Data Warehouse Design to Support Customer Relationship Management Analyses

Colleen Cunningham, Drexel University, USA

Il-Yeol Song, Drexel University, USA

Peter P. Chen, Lousiana State University, USA

ABSTRACT

CRM is a strategy that integrates concepts of knowledge management, data mining, and data warehousing in order to support an organization's decision-making process to retain long-term and profitable relationships with its customers. This research is part of a long-term study to examine systematically CRM factors that affect design decisions for CRM data warehouses in order to build a taxonomy of CRM analyses and to determine the impact of those analyses on CRM data warehousing design decisions. This article presents the design implications that CRM poses to data warehousing and then proposes a robust multidimensional starter model that supports CRM analyses. Additional research contributions include the introduction of two new measures, percent success ratio and CRM suitability ratio by which CRM models can be evaluated, the identification of and classification of CRM queries, and a preliminary heuristic for designing data warehouses to support CRM analyses.

Keywords: *customer relationship management; data warehouse*

INTRODUCTION

It is far more expensive for companies to acquire new customers than it is to retain existing customers. In fact, acquiring new customers can cost five times more than it costs to retain current customers (Massey, Montoya-Weiss & Holcom, 2001). Furthermore, according to Winer (2001), repeat customers can generate more than twice as much gross income as new customers. Companies have realized that instead of treating all customers equally, it is more effective to invest in customers that are valuable or potentially valuable, while limiting their investments in non-valuable customers (i.e., not all relationships are profitable or desirable). As a result of these types of findings as well as the fact that customers want to be served according to their individual and unique needs,

Table 1. Customer segments

		Historic Value	
		Low	High
Future Value	High	II. Re-Engineer	IV. Invest
	Low	I. Eliminate	III. Engage

Table 2. Corresponding segmentation strategies

		Historic Value	
		Low	High
Future Value	High	Up-sell & cross-sell activities and add value	Treat with priority and preferential
	Low	Reduce costs and increase prices	Engage customer to find new opportunities in order to sustain loyalty

companies need to develop and manage their relationships with their customers such that the relationships are long-term and profitable. Therefore, companies are turning to Customer Relationship Management (CRM) techniques and CRM-supported technologies.

In our earlier work (Cunningham, Song, Jung, & Chen, 2003), we defined CRM as a strategy that utilizes organizational knowledge and technology in order to enable proactive and profitable long-term relationships with customers. It integrates the use of knowledge management, or organizational knowledge, and technologies to enable organizations to make decisions about, among other things, product offerings, marketing strategies, and customer interactions. By utilizing a data warehouse, companies can make decisions about customer-specific strategies such as customer profiling, customer segmentation, and cross-selling analysis. For example, a company can use a data warehouse to determine its customers' historic and future values and to segment its customer base. shows four quadrants of customer segmentation: (1) customers that should be eliminated (i.e., they cost more than what they generate in revenues); (2) customers with whom the relationship should be re-engineered (i.e., those that have the potential to be valuable, but may require the company's encouragement, cooperation, and/or management); (3) customers that the company should engage; and (4) customers in which the company should invest (Buttle, 1999; Verhoef & Donkers, 2001). The company then could use the corresponding strategies, as depicted in Table 2, to manage the customer relationships. Table 1 and Table 2 are only examples of the types of segmentation that can be performed with a data warehouse. However, if used, a word of caution should be taken before categorizing a customer into Segment I, because that segment can be further segmented into (a) those customers that serve as benchmarks for more valuable customers, (b) those customers that provide the company with ideas for product improvements or efficiency improvements, and (c) those customers that do not have any value to the company.

It is important to point out that customer segmentation can be further complicated by the concept of extended households. The term *ex-*

tended household refers to the relationship that exists between companies (e.g., parent company and subsidiary). The analysis of the relationships that exist between customers (i.e., lines of potential customer influence) is known as household analysis. It is important to understand and manage extended households, because a company's decision to treat a member of one segment potentially could have a negative impact on a related customer. For example, if a customer is in a non-profitable segment, then the company may decide to increase the customer's price. However, if the company is aware that the same non-profitable customer has influence over another customer (e.g., a parent or small business) that is in a more profitable segment, then the company may decide to not increase the customer's price rather than to risk losing both of the customers. Clearly, these social networks of influence are important for companies to identify and manage because of the impact that they can have on the company's ability to retain customers.

Currently, however, there are no agreed upon standardized rules for how to design a data warehouse to support CRM. Yet, the design of the CRM data warehouse model directly impacts an organization's ability to readily perform analyses that are specific to CRM. Subsequently, the design of the CRM data warehouse model contributes to the success or failure of CRM. In fact, recent statistics indicate that between 50% and 80% of CRM initiatives fail due to inappropriate or incomplete CRM processes and poor selection of technologies (Myron & Ganeshram, 2002; Panker, 2002). Thus, the ultimate long-term purpose of our study is to systematically examine CRM factors that affect design decisions for CRM data warehouses in order to build a taxonomy of CRM analyses and to determine the impact of those analyses on CRM data warehousing design decisions.

The taxonomy and heuristics for CRM data warehousing design decisions then could be used to guide CRM initiatives and to design and implement CRM data warehouses. The taxonomy also could be used to customize a starter model for a company's specific CRM requirements within a given industry. Furthermore, that taxonomy also would serve as a guideline for companies in the selection and evaluation of CRM data warehouses and related technologies.

In order to objectively quantify the completeness and suitability of the proposed CRM model (and alternative models), we propose two new metrics: *CRM success ratio* ($r_{success}$) and *CRM suitability ratio* ($r_{suitability}$). The *CRM success ratio* ($r_{success}$) is defined as the ratio of queries that successfully executed to the total number of queries issued against the model. A query is executed successfully if the results that are returned are meaningful to the analyst. The *CRM success ratio* cannot be used only to evaluate our proposed CRM model, but it also can be used to evaluate other CRM data warehouse models, as well. The range of values for $r_{success}$ is between 0 and 1. The larger the value of $r_{success}$, the more successful the model. The following equation defines the CRM success ratio:

$$r_{success} = Q_p / Q_n \qquad (1)$$

where Q_p is the total number of queries that successfully executed against the model, and Q_n is the total number of queries issued against the model.

The *CRM suitability ratio* ($r_{suitability}$) is defined as the ratio of the sum of the individual suitability scores to the sum of the number of applicable categories. The following equation defines the CRM suitability ratio:

$$r_{suitability} = \sum_{i=1}^{N} (\mathrm{X_i C_i}) / \mathrm{N} \qquad (2)$$

where N is the total number of applicable analysis criteria, C is the individual score for each analysis capability, and X is the weight assigned to each analysis capability.

The range of values for the $r_{suitability}$ ratio is between 0 and 1, with values closer to 1 being more suitable. Unlike the $r_{success}$ ratio, which can be used to evaluate and compare the richness and completeness of CRM data warehouse models, the $r_{suitability}$ ratio, however, can be used to help companies to determine the suitability of the model based upon the contextual priorities of the decision makers (i.e., based upon the company-specific CRM needs). We utilize the two metrics to evaluate the proposed CRM data warehouse model in our case study implementation.

