CS Masters' Thesis Defense

Title:	*Importance of Metadata in Data Warehousing*
Speaker:	Abhinav Dhiman
Date:	Monday, May 7, 2012
Time:	2:00 p.m.
Location:	GMCS 405
Thesis advisor:	Dr Joseph Lewis

Abstract:

Data warehouses are an integral component for business decision making when a large volume of data has to be analyzed. Data in data warehouses come from sources including, but not limited to relational databases, legacy systems, and flat files. The data from these sources have to be processed to conform to the data model and type of the data warehouse. Extract, Transform, and Load (ETL) tools are used to take data from multiple sources to populate a data warehouse, typically into a star schema. After the data is loaded into the star schema, it can be used for query from reporting tools, also known as the reporting layer of data warehouse.

There are three parts to the data warehousing, first is the source systems, second is the ETL processing, and third is the reporting part. In order for well-functioning of all the parts, they need to work together comprehensively. Metadata, usually called data about the data is a vital component necessary in to make sure the smooth working of data warehouse. Metadata ties all the loose ends of the total parts involved in data warehouse. Metadata has to be generated at all the levels and stages of data warehouse.

The main purpose of this research is to bring up the importance of modeling of metadata model while designing data warehouse. This thesis presents the metadata model for source metadata and for ETL metadata which can be applied to any data warehouse. In most of the design scenarios the metadata is not focused during the development phase of data warehouse. According to Ralph Kimball, there is no standard way of doing the metadata implementation in Data warehouse. By presenting a metadata model, the main aim to understand the importance of metadata modeling and a standard way of implementing the metadata in data warehouse.