

Warehouse Creator: A Generic Enterprise Solution

¹, Majid Zaman, ²Muheet Ahmed Butt

¹, Scientist, Directorate of IT & SS, University of Kashmir, J&K, India

², Scientist, PG Department of Computer Science, University of Kashmir, Srinagar, J&K, India

Abstract : With the advent of 21st century, enterprises developed warehousing solutions and the early solutions were such that multiple warehouses were created within a single enterprise. Though data warehousing has become a benchmark of data integration yet no generic tool cut across type, complexity, enterprise has been proposed as of yet. In this paper, we propose a tool which can create warehouse irrespective of type, complexity, enterprise work etc. Proposed tool creates dimension tables and fact table by evaluation the data present in various data sets (data sources). User/Administration intervention required is minimal; however access required for data sources is essential.

Keywords: OLAP, Data Warehouse, Enterprise, Data Sources

I. INTRODUCTION

1980s-1990s was era of automation of activities and to ease up the industry/office work. In this era the primary goal was the digitization of data and much was not thought about for the integration and centralization of data and resources. With the advent of 21st century industry realized that enterprises had multiple heterogeneous [3][1][2] data sources and allied infrastructures. In 1996 Ralph Kimball published the book "Datawarehouse Toolkit" but it was not till 2000 onwards people realized the essence of warehousing and emergence of its implementation. The motivation behind this project is everyday data crisis. Every now and then enterprises across the globe are spending millions of rupees to manage and integrate the data, however yet there is no single industry/enterprise data integration tool/regulation and enterprises across the globe are still in the process of standardizing data integration making use of warehouse as tool kit.

The Figure 1 illustrates the existing Enterprise solution, where in n heterogeneous data sources are interconnected via internet/intranet as the case may be.

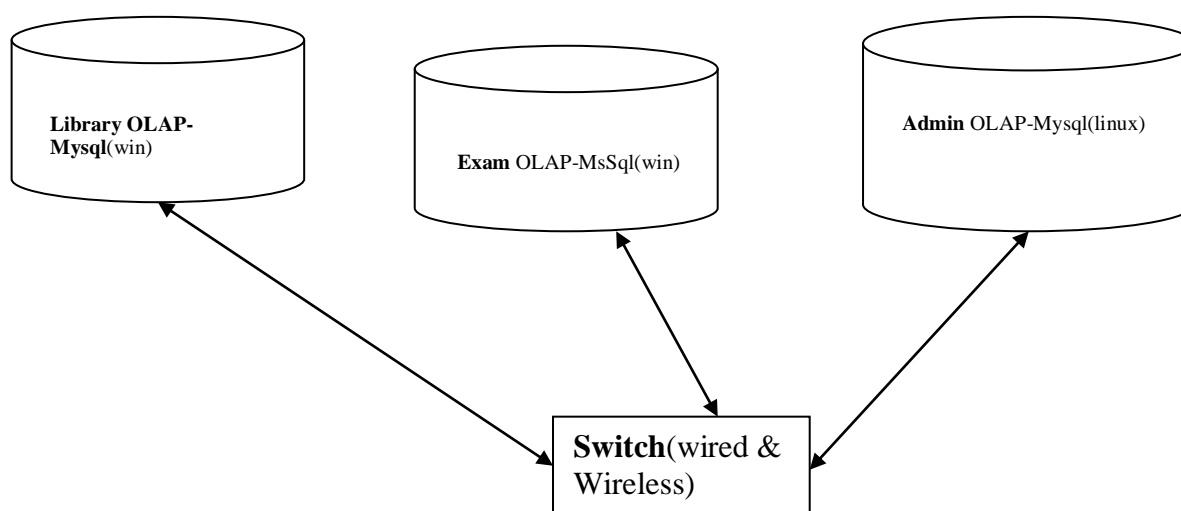


Figure 1: Warehouse Enterprise Solution

Enterprise users have to access each data source in order to retrieve information of his/her choice; problems encountered by enterprise user are listed below

- [1] User interested in information retrieval is required to know query language in order to access information from Data Source.
- [2] User should be well aware of database architecture
- [3] User should be well versed with information source, as in where information of his/her interest is saved.
- [4] Access rights of enterprise users are area of concern.
- [5] Geographical distance of Database servers can become reason of worry, besides other issues.

