

Study of Slowly Changing Dimension to Dependent Data Mart to Manage Finance Data

Arvind Singh¹, Surya Prakash Pandey²

^{1,2}Department of Computer Science, Awadhesh Pratap Singh University Rewa Madhya Pradesh, India

ABSTRACT

Article Info

Volume 7, Issue 6

Page Number : 190-195

Publication Issue :

November-December-2021

Article History

Accepted : 01 Dec 2021

Published : 07 Dec 2021

Financial institutions face many challenges of managing and marketing campaigns which leads in its data warehouse. The management of marketing campaign leads in dependent data mart with real time updating and recording difficulties especially when many campaigns are running parallel ways. To securing the customers from being contacted too often for sales-based marketing contacts, the concept of novelty skeleton are introduced to clamp the customers who have been targeted in Sales based campaign for a specified time period. During the novelty Frame, the customer cannot be targeted by other Sales based campaign categorized under the same channel. The introduction of novelty skeleton has increased the difficulties of campaign management and data management. The difficulties of data management include timely update and robust storage system of campaign leads. In this paper, we explained represent the concept of slowly changing dimension on dependent data mart and also studied how it can be used in the data mart of financial institutions to update and maintain marketing campaign records of customers.

Keywords : Data Mart, Dependent Data Mart, SCD, marketing campaigns, DCHT

I. INTRODUCTION

The Financial institutions generate huge volume of data daily due multiple transactions and operations. These data are analyzed by the financial institutions such as bank to provide valuable customer insights such as the customer behavior, needs and desires. Unexpected ways of campaign and management may result in customer exasperation. Robust and well-managed data Mart that stores and tracks customer contacts history which can assist in effective campaign management. However, with the set of activities and

presence of large amount and volume of data, the institutions face lot of difficulty to drive the dependent data mart.

We can think two main categories of campaigns in the financial institutions which are based on sales and non-sales in nature.

The sales based campaigns target selected set of customers. A campaign leads to market for selling financial institution's products and their services. Non-sales based campaigns serve and provide

customers with fulfillment and notification services. These campaigns are break into two types of medium for campaign distribution- calling and non-calling types [5]. The calling types include telemarketer call, relationship manager call and branch officer calls. Non-calling types include direct mailer, Email blast, electronic direct mailer and short messages [5]. Analysts select and target the most appropriate segments of customers for marketing campaigns from the dependent data mart. Below figure shows that the source data goes to the data warehouse for storage purposes then it create the data mart.

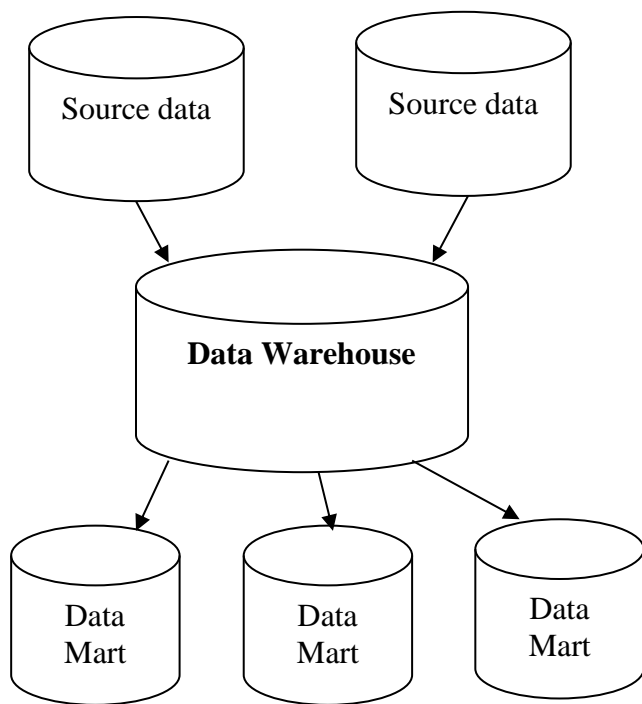


Figure 1.1 Dependent Data Mart

Dependent data marts are repositories that are controlled by data warehouse and source. Quite often, a data warehouse has always connected with data marts that reside in other business units. Even metadata (data about data) for the data mart may be dependent from metadata captured for data marts by other business units.

Source data for the dependent data mart is extract from operational systems as well as systems that may be external to the business unit for which the data

mart constructed. Source data may also be extracted for Enterprise Data Warehouse· if any exists [4].

Customers who are targeted for many campaigns may be over contacted during a period of time which may take result in unhappy from the institution and may spread bad experience. Proper handling process is needed to ensure that the result does not take place unhappy experience.

A powerful data management process will rule an efficient data operating process to manage present and historical data over time in the data mart for any industry.

Slowly Changing Dimension (Kimball, 2008) is the name of a data management process that loads data into multiple dimensions of tables which contains data. To study and adopting Slowly Changing Dimension, the data has to change slowly on an irregular and random schedule. There are three kinds of Slowly Changing Dimension methods, Type 1, Type 2, and Type 3. The most commonly practiced SCD types are 1, 2 and 3. Below we present the descriptions of the different Types of SCD [5].

