



ISSN: 2319-5967

ISO 9001:2008 Certified

International Journal of Engineering Science and Innovative Technology (IJESIT)

Volume 2, Issue 1, January 2013

Information Integration: An Enterprise Solution

Er. Majid Zaman,
Scientist,
Directorate of IT & SS,
University of Kashmir, Srinagar

Er. Muheet Ahmed Butt,
Scientist,
PG Department of Computer Science,
University of Kashmir

Abstract- Data storage is arranged in a totally decentralized way and information retrieval might involve querying many data sources. In large organization for example, where decisions are commonly made based on information measurement, each division might support its own database scheme (HRM, Finance, Sales etc). Therefore accrued data for the whole organization requires data accumulation from various sources [4]. The decentralization is further increased if we also take into account affirmable partners, vendors or contender, whose information might be of company's interest. Another area, where this decentralization aggravates content retrieval is sizable scale scientific projects. Here not only the amount of data, but also the quality has to be taken into record. Scientists nowadays, besides fundamental domain cognition, require entree to data and results provided by others. Therefore, querying one by one different data sources leads to important unskillful ness in their work. Finally, for an effective search in Enterprise, user is needed to look up for content in multiple information sources and collect the data individually. Much engagement is set on the so called "semantic heterogeneity", which seems every time there is a more than one way to structure a body of information. Semantic heterogeneity seems to be an ineluctable load in data interdependence and use, since people tend to exemplify their data according to their own perceptive of the reality. This of course is fundamentally different for each individual. In that sense, heterogeneity is to be recovered in data models, conceptual schemas, and of course the mind of the users. In this paper data integration from these heterogeneous data sources using data warehouse is dealt upon.

I. INTRODUCTION

Enterprise globally started automation in late 1970's and early 1980s, however due non standardization of development tools and non-availability of RAD tools, department within the enterprises was automated but nothing was thought about future and data integration [1]. To just meet with the dynamic changing scenario of Information Technology an in house computerization unit was established which mostly focused on designing and developing software solutions for organizations. It mostly involved development of those software products which were used to meet day to day requirements of various departments of the enterprise like HRM, Finance, and Sales etc. in an efficient manner. At first, development activity was carried out in FoxBASE, dbase & FORTRAN as programming platforms, Unix and Windows 9x as operating systems, teams were set up initially to computerize the Enterprise related processes.

As year passed on, content explosion within the enterprise was at its prime, it was not much long before establishment administration realized that in order to manage information in an expeditious manner and render cordial accession to content, more new teams needed to be setup with enhanced budget in order to store/manage data and information in an efficient way to meet the changing demands. Newly established unit were granted nice budget along with state to choose the implement needed for development, within no time enterprises were riding on with the success of Information Technology and almost all the areas of the enterprise were computerized and independent solutions within single organization each working fine but with no possibility of integration.

As enterprises, both government and public, have to manage big magnitude of vital knowledge stored in some form of databases or files [3]. One of the main difficulty to deal with content managing is the weak interoperability between various databases and information systems. Especially this difficulty is serious when we want create collaboration between the content systems of various departments within the organization, therefore the Development which was carried out in the enterprise lagged integration of data from various heterogeneous data sources which were present at various operational levels of the organizations.

II. DATA MART & DATA WAREHOUSE

A data mart is a simplistic kind of a data warehouse that is convergent on a single subject (or functional area), such as HRM, Finance, Sales etc. Data Marts are often built and possessed by a single department within an



ISSN: 2319-5967

ISO 9001:2008 Certified

International Journal of Engineering Science and Innovative Technology (IJESIT)

Volume 2, Issue 1, January 2013

enterprise. Acknowledged their single-subject focus, data marts normally draw data from only a few sources. The sources could be internal functional systems, a fundamental data warehouse, or external data [5][7]. A data warehouse, unlike a data mart, trades with aggregates subject areas and is typically enforced and controlled by a central organizational unit such as the corporate Information Technology (IT) group. Often, it is called a central or enterprise data warehouse. Typically, a data warehouse assembles data from multiple source systems.

Nothing in these basic definitions limits the size of a data mart or the complexity of the decision-support data that it contains. Nevertheless, data marts are typically smaller and less complex than data warehouses; hence, they are typically easier to build and maintain [oracle].

