International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Using Online Analytical Processing (OLAP) in Data Warehousing

Mohieldin Mahmoud[1], Alameen Eltoum[2], Ramadan Faraj[3]

[1, 2, 3]al Asmarya Islamic University

Abstract: ***Data warehousing is needed in recent time to give information that helps in the development of a company, support decision makers, it provides an efficient ways for transforming, manipulating analyzing large amount of data from different organizations or companies. In this paper we aimed to over view, data warehousing definition; how data are gathered from multi sources in to data warehousing using a model. Important techniques were shown; how to analyze data using data warehousing, to make decisions, this technique called: Online Analytical Processing OLAP.***

Keywords: Data Warehousing, OLAP, gathering, data cleansing, OLAP FASMI Test.

1. Introduction

The concept of data warehousing returns back to the early 1980s, where was considered relational database management systems. A data warehouse is a destination (archive) of information gathered from many sources, kept under a unified schema, at a single site.

The businesses of all sizes its realized the importance of data warehouse, because there are significant benefits by implementing a Data warehouse. It is generally accepted that data warehousing provides an excellent away for transforming the large amounts of data that exist in these organizations which useful and reliable information for gives answers to their questions and to support the decision making process. A data warehouse contains powerful data analysis techniques that are available today such as data mining and multidimensional analysis, not only the traditional query and reporting, but Making use of these techniques continuously with data warehousing result easier access to the information we need for decision making.

In section 2 we will talk why data warehousing importance and what the benefits of data ware housing, in section 3 we will talking how to gathering data from multi source in to data warehousing and section 4 we will talk data cleansing and section 5we will talking about one technique of Online Analytical Processing (OLAP). Finally section 6 Conclusions and section 7 reference [1, 2, 36, 10].

2. Importance of Warehousing

Most of companies fetching large amounts of data for reporting, also many use data warehousing mechanism to clean, sort and store that large of data. The data warehouse can be used in Data recording and storage growing, and use History data for excellent expectations of the future, and gives total view of the organization to this data and As well as use for support decision-making[1, 3, 5, 10].

2.1 Benefits Data Warehouse

The benefits data ware housing support decision making, Delivers Enhanced Business Intelligence, Saves Time, Enhances Data Quality and Consistency, Provides Historical data And Generates a High ROI.

3. Gathering Data From Multi Source

The data sources send new information in to warehouse, either continuously or periodically where data warehouse periodically requests new information from data sources then keeping warehouse exactly synchronized from data sources but that is too expensive, and acts on data updates are periodically downloaded from online transaction processing (OLTP) systems[1, 7].

4. Data Cleansing

The data cannot respond to saving directly after gathering from multi data sources but in this phase the activity of noise removal from input data before bringing it to the Data Warehouses environment should made. Data cleaning is very important to the overall health of your warehouse project and ultimately affects the health of your company in correct decision making.

The process of data cleaning properly must be done by participation of a domain expert because the detection and correction of anomalies requires detailed domain knowledge.
Data cleaning is described as a process likeness automatic, but must be automatic as far as possible because of the massive amount of data that is usually be processed and a certain time required for an expert to clean it manually[1, 4].

5. Online Analytical Processing (OLAP):

Data warehouse often consisting large data and allows analysis of these data, there are many tools to analyze the data, one of tools is the Online Analytical Processing (OLAP) which enable analysis of these data kept in data warehouse online to support Decisions making and this analysis is referred to as (OLAP). Its sometimes designated as multidimensional online analytical processing (MD-OLAP) or relational online analytical processing (ROLAP).

The data analysis by OLAP obtained via the metadata that document data source, frequency of update, and location of data.

