FUNCTIONAL BEHAVIOR PATTERN FOR DATA MART BASED ON ATTRIBUTE RELATIVITY

Paulraj M and Sivaprakasam P
Department of Computer Science, Vinayaka Missions University
Salem, Tamil Nadu, India
Department of Computer Science, Sri Vasavi College
Erode , Tamil Nadu – 638 316, Tamil Nadu, India

ABSTRACT

The growing need for huge volume of data in enterprise and corporate environment, fuel the demand of data warehousing. Data warehousing collects the data at different levels (i.e., departmental, operational, functional) and stored as a collective data repository with better storage efficiency. Various data warehousing models concentrate on storing the data more efficiently and quickly. In addition accessibility of data from the warehouse needs better understanding of the structure in which the data layers are stored in the repository. However function requirements of users are not easily understood by the data warehouse model. It needs efficient decision support system to extract the required user demanded data from data warehouse. To handle the issue of functional decision support system to extract user relevant data, data marts are introduced. Data marts built separate functional data repository layers based on the departmental decision support requirements in the enterprise and corporate data applications. In our research work, we plan to built a Functional Layer Interfaced Data Mart Architecture (FLIDMA) to provide a better decision support system for larger corporate and enterprise data applications.In this work, the functional behavior of the corporate system is analyzed, based on its operational goal to build layers of data storage repositories with relevant data attributes using functional behavior pattern (FBP). An experimental evaluation is conducted with benchmark datasets from UCI repository data sets and compared with existing multi-functional data warehousing model in terms of number of functional data attributes, attribute relativity, analysis of functional behavior.

Key Words: Data warehouse, Data mart, Fuctional behavior, Functional attributes, Decision support, Inductive rule mining

1. INTRODUCTION

A data warehouse is a database utilized for exposing and analyzing the data in the database. The data gathered in the warehouse are uploaded from the equipped systems such as market place. The data may exceed throughout an equipped data hoard for further operations prior to they are used in the DW for reporting. A data warehouse created from incorporated data source systems does not need ETL, producing databases, or operational data store databases. Data warehouses are subdivided into data marts. Data marts hoard subsets of data from a warehouse. An overview of data warehouse with data mart is shown in fig 1.

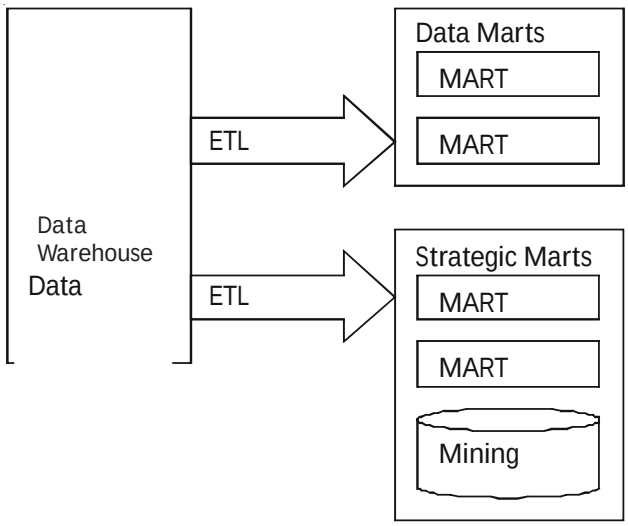

Fig 1 Data Mart in Data warehouse

A data mart is the entrée deposit of the data warehouse situation that is utilized to obtain data not in to the users. The data mart is a division of the data warehouse which is habitually oriented to a definite business line. In some instances, each subdivision or industry unit is measured the owner of its data mart together with all the hardware, software and data. This facilitates each subdivision to utilize, operate and extend their data without changing information within other data marts or the data warehouse. The functional behavior of each data mart is analyzed and used in such a way. The merits of creating data mart is as follows

- Can access regularly desired data easily
- Generates combined view by a group of users
- Enhances end-user response time
- Lesser cost than employing a complete data warehouse

Since the data mart provided an improved end user response time, the processing of query in data mart consumes less time to access it. The data mart is a subset of data warehouse and the data is stored based on its behavior and grouped under a name in such a way. To analyze the attributes in the data mart, and classified the attributes based on its functional behavior of the database, in this work, we are going to present a new technique for managing the data in data mart by analyzing the functional behavior of the at

2. LITERATURE REVIEW

Promising applications needs an efficient support of database management systems in real world for storing the data applications. Data in the data warehouse are accessed by queries. Researchers introduced a new concept [2] at receiving appropriate K answers for status responsive queries. These appropriate K answers much close and the operation is receiving the queries by using genetic algorithm (GA).

