

## Kimball Design Tip #36: To Be Or Not To Be (Centralized)

By Margy Ross and Bob Becker

Contrary to William Shakespeare and some data warehouse industry pundits, that's NOT the question.

In this article, we discuss an issue faced by maturing data mart/warehouse environments. While some organizations are newcomers to the data warehouse party, others have been at this for quite a while. As the market matures, the cause of data warehouse "pain" within the IT organization is bound to evolve. Recently, centralization has been promoted as the latest miracle elixir. Centralization is claimed to turn independent, disparate data marts into "gold" by reducing administrative costs and improving performance. While centralization "may" deliver some operational efficiency, it does not inherently address the larger issues of integration and consistency.

If your data warehouse environment has been developed without an overall strategy, you are probably dealing with multiple, independent islands of data with the following characteristics:

- Multiple, uncoordinated extracts from the same operational source
- Multiple variations of the same information with inconsistent naming conventions and business rules
- Multiple analyses illustrating inconsistent performance results

Some analysts have tarnished the reputation of data marts by attributing this multitude of data warehousing sins to the mart approach. That's a gross generalization that fails to reflect the benefits that many organizations have realized with their architected data marts. The problems we listed above are the result of a non-existent, poorly defined, or inappropriately executed strategy and can exist with any architectural approach including the enterprise data warehouse, hub-and-spoke and distributed/federated marts.

We can all agree that isolated sets of analytical data warrant attention. Clearly, they are inefficient and incapable of delivering on the business promise of data warehousing. Standalone databases may be easier to initially implement, but without a higher-level integration strategy they are dead-ends that continue to perpetuate incompatible views of the organization. Merely moving these renegade data islands onto a centralized box is no silver bullet if you dodge the real issue: data integration and consistency. A centralization approach that fails to deal with these ills is guilty of treating the symptoms rather than the disease. While it may be simpler to just brush integration and consistency under the carpet due to the political or organizational challenges associated with them, these are the tickets to true business benefit from the data warehouse. It's hard work, but the business pay-off is worth it. In the vernacular of dimensional modeling, this means focusing on the data warehouse bus architecture and conformed dimensions/facts.

As we've previously described, the data warehouse bus architecture is a tool to establish the overall data integration strategy for the organization's data warehouse. It provides the framework for integrating your organization's analytic information. The bus architecture is documented and communicated via the Data warehouse bus matrix (as Ralph described in an Intelligent Enterprise article - [www.intelligententerprise.com/db\\_area/archives/1999/990712/webhouse.shtml](http://www.intelligententerprise.com/db_area/archives/1999/990712/webhouse.shtml)). The matrix rows represent the core business processes of the organization, while the matrix columns reflect the common, conformed dimensions.

Conformed dimensions are the means for consistently describing the core characteristics of your business. They are the integration points between the disparate business processes of the organization, ensuring semantic consistency between the processes. There may be valid business reasons for not conforming dimensions. For example, if you are a diversified conglomerate that sells unique products to unique customers through unique channels. However, for most organizations, the key to integrating disparate data is organizational commitment to the creation and use of conformed dimensions throughout your data warehouse architecture, regardless of whether data is centralized or distributed physically.

As we warned earlier, centralization without integration may only throw more fuel on the pre-existing problems. Management may be convinced that buying a new box to house the myriad of existing data marts/warehouses will deliver operational efficiency. Depending on the amount of money they're willing to spend on a centralized hardware platform, it may even positively impact performance. However, these IT benefits are insignificant compared to the business potential from integrated data. Centralization without data integration and semantic consistency will distract an organization from focusing on the real crux of the problem. Inconsistent data will continue to flummox the organization's decision-making ability.

We are well aware that moving to a data warehouse bus architecture will require organizational willpower and the allocation of scarce resources. No one said it would be easy. In fact some industry analysts state that it can't be done. However, our clients' experiences prove otherwise.

We've outlined the typical tasks involved in migrating disparate data to a bus architecture with conformed dimensions. Of course, since each organization's pre-existing environment varies, the list would need to be adjusted to reflect your specific scenario.

- Document the existing data marts/warehouses in your organization, noting the inevitable data overlaps.
- Conduct a high-level assessment of the organization's unmet business requirements.
- Gather key stakeholders to develop a preliminary data warehouse bus matrix for your organization.
- Identify a dimension authority or stewardship committee for each dimension to be conformed.
- Design the core conformed dimensions by integrating and/or reconciling the existing, disparate dimension attributes. Realistically, it may be overwhelming to get everyone to agree on every attribute, but don't let that bring this process to a crashing halt. You've got to start walking down the path toward integration.
- Gain organization agreement on the master conformed dimension(s).
- Develop an incremental plan for converting to the new conformed dimension(s).

Formulating the bus architecture and deploying conformed dimensions will result in a comprehensive data warehouse for your organization that is integrated, consistent, legible and well performing. You'll be able to naturally add data marts with confidence that it will integrate with the existing data. Rather than diverting attention to data inconsistencies and reconciliations, your organization's decision-making capabilities will be empowered with consistent, integrated data.