

Data Warehouse Design to Support Customer Relationship Management Analyses

Colleen Cunningham
College of Information Science and
Technology
Drexel University
Philadelphia, PA 19104

Il-Yeol Song
College of Information Science and
Technology
Drexel University
Philadelphia, PA 19104

Peter P. Chen
Department of Computer Science
Louisiana State University
Baton Rouge, LA 70803

ABSTRACT

CRM is a strategy that integrates the concepts of Knowledge Management, Data Mining, and Data Warehousing in order to support the organization's decision-making process to retain long-term and profitable relationships with its customers. In this paper, we first present the design implications that CRM poses to data warehousing, and then propose a robust multidimensional starter model that supports CRM analyses. We then present sample CRM queries, test our starter model using those queries and define two measures (% success ratio and CRM suitability ratio) by which CRM models can be evaluated. We finally introduce a preliminary heuristic for designing data warehouses to support CRM analyses. 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.

Categories and Subject Descriptors

H.2.1 [Database Management]: Logical Design – Data models.

General Terms

Design

Keywords

Data Warehouse, Customer Relationship Management

1. INTRODUCTION

Acquiring new customers can cost five times more than it costs to retain current customers [9]. Furthermore, repeat customers can generate more than twice as much gross income as new customers [16]. 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. 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, 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.

CRM can be defined as a strategy that utilizes organizational knowledge and technology to enable proactive and profitable long-term relationships with customers [3]. It is a means of using knowledge management, or organizational knowledge, and technologies to enable organizations to make decisions about, among other things, product offerings, marketing strategies and how they interact with their customers. By utilizing a data warehouse, companies can make decisions about customer-specific strategies such as customer profiling, customer segmentation, cross-selling analysis, etc. For example, a company can use a data warehouse to determine its customers' historic and future values and to segment its customer base. Table 1 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 [2, 15]. The company could then 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 that do not have any value to the company.

Table 1: Customer Segments

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

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Dolap '04, November 12–13, 2004, Washington, DC, USA.
Copyright 2004 ACM 1-58113-977-2/04/0011...\$5.00.

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

Currently, however, there are no agreed upon standardized rules for how to design the 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 [10, 11]. 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 could then be used to guide CRM initiatives and to design and implement CRM data warehouses. The taxonomy could also be used to customize the starter model for a company's specific CRM requirements within a given industry. Furthermore, that taxonomy would also serve as a guideline for companies in the selection and evaluation of CRM data warehouses and related technologies.

In order to objectively quantify the success 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 successfully executed if the results that are returned are meaningful to the analyst. The *CRM success ratio* can not only be used to evaluate our proposed CRM model, but it can also 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.

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

Q_p : the total number of queries that successfully executed against the model

Q_n : 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.

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

N: the total number of applicable analysis criteria

C: individual score for each analysis capability

X: 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 determine the suitability of the model to their individually different CRM needs. We utilize the two metrics to evaluate the proposed CRM data warehouse model in our case study implementation.

The remainder of this paper is organized as follows: a brief review of CRM literature is presented in Section 2; the schema design is presented in Section 3; the experiment is described in Section 4; the results presented in Section 5 followed by the discussion in Section 6 and the conclusions in Section 7.

2. CRM LITERATURE REVIEW

For the analysis of CRM, we need to classify customers into one of the four quadrants in Table 1 and subsequently apply the appropriate strategy. In literature, researchers used the total historical value, total potential future value, and customer lifetime value (CLV). There are many ways of defining and calculating those measures [4, 5, 6, 12]. For the purposes of this paper, CLV will be the sum of the total historical value and the total potential value for each customer. The *total historical value* will be computed as follows:

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

j: 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, overhead cost, etc. The *potential future value* for a customer will be defined using the same equation as defined by [5]:

$$\text{Potential Future Value} = \sum_{j=1}^N (\text{Probability}_j \times \text{Profitability}_j) \quad (4)$$

j: individual products that the customer could potentially 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* [13].

One of the goals of companies should be to increase its 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 its 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 would include, but are not limited to: margins, on-time deliveries, late-deliveries and customer retention rates. Other KPIs that are relevant to CRM would include, but are not limited to: marketing cost, number and value of new customers gained, complaint numbers and customer satisfaction rates [7].

3. SCHEMA DESIGN FOR CRM

The first step to designing the schema for the CRM data warehouse is to identify the different types of categories of analyses that are relevant to CRM. Specific data points of interest are then identified from both experience and literature [1, 7, 13]. 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. It should be noted that there is no significance to the order in which the items are listed in Table 3.

