Paper No. 239- 2013

Implementation of Slowly Changing Dimension to Data Warehouse to Manage Marketing Campaigns in Banks

Wang Lihui
Murphy Choy
Michelle L.F. Cheong

ABSTRACT

Banks faces the challenges of managing marketing campaign leads in its data warehouse. The management of marketing campaign leads in data warehouse with real time updating and recording is complex especially when multiple campaigns are active simultaneously. To prevent the customers from being contacted too often for sales-based marketing contacts, the concept of Recency Frame is introduced to suppress the customers who have been targeted in Sales-based campaign for a specified time period. During the Recency Frame, the customer cannot be targeted by other Sales-based campaign categorized under the same channel. The introduction of Recency Frame 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 illustrate the concept of slowly changing dimension and how it could be utilized in the data warehouse of banks to update and maintain campaign records of customers.

INTRODUCTION

The banks generate huge volume of data daily due multiple transactions and operations. These data are analysed by the bank to provide valuable customer insights such as the customer behavior, needs and desires. Banks use such insights for strategic campaign management. Several marketing campaigns could be launched at any one point of time to promote the wide array of banking products and services. Unmethodical campaign management may result in customer resentment or backlash when customers are overly contacted by simultaneous marketing campaigns. Robust and well-managed data warehouse that stores and tracks customer contact history assist in effective campaign management. However, with the maelstrom of activities and presence of huge volume of data, the bank faces tremendous difficulty to manage the data warehouse.

There are 2 main categories of campaigns in the bank – sales based and non-sales based. The sales based campaigns target selected group of customers as leads to market and sell bank products and services. Non-sales based campaigns provide customers with notification and fulfillment service. The campaigns are divided into two types of channels for campaign distribution- calling and non-calling types. The calling types include telemarketer call, relationship manager call and branch officer calls. Non-calling types include direct mailer (DM), Email blast, electronic direct mailer (eDM) and short messages (SMS). Analysts select and target the most appropriate segments of customers for marketing campaigns. Customers who are targeted for many campaigns may be over contacted during a period of time which may result in unpleasant banking experience for the customers. Proper operating process is needed to ensure such a scenario does not occur.

A robust data management process will dictate an efficient data operating process to manage current and historical data over time in the data warehouse for any industry. The purpose of this paper is to present the innovative use of implementing Slowly Changing Dimension to the data management of recording and maintaining customers' contact history to facilitate contact management.

SLOWLY CHANGING DIMENSION (SCD)

Slowly Changing Dimension (Kimball, 2008) is the name of a data management process that loads data into dimension tables which contains data. To adopt SCD, the data has to change slowly on an irregular, random and variable schedule.

There are 6 current types of SCD methodologies, namely Type 0, Type 1, Type 2, Type 3, Type 4, Type 6. The

most commonly practiced SCD types are 1, 2 and 3. Below we present the descriptions of the different Types of SCD

Type 0 - Type 0 SCD does not update the changes in the data. Original values of the record remain in the dimension that was initially created.

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.

Type 2 – Type 2 SCD updates the record by inserting new observation while preserving the historical observations. Unlimited historical observations can be preserve with this type.

Type 3 – Type 3 SCD updates the record by creating new dimensions to the table structure. It preserves limited history - only the previous record could be preserve.

Type 4 – Type 4 SCD updates the record in the current data table and preserves all or some historical observations in an archive table.

Type 6 – Type 6 SCD is a hybrid of the methodologies of Type 1, 2 and 3. It incorporates Type 3 SCD by creating additional dimensions to the table structure while preserving historical observation. It incorporated Type 2 SCD by creating new dimensions to include the different version number. Type 1 SCD is incorporated by updating the latest record in the observation.

The type of SCD to implement largely depends on the business requirements and environment.

SCD TYPE 2

SCD Type 2, inserts new record to the existing dimension table and leaves the old records intact. The advantage to this is that it allows one to track historical record readily. 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 customers' passport number is changed from time to time, SCD Type 2 preserves the old record while inserting 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.

