International Journal on Recent and Innovation Trends in Computing and Communication ISSN 2321 – 8169

Volume: 1 Issue: 5 451 – 452

HANDLING DATA, METADATA AND DESIGN FOR GLOBAL DATA WAREHOUSE

Mrs. Neeraj Sharma(Assistant Professor)

Department of Computer Science

Ideal institute of management &Technology
Ghaziabad (UP), India
neeraj_khn@yahoo.com

Abstract — Data ware house modeling is a complicated task, which involves knowledge of business process, as well as familiarity with operational information system structure and behavior. While it is universally recognized that a DW leans on multi dimensional model, there is suggested to contract a data ware house conceptual model, My analysis of these methods indicate that they are limited in their applicability to model large scale systems, as they require acquaintance with the business process and ability to select relevant transactional eliminate the technical and reporting difficulties experienced with the classic star schemes and snow flake. The objective of Meta data strategy is to provide a blueprint that will enable efficient stronger, maintenance, and delivery of timely and accurate business and technical Meta data of the organization in building a robust data warehouse environment ,meta data is considered to be a critical success factor .This work will address meta data requirement and a proposed solution for effective metadata management and deployment /delivery at organization .The purpose of data quality methodology is to validate the quality of specific business application including the source of these business application and the process that feed the business application of the EDW environment application are worthless .Business application vendors have great products to sell ,offering simplified access to complex computation models and enormous stores of corporate data, yet they often deliver below expectations. This work will help to overcome this limitation.

***Keywords*— Metadata, Data Warehousing, Data Quality methodology, Refined Schema Solution for Aforementioned Problems**

*****

I. INTRODUCTION

Data warehouse is a repository of an organization's electronically stored data.. The definition of the data warehouse focuses on data storage. However, the means to retrieve and analyze data, to extract, transform and load data, and to manage the data dictionary also considered essential components of a data warehousing system. Organization have vast amount of data but have found increasingly difficult to access it and make use of it.

II. Problem formulation

It is well known that the ultimate goal of organization, government and companies, when they accumulate and store information is the ability to process it later and take advantages of it , unfortunate ,neither the accumulation nor the storage process ,seem to be completely credible. Errors in database have been reported to be up to ten % range and even higher in a variety of application. Some experts have said that the typical data warehouse project will require companies to spend 80% of their time doing this. While the percentage may or may not be as high as 80%, one thing that you must realize is most vendors will understate the amount of time you will have to spend doing it.

1. While it is universally recognized that a DW leans on a multidimensional model, her is no agreement on the approach to conceptual modeling.
2. Data loading is very complex in case of snowflake schema
3. Star schema is not multi-lingual
4. Dimension tables are un –normalized in star schema i.e. it does not support shared dimension
5. Too many join slows down the performance in case snowflake schema
6. Star schema cant represent the semantics and operation of multi dimensional data adequately
7. Metadata management plays a crucial role in the successful implementation of data warehouse .there is no standard metadata management strategy which ensures the smooth of a data warehouse

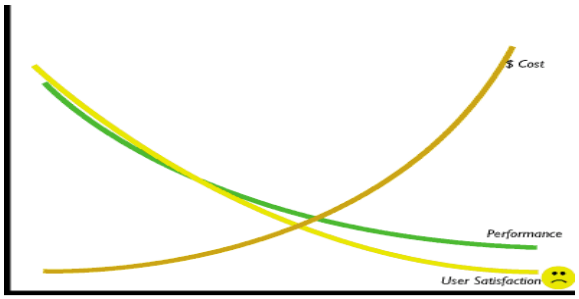

Figure: Effect on the Growth of Data Warehouse

III. Refined Schema (Solution for aforementioned problems)

This new refined schema will eliminate both technical and business reporting problems experienced with star schema. This model consist of a quantity of relational tables arranged multi-dimensionally and use metadata management strategies
.

Metadata Management Strategies

The objective of metadata strategy is to provide a blueprint that will enable efficient storage, maintains and delivery of timely and accurate business and technical metadata of the organization. Metadata helps the analyst find what data is in the warehouse and use that data effectively and efficiently. Metadata is consider to be critical success factor for both technical and business users need to understand what is the warehouse

Multidimensional Modeling

The multidimensional data model is an integral part of On-line Analytical processing, or OLAP because OLAP is on-line, it must provide answers quickly, analyst pose iterative queries during interactive sessions, not in batch jobs that run overnight. And because OLAP is also analytic, the queries are complex. The multidimensional data model is composed of logical cubes, measure, dimension, hierarchies' levels and attributes. the simplicity of the model is inherent because it define object that represent real world business entities ,analysts know which business measure they are interested in examine , which dimension and attributes make the data meaningful, and how the dimension of the business are organized into level and hierarchies

This refined multidimensional model consists of the following tables:
The center of a star forms the fact table containing the key figure.
The fact table is surrounded by several dimensions
A dimension consists of different table types:

Dimension Table: The attributes of the dimension tables are called characteristics

Master Tables: Master data tables: dependent attributes of a characteristic can be stored in a separate table called the master data tables for that characteristic

Text Tables: Textual description of a character tics is stored in a separate text tables.

Hierarchy Tables: Hierarchies of character tics or attributes may be stored in separate hierarchy tables .Hierarchies are used in analysis to describe alternative views of t he data.

Connecting Master Tables to form Star

In hybrid of star and snow schema the dimension tables don't contain master data information .The master data information is stored in separate tables, called master data tables which are being linked to fact table using shared ID's .Master tables carrying texts like cost center description or name are also being stored separately which will give us benefit of storing description in multiple language within same model by just maintaining language key as mandatory field in this table.

Hence following will the advantage of refined schema:-

- Multingual capability
- Cross –cube use of mater data
- Fewer joins

IV. Result and Discussion

A new refined schema eliminating the problems faced with other schemas. In this case of star schema there are redundant entries in the dimension tables and modeling hierarchy types in a dimension leads to anomalies .It was very cumbersome to get data in different language as that resulted in redundancy of data .Also ,since basic fact data and aggregates are stored in the same fact tables ,query performance was very poor. Keeping these problem in the mind this thesis propose a new refined schema in which quantity of relational tables are arranged multi-dimensionally , meaning that it consist of a central fact table surrounded by several dimension tables and future shared ID tables are used to link these dimension tables to their respective master data tables such as attributes tables., text table and hierarchy table which contain master data for the corresponding dimension .Now as all different kinds of information are stored separately in its receptive master

References

[1] Horowitz, A. "Ensuring the Integrity of Your Data", Beyond Computing.
[2] Kimball, R. , "Dealing with Dirty Data" DBMS Online. Strange, K. "A Taxonomy of Data Quality", Gartner Group Watterson, K, "Dirty Data, Dumb Decisions", DM Review Magazine. Williams, J. "Tools for Traveling Data" DBMS Online.
[3] R .Aggrawal , A Gupta , S. Sarawagi "Modeling multidimensional database "
[4] Chaudhuri , U .Dayal "An Overview of Data Warehousing And OLAP Technology"
[5] Raden ,N "Modeling a Data Warehouse"
[6] Forrestor Quarterly Data Warehousing Survey, Jan 2010
[7] Kimball ,R. :The data warehouse toolkits ,New York