



Kimball Design Tip #50: Factless Fact Tables? Sounds Like Jumbo Shrimp?

By Bob Becker

Factless fact tables appear to be an oxymoron, similar to jumbo shrimp. How can you have a fact table that doesn't have any facts? We've discussed the basics of factless fact tables several times in our books and articles. In this design tip, we use a factless fact table to complement our slowly changing dimension strategies.

As you probably recall, a factless fact table captures the many-to-many relationships between dimensions, but contains no numeric or textual facts. They are often used to record events or coverage information. Common examples of factless fact tables include:

- Identifying product promotion events (to determine promoted products that didn't sell)
- Tracking student attendance or registration events
- Tracking insurance-related accident events
- Identifying building, facility, and equipment schedules for a hospital or university

For more information on factless fact tables, see Ralph's earlier articles at

http://www.intelligententerprise.com/db_area/archives/1999/991602/warehouse.shtml and <http://www.dbmsmag.com/9609d05.html>.

In today's design tip, imagine we are working on a design for a large business-to-consumer company (pick your favorite consumer-oriented industry – airline, insurance, credit card, banking, communications, or web retailer). The company does business with tens of millions of customers. In addition to the typical requirements for transaction schema to track consumer behavior and periodic snapshot schema to trend our consumer relationships over time, our business partners need the ability to see a customer's exact profile (including dozens of attributes) at any point in time.

Long-time readers may remember Ralph discussing a similar situation in Design Tip 13 (<http://ralphkimball.com/html/designtips/2000/designtip13.html>). He outlined a technique where the dimension itself captures profile change events as a slowly changing dimension Type 2, rather than creating a fact table to capture the profile transactions. However, we are not likely to use the DT #13 technique in the current scenario given the huge data volumes (millions of customer rows) and potentially volatile changes (dozens of attributes).

Let's assume we design a base customer dimension (with minimal SCD Type 2 attributes), along with four "mini" dimensions to track changes to customer credit attributes, customer preferences, market segmentation/propensities, and customer geography. The five foreign keys are included in the transaction-grained fact table, as well as the monthly snapshot. These foreign keys represent the customer's "state" when the fact row is loaded. So far so good, but we still need to support customer profiling at any point in time. We consider using another periodic snapshot fact table, loaded daily for every customer to capture the point-in-time relationship of the customer dimension and associated mini-dimensions. This translates into loading tens of millions of snapshots nightly with several years of history. We quickly do the math and decide to evaluate other alternatives.

About now you're thinking "That's great, but what about the jumbo shrimp?" We can use a factless fact table to capture the relationship between the customer dimension and mini-dimensions over time. We load a fact row in the factless fact table whenever there is a Type 2 change to the base customer dimension or a change in the relationship between the base dimension and the mini-dimensions. The factless fact table contains foreign keys for the base customer dimension and each of the four mini-dimensions when the row is loaded. We then embellish this design with two dates, row effective and row expiration, to locate a customer's profile at any point in time. We might also add a simple dimension to flag the current customer profile, in addition to a change reason dimension to indicate what caused a new row to be loaded into the factless fact table.