UNSTRUCTURED MULTIDIMENSIONAL ARRAY MULTIMEDIA RETRIVAL MODEL BASED XML DATABASE

Dattatray V. Meshram[1], D. M. Dakhane[2]

1 ME C.S.E. (IV Semester), SIPNA's College of Engineering & Technology, Amravati (MS)   INDIA,
2 Associate Professor, Dept of CSE, SIPNA's College of Engineering & Technology, Amravati (MS) INDIA
dv.meshram@rediffmail.com, d.dakhane@gmail.com

Abstract
Unstructured Data derived from the thought of data warehouse, data cube and xml, this paper presents a new database structure model which organizes the unstructured data in a multidimensional data cube based on XML Database. In this data cube of XML, clustered data are stored in instance table. A leading data corresponding are stored in dimension table. The relational model is helpful to construct data model, but it lacks flexibility, now the new data model can complement the defect of relational model. When querying, a leading data is gained from dimension table of XML then receiving the unstructured data through XQuery. Thus we increase the flexibility of XML database.

Keywords: XML, multimedia, Multi-dimension, Database, Retrieval Model, multidimensional array, unstructured data.

--***---

I. INTRODUCTION

Development of by using XML, all kinds of data can be stored and retrieved by XML itself. Self-descriptiveness, the separation of content and display, across platforms and extension characterizes XML. XML provides many tools that the database must possess.

Recently the multimedia data increase surprisingly across network. People enjoy large multimedia information. Meanwhile, they also face following problems such as storage, management and retrievals. Content-based multimedia indexing and retrieval has more and more captured the attention of practitioners and researchers. Data warehouse is a kind of data collections which are subject oriented, integrated, and relatively nonvolatile and time variant. In recent years, it obtained rapid development in the domain of information management and database.

The application of data warehouse lies in its high efficiency in data management, suits to generous database application. So many researchers have already attempted to manage multimedia data with data warehouse. The widespread application of data warehouse is mainly due to its interior data organization structure-data cube. By data cube, one can analyze and process data from multi angles and multi-levels.

In the management of multimedia data, it contains exactly mass data. So, based on data cube technique of XML database and building proper multimedia data cube model, one can improve management efficiency for multimedia data or unstructured data.

XML introduction:

XML (Extensible Markup Language) is a subset of SGML (Standard Generalized Markup Language). First of all, XML is a Meta markup language by which the developers can define the tags according to their own requirements. Next, XML is a semantic structured language which describes the structure and the semantic meaning of a document. We can use the feature to describe high-semantic about multimedia.

The data described by XML has character of semi-structure. The traditional way of multimedia data management is challenged by XML. The traditional way of data management based on the relational database is completely different from the way of data seeking and information index described by XML. XML has the character of tree hierarchical structure and the seeking result of XML is more satisfactory than the traditional database.

Moreover, XML has the following advantages at aspect of data application:
1) The XML document is a pure text file which is not limited by the operating system and software platform;
2) XML has the function of self-descriptive meaning based on schema.
3) XML can not only describe the structural data, but also effectively describe the semi-structural data and even the unstructured data.

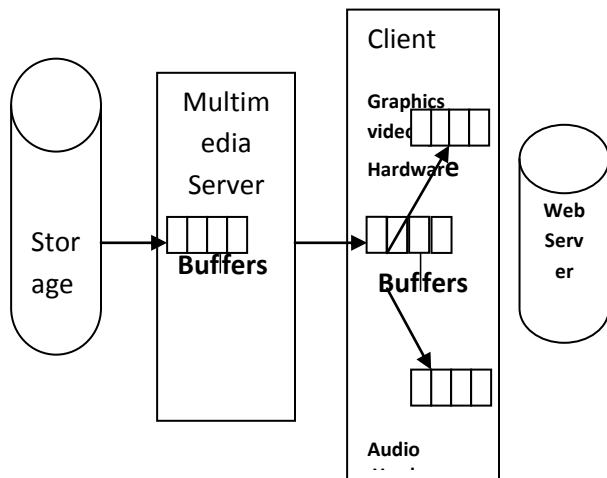## 2. LITERATURE REVIEW & RELATED WORK

Figure 1: Structure of Multimedia Database

Multi-dimensional data model is a logical concept, which is mainly to deal with how to query quickly and show with various points of view. By the concept of dimension, level and measurement, it considers the information as a conceptual cube.

