# Slowly Changing Dimensions [ SCD]

# SUGGESTED SOLUTION

|  |
| --- |
| **QUICK STUDY** |
| Business analysts need to track changes in dimension attributes. Reevaluation of the customer marketing segment is an example of what might prompt a change. There are three fundamental techniques.Type 1 is most appropriate when processing corrections; this technique won’t preserve historically accurate associations. The changed attribute is simply updated (overwritten) to reflect the most current value.With a type 2 change, a new row with a new surrogate primary key is inserted into the dimension table to capture changes. Both the prior and new rows contain as attributes the natural key (or durable identifier), the most-recent-row flag and the row effective and expiration dates.With type 3, another attribute is added to the existing dimension row to support analysis based on either the new or prior attribute value. This is the least commonly needed technique |

<http://www.kimballgroup.com/2005/03/10/slowly-changing-dimensions-are-not-always-as-easy-as-1-2-3/>

Using a lecture notes to help you with this tutorial, to decide which SDC type would you use for each task,1, 2 or 3.

# Scenario 1:

Orions' development team want to understand the implications of handling data changes in their data warehouse by looking at the impact on CustomerStar1.



Orion has been bought up by a Star Company and they would like to use the new companies name to indicate a type of customers they have, as shown in the table below.

|  |  |
| --- | --- |
| **customer\_type** | **customer\_type** |
| Orion Club members | Star Club members |
| Orion Club Gold members | Star Club Gold members |
| Internet/Catalog Customers | Start Internet/Catalog Customers |

These updates have been made in a customer type table, within the operational database, and it looks like this:

|  |  |
| --- | --- |
| ID | Customer TypeGroup |
| 1010 | Star Club members |
| 1020 | Star Club members |
| 1030 | Star Club members |
| 1040 | Star Club members |
| 2010 | Star Club Gold members |
| 2020 | Star Club Gold members |
| 2030 | Star Club Gold members |
| 3010 | Star Internet/Catalog Customers |

In the CustomerStar1, CustomerDim table data currently looks like this, and is soon going to have data loads coming through of the “new” customer type…

|  |  |  |  |
| --- | --- | --- | --- |
| **CustomerID** | **Name** | **customer\_type** | **Location** |
| **4** | **James Kvarniq** | **Orion Club members** | **US** |
| **13** | **Markus Sepke** | **Orion Club Gold members** | **DE** |
| **16** | **Ulrich Heyde** | **Internet/Catalog Customers** | **DE** |

Management has decided that they will use the new customer type names throughout the data warehouse and for ALL reporting from the warehouse.

# SUGGESTED SOLUTION

* Which Type of SCD should they use fro the Customer Dim? TYPE 1
* Why is this (????) an appropriate type? MANAGEMENT WANT TO SE ONLY THE NEW NAMES THROUGHOUT THE DW AND IT’S REPORTING. OLD NAMES NOT NEEDED TO BE KEPT
* Give an example of how would the data look like in a customer dim table

|  |  |  |  |
| --- | --- | --- | --- |
| **CustomerID** | **Name** | **customer\_type** | **Location** |
| **4** | **James Kvarniq** | **Star Club members** | **US** |
| **13** | **Markus Sepke** | **Star Club Gold members** | **DE** |
| **16** | **Ulrich Heyde** | **Star Internet/Catalog Customers** | **DE** |

* What are advantages and disadvantages using this SCD type?

Advantages – this change will not have an effect on the business on day-to-day basis; and change is quick.

Disadvantages - they lose the ability to see old and new customers, could have an impact on deciding customers type.

# Scenario 2:

In this scenario, as in the first, we are working with same example where Orion has been bought up by a Star Company and they would like to use the new companies name to indicate a type of customers they have, as shown in the table below.

|  |  |
| --- | --- |
| **customer\_type** | **customer\_type** |
| Orion Club members | Star Club members |
| Orion Club Gold members | Star Club Gold members |
| Internet/Catalog Customers | Start Internet/Catalog Customers |

However, this time management has decided that they would like to keep the ability to report by the old company name in order to monitor possible effects and the success of their newly acquired business.

In the CustomerStar1, CustomerDim table data currently looks like this, and is soon going to have data loads coming through of the “new” customer type…

|  |  |  |  |
| --- | --- | --- | --- |
| **CustomerID** | **Name** | **customer\_type** | **Location** |
| **4** | **James Kvarniq** | **Orion Club members** | **US** |
| **13** | **Markus Sepke** | **Orion Club Gold members** | **DE** |
| **16** | **Ulrich Heyde** | **Internet/Catalog Customers** | **DE** |

# SUGGESTED SOLUTION

* Which Type of SCD should they use fro the Customer Dim? TYPE 3
* Why is this (????) an appropriate type?
  + The old values till important in order to monitor possible effects of their business. Type 3 gives the opportunity to hold a pre-, and post- changed name.
* Give an example of how would the data look like in a CustomerDim table, CustomerStart1, to facilitate reporting needs.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **CustomerID** | **Name** | **customer\_type** | **Old\_customer\_type** | **Effective\_date** | **Location** |
| **4** | **James Kvarniq** | **Star Club members** | **Orion Club members** | **1/1/2015** | **US** |
| **13** | **Markus Sepke** | **Star Club Gold members** | **Orion Club Gold members** | **1/1/2015** | **DE** |
| **16** | **Ulrich Heyde** | **Star Internet/Catalog Customers** | **Internet/Catalog Customers** | **1/1/2015** | **DE** |

Good reporting tools would give columns on graphs the correct label. E.g graphs that discuss data that relates to before the change could have a column title of “Orion Club members”. The same graph, but predicting data the year after could have a column title of “Star Club members”. And a graph that had data for all years could have a column title of Orion Club members / Star Club members”.

