Relational Database System and Native XML Database System
for Publishing Production

M.Sc Marinko Žagar
Školska knjiga d.d, Information technology department
marinko.zagar@skolskaknjiga.hr
Ph.D Klaudio Pap
Faculty of Graphics Art
klaudio.pap@zg.t-com.hr
Ph.D Mario Barišić
Vjesnik d.d.
mario.barisic@vjesnik.hr

Abstract. *Databases in publishing production processes are seldom used for storing contents or image material used for book making, regardless of the production form (printing, WEB, CD). The possibilities provided by databases are those that make the necessary conditions for implementing databases in the center of publishing production processes. By applying databases as the center of publishing production processes digitalization of all production phases is possible regardless of the end product type. The question that must be answered before bringing a decision on database implementation is what database should be chosen as the center of the publishing production system? Should one choose the RDBMS database system or the Native XML database? This paper aims at giving an answer to this question while arguing on the implementation of the first or second system based on an actual example.*

Keywords. Publishing, RDBMS, Native XML Database

1. Introduction

Entire publishing production workflow imply data transfer between phases which include author, editor, graphical art editor, and printinh house. Besides listed participants, texts, images and other materials is checked by lector, kolector, art director etc. Market demands, technological opportunities, and competition enforce publisher to publish beside book in printing edition also CD edition, WEB pages and sometimes WAP informations.

It is seems that business processes hide big redundancy in different resources. How it is likely to organize the total workflow in a way that there is no multiplications of resources independently because of the kind of the final publishing product? The answer is in use databases. Use of databases in publishing-printing production process for the preparing of contents of titles still is rarity, especially when is in consideration of the book publishing. Choice of database for the central depository of contents of titles depends on the kind and way in which we wont to save data in the database.

2. XML and Relational Databases

When speak about publishing printing business and information system, we see that this is a system in which is present different information and production technology in different segments of system. The large problem with which publishers and printing houses meet is the choice of technology which is able to integrate all technologies used in different business processes through the one information system. This is a information system which will give oportunity to control business system in all preproduction, production and postproduction phases of book production. Technology which intrudes as the solution for the uniting of all business processes in the unique information system, regardless about which device is used, is XML technology. XML technology is the communication bridge between different information technologies, and in the beginning of development, XML has not been imagined as database or the tool for the work with the database.

Question is: How data which has been formatted in XML form to save in the database, in order to easily and simply search and make to be secure and accessible to all parts of information system?
Information systems have basically the relational database in whom save data via determined applications. Structured data from the database use applications which are not based on XML technology. For example, If we want to save data which exist on Web sites of publisher in XML form in the database or search through data on Web, for this we would need XML database or some Content managment program. During the choice of databases it is important to determine that databases will be used for the saving of documents or data. XML will be to serve only as the transportation media between the database and applications or XML document will be saved as a whole in the database. Question is very important because data oriented documents (Data-centric document) have many common characteristics,

while for documents who are not data oriented than important as a whole (Document-centric document) have not common characteristics.

Data oriented documents use XMLHave been designed for the transfer between devices and alone XML structure in this documents is not important. Databases and XML basically are incompatible platforms, they are created with the complete different purposePopularity of XML and growing use, because of his simplicity, produce the need for the connecting and integration these platforms. Understanding differences between databases and XML is necessary bacause we can use good sides of every technology during integrations.

3. Native XML Database (NXD)

All known databases has supplements (extensions) for transporting data from XML format. Because every conversion and transporting of data between different databases and systems requires the writing of additional code, routines and controls, and connecting with relational databases not always is possibly, most naturally in XML technological surroundings is saving data in XML database. Native XML databases imposes as the technological solution especially for organizations whose business is closely related with Internet. Publishing-printing business systems use Internet as a part of infrastructure and use native XML databases will be soon the imperative.

Name Native come from the marketing campaigns of German company Software AG for her Nativ XML base Tamino. Name is later enter in the usage without the formal technical definition.

