Case Study:
How to Make Telecom Pricing Strategy Using Data Warehouse Approach

Seungjae Shin\textsuperscript{1,2}, Gilju Park\textsuperscript{3}, Wonjun Lee\textsuperscript{4}, Sunmi Lee\textsuperscript{5}

\textit{Korea Telecom R&D Group, Telecom Economic Research Lab.}

shinsj@rcunix.kotel.co.kr

Abstract

SIMS (Strategic Information Management System) is a large scale DSS for telecommunications industry. Its goal is developing pricing strategy. SIMS is developed based upon the huge volume of CDR (Call Data Record). Its two main functions are CBA (Call Behavior Analysis) and DB simulation. Once decision-makers fully analyze the call behavior pattern, they can make any assumption for their pricing strategy, and then they can make simulation for their own scenario. After several simulations, they can find which is the optimal solution under their environments. It is very important to know telephone customers’ behaviors because telecommunications market moves quickly to high competitive one. Telephone customers have an influence on the market. So, we need the customer oriented pricing strategy because the market itself turned into customer oriented one. To get the customer oriented strategy, we have to analyze thoroughly customers’ call behavior pattern. SIMS is the strong tool to get the customers’ call behavior pattern and to conduct simulation based on this knowledge.

1. Introduction

Everyday the telecommunications operating company generates a large number of data from switching systems, which are produced for every telephone call through a telephone network. CDR is the fingerprint of how many seconds and at what time a customer is using telephone. Therefore, CDR displays map of telephone customers’ behavior. If we knew the customers’ call behavior, we could make better decisions for telecommunications marketing strategy. However, the monthly size of CDR per a switching system may be near 1 giga bytes in general. In the past, it may not be easy to load all the company’s data at one time because of limitation of hardware capacity. Today, it is possible to access large amount of CDR according to the emergence of large scale data storage devices, symmetrical multiprocessing, and high performance databases. By using Data Warehouse technique, we can produce useful information for telecommunications business by analyzing CDR.

Korea Telecom’s Call Data Analysis Team made SIMS, a prototype Data Warehouse System for telecommunications pricing strategy. This paper demonstrates how to make pricing strategy using Data Warehouse technique. There are two main points in this job: CBA and DB simulation. CBA is a good indicator for developing marketing strategies, and DB simulation acts as conforming these strategies. The goal of developing SIMS is to produce reports which are applicable to building pricing strategy and to support a various level of analysis. This system can be a useful tool to survive in a highly competitive telecommunications marketplace.

\textsuperscript{1} Technical Research Staff, Call Data Analysis Team, Korea Telecom Economic Research Lab., R&D Group, Korea Telecom, 17 Woomyeon-Dong, Seocho-Gu, Seoul, 137-792, Korea (South)

\textsuperscript{2} shinsj@rcunix.kotel.co.kr, (Fax) +82-2-526-5396, (voice) +82-2-526-5316

\textsuperscript{3} Senior Researcher, Leader of CBA Team

\textsuperscript{4} Technical Research Staff, CBA Team

\textsuperscript{5} Researcher, CBA Team
2. What is a Data Warehouse?

A Data Warehouse is a database designed for decision support applications. Warehouse data is taken from operational systems, and is modified and combined to make it suitable for analysis by business-oriented users.6

The technologies used in the Data Warehouse System are multi-dimensional databases, client-server for data access, and metadata management. Using these techniques, a Data Warehouse System enables the integration of multiple databases into a single database designed for decision support. End users of a Data Warehouse System are various: a powerful data analyst and a high-level decision-maker. There are various levels of data according to each level of users. For a decision-maker, data are highly summarized and for an operational manager, the level of data is very low.

According to rules of William Inmon, who devised the concept of a Data Warehouse, the characteristics of data in a Data Warehouse should be:

- subject-oriented: organized according to the data's meaning, not according to the data structure of originating transactional systems
- integrated: store and encode consistently irrespective of the data sources
- time-variant: the data is not just a current snapshot but the data is over a long time horizon
- non-volatile: as a record, the historic data is never changed once loaded

The need for business information has been increased by the change of business environments. The deregulation and globalization make business environments competitive and rapidly changing especially in the telecommunications industry. This change makes the decision-makers need a good Data Warehouse System. Several major telephone and communications companies already set up the Data Warehouse System to have the best information to serve their customers at the lowest cost.9

3. SIMS : A Data Warehouse System

The typical Data Warehouse System is composed of three parts: warehouse population, Data Warehouse, and data access.10 Figure[1] shows the components of SIMS and their relationship.