A brief review of CRM literature is presented in the next section. The section on schema design introduces the analytical CRM analyses requirements that the data warehouse must support as well as provides guidelines for designing the fact tables and the dimensions. The experiment, which is subsequently described with the results in the following section, tests the completeness of the model. The flexibility of the model, the utilization of the CRM analyses, as well as the initial heuristics for designing a CRM data warehouse are presented in the discussion. Finally, the research contributions and future work are discussed in the conclusions

CRM LITERATURE REVIEW

The shift in marketing paradigms from mass marketing to target marketing to the customer-centric one-to-one marketing (known as relationship marketing) is driving CRM (Bose, 2002). Mass marketing is a product-focused approach that allows companies to reach a wide audience with little or no research, irrespective of the consumer's individual needs. Unlike mass marketing, target marketing focuses on marketing to segmented groups that share a similar set of characteristics (e.g., demographic information and purchasing habits). While both approaches are cost-effective, they do not allow for personalization. On the other hand, one-to-one marketing (relationship marketing) enables companies to treat customers individually according to their unique needs. Since not all relationships are profitable or desirable, relationship marketing allows companies to focus on customers that have the best potential lifetime value. In order to identify the appropriate customer-specific approach for managing individual customers, we first must classify customers into one of the four quadrants in Table 1 and subsequently apply the appropriate strategy. In the literature, researchers use the total historical value, total potential future value, and customer lifetime value (CLV). In fact, managing the CLV is essential to the success of CRM strategies (Bose, 2002), because companies that understand and utilize CLV are 60% more profitable than those that do not (Kale, 2004). There are many ways to define and calculate those measures (Hawkes, 2000; Hwang, Jung & Suh, 2004; Jain & Singh, 2002; Rosset, Neumann, Eick & Vatnik, 2003). For the purposes of this article, CLV is the sum of the total historical value and the total potential value for each customer. The following equation defines the *total historical value*:

$$\text{Historical Value} = \sum_{i=1}^{N} (\text{Revenue}_j - \text{Cost}_j) \tag{3}$$

where j is the individual products that the customer has purchased.

In Equation (3), the historical value is computed by summing the difference between the revenue and total cost over every product (j) that the customer has purchased in the past. The cost would include such things as product cost, distribution cost, and overhead cost. Using the same calculation as defined by Hwang et al. (2004), the following equation defines the *potential future value* for a customer:

$$\text{Potential Future Value} = \sum_{i=1}^{N} (\text{Probability}_j \text{ X } \text{Profitability}_j) \tag{4}$$

where j is the individual products that the customer potentially could purchase.

In Equation (4), the profitability represents the expected revenues minus the sum of the expected costs that would be incurred in order to gain the additional revenues. The probability represents the likelihood that the customer would purchase the product. Thus, the total potential future value would be the sum of individual potential future value of each product that the customer could potentially purchase. The sum of all of the individual customer lifetime values is known as *customer equity* (Rust, Lemon & Zeithaml, 2004).

One of the goals of companies should be to increase their customer equity from one year to the next. By incorporating the ability to compute the CLV into the CRM data warehouse, companies can utilize the CRM data warehouse to determine their customer growth. Additionally, companies can use key performance indicators (KPIs) to identify areas that could be improved. Specific KPIs should relate to the goals of the organization. For example, if a company wants to minimize the number of late deliveries, then an on-time delivery KPI should be selected. Some known KPIs that are relevant to CRM include, but are not limited to, margins, on-time deliveries, late-deliveries, and customer retention rates. Other KPIs that are relevant to CRM include, but are not limited to, marketing cost, number and value of new customers gained, complaint numbers, and customer satisfaction rates (Kellen, 2002).

SCHEMA DESIGN FOR CRM

The first step in any design methodology is to understand the requirements. As such, the minimum requirements for CRM analyses are presented in the CRM analysis requirements section. The specific CRM analysis requirements as well as the need to classify customers according to the four CRM quadrants presented in Table 1 then are used to identify the specific fact tables and dimensions. The heuristics (or guidelines) for modeling the fact tables and dimensions then are explored in the design rationale for the fact tables and design rationale for the dimensions subsections.

CRM Analysis Requirements

The purpose of a data warehouse is not just to store data but rather to facilitate decision making. As such, the first step to designing the schema for the CRM data warehouse is to identify the different types of analyses that are relevant to CRM. For example, some typical CRM analyses that have been identified include customer profitability analysis, churn analysis, channel analysis, product profitability analysis, customer scoring, and campaign management.

In addition to identifying what CRM analyses the data warehouse needs to support, we also must understand how the data analyses are used by the business users. Often, understanding the business use of the data analyses provides additional insights as to how the data should be structured, including the identification of additional attributes that should be included in the model.

Once the specific types of CRM analyses as well as the intended uses of those analyses have been identified, they can be decomposed into the data points that are needed to support the analyses. Moreover, additional data points also can be identified from both experience and literature (Boon, Corbitt, & Parker, 2002; Kellen, 2002; Rust et al., 2004). It should be noted that the additional data points could include non-transactional information such as customer complaints, support calls, and other useful information that is relevant for managing the customer relationships. Furthermore, the non-transactional information could exist in a variety of formats, such as video and graphics (Bose, 2002). Such data formats are beyond the scope of this article. Table 3 identifies the types of analyses that are relevant to CRM as well as some of the data maintenance issues that must be considered. In other words, Table 3 identifies the minimum design requirements for a CRM data warehouse (DW). It should be noted that there is no significance to the order in which the items are listed in Table 3. The design rationale in the following section is based on the minimum design requirements in Table 3.

Design Rationale for the Fact Tables

The model needs to have fact tables that can be used to compute the historical and future values for each customer, because they are used to classify customers. As such, the model consists of a profitability fact table, a future value fact table, a customer service fact table, and various dimensions, which are defined in Table 4. We note that not all of the fact tables and dimensions are included in Figure 1.

Table 3. Minimum design requirements for CRM DWs

No.	Analysis Type/Data Maintenance	Description
3.1	Customer Profitability	Ability to determine profitability of each customer
3.2	Product Profitability	Ability to determine profitability of each product
3.3	Market Profitability	Ability to determine profitability of each market
3.4	Campaign Analysis	Ability to evaluate different campaigns and responses over time
3.5	Channel Analysis	Ability to evaluate the profitability of each channel (e.g., stores, Web, and phone)
3.6	Customer Retention	Ability to track customer retention
3.7	Customer Attrition	Ability to identify root causes for customer attrition
3.8	Customer Scoring	Ability to score customers
3.9	Household Analysis	Ability to associate customers with multiple extended household accounts
3.10	Customer Segmentation	Ability to segment customers into multiple customer segmentations
3.11	Customer Loyalty	Ability to understand loyalty patterns among different relationship groups
3.12	Demographic Analysis	Ability to perform demographic analysis
3.13	Trend Analysis	Ability to perform trend analysis
3.14	Product Delivery Performance	Ability to evaluate on-time, late, and early product deliveries
3.15	Product Returns	Ability to analyze the reasons for and the impact of products being returned
3.16	Customer Service Analysis	Ability to track and analyze customer satisfaction, the average cost of interacting with the customer, and the time it takes to resolve customer complaints
3.17	Up-Selling Analysis	Ability to analyze opportunities for customers to buy larger volumes of a product or a product with a higher profitability margin
3.18	Cross-Selling Analysis	Ability to identify additional types of products that customers could purchase, which they currently are not purchasing
3.19	Web Analysis	Ability to analyze metrics for Web site
3.20	Data Maintenance	Ability to maintain the history of customer segments and scores
3.21	Data Maintenance	Ability to integrate data from multiple sources, including external sources
3.22	Data Maintenance	Ability to efficiently update/maintain data

