



# INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

## REAL TIME DATA WAREHOUSING : AN OVERVIEW

MS. DEEPASHRI S. KHAWASE<sup>1</sup>, MR. H .R. DESHMUKH<sup>2</sup>, S. H. KUCHE<sup>3</sup>, A. S. MAHALLE<sup>3</sup>

1. M.E I Year, IBSS COE, Amravati.
2. Prof and Head, CSE, IBSS COE Amravati.
3. Asst. Prof., CSE, IBSS COE Amravati.

Accepted Date: 27/02/2014 ; Published Date: 01/05/2014

**Abstract:** Today organizations are in an insatiable need for information in this competitive era. Knowing what your customers want is the key to maintaining a competitive advantage. Organizations poised to experience the greatest success will be the firms that can leverage their data to meet organizational needs, build healthy relationships with stakeholders and meet the demands of customers. An effective and efficient real-time business intelligence infrastructure that leverages the power of a data warehouse is able to deliver value by helping organizations enhance their customer experiences. A real-time data warehouse can eliminate the data availability gap and can enable organizations to concentrate on processing their valuable customer data. This paper provides an overview on informative approach to evaluate real-time replenishment software for feeding a data warehouse outlining real-time data transformation and integration requirements and highlighting how a business succeed quickly and exceed the ever-changing needs and expectations of the customers.

**Keywords:** Real Time Data Warehousing, Business intelligence, CTF (Capture, transform, flow) tool



PAPER-QR CODE

Corresponding Author: MS. DEEPASHRI S. KHAWASE

Access Online On:

[www.ijpret.com](http://www.ijpret.com)

How to Cite This Article:

Deepashri Khawase, IJPRET, 2014; Volume 2 (9): 673-678

## INTRODUCTION

Data warehouse is subject Oriented, Integrated, Time-Variant and non-volatile collection of data. This data helps in supporting decision making process by analyst in an organization. The Data Warehouses provide us generalized and consolidated data in multidimensional view. Along with generalize and consolidated view of data the Data Warehouses also provide us Online Analytical Processing (OLAP) tools. These tools help us in interactive and effective analysis of data in multidimensional space. This analysis results in data generalization and data mining. The data mining functions like association, clustering, classification, prediction can be integrated with OLAP operations to enhance interactive mining of knowledge at multiple level of abstraction. Hence data warehouse has now become important platform for data analysis and online analytical processing. Companies use data warehouses to store information for marketing, sales and manufacturing to help managers run the organization more effectively. The ability to manage and effectively present the volume of data tracked in today's business is the cornerstone of data warehousing. But when the data warehouse is replenished in real-time it empowers users by providing them with the most up-to-date information possible. Currently, the dominant method of replenishing data warehouses and data marts is to use extraction, transformation and load (ETL) tools that "pull" data from source systems periodically – at the end of a day, week, or month – and provide a "snapshot" of your business data at a given moment in time. That batch data is then loaded into a data warehouse table. During each cycle, the warehouse table is completely refreshed and the process is repeated no matter whether the data has changed or not. Businesses need tools that can provide real-time business intelligence and an absolutely current and comprehensive picture of their organization and their customers – not last week or last month, but right now.

## II. Operation of Real-Time Data Warehousing:

An intelligent warehousing solution and framework can be divided into three fundamental tiers with data flows between them. The three layers are Presentation Layer, Architecture Layer, and Middleware Layer. The three layers have their particular significance .

*Presentation Layer:* The presentation layer manages the flow of information from the warehouse to the analyst, providing an interface that makes it easier for the analyst to view and work with the data. This layer is where graphical user interface (GUI) tools are most important.

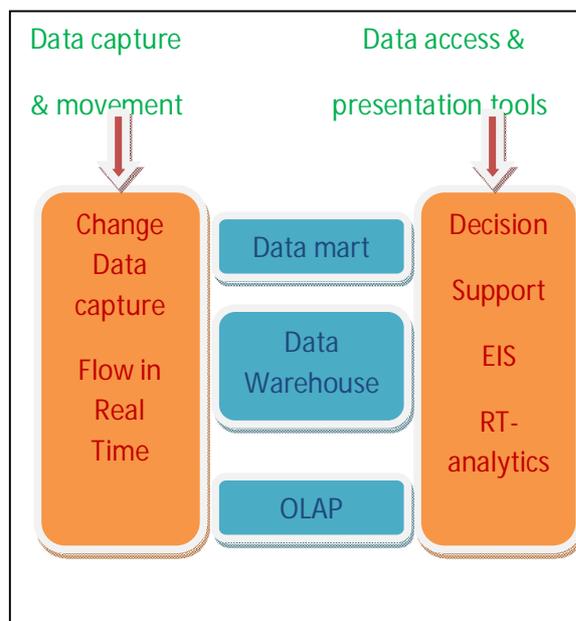
*Architecture Layer:* The architecture layer describes the structure of the data in the warehouse. An important component of the architecture layer is flexibility. The level of flexibility is measured in terms ease of analyst to break out of the standard representation of information offered by the warehouse in order for custom analysis.

*The middleware layer:* This layer is the glue that holds the data warehouse together. It integrates the data ware house with production and operational systems. Real-time integration and replenishment tools that help businesses deal with the data management issues of implementing a data warehouse can be valuable.