In order to convene the high demand of applications, Active Data Warehousing has appeared as a substitute to conservative warehousing practices [1]. Different data warehouse system architectures like independent data mart, dependent data mart, and heterogeneous distributed data warehouse [8] are explained and compared under different systems. The issues of "increasingly" union relations whose records are constantly recovered from remote sources during an uneven network that may acquire momentary failures are considered. To overcome the issues raised in the different warehouses, a new technique [3] is developed RPJ (Rate-based Progressive Join) based on solid hypothetical investigation. A data warehouse development and operations are carried over by several techniques [10]. A necessitate for on-line warehouse possessed several challenges and focused on regularly encountered procedure in this framework, specifically, the join [4] of a fast stream of source updates with a disk-based relation. The design and development of a data warehouse model is presented in [11]. To integrate information source reliability [12], real world data has been used. The data warehouse is subdivided into data marts for effective storage repositories.

Since the database is only applied to the raw data, [5] presented a multidimensional data model and execute the resolution of data mart. A development in IT started to utilize the databases are called Data Marts (DM). The information in Data mart is examined [6] and how real time information established and contacted through decision making process. In data mart architecture [7], web log session analyzer is used for identifying the interactive patterns of web users. DaWall (Data Warehouse IntegratIon) [9] is a tool supporting the diverse behaviors linked to the incorporation of multidimensional data present in the data marts. Previous literary works are generally discussed the problems of data marts. To identify each functional behavior of data mart, in this work, we are going to implement a new technique functional behavior pattern for data mart based on attribute relativity.

3. FUNCTIONAL BEHAVIOR PATTERN FOR DATA MART BASED ON ATTRIBUTE RELATIVITY

The proposed FBP of data mart is efficiently desgined for building the data mart components with a well-known arributes behavior of the data stored in the data warehouse. A data warehouse maintained the data in a arbitrary manner and a data mart which stores the data based depends upon the functional behavior of the data used in the data warehouse environment. The functional behavior of the data is identified based on choosing the relevant attributes of the system. The architecture diagram of the proposed FBP for data mart based on attribute relativity is shown in fig 3.1.

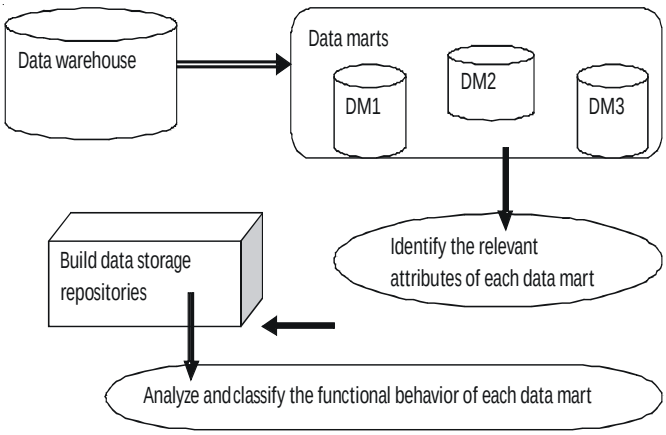

Fig 3.1 Architecture diagram of FBP for data mart

3.1 Overview of Data Mart

A data mart is a straightforward structure of a data warehouse that is attended on a distinct subject (or functional area), such as Finance, Sales, or Marketing. Data marts are frequently constructed and forbidden by a particular division within an organization. Specified their single-subject hub, data marts regularly illustrate data from only a small amount of sources. The sources could be inner outfitted systems, a central data warehouse, or external data. The data mart architecture is shown in fig 3.2. The major steps in implementing a data mart are

i) To design the schema,
ii) Construct the physical storage,
iii) Populate the data mart with data from source systems,
iv) Access it to make informed decisions, and
v) Manage it over time.