Table 4. Starter model dimension definitions

Dimension Name	Dimension Definition
Channel Dimension	Stores the different modes for interacting with customers
Customer Dimension	Stores the static information about the customer
Customer Behavior Dimension	Stores the dynamic scoring attributes of the customer
Customer Demographics Dimension	Stores the dynamic demographic characteristics of the customer
Customer Existence	Tracks the periods in which the customer is valid
CustomerMarket	Tracks changes in the relationship between the customer and market dimensions
Comments Dimension	Stores the reasons for customer attrition and product returns
Company Representative	Stores the company representatives (sales representatives)
County Demographics Dimension	Stores external demographics about the counties
Extended Household	Represents the fact that the customer may belong to one or more extended households
Market Dimension	The organizational hierarchy and regions in which the customer belongs
Product Dimension	Represents the products that the company sells
Product Existence	Tracks the periods in which the products are valid
Promotion Dimension	Represents the promotions that the company offers
Prospect	Stores information about prospects
Scenario Dimension	Used to analyze hypothetical up-selling and cross-selling scenarios
Supplier Dimension	Represents the vendors that supply the products
sTime Dimension	The universal times used throughout the schema
Time Dimension	Universal dates used throughout the schema

The profitability fact table includes the attributes (e.g., revenues and all costs—distribution, marketing, overhead, and product) that are required to compute the historical profitability of each transaction in the profitability fact table with the minimum number of joins. That, in turn, improves the performance when querying the data warehouse. Additionally, storing the detailed transactions facilitates the ability to compute the CLV for each customer across each product. Moreover, the model depicted in Figure 1 can be used to calculate KPIs for delivery, such as the number of on-time items and the number of damage-free items. The complement measures are calculated by subtracting the explicitly stored KPI measures from the total quantity. These KPIs are important to track and manage, because they can help organizations to identify internal areas for process improvements and ultimately influence customer satisfaction and possibly customer retention.

The customer service fact table contains information about each interaction with the customer, including the cost of the interaction, the time to resolve the complaint, and a count of customer satisfaction or dissatisfaction. The total historical value of each customer is computed by summing the historical value of each transaction (i.e., the net revenue from the profitability fact table) and then subtracting the sum of the cost of interacting with the customer (i.e., the service cost from the customer service fact table).

In accordance with Equation 4, the future value fact table stores measures that are needed to compute the potential future lifetime value for each customer. For example, among other things, the future value fact table contains the expected gross revenue, costs, expected purchasing frequency, and the probability of gaining additional revenue. It also contains other descriptive attributes that can be used to analyze and categorize the customer's future lifetime value. The customer lifetime value, which is used to classify each customer in one of the four quadrants in Table 1, is computed by summing the historical value for each customer and the future value for each customer.

Design Rationale for the Dimensions

Dimensions are very important to a data warehouse, because they allow the users to easily browse the content of the data warehouse. Special treatment of certain types of dimensions must be taken into consideration for CRM analyses. Each of those dimension types and their special treatments are discussed in the following subsections.

Existence Dimensions and Time

Customer Relationship Management is a process. As with any business process, the CRM process needs to be changed periodically to reflect changes in and additions to the business process (e.g., organizational restructuring due to territory realignments or mergers and acquisitions, new or modified business rules, changes in strategic focus, and modified or new analysis requirements). Thus, time is an inherent part of business systems and must be modeled in the data warehouse. Traditionally, the time dimension primarily participates in a relationship with the fact tables only. Additionally, there are two ways of handling temporal changes: tuple versioning and attribute versioning (Allen & March, 2003). Tuple versioning (or row time stamping) is used in multiple ways to record (1) changes in the active state of a dimension, (2) changes to the values of attributes, and (3) changes in relationships (Todman, 2001). As such, traditional tuple versioning has limitations within the context of CRM. For example, periods of customer inactivity can be determined only by identifying two consecutive tuples where there is a gap in the timestamp. Additionally, queries that involve durations may be spread over many tuples, which would make the SQL statement complex with slow response times (Todman, 2001).

In order to alleviate the issues with traditional time stamping in the context of CRM, each dimension is examined carefully to determine if the dimension (1) contains attributes whose complete set of historical values have to be maintained, or (2) is subject to discontinuous existence (i.e., only valid for specific periods).

If either (1) or (2) is applicable, then a separate dimension is created called an existence dimension. The existence dimensions are implemented as outriggers, and two relationships are created between the time dimension and each outrigger dimension. The two relationships are formed in order to record the date period in which the data instances are valid. In doing so, this facilitates the ability to perform state duration queries and transition detection queries (Todman, 2001). State duration queries contain a time period (start date and end date) in the *where* clause of the query, whereas transition detection queries identify a change by identifying consecutive periods for the same dimension (Todman, 2001).

Careful consideration is given to this step in the design process, because the fact table only can capture historical values when a transaction occurs. Unfortunately, the reality is that there may be periods of inactivity, which would mean that any changes that occur during those periods of inactivity would not be recorded in the data warehouse. This would, in turn, impact the types of analyses that could be done, since one cannot analyze data that one has not recorded.

Mini-Dimensions

If a dimension contains attributes that are likely to change at a different rate than the other attributes within the dimension, then a separate dimension is created as a mini-dimension. The new dimensions are implemented as mini-dimensions as opposed to outriggers in order to allow the user to readily browse the fact table. One benefit of this approach is that the history of the changes in the customer's behavior scores and demographics are stored as part of the fact table, which facilitates robust analyses without requiring the use of Type 1, 2, or 3 techniques (Kimball & Ross, 2002) for the Customer Demographics or Customer Behavior dimensions.

Customer Dimension

The customer must be at the heart of the customer-centric data warehouse. As such, careful attention must be given to the design of the customer dimension, which will force attention on the customer. Direct relationships are formed between the Customer dimension and the Sales Representative, Market, Comment, and Time dimensions in order to allow the user to readily determine the most current values for the sales representative, market, activation date, attrition date, and attrition comments by simply browsing the Customer dimension without having to include a time constraint in the query statement.

Other Dimensions

There is a Household dimension as well as an extended household dimension in order to analyze the potential lines of influence that exists between customers. In accordance with the types of CRM analyses that the data warehouse must support, other dimensions are identified according to the dimensions along which the fact tables are analyzed. For example, other dimensions include the Product, Supplier, Channel, Promotion, Market, and Sales Representative dimensions in order to facilitate the CRM analyses described in the CRM Analysis Requirements section.

