http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 5 (5): 964-969, 2006
ISSN 1812-5638

Data Warehouse Design for Sales Performance Analysis

J.A. Nasir, M.K. Shahzad and M.A. Pasha
Intelligent Information Systems Research Group, Punjab University
College of Information Technology, University of Punjab, Lahore, Pakistan

Abstract: Sales-Performance Analysis strategy integrates the concepts of Data Mining, Analytical Processing and Data Warehouse to support organization's decision-making process. In this study we have presented analytical-requirements and design-issues of sales-performance analysis for data warehouse design. For these requirements and issues, modern decision-design based experimental model has been anticipated and evaluated. For evaluation, two parameters (success, suitability ratio) has been defined and used. The present study shows, 1) The experimental model provides better scalability, for analysis, mining and reporting, than the formal decision designs 2) Experimental model can be used for various Sales-Performance Analyses like product profitability, channel analysis, product sales analysis etc.

Key words: Scalability, sales-performance, data warehouse, experimental model

INTRODUCTION

Data Warehouse (DW) provides a foundation for various types of sales forecast reporting and analysis (Torben and Christian, 1999). This forecast reporting and analysis provide visibility into a company's sales pipeline, integrating information from sales, customer and financial sources for a complete picture of sales performance.

Business intelligence-enabled sales forecasts allow sales-management to monitor and act on individual opportunities, more accurately forecast, current and future period revenues and understand the drivers that distinguish won versus lost deals. Executives can use graphical dashboards to quickly assess actual sales performance versus corporate targets and sales management forecasts. Marketing users can analyze lead progression through each stage of the sales cycle to quantify the effectiveness and revenue impact of marketing efforts. Business Intelligence enables organization to associate sales pipeline data with financial, marketing and customer information to make informed, strategic decisions to improve sales effectiveness.

Retailers commonly use Sales-Performance Analysis (SPA), which enables the continuous monitoring of sales performance to uncover sales trends, investigate product demand and optimize merchandising strategies. Various levels of analysis from summary reporting to statistical trending are required by executives, decision makers, sales manager as well as marketing analysts to analyze company's sales performance.

Our proposed EM provides a foundation to solve diverse business challenges: a) EM provide knowledge of when where why and what sold to ensure optimal stocking, merchandising and sales. b) Provide support to develop strategies to detect and mitigate fraud by statistical analysis c) Can support manager to analyze the effectiveness of existing programs to optimize their marketing spend and better merchandising. d) EM also has the ability to act as a plinth for altering managers to potential lost sales. e) Provides support for analyzing sales pipeline data from different business prospective f) Provide support for forecasting effectively through trend analysis and close probabilities.

The design of SPA-DW has direct impact on organization's ability to eagerly perform analysis specific to product sales. However, there is no standardized rule for how to design DW for SPA. Subsequently, the design of the DW model contributes to the success or failure of PSA. Various types of schemas In fact, recent statistics indicate that between 50 and 80% of the initiatives fails due to inappropriate or incomplete SPA processes and poor selection of technologies. Thus, the ultimate long-term purpose of our studying is to systematically examine SPA factors that affect design decisions for SPA data warehouse in order to determine the impact of those analyses on SPA data warehouse design decisions.

To quantify the success of the proposed EM, two metrics are recommended:

- PSA Success Ratio (SCR)
- PSA Suitability Ratio (SUR).

Corresponding Author: M.K. Shahzad, Intelligent Information Systems Research Group,
Punjab University College of Information Technology, University of Punjab, Lahore, Pakistan

The PSA success ratio is defined as 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 PSA success ratio (SCR) cannot only be used to evaluate our proposed Starter model. But it can also be used to evaluate other SPA-DW models as well. The range of values for SCR is between 0 and 1.

The larger the value of SCR, the more successful the model.

$$SCR = Qp/Qn$$

Qp: The total number queries that executed successfully against the model

Qn: The total number of queries issued against the model

The PSM suitability ratio (SUR) is defined as the ratio of the sum of the individual suitability scores to the sum of the number of applicable categories.

The larger the value of SCR, the more successful the model

$$SUR = \sum\nolimits^{n}_{i=1} (X_{iy} C_{ik}) / \sum\nolimits^{n}_{i=1} (X_{iq} C_{iz})$$

$$; y = \{y \mid y \in N \wedge 0 \leq y \leq 5\}$$
$$\text{and } 0 \leq k \leq z$$

- i : is the question number posed to evaluate the model.
- y : is the weight assigned to the question. The weight of question can vary from 0 to 5.
- k : is the max. possible grade which can be given against the question Q_i.
- q : is the max. possible weight which can be given to the question Q_i.
- X_{iy} : is the weight assigned to each question Q_i to evaluate the model.
- C_{iz} : is the grade given for question Q_i. Its value depend upon the amount to which model meet the requirement of question.
- C_{ik} : is the maximum possible score of the question Q_i.