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, phone, etc.)
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, the time it takes to resolve customer complaints, etc.
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.

3.1 Design Rationale for the Fact Tables

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 were included in Figure 1. The *profitability fact table* includes the attributes (e.g. revenues and all costs: distribution, marketing, overhead, product, etc.) 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, would improve 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 can be 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 identify internal areas for process improvements and ultimately influence customer satisfaction and possibly customer retention.

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	Stores the dynamic demographic characteristics of the customer.
CustomerExistence	Tracks the periods in which the customer is a 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 & regions in which the customer belongs
Product Dimension	Represents the products that the company sells
ProductExistence	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 *customer service fact table* contains information about each interaction with the customer, including the cost of the interaction, the time to resolve the complaint, a count of customer satisfaction or dissatisfaction, etc. The total historical value of each customer can be 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 (e.g. expected gross revenue, costs, expected purchasing frequency, probability of gaining additional revenue, etc.) that are needed to compute the potential future lifetime value for each customer. 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 can be computed by summing the historical value for each customer and the future value for each customer.

3.2 Design Rationale for the Dimensions

Initially, the dimensions were identified according to the dimensions along which the fact tables should be analyzed (e.g. Customer, Product, Supplier, Channel, Promotion, Market, Sales Representative and Time). Each dimension was carefully examined to determine if the dimension:

- a. contained attributes that were likely to change at a different rate than the other attributes within the dimension
- b. contained attributes whose complete set of historical values had to be maintained, or
- c. was subject to discontinuous existence (i.e. only valid for specific periods).

If any of the above properties was applicable, then a separate dimension was created (called an *existence* dimension in the case of (b) and (c)). Furthermore, in the case of (a), the new dimensions were implemented as minidimensions as opposed to outriggers in order to allow the user to readily browse the fact table. An additional benefit of this approach was that the history of the changes in the customer's behavior scores and demographics will be stored as part of the fact table, which would facilitate robust analyses without requiring the use of Type 1, 2 or 3 techniques [8] for the Customer Demographics or Customer Behavior dimensions.

In the case of (b) and (c), the dimensions were implemented as outriggers, and two relationships were created between the Time dimension and each trigger dimension. The two relationships were formed in order to record the date period in which the data instances were valid. In doing so, this facilitated the ability to perform state duration queries and transition detection queries [14]. 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 [14].

Careful consideration was given to this step in the design process because the fact table can only 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 which one has not recorded.

Direct relationships were formed between the Customer dimension and the Sales Representative, Market, Comment and Time dimensions. This was done 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. 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.

4. EXPERIMENT

4.1 Implementation

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

4.2 Methodology

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

4.3 Selection of Queries to Test

It is believed that the proposed data warehouse schema would have a positive impact on the ability to perform CRM analyses. Therefore, in order to avoid any biases in the types of queries used to test the model, stratified random sampling was used to select the specific queries used in the experiment. The stratified random sampling was conducted as follows: (1) representative queries used in CRM were gathered from independent associations; (2) the queries were grouped into categories based upon the nature of the query; (3) within each category, each query was numbered; (4) a random number generator was used to select queries from each category; (5) the queries whose assigned number corresponded to the number generated by the random number generator were selected. The specific queries that were selected are listed in Table 5.