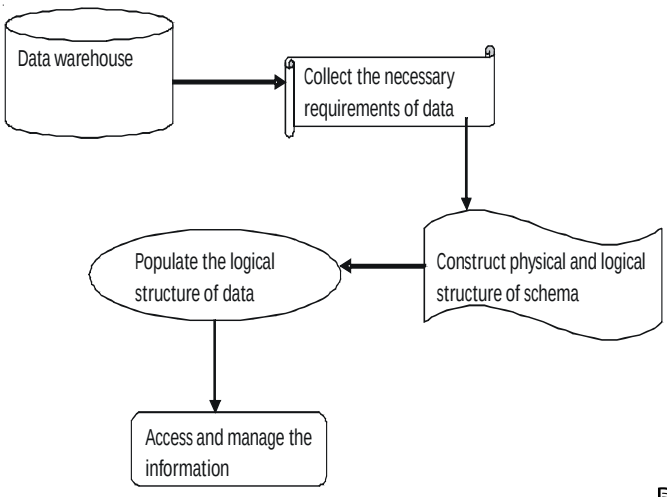

Fig 3.2 Overview of Data Mart

A design schema of the data mart is done by collecting all the necessary requirements of the user and identifying the data sources of the collected data. After identifying the sources of data, an appropriate subset of data is chose and the physical and logical sructure of data mart is analyzed and designed. Creating the physical and the logical construction connected with the data mart to present fast and competent access to the data in data mart.

A construction of physical storage is done by identifying schema objects, such as tables and indexes defined in the design step and determining the design structures of the data stored in the data mart. The populating step specifies all the tasks associated to receiving the data from the resource, transforming it to the exact layout and affecting it into the data mart. The populating step involves the following tasks:

- Planning data sources to objective data structures
- Retrieving data
- Transforming the data
- Loading data into the data mart
- Producing and hoarding metadata

The accessing scheme involved querying the data, examining it, producing reports, charts, and graphs. Normally, the end user utilizes a graphical front-end device to suggest queries to the database and exhibit the outcomes of the queries. The accessing step performed the following tasks:

- Set up a transitional deposit for the front-end tool to utilize. This interprets database structures and object names into business terms.
- Preserve and handle business interfaces.

The management tasks of the data mart is performed using the following tasks by

- Giving protected access to the data
- Organizing the development of the data and Optimizing the system for enhanced performance
- Ensuring the accessibility of data still with system failures

Finally, the data mart is efficiently analyzed and processed over different steps in order to obtain a data mart with a specified list of attributes.

3.2 Identifying the functional behavior pattern

After managing the information in the data mart, the relavant attributes of each data mart is chosen. Based on those relavant attributes of data mart, almost the operational goal of the data mart is identified. Based on the operational goal, the functional behavior of the data mart is analyzed and stored as a data repositories in the layer. The procedure below describes the analysis of the proposed functional behavior pattern of the data mart.

Given: Data warehouse (DW)

Step1 : Sub-divide the DW into a set of Data marts (DM)

Step2 : Let the data mart be sales

Step3 : DW is formed using a set of schemas

Step4 : [described briefly in section 3.1]

Step5 : After the formation of data mart (sales details)

Step6 : Identify the relevant attributes of data mart such as product details, price etc.,

Step7 : Store the relevant attributes of particular data mart in a data repository

Step8 : Based on relevant attributes,

Step9 : Identify the functional behavior of the data

Step10 : Based on functional behavior of data,

Step11 : Classify the data mart

Step12 : End

Assume the data mart be sales. The main attributes of sales are normally be sales details i.e., product name, product ID, Product description, Price and so on. Identify the relevant attributes of sales i.e., product descrption, ID and price. Based on choosing the relevant attributes, the functional behavior is analyzed and stored in data repository for further analysis of requirements. The functional behavior of the data mart is efficiently analyzed and the relevant attributes are also taken from the data warehouse. Through FBP, the analysis is done throught he selection of attributes which are closely related with the data description.

