

Call Management Data Warehousing

Pooja Jaiswal

Master of Computer Application, V.E.S Institute of Technology, Mumbai - 74, India

Abstract: In computing, a data warehouse is a system used for reporting and data analysis. DWs are central repositories of integrated data from one or more disparate sources. Data warehouse acting as a decision support systems. Data warehouses standardize the data across the organization so that there will be one view of information. Data warehouses can provide the information required by the decision makers. The data warehousing module is intended for centralizing the storage location of the organization's data. This allows the business to maintain one version of the truth for reporting and analysis.

The paper focuses on different components in the Data warehouse and its schemas. The main components include the ETL process, OLAP and other client applications residing on the Data warehouse. Extraction–transformation–loading (ETL) tools are pieces of software responsible for the extraction of data from several sources, its cleansing, customization, reformatting, integration, and insertion into a data warehouse. Building the ETL process is potentially one of the biggest tasks of building a warehouse; it is complex, time consuming, and consumes most of data warehouse project's implementation efforts, costs, and resources. Building a data warehouse requires focusing closely on understanding three main areas: the source area, the destination area, and the mapping area (ETL processes).

OLAP is a trend in database technology, based on the multidimensional view of data. Multidimensional data (called multicubes) form the basic logical data model for OLAP applications and a significant number of proposals exist for cube models, without – in most cases – the distinct description of an equivalent “presentational” layer or model for the presentation of multidimensional data and their hierarchies.

Keywords: ETL process, OLAP, Call Management Data Warehousing.

1. INTRODUCTION

A data warehouse (DW) is a collection of technologies aimed at enabling the decision maker to make better and faster decisions. Data warehouses differ from operational databases in that they are subject oriented, integrated, time variant, non-volatile, summarized, larger, not normalized, and perform OLAP. The generic data warehouse architecture consists of three layers (data sources, data staging area, and primary data warehouse). Although ETL processes area is very important, it has little research. This is because of its difficulty and lack of formal model for representing ETL activities that map the incoming data from different DSs to be in a suitable format for loading to the target DW. To build a DW we must run the ETL tool which has three tasks:

- (1) Data is extracted from different data sources,
- (2) Propagated to the data staging area where it is transformed and cleansed, and then
- (3) Loaded to the data warehouse.

ETL tools are a category of specialized tools with the task of dealing with data warehouse homogeneity, cleaning, transforming, and loading problems. This research will try to find a formal representation model for capturing the ETL processes that map the incoming data from different data sources to be in a suitable format for loading to the target DW.

OLAP and data warehousing has become a major research area in the database community. Data warehouses might be implemented on standard or extended relational DBMSs, called Relational OLAP (ROLAP) servers. These servers assume that data is stored in relational databases, and they support extensions to SQL and special access and implementation methods to efficiently implement the multidimensional data model and operations. In contrast, multidimensional OLAP (MOLAP) servers are servers that directly store multidimensional data in special data structures (e.g., arrays) and implement the OLAP operations over these special data structures.