In order to overcome all the above mentioned problem besides other problems Data Warehousing is most feasible solution, besides enterprise users can be give web bases access to retrieve information of his/her interest without having right to update any piece of data/information from any data source.

II. PROBLEM STATEMENT

In order to develop Data Warehousing tool which can integrate data and create fully functional warehouse following requirements have to be fulfilled

- [1] The tool should have the ability to create warehouse from existing OLAP.
- [2] This solution can be installed in any enterprise and warehouse can be created.
- [3] Data is extracted from Heterogeneous sources.
- [4] Network is both wired and wireless
- [5] New Data sources can be added at any time and incorporated into warehouse.
- [6] Web based interface has to be provided with warehouse in order to give user information retrieval access based on the key words.
- [7] Warehouse creation should not manipulate table data on the remote server.

III. DATA SETS

Data sets are composed of data from databases and legacy sources which can be any other source. Data warehousing, and data marts are designed to assist individuals and organizations in managing and extracting meaning from enormous amounts of data. Data mining is used to analyze these data sets and predict future trends and forecast. Data warehouses and data marts are used to store and analyse historical data present in the warehouse in order to make better choices and estimates about the future. The purpose of many of these activities and approaches is to relate data sets to each other, group related data together [5], and ensure the ability of users to access the data they need.

3.1 Data In Databases

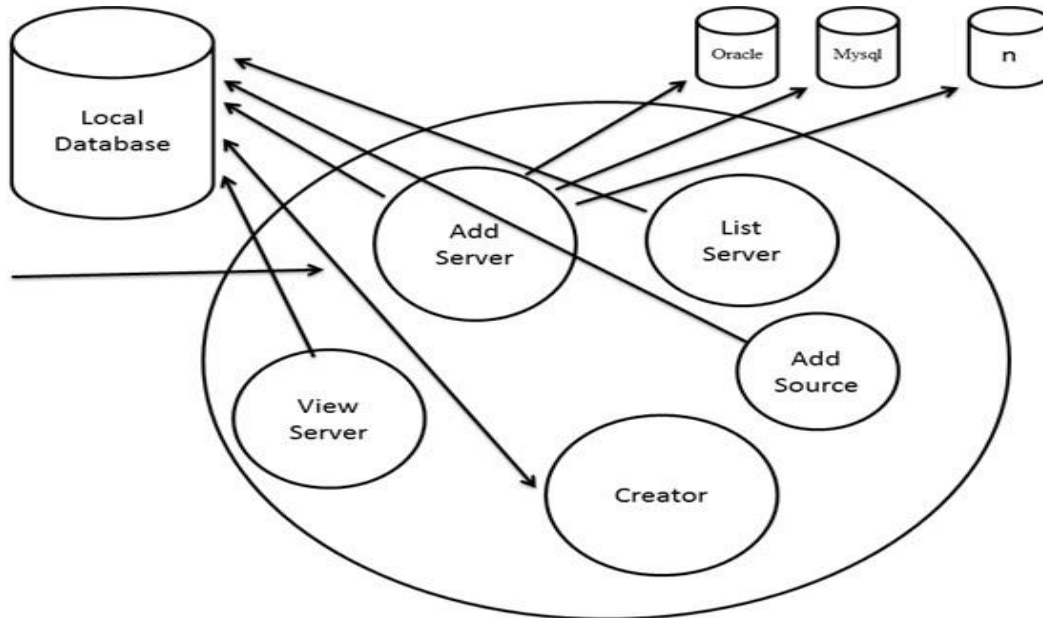
Data from different databases is extracted and placed into the warehouse, data resides in [3][4] heterogeneous databases like MySQL, MSSQL etc.

3.2 Legacy Sources

Legacy sources like text files also form a source of data. Data from these sources could be extracted and placed into the warehouse.

