

# LITERATURE SURVEY ON DATA WAREHOUSE AND ITS TECHNIQUES

MUHAMMAD KHALEEL (0912125)

SZABIST KARACHI CAMPUS

**Abstract.** Data warehouse and online analytical processing (OLAP) both are core component for decision making and support which had become focus of attraction in the data base management industry. By increasing demand of these two thing in data base there are now there is much these type of products in commercial and user base and as well as this type of services are available in market as well as to the vendors of data base management systems has offered in these area. Whereas the decision making support have some different requirement in database technology as we can compare with traditional online transaction processing application. In this literature survey data warehousing and OLAP technologies have been gathered with new requirement. The literature survey I have also described about the back end tools which are used for extraction, cleaning and loading into data warehouse and making multidimensional data models typically of the OLAP. Where as in the front end client tools are of the query execution and it also include data analysis where as the server extension are efficient in query processing also in other type of tools for metadata management as well but also for the managing that data ware house.

## 1 BACKGROUND

Data warehousing is usually says about the collection of data for decision making and the support technology but aimed for enabling the knowledge to make the better and faster decision. Whereas the data warehouse mostly narrow means the data store itself which has or emphasize on the aggregated data over the detailed and the individual records which are usually found in transactional industry. We can see for the past several years in the industry have seen a high rate of growth as for product numbers available, as well as the services that is offered and the adaption of such kind of this type of technology by this industry. The Meta group had a survey which states that data warehousing and its market which includes the hard ware, database software the tool have been project to growth from two to eight billion for span of the year.

## 2 INTRODUCTION

Whereas we can see the data ware housing and its technologies has been able to successfully deployed in many technologies and as well as has being able to

successfully deploy in other many industries like manufacturing, the retail, financial tool services, transportation and many other important industries of day to day life. The data warehouse as said before supporting the online analytical processing which is (OLAP) that can be for functional with performance wise requirement is having This is somehow which is quite different from such these of type of online transaction processing (OLTP) and other applications that is traditional and supported by other operational databases. The main online transaction processing (OLTP) are usually in the clerical data processing step like in the order entry and as well as in the banks for transaction which is one of the most important for day to day operation for any organization If we look into this task some of them are usually structured as well as repetitive and usually among them consist of some small isolated type transaction. These transaction in usual requires usually require the detail of the up to date data as well of usually they read and in some cases they update only just few records and usually they are accessed by their own primary key. The operational data base is usually tended to be for the hundred of megabyte to gigabyte in size. In every part the consistency and recoverability of any data base is critical. The most consequently the database are usually define for the reflective some operational semantic for some known application and as well as for particular for concurrency conflicts. A data warehouse has its own contrast for target and decision support as well. The historical data, the summarize data and consolidated data in this is important then individual and detailed record. As data ware house contain consolidated data. The database and data warehouse as contain consolidated data but for operational database the data base are in gigabyte and terabyte. We can say that data warehouse can be implemented for some standard or as well as for some this type of extended in relational data base management system These type servers usually assumes that the data which is stored among these relational databases have also support some extension to sequential language for special access and as well as for implementation method one of the multidimensional data models and its operating from these contrast the multidimensional online analytical processing (MOLAP) serves are one of the those type of servers that usually direct stores multi dimensional data like these special data structures are arrays and many more which can also be implement in these online analytical processing operations on over like these one of the data structure.

### **3 LITERATURE SURVEY**

#### **3.1 DATA WARE HOUSE ARCHITECTURE**

The architecture of data ware housing includes tools one of those tools which can extract the data base from different and multiple operational database servers and from different external data base and surely which is used for cleaning the data, transforming and to integrate the data by loading data to a ware house. Whereas the data in the ware house and as well as in data marts also stores and as well as managed by these more than single ware house servers .The data in that can be presented in multidimensional view and from variety of front one of

the end tool as well as query tools for these such report writing analyzing tools and at the last for the data mining tools. As we also know that there is also repository for storing the data and as well as for staging the data and as well as the staging the metadata and the last also the tools which is represent that the monitoring and administering the warehouse system.

