Chapter 1

Introduction

There has been a rigorous effort by researchers to improve storage and retrieval efficiencies while storing increasing volume of data. The challenge became more complex with the advent of multimedia data particularly in increasing and large volume. This area has drawn focused attention of researchers to address the efficiency of storage, retrieval and analysis.

A Multimedia data warehouse is a collection of large volume of text, static digital images and images in motion, audio and video data. It is built by integrating large volume of multimedia data from multiple heterogeneous or homogeneous multimedia data sources. It is organized in a way to provide vital strategic information. Information hidden in the data, housed in a warehouse. A data warehouse is organized around major subject area(s). It is modeled in a way that its implementation supports pre-computation and fast access to summarized data. It needs to be designed in a way that query relevant data can be extracted and analyzed with speed and accuracy. Following figure depicts the generic architecture of multimedia data warehouse.

![Multimedia Data Warehouse]

Figure 1.1 Multimedia Data Warehouse
Multimedia databases store and retrieve large quantities of multimedia data. As the usage of multimedia data grows in organizations, the users require faster processing and analysis of multimedia data by way of innovative approaches. These innovative approaches in addition are needed to enhance value to analytic outcome.

Data warehouse is used to analyze data. The data modeling paradigm for a data warehouse must comply with requirements that are totally different from the data models in OLTP environments. Data warehouse uses multidimensional model. Multidimensional models consider dimensions as static entities as members of dimensions are computed once, and in a single way during the ETL step and facts as the dynamic part of data warehouse. However, this can restrict the analysis of data, particularly in the case of multimedia data. Indeed, multimedia data is a data, with various formats like text, image, video and audio. This data is generally described by different levels of features. Low level feature includes color, texture and shape. Mid level feature includes objects in data. High level feature includes objects, events, emotions, topic etc. Another type of feature is data-specific, i.e. features computed by specific processes applied on the multimedia data, such as the various measurements calculated on any image. Warehousing multimedia data require a specific extraction process which extracts domain specific features, tools for visualization and multimedia aggregation functions. Proposed work uses the different levels of features according to the requirement of application domain and analysis goal. The use of different levels of features and the computed features can improve characterization and the analysis of data.

The data model of the multimedia data warehouse must provide easy and speedy access to data as well as its analytical findings through an interface provided to the user. This approach is to provide user oriented interface that facilitates easy interaction with the interface by way of writing queries to satisfy information retrieval needs with higher level of query execution efficiency.

To efficiently store, access and analyze multimedia data in well organized way in a data warehouse, efficient approaches for storage, retrieval, and management of voluminous data are required. Storage and access of multimedia data is a critical issue for the overall system's performance and functionalities. These targets require design,
development and deployment of improved models and subsequent techniques to store, retrieve and analyze multimedia data.

The research study in the area of Multimedia data warehouse is rooted in traditional areas of multimedia analysis and Data warehouse that has its starting in late 1990s and has its continuity of development till date. The research outcome is evolving with more and more improved models, architectures and frameworks. Because of exponential increase in volume of data, the complexity of storage and retrieval mechanism also grows proportionately as a result. The research and development track has still opening for further improvement with regard to the current research outcomes. The challenges still prevail in storage, access and analysis of multimedia data that are housed in multimedia data warehouse which leads to targeted storage and retrieval efficiencies for volume of data rapidly increasing exponentially.

1.1 Motivation of research work

Majority of the data warehouse implementation had numeric and text data stored in the warehouse. Major work in building efficient data warehouse and access mechanism has been contributed in respect of data warehouse housing numbers and texts has been reported but comparatively less work has been reported for efficient storage, access and retrieval of multimedia data. The success of data warehouses in the corporate business world motivates to provide effective model that leads for the building of multimedia data warehouse. Multimedia data are voluminous, semi structured or unstructured in nature which requires efficient methods to integrate and manage them. The storage methodologies and access methodologies are needed to be optimized for multimedia data warehouse.

1.2 Objective of research work

The objective of this research is to address the issue of modeling and prototyping multimedia data warehouse which addresses the provision of efficient data storage and access mechanism. Multimedia data warehouse needs optimization in the storage structure and needs the provision of design for improvement in access latency occurring in the query processing.
The challenge from a data modeling and analysis point of view is to develop a model that stores multimedia data in such a way that they can effectively and accurately represent multimedia data and at the same time manages the complexity and richness of multimedia data. The challenge from an architectural framework point of view is to creating an efficient framework for storing, querying, and analyzing multimedia data.

1.3 Organization of Thesis

The organization of thesis is in the form of sequence of five chapters.

Chapter 1 specifies the research track as introduction. Chapter also explains the motivation and objective of the research.

Chapter 2 is devoted to extensive review of relevant literatures particularly the research papers of peer reviewed journals. Almost tentatively eighty seven research papers are reviewed specifying review of targeted individual research paper. The reviews are summarized in review findings specifying what work has been done in the undertaken research work and subsequently what challenges are needed to be solved to achieve targeted outcomes. The later half of the literature review findings shapes the problem formulation. The problem formulation truly represents the title of the research. The research problem statement has also been stated. The targeted outcomes specified in problem formulation along with the research issue to be addressed are to be achieved and for that proposed work is specified in chapter 4.

Chapter 3 specifies the foundation on which the proposed work begins with justification. Chapter also presents the technological background and performance parameters.

Chapter 4 provides the proposed architectural framework that is applied to three dimensions: biometrics image data, geographic image data and e-learning video data. The proposed framework implementation applied in each considered dimension has its process divided into three phases. These phases are: Multimedia data extraction and integration phase, Multimedia data modeling and warehousing phase and
Multimedia data analysis phase. In case of building multimedia data warehouse, the first phase works on multimedia data. Multimedia data extraction and integration phase extracts and integrates different levels of features according to the business domain and analysis goal. Second phase works on multimedia data modeling and warehousing phase which supports the physical implementation of the logical data model by dimensional modeling. Third phase works on analysis phase which facilitates accesses and analysis of data. The similar implementation is extended to fit the proposed model based prototype for Biometric & Geographic image data and e-learning video data. The data of all the three dimensions are used in proposed prototype.

Chapter 5 describes the result analysis and conclusion with further possible extension of the proposed research work. The analysis of results carried out on data and subsequent results in three dimensions has led to the justification of research targets and the same is summarized in the conclusion.