A Survey on Heterogeneous Data Exchange Using XML

Kartikey Borle[1], Amit Bendarkar[2], Rajan Bhirud[3], Aditya Bambole[4] , Prof. Rachna Somkunwar[5]

[1]*Department of Computer Engineering, DIT, Pimpri, Pune*

[2]*Department of Computer Engineering, DIT, Pimpri, Pune*

[3]*Department of Computer Engineering, DIT, Pimpri, Pune*

[4]*Department of Computer Engineering, DIT, Pimpri, Pune*

[5] *Assistant Professor, Department of Computer Engineering, DIT, Pimpri, Pune*

---***---

Abstract - *XML is a popular and highly used markup language which is used because of its simplicity, generality, and usability. As a result, it is widely used in modern computing. Integrating multiple data sources has become important. This paper explore heterogeneous data sharing to acquire transparent operation and continuous heterogeneous data integration, and also focus on the basic properties of XML data exchange based on available heterogeneous database integration systems.*

Key Words: **Data export, Data import, Heterogeneous databases, XML, XML parsing**

1. INTRODUCTION

Nowadays, internet is growing rapidly and there is a wide use of database system, many numbers of independent databases has been planned and created in distributed environment. All these miscellaneous or different databases are different in structures. To eliminate the conflict between heterogeneous data, data exchange takes place in heterogeneous databases. However, in certain cases there is necessity of information transmission in many application systems. Therefore, in heterogeneous databases system, data exchange between them becomes an important point of discussion. Big amount of semi-structured and structured data is stored with the help of XML technique. XML technique is also used for exchanging data from source to destination. Whenever large amount of information is stored, data transformation and data manipulation is necessary. The database that are maintained here are mostly heterogeneous and distributed. XML is designed to represent data in the form of tags (elements); it allows information to be expressed in such ways that match better for business. It brings powerful properties such as flexibility, heterogeneity, and extensibility to information modelling. That is the reason, XML has become a standard data format widely used for data transmission over the Internet as well as in these organizations and a common language. This leads to an increasing need for exchanging and integrating the heterogonous XML data sources and schemas between different application systems. Therefore, data exchange between different and heterogeneous databases becomes

very crucial point to work. Different languages such as Document Type Definition (DTD), eXtensible Stylesheet Language Transformations (XSLT). XML documents can also be restructured using Document Type Definition (DTD). For creating a mediate architecture of XML schemas, eXtensible Stylesheet Language Transformations (XSLT) can be used. XML (eXtensible Marked Language) provides solution for data integration of information. As there are large, complex and different types of queries, a fixed mapping is not possible to work. An XML document contains a root element following the nested elements. Elements can be either attributes or sub-elements.

2. RELATED WORK

The purpose of data exchange is to realize data share among systems. The aim of XML tag is to spot every XML element unambiguously and to show constructive relationship. The constructive symbol called XML tag allotted for every node in XML document is extra knowledge to the subject. Post Order Numbering (PON) mentioned follows post order traversal requires a lot of space for tag that is a significant disadvantage. A famous coding method for trees known as Dewey coding is mentioned. M-IBSL (Modified Improved Binary String Labeling) is explained within that partition XML DOM nodes into sensitive and public nodes[1]. The content of public nodes is additionally sent at the same time when the encrypted sensitive nodes is being send, which is an extra information needed to reconstruct the XML document at the buyer. Therefore, leakage of extra information is not prevented. XML-based information exchange has gained researchers attention with development of internet technologies.

Work to be done by researchers here was to study the conversion of XML document to relative information and transformation methods between relational models and XML. In [2], researchers studied the related theories to XML and relative models and proposed a data-exchange model in e-commerce system. Data exchange between different ERP systems is needed to be designed and implement in an interactive way. They proposed a tool to map elements of one schema to another. They studied linguistic conflicts in heterogeneous database systems and divided them in

subsequent different types and mentioned different causes for them.

Different RDBMS suppliers like Sybase, Db2, SQL and Oracle have their own tools for data exchange between their databases and XML. XML processing has become more easy and efficient because of this.

In comparison to this, native XML dbs store documents in XML format, however their capacity of processing data is limited. Some skilled middleware programs are used for exchanging data between different XML documents and different XML files using the procedures in respective applications.

Application-oriented data integration systems are designed and developed at the same time. Such systems are quite common in tiny enterprise information systems. They are all used underneath specific application atmosphere, for instance, HR management. In e-commerce environment, management information system(MIS) is back-end for enterprises daily operation. User is shown a website for browsing and buying products with different facilities for smooth operation, which is the front-end in this case. Therefore, they decided to visualize and design a system for exchange of data from databases of front end to back end and vice versa.