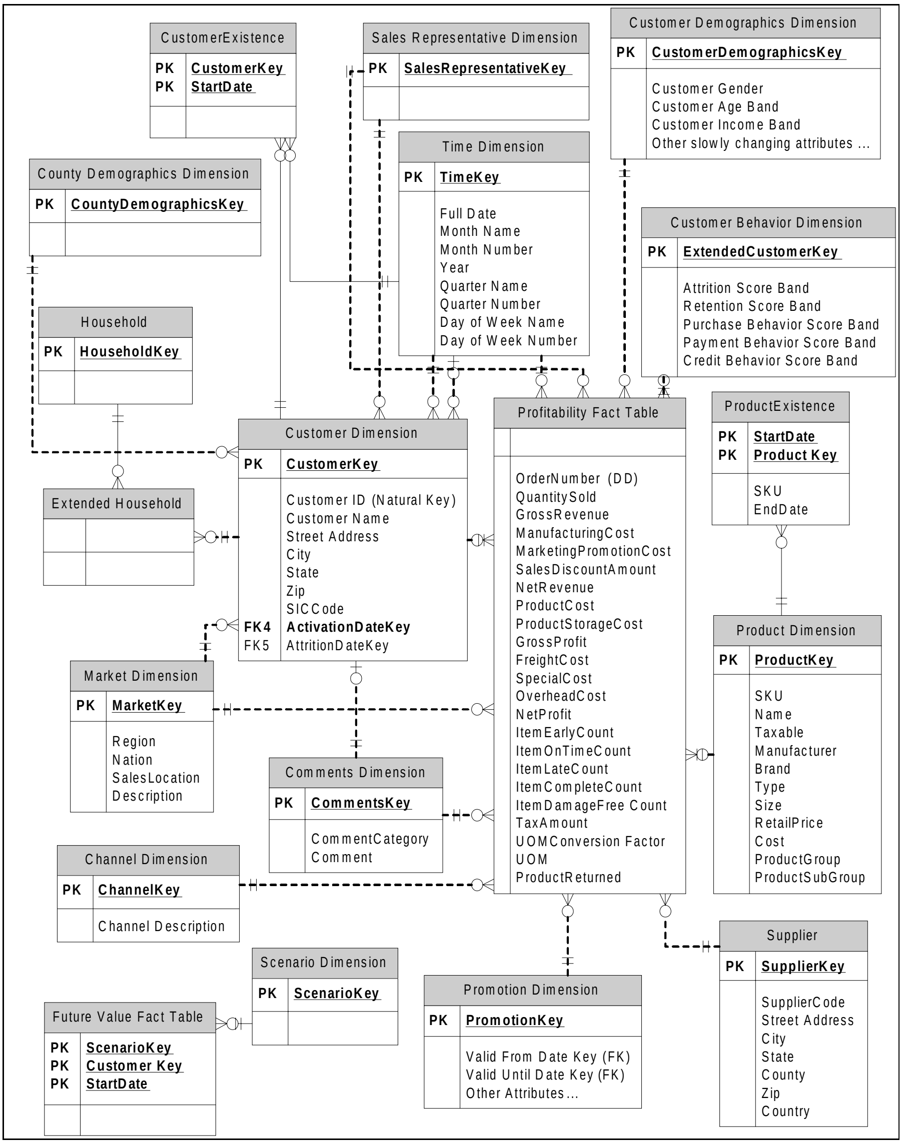
As a result of this approach to modeling the dimensions, the only slowly changing dimensions in the model are the County Demographics dimension, the Product dimension, the Supplier dimension, and the Customer dimension.

The model depicted in Figure 1, which is based upon the minimum design requirements and the design rationale presented in this section, is tested to determine its completeness and flexibility for CRM analyses. The experiment that is used to test the model is described in the following section.

EXPERIMENT

The purpose of the experiment is to test the completeness and flexibility of the proposed CRM data warehouse model. Our hypothesis is that the proposed data warehouse starter model has a positive impact on the ability to perform CRM analyses. The implementation, methodology, and selection of the queries that

Figure 1. Proposed CRM data warehouse model



are used in the experiment to test our hypothesis as well as the results are discussed in the specific subsections that follow.

Implementation

We perform a case study to test the validity of our proposed starter model. The proposed CRM data warehouse model is implemented in SQL Server 2000 running on a Windows 2000 server. The hardware computer is a DELL 1650 database server with a single processor and 2.0 MHz. The schema is populated with 1,685,809 rows of data from a manufacturing company.

Methodology

In the experiment, a series of CRM queries are executed against the proposed data warehouse schema. The success rate of the proposed schema is computed as a ratio of the number of successful queries executed divided by the total number of queries used in the investigation. Furthermore, the proposed CRM data warehouse model is tested to determine if it could or could not perform the analyses listed in Table 3. For each analysis in Table 3 that the model could perform, it is given a score of one point; otherwise, the model is given a score of zero points. The sum of the points for the model is computed in order to determine an overall CRM-analysis capability score. The selection of the queries that are used to study the model is discussed in the following section.

Selection of Queries to Test

Since we believe that the proposed data warehouse starter model has a positive impact on the ability to perform CRM analyses, special care was taken in the selection of the queries used for testing in order to avoid any biases in the types of queries used to test the model. Stratified random sampling is used to select the specific queries for the experiment. The stratified random sampling is conducted as follows: (1) representative queries for CRM are gathered from literature and experience; (2) the queries are grouped into categories based upon the nature of the query; (3) within each category, each query is numbered; (4) a random number generator is used to select queries from each category; and (5) the queries whose assigned number corresponds to the number generated by the random number generator are selected. The specific queries that are selected are listed in Table 5. It is important to note that since the queries are randomly selected from a pool of CRM-related queries, it is possible that the $r_{success}$ ratio can be less than one for our proposed model. It is also important to note that the representative CRM queries are queries that equally apply to different industries and not queries that are specific to only one industry. This aspect of the sampling procedure is important in order to make generalizations about the characteristics of the data warehouse schema that should be present in order to perform CRM analyses across different industries.

Results

Our preliminary finding is that the proposed CRM data warehouse model can be used to successfully perform CRM analyses. Based upon the sample queries, our model has a value of 1 and 0.93 for the $r_{success}$ and $r_{suitability}$ ratios, respectively. The individual scores for successfully executing the queries against the model are listed in Table 5. The individual and cumulative scores for the suitability of the proposed CRM data warehouse model are listed in Table 6. It should be noted that there is no significance to the order in which the items are listed in the table.

The scores for items 6.1 through 6.11 in Table 6 are based upon whether or not queries are successfully executed in those categories. The scores for items 14 and 15 are determined while loading data from multiple sources and updating customer scores. Each of the queries that were successfully executed in the experiment is discussed in further detail in the following section in order to highlight the completeness and flexibility of the model.

DISCUSSION

In addition to discussing the completeness and flexibility of the model, this section also presents potential uses of the specific analyses, including KPIs. This section also describes the data quality issues pertaining to CRM data warehouses before presenting a summary of the heuristics for designing data warehouses to support CRM analyses.

Model Completeness and Flexibility

The starter model depicted in Figure 1 can be used for a variety of CRM analyses, including customer profitability analysis, household profitability analysis, demographics profitability analysis, product profitability analysis, channel profitability analysis, and promotion profitability analysis simply by including the appropriate dimensions in the query statement. Furthermore, each query can be modified to include additional measures and descriptions simply by including additional fields from the fact table and the dimensions. Some of those queries are discussed next.