Customer Key	Customer Name	Passport Number
101	ABC	X1Y2000
102	ABC	A1D2002

Additional variables could be introduced to the table for tracking purpose. The non-exhaustive variables are Effective Start Date, Effective End Date and Convenient Flag. The Effective Start Date is the date where the change becomes effective and the Effective End Date is the date where the change ceases to be effective. Finally, the Convenient Flag is a dummy variable in 1 or 0 (1 as most current) that provides a rapid way of identifying the most current record. Based on the previous example, we can take a look at how these variables are introduced to better manage the records.

Customer Key	Customer Name	Passport Number	Effective Start Date	Effective End Date	Convenient Flag
101	ABC	X1Y2000	1 January, 2005	12 May, 2012	0
102	ABC	A1D2002	13 May, 2012		1

The new record with Customer Key 102 has an unknown Effective End Date which is in null value. Alternatively, a standardized surrogate high date (e.g. 31 May 9999) may be used as an 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 suppliers' information etc.

In a banking scenario, the bank is required to retain customer information for a number of years as per regulatory requirement. SCD Type 2 is capable of retaining unlimited number of historical records and current records thus meeting the requirement needed by the bank.

SCD IMPLEMENTATION IN BANKS

When customers are over-contacted by the bank for marketing campaigns would leave the customer an unpleasant banking experience. Therefore, it is necessary for the bank to implement marketing campaign Recency Rule to all sales-based campaigns. The Recency Framework will "lock" the customers who are targeted in Sales-based campaign for a time period known as the Recency Frame. During the Recency Frame, the customers will not be targeted by other Sales-based campaign of the same channel. The Recency Frame will be distinctive at campaign channel level.

Analysts select customers as marketing leads based on their campaigns' criteria. With the Recency Rules in place, analyst has to manually track customers' end of recency date and drop customers who are still within the Recency Frame from the new campaign leads. The painful process is not time efficient and would leave room for human negligence. Figure 1 shows the current operation process.