Other definitions:

- Online analytical processing is considered a fast analysis of shared multidimensional information and usefulness to mining qualitative, quantitative from warehouse. Also support as better level of analysis for data warehouse and this technique is used to fetch promptly, report and knowledgebase analysis of traditional knowledge system [1, 2].
- Online analytical processing (OLAP) is executed the complex queries and large list selections in a small number of minutes And essentially element to decision support [1, 3].
- OLAP schema is designed facilities retrieval or return of data and contains of number of large tables called fact tables. These massive tables are surrounded by a number of small tables called Dimensions. Fact table consist data that can be used arithmetically to the analysis of objects. Every dimension table gives the reference to data in fact table [2, 3].

OLAP FASMI Test:

- **Fast:** deliver information to end user in 5 second any few time or most query is answered in that time.
- **Analysis:** applies basic numerical and statistical analysis on this data, which allows user to execute new operational calculations or queries as part from analysis to view the result desired from data[3, 5].
- **Shared:** consider system does not stand-alone and must be implementing all security requirements necessary for confidentiality or sharing potentially confidential data for a multi-user environment.
- **Multi-dimensional:** we will talking it later but consider essential characteristic OLAP.[1, 5]
- **Information: accesses for** all data and derived information required and relevant application, to reside and not limited size.[3, 9]

5.1 Classifications of Online Analytical Processing (OLAP):

Classifies of OLAP for three types:

5.1.1 Relational (ROLAP):

Allows with storage data in relational database and works ROLAP server to create multi dimensions to this data[1, 7].

5.1.2 Multidimensional (MOLAP):

Allows by modeled and view and storage this data in to database in Multi dimensions, in this case we need very large size storage and use to fast processing.[3, 5].

Data cube consists from:

Dimensions consider a collection of members represented by an axis or units from same type which identify form multi dimensions and storage data in multi dimensions focus on numeric data such as values, counts, weights, balances. And each data point associated only one member from multi dimensions. And dimensions hierarchy is arranged member of dimension as example: time dimension contains (all months, quarter, year)

Each hierarchy contains multi hierarchy levels and so on.

a) **Measurement** considers numerical of values the actual numbers called as variables.
b) **Members** group a distinct names or identifier in every dimension to determining data items position.

5.1.3 Hybrid (HOLAP):

There are other types to called Hybrid OLAP (HOLAP) in which combination between them and takes features MOLAP in view data in multi dimensions and fast processing and also takes features ROLAP in storage data in relational database [1, 5, 9].

5.2 Basic operations on OLAP:

Slice

Cuts through of cube or option particular slice from cube and do users focus on some specific dimensions.

Dice:

Is operation focuses on the calculations for particular collection from members dimension instead from all members.

Drill-down:

Is operation for moving down on along the dimensional hierarchy levels and can be gives user higher levels of details and focuses on data to performance more details calculations.

Roll-up:

Is operation for moving up on along the dimensional hierarchy levels and can be gives users can see a summarized level of data and its focuses on data and more general calculations. Look at the following tables for illustrated two operations above (roll-up, drill-down)[6, 7, 9].

4. Conclusions

Data warehouse environment is considered typical for data analysis to supporting decision making, also kept historical data not volatile but statistics data and also its supporting online analytical processing (OLAP) for data analysis so put this data in form multi dimensions after that procedure some operations to gives reports and charts and data in multi dimensions its usually focus on numeric data such as values, counts, weights, balances. OLAP needs for large size and achieve fast processing and answers on questions from data warehouse to presents for discussion maker. We note OLAP useful for development of company and economic etc.

5. Recommendation

The study strongly recommended that all companies, institution should make use of warehousing data and the OLAP for ensuring development.

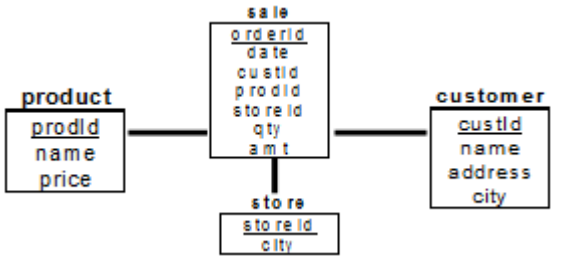

Figure 1: Star schema. Sale is fact table, and product, customer, store dimensions table.