Definition Data Cube Model.

Data Cube is a 4-tuple (D,M, F, AGG), where as
D is dimension set. D= { D1, D2,…, Dn }

Di = (DHi ,π) where represents a dimension. DHi represents an ordered set, composed of all the levels of dimension.

Figure 2: System Framework

In previous research, multi-dimensional database is mainly constructed based on relational database. The representation and storage of multi-dimensional data are implemented by relational structure. Multi-dimensional data are stored in two kinds of table: one is instance table, which is used to store data and dimension keywords; the other is dimension table. For every dimension, it needs at least a table to store corresponding descriptive information like the level of dimension and category number.

The dimension table and instance table are associated by primary key and external key, and then form a star schema. For the dimension with complicated levels, to avoid occupying excessive storage space, more than one table is used to describe data. This kind of extended star schema is called snowflake schema. This representation manner of multidimensional database can let user find out their preferred information simply, also it can save query time.

A star schema is composed of the instance table and dimension table. The instance table is the nucleus of the whole start structure, which is made up of external key and the measurement value that the user preferred.

The instance measurement value is the aggregated data that the user wants to find out and analyze in the application. The dimension table maintains the meta-data information used by a user in query like level relation, category attributes of member and so on. These are static, and can be used to analyze data. Every dimension table is linked by a primary key to a corresponding external key in the instance table.

The instance table has the following characteristics:
1) Contains a great deal of data column and the storage capacity can be up to T bits.
2) Includes mainly value information, with few texts and multimedia information.
3) With external key used to link dimension table.
4) Contains static data and aggregated data.

3. ANALYSIS OF PROBLEM

With this technique, we can maintain, store and retrieve the unstructured or multimedia multidimensional data with more flexible and efficient way. The relational model is helpful to construct data model, but it lacks flexibility, now the new data model can complement the defect of relational model. When querying, a leading data is gained from dimension table of XML, then receiving the multimedia data through XQuery.

The widespread application of data warehouse is mainly due to its interior data organization structure-data cube. By data cube, one can analyze and process data from multi angles and multi-levels. In the management of multimedia data, what one faces is exactly mass data.
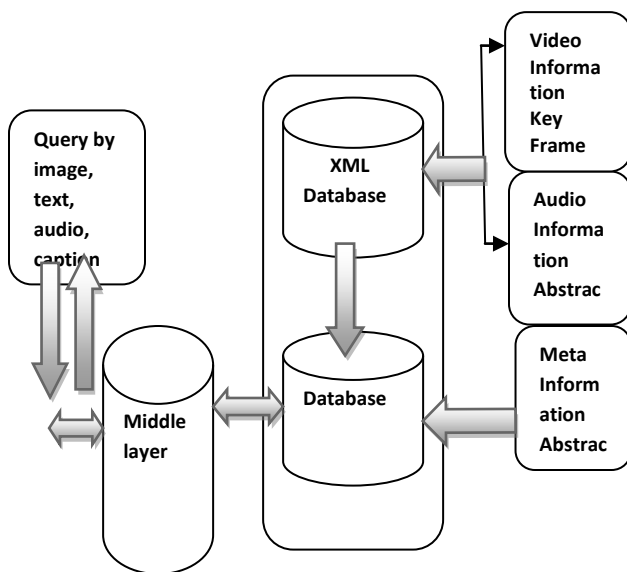
So, based on data cube technique of XML database and building proper multimedia data cube model, one can improve management efficiency for multimedia data. The advantage of data warehouse lies in its high efficiency in data management, suits to generous database application.