Each Data Mart can incorporate various combinations of tables, columns and rows from the Enterprise Data Warehouse [8]. For example, a concern unit or user group that doesn't necessitates a good deal of historical data might only need transactions from the actual calendar year in the database. The Personnel Department might need to see all details about employees, whereas data such as "salary" or "home address" might not be appropriate for a Data Mart that focuses on Sales.

III. DATA MART DESIGN

Let us take an example of a Accounting System and HRM System. Existing Account Data Mart has Employee ID Dimension, Pay Dimension, GP FUND Dimension, Allowances Dimension, Loan Dimension, Insurance Dimension, Appraisal Dimension, Advances Dimension etc. Some of the dimensions mentioned here are common in most of the enterprise [4]. Like wise there is Employee Data Mart were in there is Employee ID Dimension which in itself is replication and is also present as a part of Account Data mart, besides their are other common dimensions like Employee Name Dimension, Employee Address Dimension again these are some of the most common dimension pertaining to employees.

Table 1: Data Marts for Account and Employee

Account Data Mart	Employee Data Mart
Employee ID	Employee ID	...
Pay	Employee Name	...
GP Fund	Employee Address	...
Allowances	Job Grade	...
Loan	Salary	...
Insurance	Education	...
Appraisal	Appraisal	...
Advances	DOB	...
...
...

On the same line other marts are created, warehouse administrator has capability to pick and choose data from various data sources. These will be followed by performing ETL on these newly created Data Marts. The Account and Employee Data Marts is shown in the figure below

IV. DATA WAREHOUSE DESIGN

In our proposed solution the data integration into a single data warehouse with data marts is the prime focus. The solution thus proposed must make use of Linux-Operating System, Oracle Database Management System, Apache TomCat Web Server, Java/JSP server with above mentioned software configuration be set up. This server is part of enterprise network where in it has been given a controlled access to all existing solutions in enterprise e.g accounts, employees , HRM, Finance, Sales etc in other words this server is made to perform all the ETL functions on all the heterogeneous data sources and place extracted data into data marts. The basic idea of this solution is to create Data Mart for every autonomous information source and then integrate these data marts to have single warehouse which could be named as Account Mart, Employee Mart, Academic Mart, HRM Mart, Budget Data Mart, Finance Data Mart etc. Enterprise now a days are interconnected by a Local Area Network (LAN) in which each operational section/division has a separate working software Application and Data Source. While as the Account solution may have been developed on LAMP (Linux, Apache, My SQL, PHP), HRM system is developed on Microsoft technologies MSSQL,.Net framework etc, at the same time budget is still very much legacy system developed in dBase 4.0. The newly created marts are connected to create a single warehouse.

V. CONCLUSION

Enterprises across the world while centering on automation of division did not pay much emphasis upon



ISSN: 2319-5967

ISO 9001:2008 Certified

International Journal of Engineering Science and Innovative Technology (IJESIT)

Volume 2, Issue 1, January 2013

similarity and uniformity of data, and almost all the enterprises across the world ended up with many heterogeneous data sources. While data in warehouse must be credible, it must be carefully put together from a variety of sources around the enterprise. Data Warehouse not only makes enterprise content easily accessible but has become tool for data integration.

REFERENCES

- [1]. Jorge Bernardino, Pedro Furtado, Henrique Madeira, "A Cost Effective Approach for Very Large Data Warehouses", Proceedings of the International Database Engineering and Applications Symposium, 2002.
- [2]. Kimball, Ralph, "The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses", John Wiley & Sons, Inc, 1996.
- [3]. Date, C. J. (1995), an Introduction to Database Systems, Addison-Wesley Publishing Company, Inc.
- [4]. E.Bertino, P.Bonatti, E.Ferrari "TRBAC: A Temporal Role Based Access Control", Information and System Security, 2001.
- [5]. J. Gray and A. Reuter. Transaction Processing – Concepts and Techniques. Morgan Kaufmann Publishers, 1993.
- [6]. http://docs.oracle.com/html/E10312_01/dm_concepts.htm.
- [7]. W. J. Labio, Y. Zhuge, J. N. Wiener, H. Gupta, H. Garcia-Molina, J. Widom. Stanford University. "The WHIPS Prototype for Data Warehouse Creation and Maintenance". SIGMOD 1997.
- [8]. Marotta. "Transformations based approach for designing Data Warehouses Internal Report." InCo. Universidad de la República, Montevideo, Uruguay. 1999.