4. EXPERIMENTAL EVALUATION

The proposed functional behavior pattern is efficiently analyzed for data mart based on its attribute relativity. The experiments were run on an Intel P-IV machine with 2 GB memory and 3 GHz dual processor CPU. We present an experimental study to estimate the effectiveness and performance of the proposed functional behavior pattern for data mart based on attribute relations. The effectiveness of the proposed functional behavior pattern for data mart is estimated on benchmark datasets from UCI repository with varying characteristics.

To the best of our knowledge, there is none other technique which described the functional behavioral pattern of data mart based on relevant attributes. The attributes are chosen which are closely related with the organization. Based on those attributes, the operational goal of the organization is ef

analyzed and the functional behavior is identified. The performance of the proposed functional behavior pattern for data mart based on attribute relativity is measured in terms of

i) Number of functional data attributes,
ii) Attribute relativity,
iii) Analysis of functional behavior

Functional data attributes are attributes which can be allotted a function as a substitute of a value, and the function will be identified any time a value for that attribute is desired. Attribute relativity specifies the rate of attributes present in the data mart are closely related with the operational goal of organization/data mart. Based on attribute relativity, the functional behavior of data mart is easily analyzed.

5. RESULTS AND DISCUSSION

When compared to an existing multi-functional data warehousing model, the proposed functional behavior pattern for data mart based on attribute relativity is effective in terms of number of functional data attributes, attribute relativity and analysis of functional behavior of the data mart by maintaining the attributes in the data mart in a successful manner. Since we are using a relevant attributed in data mart, the functional behavioral pattern of the data mart/ organization is efficiently analyzed and viewed. The experiments are conducted with benchmark data sets to estimate the performance. The below table and graph describes the performance of the proposed function behavior pattern for data mart based on attribute relativity with an existing multi-functional data warehousing model.

No. of Data Marts	Number of functional data attributes	
	Proposed FBP for data mart	Existing multi-functional model in DW
1	5	2
2	9	5
3	12	8
4	13	10
5	15	11

Table 5.1 no. of data marts vs. no. of functional data attributes

Table 5.1 described the identification of functional data attributes present in the data mart for analyzing the functional behavior. The outcome of the proposed FBP for data mart is compared with an existing multi-functional model in DW.

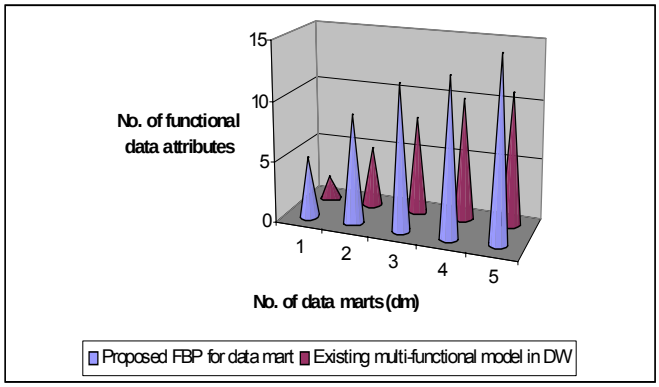

Fig 5.1 no. of data marts vs. no. of functional data attributes

Fig 5.1 describes the process of identifying the functional data attributes present in the data mart. The proposed Functional Behavior Pattern which has been used here had chosen the best relevant attributes and the functional data attributes are chosen for further experimentation. Existing multi-functional data model only identified the process of data mart but not the operational goal of the data mart. In the proposed FBP for data mart efficiently identified the operational goal of the data mart by choosing the functional data attributes. Based on the number of data marts, the functional data attributes present in the data mart is high in the proposed FBP for data mart contrast to an existing multi-functional model in DW.

No. of (a) Attributes	Attribute Relativity (%)	
	Proposed FBP for data mart	Existing multi-functional model in DW
5	10	5
10	13	12
15	18	10
20	20	15
25	23	18

Table 5.2 no. of attributes vs. Attribute relativity (%)

Table 5.2 described the attribute relativity of functional data attributes present in the data mart for analyzing the functional behavior. The outcome of the proposed FBP for data mart is compared with an existing multi-functional model in DW using attribute relativity.