Multi-Dimensional Analysis is an Informational Analysis on data which takes into account many different relationships, each of which represents a dimension. For example, a CRM analyst may want to understand the relationships among customer by region, by quarter, by demographic distribution (income, education level, gender), by product. Multi-dimensional analysis will yield results for these complex relationships.

Multi-Dimensional Analysis is generally used in statistics, econometrics and other related fields and the results of this kind of analysis used in the different fields can be further applied to different fields like business enterprise. Multi-dimensional analysis actually is a process which groups data into two basic categories which are the data dimension category and the measurement category.

Two dimensional data sets are also called panel data in other disciplines. Logically, any two or more dimensional data sets could actually be considered as multidimensional data but the term multidimensional data tends to be applied on data sets only with three or more dimensions.

For instance, there are data sets used for forecasting which provide forecasts for various target periods and these are carried out by multiple forecasters made at multiple horizons. All three dimensions can provide for better information which can be gleaned from two dimensional panel data sets.

In a multidimensional the term dimension refers to a structural attribute of a data cube. The dimension is composed or related and hierarchical members. For example, the "Time" dimension may have the members like years, quarters, months, weeks, day, and hour and so on. In the same manner, the "Geography" dimension may have members like regions, countries, cities and so on.

A dimension member is an element of any given dimension just like in the example above where year like years, quarters, months and weeks are members of the "Time" dimension.

4. OBJECTIVES

Aims and objectives of this proposed work are summarized as follow represent a multimedia data cube model based on XML Database. The relational model has the defect of lacking flexibility for constructing data model. Organize multimedia data, the new data model based on XML database can construct more flexible data model.

Encapsulate new result into element or attribute, use many theoretical of tree to optimize the retrieval of the element or attribute by strong XQuery, which has rich expressions and algorithm. Improve the efficiency of retrieval about multi-dimension multimedia database.

5. PROPOSED WORKS

We considering the implementation of multi-dimensional multimedia model based on XML database consisting of five basic modules. The Java XML API would be used to read/write XML Files. The following would be the modules in my project,

XML Reader/Writer module:

This would be used to read/write to the XML File. Through this modules user can input the query and received the output from this module. This is the basic module for user interaction with XML database.

Multi-dimension Database module:

This would be a MySQL Implementation of the Multi-dimension database for multimedia content. Here all the input data would be stored and process as per the user request or query.

Database access module:

This module would be Java code files that use Type 4 driver for connection with the MySQL Database. These classes would be used to retrieve the data from the MySQL Database

Query module:

Here the user would be giving an input query to the system which would be processed to give an appropriately fetched data from the database

Results module:

All the results would be displayed in a proper multimedia format, in the results module

6. DESIRED IMPLICATIONS

Multimedia data is composed of text, image, graphic, video, audio and so on. Among them video contains most abundant information. So in this paper, the research is focused on video, and other media formats are left to future.

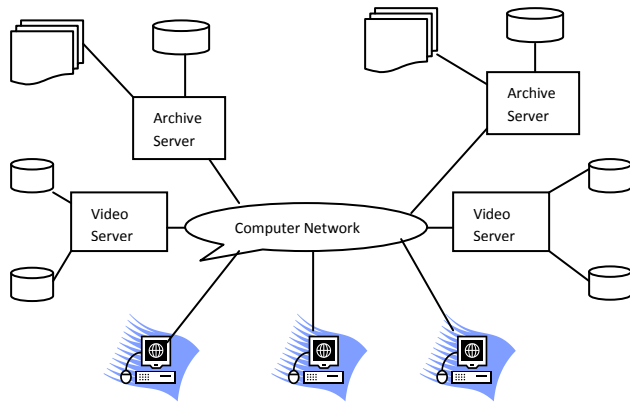Tertiary Storage Metadata Database Tertiary Storage   Metadata Database

Figure 3: 2 Star structure of a multimedia database

A. Video data:

Video data is a kind of special data used to store image sequence. The information it delivered is much huger than that static text and image does. Generally speaking it has the following characteristics:

1) Video data has relatively high information resolution, containing much more information.
2) Video data is non-structure data and the relation of which is complicated. There are both time relation and space relation between video segments and objects in a video segment. Moreover, the video data is relative to some special domain.
3) Unlike character and numeral for which the interpretation is relatively objective, the interpretation for video data is various, and usually has some human subjective description in it. When search some information, the query method based on keywords originally used in traditional database could not be used more, a new similarity query method should be used.