The range of values for the SUR ratio is minimum 0, with values away from 0 suitability is increased. Unlike the SCR ratio, which can be used to evaluate and compare the richness and completeness of PSA data warehouse models, the SUR ratio, however, can be used to help companies determine the suitability of the model to their individually different PSA needs. We can utilize the two metrics to evaluate the proposed SPA DW model in any case study implementation.

ANALYTICAL DESIGN REQUIREMENTS

The pre-requisite step to SPA DW schema-design is to identify the different types and categories of analyses that are relevant to sales performance (International PM, 2005). Here we are identifying the types of analyses that are relevant to SPA as well as some of the data maintenance issues that must be considered. In other words, Table 1 identifies the minimum design requirements for a SPA-DW. It should be noted that there is no significance to the order in which the items are listed in Table 1 and some analysis (1.12 and 1.13) are common in customer relationship and sales performance. So they are directly taken from (Colleen *et al*., 2004).

Table 1: Minimum design requirements for SPA-DW

Analysis type/date maintenance	Description
Product profitability	Ability to determine profitability of each market
Market analysis	Ability to determine profitability of each market
Inventory analysis	Ability to track customer retention
Merchandizing analysis	Ability to identify root causes for customer attrition
Product score	Ability to score customers
Loss prevention analysis	Ability to associate customers with multiple extended household accounts
Fraud and theft analysis	Ability to segment customer into multiple customer segmentations
Demographic analysis	Ability to perform demographic analysis
Trend analysis	Ability to perform trend analysis
Product sales performance	Ability to evaluate on-time, late and early product deliveries
Suppliers return	Ability to analyze the reasons for and the impact of products being returned
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.
Up-selling analysis	Ability to analyze opportunities for customers to buy larger volumes of a product or a product with a higher profitability margin
Cross-selling analysis	Ability to identify additional types of products that customers that customers could purchase, which they currently are not purchasing.
Web based analysis	Ability to analyze metrics for web site
Data storage	Ability to maintain the history of customer segments and scores.
Data integration	Ability to integrate data from multiple sources, including external sources
Sales pipeline analysis	Ability to efficiently update/maintain data.
Competitor analysis	
Sales force analysis	

SALES PERFORMANCE DESIGN ISSUES

Until now, we have identified the design requirements of the SPA DW schema-design. Before designing the schema it is required to consider the following design issues:

a. Dimensions contain attributes that were likely to change at a different rate than the other attributes within the dimension.
b. Dimension may contain attributes whose complete set of historical values had to be maintained.
c. It may subject to discontinuous existence (i.e., only valid for specific periods).
d. Schema may contain facts, which would act as dimensions (i.e., sales a fact can also act as dimension if user want to analyze profit on the basis of sales).
e. Many to many relationships, between facts and dimensions can also exist in the schema.
f. Schema should handle strict and non-strict hierarchies.
g. Changes in data overtime e.g., changes in the organizational hierarchies should be accommodated.
h. Multiple hierarchies may exist in each dimension (different aggregation paths).
i. Non-onto hierarchies (non balanced instances tree).
j. Explicit hierarchies in dimensions.
k. Symmetric treatment of dimensions and measures

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 mini-dimensions 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 (Paulraj, 2000) 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 outrigger 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 (Colleen *et al*., 2004). 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.

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.

DATA WAREHOUSE SCHEMA FOR SALES PERFORMANCE ANALYSIS

The model consists of a profitability fact table, a future value fact table, a product sales fact table and various dimensions, which are defined in Table 2. We note that not all of the fact tables and dimensions were included in Fig. 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. Moreover, the model depicted in Fig. 1 can be used to calculate KPIs for

Table 2: EM dimensions for SPA

Dimension name	Dimension definition
Channel dimension	Keeps track of communication channels and their description
Customer dimension	Has basic information about the customer
Customer behavior (level)	Contains behavioral information about customer
Customer demographics	Contains personal information about the customer, e.g. age, sex, income etc.
Customer start (level)	Level has the joining date of the customer.
Market dimension	Gives complete detail of the existing and target market.
Comments	For better customer analysis, description of individual comments is stored in this dimension.
Company representative	Provides a base for reps information, details can be added in the same or new level.
Product dimension	Has product details about a product e.g., cost, category, price
Product existence (level)	In order to keep track of the starting and ending date of the product this level is used.
Promotion dimension	Has promotions given along with their validity durations. i.e., from and to date.
Supplier dimension	Tracks name, address, city, state and conntry of the supplier
Time dimension	Is the standard dimension of time