IV. DESIGN

The detailed design is diagrammatically described below. The DFD shown below enlists n modules required for creation of the warehouse. The local data source is where data warehouse will be created from n heterogeneous database e.g Oracle, MySql, MsSql etc. The local data source can be Oracle, MySql, MsSql, Db2 etc depending upon system requirements. The modules listed below are required for enlisting of various data sources and subsequently retrieving data from the said sources. The data is only read from data sources and is not allowed to modify the contents of OLAPs/ transnational servers which continue to work for the designated purpose.



V. IMPLEMENTATION

The solution has two distant parts:

5.1 Creation of Warehouse:

- [1] Identification and addition of servers based on
- [2] ip address
- [3] username
- [4] Password
- [5] view list of database servers added
- [6] Remove list of database servers added.
- [7] Add source where source are the tables which will be required for the creation of warehouse.
- [8] Create Warehouse: Associate name with the data warehouse to be created and select a primary table.
Primary table is the table from where the creation starts.

In the back end data warehouse is created along with the fact table and dimension tables. Dimension tables are populated by extracting and cleaning information from tables added at step(IV). Once the data is fetched into dimension table, data as per structure of fact table is fetched and stored in fact table.

5.2 Use of Warehouse:

Once warehouse is created for enterprise, it is made functional.

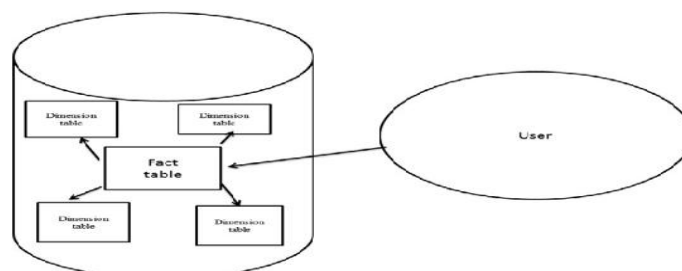


Figure:

- [1] GUI for users accepts search string.
- [2] Search is made on fact table.
- [3] Primary information extracted from fact table is displayed.
- [4] Depending upon the user requirement, user can view information of his/her choice.

VI. CONCLUSION

Data Warehouse creation methodologies are available to enhance the quality of the data at several stages in the process of developing a data warehouse. Data warehouse creation tools can be useful in automating many of the activities that are involved in cleansing the data- parsing, standardizing, correction, matching, transformation ect. Many of the tools specialize in auditing the data, detecting patterns in the data, and comparing the data to business rules[4]. Once these problems have been isolated, the warehouse builder could determine which features of the data quality tools address the specific needs of the data sources to be used. The association of the data quality tool at every step of the data cleansing process plays a very important role in ensuring that sanity of the data is protected at every manner.

REFERENCES

- [1]. EMA Butt, SMK Quadri, EM Zaman, "Star Schema Implementation for Automation of Examination Records", WORLDCOMP, 2012 FEC3568, Las Vegas USA.
- [2]. MA Butt, SMK Quadri, M Zaman," Data Warehouse Implementation of Examination Databases", International Journal of Computer Applications 44 (5), 18-23.
- [3]. M Zaman, MA Butt, DSMK Quadri, "User Desired Information Translation" Journal of Global Research in Computer Science 3 (6), 51-53
- [4]. Neill, P. (1998). "It's a Dirty Job: Cleaning Data in the Warehouse", Gartner Group, January 12, 1998
- [5]. EMA Butt, SMK Quadri, EM Zaman, "Star Schema Implementation for Automation of Examination Records", WORLDCOMP, 2012 FEC3568, Las Vegas USA
- [6]. Zaman, Majid, S. MK Quadri, and Muheet Ahmed Butt. "Generic Search Optimization for Heterogeneous Data Sources." International Journal of Computer Applications 44.5 (2012): 14-17.
- [7]. EM Zaman, EMA Butt. "Information Integration: An Enterprise Solution" International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 2, Issue 1, January 2013: 315-317