



Kimball Design Tip #6: Showing The Correlation Between Dimensions

By Ralph Kimball

One of the questions I get asked frequently is "how can I represent the correlation between two dimensions without going through the fact table?" Often the designer follows up with the question "can I create a little joiner table with just the two dimension keys and then connect THAT table to the fact table?"

Of course, in a classic dimensional model, we only have two choices. Either the dimensions are modeled independently and the two dimension keys occur together ONLY in the fact table, or else the two dimensions are combined into a single super-dimension with a single key. So when does the designer choose separate dimensions and when does the designer combine the dimensions?

To be more concrete, let us imagine that the two dimensions are Product and Market in a retail setting. Suppose that the fact table of interest records actual sales of Products in the various Markets over Time. Our desire to represent the correlation between the Product and Market dimensions is based on the suspicion that "Products are highly correlated with Markets in our business". This sentence is the key to the whole design question.

If Products are extremely correlated with Markets, then there may be a 1-to-1 or a many-to-1 relationship between Products and Markets. In this case, combining the two dimensions makes eminent sense. The combined dimension is only as big as the larger of the two dimensions. Browsing (looking at the combinations of values) within the combined dimension would be useful and fast. Interesting patterns would be apparent.

But rarely do Product and Market have such a nice relationship. At least three factors intrude that eventually make us pull these two dimensions apart.

1. The 1-to-1 or many-to-1 relationship may not literally be true. We may have to admit that the relationship is really many-to-many. When most Products are sold in most Markets, it becomes obvious that we need two dimensions because otherwise our combined dimension becomes huge and starts to look like a Cartesian product of the original dimensions. Browsing doesn't yield much insight.
2. If the relationship between Product and Market varies over Time, or under the influence of a fourth dimension like Promotion, then we have to admit that the combined dimension is in reality some kind of fact table itself!
3. There are more relationships between Product and Market than simply the retail sales. Each business process involving Product and Market will turn out to generate its own fact table. Good candidates include Promotion Coverage, Advertising, Distribution, and Manufacturing. Creating a combined Product-Market dimension exclusively around retail sales would make some of these other processes impossible to express.

The point of this Design Tip is to encourage you to visualize the relationship between the entities you choose as dimensions. When entities have a fixed, time-invariant strongly correlated relationship, they should be modeled as a single dimension. In most other cases, your design will be simpler and smaller when you separate the entities into two dimensions.

Don't avoid the fact table! Fact tables are incredibly efficient. They contain only dimension keys and measurements. They contain only the combinations of dimensions that occur in a particular process. So when you want to represent the correlation between dimensions, remember that the fact table was created exactly for this purpose.