Type 1: Type 1 SCD overwrites old records with new records. Type 1 SCD is easy to maintain. However, no historical observations are kept in the data warehouse [5].

Type 2: Type 2 SCD updates the record by inserting new observation while preserving the historical observations. Unlimited historical observations can be preserving with this type [5].

Type 3: Type 3 SCD updates the record by creating new dimensions to the table structure. It preserves history - only the previous record could be preserve [5].

The observation covers Type 1 SCD by updating the latest record.

The type of SCD to implement largely depends on the business and institutional requirements with environment. SCD TYPE 2 SCD Type 2 puts new record to the existing available dimension table and leaves the old records. Its advantage is that it allows historical records to be easily tracked.

In SCD Type 2, surrogate keys or/and version numbers are introduced to track historical records where surrogate keys are artificial primary keys in sequential numbers. In a scenario where customer's passport number is changed from time to time, SCD Type 2 preserves the old record while putting new records to the existing table as shown below. In this scenario, Customer Key is a surrogate key in sequence and Customer Name is the primary key [5].

Customer Key Id	Customer Name	Passport Number
202001	Rajesh Singh	20XPNA001
202002	Ram Prakash	20XPNA020

Table 1.1

Customer Key Id	Customer Name	Passport Number	Effective Start Date	Effective End Date	Convenient Flag
202001	Rajesh Singh	20PN001	1 February 2006	13 June 2013	0
202002	Ram Prakash	20PN020	14 July 2013	null	1

Table 1.2

In a financial institution prospective, the financial institution is required to retain more customer information for a many number of years as per requirement. SCD Type 2 is able to retaining endless number of historical records and present records hence meeting the requirement needed by the institution.

Some other variables/fields could be introduced to the table for tracking purpose. The non-wide variables are the effective start date, the effective end date, and the convenient flag.

The effective start date is the date the change becomes effective and the effective end date is the date the change ceases to be effective.

And last, the Convenient Flag is a dummy variable/field in 0 or 1 (1 means most latest) that give a fast way to identifying the latest record. As based on the past example, we can see how these variables/fields are introduced to better management of records as shown below in Table 1.2. The new record with customer key 202002 has an unknown effective expiration/End date which is null in value.

In the Alternative, a standardized fugglemen high date (e.g. 30 Dec 9999) can be used as the end date.

SCD Type 2 is relatively simple in implementation. The ability to retrieve current and historical data with Type 2 is also useful in reporting purpose. In view of that, SCD Type 2 is widely used in scenarios such as capturing customers' information and supplier's information etc.

II. SCD STUDY IN FINANCE DATA

When customers are over contacted by the institution for marketing campaigns purpose and would leave the customer an unhappy financial experience. So, it is necessary for the financial institution like bank and insurance companies to implement marketing campaign novelty Rule for all sales-based campaigns purpose. The novelty skeletons will lock-on the customers who are targeted in Sales-based campaign

for a time period known as the novelty skeletons. During the novelty skeleton, the customers will not be targeted by other Sales-based campaign by the same medium. The novelty skeleton will be distinctive at campaign channel level.

Analysts select customers as marketing leads based on their campaigns' criteria. With the novelty Rules in place, analyst has to manually track customer's end of novelty date and drop customers who are still within the novelty skeleton from the new campaign leads. The painful process is not time efficient and would leave room for human negligence [5].

With the study of Slowly Changing Data, the contact history and present contact records of customers will be loaded onto a common dimension table Data Contact History Table (DCHT). The DCHT will be capable of capturing the contact records of institution-wide customers. The DCHT will embed the novelty Rule and the distinctive novelty skeleton of each campaign at channel level. The end of novelty date of each lead will be automatically populated in the DCHT.

Individual analysts load his or her marketing leads onto a table known as the Analyst Table for every new campaign. Records in the Analyst Table will be uploaded to the CHT. Leads that are still within the novelty skeleton will be released out of the DCHT. The Analysts does not need to perform manual checking which improves performance efficiency and discard human intervention. Improved operational processes allow for better control of the innovation framework of campaigns.

III. HOW IS SCD STUDY IN FINANCE DATA

As mentioned earlier, DCHT will link the novelty Rules of every campaign channel level. The novelty skeleton will be notable at campaign channel level. For non-calling campaign channels, the novelty skeleton is $(T + X)$ days, where T is the lead time blast date and X is the lapse time period in days. For calling campaign channels, the novelty Frame is determined by successful contact of the leads. In the prospective where lead has not been successfully contacted, the novelty Frame will be $(T + Y)$ days, where T is the lead time blast date and Y is the lapse time period in days. In the prospective where the lead has been successfully contacted, the novelty skeleton is $(Z+Y)$ days, where Z is the lead contacted date and Y is the lapse time period in days [5].

