

Kimball Design Tip #53: Dimension Embellishments

By Bob Becker

When developing dimensional models, we strive to create robust dimension tables decorated with a rich set of descriptive attributes. The more relevant attributes we pack into dimensions, the greater the users' ability to evaluate their business in new and creative ways. This is especially true when building a customer-centric dimension.

We encourage you to embed intellectual capital in dimensional models. Rather than applying business rules to the data at the analytical layer (often using Excel), derivations and groupings required by the business should be captured in the data so they're consistent and easily shared across analysts regardless of their tools. Of course, this necessitates understanding what the business is doing with data above and beyond what's captured in the operational source. However, it's through this understanding and inclusion of derived attributes (and metrics) that the data warehouse adds value.

As we deliver a wide variety of analytic goodies in the customer dimension, we sometimes become victims of our own success. Inevitably, the business wants to track changes for all these interesting attributes. Assuming we have a customer dimension with millions of rows, we need to use mini-dimensions to track customer attribute changes. Our old friend, the type 2 slowly changing dimension technique, isn't effective due to the large number of additional rows required to support all the change.

The mini-dimension technique uses a separate dimension(s) for the attributes that frequently change. We might build a mini-dimension for customer demographic attributes, such as own/rent home, presence of children, and income level. This dimension would contain a row for every unique combination of these attributes observed in the data. The static and less frequently changing attributes are kept in our large base customer dimension. The fact table captures the relationship of the base customer dimension and demographic mini-dimension as the fact rows are loaded.

It is not unusual for organizations dealing with consumer-level data to create a series of related mini-dimensions. A financial services organization might have mini-dimensions for customer scores, delinquency statuses, behavior segmentations, and credit bureau attributes. The appropriate mini-dimensions along with the base customer dimension are tied together via their foreign key relationship in the fact table rows. The mini-dimensions effectively track changes and also provide smaller points of entry into the fact tables. They are particularly useful when analysis does not require consumer-specific detail.

Users often want to analyze customers without analyzing metrics in a fact table, especially when comparing customer counts based on specific attribute criteria. It's often advantageous to include the currently-assigned surrogate keys for the customer mini-dimensions in the base customer dimension to facilitate this analysis without requiring joins to the fact table. A simple database view or materialized view provides a complete picture of the current view of the customer dimension. In this case, be careful not to attempt to track the mini-dimension surrogate keys as type 2 slowly changing dimension attributes. This will put you right back at the beginning with a large customer dimension growing out of control with too frequent type 2 changes.

Another dimension embellishment is to add aggregated performance metrics to the customer dimension, such as total net purchases last year. While we normally consider performance metrics to be best handled as facts in fact tables (and they should certainly be there!), we are populating them

in the dimension to support constraining and labeling, not for use in numeric calculations. Business users will appreciate the inclusion of these metrics for analyses. Of course, populating these attributes in our dimension table places additional demands on the data staging system. We must ensure these aggregated attributes are accurate and consistent.

An alternative and/or complementary approach to storing the actual aggregated performance metrics is grouping the aggregated values into range buckets or segments, such as identifying a credit card customer as a balance revolver or transactor. This is likely to be of greater analytic value than the actual aggregated values and has the added benefit of assuring a consistent segment definition across the organization. This approach works particular well in combination with the mini-dimension technique.