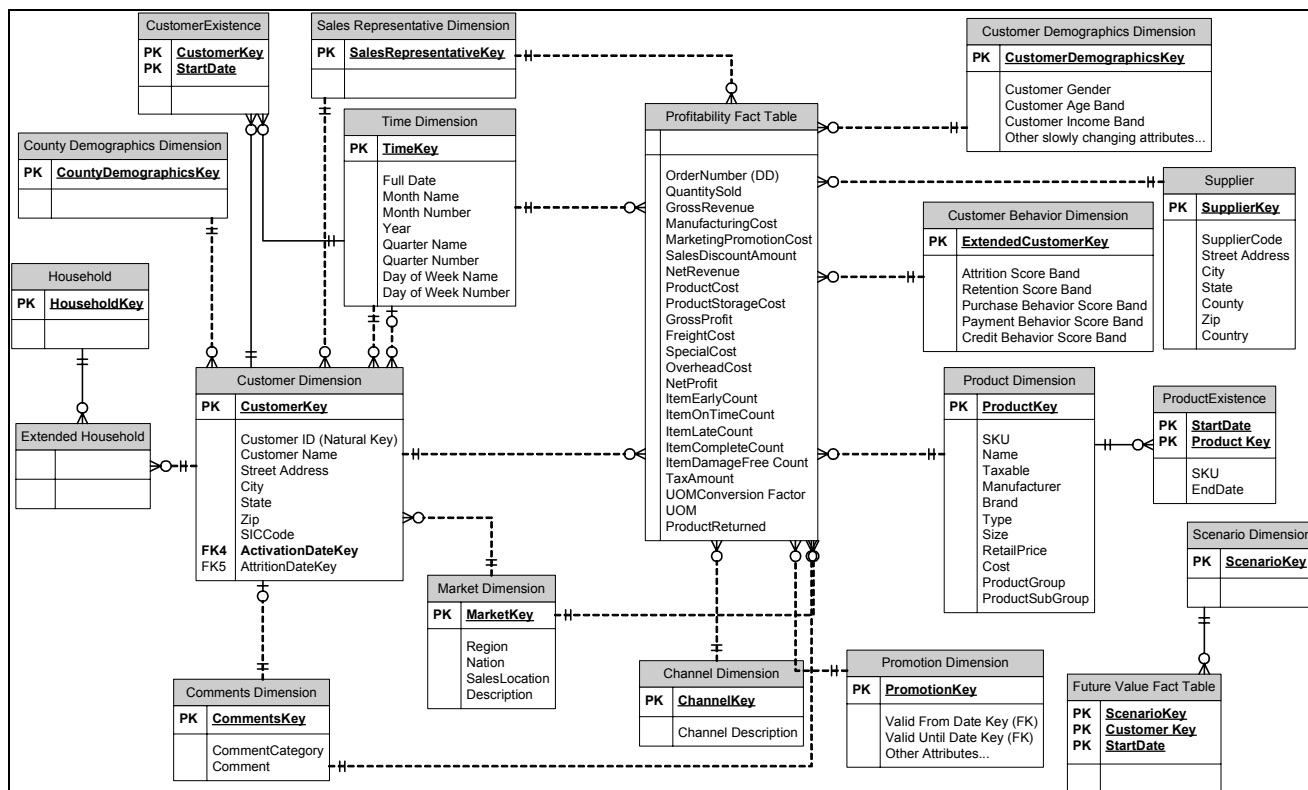


Figure 1: Proposed CRM Data Warehouse Model

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 & 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

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	1
6.5	Ability to segment customers into multiple customer	1
6.6	Ability to maintain the history of customer segments and	1
6.7	Ability to evaluate different campaigns and responses	1
6.8	Ability to analyze metrics for website	0
6.9	Ability to understand loyalty patterns among different	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	1
6.15	Ability to efficiently update/maintain data.	1
Total		14

It is important to note that since the queries were 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 were queries that equally applied to different industries and not queries that were specific to only one industry. This aspect of the sampling procedure was 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.

5. 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 had 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 – 6.11 in Table 6 were based upon whether or not queries were successfully executed in those categories. The scores for items 14 and 15 were determined while loading data from multiple sources and updating customer scores.

6. DISCUSSION

6.1 General Discussion

It was determined that the starter model depicted in Figure 1 could 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 by simply including the appropriate dimensions in the query

statement. Furthermore, each query could be modified to include additional measures and descriptions by simply including additional fields from the fact table and the dimensions. Some of those queries are discussed below.

The following SQL statement was used to identify the most profitable customers based upon total revenue and gross margin. By excluding the time dimension, the customer profitability SQL statement identified 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, can be used to classify customers in one of the four CRM quadrants (see Table 1), which can be subsequently used to determine the appropriate strategy for managing the customer.

```
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 2: Customer Profitability Analysis Query - Which customers are most profitable based upon gross margin and revenue?

The SQL statement in Figure 3 could be used to determine the margins for each product, and subsequently used to identify products that potentially may be eliminated from the company’s product line.

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

```
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 3: Product Profitability Analysis Query - Which products in which markets are most profitable?

In terms of items 5.11 – 5.13 in Table 5, not only could the return analyses be used to monitor the impact of corrective actions, but they could also be used to identify improvement targets, which could be tied to employee (and or departmental) performance goals. Additionally, item 5.13 from Table 5 could be used to identify whether or not a problem is systematic across all customers, many customers or a few specific customers. This query could also be used to help management make an informed decision with respect to allocating resources to address problems that lead to customers returning products. Moreover, the results

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 & 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
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	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

could be used by the sales team to gain further insight into why their customers have returned products. The sales team could potentially use that information to work with the customer to resolve the issue(s) in cases where the customer has repeatedly returned products for reasons that cannot be considered the

company's mistake. Alternatively, the sales team could use the results to identify accounts that could (should) be charged additional fees if the customer repeatedly returns products. Table 7 summarizes some of the possible uses for the CRM analyses that were presented in Table 5.