Current Operation Process

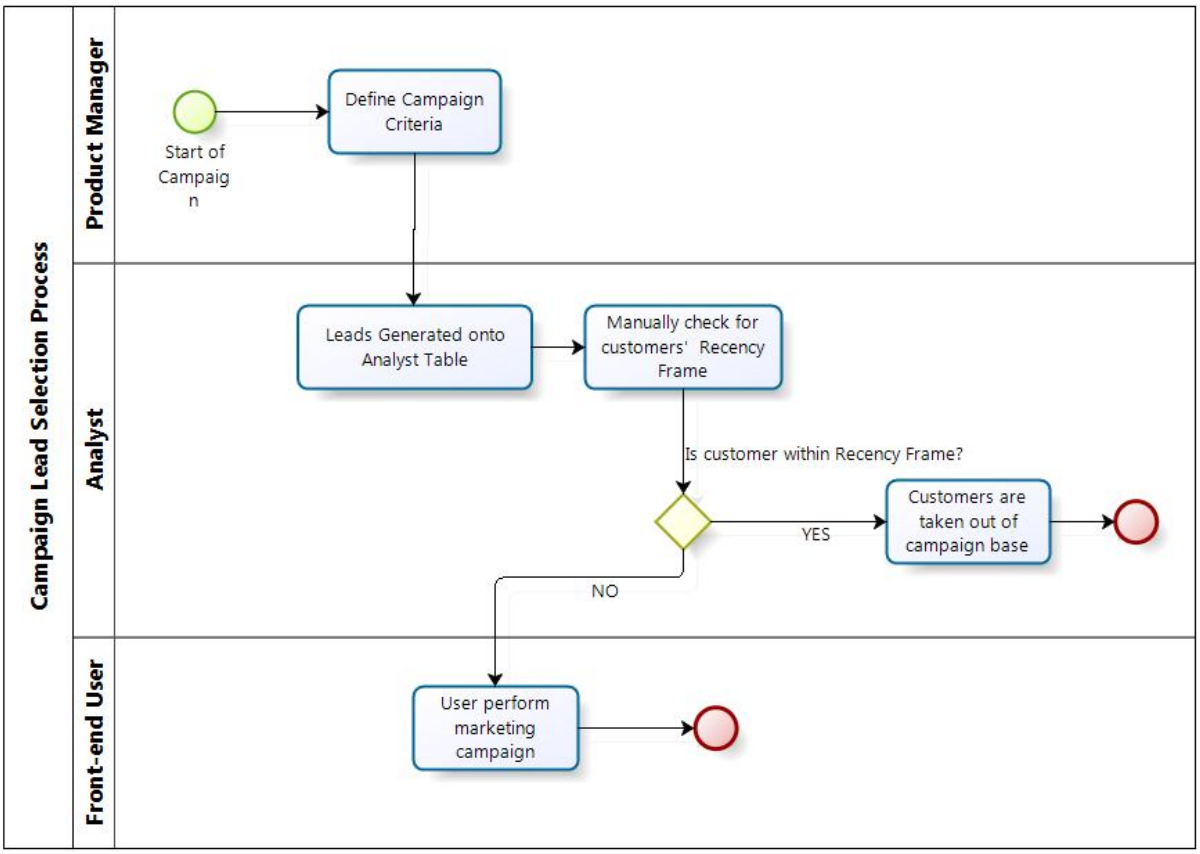

Figure 1 - Current Operation Process

With the implementation of SCD, the contact history and current contact records of customers will be loaded onto a common dimension table- Contact History Table (CHT). The CHT will be able capture the contact records of bank-wide customers. The CHT will embed the Recency Rule and the distinctive Recency Frame of each campaign at channel level. The end of recency date of each lead will be auto-populated in the CHT. 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 Recency Frame will be "thrown" out of the CHT. As such, the Analysts do not need to perform manual check which improves performance efficiency and eliminate human negligence. The improved operation process allows better control of the Recency Framework of campaigns and the process flow is shown in Figure 2.

