

Syed Mushtaq DATA WAREHOUSE ASSESSMENT

Q1.Category of a product may change over a period of time .Historical category information (current categories as well as old categories) has to be stored.Which SCD type will be suitable to implement this requirement ?
What kind of structure changes are required in dimension table to implement SCD type 2 and type 3?

Ans. we can use **SCD TYPE 2** because it will store all the previous history information. When the value of a chosen attribute changes, the current record is closed. A new record is created with the changed data values and this new record becomes the current record. Each record contains the effective time and expiration time to identify the time period between which the record was active.

SCD TYPE 2:

The methodology is preserving the history by adding new rows and there is no limit on number of keeping changes. History is maintained for all the changes that an attribute is undergoing.

CustomerKey	CustomerId	CustomerName	City	BirthDate
1003	5873	John Smith	Birmingham	28/02/1978

Current Table

CustomerKey	CustomerId	CustomerName	City	BirthDate	StartDate	EndDate
1001	5873	John Smith	London	28/02/1978	01/01/1970	27/12/2004
1002	5873	John Smith	Leeds	28/02/1978	28/12/2004	08/01/2017
1003	5873	John Smith	Birmingham	28/02/1978	09/01/2017	31/12/9999

After Changes .

We should create (ADD) two columns as Start_date and End_date to keep the history of previous information.

SCD TYPE 3:

In this approach you can manage a limited amount of history by adding additional columns for tracking attributes.

Wanting to have previous and current city of customer you can implement columns: PreviousCity and CurrentCity. There will be only one row in Data Warehouse (DW) per source database therefore Type 3 has limited usability and is less popular than Type 1 and 2.

CustomerKey	CustomerId	CustomerName	PreviousCity	CurrentCity	BirthDate	UpdateDate
1001	5873	John Smith	London	London	28/02/1978	27/12/2004
1002	5873	John Smith	London	Leeds	28/02/1978	08/01/2017
1003	5873	John Smith	Leeds	Birmingham	28/02/1978	08/01/2017

here we have current and previous city.

Q2 .What is surrogate key? Why it is required?

Ans . A surrogate key is a key which does not have any contextual or business meaning. It is manufactured “artificially” and only for the purposes of data analysis. The most frequently used version of a surrogate key is an increasing sequential integer or “counter” value (i.e. 1, 2, 3). Surrogate keys can also include the current system date/time stamp, or a random alphanumeric string.

There are various reasons why we cannot simply reuse our existing natural or business keys.

First, Business keys usually have a business meaning in the OLTP systems, such as social security number for Employee and VAT number for Company. Hence, they are tied to the business setting and requirements and if these would change (e.g., due to a merger or acquisition, a new legislation) then all tables using those keys need to be updated, which may be a resource intensive operation in a data warehouse environment, because not only the current state is stored, but also historical data.

Business keys are usually bigger in size which will result in big indexes and slow down index traversal and, consequently, query execution time. Therefore, using surrogate keys will save space and improve performance.

Q3. Stores are grouped in to multiple clusters . A store can be part of one or more clusters .Design tables to store this store-cluster mapping info?

Ans Stores:

Storeid	Locid	Storename	Sizeid	ReveueId	Categoryid
101	1	Levis	102	1013	702
102	2	Reebok	102	1013	703
103	2	KFC	101	1012	701
104	2	Burger King	101	1011	701

Location Cluster :

Locid	LocName
1	Bangalore North
2	Bangalore South

Size Cluster :

sizeid	SizeRange
101	Less than 500 SqFt.
102	More than 500 SqFt.

Revenue Cluster :

Revenueid	Revenue Range (Yearly)
1011	Less than 1 Million
1012	Less than 10 Million
1013	Less than 50 Million

Category Cluster :

Categoryid	Category_Name
701	Food and Beverages
702	Clothes
703	Shoes

Q4. What is semi-additive measures? Give an Example ?

Ans . Semi Additive measures are values that we cannot be summarised .

But some analytical functions can be used .

One of the common Example is date .

Another Example is account balance.

Transaction_ID	Customer_ID	Date	Account_No	Transaction Type	Balance_Amount
12654	727598456	3/1/2015	0005437675423	Credit	20000
12655	727598456	3/1/2015	0005437675423	Debit	18000
12656	727598456	3/5/2015	0005437675423	Credit	21000
12657	727598456	3/5/2015	0005437675423	Debit	15000
12658	727598456	3/5/2015	0005437675423	Credit	32000
12659	727598456	3/5/2015	0005437675423	Debit	10000