XML is developed by W3C in 1998. It is a universal language standard. It is the simplest SGML subgroup and was specifically made for internet purpose. The most important strength is that it has robust capability of data information and transmission. XML is layered architecture. Using Document Type Definition (DTD) or Schema makes XML document more structured for the reason of having robust information capability in order that it's simpler to retrieve information from XML document. XML document viewing automation. W3C recommends Extensible System Language (XSL) style sheet standard. To retrieve the attributes or elements in the XML document, it will be taking total utilization of the matching of pattern methodology, and so developing switching rules for clearly defined item. In the meantime, new attributes or elements can be added in keeping with the actual condition, and developing programs to control them to fulfill that conditions. There are two important communication technology for the XML document: SAX (Simple API for XML) and DOM (document object module). Taking XML as intermediary file, and inspect the data within the XML document using SAX and DOM. The XML parser supported on SAX depends upon event-driven interface while the one supported upon DOM changes data structure in XML documents in a group of item tree structure[3].

Parsing XML information, the important tools are DOM and SAX. DOM is appropriate for computing complex data structures. SAX Parser parses the XML record line by line and triggers occasions once it experiences opening label, shutting tag or character in XML record.

Nowadays, the XML technology is being used widely in dissimilar database information turning, and have become a public technology to as certain information trading standards

within the actual application. The information conversion model that is set depending on XML technology as the functional representation of the general information gives a convenient way to information unification and turning between relational databases and different representation of information. There are many database by-products relating to XML export and import tools, like Oracle, Sybase, Microsoft and IBM, etc. with extra tools to remodel XML in database tables[3].
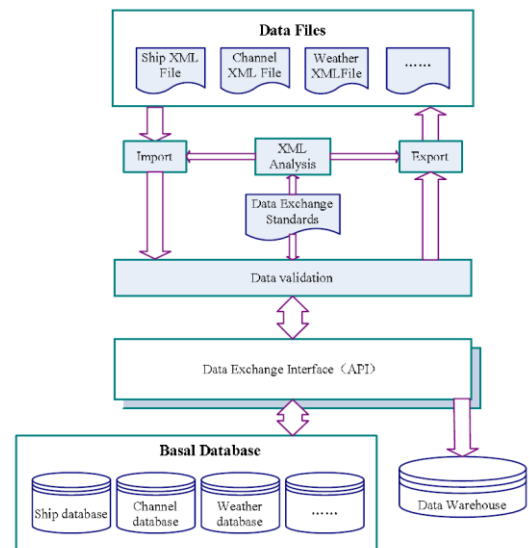

Fig -1: XML based data exchange structure [3]

At whatever point huge measure of information from mechanical applications required to be put away, change and control of data can be a confounded issue. Vast measure of information consolidated with exclusive information positions, produced from very surprising information sources, makes an incredible test for information taking care of. To deal with this issue, a completely incorporated arrangement that grants stockpiling and change of subjective information configurations is presented. The taking after Fig.2, delineates how before utilizing information is kept and changed over [4].
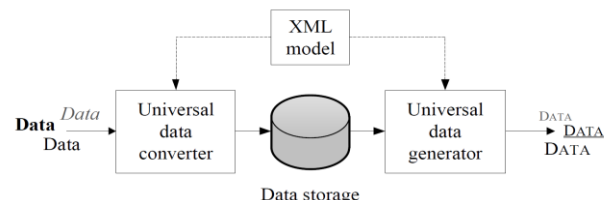

Fig -2: Data storage model

2.1 Heterogeneous Database Data Exchange Build Up On The Xml Technology

In [5] author says Extensible Mark-up Language has the marvelous information execution capacity, solid capacity of expandability and self-portrayal, while the comprehension of the partition of information and the diverse assortment of

expression. These attributes discover that XML are frequently utilized as medium of data trade, so acknowledging heterogeneous framework information trade.

The investigation of data trade of heterogeneous database in view of XML is to truly concentrate the common change between social database construction and the XML demonstrate. Along these lines, setting up a mapping connection between the database model and XML record demonstrate and comprehend information transformation between them.

2.2 Document of XML

A XML database is built from one or numerous XML file. The database is developed out of large giant XML reports. When utilizing the term record, we tend to mean the input XML document. As in the accompanying Fig.3, it somehow looks like Hypertext mark-up Language (HTML) aside from that XML has extra strict manage regarding opening and shutting labels. It means that once a tag is opened (e.g. <title>), it needs to be closed (using </title>).

```
<book>
   <title>XML</title>
    <chapter>
      <section>
         <head>introduction</head>
            <subsection>data model</subsection>
      </section>
      <section>
         <head>tree pattern</head>
            <subsection>twig</subsection>
      </section>
      <section>
         <head>search</head>
      </section>
      <author>Lu</author>
      <year>2013</year>
    </chapter>
</book>
```

Fig -3: XML document example

3. UNDERSTANDING THE XML

Supporting JDOM to inspect, produce and serialize the XML document there is adoption of Pure Java technology. JDOM is developed by Jason Hunter and Brett McLaughlin in 2000. It is open source. It combined with DOM and SAX to overcome some insufficiency. To attain data unification, the XML inspection contains three point: data importation, data exportation and transmitting data.