Improved Operation Process

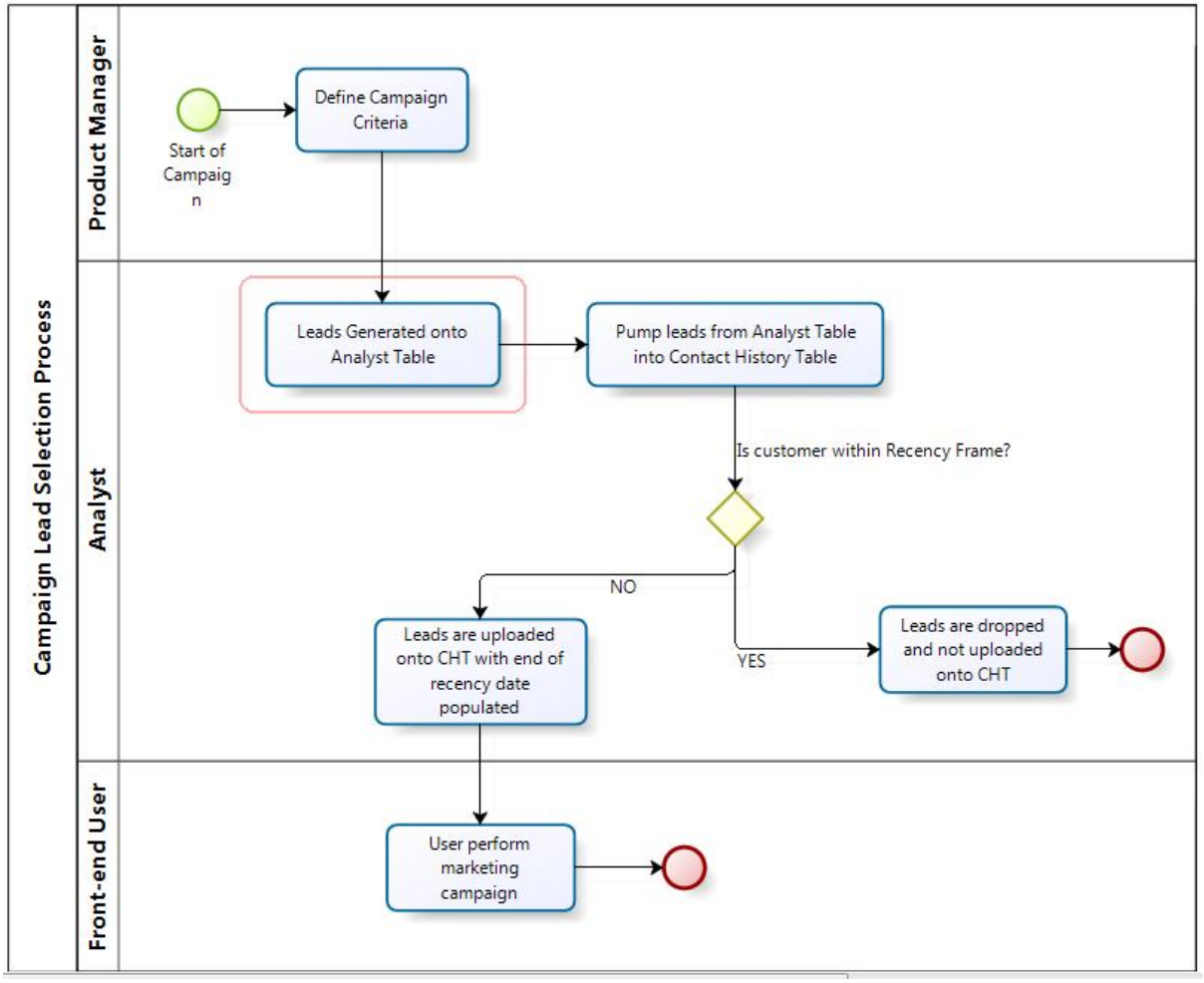

Figure 2 - Improved Operation Process

HOW IS SCD IMPLEMENTED

As mentioned in Section 4, CHT will embed the Recency Rules of each campaign channel level. The Recency Frame will be distinctive at campaign channel level. For non-calling campaign channels, the Recency Frame is (T + X) days, where T is the lead blast date and X is the lapse period in days. For calling campaign channels, the Recency Frame is determined by successful contact of the leads. In the scenario where lead has not been successfully contacted, the Recency Frame will be (T + Y) days, where T is the lead blast date and Y is the lapse period in days. In the scenario where the lead has been successfully contacted, the Recency Frame is (Z+Y) days, where Z is the lead contacted date and Y is the lapse period in days.

The CHT checks on the status of the leads. Below are a few scenarios of the leads:

1) Fresh leads who has not been targeted in any past campaigns
2) Existing leads who have been targeted in past campaign and are out of Recency Frame
3) Existing leads who have been targeted in past campaign and are still within the Recency Frame

   3.1) Existing channel of campaign differs from the newly targeted channel of campaign

   3.2) Existing channel of campaign is similar to the newly targeted channel of campaign

In scenario 1 and 2, where fresh leads who have not been targeted in past campaigns and existing leads who are out of Recency Frame will be successful uploaded onto CHT. CHT will only consist of actionable marketing leads. In scenario 3, where existing leads who have been targeted in past campaigns and are still within the Recency Frame will be categorized into 2 sub-categories. For existing leads where its current channel of campaign differs from the newly targeted channel of campaign as explained in 3.1, these leads will be successful uploaded onto CHT. For existing leads where its current channel of campaign is similar to the newly targeted channel of campaign as explained in 3.2, these leads will not be uploaded onto CHT. This is because Recency Frame is distinctive at channel level. Figure 3 will illustrate the derivation of CHT from multiple Analyst Tables.

Derivation of CHT

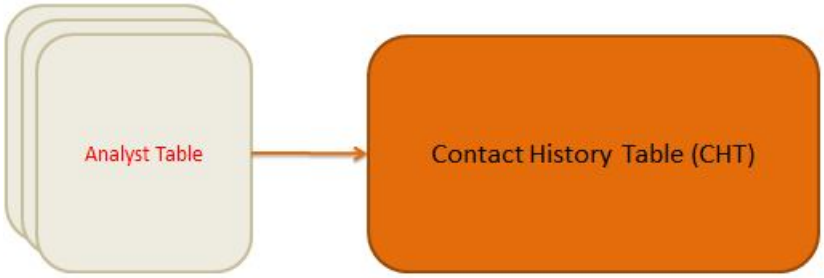

Figure 3- Derivation of CHT

METHODOLOGY

The Analyst Table is shown in Figure 4. The Analyst Table document the campaign details – campaign identification number (Campaign ID), campaign information, channel of marketing (Channel Class), customers' particulars and Analyst name. The bank runs multiple campaigns concurrently, different analyst might end up with similar leads on their campaigns selection. Analyst will then load leads from its Analyst Table onto the CHT. The analyst who uploads the leads onto the CHT first, gets to retain the customer in his/her base. Leads secured in CHT are based on the first in first out processing technique.

