SCD - Slowly Changing Dimensions

by Sudheer Sharma - Tuesday, July 28, 2009

http://dwhnotes.com/data-warehousing/scd

What do you think of this post?

Awesome (10) Interesting (6) Useful (11)

Implementing SCD's is common concern in Data Warehouse design. Let me brief about different types of SCD's first then we can step in to SCD's design.

Dimensions that change over time are called Slowly Changing Dimensions. For instance, a product price changes over time; People change their names for some reason; Country and State names may change over time. These are a few examples of Slowly Changing Dimensions since some changes are happening to them over a period of time.

Everybody knows that 'Slowly Changing Dimensions' is the acronym for SCD. We have 4 types of SCD's in Data Warehouse

Let's say I have a customer dimension with these columns mainly (Customer Id, Customer First Name, Customer Last Name, Customer Country)

Customer Id	Customer First Name	Customer Last Name	Customer Country
1	Sudheer	Sharma	India

Now, this guy moved to US. In source the country name has been changed to US, we need to update that in our target dimension to reflect this change.

SCD Type 1: The new incoming record (changed/modified data set) replaces the existing old record in target.

Customer Id	Customer First Name	Customer Last Name	Customer Country
1	Sudheer	Sharma	US

Old value (India) is overwritten by the new value (US) and there is no way to find out the old version of data. It holds only the current version of data.

SCD Type 2: In this case, an additional record is added into the customer dimension. The beauty of this approach is it will maintain two versions, you will find two records the older version and the current version. In other words it maintains history. Again we can implement Type 2 in following methods

1. Versioning
2. Effective Dates
3. By setting Current Flag values/Record Indicators.

Method 1: Versioning

Customer Id	Customer First Name	Customer Last Name	Customer Country	Effective Year	Version
1	Sudheer	Sharma	India	2008	0
1	Sudheer	Sharma	US	2009	1

Method 2: Effective Dates

Customer Id	Customer First Name	Customer Last Name	Customer Country	Effective Start Date	Effective EndDate
1	Sudheer	Sharma	India	01/01/2008	12/31/2008
1	Sudheer	Sharma	US	01/01/2009	tilldate

Method 3: Effective Dates & Current Record Indicators

Customer Id	Customer First Name	Customer Last Name	Customer Country	Effective Start Date	Effective End Date	Current Record IND
1	Sudheer	Sharma	India	01/01/2008	12/31/2008	N
1	Sudheer	Sharma	US	01/01/2009	tilldate	Y

SCD Type 3: In this approach, only the information about a previous value of a dimension is written into the database. An ‘old ‘or ‘previous’ column is created which stores the immediate previous attribute.

Product ID	Product Name	Current Year	Current Price	Previous Year	Previous Price
1	Close-up	2008	50.00	2007	45.00

The problem with this approach is over years, if the product price continuously changes, then the complete history may not be stored, only the latest change will be stored. For example, in year 2009, if the product price changes to 60, then we would not be able to see the complete history of 2007 prices, since the old values would have been updated with 2008 information.

SCD Type 4: In this approach, one table hold current data and another table keeps historical data for each

dimension.

Customer Dimension

Customer Id	Customer First Name	Customer Last Name	Customer Country
1	Sudheer	Sharma	US

Customer History Table

Customer Id	Customer First Name	Customer Last Name	Customer Country	Effective Year
1	Sudheer	Sharma	India	2008

What do you think of this post?

Awesome (10) Interesting (6) Useful (11)
