

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2015

Modelling Architecture for Multimedia Data Warehouse

Mital Vora¹, Jelam Vora², Dr. N. N. Jani³

Assistant Professor, Department of Computer Science, T. N. Rao College of I.T., Rajkot, Gujarat, India¹

Assistant Professor, Department of Computer Science, T. N. Rao College of I.T., Rajkot, Gujarat, India²

Director, SKPIMCS - MCA, Gandhinagar, Gujarat, India³

ABSTRACT: Data Warehouse is an information system mainly used to support strategic decision. During last few years there is a need arise to manage multimedia data in decision making process in business industry which leads to build Multimedia data warehouse. Multimedia data warehouse is a collection of large volume of image, audio, video and text data. To efficiently store, access and analyse such data there is a need arise to manage these data. Data management includes the access and storage mechanisms that support the data warehouse. Storage and retrieval of multimedia data is a critical issue for the overall system's performance and functionality. Multimedia data warehouse must be studied in order to provide an efficient environment in which data can be efficiently stored, retrieved and analyzed. In this paper, we propose the architectural framework to build multimedia data warehouse with the aim to provide better performance. To achieve better storage, access and analysis performance certain techniques are incorporated. Storage efficiency is improved by using provided compression technique and partitioning method. Access and analysis efficiency is improved by representing multimedia data by multilevel features and by applying indexing technique.

KEYWORDS: Multimedia Data Warehouse, Multimedia Analysis, Data integration, Data processing, Dimensional Modelling

I. INTRODUCTION

Data warehouse is subject-oriented, integrated, non volatile and time variant collection of data that support in management's decision making process [7]. A Data warehouse built by integrating large amount of data from multiple heterogeneous sources and is organized in a way to provide vital strategic information. Data warehouse uses dimensional modelling to store large amount of integrated data. Multidimensional modelling uses star schema or snow flake schema to store data in warehouse. Data warehouse supports analytical reporting, ad hoc queries and decision making. Data warehouses used to store numeric and textual data for decision making process and most commercial applications are designed to operate with data warehouse of this nature. Majority of the Data warehouse systems helps in analysing numeric data. Much research work has been done to design data warehouse for storing, aggregating and summarizing these data. Data warehouse technology with numeric data is considered to be matured [3].

Multimedia data warehouse study is rooted in traditional areas of multimedia analysis and Data warehouse, which started in the late 1990s to early 2000s. In today's business scenario the type of data is not limited to numeric or textual data but it includes wide varieties of images, audio, video, maps etc. Till date, new models, architectures and framework have continued to emerged and proposed in the multimedia data warehouse research and development (R&D) community to efficiently store, access and process multimedia data in warehouse environment.

Multimedia data is widely used in the field of information in science, engineering, medicine, modern biology, geography, biometrics, weather forecast, digital libraries, manufacturing and retailing, art and entertainment, journalism, social sciences and distance learning. These data comprise of various formats like image, audio, video, text and signal data. To efficiently access and analyse such data there is a need arise to manage these data tremendously. Data management includes the storage and access mechanisms that support multimedia data warehouse. Storage and access of multimedia data is a critical issue for the overall system's performance and functionality. Hence deployment of new

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2015

techniques to store, retrieve and process the multimedia data is essential and imperative. There is much to do in regard to complex, multimedia data warehousing [6].

The focus of this paper is to address the issue of modelling multimedia data warehouse architectural framework which address the provision of efficient data storage and access mechanism. This needs optimizations in the storage structure and needs the provision of design for improvement in access latency occurring in the query processing. For the experimental purpose we have tested biometric face image data, geographic image data and video data for the proposed multimedia data warehouse model.

