Paper 125-25

Riding the Information Wave to Marketing Success: Web Enabled On-Line Analytical Processing from Your Data Warehouse

George Bukhbinder, Palisades Research Inc., Bernardsville, NJ
Peter Babigian, AT&T, Basking Ridge, NJ

Abstract

Advanced decision support systems are widely used to target the appropriate customer base and measure the results of marketing efforts. A data warehouse is the foundation of such a system. This paper addresses several factors important to the rapid and successful completion of a large-scale data warehouse and decision support system development project in the telecommunications industry. These factors include:

1. Data model design.
2. Metadata maintenance and production automation using SAS/Warehouse Administrator® software.
3. Dimensions and hierarchies required for building a multidimensional database.
4. Internet-enabled OLAP.
5. Customer profiling and Customer Relationship Management using data mining in conjunction with data warehousing.

The presentation includes a demonstration of Internet-enabled OLAP used as an interface to a data warehouse and describes the business benefits of building a powerful decision support system for marketing.

Introduction

During the last decade, the focus in marketing has shifted from mass to direct/target marketing. The late 1990s saw the emergence of a new strategic management tool - Customer Relationship Management (CRM) - which is used as a framework for mass customization or one-to-one marketing. CRM combines marketing and technology to more effectively focus on each customer as an individual, generating quick, positive ROI and increased profitability.

This customer focus can be realized through the development and integration of IT technologies such as data warehouses and sophisticated analytical methods including data mining, On-Line Analytical Processing (OLAP), and Multi-Dimensional Data Bases (MDDB). The creation and enhancement of these tools along with the dramatic expansion of the Internet made CRM economically and technically possible. Because CRM relies heavily on computer technology, positive ROI can be expected only with rapid and cost-efficient development of the data warehouse and associated analytical tools.

This paper discusses the techniques used to create a data warehouse and Internet-enabled OLAP, which form the foundation for the successful development and implementation of CRM and a decision support system for database marketing.

To ensure non-disclosure of any confidential information, all data shown in this presentation are either publicly available or fictitious.

Project Objectives and Measurements

Prior to this project, IT support of this market segment was based on mainframe flat file processing. Reports had to be obtained through

ad hoc request capabilities. Because of the diverse structure of information from multiple data sources, the time required to perform the average ad hoc request significantly exceeded the expectations of marketing managers. In addition to very poor performance, the old system suffered from unreliable identification of individuals belonging to this market segment, because source data was not properly linked, cleaned and consolidated. These problems impeded the organization's ability to support decision making.

Consultants from Palisades Research, Inc. have teamed with AT&T to develop a Decision Support System (DSS) to meet the information needs of marketing to a specific segment of the company's customer base. For this purpose, a Data Warehouse and Internet-enabled OLAP system were built and used under the UNIX operating system, interfacing with mainframe and PC platforms. The DSS was designed to meet the following business objectives:

1. Provide reliable identification of individuals in the market segment.
2. Improve customer acquisition and retention rates.
3. Maximize customer profitability.
4. Track results on line.
5. Provide more cost-efficient communications.
6. Streamline marketing & product decision-making processes.
7. Secure positive project-based ROI .

The success of the project is measured by:

1. ROI.
2. Customer satisfaction, measured by survey.
3. Time to market.

Data Model Design

The DSS uses two distinct categories of data: Customer information and product information. Data sources are internal and external:

- Internal information is derived from company sources that contain customer account data, marketing, sales and promotional history. This data is combined with network usage information from billing systems.

- External information comes from customer lists that the company purchases from a number of vendors.

The structure of detail data in the warehouse is determined to a large extent by the needs of the business unit it supports. We adopted the star schema data model. The central point of the data warehouse is a customer account table that is linked to a promotion history table, service history table, and usage and revenue tables with data presented for each telecommunications product. In addition to these tables, which contain data applicable to telecommunications customers in general, there are a number of tables that describe specific characteristics of customers to whom this business unit markets services.

Although the data model described in the previous paragraph was implemented for a telecommunications company, a similar design approach is applicable to a wide range of industries, including banking, utilities and insurance. These industries share critical business characteristics with telecommunications, such as a fairly stable customer base and monthly billing.