2. RELATED WORK

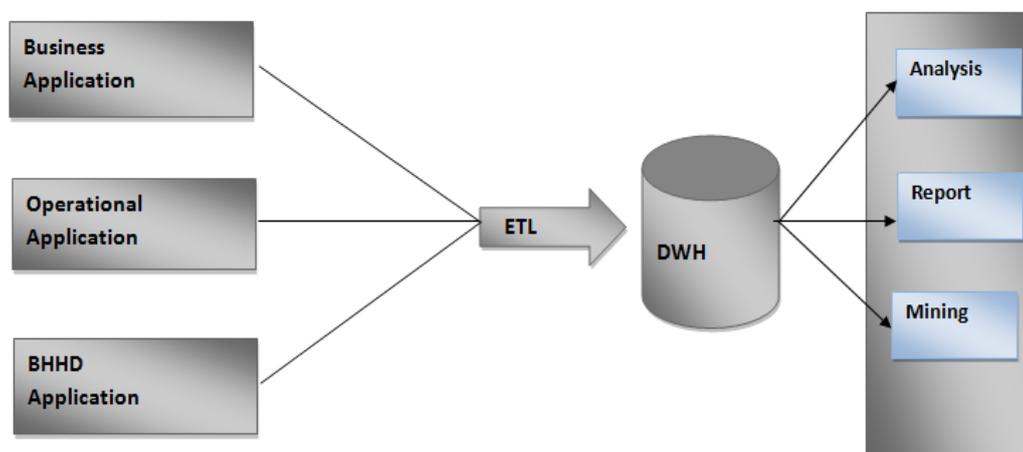
This paper summarizes the experience in designing and modeling an data warehouse. Existing facilities and databases affect the chosen data warehouse that brings them together to support decisional activities for IT Department. The choice to develop a dedicated system is mainly forced by the peculiar information type that defines the basic information in data warehouse widely different from organization to organization. Data Mart can hold information which addresses both strategic and tactical information needs and provides information which allows key operating functions to effectively manage performance. It unifies information from various databases into a single database. Different varieties of approaches for the integration of ETL tool in data warehouse have been proposed.

The ETL process, in data warehouse, is a hot point of research because of its importance and cost in data warehouse project building and maintenance. The method is a systematic review to identify, extract and analyze the main proposals on modeling conceptual ETL processes for Data warehouse. The main proposals were identified and compared based on the features, activities and notation of ETL processes and concluded the study by reflecting on the approaches being studied and providing an update skeleton for future study.

3. PROPOSED DATA WAREHOUSE ENVIRONMENT

The data mining from data warehouse can be a ready and effective system for the decision makers. The reputed company Mahindra Finance has been taken for this study. Figure shows the DWH architecture of Mahindra Finance. The information is spread across diverse platforms, data from different sources has to be taken from different sources and then consolidated to produce required report.

ETL activities are performed to extract the data from heterogeneous sources and load into staging and then load the data into dimension and fact tables as per the schedules. The purpose of this paper was to investigate current system of information delivery and proposing a better system for timely, accurate and consistent information delivery to the decision makers.