II. RELATED WORK

In the late 1990s, W Lee et al [13] developed multimedia data warehouse for EoD (Education on Demand) systems. The system is implemented for video data, the relevant shot is pre processed for its physical and semantic structure by providing appropriate indexing and retrieval characteristics. For data storage they use star schema model. [11] proposed multi-tier image data warehouse framework based on the OOAD and component based development and have not described modelling technique much. Researchers have built multimedia data warehouse which can analyse data coming from heterogeneous and distributed sources [12, 5]. [12] provides materialized views to use in the analysis of multimedia data. Multimedia data can be represented by different level of features of abstraction. Researchers have suggested multiversion multidimensional model [2, 3] which describes and stores cardiology ECG data with content based or description based descriptors. They also use aggregation functions for multimedia data that are integrated into the data warehouse and in the OLAP engine. The limitation of their work is efficiency of storage and optimization of processes. Because of the interoperability and flexibility of the XML, researchers also proposed model which is used to build XML based data warehouse [5]. [5] uses XML technology to build framework for web enabled multimedia data warehouse which supports content-based integration and retrieval of multimedia data, and manages changes of data sources efficiently in a distributed environment. [1,14] presented the model that uses semantic based data. [1] proposed hierarchical way to store semantic data, they use two repositories of metadata which describes data in hierarchical manner in XML files. The authors of [4] have built a data warehouse which has two ontologies, one for the specific business terms and one for the technical terms, specific to the aggregation and knowledge extraction tools. This requires a one-time collaboration between the business experts and data warehouse designers, to produce a mapping between two ontologies. For multimedia data storage and representation [14] designed visual cube which uses two types of dimension. One type of dimension is meta information dimension which stores meta information regarding image and other is visual dimension which stores data based on image visual features. They also propose the idea of dynamic aggregation selection. The concept of content server is also proposed [9], which stores content of multimedia data in content server. proposed Content Server will maintain the indexes for metadata of the stored multimedia content as well as will have a repository of multimedia content. In [8] presents content based image retrieval using dynamic indexing and guided search combined with data mining and data warehousing techniques. They have developed wavelet based scheme for multiple feature extraction and developed multimedia starflake schema for image data warehouse, which support multiple feature integration and dynamic image indexing.

III. PROPOSED ARCHITECTURAL FRAMEWORK FOR BUILDING MULTIMEDIA DATA WAREHOUSE

Proposed architectural framework is aimed to provide further enhancement of the already proposed system for building multimedia data warehouse. As the storage and retrieval of multimedia data efficiently is a critical issue for the overall system's performance and functionality, the proposed work uses existing compression technique and partitioning for storage efficiency and for efficient retrieval proposed work focuses on the representation of multimedia data, indexing and partitioning mechanism. The proposed architectural framework comprises of mainly three phases: These three phases are: Multimedia data extraction and Integration phase, Multimedia data modeling phase and Data access and analysis phase. Following figure shows the generic architecture for multimedia data warehouse. Multimedia data is extracted from the source system. These data are then processed to acquire different levels of feature. These data are then stored in data warehouse for data analytics and data retrieval.