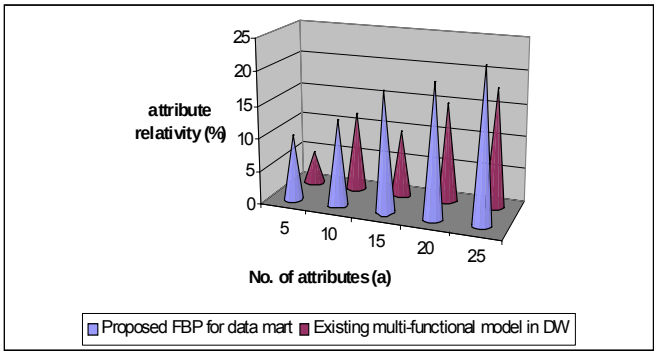

Fig 5.2 no. of attributes vs. Attribute relati

Fig 5.2 describes the attribute relativity of the functional data attributes present in the data mart. The proposed Functional Behavior Pattern chosen the relevant attributes based on the operational goal of the data mart and the functional data attributes are analyzed efficiently based on relevant attributes of the data repositories. Existing multi-functional data model only identified the process of data mart but the proposed FBP for data mart efficiently identified the operational goal of the data mart by choosing the functional data attributes and the attribute relativity is based on the number of attributes which are closely related with the operational goal of the data mart. Based on the number of attributes present in the data marts, the attribute relativity present in the data mart is high in the proposed FBP for data mart contrast to an existing multi-functional model in DW.

No. of Data Marts	Analysis of functional behavior	
	Proposed FBP for data mart	Existing multi-functional model in DW
1	10	5
2	15.3	8.4
3	19.2	12.5
4	21.5	15.7
5	24	18

Table 5.3 no. of data marts vs. Analysis of functional behavior

Table 5.3 described the functional behavioral analysis of the data mart and the outcome of the proposed FBP for data mart is compared with an existing multi-functional model in DW using attribute relativity.

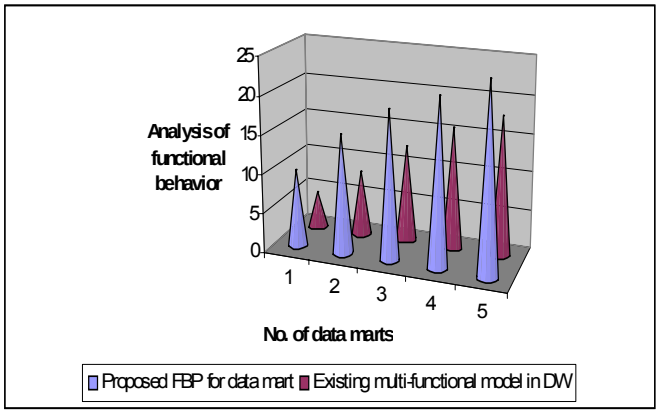

Fig 5.3 no. of data marts vs. Analysis of functional behavior

Fig 5.3 describes the efficiency of analysis of functional behavior of the data mart. The proposed Functional Behavior Pattern chosen the relevant attributes based on the operational goal of the data mart and the fig 5.2 showed the attribute relativity of the proposed FBP for data mart is high. Since the relativity of attribute is high, the efficiency of the proposed FBP for data mart is also being high. Existing multi-functional data model does not analyze the functional behavior of data mart but the proposed FBP does. Based on the number of data marts, the analysis of functional behavior in the data mart is high in the proposed FBP for data mart contrast to an existing multi-functional model in DW.

Finally, it is being observed that the proposed Functional Behavior pattern technique efficiently analyzed the functional behavior of the data mart based on the selection of relevant attributes of the data mart by building layers of data storage repositories.

6. CONCLUSION

In existing multi-functional models in DW, the analysis of functional behavior does not analyzed only the process of data mart is identified and processed. The issues raised over existing multi-functional models are efficiently handled by the proposed Functional Behavior Pattern for Data mart based on attribute relativity. The proposed Functional behavior, at first, the relevant attributes of the particular data mart is chosen which specifies the detailed description of the operational goal of the data mart. The proposed Functional Behavior pattern for data mart achieved the analysis of functional behavioral of the data mart based on attribute relativity. An efficiently developed Functional Behavior Pattern for data mart using attribute relativity performed the functional analysis and the experimental results showed that the proposed functional behavioral pattern for data mart analyzed efficiently by using the attribute relativity, and efficient in terms of analysis of functional behavior compared to an existing multi-functional models in DW.

