# Design Tip #152 Slowly Changing Dimension Types 0, 4, 5, 6 and 7

By Margy
Ross

Ralph introduced the concept of slowly changing dimension (SCD) attributes in 1996. Dimensional modelers, in conjunction with the business's data governance representatives, must specify the data warehouse's response to operational attribute value changes. Most Kimball readers are familiar with the core SCD approaches: type 1 (overwrite), type 2 (add a row), and type 3 (add a column). Since legibility is a key component of the Kimball mantra, we sometimes wish Ralph had given these techniques more descriptive names, such as "overwrite" instead of "type 1." But at this point, the SCD type numbers are part of our industry's vernacular.

We have written about more advanced SCD patterns, such as the 2005 article entitled "Slowly Changing Dimensions are Not Always as Easy as 1, 2, 3." However, we've not consistently named the more advanced and hybrid techniques. With the third edition of *The Data Warehouse Toolkit* (Wiley, 2013), we've decided to assign "type numbers" to several techniques that have been described, but not precisely labeled in the past. Our hope is that more specific technique names will facilitate clearer communication between DW/BI team members.

**Type 0: Retain Original**

With type 0, the dimension attribute value never changes, so facts are always grouped by this original value. Type 0 is appropriate for any attribute labeled "original," such as a customer's original credit score, or any durable identifiers. Type 0 also applies to most date dimension attributes.

**Type 4: Add Mini-Dimension**

The type 4 technique is used when a group of dimension attributes are split off into a separate mini-dimension. This approach is useful when dimension attribute values are relatively volatile. Frequently-used attributes in multi-million row dimension tables are also mini-dimension design candidates, even if they don't change frequently. A surrogate key is assigned to each unique profile or combination of attribute values in the mini-dimension. The surrogate keys of both the base dimension and mini-dimension profile are captured as foreign keys in the fact table.
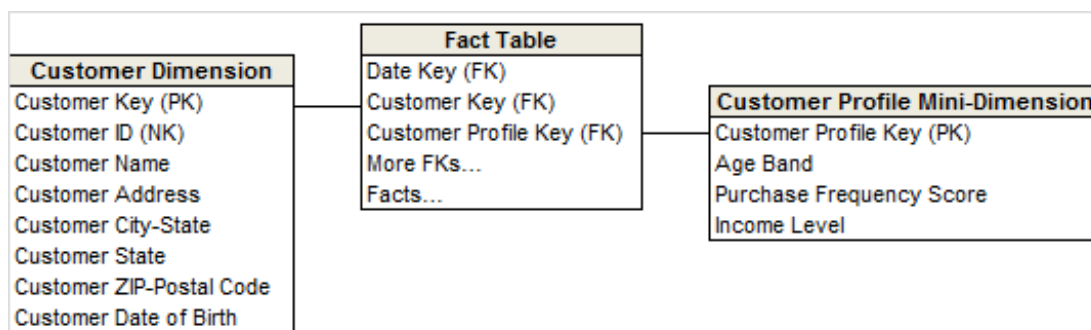
The following type 5, 6, and 7 techniques are hybrids that combine the basics to support the common requirement to both accurately preserve historical attribute values, plus report historical facts according to current attribute values. The hybrid approaches provide more analytic flexibility, albeit with greater complexity.

## Type 5: Add Mini-Dimension and Type 1 Outrigger

The type 5 technique builds on the type 4 mini-dimension by embedding a "current profile" mini-dimension key in the base dimension that's overwritten as a type 1 attribute. This approach, called type 5 because 4 + 1 equals 5, allows the currently-assigned mini-dimension attribute values to be accessed along with the base dimension's others without linking through a fact table. Logically, we typically represent the base dimension and current mini-dimension profile outrigger as a single table in the presentation layer. The outrigger attributes should have distinct column names, like "Current Income Level," to differentiate them from attributes in the mini-dimension linked to the fact table. The ETL team must update/overwrite the type 1 mini-dimension reference whenever the current mini-dimension changes over time. If the outrigger approach does not deliver satisfactory query performance, then the mini-dimension attributes could be physically embedded (and updated) in the base dimension.
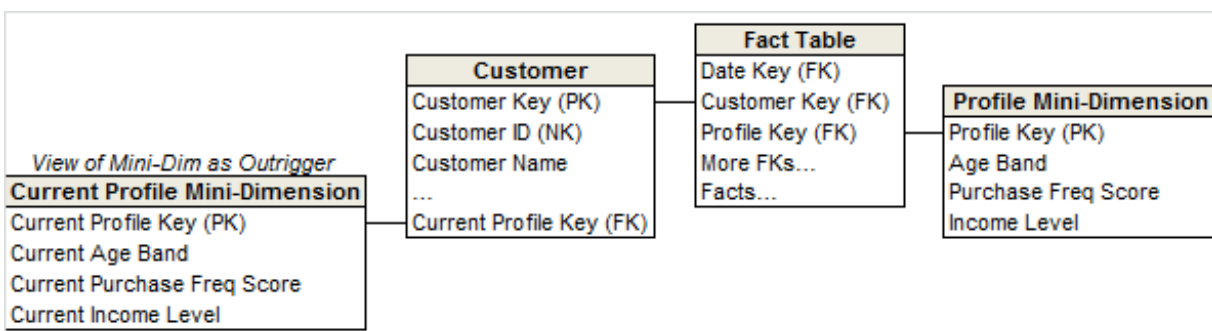