The population of SIMS is 2.77% sampling data11 from the line capacity of switching systems of all country. The size of CDR12 in this population is near 9 giga bytes per month which amount to 0.15 billion call records. The total number of CDR over 1 year is near 110 giga bytes. Through extraction & transformation tools programmed by C-language, CDR can be transformed into DB. The Data Warehouse has Oracle 7.3 RDBMS and HP9000/K460 high performance computer machine, and large volume of processed data. We chose powerbuilder 5.0 as a data access tool to make an interface between user and a Data Warehouse.

The type of SIMS as a Data Warehouse is specialized warehouse.13 The application of SIMS is specialized in the telecommunications industry because SIMS uses CDR. The characteristics of SIMS are:

- SIMS is based on the star schema to support multi-dimensional queries efficiently.

---

7 consistent naming conventions, consistent measurement of variables, consistent physical attributes of data
8 There are only two kinds of operations: the initial loading of data and the access of data
9 British Telecom, Pacific Bell, US West, MCI
11 Sampling number of line capacity is 550,000.
12 The data contains only domestic call (local, long distance, and pager)
SIMS is the large, high-end server based DSS. SIMS can handle large historic data\textsuperscript{14}, SIMS can store CDR to the Optical Disk and DAT, SIMS can process the CDR from the various types of switching systems,\textsuperscript{15} SIMS can handle subject-oriented data from the several points of view.\textsuperscript{16}

4. DB Design

The DB design of SIMS is a simple star schema. The DB Design of a Data Warehouse has two types of tables: fact table and dimension table. There are three fact tables and four dimension tables in SIMS. The basic schema is shown in Figure[2].

![Figure[2] SIMS's DB Design](image)

In spite of increasing performance/price ratio of computer system, we were not able to handle all the CDR in the form of DB because of its size\textsuperscript{17}. So, we made three types of statistic tables for database which have the transformed data from the CDR: subscriber_number, calling_time, and calling_duration.

- **Subscriber_Number**: this contains number_of_calls and calling_time_by_seconds\textsuperscript{18} at each discount time zone\textsuperscript{19} according to each subscriber.
- **Calling_Time**: this contains number_of_calls and calling_time_by_seconds according to the time when the user uses telephone.
- **Calling_Duration**: this contains counts_of_call and number_of_calling_unit by seconds. For example, there are 100 calls of 1 second calling_unit\textsuperscript{20}, 200 calls of 2 second calling_unit, ... at a certain day.

The tariff structure of Telecommunications industry depends on the distance between calling party and called party, the time when user uses telephone, type of telephone and type of service which user uses, and the price rate. The dimension tables such as Distance_table, Discount_table, Class_table, Tariff_table contain these kinds of information. There are unique identifiable key to link between fact tables and dimension tables.

After statistic data are made, these data are loaded into DB. Summary data are made from the statistic DB according to each dimension. Among summary data, 1\textsuperscript{st} summary data are lightly summarized data and 2\textsuperscript{nd} summary data are highly summarized data. For example, what weekly_summarized data is to monthly_summarized data, lightly is to highly. Figure[3] shows the hierarchy of data used in SIMS.

![Figure[3] Hierarchy of Data](image)

5. Procedure for Making Billing Strategy

The procedure for making billing strategy is made of five steps. The Figure[4] shows the sequence of five steps.

\textsuperscript{14} By Inmon’s rule, the data in a Data Warehouse should be subject-oriented, integrated, time-variant, and non-volatile.
\textsuperscript{15} There are six types of switching systems in Korea: TDX, 5ESS, No1A, M10CN, S1240, AXE. Now, SIMS can handle CDR of TDX series switching systems and No1A switching system.
\textsuperscript{16} By dimension tables, we can view the data according to distance, time, and type of service.
\textsuperscript{17} The size of statistic data is 1/15 of CDR.
\textsuperscript{18} Time used by each call measured by seconds.
\textsuperscript{19} There are three discount time zone: standard rate time zone, economic rate time zone, and night rate time zone.
\textsuperscript{20} 1 calling unit = 1 second, 2 calling units = 2 seconds, 10 calling units = 10 seconds.
5.1. Step 1: Data Collection & Processing

CDR have been collected from the various types of switching systems. First of all, we have to decide the sample switching systems from each area. Unfortunately, CDR’s format of each switching system is different, so we have to start with code-conversion & standardization process. Code-conversion job is to make CDR of switching system readable by UNIX System and standardization job is to make different format of each switching system identical. After code_conversion & standardization jobs, data extraction is the next job, which means extracting and cleansing CDR for Data Warehouse.

