

Kimball Design Tip #46: Another Look At Degenerate Dimensions

By Bob Becker

We are often asked about degenerate dimensions in our modeling workshops. Degenerate dimensions cause confusion since they don't look or feel like normal dimensions. It's helpful to remember that according to Webster, "degenerate" refers to something that's 1) declined from the standard norm, or 2) is mathematically simpler.

A degenerate dimension (DD) acts as a dimension key in the fact table, however does not join to a corresponding dimension table because all its interesting attributes have already been placed in other analytic dimensions. Sometimes people want to refer to degenerate dimensions as textual facts, however they're not facts since the fact table's primary key often consists of the DD combined with one or more additional dimension foreign keys.

Degenerate dimensions commonly occur when the fact table's grain is a single transaction (or transaction line). Transaction control header numbers assigned by the operational business process are typically degenerate dimensions, such as order, ticket, credit card transaction, or check numbers. These degenerate dimensions are natural keys of the "parents" of the line items.

Even though there is no corresponding dimension table of attributes, degenerate dimensions can be quite useful for grouping together related fact tables rows. For example, retail point-of-sale transaction numbers tie all the individual items purchased together into a single market basket. In health care, degenerate dimensions can group the claims items related to a single hospital stay or episode of care.

We sometimes encounter more than one DD in a fact table. For example, an insurance claim line fact table typically includes both claim and policy numbers as degenerate dimensions. A manufacturer could include degenerate dimensions for the quote, order, and bill of lading numbers in the shipments fact table.

Degenerate dimensions also serve as a helpful tie-back to the operational world. This can be especially useful during data staging development to align fact table rows to the operational system for quality assurance and integrity checking.

We typically don't implement a surrogate key for a DD. Usually the values for the degenerate dimension are unique and reasonably sized; they don't warrant the assignment of a surrogate key. However, if the operational identifier is a unwieldy alpha-numeric, a surrogate key might conserve significant space, especially if the fact table has a large number of rows. Likewise, a surrogate key is necessary if the operational ID is not unique over time or facilities. Of course, if you join this surrogate key to a dimension table, then the dimension is no longer degenerate.

During design reviews, we sometimes find a dimension table growing proportionately with the fact table. As rows are inserted into the fact table, new rows are also inserted into a related dimension table, often at the same rate as rows were added to the fact table. This situation should send a red flag waving. Usually when a dimension table is growing at roughly the same rate as the fact table, there is a degenerate dimension lurking that has been missed in the initial design.