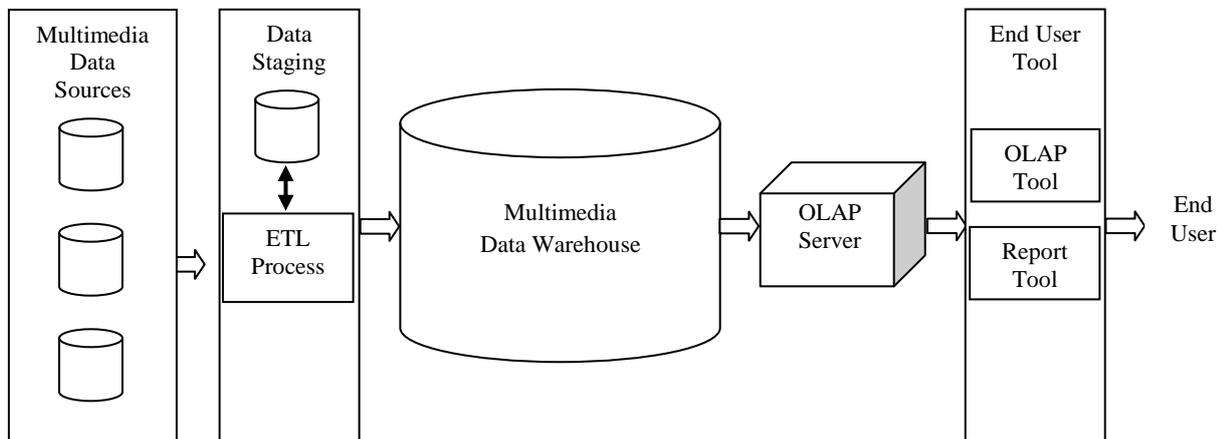


Fig. 1 Multimedia Data Warehouse Architecture

3.1 Multimedia data extraction and Integration phase

In this phase, multimedia data is extracted from the operational sources. The relevant characteristic of multimedia data should be extracted according to the analysis goal. Multimedia data is usually described by different levels of feature of abstraction. Low level features (color, texture, shape, etc) are widely used to describe multimedia data as these data can be extracted automatically using program. These features seldom represent the semantic content of the multimedia object. With regard to the data access and analysis, the high level semantic content is important. For this reason, this work has included low level features descriptor, high level semantic feature descriptor and calculated feature descriptors of multimedia content. Calculated features are features which can be extracted from multimedia data processing and can also be derived from the low level and high level features. for example, face nodal points and distance between major nodal points can be extracted from face image. These extracted features are known as calculated feature. After extracting multimedia data, low level feature extraction process takes place which is done automatically by using program and high level feature extraction process carried out manually. Some basic characteristics – file size, filename, author name, format, compression rate, duration for videos or sounds and resolution for images are also extracted automatically using program. After extracting these data, multimedia data is compressed using existing lossy technique. After preparing the data in extraction part data is ready to get loaded in data warehouse. Following figure shows the low level and high level feature extraction process.

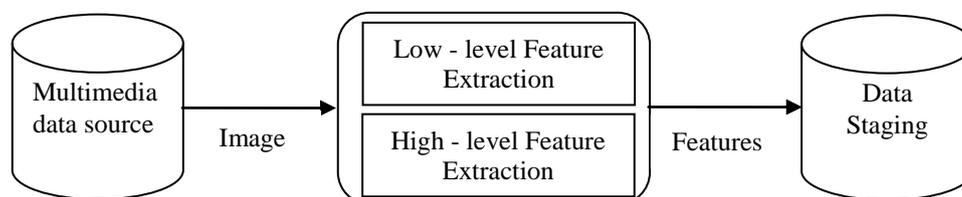


Fig. 2 Feature Extraction process

3.2 Multimedia data modeling phase

Data warehouse allows the data to be modeled in multidimensional way and to be observed from different perspective. Dimensional model of warehouse allows the creation of appropriate analysis contexts and the preparation of data for analysis. This requires to build multimedia data cubes on which OLAP operation are performed.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2015

To design multimedia data warehouse for multimedia data, logical and physical model has been developed. Logical dimensional model shows Main entities and their relationships in a logically sound manner, to serve as model for physical implementation. Physical dimensional model shows the actual representation of dimension and facts in data warehouse as they are implemented. Proposed architectural framework uses star schema technique of multidimensional model. Star schema contains one fact table that is surrounded by number of dimension tables. Facts are considered as dynamic part of warehouse and dimensions are considered as static entities because dimensions are computed once during the ETL process. Proposed schema includes a fact containing measures for data and number of dimensions is created for low level and high level features of multimedia data, data related to analyze multimedia data and a number of other data for the targeted application. After storing these data, indexes are created for feature attributes to speed up the query processing. For better storage, management and for better processing performance, partitioning technique is applied on the year based criteria.

3.4 Data access and analysis phase

In the data analysis phase, an end user tool has been created that accesses and analyzes multimedia data from the multimedia data warehouse. End user can analyze multimedia data by provided multiple features. Storage and access efficiency of multimedia data is studied.

III. EXPERIMENTAL RESULTS

To validate our proposed architectural framework for multimedia data warehouse, experiment has been performed on small scale OLAP environment. Data compression is applied only on multimedia data. To shows the query duration for simple, middle and complex level queries, two sample queries for each level is evaluated. Following figure shows the query performance for compressed image data without using indexing and partitioning approach.

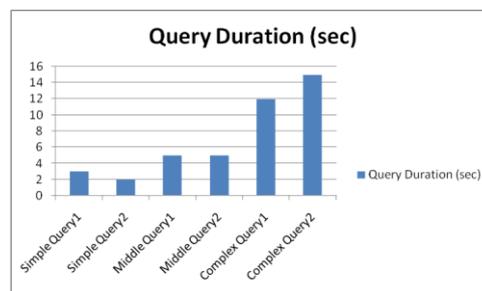


Fig. 3 Query duration for set of queries

To shows the query duration for simple, middle and complex level queries, the same two sample queries for each level is evaluated on small scale OLAP environment. Following figure shows the query performance for compressed image data with using indexing approach.

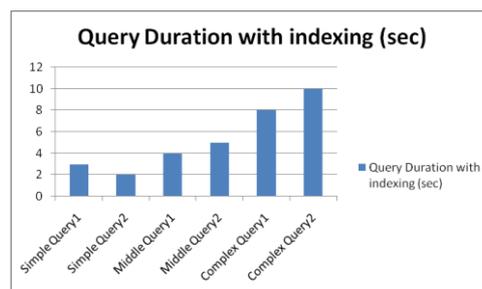


Fig. 4 Query duration for set of queries by using indexing

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2015

To shows the query duration for simple, middle and complex level queries, the same two sample queries for each level is evaluated on small scale OLAP environment. Following figure shows the query performance for compressed image data with using indexing and partitioning approach together.