The DCHT checks on the status of the leads. Below are a few prospective of the leads:

- (1) Fresh leads which was not targeted in any past campaigns
- (2) Already availing leads which was targeted in past campaign and are out of novelty skeleton
- (3) Already availing leads which was targeted in past campaign and are still within the novelty skeleton

Existing channel of campaign differs from the newly targeted channel of campaign.

In prospective 1 and 2, where fresh leads which were not targeted in past campaigns and existing leads which were out of novelty skeleton will be successful uploaded onto DCHT. DCHT will only consist of actionable marketing leads. In prospective 3, where existing leads which were targeted in past campaigns and are still within the novelty skeleton will be categorized two sub-categories. Existing leads where present medium of campaign is different from the new targeted channel of campaign. These leads will be successful uploaded onto DCHT. Existing leads where

present medium of campaign is similar to the new targeted medium of campaign. These leads will not be uploaded onto DCHT. This is because novelty skeleton is distinctive at channel level. Figure 3.1 is explaining the derivation of DCHT from multiple Analyst Tables.

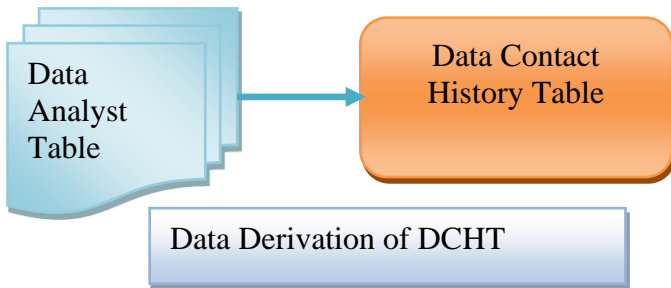


Figure 3.1

IV. METHODOLOGY

The Analyst Table is shown in Figure 3.1. The Analyst Table documented the campaign details such as: ID (Campaign ID), campaign information, medium of marketing (Channel Class), customer’s particulars and Analyst name. The financial institution runs multiple campaigns simultaneously. Different analyst might end up with similar leads on their campaigns selection. First Analyst take leads from DCHT. Second Analyst will then load leads from its Analyst Table onto the DCHT. Third the analyst who uploads the leads onto the DCHT then, gets to remain the customer in his/her base. And last Leads secured in DCHT which are based on the first come first serve processing.

Slowly Changing Dimension Type 2 updates the new campaign of already available leads by putting new observation while securing the historical observations. A new binary variable/field labeled as present is created to provide a fast way of identifying the most present record. The variable Current labels 1 as the latest observation and 0 as the past observations of campaign records.

The contact histories of customers are remaining as unlimited historical observations can be secured with

SCD type 2. After some time, the contact histories of customers are continuously built on a single database and may cause the database to spiral out of data space.

V. IMPROVEMENT

The SCD Type 2 is that the data is regularly built on a single type of database causing the database to expand fast way in overtime. My suggestion is that for diverting historical data onto an Archive Table and remaining current campaign in DCHT. Records that are 0 in the present variable/field will be archived into Archive Table. When Analysts load Analyst Table then the Archive Table will be updated onto DCHT. Figure 5.1 showing that the how improved data management are uploading leads onto the DCHT and Archive Table. And also Figure 5.1 shows the improved workflow of campaign lead and their selection process.



Figure 5.1 Improved Data Management

VI. CONCLUSION

Management of multiple campaigns records of a large amount of customer base is a difficult task for the financial institutions. Business rules such as novelty skeleton at campaign channels have compounded the level of difficulty.

DCHT is a common dimension table, Analysts uses DCHT to upload new campaign leads. Each Analyst creates a new leads to be captured onto Analyst Table for every new marketing campaign.

The study of SCD to data mart to manage the marketing campaigns in financial institution has improves the data management and campaign management efficiency effectively. The automated checks and updates of data mart reduce impossible tasks.

VII. REFERENCES

- [1]. Ralph Kimball (2008, September 22). Slowly Changing Dimensions, Part 2. Kimball Group. January 20, 2013
- [2]. Margy Ross (2008, September 29). Practical Steps for Designing a Dimensional Model. Kimball Group. January 20, 2013
- [3]. Shauna Trainor (2013, February 13). Market In Modesty To Prevent Resentment. Covenant Group. February 13,2013
- [4]. Inmon, W.H., Building the Data Warehouse, New York: Wiley Computer Publishing, Second Edition, 1996
- [5]. Lihui WANG, Junyu CHOY, Michelle L. F. CHEONG, "Implementation of Slowly Changing Dimension to Data Warehouse to Manage Marketing Campaigns in Banks", Research Collection School Of Computing and Information Systems, Paper No. 239, 2013

Cite this article as :

Arvind Singh, Surya Prakash Pandey, "Study of Slowly Changing Dimension to Dependent Data Mart to Manage Finance Data", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 6, pp. 190-195, November-December 2021. Available at doi : <https://doi.org/10.32628/CSEIT217657>
Journal URL : <https://ijsrcseit.com/CSEIT217657>