Native XML databases are databases specially designed for the saving XML documents. As well as other databases, XML database supports transactions, security, multiuser access to the documents, queries etc XML databases largest use finds for the preparing of documents which are not data orientated. They are important as a whole (Document-centric document). How XML databases support queries, for example, it is possibly with simple query find all thirdly paragraphs in all documents which starts with italic letters. Such queries are not simple to make in standard SQL language.

Native XML database define the logical model of XML document for saving and searching of data towards this model. As the minimum, model must contain the elements, attributes, PCDATA and structure of document. Such as the record is basic unit in relational databases, the XML document is basic units in XML database. This fact does not eliminate the possibility of reach of logical part of XML document like the individual element, such as in relational databases is case with the reach of any column from raws. Towards the architecture Native XML databases are divided on two categories: text XML databases and model XML databases.

3.1. Text XML databases

Text oriented XML databases are databases who XML document save in the text format. This can be file in file system, BLOB in the relational database or some own format from manufacturer.

3.2. Model XML databases

Instead of the saving XML documents in the text format, model XML database builds the internal object model from the document and saves this model. Some databases save the model in the relational form, and some in the object-oriented form.

Characteristics of Native XML databases:

- Document collection – Native XML databases can save documents in collections, like relational databases which save data in tables.

- Queri language – Native XML databases supports query language, and most popular is XPath with the extension for query on many XML documents and XQuery.

- Refreshing of data (the update) and deletion – standard characteristic of the nativ XML databases. Two standard languages for the refresing of data, besides the language which produce manufacturers of the some XML databases, are XUpdate and XQuery. XUpdate is based on Xpath language.

- Transactions, locking and concurency All native XML databases support transactions (rollback), while most often locks are on the level of whole document. Therefore the multiuser work may be slowed down.

- Program interface towards the database – All native XML databases have the program interface towards the database. Besides manufacturers of databases which offers its interface exist products independent of database.

- Round-tripping – The very important characteristic in view of documents which has been orientated towards contents (Document-centric document). It is supported in all Native XML databases, and the sense is that saved document can be opened in the original form. Text orientated native XML databases support round-tripp in the original form, while the model orientated native XML databases support this on the level of document model.

- Connecting with distant data – Some native XML databases can import distant data with XML document. Usually these data have been reached via ODBC from relational databases.

- Indexing – Native XML databases support three types of indexes. Index on the value level (value index), structural index i full-text index. Value index is placed on the value of element or the attribute. Structural index is placed on tag level. Full-text index is applied for the search of determined text regardless of role in the document.

4. Relational Databases

Information systems of the publishing business systems are organized around relational databases. Data about transactions are collecting from all parts of business system via applications designed for that purpose and save in the database for the examination, reporting or analysing and for managing business system. The largest advantage relational databases in respect to XML databases is just the large number of people which know relational databases.

- The physical space – Database controls the secure accommodation of data
- Security - Security is designed on the database level, integrated in the operating system or the combination of the quoted
- Data model - Relational model, presented through tables with raws and columns
- Data types – Relational databases support all data types
- Data relations - Definitions of colons can make correlation between the colon towards DDL rul
- Scheme - relational database scheme consists data about all connections between data.
- The referential integrity - The typical functionality of RDBMS by the retaining and propagating of exchanges in connected columns.
- Supplement validation - Use of triggers and stored procedures spends the future validation
- Query language - Majorities of commercial RDBMS supports standard SQL language. Many manufacturers have one's supplements for the purpose of enlargements of functionalities.
- Manipulation of querie result - SQL has parameters which can group, sort and determine how to organize query results.
- Query on many databases - Some RDBMS enable queries on many different databases.
- Indexing - Index system in RDBMS is very functional and it is able to set on the columns level.-
- Transactions - Majorities of RDBMS supports transactions and rollback and tsupport ACID characteristics.
- Multiuser work on locked data - RDBMS systems insure the lock-up of page and raw.
- Dependence of the platform for the data storage - Relational databases are commercial products and different versions work for different types of machines and operating systems. Data transfer between systems is not problem.
- Scheme dependence - Database is dependent on the database scheme, and schemes is the component of database.
- Reuse schemes - Scheme is the component of database and is not simply the repeated use by other databases without suitable tools.
- Nesting Columns from tables of database cannot be used for the real nesting.