Table 8: Initial Heuristics for Designing CRM DWs

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

6.2 Initial Heuristics for Designing CRM Data Warehouses

Once the types of CRM analyses that the data warehouse needs to be able to support has been identified, the data points identified and the granularity 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 our initial heuristics for designing a data warehouse to successfully support CRM analyses.

7. CONCLUSIONS

In this paper, we first presented the design implications that CRM poses to data warehousing and then proposed a robust multidimensional starter model that supports CRM analyses. Based upon the sample queries, our model had a value of 1 and 0.93 for the $r_{success}$ and $r_{suitability}$ ratios, respectively. Our study showed that our starter model could be used to analyze various

CRM analyses, including, but not limited to: 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 by simply 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 are easily understood by the user. It should be noted that such “rich” information could then 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 included 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 was tested; the introduction of two measures (%)

success ratio and CRM suitability ratio) by which CRM data warehouse models can be evaluated; and the identification of the initial heuristics for designing a data warehouse to support CRM.

8. ACKNOWLEDGEMENT

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.

9. REFERENCES

- [1] Boon, Olaf, Corbitt, Brian and Parker, Craig (2002). Conceptualising the requirements of CRM from an organizational perspective: a review of the literature. *Proceedings of 7th Australian Workshop on Requirements Engineering (AWRE2002)*, December 2-3, Melbourne, Australia.
- [2] Buttle, Francis (1999), The S.C.O.P.E. of Customer Relationship Management. *International Journal of Customer Relationship Management* **1** (4), 327-337.
- [3] Cunningham, Colleen, Il-Yeol Song, Jin Taek Jung and Peter Chen (2003), Design and Research Implications of Customer Relationship Management on Data Warehousing and CRM Decisions. In *Proceedings of the 2003 Information Resources Management Association International Conference (IRMA 2003)*, 82 – 85.
- [4] Hawkes, Valoris Abram (2000). The Heart of the Matter: The Challenge of Customer Lifetime Value. *CRM Forum Resources*, Issue 13.
- [5] Hwang, Hyunseok, Jung, Taesoo and Suh, Euiho (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.
- [6] Jain, Dipak and Singh, Siddhartha (2002). Customer Lifetime Value Research In Marketing: A Review and Future Direction. *Journal of Interactive Marketing* **16** (2), 34-46.
- [7] Kellen, Vince (2002). CRM Measurement Frameworks. <http://www.kellen.net/crmmeas.htm>.
- [8] Kimball, Ralph and Margy Ross (2002), *The Data Warehouse Toolkit* (second edition), Wiley Computer Publishing, New York.
- [9] Massey, Anne P., Mitzi M. Montoya-Weiss and Kent Holcom (2001), Re-engineering the customer relationship: leveraging knowledge assets at IBM. *Decision Support Systems* **32** 2, 155 – 170.
- [10] Myron, David and Ganeshram, Ramin (2002). The Truth About CRM Success & Failure. *CRM Magazine*, July 2002.
- [11] Panker, Jon (2002). Are Reports of CRM Failure Greatly Exaggerated? SearchCRM.com, June 20, 2002 http://searchcrm.techtarget.com/originalContent/0,289142,si_d11_gci834332,00.html.
- [12] Rosset, Saharon, Neumann, Einat, Eick, Uri, Vatnik, Nurit (2003). Customer Lifetime Value Models for Decision Support. *Data Mining and Knowledge Discovery*, **7**, 321-339.
- [13] Rust, Roland T., Lemon, Katherine N. and Zeithaml, Valarie A. (2004). Return on Marketing: Using Customer Equity to Focus Marketing Strategy. *Journal of Marketing*, **68**, 1, 109 – 139.
- [14] Todman, Chris (2001). *Designing a Data Warehouse*, Upper Saddle River, NJ: Prentice Hall.
- [15] Verhoef, Peter C. and Bas Donkers (2001), Predicting customer potential value an application in the insurance industry. *Decision Support Systems*, **32**, 2, 189 – 199.
- [16] Winer, Russell S. (2001), A framework for customer relationship management. *California Management Review*, **43**, 4, 89 – 108.