SCD Type 2 updates the new campaign of existing leads by inserting new observation while preserving the historical observations. A new binary variable labeled as Current is created to provide a rapid way of identifying the most current record. The variable Current labels '1' as the most current observation and '0' as the past observations of campaign records.

As per regulatory requirements, the contact histories of customers are retained as unlimited historical observations can be preserved with SCD type 2. Overtime, the contact histories of customers are continuously built on a single database and may cause the database to spiral out of data space. Hence, we recommend the following improvement to be made.

IMPROVEMENT

The downside to SCD Type 2 is that the data is continuously built on a single database causing the database to expand rapidly overtime. We suggest diverting historical data onto an Archive Table and remaining current campaign in CHT. Records that are '0' in the Current variable will be archived into Archive Table. The Archive Table will be updated when Analysts load Analyst Table onto CHT. Based on the second scenario described in Section 5, the historical records of leads who are out of Recency Frame and are targeted for new campaigns leads will be purged onto Archive Table. With these changes, the CHT is a SCD Type 4.

Figure 4 shows the improved data management of uploading leads onto the CHT and Archive Table. Figure 5 shows the improved workflow of campaign lead-selection process.

Improved Data Management

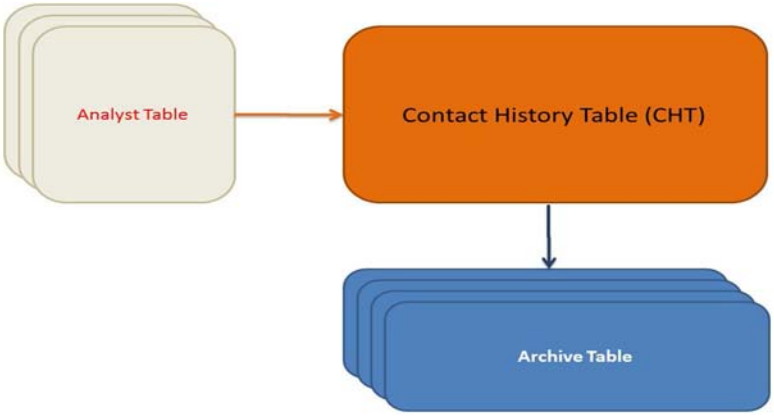

Figure 4- Improved Data Management

Improved Workflow of Campaign Lead Selection Process

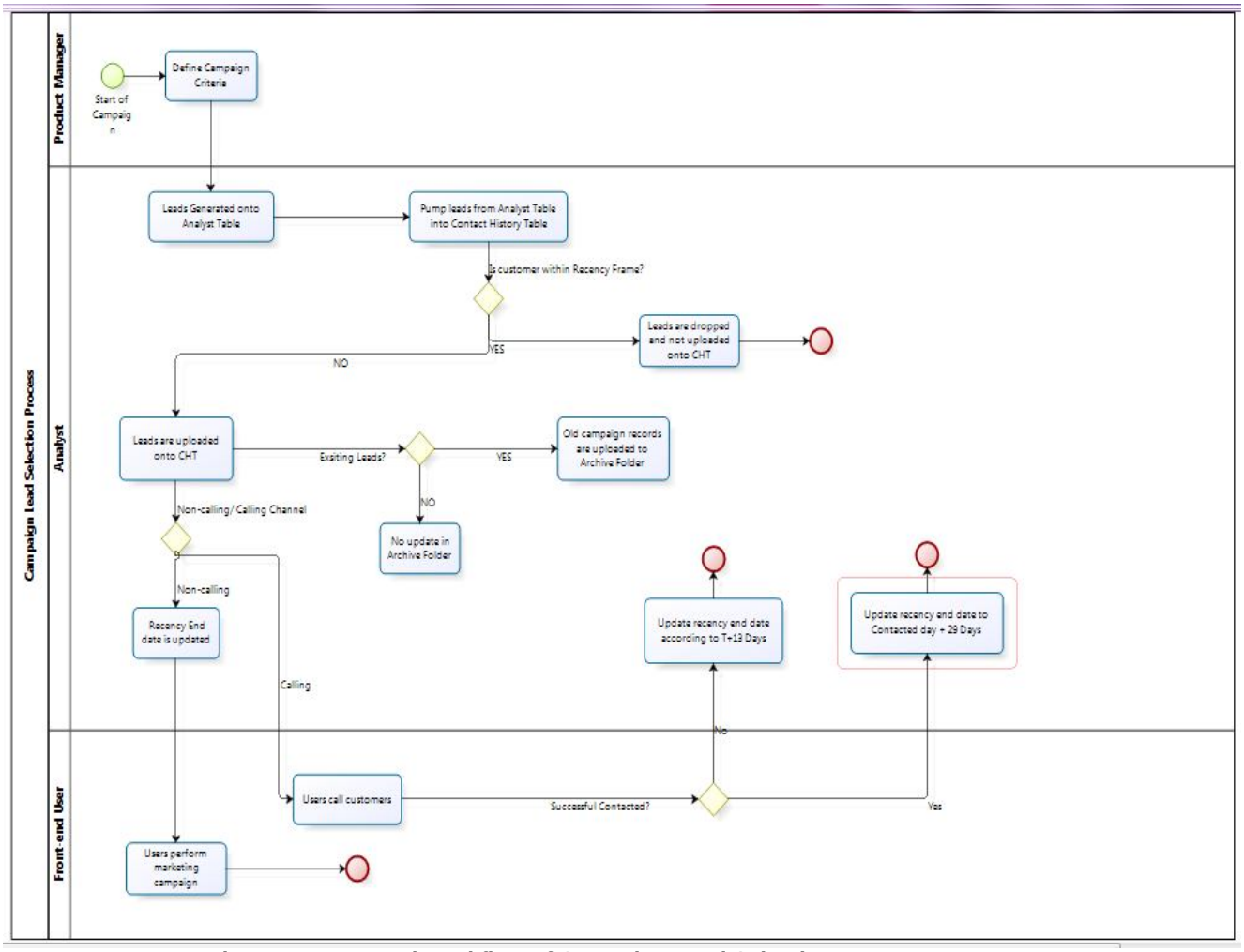

Figure 5 - Improved Workflow of Campaign Lead Selection Process

CONCLUSION

Managing multiple campaigns records of a huge customer base is a tedious task for the banks. By incorporating the business rules such as Recency Frame at distinctive campaign channels has compounded the level of difficulty.