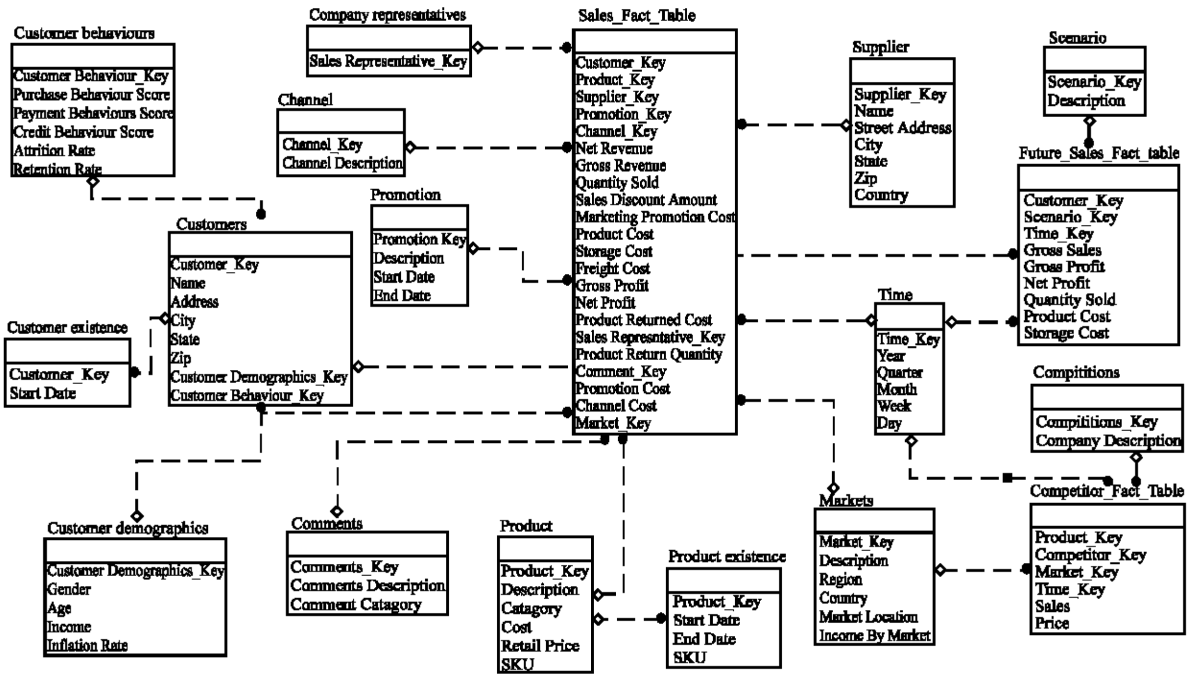

Fig. 1: Proposed EM for sales performance analysis

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 sales increase.

The sales 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 Eq. 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.

EXPERIMENT

We performed a study to test the scalability of the proposed EM. Proposed model was implemented in SQL Server 2000 running on Windows 2000. For testing purpose, a series of SPA queries were executed against the proposed 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. Proposed EM was tested to determine if it could or could not perform the tasks listed in Table 3.

For each task the model could perform, it was given a score from 0-5.

By using the above-mentioned values and suitability ratio formula SUR can be calculated.

$$SUR = \sum^{n}_{i=1} (X_{iy} C_{ik}) / \sum^{n}_{i=1} (X_{iq} C_{iz})$$

Table 3: Sample suitability scores

Criteria	Score	W
Trend analysis ability	5	5
Product profitability analysis	4	3
Integration ability	4	5
Efficiently update/maintain data	5	2
Demographic analysis ability	2	4
Ability to analyze metrics for website	0	4
Campaigns and responses analysis ability	3	3
Customer history maintenance	5	5
Total	7	31

Table 4: Sample success ratio

Category	Analysis	Pass/Fail
Product profitability analysis	Which products are the most profitable?	1
Product profitability analysis	What is the lifetime value of each product?	1
Market profitability analysis	Which products in which markets are most profitable?	0
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
Channel analysis	Which distribution channels contribute the greatest revenue and gross margin?	1
Returns analysis	What are the top 10 reasons that customers return products?	1
	Total	5

Table 5: Sample analysis queries for comparative scalability

Analysis queries	MDD-S OLAP	FDD-S OLAP
Which customers are very profitable?	1	1
Amount of customer at a certain date?	1	1
Variation of the customers who are adding value to business	1	0
Dynamics of product quantity in stock?	1	1
Customer behavior Analysis	0	1
Employee's promotion and returns performance	1	1
Reliable suppliers of the area	1	1
Channel variation and utilization performance	1	1
Competitors policy analysis and their reflections	1	0
Area wise sales performance	1	1
Sales performance analysis over periods of time	1	0
Sales representative performance analysis	1	0
Detailed product return analysis	0	1
Brief product return analysis	1	0
Customer demographics analysis	1	1
Customer demographics analysis with competitive sales	1	0