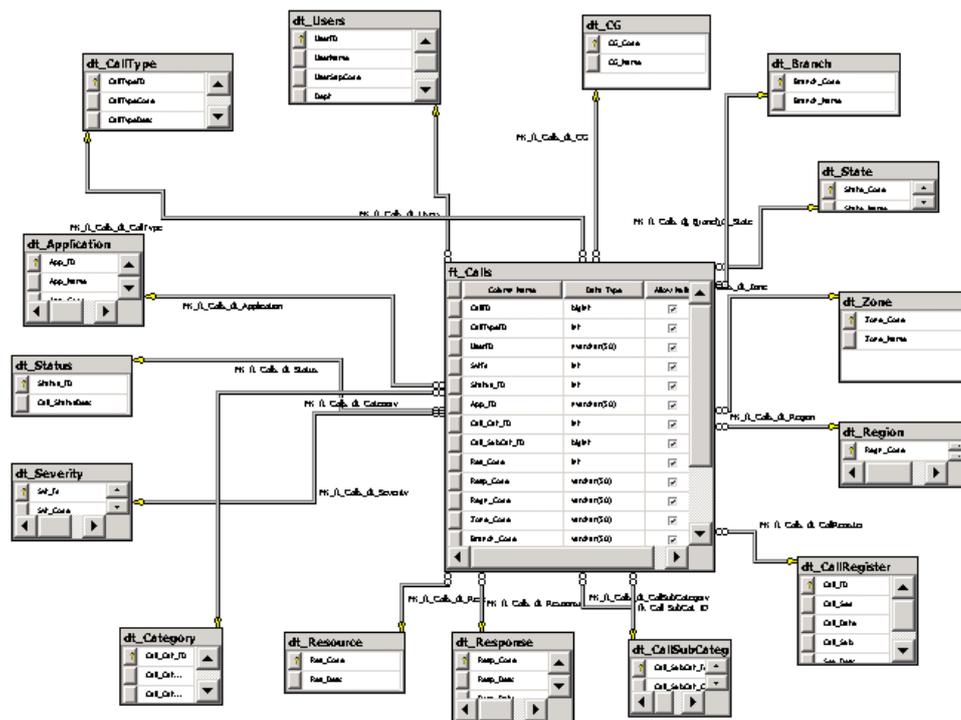


3.1.1. Data Modeling Design:

Data modeling is a technique for defining business requirements for a database. It is sometimes called database modeling. A database model is a specification describing how a database is structured and used. Here we have decided to have Star schema.

3.1.2. Star schema:

A star schema is a modeling paradigm in which the data warehouse contains few "fact tables" (possibly only one, justifying the name) referencing any number of "dimension tables". The fact table contains the detailed summary data. Its primary key has one key per dimension. Each dimension is a single, highly de-normalized table. Every tuple in the fact table consists of fact or subject of interest, and dimension that provide the fact. The dimension table consists of columns that correspond to the attributes of the dimension. Star Schema designed for data marts.



3.1.3. ETL Activity:

There are several transformations available to transform the data from source files to target tables. Expression transformation is used to transform the format of date or calculate anything if needed before loading the data into target. Filter transformation filters the records as specified in the condition part. Update strategy is used to update the records. Sequence generator transformation generates continuous values, which can be used to generate surrogate keys. Router is an advanced filter which is used to direct the output to two or more different tables based on the condition specified. Joiner transformation is used to join tables from two or more heterogeneous sources. Lookup transformation is used to check the content of one or more attributes of target table before loading the data from source.

List of the Source Systems that are used in Mahindra Finance

1. EMLAP - Enquiry Management Loan Application Processing
2. PLAP - Personal Loan Application Processing
3. GLAP - Gold Loan Application Processing

3.1.3.1. Data Extraction process:

In this step data is extracted from source tables to respective staging tables. Data Extraction is done by PL/SQL scripts.

There are 25 source tables of EMLAP and 26 source tables of PLAP are used for extraction purpose.

3.1.3.2. Data Transformation process:

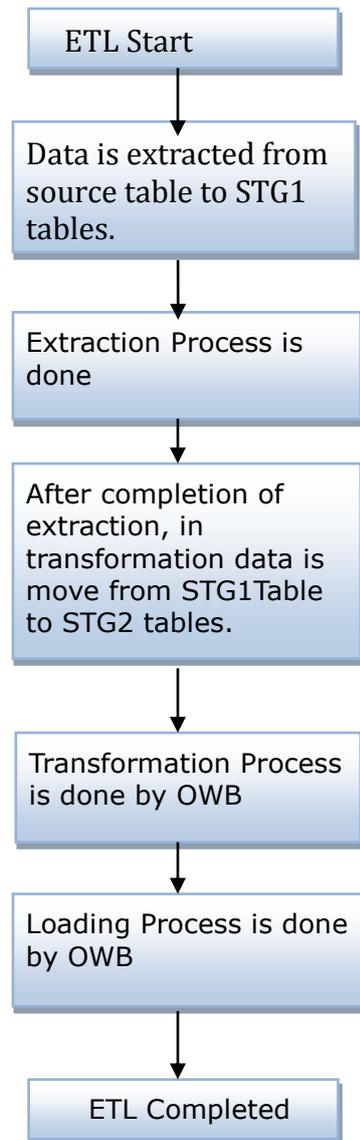
After successful completion of Extraction process, Transformation process starts. Data transformation is done by OWB tool. In transformation, data is moving from 'Staging 1' table to 'Staging 2' table with all necessary data type conversion.

3.1.3.3. Data loading process:

Once the transformation phase completes, data loading is started using OWB tool.

During this process all required mapping of Dimensions and Facts need to run. All mapping are scheduled as job.

