

Kimball Design Tip #61: Handling All The Dates

By Bob Becker

It's not unusual to identify dozens of different dates, each with business significance that must be included in a dimensional design. For example, in a financial services organization you might be dealing with deposit date, withdrawal date, funding date, check written date, check processed date, account opened date, card issued date, product introduction date, promotion begin date, customer birth date, row effective date, row load date and statement month.

The first thing to know is not all dates are created equal and handled the same way. Many dates end up as date dimension foreign keys in the fact tables. Most of the remaining dates become attributes of other dimensions. Finally, some dates are included in the design to facilitate ETL processing and/or auditing capabilities.

Assume our financial services company is designing a fact table integrating checking account transactions, such as deposit, ATM and check transactions. Each fact row includes a transaction type dimension to identify the transaction it represents, as well as a transaction date dimension. The business meaning of the date (such as check transaction date, ATM transaction date or deposit transaction date) is defined by the transaction type dimension. In this case, we would not include three separate date keys in the fact table since only one would be valid for a given row.

In other situations, a single transaction represented by one row in the fact table can be defined by multiple dates, such as the transaction event date and transaction posted date. In this case, both dates will be included as uniquely-named dimension foreign keys. We would use role playing to physically build one date dimension with views to present logically unique date dimensions.

A recent client "generalized" their transaction schema to include all transactions across multiple business processes. The schema for this generalized design propagated a number of date dimensions into the transaction fact table having a null or not applicable value. In our financial services case study, this would be akin to generalizing all checking, credit card, savings and mortgage transactions into a single fact table.

All these fact rows represent transactions in a general sense, but they result from different business measurement processes. Remember, we advocate designing your schema by business process, usually a fact table per business process. Checking transactions are very different from funding a mortgage loan. The two processes have unique metrics and dimensionality, resulting in separate fact tables, each with its own uniquely defined and labeled dates associated with it.

Obviously, we also include a date dimension in periodic snapshot schema reflecting the time period for the row, such as snapshot month. These are shrunken subset date dimensions that conform with our core date dimension.

Many business-significant dates will be included as attributes in dimension tables. Account open date would be included in the account dimension. Likewise, customer birth date, product introduction date and promotion begin date belong in their respective dimensions of customer, product and promotion.

When dates are dimension table attributes, we need to consider their reporting and analysis usage.

Is it enough to know the actual date an account was opened or should we also include attributes for account opened year, account opened month, and account opened month/year? These additional attributes improve the business users' ability to ask interesting analytic questions by grouping accounts based on the year and/or month the account was opened.

In order to support more extensive date-related analysis of these dimension attributes, you can incorporate a robust date dimension as an outrigger to the dimension table. In this case, we include the surrogate key for the applicable date in our dimension rather than the date itself, then use a view to declare unique business-appropriate column labels. This technique opens up all of the rich attributes of our core date dimension for analysis. However, remember that extensive use of outrigger dimensions can compromise usability and performance. Also, be careful that all outrigger dates fall within the date range stored in the standard date dimension table.

There are additional dates to help the data warehouse team manage the ETL process and support audit-ability of the data. Dates such as row effective date, row expiration date, row loaded date or row last updated date should be included in each dimension table. While these dates may not need to be user accessible, they can prove invaluable to the data warehouse team.