

DATAWAREHOUSING ASSIGNMENT-2

DATAWAREHOUSING ASSIGNMENT-2

1.

a) Category of a product may change over a period of time. Historical category information

(current category as well as all old categories) has to be stored. Which SCD type will be suitable

to implement this requirement? What kind of structure changes are required in a dimension

table to implement SCD type 2 and type 3.

ANS - For the above scenario SCD type-2 is suitable

Justification - In SCD type-2 Dimension, a new record is added to the table to represent the new information. Therefore, both the original and the new record will be present. The new record gets its own primary key. This allows us to accurately keep all historical information. History is maintained for all the changes that an attribute is undergoing

SCD Type-2

Original Table

| Customer_ID | Customer_name | City | Phone_num |
|-------------|---------------|----------|------------|
| 1 | Suraj | Banglore | 7578578954 |
| 2 | Kartik | Mysore | 7352525527 |
| 3 | Udit | Mumbai | 7352345235 |
| 4 | Virat | Delhi | 7523552527 |

| Customer_ID | Customer_name | City | Phone_num | Start_date | End_date |
|-------------|---------------|-----------------|-------------------|------------------|-----------|
| 1 | Suraj | Banglore | 7578578954 | 5/7/2017 | |
| 2 | Kartik | Mysore | 7352525527 | 17/9/2017 | |
| 3 | Udit | Mumbai | 7352345235 | 18/1/2016 | |
| 4 | Virat | Delhi | 7523552527 | 14/5/2017 | 23/7/2017 |
| 5 | Virat | Banglore | 7523552527 | 23/7/2017 | |

Start_date and End_date are new version columns

5 th row is new added row with latest data

DATAWAREHOUSING ASSIGNMENT-2

SCD TYPE-3

Original Table

| Customer_ID | Customer_name | City | Phone_num |
|-------------|---------------|----------|------------|
| 1 | Suraj | Banglore | 7578578954 |
| 2 | Kartik | Mysore | 7352525527 |
| 3 | Udit | Mumbai | 7352345235 |
| 4 | Virat | Delhi | 7523552527 |

| Customer_ID | Customer_name | Current_City | Previous_city | Phone_num |
|-------------|---------------|--------------|---------------|------------|
| 1 | Suraj | Banglore | | 7578578954 |
| 2 | Kartik | Mysore | | 7352525527 |
| 3 | Udit | Mumbai | | 7352345235 |
| 4 | Virat | Banglore | Delhi | 7523552527 |

B) What is surrogate key? Why it is required?

ANS-

Surrogate key - A surrogate key is a unique identifier used in databases for a modeled entity or an object. It is a unique key whose only significance is to act as the primary identifier of an object or entity and is not derived from any other data in the database and may or may not be used as the primary key. The usual surrogate key used is a unique sequential number

It is required because

- > With help of surrogate keys, you can integrate heterogeneous data sources to data warehouse if they don't have natural or business keys.
- > Joining tables (fact and dimensions) using surrogate key is faster hence better performance
- > Surrogate keys are very helpful for ETL transformations.
- > Data warehouse Surrogate keys are usually small integer numbers that makes smaller index and better performance
- > Surrogate keys are required if you are implementing SCD(Slowly Changing Dimensions)

DATAWAREHOUSING ASSIGNMENT-2

C) 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 information.

Consider clothing brands like Lee cooper, Levis, Allen solly, united colors of benetton which maybe present in different stores in a city like Reliance trends, Brand Factory, Mega mart and westside and these stores may be present in multiple locations like Bangalore, Mumbai, Delhi and Chennai. So for example Lee cooper may be present in two different stores in bangalore and in one store in mumbai.

Location table

| Location_Key | Location_Name |
|--------------|---------------|
| 100 | Bangalore |
| 101 | Mumbai |
| 102 | Delhi |
| 103 | Chennai |

Store table

| Store_Key | Store_Name |
|-----------|-----------------|
| 100 | Reliance trends |
| 101 | Brand Factory |
| 102 | Westside |
| 103 | Mega mart |

Product table

| Product_Key | Product_Name |
|-------------|---------------------------|
| 200 | Lee Cooper |
| 201 | Levis |
| 202 | Allen Solly |
| 203 | United Colors of benetton |

Mapping Table

| Mapping_Key | Product_Key | Store_key | Location_Key |
|-------------|-------------|-----------|--------------|
| 1 | 200 | 100 | 100 |
| 2 | 200 | 103 | 100 |
| 3 | 200 | 101 | 101 |
| 4 | 201 | 102 | 102 |

DATAWAREHOUSING ASSIGNMENT-2

D) What is a semi-additive measure? Give an example

ANS-

Semi-additive measures - Semi-additive measures are the measures can be summed across some dimensions, but not all.

Example-

1. checking account or savings account balance amounts are common semi-additive facts. You can recreate a balance amount from the transactions file, but it doesn't make any sense to add the balance amounts from October, November, and December (across the time dimension).
2. You can aggregate department headcounts to give an organization total, but you cannot aggregate them over time, so the Sales department headcount for March 31 maybe 20 employees, and for April 30 the headcount maybe 23, but that does not mean that the total headcount at the end of April is 43