Table 5. Sample CRM analyses

No.	Category	Analysis	Pass	Fail
5.1	Channel Analysis	Which distribution channels contribute the greatest revenue and gross margin?	1	0
5.2	Order Delivery Performance	How do early, on-time, and late-order shipment rates for this year compare to last year?	1	0
5.3	Order Delivery Performance and Channel Analysis	How do order shipment rates (early, on time, late) for this year compare to last year by channel?	1	0
5.4	Customer Profitability Analysis	Which customers are most profitable based upon gross margin and revenue?	1	0
5.5	Customer Profitability Analysis	What are the customers' sales and margin trends?	1	0
5.6	Customer Retention	How many unique customers are purchasing this year compared to last year?	1	0
5.7	Market Profitability Analysis	Which markets are most profitable overall?	1	0
5.8	Market Profitability Analysis	Which products in which markets are most profitable?	1	0
5.9	Product Profitability Analysis	Which products are the most profitable?	1	0
5.10	Product Profitability Analysis	What is the lifetime value of each product?	1	0
5.11	Returns Analysis	What are the top 10 reasons that customers return products?	1	0
5.12	Returns Analysis	What is the impact of the value of the returned products on revenues?	1	0
5.13	Returns Analysis	What is the trend for product returns by customers by product by reason?	1	0
5.14	Customer Attrition	What are the top 10 reasons for customer attrition?	1	0
5.15	Customer Attrition	What is the impact of the value of the customers that have left on revenues?	1	0

The SQL statement in Figure 2 is used to identify the most profitable customers based upon total revenue and gross margin. By excluding the time dimension, the customer profitability SQL statement identifies the customer's historical lifetime value to the company. This is an important analysis that, in conjunction with the customer's future value and the customer service interaction costs, is used to classify customers in one of the four CRM quadrants (Table 1), which subsequently can be used to determine the appropriate strategy for managing the customer.

The SQL statement in Figure 3 is used to determine the margins for each product and subsequently identifies products that potentially may be eliminated from the company's product line. The ability to be able to determine

Table 6. Sample suitability for CRM analyses scores

No.	Criteria	Score
6.1	Ability to track retention	1
6.2	Ability to identify root causes for customer attrition	1
6.3	Ability to score customers	1
6.4	Ability to associate customers with multiple extended household accounts	1
6.5	Ability to segment customers into multiple customer segmentations	1
6.6	Ability to maintain the history of customer segments and scores	1
6.7	Ability to evaluate different campaigns and responses over time	1
6.8	Ability to analyze metrics for Web site	0
6.9	Ability to understand loyalty patterns among different relationship groups	1
6.10	Ability to perform demographic analysis	1
6.11	Ability to perform trend analysis	1
6.12	Ability to perform customer profitability analysis	1
6.13	Ability to perform product profitability analysis	1
6.14	Ability to integrate data from multiple sources, including external sources	1
6.15	Ability to efficiently update/maintain data	1
	Total	**14**

Figure 2. Customer profitability analysis query — Which customers are most profitable based upon gross margin and revenue?

```
SELECT b.CustomerKey, b.CustomerName, Sum(a.GrossRevenue) AS TotalRevenue,
Sum(a.GrossProfit) AS TotalGrossProfit, TotalGrossProfit/TotalRevenue AS GrossMargin
FROM tblProfitabilityFactTable a, tblCustomer b
WHERE b.CustomerKey=a.CustomerKey
GROUP BY b.CustomerKey, b.CustomerName
ORDER BY Sum(a.GrossRevenue) DESC;
```

Figure 3. Product profitability analysis query — Which products in which markets are most profitable?

```
SELECT c.Year, b.MarketKey, b.LocationCode, b.Location, b.Description,
b.CompetitorName, d.ProductCode, d.Name, Sum(a.GrossRevenue) AS TotalRevenue,
Sum(a.GrossProfit) AS TotalGrossProfit, TotalGrossProfit/TotalRevenue AS GrossMargin
FROM tblProfitabilityFactTable a, tblMarket b, tblTimeDimension c, tblProductDimension
d
WHERE b.MarketKey=a.MarketKey And a.TimeKey=c.TimeKey And
a.ProductKey=d.ProductKey
GROUP BY c.Year, b.MarketKey, b.LocationCode, b.Location, b.Description,
b.CompetitorName, d.ProductKey, d.ProductCode, d.Name, b.MarketKey
ORDER BY Sum(a.GrossRevenue) DESC;
```

Figure 4. Order delivery performance query — How do early, on time, and late order shipment rates for this year compare to last year?

```
SELECT b.Year, Sum(a.ItemOnTimeCount) AS OnTime, Sum(a.ItemEarlyCount) AS
Early, Sum(a.ItemLateCount) AS Late,
Sum(a.ItemOnTimeCount+a.ItemEarlyCount+a.ItemLateCount) AS TotalCount,
OnTime/Late*100 AS PercentOnTime, Early/TotalCount*100 AS PercentEarly,
Late/TotalCount*100 AS PercentLate
FROM tblProfitabilityFactTable a, tblTimeDimension b
WHERE b.TimeKey = a.TimeKey
GROUP BY b.Year;
```

Figure 5. Order delivery performance and channel analysis query — How do order shipment rates (early, on-time, late) for this year compare to last year by channel?

```
SELECT b.Year, c.ChannelCode, Sum(a.ItemOnTimeCount) AS OnTime,
Sum(a.ItemEarlyCount) AS Early, Sum(a.ItemLateCount) AS Late,
Sum(a.ItemOnTimeCount+a.ItemEarlyCount+a.ItemLateCount) AS TotalCount,
OnTime/Late*100 AS PercentOnTime, Early/TotalCount*100 AS PercentEarly,
Late/TotalCount*100 AS PercentLate
FROM tblProfitabilityFactTable a, tblTimeDimension b, tblChannelDimension c
WHERE b.TimeKey=a.TimeKey And c.ChannelKey = a.ChannelKey
GROUP BY b.Year, c.ChannelCode;
```

Figure 6. Market profitability analysis query — Which markets are most profitable overall?

```
SELECT c.Year, b.MarketKey, b.LocationCode, b.Location, b.Description,
b.CompetitorName, Sum(a.GrossRevenue) AS TotalRevenue, Sum(a.GrossProfit) AS
TotalGrossProfit, TotalGrossProfit/TotalRevenue AS GrossMargin
FROM tblProfitabilityFactTable a, tblMarket b, tblTimeDimension c
WHERE b.MarketKey=a.MarketKey And a.TimeKey=c.TimeKey
GROUP BY c.Year, b.MarketKey, b.LocationCode, b.Location, b.Description,
b.CompetitorName, b.MarketKey
ORDER BY Sum(a.GrossRevenue) DESC;
```

the lifetime value of each product (irrespective of market) merely by modifying the SQL statement in Figure 3 to exclude the product code further illustrates the flexibility and robustness of the proposed CRM model.

The SQL statement in Figure 4 is used to determine and compare Key Performance Indicators (KPIs) for overall on-time, early, and late shipment percentages for different years.

By modifying the statement in Figure 4 to include the Channel Dimension, the performance of each channel from one year to the next is determined. The modified SQL statement can be seen in Figure 5.

The SQL statement in Figure 6 is used to determine the overall profitability of each market. By eliminating the market key from the SQL statement, the profitability for each location is obtained for each location within the organizational hierarchy that is defined in the market dimension.

