied with the ability to execute predefined applications that query, analyze, and present information from the dimensional model. There is a broad spectrum of *BI application* capabilities, from a set of canned static reports to analytic applications that directly interact with the operational transaction systems. In all cases, the goal is to deliver capabilities that are accepted by the business to support and enhance their decision making. Previously, we referred to these templates and applications as *end user applications*, but have since adopted the more current BI application terminology.

While our vocabulary has evolved slightly over the last 13 years, the underlying concepts have held steady. This is a testament to the permanency of our mission: bringing data effectively to business users to help them make decisions. Considering all the other changes in your world during this same timeframe, I think you’ll agree that the evolution of our vocabulary has been very slowly changing in comparison.