REFERENCES

[1] Polyzotis, N.; Skiadopoulos, S et. Al., 'Meshing Streaming Updates with Persistent Data in an Active Data Warehouse', IEEE Transactions on Knowledge and Data Engineering, 2008

[2] Perez, J.M. Berlanga, R. Aramburu, et. Al., "Integrating Data Warehouses with Web Data: A Survey", IEEE Transactions on Knowledge and Data Engineering, 2008

[3] Y. Tao, M. Yiu, D. Papadias, M. Hadjieleftheriou, and N. Mamoulis, "RPJ: Producing fast join results on streams through rate-based optimization."in Proc. of SIGMOD, 2005.

[4] A. Polyzotis, S. Skiadopoulos, P. Vassiliadis, A. Simitsis, and N.-E. Frantzell, "Supporting Streaming Updates in an Active Data Warehouse,"University of California Santa Cruz," Technical Report, 2006.

[5] Li Jian , Wang Bing et. Al., „Multidimensional Data Model Research for Data Mart Based Analysis System of Drilling Fluid“, 2010 Second WRI Global Congress on Intelligent Systems (GCIS).

[6] Sassi, R.J. et. Al., 'Zero Latency applied on a commercial Data Mart: Real-time information in support of decision making', 2011 IEEE 3rd International Conference on Communication Software and Networks (ICCSN)

[7] Desmarais, M.C. , " Web log session analyzer: integrating parsing and logic programming into a data mart architecture", :

The 2005 IEEE/WIC/ACM International Conference on Web Intelligence, 2005. Proceedings.

[8] Hajmoosaei, A. et. Al., "Comparison plan for data warehouse system architectures", 2011 3rd International Conference on Data Mining and Intelligent Information Technology Applications (ICMiA).

[9] Cabibbo, L. et. Al., "DaWaII: a Tool for the Integration of Autonomous Data Marts", Proceedings of the 22nd International Conference on Data Engineering, 2006. ICDE '06.

[10] B.-K. Park, H. Han, and I.-Y. Song. XML-OLAP: A Multidimensional Analysis Framework for XMLWarehouses. In Proceedings of the 6th International Conference on Data Warehousing and Knowledge Discovery, pages 32–42. Springer, Berlin, 2005.

[11] Sen, A. et. Al., "A Model of Data Warehousing Process Maturity", IEEE Transactions on Software Engineering, 2012

[12] Destercke, S. et. Al., "Evaluating Data Reliability: An Evidential Answer with Application to a Web-Enabled Data Warehouse", IEEE Transactions on Knowledge and Data Engineering, 2011

ABOUT THE AUTHOR

PAULRAJ M received a Master Degree from Madurai Kamaraj University, India, Post Graduate Diploma in Business Administration from Annamalai University, India and Master of Computer Application (M.C.A) degree from Indira Gandhi National Open University, New Delhi, India, in 2002. He joined the Computer Centre , Department of Statistics, Ministry of Planning, New Delhi, and He joined the National Crime Records Bureau, Ministry of Home Affairs, New Delhi, where he was a Junior Staff Officer(System Analyst). He is currently Senior Systems Manager of Department of Information Technology at Canara Bank, Head Office, Bangalore, India, his job interests include introducing new technology and projects related to financial systems & analysis. He currently preparing a Ph.D in Computer Science, Research Scholar, thesis titled "Performance Analysis for Data Mart" in a reputed University at Tamil Nadu, India.

CO-AUTHOR

Dr. SIVAPRAKASAM P, M.Sc., M.Phil., Ph.D is associate professor, Department of Computer Science, at Sri Vasavi College, Erode – 638 316, Tamil Nadu, India. He obtained Ph.D. from Periyar University, Salem, Tamil Nadu, India in Analysis of Web Caching Architecture.

***