Major Design Concepts

(see Application Design Chart - figure 1)

DATA LOADING, VALIDATION and TRANSFORMATION

The update frequency should be determined by business needs. The operational data is updated constantly, but based on the costs and benefits of having more recent data we determined that the data in the warehouse should be updated monthly.

The solution you choose must be flexible enough to load data in various formats, coming from diverse hardware platforms. For the system in this project, data is loaded onto Unix from mainframe, PCs and Unix. The native format of the data might be SAS, other ODBC compliant databases, flat files, Terradata, Excel, or Access.

The system performs data validation by collecting frequencies and other statistics on a monthly basis for all data sources, and analyzing this data over time. Transformation is necessary for data consistency. There are three major kinds of inconsistency:

1. Tables with different structures. Example: One table using separate fields for last name and first name and another using a single "name" field with the format (Smith, John).

2. Different sets of values for the same field. Example: One table categorizing age groups as 18-30, 31-55, and 56+, and another using 18-25, 26-40, 41-59 and 60+.

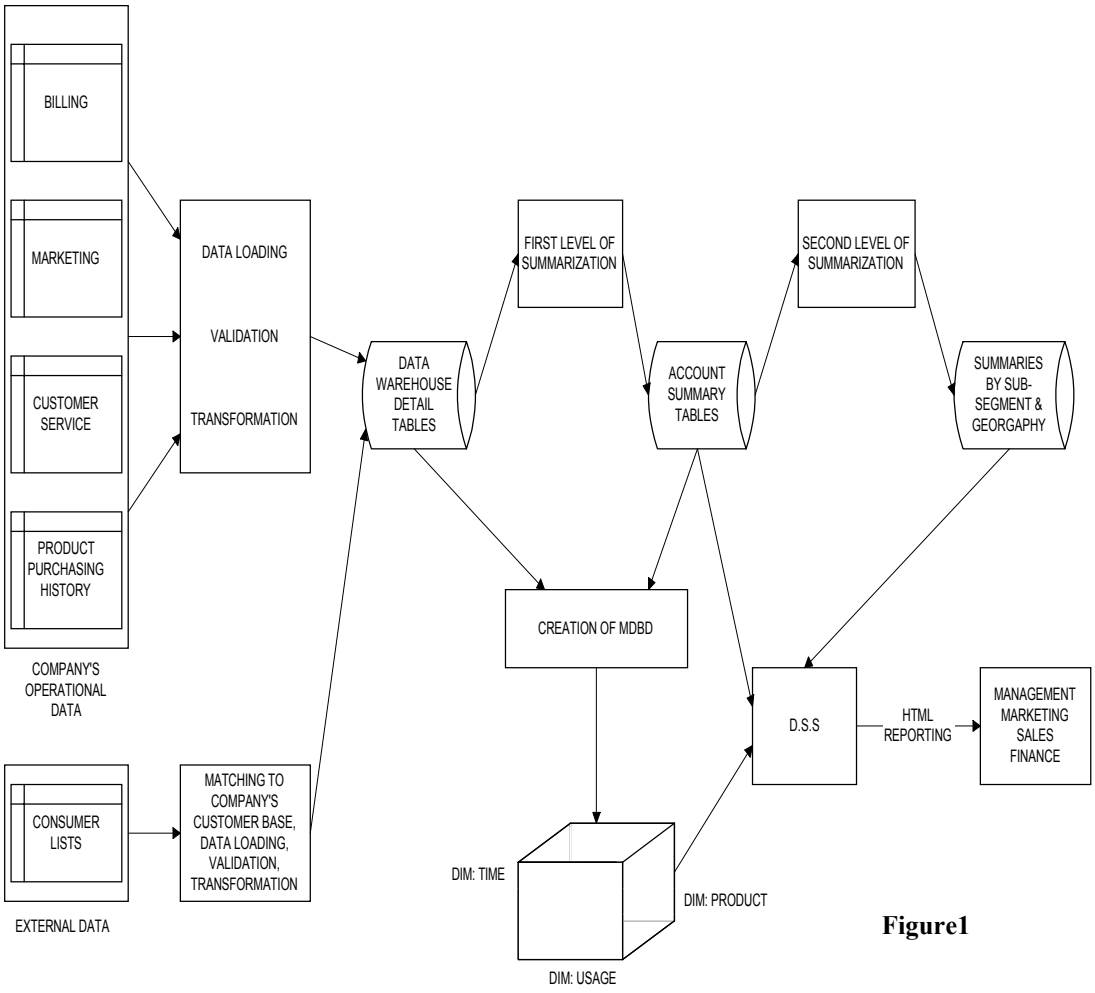

Figure1

3. Contradictory data from different sources. Example: Two tables having different addresses for the same person.

Removing all the inconsistencies is one of the most challenging parts of the project.

SUMMERIZATION

Most decisions are based on summarized data. To insure system performance, it is important to create the appropriate number of summary layers. It is useless to create summaries at higher levels of granularity than the business can use. In our data warehouse, the first layer of summarization represents an individual account by and across months. The next level of summarization represents a subsegment of the customer base or a product. A very important summarization object is an MDDB, which provides reach-through capabilities.

OLAP

OLAP systems enable managers and business analysts to examine business data and perform powerful comparison analyses on very large data volumes. OLAP technology often requires the creation of MDDBs. An MDDB is a specialized storage facility that allows analysts to retrieve information quickly because the data has already been logically organized in a matrix-like format. Using a SAS Internet-enabled multidimensional viewer, users gain drill-down capabilities to analyze data from their desktop. An intranet-based solution allows us to inexpensively provide access to a geographically disbursed user population without any special software or additional LAN connection from their desktops.

METADATA

Metadata is data about data. You can use it to audit or verify your data as well as apply common data definitions which can be shared by all users of the data warehouse. For example, the metadata will "remember" whether members of the customer base who do not purchase the business unit's products are included in average revenue calculations. The metadata resides on a Windows-based computer and is implemented using SAS warehouse administrator software. This software automatically tracks data transformation taking place in the warehouse and continuously updates the metadata.