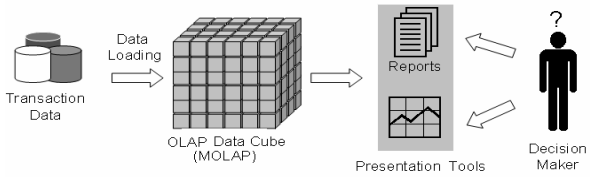

Figure 2: Three dimensions data cube: Represented viewed data in data cube

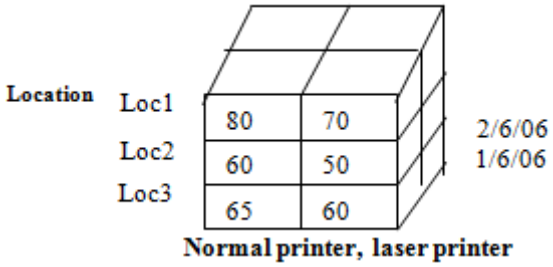

Figure 3: Three dimensions data cube

Table 1: Representation of table 1

Product			
Location	laser Printer	Normal printer	Total
Loc1	80	70	150
Loc2	60	50	110
Loc3	65	60	115

Table 2: roll-up, drill-down. Note drill-down considers reverse to roll-up

Volume of prod		1996			
		Q1	Q2	Q3	Q4
west	San Jose	78	45	34	56
	bouder.	90	67	87	91

Roll-up Dimension: time

Volume of prod		Quarter1		
		jan	Feb	mar
west	San Jose	30	26	22
	bouder	28	30	32

References

[1] Chuck Ballard, Dirk Herreman, Don Schau, Rhonda Bell, Eunsaeng Kim and Ann Valencic. (1998)-"Data Modeling Techniques for Data Warehousing", "International Technical Support Organization".

[2] Raghu Ramakrishnam and Johannes Gehrke(1999) "Data base management systems second edition", July 2002, http://pages.cs.wisc.edu/~dbbook/..

[3] (1998).Spatial Data Warehousing for Hospital Organizations an ESRI White Paper March 1998, Environmental Systems Research Institute.

[4] John Wang(2008), "Encyclopedia of Data Warehousing and Mining Second Edition", Montclair State University, USA, Vol II, Published in the United States of America by Information Science Reference (an imprint of IGI Global).

[5] Jiawei Han, Micheline Kamber, "Data Mining: Concepts and Techniques".

[6] Deepak K Pande, "Data warehouse techniques in traditional knowledge systems", e mail: pandedk1181@yahoo.com, received 6/july/2004, indian journal of traditional knowledge ;vol4(4) received 9/july/2005.

[7] G.N. Wikramanayake and J.S. Goonetillake(2007)."Managing Very Large Databases and Data Warehousing", Sri Lankan Journal of Librarianship and Information Management Volume 2, No.1 pp.22-29.

[8] University of Colombo School of Computing 35 Reid Avenue, Colombo 07, Sri Lanka, Tel. +94-11-2591065, Fax: +94-11-2587239 email: {gnw, jsg}@ucsc.cmb.ac.lk

[9] Sarina Sulaiman1, Siti Mariyam Shamsuddin1 and Ajith Abraham2."Data Warehousing for Rough Web Caching and Pre-fetching". 1Soft Computing Research Group (SCRG) K-Economy Research Alliance (RAKE) Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia 2Machine Intelligence Research Labs (MIR Labs), Washington 98092, USA. http://www.mirlabs.org sarina@utm.my, mariyam@utm.my, ajith.abraham@ieee.org.

[10] JishaJ Dell (2009)."Database Solutions Engineering". Product Group (July 2009).

[11] George Lumpkin(2009)." Oracle Database 11g for Data Warehousing and Business Intelligence", Oracle Corporation, September 2009.