

## Design Tip #140 Is it a Dimension, a Fact, or Both?

By Joy Mundy

For most subject areas, it's pretty easy to identify the major dimensions: Product, Customer Account, Student, Employee, and Organization are all easily understood as descriptive dimensions. A store's sales, a telecommunication company's phone calls, and a college's course registrations are all clearly facts.

However, for some subject areas, it can be challenging – especially for the new dimensional modeler – to identify whether an entity is a dimension or a fact. For example, an insurance company's claims processing unit wants to analyze and report on their open claims. "Claim" feels like a dimension, but at the same time, it can behave like a fact table. A similar situation arises with software companies with extended sales cycles: is the sales opportunity a dimension, a fact, or both?

In most cases, the design puzzle is solved by recognizing that the business event you're trying to represent in the fact table is actually a long-lived process or lifecycle. Often, the business users are most interested in seeing the current state of the process. A table with one row per process – one row per claim or sales opportunity, for example – sounds like a dimension table. But if you distinguish between the entity (claim or sales opportunity) and the process (claim settlement or sales pipeline), it becomes clearer. We need a fact table to measure the process. And many dimension tables to describe the attributes of the entity measured in that process.

This type of schema is implemented as an accumulating snapshot. The accumulating snapshot is less common than transactional and periodic snapshot fact tables. The grain of this type of fact table is one row per process; it has many roles of the date dimension; and the fact table rows are updated multiple times over the life of the process (hence the name accumulating snapshot). You can read more about accumulating snapshot fact tables in *The Data Warehouse Toolkit*, pages 128-134.

Many of the core dimensions of an accumulating snapshot schema are easy to identify, but there are some challenges to these designs. Long-lived processes tend to have a lot of little flags and codes from the source system that signal various statuses and conditions in the process. These are great candidates for junk dimensions. Expect your accumulating snapshot schema to have several junk dimensions.

I try to avoid creating a dimension with the same number of rows as the accumulating snapshot fact table. This can happen if you have not separated your junk dimensions into logically correlated groupings. In the earlier examples, I also resist designing a Claim or Sales Opportunity text description dimension for as long as I can. Realistically, though, there is usually some bit of detail that the business users absolutely need to see, such as the accident report or description of the sales opportunity. Although such a dimension may be huge, it is likely to be accessed by the business users only after other dimensions have been tightly constrained since there may be no predictable structure to the text entries.