B. Model for dimension and measurement of multi-media data cube:

According to the concept of data cube described in section 2, multimedia data cube can be defined. The key frame, caption, text and time can be considered as the dimension of a multimedia data cube.

Both dimension and measurement of data cube have space information and non-space information, so there are three types of dimension in data cube as follows.

1) Non-space dimension. There is only non-space data in it, for example time dimension.
 2) Space—non-space dimensions. In this dimension, the lower level data is space data. By summarized, the data over some level will be non-space data.
3) Space—space dimension. The whole level data in this dimension are space dimension.

CONCLUSIONS

Through this paper, we present a multimedia data cube model based on XML Database. The relational model has the defect of lacking flexibility for constructing data model. With the data cube to organize multimedia data, the new data model based on XML database can construct more flexible data model. We give the implementation by using xml.

Meanwhile, according to the characteristic of XML, we can encapsulate new result into element or attribute, use many theoretical of tree to optimize the retrieval of the element or attribute by strong XQuery, which has rich expressions and algorithm. We discussed only the implementation model of multi-dimension multimedia database based on XML database. In the future, we'll give several implementation methods of XQuery algorithm to improve the efficiency of retrieval about multi-dimension multimedia database.

ACKNOWLEDGEMENTS

REFERENCES

[1] Djeraba C.," Content based multimedia indexing and Retrieval", IEEE Multimedia Volume 9 No. 2 page 18-22, 2002.
[2] Nicu Sebe, Qi Tian, "Personalized Multimedia Retrieval: The New Trend?". ACM Multimedia Information Retrieval Workshop, Augsburg, Germany, September, page 299-306, 2007.
[3] M. Lew, N. Sebe, C. Djeraba, R. Jain, "Content-based Multimedia Information Retrieval: State of the Art and Challenges", ACM Transactions on Multimedia Computing, communication, and Applications, Volume 2, No. 1, page 1-19, February 2006.
[4] Mathias Jarks, Fundamental of Data Warehouse (Second Edition), the Association for Computing Machinery, 2001.
[5] Hyon Hee Kim, Seung Soo Park, "Building a Web-Enabled Multimedia Data Warehouse", HIS 2003, LNCS 2713, Page 594-600.
[6] Anne-Muriel Agigon, Anne Tchounikine and Maryvonne Miquel, "Handling Multiple point of view in Multimedia Data Warehouse", ACM Transaction on Multimedia Computing, Communications and Applications, Volume 2 No.3, page 19-21, August 2006.
[7] A. Maniatis, P. Vassiliadis, S. Skiadopoulos, Y. Vassilioul, "CPM: A Cube Presentation model for OLAP", DaWaK, LNCS 2737, page 4-13,2003.

[8]   Athena Vakali B Catania, A. Maddalena, "XML Data Stores: Emerging Practices", IEEE Internet Computing Volume 9(2), page 62-69, March 2005.

[9]   Jonathan Rabie Red Hat "XML Processing and Data Integration with XQuery", IEEE Internet Computing Volume 11(4), page 62-67, July 2007.

[10]   C. Mathis, T. Harder, M. Huastein. Locking-Aware Structural Join Operators for XML Query Processing. ACM SIGMOD, 2006.

[11]   R. Goldman, J. Widom. DataGuides: Enabling Query Formulation and Optimization in Semistructured Databases. Proceedings of the International Conference on Data Engineering (ICDE), 2002.

[12]   R. Bordawekar, O. Shmueli. Flexible Workload-Aware Clustering of XML Documents. Database and XML Technologies, Second International XML Database Symposium, XSym, 2004.