The SQL statement in Figure 7 demonstrates that by including the product code from the product dimension in the previous SQL

Figure 7. Product profitability analysis query — Which products in which markets are most profitable?

```
SELECT c.Year, b.MarketKey, b.LocationCode, b.Location, b.Description,
b.CompetitorName, d.ProductCode, d.Name, Sum(a.GrossRevenue) AS TotalRevenue,
Sum(a.GrossProfit) AS TotalGrossProfit, TotalGrossProfit/TotalRevenue AS
GrossMargin
FROM tblProfitabilityFactTable a, tblMarket b, tblTimeDimension c,
tblProductDimension d
WHERE b.MarketKey=a.MarketKey And a.TimeKey=c.TimeKey And
a.ProductKey=d.ProductKey
GROUP BY c.Year, b.MarketKey, b.LocationCode, b.Location, b.Description,
b.CompetitorName, d.ProductKey, d.ProductCode, d.Name, b.MarketKey
ORDER BY Sum(a.GrossRevenue) DESC;
```

Figure 8. Returns analysis — What are the top reasons that customers return products?

```
SELECT b.CommentsKey, c.ProductCode, c.Name, d.Comment, Sum(a.GrossRevenue)
AS TotalRevenue, Sum(a.GrossProfit) AS TotalGrossProfit,
TotalGrossProfit/TotalRevenue AS GrossMargin, Count(*) AS MembershipCount
FROM tblProfitabilityFactTable a, tblTimeDimension b, tblProductDimension c,
tblCommentDimension d
WHERE a.TimeKey=b.TimeKey And a.ProductKey=c.ProductKey And
a.CommentsKey=d.CommentsKey And a.ProductReturned=Yes
GROUP BY d.CommentsKey, c.ProductCode, c.Name, d.Comment, c.ProductKey
ORDER BY Count(*) DESC, Sum(a.GrossProfit) DESC;
ORDER BY Sum(a.GrossRevenue) DESC;
```

statement, the profitability of each product by market is obtained.

The SQL statement in Figure 8 is used to determine the top reasons for product returns. In this case, the basis for the top reasons is merely the count of the number of reasons that products are returned. By first grouping the products that are returned according to the reason for their return and the product code and then including the number of returns, revenue, gross profit, and gross margin for each group, the SQL statement in Figure 8 identifies areas upon which the company should improve in order to minimize the number of returns and to improve overall customer satisfaction. Specifically, since companies have limited resources, a company can use the result set to create Pareto charts according to the most frequently occurring problems that have the largest associated gross profits. Management teams then can use the Pareto charts to determine which problems to address first with corrective actions. It should be noted that in order to facilitate quick identification of the most frequently occurring problems that have the largest associated gross profits, the SQL statement in Figure 8 includes an ORDER BY clause.

Furthermore, simply by modifying the SQL statement in Figure 8 to include the year of the transaction from the profitability fact table in the SELECT clause and the GROUP BY clause, companies can use the results of the modified query to monitor the trend of return reasons over time. Stated differently, companies can use the results of the modified query statement to monitor the impact of the corrective actions over time. Not only can the return analyses be used to monitor the impact of corrective actions, but they also can be used to identify improvement

Figure 9. What is the impact of the value of the returned products on revenues?

```
SELECT b.CommentsKey, c.ProductCode, c.Name, d.Comment, Sum(a.GrossRevenue)
AS TotalRevenue, Sum(a.GrossProfit) AS TotalGrossProfit,
TotalGrossProfit/TotalRevenue AS GrossMargin, Count(*) AS MembershipCount
FROM tblProfitabilityFactTable a, tblTimeDimension b, tblProductDimension c,
tblCommentDimension d
WHERE a.TimeKey=b.TimeKey And a.ProductKey=c.ProductKey And
a.CommentsKey=d.CommentsKey And a.ProductReturned=Yes
GROUP BY d.CommentsKey, c.ProductCode, c.Name, d.Comment, c.ProductKey
ORDER BY Count(*) DESC;
ORDER BY Sum(a.GrossRevenue) DESC;
```

Figure 10. What is the trend for product returns by customers by product by reason?

```
SELECT e.CustomerName, b.Year, b.CommentsKey, c.ProductCode, c.Name,
d.Comment, Sum(a.GrossRevenue) AS TotalRevenue, Sum(a.GrossProfit) AS
TotalGrossProfit, TotalGrossProfit/TotalRevenue AS GrossMargin, Count(*) AS
MembershipCount
FROM tblProfitabilityFactTable a, tblTimeDimension b, tblProductDimension c,
tblCommentDimension d, tblCustomerDimension e
WHERE a.TimeKey=b.TimeKey And a.ProductKey=c.ProductKey And
a.CommentsKey=d.CommentsKey And a.ProductReturned=Yes
GROUP BY e.CustomerName, b.Year, d.CommentsKey, c.ProductCode, c.Name,
d.Comment, c.ProductKey
ORDER BY Count(*) DESC, Sum(a.GrossProfit) DESC;
ORDER BY Sum(a.GrossRevenue) DESC;
```

targets, which can be tied to employee (and/or departmental) performance goals.

The SQL statement listed in Figure 9 is used to determine the impact of the returned products on revenues.

The SQL statement listed in Figure 10 is used to identify the trend for product returns by customer, by product, and by reason. The results can be used to identify whether or not a problem is systematic across all customers, many customers, or a few specific customers. This query also can be used to help management make an informed decision with respect to allocating resources to address problems that lead to customers returning products. Additionally, the results can be used by the sales team to gain further insights into why their customers have returned products. The sales team potentially can use that information to work with the customer to resolve the issue(s) in cases where the customer repeatedly has returned products for reasons that cannot be considered the company's mistake. Alternatively, the sales team can use the results to identify accounts that could (should) be charged additional fees if the customer repeatedly returns products.

The SQL statement in Figure 11 is used to identify the top reasons for customer attrition. Figure 12 is used to analyze the impact of customer attrition on the total revenues. By analyzing customer attrition, companies can gain further insights into areas for improvement in order to reduce the attrition rate and thereby improve its overall company value.

Table 7 summarizes some of the possible uses for the CRM analyses that are presented in Table 5.

Data Quality

Given the range of decisions that the CRM data warehouse must be able to support

Figure 11. What are the top 10 reasons for customer attrition?

```
SELECT b.Comment, Count(a.CommentsKey) AS CountReasons
FROM tblCustomer AS a, tblCommentDimension AS b
WHERE a.CommentsKey=b.CommentsKey
GROUP BY b.Comment
ORDER BY Count(a.CommentsKey) DESC;
```

Figure 12. What is the impact of the value of the customers that have left on revenues?

```
SELECT b.Comment, Count(a.CommentsKey) AS NumberOfTransactions,
Sum(c.GrossRevenue) AS TotalGrossRevenue
FROM tblCustomer AS a, tblCommentDimension AS b, tblProfitabilityFactTable
AS c
WHERE a.CommentsKey=b.CommentsKey AND c.CustomerKey=a.CustomerKey
GROUP BY b.Comment
ORDER BY Count(a.CommentsKey) DESC;
```

and given the potential impact on companies' profitability, it is imperative that the data are accurate. The data that are used in the CRM analyses originate from disparate data sources that must be integrated into the data warehouse. Under such circumstances, the issue of dirty data arises. Analyzing dirty data, particularly in the context of systems that support corporate decision-making processes (e.g., CRM analyses and subsequent decisions), would result in unreliable results and potentially inappropriate decisions. As such, ensuring data quality within the data warehouse is important to the overall success of subsequent CRM analyses and decisions. Data quality should not be considered a one-time exercise conducted only when data are loaded into the data warehouse. Rather, there should be a continuous and systematic data quality improvement process (Lee, Pipino, Strong, & Wang, 2004; Shankaranarayan, Ziad, & Wang, 2003).

