**XML BASED DATABASE MIGRATION**

##### A PROJECT REPORT

###### ***Submitted by***

**KUSMA THUMMAGUNTA**

**SHAMA HINA.D**

**Under the guidance of**

**Mrs. J. BRINDHA MERIN, AP(Sr.G)/CSE**

**ABSTRACT**

Data migration using xml as a conversion tool is a system that accepts user input for migrating tables in one database to a different database. Various Databases like MySQL Server, PostgreSQL, MS Access, SQLite are considered for the project. The Databases can also be Oracle or MS SQL Server etc. but not considered as the process would be similar for all of these Databases. All these databases are in various encoding systems and can't be directly migrated. Therefore XML is used for migration of these tables from one database to another database as XML is platform independent and supports all encoding schemes. XML is designed to structure the data and carry it in a rational and sensible manner. This helps web developers and programmers to manipulate the data effectively in a easy way. But, XML cant be the solution by itself. XML is just a way which is very convenient way to structure and carry data. During the conversion of the data in the tables, the database tables are not changed and stay unaffected. The fields which are retrieved from the tables are next processed for converting into XML. Migration of databases using XML is a project which helps in converting the native Database like My SQL and MS-Access into XML which is a text formatted flat file.This gives the ability to store different kinds of application system data. Using XML for the backend storage helps in reducing the occupied memory space and hence it is not required to pay any additional amount in creating and maintaining the XML database because it is just a file and so requires very less memory space.

Existing Systems of Migration are difficult or complicated processes as that requires the conversion of tables from one database into another database by exporting them as excel sheets or csv files. The disadvantages of it are : it's a Slow and tedious migration process, requires a great deal of expenditure and it's not very user-friendly. The proposed system is considered to be fast migration tool to be implemented using JSP and XML. This system migrates any RDBMS into another RDBMS with an easy GUI. Advantages of it are: it's a really fast migration process, it's completely free as all the technologies are open-source and it's very user friendly.

**Chapter 1**

**INTRODUCTION**

Database Migration through XML is a project through which the prevailing databases are migrated to another destination database using XML as a tool. By converting thesee databases into the XML file format, the databases like Oracle and MS Access for the generation of the XML code are conneced inside the network.

Example : Let us consider, Database Administrator holds a firm with multiple companies with each of them having different database. When the database Administrator desires to get all different database’s of that firm together in to one single database then the database administrator has to migrate all these different databases in to one single database. Migrating one database into a different database may be a complicated and hectic huge process.For situations like this our project can often be used, as the entire database is converted into a XML file first and then into to the desired destination database. Doing this makes the conversion easier and therefore the source database stays unchanged and protected. To access these databases is allowed by only authorized persons. In order to choose the different kinds of databases to be migrated, the database administrator is initially prompted for username and password and only after the authentication, the administrator is allowed to access them. The database administrator will be asked for the database name and then to select all the desired tables present in the database. The database is displayed to the administrator after selection, from which he/she can select the table which is wanted to be converted.

Before all this conversion begins, the database Administrator can check if the proper tables are being selected by viewing the tables in the selected database and can make desired modifications. The databases are then converted and the database administrator is received with a message saying migration is complete. A new file is then created and the converted file is going to be in the form of XML format in that file and can be previewed to the database administrator in a browser and then viewed completely in the database format once migration is complete.During the conversion of the table data, the database containing the tables stays unaffected with out any changes and therefore the fields which are retrieved from the tables are subjected for the XML conversion. The document overview is a migration technique for converting various databases of various storage formats like Oracle, My SQL, MS SQL Server etc. to the single XML Format as XML is convenient and faster cross-platform application format.

1. **Motivation**

The motivation of database migration project is to analyze the probability of Database migration from different databases into XML Format and to explore the possibility of using XML files as database tables inside the web page. By making use of this project we will convert any database format into one single format called XML, because it is a cross platform, machine independent file format . When the database formats like .mdb, .sql, .dmp, .dbf etc.. Cannot interact with one another directly, xml comes for rescue. Moreover, it can easily communicate with front-end languages like C#.NET,VB.NET or Java etc.., by serving as a database for these coding languages. It might be secure but may include constraints also. In our project, we used XML as a database after converting various database table formats from Oracle, My-SQL Server, Derby DB, MS-Access to the single XML Format.

1. **Problem Definition**

Database Migration through XML is first by converting the source databases like Oracle and MS-Access into the XML database and then to the destination databases like MySql and Postgresql. This project has a capability to store any sort of data from the appliance system since the storage end is a XML file which is in text format. There is great reduction in the storage space and very less memory is occupies since XML is used as a backend. And it doesn’t incur aditional costs as XML database is in file format and is easy to maintain with high security using encryption if required. Comparatively, over the other databases XML databases has more advantages and comfort.