CHT is a common dimension table used by Analysts to upload new campaign leads. Individual Analyst generates new leads to be captured onto Analyst Table for every new marketing campaign. The records on Analyst Table are uploaded onto CHT where a Recency Check Automation is embedded. This function runs an Automated Recency Check on new leads that are uploaded the CHT and purged out any campaign lead that violates the Recency Rule. The CHT only captures leads that fulfil the Recency Rule for marketers to follow up on.

SCD is implemented to update marketing records of bank-wide customers onto CHT. SCD Type 2 inserts new marketing records which fulfill the Recency Rule onto CHT and retains historical records from past marketing campaign. The records are continuously built on a single database- CHT, which may cause the database to expand rapidly overtime and spiral out of data space. Hence we suggest implementing a SCD Type 4 where current records are updated onto CHT and historical data are purged onto an Archive Table.

The implementation of SCD to data warehouse to manage marketing campaigns in banks has improves the data management and campaign management efficiency. The automated checks and updates of data warehouse greatly reduce cumbersome tasks performed by manual labour and provide a timely update and maintenance of the data.

REFERENCES

1. Ralph Kimball (2008, September 22). Slowly Changing Dimensions, Part 2. Kimball Group. January 20, 2013 from http://www.kimballgroup.com/2008/09/22/slowly-changing-dimensions-part-2/print/

2. Margy Ross (2008, September 29). Practical Steps for Designing a Dimensional Model. Kimball Group. January 20, 2013 from http://www.kimballgroup.com/2008/09/29/practical-steps-for-designing-a-dimensional-model/print/

3. Shauna Trainor (2013, February 13). Market In Modesty To Prevent Resentment. Covenant Group. February 13,2013 from http://www.covenantgroup.com/blog/market-in-modesty-to-prevent-resentment/

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Wang Lihui
Singapore Management University
Singapore
lihui.wang.2012@mitb.smu.edu.sg