### **3.2 DATA WARE HOUSE BACK END TOOLS AND ITS UTILITIES**

As we know that data ware housing system usually its uses variety for data extraction, as well as for Tools for data cleaning , to load as well as for refreshing utilities for other populating ware houses. As for the data extraction for different sources and other locations are some usually implanted though special gateway as well as for standard interfaces. The data ware houses are usually used in decision making so that it is one of the most important thing that just the data should be correct. The other data of large volume are extracted from multiple sources so that there are chances of high risk of errors in the data. Tools are available which help in detect data and to correct the data. As data cleaning there are three classes. The number one in these three is data migration by this term which is usually allow simple rule to transform with specification and specified location. The second among three is data scrub ling tool which usually have domain specific knowledge have scrambling the data. The third and last is the data auditing is one of those tools which discuss and make the possible rules and the relationship by summing the data. This tool might and looks like the data mining tools. In data base after extracting the data, the cleaning the data cleaning and transforming the data into warehouse. The back load activities usually use for these purposes the warehouse. For such the load utility shortcut that allows the system administrator for monitoring the status so that if he want to cancel he cancel or approve. It can suspend to resume and to load and to restart when failure occur and there should be no live of data integrating. The warehouse should be refreshing in updating the source data as well as the base data and data should be derived from data stored in the data warehouse. As the refreshing techniques that may also depend for on the characteristics from the other source and other capabilities for data servers. The extraction of an entire file on the data base is always expensive and the choice for the legacy the data source. Maximum contemporary data base usually provides the other replication servers for the support of incremental techniques just for update from a primary database for one or more replication. As like these type of replication server this has always been used in refreshing for data warehouses. As we know that these refreshing cycles have to be properly choosing so that these types of data should not be exhausted and trouble shoot in the load utility any more.

### 3.3 TOOLS FOR THE FRONT END ANDECONCEPTUAL MODW

One of the most popular models which are conceptual which influences front end tools the data base and its design and query engines for online analytical processing (OLAP) of the multidimensional view of data in warehouse. Whereas all the multidimensional data a model has always a set of numerical measure are included that have objects of analysis. The other distinctive feature for conceptual model for OLAP have stress in the aggregation for measure for one or more dimensions has one of the key operations for this computing , the reading and calculating for total sales for each country and as well that is for each year also. In this another most popular operation has some comparing any two type of operation measures have aggregated some dimensions. Times is also one of the dimensions and have a great significance support for decision support. Have to build in ware houses of considerable and other as aspect of time dimensions. As we know that the Multidimensional data model usually grew out to be in the views just for business data which are popularized in spreadsheet which is presented on computers and most of them are usually used for business analysis. As the spreadsheet up till now most liked front end application for online analytical processing (OLAP). There is a challenge for the support queries environment just for the purpose of online analytical processing (OLAP) that are purely crudely summarized as for supporting the spread sheet operation for such a large multidimensional gigabyte or for the multidimensional terabyte databases. As we know that the products of era base for arbor cooperation are used in Microsoft excel as the front end tools as we know only for multidimensional engines. The other different operators related for the pivoting are some rollups or drill down, corresponding the roll up for taking some current data for objects and as well as doing for further grouped by one for the dimensions and for the other rollup it is possible to use sales data. As we know that a multidimensional spread sheet have attracts for a lots of interest from when it does empowers for the user end for analyzing its business data which is present, by this have not replaced traditional analysis which is usually managed by a query environment. As these environments usually uses a stored procedure and as well as for predefined such complex queries for providing packaged analysis tool. These type of tools we know that it makes that possible just by the end users to have query by the term for domain specification for businesses and its data. As this type of application usually often time it uses a row access of data tools and as well as optimize for the access pattern which are data base servers. We can add there are query environments for helping in building ad hoc sql queries just for the point and to click. As the last now that are many varieties in different data mining which are usually used just for user end tools that for data warehouses.

### References

1. <http://db.cs.berkeley.edu/claremont/claremontreport08.pdf>

2. <http://research.microsoft.com/pubs/76058/sigrecord.pdf>
3. <http://infolab.stanford.edu/pub/papers/warehouse-research.ps>
4. <http://pages.cs.wisc.edu/~ahollowa/paper377.pdf>