Flowchart of ETL:



4. RESULT

Some of the results observed after querying the data marts are documented below. The results are cross checked with the requirements specified by the different types of users. The three different data marts are queried using SQL query. The results returned by the queries are found accurate and meeting users demands. The sample screen shots of queries and the results are shown.

Below query shows application wise call count with resolution. The information required was the number of closed calls.

```
SQL> Select TOP 10(f_CallCatDesc),  
COUNT(f_CallCatDesc) AS 'Call Count'  
from [172.30.0.222\mfgponl].[MmfsI_App_Call_Mgmt].[dbo].[t_CallRegister_Hdr]
```

where f_Lst_RespTpDesc='Closed'

AND f_CallDate between '10-01-2014' and '12-31-2014'

GROUP BY f_CallCatDesc

ORDER BY 2 DESC

	f_CallCatDesc	Call Count
1	HHD	6543
2	BHHD	6041
3	Other	4164
4	SMS SEVA	2350
5	Desktop / Laptop	1468
6	Bits Specific Request	1056
7	New Requirements	980
8	EMLAP	848
9	Intranet Legacy Solutions	739
10	Printer	635

Below query shows application wise, priority wise and status wise call count. It gives the count of specific periods.

Select TOP 10(f_CallCatDesc),COUNT(f_SvrtDesc) as Count,

f_SvrtDesc,f_Lst_RespTpDesc

from [172.30.0.222\mfgponl].[MmfsI_App_Call_Mgmt].[dbo].[t_CallRegister_Hdr]

where f_CallDate between '10-01-2014' and '12-31-2014'

group by f_SvrtDesc, f_SvrtDesc,f_Lst_RespTpDesc, f_CallCatDesc

ORDER BY 2 DESC

	f_CallCatDesc	Count	f_SvrtDesc	f_Lst_RespTpDesc
1	HHD	6240	Severity 2	Closed
2	BHHD	5701	Severity 2	Closed
3	Other	3707	Severity 2	Closed
4	SMS SEVA	2349	Severity 2	Closed
5	Bits Specific Request	932	Severity 3	Closed
6	Desktop / Laptop	889	Severity 2	Closed
7	New Requirements	864	Severity 2	Closed
8	EMLAP	748	Severity 2	Closed
9	Intranet Legacy Solutions	697	Severity 2	Closed
10	BHHD	668	Severity 2	NULL

5. CONCLUSION

Data warehousing is the leading and most reliable technology used today by companies for planning, forecasting, and management.

The Extraction, Transformation and Load (ETL) process is the most critical component of a data warehouse, however getting this right does not conclude that the implementation is a success. ETL is a continuous process as the source database will continue to change and when this occurs, the business rules or transformation logic will also need to be changed and thus, code changes may be required. Data Warehouse teams cannot “set and forget” ETL routines after a successful iteration of the ETL routine. This routine needs to be fine-tuned or refined regularly to protect the integrity of the data warehouse.

6. FUTURE SCOPE

Data Warehousing is such a new field that it is difficult to estimate what new developments are likely to most affect it. Future scope for this research paper is as follows:

1. Management Dashboard system
2. User Dashboards
3. Fine tuning of ETL jobs as data will grow every year
4. ETL enhancements e.g. tracking of each activity of ETL in database, start time, end time, failure notification
5. Capturing Audit related information e.g. logged in user details, IP address, Host time and date and time
6. Schedule an mailer
7. Sending the important Information by email.

ACKNOWLEDGEMENTS

This paper is prepared through experimental results of data warehouse of reputed company, queried with all possible combinations and consulted the decision makers for the usability.

REFERENCES

- [1] Raghu Ramakrishnan, Johannes Gerhke, "Database Management Systems" McGraw Hill
- [2] Ralph Kimball, "The Data Warehouse Lifecycle Toolkit", John Wiley.
- [3] Alex Berson, Data Warehousing Data Mining & OLAP, Computing Mcgraw-Hill, November 5, 1997.
- [4] Korth, Silberchatz, Sudarshan, "Database System Concepts" McGraw Hill.