DATA MINING

Data mining is the process of selecting, exploring, and modeling large amounts of data to uncover previously unknown patterns to achieve business advantage. This technology integrates different statistical and pattern recognition techniques such as neural networks, tree-based models, churn analysis, and traditional statistics, allowing the discovery of patterns, trends, exceptions, relationships, and anomalies that might otherwise stay hidden. In our data warehouse we applied data mining methodology to address following business problems:

- Demographic profiling of high value customers using decision tree models.
- Response modeling using a combination of regression methods, selecting those that produce the best gain chart.
- Determining of the hierarchy structure of the MDDB and defining the drill-down pass in OLAP applications. We exploited decision tree models using the dimensions of the data cube as independent variables and a high-value customer flag as a response variable.

Physical Architecture and Software (see figure 2)

The following company operational systems supplied data to the warehouse:

- Billing
- Account maintenance
- Customer service center databases
- Product purchasing history
- Marketing department information systems (promotion and contact history)

The warehouse development and maintenance is performed using the following SAS software products:

- Base SAS® System
- SAS Connect®
- SAS MDDB™ server
- SAS Intrnet™ software and SAS Internet multidimensional viewer
- SAS Warehouse Administrator®

Conclusion

Implementing a marketing decision support system built on top of a data warehouse using OLAP technology enhances current business processes while exploiting existing data. The following factors are important to the successful completion of this expensive development project and securing quick positive ROI:

1. Limiting the scope and keeping it simple.
2. Planning your data warehouse for the big picture but implementing it in small, concrete steps.
3. Selecting the software to build the system efficiently. This software should provide capabilities for accessing data across different platforms within the enterprise, data management and summarization, building and maintaining metadata, statistical and data analysis, and MDDB and OLAP development. For the most part, we used SAS software to perform these functions.
4. Clearly identifying system users to obtain different points of view. Review carefully users' questions to ensure that the DSS will address them.
5. Identifying the right data elements to store relevant information in the data warehouse.
6. Not spending too much time expecting users to provide complete information about their needs. Start to build the system at a reasonable point even if it appears that not all the required data is available. In all probability, it will be necessary at some point during development to recreate data tables and update some processes.
7. Get senior management's commitment to support the project, in order to deal efficiently with numerous problems that will arise during project development, particularly when interfacing with suppliers of data within the organization.