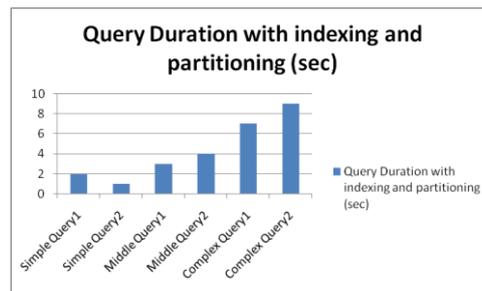


Fig. 5 Query duration for set of queries by using indexing and partitioning

Above figure shows the improved efficiency in data retrieval by using indexing and partitioning as compared to the data retrieval without using indexing and partitioning approach.

IV. CONCLUSION

Our paper has presented a systematic approach on building architectural framework for multimedia data warehouse in a generic way, so that the techniques can be applied to a wide range of multimedia data warehouse models and implementations. Implementation of these technique helps to build better multimedia model. Storage efficiency is improved by using provided compression and partitioning technique. Access and analysis efficiency is improved by representing multimedia data by multilevel features, by applying indexing technique and by using partitioning technique. By using this proposed approach, we not only obtain an efficient storage and processing of multimedia data warehouse but we also analyse multimedia data better.

REFERENCES

1. Andrei Vanea, Rodica Potolea, "A Hierarchical Semantically Enhanced Multimedia Data Warehouse" IEEE 2006.
2. Anne-Muriel Arigon, Anne Tchounikine, Maryvonne Miquel, "Handling Multiple Points of View in a Multimedia Data Warehouse", ACM, Vol. 2, No. 3, pp 199–218, August 2006.
3. Anne-Muriel Arigon, Maryvonne Miquel and Anne Tchounikine, "Multimedia data warehouses: a multiversion model and a medical application". Springer Science, 2007.
4. G. Xie, Y. Yang, S. Liu, Z. Qiu, Y. Pan, X. Zhou, EIAW: "Towards a Business-friendly Data Warehouse Using Semantic Web Technologies, The Semantic Web", 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007.
5. Hyon Hee Kim, Seung Soo Park, "Building a Web-Enabled Multimedia Data Warehouse", Springer Volume 2713, pp 594-600, 2003.
6. H. Mahboubi, J.C. Ralaivao, S. Loudcher, O. Boussaid, F. Bentayeb, J. Darmont, X-WACoDa: An XML-based approach for Warehousing and Analyzing Complex Data, Advances in Data Warehousing and Mining, IGI Publishing, 2009.
7. Inmon W., Hackathorn, R., "Using the data warehouse," Wiley-QED Publishing, Somerset, NJ, USA , 1994.
8. Jane You, Qin Li, Jinghua Wang, "Image retrieval by dynamic indexing and guided search", proceedings of the 8th IEEE international conference on cognitive informatics, 2009.
9. Meenakshi Srivastava, Dr. S. K. Singh, Dr. S. Q. Abbas, "An Architecture for Creation of Multimedia Data Warehouse" IJESIT, volume 2, issue 4, pp. 309-315, July 2013.
10. Mohd. Fraz, Ajay Indian, Hina Saxena, Saurabh Verma, "Improving Compression Efficiency of Data Warehouse," International Journal of Scientific & Engineering Research (IJSER), pp. 1575-1578 Volume 4, Issue7, Jul 2013.
11. Stephen T C Wong, Kent Soo Hoo Jr, Robert C Knowlton, Kenneth D Laxer, Xinhau Cao, Randall A Hawkins, William P Dillon, Ronald L Arenson, "Design and Applications of a Multimodality Image Data Warehouse Framework," Journal of the American Medical Informatics Association Volume 9 No. 3, pp. 239-254, May / Jun 2002.
12. Tania Cerquitelli, Genoveva Vargas Solar, Jose Luis Zechinelli Martini, "Building Multimedia Data Warehouse from Distributed Data", e-Gnosis Vol-2, 2004.
13. Wookey Lee, Yongkyu Kim, Yunsun Lee, Jinho Kim, "Developing multimedia data warehouse of education on-demand systems", pp. 942-945, IEEE 1999.
14. Xin Jin, Jiawei Han, Liangliang Cao, Jiebo Luo, Bolin Ding, Cindy Xide Lin, "Visual Cube and On-Line Analytical Processing of Images", ACM 2010.