# Comprehensive Study For Design in Data Warehouse Using UML

**Mrigyanka Mudgal\*, Pravin S. Metkewar\*\* and Monika Sharma\*\*\***

### ABSTRACT

Data warehousing is the concept which is widely blooming. Despite the fact that data warehousing requires strategies that are not as similar as the operational data warehouse, no logical method has been proposed so that the complete philosophy can be relied upon in the data warehouse. In this paper, we plot a general methodology system for DW that forecast and examines the connections between the various steps which distinguish the data warehouse from operational data warehouse. Various methods and approaches have been provided to design data warehouse. In this paper, we are conversing and discussing about how the data can be used in UML methodology. As the data has to go through various steps and then it is stored in the data warehouse. In order to make this a reliable process we are using UML methodologies.[1]

*Keywords:* UML, Data Warehouse Design, Multidimensional Modeling, Database, Data Warehouse

## 1. INTRODUCTION

Before starting with the design of the data warehouse. We should have knowledge about what data warehouse is, "Data warehousing is the procedure of building and utilizing a data and is a repository for storing data for future predictions.[1] The Data warehousing center is built by coordinating information from different sources that supports analytical reporting, organized and/or ad-hoc inquiries, and decision making." As the database group is dedicating and expanding, it focuses on certain issues regarding the research topic concerning data warehousing centers (DWs) .DWs concentrates on particular issues, for example, multidimensional data models. In this paper, my exploration is gone for characterizing the essential steps required for a best DW plan. The study is made on the Object Oriented Design using UML (Unified Modeling Language), it allows various users to tackle most effective data warehousing design. Unlike other approaches such as multidimensional analysis or data warehouse design. To the best of my knowledge, not much of the exertion has been made so far to build up a fully structured and reliable configuration technique.[2] This paper is based on the process of building and utilizing the data warehouse. With the fast advancement of data and correspondence innovations and spread of technological use, the database outline comprises of taking different stages, the data is collected from various sources that is internal and external data sources and the data is transferred to the ETL process and the data that is required is extracted and then is put into the staging area, after that it is transferred to transformation of the data and then loaded. The main stage is Investigation of operational data warehouse whose main criteria is to gather the data concerning the already accessible and existing data warehouse framework which includes the originator, alongside the general population included in overseeing the data warehouse and produces in yield the (theoretical or sensible) plans of either the entire or part of the data warehouse. The following stage comprises in social occasion and shifting the client prerequisites. It includes the developer and clients of the DW, and produces its output. Here, we are discussing that the stage comprises in social occasion and shifting the client prerequisites.

\* SICSR Symbiosis International University Pune, India, *Email: Mrigyanka.mudgal@gmail.com*

\*\* Assoc. Professor, SICSR Symbiosis International University Pune, India, *Email: pravin.metkewar@sicsr.ac.in*

\*\*\* Amity Institute of Information Technology, Amity University, Uttar Pradesh, Noida, India, *Email: Msharma5@amity.edu*

It includes the developer and clients of the DW, and produces in output, we have also discussed that the datawarehouse is designed in various layers that are divided and independent but interact with each other, the data is gathered from various sources and then ETL is applied to it, this paper consists of all the processes which are applied when data travels from ODS to DWSS.

## 2. LITERATURE REVIEW

In the blooming industry, data is huge and in order to store and provide vital information various researches are being done. In this literature, we have referred various papers published by many scholars, the procedure of adding to a data warehouse begins with recognizing and assembling prerequisites, outlining the dimensional model created after the testing and designing. [4] The outline stage is excessively imperative action in the effective working of a data warehouse. In this paper, I reviewed and assessed the literature with the different data warehousing center outline approaches on the premise of configuration criteria and propose a summed up article situated reasonable outline system in light of UML that meets a wide range of client needs. Various authors have proposed many frameworks to create the proper research and they however propose how to build the schema from data sources and different other approaches such as top down and bottom up approach. As per the research paper, after researching a lot we have studies about the data warehouse designing steps and how data passes from various phases and is stored in the data warehouse.

## 3. DESIGN OF A DATAWAREHOUSE

Table 1 show the peak signal to noise ratio of performance of our proposed method of watermarked image and original image with various watermark image, where our watermarked images peak signal to noise ratio has a better performance than others.

The architecture of the data warehouse has several layers, and each layer is divided in such a way that they interact with each other. Each layer is dependent on one another:

ODS (Operational Data Source) - It defines the structure of the operations and sources, there are two types of sources, external and internal sources.