MDD-S OLAP: Scalability of Modern Decision Design for OLAP, FDD-S OLAP: Scalability of Formal Decision Design for OLAP

Table 6: Sample suitability scores

	No. of SA	No. of Non-SA	Scal (%)
MDD			
OLAP analysis	74	6	92.5
KM	70	10	87.5
Reporting	48	32	60
FDD			
OLAP analysis	57	23	71.25
KM	49	31	61.25
Reporting	44	36	55

MDD: Is for Modern decision design, FDD: Is for Formal decision design, No. of SA: Is for number of queries for which design is found scalable, No. of Non-SA: Is for number of queries for which design is not found scalable, Scal %: Gives the scalability percentage

$$SUR = (5*5)+(4*3)+(4*5)+(5*2)+(2*4)+(0*4)+(3*3)+(5*5)/ (5*5)+(3*5)+(5*5)+(2*5)+(4*5)+(4*5)+(3*5)+(5*5)$$

$$SUR = 109/155 = 0.703$$

It is important to note that since the queries were randomly selected from a pool of related queries. It is also important to note that the representative SPA queries were queries that equally applied to different industries and not queries that were specific to only one industry. The success score is 5; while the ratio is 0.83 (Table 4).

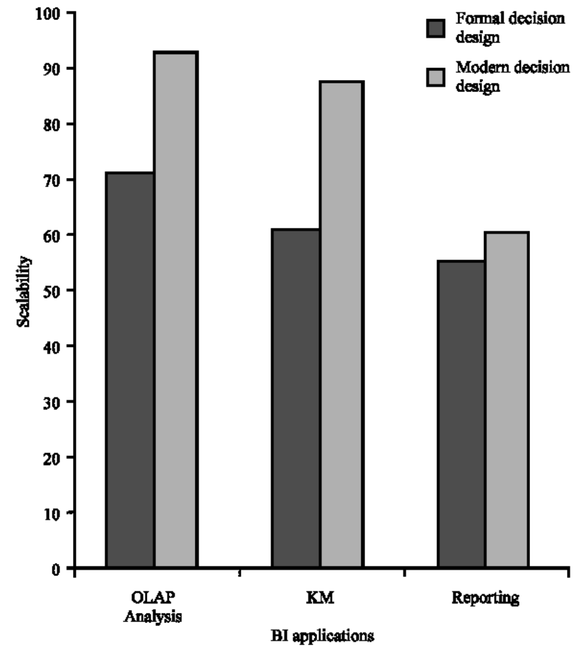

Fig. 2: Decision design comparison

Scalability is also the strength of modern decision design. For scalability comparison of modern and formal decision design, we have selected a pool of 80 most widely used SPA analysis queries for all three types of BI applications (analysis, mining and reporting). These queries were run, straightforwardly and with changes, on formal and modern decision design in order to evaluate their scalability and then compare them.

Some of queries and their scalability/non-scalability are given in Table 5. In the Table 5 MDD-S is the modern decision design scalability, while FDD-S is the formal decision design scalability. These analysis queries were checked for all three BI applications and sample's evaluation is given for OLAP analysis. Value of scalability is considered to be 0 or 1, where 0 indicates that, design does not support provides scalability. Similarly, 1 shows that analysis queries can be executed in the presence of changes and design is flexible (Table 5).

Table 6 shows the number of scalable analysis, non-scalable analysis and percentage of scalability. It is to be noted that these facts are based on our study of 80 queries. Also, the success ratio formula can be used to calculate the suitability ratio, by assigning different values to questions.

Figure 2 shows the comparison of the percentage scalability of modern and formal decision design. It clearly indicates the increased scalability of modern decision design for business intelligence i.e., for OLAP analysis, KM and reporting.

CONCLUSIONS

In present study, we first presented the design implications and then proposed EM that supports SPA analysis. Based on sample queries, our model's value has been calculated. The present study shows that this model provides better scalability than the formal decision designs and EM can be used to analyze various SPA's. Also it can be made flexible by using the values of SUR and SCR.

REFERENCES

Colleen, C., II-Y. Song and P.P. Chen, 2004. Data warehouse design to support customer relationship management analysis. Proceedings of DOLAP'04, Washington, DC, USA.

International Public Management Association, 2005. Best practices in HR. http://www.ipma-hr.org/index.cfm?navid = 127 Accessed on 9 August 2005.

Paulraj, Ponniah, 2000. Data Warehousing Fundamentals. John and Willey Publishers, pp: 38.

Torben, B.P. and S.J. Christian, 1999. Multidimensional data modeling for complex data. Proceedings of 15th ICDE, Sydney, Australia.