Note that Type 3 doesn’t cope with more than one change. That’s why we didn’t use this example in the scenario 1.

* Has Fact table been changed to accommodate the new requirements?
  + It hasn’t…
* What are advantages and disadvantages using this SCD type?

Advantages: No change to dimension or fact table keys. Small changes needed to make queries to list new & old names.

Disadvantages: Type 3 can record only single change

# Scenario 3:



Changes in the products sold have made the current Product Group structure inappropriate for analysis and management purposes. The following shows the structure and sample content of the Product dimension table.

|  |  |  |  |
| --- | --- | --- | --- |
| Product\_ID | Product\_Category | Product\_Group | Product\_Name |
| 101 | Children Sports | A-Team, Kids | Kids Sweat Round Neck |
| 440 | Outdoors | Backpacks | Sioux Men's Backpack 26 Litre. |
| 650 | Golf | Golf | Rosefinch Cart 8 1/2 Black |
| 770 | Sports | Eclipse, Adult Clothes | Hsc Dutch Player Shirt |

The following listing shows some of the fact table data associated with the Product data.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Customer\_ID | Time\_ID | Order\_ID | Product\_ID | Quantity | Total\_Retail\_Price |
| 43 | 101 | 11 | 101 | 11 | $16.50 |
| 13 | 102 | 12 | 101 | 11 | $247.50 |
| 16 | 103 | 13 | 440 | 11 | $28.30 |
| 17 | 104 | 14 | 440 | 21 | $32.00 |
| 13 | 105 | 15 | 770 | 31 | $63.60 |
| 16 | 106 | 16 | 650 | 11 | $234.60 |
| 43 | 107 | 14 | 650 | 11 | $35.40 |

Management has decided to move to the new Product Group structure from the start of the next calendar year for new business but wish to retain all existing data according to the old Product Groups.

|  |  |  |
| --- | --- | --- |
| Product\_Category | NEWProduct\_Category | Product\_Group |
| Children Sports | Sports | A-Team, Kids |
| Outdoors | Outdoors | Backpacks |
| Golf | Outdoors | Golf |
| Sports | Sports | Eclipse, Adult Clothes |

# Suggested Answers:

* Which Type of SCD should they use fro the Product Dim? Type 2
* Why is this (????) an appropriate type? Wish to retain all existing data .
* Give an example of how would the data look like in **a Product Dim** table, ProductSales1, to facilitate reporting needs.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Surrogate | Product\_ID | Product\_Category | Product\_Group | Product\_Name | Effective Date |
| 1 | 101 | Children Sports | A-Team, Kids | Kids Sweat Round Neck | 01/01/0001 |
| 2 | 101 | Sports | A-Team, Kids | Kids Sweat Round Neck | **01/01/2015** |
| 3 | 440 | Outdoors | Backpacks | Sioux Men's Backpack 26 Litre. | 01/01/0001 |
| 4 | 650 | Golf | Golf | Rosefinch Cart 8 1/2 Black | 01/01/0001 |
| 5 | 650 | Outdoors | Golf | Rosefinch Cart 8 1/2 Black | **01/01/2015** |
| 6 | 770 | Sports | Eclipse, Adult Clothes | Hsc Dutch Player Shirt | 01/01/0001 |

* In what way has the fact table is changed to accommodate these new requirements?

Draw the table showing its structure and data content.

Fact table key for product id is now a surrogate key instead, the prod id

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Customer\_ID | Time\_ID | Order\_ID | Product\_ID | Quantity | Total\_Retail\_Price |
| 43 | 101 | 11 | 1 | 11 | $16.50 |
| 13 | 102 | 12 | 1 | 11 | $247.50 |
| 16 | 103 | 13 | 3 | 11 | $28.30 |
| 17 | 104 | 14 | 3 | 21 | $32.00 |
| 13 | 105 | 15 | 6 | 31 | $63.60 |
| 16 | 106 | 16 | 4 | 11 | $234.60 |
| 43 | 107 | 14 | 4 | 11 | $35.40 |

And then the Facts table now points to the Surrogate keys. At the time the sale was made, the product being sold is describable by line 4 of the Product Dim table.

If an analyst wants to report on **exactly** “product 4” then that’s easy. But if they want to report on all versions of the product the DW has to first look up what the Product ID for this product (it is “650”), and then report on all the sales of Product 650.

So the DW has to translate the query from the Product ID to Product No.

This sounds messy. But the DW tool will do this transparently, so it is only messy for the tool writer, not the tool users.

* It is now required to add a totally new transaction to the ProductSales1 fact table. This is for Golf, Time\_Id = 108, Cust\_id = 63, Order\_id = 17, Quantity = 12, Total\_Retail\_Price = $234.60

556.00. Show how this data will look when added to the fact table.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Customer\_ID | Time\_ID | Order\_ID | Product\_ID | Quantity | Total\_Retail\_Price |
| 43 | 101 | 11 | 1 | 11 | $16.50 |
| 13 | 102 | 12 | 1 | 11 | $247.50 |
| 16 | 103 | 13 | 3 | 11 | $28.30 |
| 17 | 104 | 14 | 3 | 21 | $32.00 |
| 13 | 105 | 15 | 7 | 31 | $63.60 |
| 16 | 106 | 16 | 5 | 11 | $234.60 |
| 43 | 107 | 14 | 5 | 11 | $35.40 |
| **63** | **108** | **17** | **4** | **12** | **$234.60** |

* What are advantages and disadvantages using this SCD type.

The company will be able to see all existing data according to the old and new product group. To do this they will have to have SCD built into the system of the DW.

Investigated how is SCD supported in SAS.

* To what degree can the analysts perform year on year comparisons by Product and/or Product Group?

SCD 2 gives full reporting to the analysts.