3.1 Data Exportation

For exporting data, firstly the required data is to be extracted from the source database in accordance with the data rules, and assembling it according to the logical structure requirements of original application.

3.2 Data Importation

Keeping in mind the need for the maintenance, three steps are required to save the XML data files retrieved from dissimilar system into our system are: Validating XML File, assembling data and store in database.

3.2.1 Validating XML File

During data transfer process, the information can be artificially remodel, thus testing the XML data file is more important. The test method is divided into two parts: a) Making the schema file for data standards to validate the data b) To write program for verification of unverified data by DTD.

Failure in validation means, with the data standard, the arrangement of information is not compatible. After the completion of this step, the specific explanation will be given. The assembling of data will go on if test is passed.

3.2.2 Assembling Data

In assembling process, there is data extraction and transformation. It is mainly about the establishing of the mapping between database table and XML data, and transferring data in the expected formatting of destination db.

3.2.3 Save in database

After assembling data supporting the target database standards, use the hibernateDao.updateObject or hibernateDao.saveObject to update existed data or save the new data.

3.3 Transmitting Data

If the source and destination dbs exist on the same machine (here server), data transmission will be abandoned. However, if both of them are running on different servers, it is necessary to focus on way to deal with the data conveyance issue. In the process of data transmission, some data file might be missing or there can be damage problem, thus some security policies can be adopted to ensure the safe transmission avoiding other users copying or accessing data, such as key management, digital signature and file encryption. Thus, ensuring that data cannot be modified throughout the data conveyance process and providing some facility that can identify content and sender of the document.

4. CONCLUSIONS

To conceive information trade and transmission ability, XML is the most vital and frequently used information trade innovation. The most effective method to utilize information trade procedures in heterogeneous database framework is

essentially investigated in this paper. In this paper, we mainly studied data exchange between different types of dbs via an exchange technology called as XML. Different ways of exporting and importing the information were studied in this paper, which involves different steps like in intermediary file, assembling, validating, and transmitting the data. We have also focused and discussed on SAX parser and DOM parser and their potentiality. We have come to know from aggregate usage procedure of the completely dissimilar type of database trade that utilizing XML as an information trading medium we can well accomplish the information transformation between the different databases, consequently accomplishing the reason for data trade and information sharing.

REFERENCES

[1] Varma,Ravi, and G. Venkata Rami Reddy. "A Review of XML Parallel Parsing Techniques." International Journal of Computer Science and Mobile Computing 3.4 (2014).

[2] Wu, Mingli, and Yebai Li. "Investigations on XML-based Data Exchange between Heterogeneous Databases." Web Information Systems and Applications Conference (WISA), 2012 Ninth. IEEE, 2012.

[3] Li, Taoying, Yan Chen, and Yingying Yu. "Research on XML Technology in Data Integration of Maritime." Computer Science & Service System (CSSS), 2012 International Conference on. IEEE, 2012.

[4] Unterweger, Andreas, et al. "A generic model for universal data storage and conversion." Applied Electrical Engineering and Computing Technologies (AEECT), 2011 IEEE Jordan Conference on. IEEE, 2011.

[5] Zhao, Jia Hua. "Heterogeneous Data Exchange Technology Based on XML." Applied Mechanics and Materials. Vol. 347. Trans Tech Publications, 2013.

[6] Haw, Su-Cheng, Chien-Sing Lee, and Hau-Lee Tong. "Effective XML data storage and distributed query retrieval system." Future Computer and Communication (ICFCC), 2010 2nd International Conference on. Vol. 3. IEEE, 2010.

[7] Xu, Honghua, et al. "A schema of data exchange for heterogeneous data." Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), 2011 2nd International Conference on. IEEE, 2011.

[8] Muralidhar, A., and V. Pattabiraman. "An Efficient Association Rule Based Clustering of XML Documents." Procedia Computer Science 50 (2015): 401-407.

[9] Rochlani, Yogesh R., and A. R. Itkikar. "Integrating heterogeneous data sources using XML mediator." International journal of computer science and network 3 (2012).

[10] Kharade, Sonam S., and Chetan J. Awati. "Techniques of Mapping XML Data: A Survey." International Journal 3.5 (2015).