One way of minimizing data quality issues is to carefully document the business rules and data formats that then can be used to ensure that those requirements are enforced. Too often, however, thorough documentation of the business rules and data formats is not available in a corporate setting. Therefore, routine data quality audits should be performed on the data in the CRM model in order to identify data quality issues that are not addressed during the ETL process. For example, missing data can be identified during data quality audits and consequently addressed by consulting with a domain expert.

Some forms of dirty data (e.g., outliers) can be identified using data mining techniques and statistical analysis, while other forms of dirty data (e.g., missing data values) are more problematic. Although Dasu, Vesonder, and Wright (2003) assert that data quality issues are application-specific, Kim, Choi, Kim, and Lee (2003) developed a taxonomy of dirty data and identified methods for addressing dirty data based upon their taxonomy. Kim et al. (2003) identified three broad categories of dirty data: (1) missing data; (2) not missing, but wrong data; and (3) not missing and not wrong, but unusable data. They then further decompose each category of dirty data. However, they did not include composite types of dirty data. They provided some suggestions to address the issue of dirty data that can be used to clean dirty data in the CRM model during the ETL process. For example, the use of constraints can be valuable for avoiding instances of missing data (e.g., not null constraints) or incorrect data (e.g., domain ranges and check constraints). It

Table 7. Initial taxonomy of CRM analyses (S = strategic and T = tactical)

#	Decision Class	Category	Analysis	Potential Use(s)	KPI
1	S	Channel Analysis	Which distribution channels contribute the greatest revenue and gross margin?	Resource allocation	
2	S, T	Order Delivery Performance	How do early, on-time, and late order shipment rates for this year compare to last year?	Setting performance goals	early delivery, on-time delivery, late delivery
3	S	Order Delivery Performance and Channel Analysis	How do order shipment rates (early, on-time, late) for this year compare to last year by channel?	Setting performance goals, monitoring trends	early delivery, on-time delivery, late delivery
4	S	Customer Profitability Analysis	Which customers are most profitable based upon gross margin and revenue?	Classify customers	gross margin, revenue
5	S	Customer Profitability Analysis	What are the customers' sales and margin trends?	Classify customers	gross margin, revenue
6	S	Customer Retention	How many unique customers are purchasing this year compared to last year?	Identify the threshold to overcome with new customers	unique customers/year
7	S, T	Market Profitability Analysis	Which markets are most profitable overall?	Setting performance goals, allocate marketing resources	gross margin/ market
8	S, T	Market Profitability Analysis	Which products in which markets are most profitable?	Setting performance goals, allocate marketing resources	gross margin/ products/ market
9	S, T	Product Profitability Analysis	Which products are the most profitable?	Managing product cost constraints, identify products to potentially eliminate from product line	gross margin/ product
10	S, T	Product Profitability Analysis	What is the lifetime value of each product?	Managing product cost constraints, identify products to potentially eliminate from product line	gross margin/ product
11	S, T	Returns Analysis	What are the top 10 reasons that customers return products?	Create Pareto charts to identify problems to correct, setting performance goals	count
12	S, T	Returns Analysis	What is the impact of the value of the returned products on revenues?	Create Pareto charts to identify problems to correct, setting performance goals	count, revenue, profit

Table 7. Initial taxonomy of CRM analyses (S = strategic and T = tactical) (cont.)

#	Decision Class	Category	Analysis	Potential Use(s)	KPI
13	S, T	Returns Analysis	What is the trend for product returns by customers by product by reason?	Create Pareto charts to identify problems to correct, setting performance goals, identify problematic accounts (identify customers that may leave), assess additional service fees	count, revenue, profit
14	S, T	Customer Attrition	What are the top 10 reasons for customer attrition?	Insights for process improvements	attrition rate
15	S, T	Customer Attrition	What is the impact of the value of the customers that have left on revenues?	Insights for process improvements	attrition rate

is important to point out that the ability to enforce integrity constraints on inconsistent spatial data (e.g., geographical data such as sales territory alignments) and outdated temporal data is not supported in current database management systems.

Initial Heuristics for Designing CRM Data Warehouses

Once the types of CRM analyses that the data warehouse needs to be able to support have been identified, the data points have been identified, and the granularity has been selected, the next step is designing the data warehouse model to support the analyses that were identified. Based upon our initial findings, Table 8 lists initial heuristics for designing a data warehouse in order to successfully support CRM analyses.

CONCLUSION

In this article, we first present the design implications that CRM poses to data warehousing and then propose a robust multidimensional starter model that supports CRM analyses. Based upon sample queries, our model has a value of 1 and 0.93 for the $r_{success}$ and $r_{suitability}$ ratios, respectively. Our study shows that our starter model can be used to analyze various profitability analyses such as customer profitability analysis, market profitability analysis, product profitability analysis, and channel profitability analysis. In fact, the model has the flexibility to analyze both trends and overall lifetime value of customers, markets, channels, and products simply by including or excluding the time dimension in the SQL statements. Since the model captures rich descriptive non-numeric information that can be included in the query statement, the proposed model can return results that the user easily can understand. It should be noted that such rich information then can be used in data mining algorithms for such things as category labels. As such, we have demonstrated that the robust proposed model can be used to perform CRM analyses.

Our contributions also include the identification of and classification of CRM queries and their uses, including KPIs; the introduction of a sampling technique to select the queries with which the model is tested; the introduction of two measures (percent success ratio and CRM suitability ratio) by which CRM data warehouse models can be evaluated; and

Table 8. Initial heuristics for designing CRM DWs

#	Heuristic	Benefit
1	Identify the types of CRM analyses, their uses, and the data elements required to perform the analyses	The model will be able to support the intended purpose of the analyses
2	Include all attributes required to compute the profitability of each individual transaction in the fact table(s)	The ability to generate a profit and loss statement for each transaction, which then can be analyzed along any dimension
3	Each dimension that will be used to analyze the Profitability fact table should be related directly to the fact table	Provides improved query performance by allowing the use of simplified queries (i.e., support browsing data)
4	Pay careful attention to the Customer dimension	It forces attention to the customer to the center of CRM
5	Create a relationship between the Customer dimension and the Market and Sales Representative dimensions	Provides the ability to quickly determine the current market and Sales Representative for the customer by merely browsing the Customer dimension
6	Include the attrition date and reason for attrition attributes in the Customer dimension	Provides the ability to determine quickly if a customer is no longer a customer by browsing the Customer dimension only
7	Attributes that are likely to change at a different rate than other attributes in the same dimension should be in a separate dimension	Minimize the number of updates
8	Create a separate *existence* dimension for any entity that can have a discontinuous existence	Provides the ability to track the periods in which the instance of the entity is valid (needed to support some temporal queries)
9	Create a separate *existence* dimension for any attribute whose historical values must be kept	Provides the ability to track accurate historical values, even during periods of inactivity
10	Create a relationship between the Time dimension and each *existence* dimension	Provides the ability to perform temporal queries efficiently using descriptive attributes of the Time dimension
11	*Existence* dimensions should be in a direct relationship with their respective original dimensions	
12	There always should be a CustomerExistence dimension	The ability to track and perform analyses on customer attrition
13	If some products are either seasonal or if it is necessary to determine when products were discontinued, then create a Product Existence dimension	The ability to perform analyses for seasonal and discontinued products
14	There should be a Household dimension and an ExtendedHousehold dimension	Provides the ability to perform Household analyses
15	The organizational hierarchical structure can be contained in one *Market* dimension	Provides the ability to maintain a history of the organizational changes, and the ability to perform analyses according to the organizational structure

the identification of the initial heuristics for designing a data warehouse to support CRM. Finally, in terms of future work, we plan to classify and test additional CRM analyses, evaluate alternative models using the same set of queries and the $r_{success}$ and $r_{suitability}$ ratios, identify materialized views that are relevant to CRM, and explore CRM query optimization.