DWCS (Data Warehouse Conceptual Schema) – It defines the conceptual schema of the data warehouse that consist of facts, dimensions, hierarchy etc.
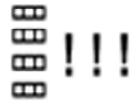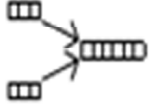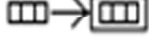
DWSS (Data Warehouse Storage Schema) – It defines physical storage of DW depending upon the targeted platform. In order to cover the schema, there are various activities that happen:

E (Extracted) T (Transformation) L (Loading).ETL process defines the mapping between ODS and DWCS.

Hence, it shows the mappings that is required for designing a datawarehouse. In the above diagram, the operational and external data sources are defined in ODS and from ODS to DWCS, ETL process took place that defines the mapping between them. Therefore different schemas are related to each other by UML dependencies. Following dependencies are shown:

$\rightarrow$ ETL $\longrightarrow$ {ODS, DWCS}

$\rightarrow$ EXPORTATION $\longrightarrow$ {DWSS, DWCS}

$\rightarrow$ BM $\longrightarrow$ {DWCS}

With the help of UML, the design is simplified as the data is very huge. The DW designer achieve the design from different levels of details. The different schema are dependent on each other using dependencies. As previously mentioned the ETL process defines the mapping between ODS and DWCS, it can be performed using UML technique as it provides an easy mechanism for the processes that occur such as integration of

| ETL Mechanism (Stereotype) | Description | Icon |
|---|---|---|
| Aggregation | Aggregates data based on some criteria | |
| Conversion | Changes data type and format or derives new data from existing data | |
| Filter | Filters and verifies data | |
| Incorrect | Reroutes incorrect data | |
| Join | Joins two data sources related to each other with some attributes | |
| Loader | Loads data into the target of an ETL process | |
| Log | Logs activity of an ETL mechanism | |
| Merge | Integrates two or more data sources with compatible attributes | |
| Surrogate | Generates unique surrogate keys | |
| Wrapper | Transforms a native data source into a record based data source | |

sources from both internal and external sources. The transformation is done from the integrated source to the target. After the mentioned activities are performed, then ODS processes are designed and they provide the transformation and loading into datawarehouse. In this research paper, an ETL process is composed of UML packages, each process in ETL is represented by various stereotypes, and thus various icons are defined for each ETL process.[3]

## 4. CONCLUSION

After studying about a data warehouse design that is based on the use of UML diagrams. I believe the best advantage is that even if we have large data as we are dealing with big data we are able to process it using ETL, so the design can be used for huge data as well and the notations will be same of designing the different DW schemas. With the help of UML diagrams we can give directions and can provides with easy mechanism to handle large amount of dataTable Styles

## 5.   FUTURE SCOPE

Building a data warehouse is a critical procedure that could end in a problem situation if took care disgracefully. There are a few obstacles in the process that should be overcome to make progress successful. These obstacles regularly take a broad measure of time to overcome, particularly the first occasion when they're experienced. Thus, in future this architecture with use of UML diagrams can outline can change the picture of designing of data warehouse. In future, will provide the algorithms and that should be applied for designing of data warehouse.

## REFERENCES

[1]   Journal of Computer Science and Information Management, Vol. 2, N. 3, 1999 Copyright@ 1999, Maximilian Press Publisher

[2]   International Journal of Database Management Systems ( IJDMS ) Vol.4, No.1, February 2012 DOI: 10.5121/ ijdms.2012.4104 33 COMPARATIVE STUDY OF DATA WAREHOUSE

[3]   Luján-Mora, S., Trujillo, J., Song, I.: Multidimensional Modeling with UML Package Diagrams. In: Proc. of the 21st Intl. Conf. on Conceptual Modeling (ER'02). Volume 2503 of LNCS., Tampere, Finland (2002) 199213

[4]   A UML Based Approach for Modeling ETL Processes in Data Warehouses Juan Trujillo, Sergio Luján-Mora Departamento de Lenguajes y Sistemas Informáticos Universidad de Alicante {slujan, jtrujillo}.

[5]   Golfarelli, M., Rizzi, S.: A methodological Framework for Data Warehouse Design. In: Proc. of the ACM 1st Intl. Workshop on Data Warehousing and OLAP (DOLAP'98), Bethesda, USA (1998) 39 4. Sapia, C., Blaschka, M., Höing, G., Dinter, B.

[6]    F. Gonzalez and J. Hernandez, " A tutorial on Digital Watermarking ", In IEEE annual Carnahan conference on security technology, Spain, 1999.

[7]   Carneiro, L., Brayner, A.: X-META: A Methodology for Data Warehouse Design with Metadata Management. In: Proc. of the 4th Intl. Workshop on Design and Management of Data Warehouses (DMDW'02), Toronto, Canada (2002) 1322

[8]   Luján-Mora, S., Trujillo, J., Song, I.: Extending UML for Multidimensional Modeling. In: Proc. of the 5th Intl. Conf..