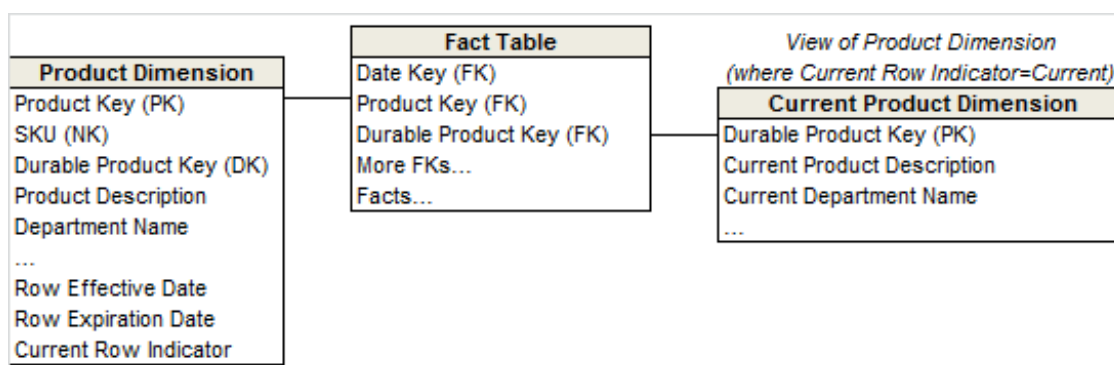## Type 6: Add Type 1 Attributes to Type 2 Dimension

Type 6 builds on the type 2 technique by also embedding current attributes in the dimension so that fact rows can be filtered or grouped by either the type 2 value in effect when the measurement occurred or the attribute's current value. The type 6 moniker was suggested by an HP engineer in 2000 because it's a type 2 row with a type 3 column that's overwritten as a type 1; both 2 + 3 + 1 and 2 x 3 x 1 equal 6. With this approach, the current attributes are updated on all prior type 2 rows associated with a particular durable key, as illustrated by the following sample rows:

Original row in Product dimension:

| Product Key | SKU (NK) | Product Description | Historic Department Name | Current Department Name | Row Effective Date | Row Expiration Date | Current Row Indicator |
|---|---|---|---|---|---|---|---|
| 12345 | ABC922-Z | IntelliKidz | Education | Education | 2012-01-01 | 9999-12-31 | Current |

Rows in Product dimension following first department reassignment:

| Product Key | SKU (NK) | Product Description | Historic Department Name | Current Department Name | Row Effective Date | Row Expiration Date | Current Row Indicator |
|---|---|---|---|---|---|---|---|
| 12345 | ABC922-Z | IntelliKidz | Education | Strategy | 2012-01-01 | 2012-12-31 | Expired |
| 25984 | ABC922-Z | IntelliKidz | Strategy | Strategy | 2013-01-01 | 9999-12-31 | Current |

Rows in Product dimension following second department reassignment:

| Product Key | SKU (NK) | Product Description | Historic Department Name | Current Department Name | Row Effective Date | Row Expiration Date | Current Row Indicator |
|---|---|---|---|---|---|---|---|
| 12345 | ABC922-Z | IntelliKidz | Education | Critical Thinking | 2012-01-01 | 2012-12-31 | Expired |
| 25984 | ABC922-Z | IntelliKidz | Strategy | Critical Thinking | 2013-01-01 | 2013-02-03 | Expired |
| 31726 | ABC922-Z | IntelliKidz | Critical Thinking | Critical Thinking | 2013-02-04 | 9999-12-31 | Current |

**Type 7: Dual Type 1 and Type 2 Dimensions**

With type 7, the fact table contains dual foreign keys for a given dimension: a surrogate key linked to the dimension table where type 2 attributes are tracked, plus the dimension's durable supernatural key linked to the current row in the type 2 dimension to present current attribute values.

Type 7 delivers the same functionality as type 6, but it's accomplished via dual keys instead of physically overwriting the current attributes with type 6. Like the other hybrid approaches, the current dimension attributes should be distinctively labeled to minimize confusion.

**Product Dimension**
Product Key (PK)
SKU (NK)
Durable Product Key (DK)
Product Description
Department Name
...
Row Effective Date
Row Expiration Date
Current Row Indicator

**Fact Table**
Date Key (FK)
Product Key (FK)
Durable Product Key (FK)
More FKs...
Facts...

View of Product Dimension
(where Current Row Indicator=Current)
**Current Product Dimension**
Durable Product Key (PK)
Current Product Description
Current Department Name
...

Finally, here's a figure from the *The Data Warehouse Toolkit, Third Edition* that highlights the implications of each slowly changing dimension technique on the analysis of fact table performance metrics. As we've warned in the past, there's more to consider than simply 1, 2 and 3!

| SCD Type | Dimension Table Action | Impact on Fact Analysis |
|---|---|---|
| Type 0 | No change to attribute value | Facts associated with attribute's original value |
| Type 1 | Overwrite attribute value | Facts associated with attribute's current value |
| Type 2 | Add new dimension row for profile with new attribute value | Facts associated with attribute value in effect when fact occurred |
| Type 3 | Add new column to preserve attribute's current and prior values | Facts associated with both current and prior attribute alternative values |
| Type 4 | Add mini-dimension table containing rapidly changing attributes | Facts associated with rapidly changing attributes in effect when fact occurred |
| Type 5 | Add type 4 mini-dimension, along with overwritten type 1 mini-dimension key in base dimension | Facts associated with rapidly changing attributes in effect when fact occurred, plus current rapidly changing attribute values |
| Type 6 | Add type 1 overwritten attributes to type 2 dimension row, and overwrite all prior dimension rows | Facts associated with attribute value in effect when fact occurred, plus current values |
| Type 7 | Add type 2 dimension row with new attribute value, plus view limited to current rows and/or attribute values | Facts associated with attribute value in effect when fact occurred, plus current values |