5. Saving XML documents in relational databases

During development of XML technologies, all known manufacturers of system for the database management have made supplements for XML. With this moment appears the real possibility that databases can be set in the centre of publishing business processes. Integration of XML databases and relational databases is possible in a few different ways, which is dependent on the result wants to be achieved. Saving of XML data in database depends on the way of saving XML document.

5.1. Saving integral XML document in the database

This method simply save total XML document in the determined table in the database. Worry for the validation and processing documents takes over the application. Adding of new table in the database does not disrupt the defined database model and therefore saving XML document in this way does not influence on the relational database model.
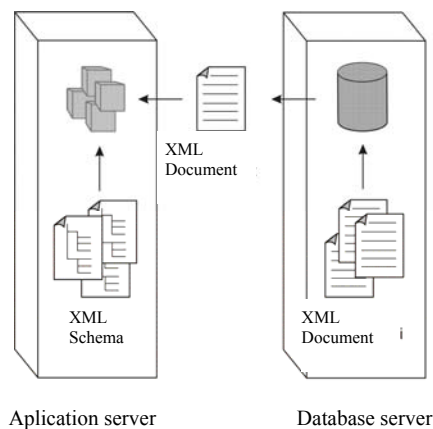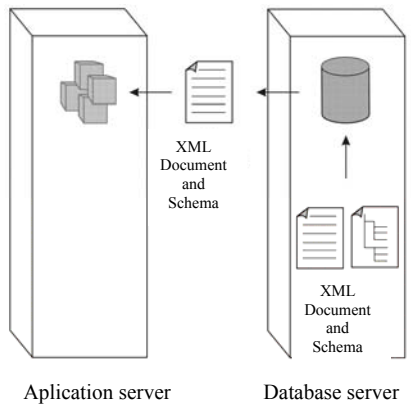

Figure 1. Saving integral XML document in the database

Suitable XML scheme in the quoted model is saved out of database and application takes care of the dynamic connecting with XML document.

5.2. Saving XML document and XML scheme in the database

XML scheme is saved in the special table in the database and with simple connection one to many is table connected with the XML document. This way savings XML documents is appropriate for the saving of new data emerged from applications which are not connected with the existing system. Saving XML documents and XML schemes in separate tables do not disrupt existing relational database model. Problem is searching which is based on the string search what is very slowly, and also distinguishing of tags from the data.

The additional problem is the examination of XML document and capture of CPU time when is XML scheme deposited in the database. The every calling XML document calls XML scheme.

Figure 2. Saving XML document and XML scheme in the database

5.3. Saving XML document segments

This model the same as previous will not disrupts existing relational data model. Instead of thesaving integral XML document, it is saved XML segments as separate logical entities in the special table. Calling of document call each segment and connect in the document via the application XML scheme is called on the application level.
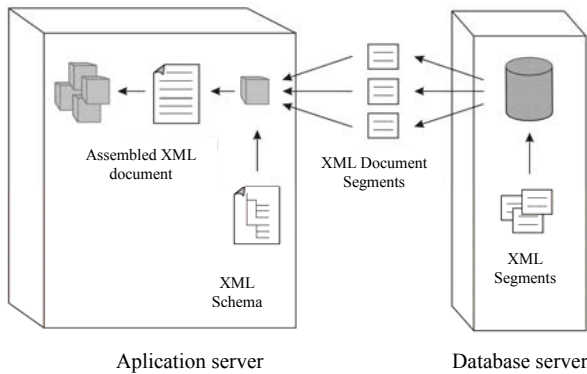

Figure 3. Saving XML document segments

5.4. Saving XML segments and XML scheme segments

Consistently the quoted architecture, XML scheme can be divided in logical parts. Divided scheme is saved in segments in the database in the special table and connects with the table in which is saved XML document.

