ABSTRACT
Designing a data warehouse system needs to integrate distributed/heterogeneous databases into a reconciled data platform for strategic use. To establish the reconciled data platform, a global schema or global data views which contain higher semantic and integrity constraints are required during the development phase. In this dissertation, we describe a novel methodology to integrate independent local logical database schemas into a global semantic database schema. These schemas are converted from the local databases to Extended Entity Relationship (EER) models. Equivalency of domain, attribute, entity, and relationship are identified between two of these EER models. While a discovery and resolution mechanism by using these equivalent definitions is applied to deal with the naming conflict and structural conflict between two EER models, a reconciled semantic view between the two EER models is established. A semantic merge mechanism by using data analysis technique is also applied to extract more semantics during the merge phase. Two local database schemas are then integrated. These steps are repeated until the schemas of all databases to be integrated have been consolidated into a single global schema.

Keywords: data warehouse; database schema; schema integration

Table of Contents
Contents Abstract.......................................................1 Contents......................................................2 List of Figure...................................................IV Chapter 1 Introduction 1.1 Overview...................................................1 1.2 Research Motivation and Objective..........................5 1.3 Organization of the Dissertation...........................6 Chapter 2 Related Work 2.1 Formulation of Schema Integration Problem..................9 2.2 Taxonomy of Schema Conflict................................9 2.3 Related Schema Integration Methodology....................10 2.3.1 Superviews: An Approach Based on Integration Operators ....................................................11 2.3.2 An Approach That Preserves Semantic Relativism......11 2.3.3 Extended ER Clustering for Integration process…………………….12 2.4 Summary…………………………………………………………………...13 Chapter 3 The Formalism 3.1 Motivation…………………………………………………………………..14 3.2 The EER Hierarchies…………………………………………………………..14 3.3 The Principles of EER Schema Integration…………………………………17 3.3.1 Attributes Level Equivalence………………………………………..18 3.3.2 Entity Level Equivalence……………………………………….21 3.3.3 Relationship Level Equivalence……………………………………..23 3.4 Summary…………………………………………………………………26 Chapter 4 A Methodology for Database Schema Integration 4.1 Phase I: Discover and resolve the database schema conflicts………………29 4.2 Phase II: Strategies for schema integration……………………………39 4.3 A Case Study………………………………………………………………..46 Chapter 5 System Implement 5.1 Overview…………………………………………………………………...50 5.2 The Meta-data……………………………………….50 5.3 The Schema Integration Tool……………………………………………….54 5.4 A Data Warehouse Case Study……………………………………………...59 5.4.1 Schema translation from various data models to EER model……….59 5.4.2 Schema integration of EER model into a global schema……………60 5.4.3 Data Conversion from existing databases to data warehouse……….63 5.5 Create Data Warehouse Application………………………………………..65 Chapter 6 Conclusion and Future Work 6.1 Conclusion………………………………………………………………….67 6.2 Future Work……………………………………………………………...68 References……………………………………………………………70 Appendix A. The Author's Publication List………………………………………….72

REFERENCES