

Name of the Scholar : **Venkata Subrahmanyam Vampugani**
Name of the Supervisor : **Prof. M. N. Doja**
Department : **Department of Computer Engineering,
Faculty of Engineering and Technology**
Title of the Thesis : **Conceptual Modeling of Data Sources of a
Data Warehouse Using UML**

Abstract

Developing a DW is a complex, expensive, time consuming, and prone to fail task. Different DW models and methods have been presented during the last few years. However, none of them addresses the whole development process in an integrated manner. In this thesis, we have presented our DW development method, based on the UML and the UP, which addressed the analysis and design of both the DW back-stage and front-end. For this task, we have extended the UML in order to accurately represent the different parts and properties of a DW. Our work provided a seamless method for developing DW and it is a great help when designing, implementing and deploying a DW. The following are the achievements of our research work:

- Done the literature review and studied all the models to design DW and a comparative chart is prepared depicting all the parameters.
- We haven't used the ER instead used the UML to model the data sources of the DW at the conceptual level as the ER modeling suffers with 3 major problems:
 - There is no standard and several variations exist.
 - ER diagrams tend to be messy and difficult to read.
 - The ER approach is weak at handling object-oriented design issues such as inheritance (subtypes and composition).
- Our method is called as the Data Warehouse Engineering Process (DWEP) based on the Unified s/w Development Process also known as UP. Three keywords make the UP unique as stated below:

- (i) Use Case(requirements) driven
 - (ii) Architecture Centric
 - (iii) Iterative and Incremental
- We used our Multidimensional Profile for the design of the DW at the conceptual level.
 - We applied our Data Mapping Profile for creating the mapping between the data sources and the DW at the conceptual level.
 - We used different UML extensions to model the data sources and the DW at the logical level: UML Profile for Database Design, a UML profile for data modeling, a UML extension for the modeling of XML documents by means of Document Type Definition (DTD) and XML Schemas, etc.
 - We designed the ETL processes that will be responsible for the gathering, transforming and uploading data from the data sources into the DW. We used our ETL Profile for this task. Finally, we used our Database Deployment Profile for taking physical design decisions.
 - Designed architectures for two case studies:
 - Case Study – 1: *Data Warehousing Architecture for an Open University System*
 - Case Study – 2: *Data Warehousing Architecture for eGovernance: Data Mart Approach*
 - Studied and presented the *Data Quality aspects in Data Warehousing*.

In a nut shell, our approach resolves the following problems of the MD modeling:

- (i) Multistar models
- (ii) Shared dimensions
- (iii) Shared hierarchy levels
- (iv) Multiple and alternative classification hierarchy levels
- (v) Heterogeneous dimensions
- (vi) Shared aggregation
- (vii) Derivation rules