Advantage of quoted model is in speed of data transfer, and the existing data model has not been disrupted.

Negative effect is the complexity of data models, when XML document and XML scheme are divided on logical parts.
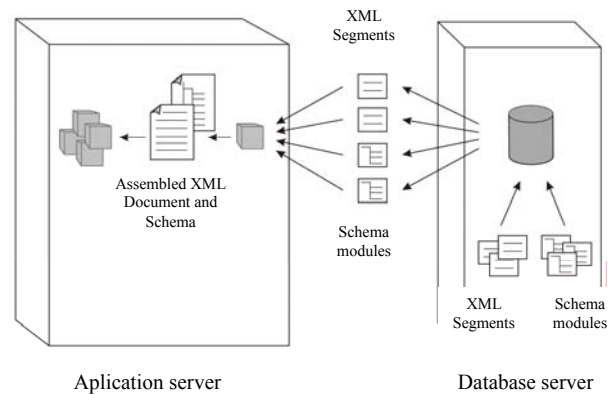
Figure 4. Saving XML segments and XML scheme segments

6. Conclusion

Testing uses relational databases in practice comes to the conclusion that in existing conditions without finished tools is necessary to make all tools or interfaces between existing tools for the processing and preparation of text and images, and databases. Relational databases can save all data types respecting all quoted advantages of saving data in the database. But, when in consideration of the implementation relational database in publishing graphic production many problems are visible for which not existed finished solutions or though finished solutions are not easily accessible.

Native XML database, on the contrary, present the determined file structure which has characteristics of database, but is easily accessible through commercial products and relatively simply is implementation in publishing graphic production surroundings. Dilemma around the choice kind of database remains still because and RDBMS and XML technology has one's advantages and defects in the relation one on other. With combination of these technologies can be used advantages from first and second technology. In this way can be achieved the system who in view of the security of data, the relational model, multiuser work is more superior from XML nativ databases, while would use advantages of XML technologies in the connecting of the different production technologies, independence in view of the different equipment and availability to data from different sources.

Implementation XML technologies in the centre of production of publishing bring following advantages:
- Once text input is not necessary additionally correct for the single type of product, because texts

find in the database and correction of text has been corrected in all types of product.

- Archiving of data is in one place for all types of product
- New versions or republications of products automatically work for all types of product
- Diminishes the number of teams necessary for the production of different types of product
- All changes on the text original are documented in the database and authors can interactive work on the text

7. References

1. Earl, Thomas; Service-Oriented Architecture: A Field Guide to Integrating XML and Web Services, Prentice Hall/PearsonPTR, 2004.

2. Akmal B. Chaudhri, Awais Rashid, Roberto Zicari, Addison Wesley,; XML Data Management: Native XML and XML-Enabled Database Systems, Addison-Wesley Pub Co, 2003.

3. Žiljak, V.; Vođenje grafičke proizvodnje XML tehnologijom, Međunarodni simpozij "Ofsetni tisak", Zagreb 2003.

4. Bates, C.; XML in Theory and Practice, John Wiley & Sons, 2003.

5. Brown, P.; Information Architecture with XML: A Management Strategy, John Wiley & Sons, 2003

6. Barišić, M., Žiljak V.; XML tehnologija u povezivanju izdavaštva i tiskarstva, Međunarodni simpozij "Ofsetni tisak", Zagreb 2003.

7. Žiljak, V.; Information system transformation after implementing XML technology, 14th International Conference on Information and Intelligent Systems, Varaždin 2003.