### III. Capture, Transform and Flow (CTF)

#### Change Data Capture

Today, more and more businesses using a data warehouse realize that they cannot achieve point-in-time consistency without continuous, real-time change data capture. Figure 1 below illustrates Typical data warehouse implementation utilizing (CTF) technology for real-time replenishment.



**Figure 1:** Typical data warehouse implementation utilizing (CTF) technology for real-time replenishment.

There are several techniques used by data integration / replenishment software to move data. Integration tools either push or pull data on an event driven or polling basis. Push integration is initiated at the source for each subscribed target. This means that as changes occur, they are captured and sent, or “pushed” across to each target. Pull integration is initiated at the target by each subscribed target. That is, the target system extracts the captured changes and “pulls” them down to the local database. Push integration is more efficient as it can better

manage system resources. As the number of targets increases, pull integration becomes resource draining on the source system, especially if that machine is a production machine that may already be overworked. With advanced CTF solutions, every time an add, change or delete occurs in the production environment, it is automatically captured and integrated or "pushed" in real-time to the data warehouse. Beyond real-time integration, change data capture can also be done periodically. Data can be captured and then stored until a predetermined integration time. Only data that has changed since the previous integration needs to be transformed and transported to the subscriber. The data warehouse can therefore be kept current and consistent with the source databases.

#### **IV. Criteria for Evaluation Of an Efficient Real-Time CTF Solution:**

In the ever-changing Internet era, companies should strive to select real-time CTF solutions that offer the following capabilities and features:

*A) Selectivity:* Business solutions like data marts and warehouses require the ability to select and filter which data is moved throughout the organization. Efficient CTF software offers an array of features including built-in data filtering and selection functions in addition to data enhancement and transformation capabilities.

#### *B) Heterogeneous Environment support*

In a change prone technology, corporate mergers, and acquisitions, most organizations bear multiple computing platforms and databases, each storing separate pockets of information that may be entirely incompatible to the next. Some integration tools, however, are capable of moving data among a wide range of databases and systems including Oracle, DB2/UDB, Sybase, and Microsoft SQL Server across OS/390, OS/400, UNIX, Linux and Microsoft Windows NT/2000. Many vendors currently offer transformational data integration tools to consolidate and synchronize heterogeneous data into a data mart or warehouse, however, many of these tools do not offer real-time capabilities.

#### *C) Ease of Administration:*

Some integration software can be awkward to install and set up. A crucial consideration during evaluation of transformational data integration tools is the work required for setup and implementation of the software. IT staff and end users seek an "out-of-the-box" experience from data warehousing tools. Organizations should ensure that no programming changes to existing applications and databases are required.

*D) Meta-data Management Capabilities:* Meta-data is information or say data about data. It allows business users as well as technical administrators to track the lineage of the data they

are using. Meta-data provides information about where the data came from, when it was delivered, what happened to it during transport, and other descriptions can all be tracked. There are two types of meta-data: technical, or administrative meta-data, and business meta-data. Both types of meta-data are needed to a successful data mart or warehouse solution.

#### *E) Scalability:*

Scalability is the ability of a computer system or a database to operate efficiently with larger quantities of data. While it is possible to predict some of the future information needs of an organization, a large portion will be unpredictable. In the Internet era, it is certain that as business environments change, so too will the kinds of decisions that need to be made and the information that influences them. Therefore, there is need of scalable solution when searching for an efficient data integration tool.

#### **V. Formulating a Real-time Data Warehouse:**

Some Organizations choose custom coding to handle the replenishment and transformation of data despite the buy-versus-make recommendations of most data warehousing analysts. Here, Businesses can write customized programs to integrate data to the data warehouse. Mapping-based tools involve built-in administration utilities that provide quick and easy definition of the integration process.

*Aggregation:* Aggregation is gathering of information in separate sets from two or more sources. Often, this data is stored in a data warehouse in a summarized form. For example, an organization may wish to summarize the data by various time periods. Aggregates are used for two primary reasons. One is to save storage space and the second reason is to improve the performance of business intelligence tools. When queries run faster, they take up less processing time and the users get their information back much more quickly.

#### **VI. CONCLUSION:**

While building a data warehouse, there one can get caught up in the technology forgetting the basics that contribute to its creation. The questions are about how fast does your company respond to new ideas, new competitive pressures and new opportunities. Possibility of branch of the organization stay informed about customers, competitors and business trends. Employees, decision-makers, partners and customers alike need access to information while it is still relevant. The Internet has significantly changed the scope and expectations for data warehousing implementations. Developing a robust, scalable data warehouse with consistent, real-time data and the ability to answer all user-information needs is not an easy goal to achieve. To keep data current is one of the most difficult challenges in managing data

warehouses and data marts. But if organizations fail to take advantage of real-time data capabilities for business intelligence, they will lose the opportunity to respond quickly to changing market trends. By understanding the relevant issues and staying informed of current practices in data warehouse development, organizations can harness the power of real-time business intelligence and attain a cost-efficient, profitable computing environment.

#### VII. REFERENCES:

1. Brobst, Stephen A. and Venkatesa, AVR, "What is Active Data Warehousing?," Abridged from "Active Warehousing," Teradata Review Spring 1999
2. Haisten, Michael, "Real-Time Data Warehouse: What is Real Time about Real-Time Data Warehousing?," DM Review Online August 2000
3. Schroeck, Michael J., "E-Analytics—The Next Generation of Data Warehousing," DM Review August 2000
4. Data Mining: Concepts and Techniques- Jiawei Han , Micheline Kamber.