The overall job can be divided by three parts: processing data, loading into DB, and running CBA and simulation. Among three parts, processing data is the most time-consumption job because of huge volume of CDR. So, it is very important to choose right data for the CBA and simulation. Figure[5] shows the flow of data process in SIMS.

5.2. Step 2: Call Behavior Analysis

After all the data are loaded into DB, we can access Data Warehouse through client PC. There are 7 functions to analyze call behaviors.

- **Subscriber’s statistics**
  This function can show each subscriber’s call behavior according to input conditions. If input conditions restrict a certain telephone number range, it shows number_of_calls and calling_time_by_seconds, and combination percentage of local call and each category of long distance call\(^{21}\). It can also show the list of ranks by the amount of telephone bill. Figure[6] displays the subscriber’s statistics.

---

\(^{21}\) Long Distance call in Korea is divided into 3 category by its distance. (1\(^{st}\) Zone= 0 ~ 30 km, 2\(^{nd}\) Zone=30~100Km, 3\(^{rd}\) Zone=100Km~ )
Calling Time statistics
This function can show the number_of_calls and calling_time_by_seconds at fixed time. If a user wants to see the call pattern at a specific day or month, it could show the call pattern graph.

Calling Duration statistics
In the DB, we already got the summarized information about the count_of_call and number_of_calling_unit by second. So we can draw a graph by count_of_call and number_of_calling_unit by seconds. Figure[7] shows the Calling Duration’s statistics.

Statistics between DDD areas
This function shows the traffic size between two DDD areas. It can show 144 * 144 matrix table which contains number of long distance calls and long distance calling_time_by_second between DDD areas.

Trend analysis
This function shows the trend of number_of_calls and calling_time_by_second at a specific time according to DDD area and type of service.

Statistics for billing category
If we know the all detailed data about the calls of each subscriber, we can calculate the each subscriber’s telecommunications cost. This function can show the table for number of subscribers belonged to a certain cost category and percentage of that category. Figure[8] shows this function.

Comparative analysis
In the case of Calling Time statistics and Calling Duration statistics, we can not distinguish one from the other unless we see them at one time. This function can show several resulting graph in one screen, so users can compare each call pattern curve at the same time. Figure[9] shows this function.

---

22 Korea has 144 local call areas.
5.3. Step 3: Hypothesis

Based on the result of Call Behavior Analysis, user can make a hypothesis for the change of tariff structure. Before conducting simulation, we have to decide input variables. There are several variables which can influence revenue: distance, zone, number of seconds per basic rate, price per one call, discount rate, hours belonged to standard, economic, and night rate, and price elasticity. The hypothesis with call behavior information is important in changing tariff structure. If we know the percentage of number of calls at daytime and nighttime, we can make different strategies in each time.

5.4. Step 4: Simulation for revenue based on the change of tariff structure

The result of the simulation shows every aspect of revenue changes according to area, time, zone, and total. Simulation makes two calculations: the one is calculation under current tariff structure and the other is calculation under new tariff structure. The new one makes recalculation based upon new input variables using same CDR. Finishing all the calculations, the new result is compared with that of the current one. Figure[10] shows simulation screen. The upper part is the input panel for new values of variables and the left bottom graph is one of the result windows and right bottom displays the current tariff structure.

5.5. Step 5: Making Billing Strategy

After doing enough times of step 3 and step 4, we can get several results. Comparing these results, we can select the optimal solution. We tried to find the scenario to give more discount rates to the telephone customers on the constraint of minimum damage to the company’s revenue. Once the best alternatives are found, it can be applied to the real world.

6. Conclusion

Data Warehouse can be defined as a large scale DSS. SIMS is a kind of Data Warehouse System for telecommunications industry. Even though implementing Data Warehouse is very tough task, successful Data Warehouse System can be a key factor to beat fierce competition in telecommunications market. It is very important to know telephone customers’ behavior because telecommunications market moves quickly to high competition. Telephone customers have an influence on the market. So, we need the customer oriented pricing strategy because the market itself turned into customer oriented one. To get the customer oriented strategy, we have to analyze call behavior pattern. SIMS is the strong tool to get the customers’ call behavior pattern and to conduct simulation based on this knowledge.

7. References