CHAPTER 1

INTRODUCTION

Research in Datawarehousing and Datamining has produced important technologies for the design, management, and use of information systems for decision support (Adelman and Moss 2000) However, despite the continued success and maturing of the field, much work remains to be done. Given the wealth of models, terminology, and definitions, the first task is to review the most important models and their treatment of the basic concepts including the notions “dimension”, “fact”, “hierarchy”, “data cube”, and many more. The intent should be to evaluate existing models based on their expressiveness, flexibility, separation of modeling aspects from implementation aspects, etc (Arun 2013). With the knowledge gained, the concept to develop an overall and comprehensive conceptual model adapted to the users’ needs and abstracting from implementation aspects is required. The integrating task is to identify existing, tools and technologies and overview of their capabilities (Leticia et al. 2013), and to learn how they can be used to manipulate multi-dimensional data (e.g., cube, roll-up, drill-across).

A hybrid consolidated fact table with fractal property is desirable. This includes sales actuals with forecasts planning to achieve the sales target in to single fact table and offers the benefit of the analyzing the sales actuals versus forecast planning (Abdulrahman 2014). However, Fast consolidated fact tables can add burden to the ETL processing though ease the Business Intelligence analytic burden on the on the applications can be favorably considered for cross-process metric.

1.1. Problem Statement and Proposed Solution

The extraction of hidden predictive information from large databases is crucial with great potential to help organization focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions (Abraham et al. 2010). The automated, prospective analyses offered
by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

Most companies already collect and refine massive quantities of data. Data mining techniques can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources, and can be integrated with new products and systems in on-line. When implemented on high performance client/server or parallel processing computers (Ashadevi and Navaneetham 2010), data mining tools can be used to analyze massive databases to deliver satisfactory answers to questions such as, "Which clients are most likely to respond to their next promotional mailing, and why?"

This thesis provides an introduction to the basic technologies of data mining. Examples of profitable applications illustrate its relevance to today’s business environment as well as a basic description of how data warehouse architectures can evolve to deliver the value of data mining to end users.

1.2 RESEARCH IN DATAWAREHOUSE ANALYSIS AND PROPOSED IMPLEMENTATION

A fundamental concept of a data warehouse is the distinction between data and information. Data is composed of observable and recordable facts that are often found in operational or transactional systems (Adelman and Moss 2000). For example, at Rutgers, these systems include the registrar’s data on students (widely known as the SRDB), human resource and payroll databases, course scheduling data, and data on financial aid. In a data warehouse environment, data only comes to have value to end-users when it is organized and presented as information. Information is an integrated collection of facts and is used as the basis for decision making. For example, an academic unit needs to have diachronic information about its extent of instructional output of its different faculty members to gauge if it is becoming more or less reliant on part-time faculty. The data warehouse is that portion of an overall Architected Data Environment that
serves as the single integrated source of data for processing information. The data warehouse has specific characteristics that include the following:

An investigation of bus architecture provides analysis and maintaining the daily data in to enterprise data warehouse .Which supports for maintaining incremental data in to enterprise DW/BI system and partitions (Anshuman and Bharti 2011). The DW/ BI planning process into concurrent parts. The main contribution on Business intelligence business process at the time of delivering integration through permanent confirmed dimensions that’s reused process. Along with an architectural frame work, the decomposing program to encourage manageable agile implementations corresponding to the rows on the supports business matrix. It’s also available in the bus architecture support both relational and OLAP dimensional structures and technology platform independent (Yaokai et al. 2011).

The pointed cells of the matrix indicate whether a dimension column is associated with a given business process row (Ralph et al. 2008). The design task identify each row to test whether a candidate dimension is well-defined for the business process and also identify each column to see where a dimension should be conformed across multiple business processes. Besides the technical design considerations, the bus matrix is used as input to prioritize DW/BI projects with business management to implement one row of the matrix at a time.

A datawarehouse aggregate fact tables are simple numeric rollups of atomic fact table data built solely to accelerate query performance. These main aggregate fact tables should be synchronized with atomic fact tables the BI layer. So that BI tools can easily select needed aggregate activities at actual query time. This process, known as aggregate navigation, must be accessible so that every report writer, query tool, and BI application harvests the same performance benefits. A properly designed set of aggregates should behave like database indexes, can be used to tune the query performance. But this kind steps are not encouraged directly into the BI applications or business users. Aggregate fact tables contain alien keys to shrink conformed dimensions, as well as aggregated facts created by summing measures from multiple atomic fact tables. Finally,
aggregate OLAP cubes with summarized measures are frequently built analogous to relational aggregates, but the OLAP cubes are meant to be accessed directly by the business users.

**Subject-Oriented:** Information is presented according to specific subjects or areas of interest, not simply as computer files. Data is manipulated to provide information about a particular subject. For example, the SRDB is not simply made accessible to end-users, but is provided structure and organized according to the specific needs.

**Integrated:** A single source of information for and about understanding multiple areas of interest. The data warehouse provides one-stop shopping and contains information about a variety of subjects. Thus, the OIRAP data warehouse has information on students, faculty and staff, instructional workload, and student outcomes.

**Non-Volatile:** Stable information that doesn’t change each time an operational process is executed. Information is consistent regardless of when the warehouse is accessed.

**Time-Variant:** Containing a history of the subject, as well as current information. Historical information is an important component of a data warehouse (Adelman and Moss 2000).