ACKNOWLEDGMENTS

The research of Peter Chen was partially supported by National Science Foundation grant ITR-0326387 and AFOSR grants F49620-03-1-0238, F49620-03-1-0239, and F49620-03-1-0241.

REFERENCES

Allen, G. N., & March, S. T. (2003). Modeling temporal dynamics for business systems. *Journal of Database Management, 14*(3), 21-36.

Boon, O., Corbitt, B., & Parker, C. (2002, December 2-3). Conceptualising the requirements of CRM from an organizational perspective: A review of the literature. In *Proceedings of 7th Australian Workshop on Requirements Engineering (AWRE2002)*, Melbourne, Australia.

Bose, R. (2002). Customer relationship management: Key components for IT success. *Industrial Management & Data Systems, 102*(2), 89-97.

Buttle, F. (1999). The S.C.O.P.E. of customer relationship management. *International Journal of Customer Relationship Management, 1*(4), 327-337.

Cunningham, C., Song, I-Y., Jung, J. T., & Chen, P. (2003, May 18-21). Design and research implications of customer relationship management on data warehousing and CRM decisions. In M. Khosrow-Pour (Ed.), *Information Technology & Organizations: Trends, Issues, Challenges & Solutions, 2003 Information Resources Management Association International Conference (IRMA 2003)* (pp. 82-85), Philadelphia, Pennsylvania, USA. Hershey, PA: Idea Group Publishing.

Dasu, T., Vesonder, G. T., & Wright, J. R. (2003, August 24-27). Data quality through knowledge engineering. In *Proceedings of SIGKDD '03*, Washington, DC.

Hawkes, V. A. (2000). The heart of the matter: The challenge of customer lifetime value. *CRM Forum Resources, 13*.

Hwang, H., Jung, T., & Suh, E. (2004). An LTV model and customer segmentation based on customer value: A case study on the wireless telecommunication industry. *Expert Systems with Applications, 26*, 181-188.

Jain, D., & Singh, S. (2002). Customer lifetime value research in marketing: A review and future direction. *Journal of Interactive Marketing, 16*(2), 34-46.

Kale, S. H. (2004, September/October). CRM failure and the seven deadly sins. *Marketing Management*, 42-46.

Kellen, V. (2002, March). *CRM measurement frameworks*. Retrieved from http://www.kellen.net/crmmeas.htm

Kim, W., Choi, B., Kim, S., & Lee, D. (2003). A taxonomy of dirty data. *Data Mining and Knowledge Discovery, 7*, 81-99.

Kimball, R., & Ross, M. (2002). *The data warehouse toolkit* (2nd ed.). New York: Wiley Computer Publishing.

Lee, Y. W., Pipino, L., Strong, D. M., & Wang, R. Y. (2004). Process-embedded data integrity. *Journal of Database Management, 15*(1), 87-103.

Massey, A. P., Montoya-Weiss, M. M., & Holcom, K. (2001). Re-engineering the customer relationship: Leveraging knowledge assets at IBM. *Decision Support Systems, 32*(2), 155-170.

Myron, D., & Ganeshram, R. (2002, July). The truth about CRM success & failure. *CRM Magazine*. Retrieved from http://www.destinationcrm.com/articles/default.asp?ArticleID=2370

Panker, J. (2002, June). Are reports of CRM failure greatly exaggerated? *SearchCRM.com*. Retrieved from http://searchcrm.techtarget.com/originalContent/0,289142,sid11_gci834332,00.html

Rosset, S., Neumann, E., Eick, U., & Vatnik, N.

(2003). Customer lifetime value models for decision support. *Data Mining and Knowledge Discovery, 7*, 321-339.

Rust, R. T., Lemon, K. N., & Zeithaml, V. A. (2004). Return on marketing: Using customer equity to focus marketing strategy. *Journal of Marketing, 68*(1), 109-139.

Shankaranarayan, G., Ziad, M., & Wang, R. Y. (2003). Managing data quality in dynamic decision environments: An information product approach. *Journal of Database Management, 14*(4), 14-32.

Todman, C. (2001). *Designing a data warehouse*. Upper Saddle River, NJ: Prentice Hall.

Verhoef, P. C., & Donkers, B. (2001). Predicting customer potential value: An application in the insurance industry. *Decision Support Systems, 32*(2), 189-199.

Winer, R. S. (2001). A framework for customer relationship management. *California Management Review, 43*(4), 89-108.

Colleen Cunningham (cmc38@drexel.edu) (master's of software engineering and master's of science in information science, Pennsylvania State University), doctoral student, College of Information Science and Technology, Drexel University. She is a full-time knowledge management analyst, PhD student, and adjunct faculty in the College of Information Science and Technology at Drexel University. Her teaching interests include database management systems and database administration. Her research interests include data warehousing to support customer relationship management (CRM), data mining, case-based reasoning (CBR), and knowledge management. Her work has appeared in the Proceedings of the 7th ACM International Workshop on Data Warehousing and OLAP*;* Advances in Case-Based Reasoning, Lecture Notes in Artificial Intelligence*;* Information Technology and Organizations*; and the* Encyclopedia of Data Warehousing and Mining.

Il-Yeol Song (song@drexel.edu) (PhD, Louisiana State University), professor, College of Information Science and Technology, Drexel University. Dr. Song's research focuses on practical application of modeling theory to real-world problems. His research interests include database modeling and design, design and performance optimization of data warehousing and OLAP, database support for Web-based systems, object-oriented analysis and design, and modeling, integration, and analysis of medical and bioinformatics data. Dr. Song published over 120 refereed papers. Dr. Song served as a PC co-chair for CIKM99, ER03, DOLAP98, and DOLAP 99. Dr. Song is a co-author of the Best Paper Award at 2004 IEEE CIBCB Symposium. Dr. Song also received three teaching awards at Drexel University, including Exemplary Teaching Award in 1992, Teaching Excellence Award in 2000, and the prestigious Lindback Distinguished Teaching Award in 2001.

Peter P. Chen (pchen@lsu.eduf) (PhD, Harvard University), Foster Distinguished Chair Professor, Computer Science Department, Louisiana State University. Dr. Chen's research focuses on developing methodologies on modeling the real world and on information system developments. His original paper on the Entity-Relationship (ER) model is one of the most cited papers in computer science and IT/IS and has been selected as one of 38 most influential papers in computer science. The ER Model is one of the most widely used models in database design, software engineering, and information system developments. He is a fellow of ACM, IEEE, and AAAS. He received many awards, including ACM/AAAI Allen Newell Award, IEEE Harry Goode Award, Stevens Award for Advancing Software Development Methods, DAMA Achievement

Award, and Taiwan's top high-tech award (Pan Wen-Yuan Outstanding Research Award). He is a member of the advisory committee of NSF Computer and Information Science and Engineering Directorate. He has been a keynote speaker for 30 international conferences. He is listed in Who's Who in America *and* Who's Who in the World*.*