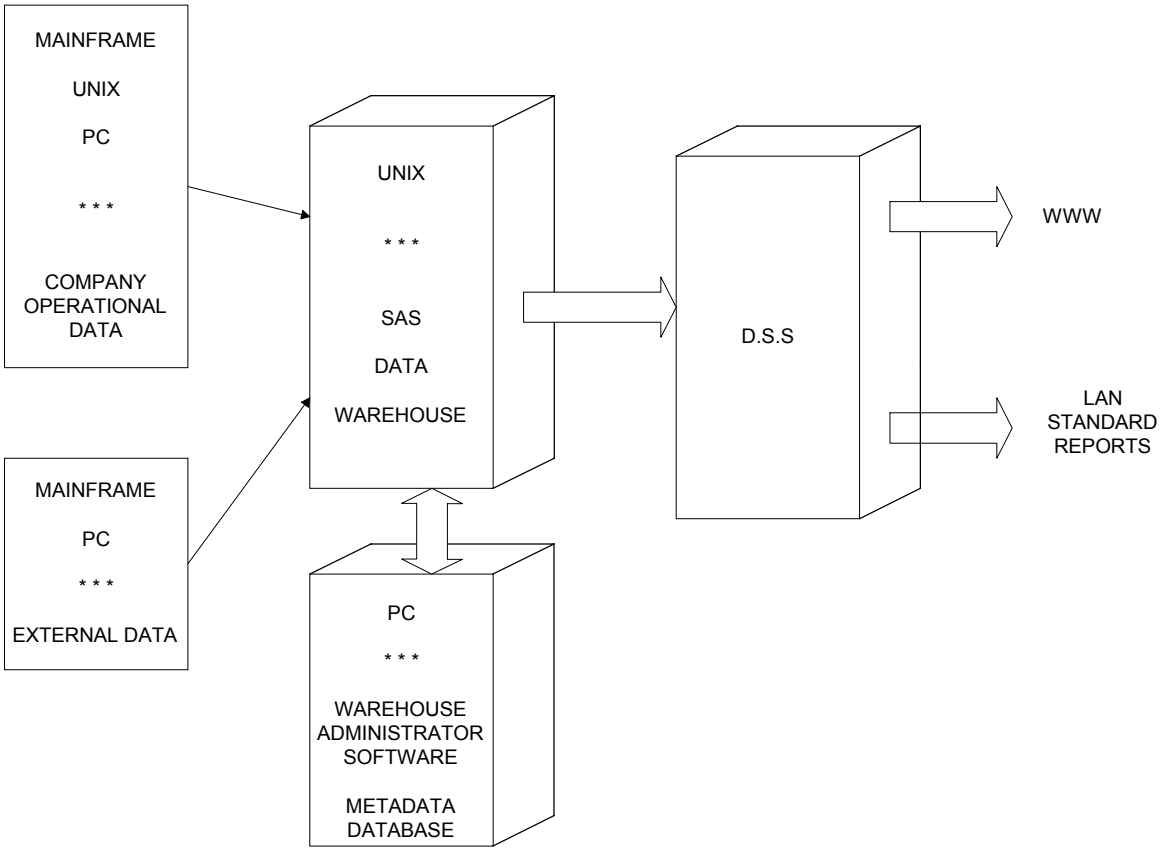

Figure 2.

Contact Information

George Bukhbinder
Palisades Research, Inc.
Suite # 309
75 Claremont Road
Bernardsville, NJ 07924
Voice: 908-953-8081
Fax: 908-953-8096
e-mail:
gbukhbinder@palisadesresearch.com

Peter Babigian
AT&T Corporation
Room # 5134A2
295 North Maple Ave.
Basking Ridge, NJ 07920
Voice: 908-221-7083
Fax: 908-221-3191
e-mail:
pbabigian@att.com