1. **Limitation**

Various databases like MySQL Server, PostgreSQL, MS Access, SQLite are considered in the project. The paid databases like Oracle or MS SQL Server etc. but not considered but the process would be similar for all these Databases.

**Chapter 3**

**METHODOLOGIES**

1. **EXISTING SYSTEM**

The Existing system is very helpful only for single database conversions. Databases are most commonly in wide usage in different resource types in these days. A part of the server memory is hence occupied by these databases and this leads to additional costs in order to maintain these databases. In web, databases are most widely in usage at the present times. The amount is based upon the factors like size of the database, and also Oracle and Access databases are platform dependent.

1. **Disadvantages in Existing system**

This process is complex and has subsequent drawbacks :

* Data corruption will occur.
* Requires huge memory space.
* Incurs additional expenses for database security.

Inside the server, a vast ammount of memory space is occupied by these databases and hence some amount of cash has to be paid to the service provider to keep up that database. The money to be paid differs accordingly depending upon the size of the database and also throws many challenges as different problems arise during the maintenance the database. There is a constant need to make database replication and database backup has to be performed constantly from time to time and it accounts to huge memory space.

1. **PROPOSED SYSTEM**

Database Migration through XML is migrating the prevailing database like Oracle and MS-Access into new desired destinatiion database with the help of XML as a conversion tool. There is great reduction in the storage space and very less memory is occupies since XML is used as a backend.

The XML has the power to be processed on almost any platform because it is platform independent and hence we are able to use this in different kinds of operating system.The data in XML format cant be restricted by any proxy server or Firewall. Here maintenance is extremely easy as there is very less memory space is occupied as it is in file format.

Earlier the maintenance of the database was a hectic task and also huge added costs. So many databases aren't supported in several platforms but XML file format supports different platforms.

1. **Advantages of Proposed system**

* Occupies very less memory space.
* No additional costs for maintaining the database.
* Since this proposed system modifies the contents of the database into a text file, there isn’t much need of administrator support.
* There is not much data loss since the database is converted into plain text format.
* This is very much compatible in any OS as it is cross platform independent.
* Human resource requirement is barely minimum.
* The accessing speed is going to be fast during the processing of data.
* Though it’s a flat file when compared to database storage it provides enhanced security. Enhanced security is assured since we offer Authentication while retrieving tables from the Databases.

1. **REQUIREMENT SPECIFICATION**

Requirement analysis and Specification has a vital importance in creating a good quality software solution for the problem encountered. To assess the clarity of the software the requirements are to br refined and analyzed very carefully. For the implementation of the successful software, it is very much needed for the requirements to be represented in very clear way. Each and every mentioned requirement should be very consistent with the overall stated objective. This project development consists of the following requirements.

* Software Requirements
* Hardware Requirements

1. **Software Requirements**

|  |  |
| --- | --- |
| Operating System | : Windows 10 |
| Coding Language | : JAVA 1.7, JSP Servlets |
| Frontend | : HTML5, JQuery, JavaScript |
| Webserver | : Apache Tomcat 7.0 |
| Database | : MySQL Server 5.1, MS Access, SQLite, PostgreSQL |

1. **Hardware Requirements (Minimum)**

|  |  |
| --- | --- |
| Processor | : Pentium-IV 2.4GHz or above |
| RAM | : 1 GB |
| Hard Disk Drive | : 160GB |

**Chapter 4**

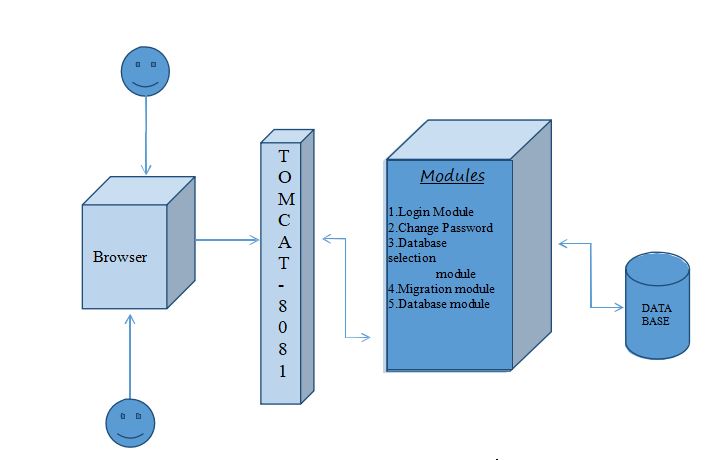
**SYSTEM DESIGN**

1. **SYSTEM ARCHITECTURE**

During System design, there is a the program structure, data structure and procedural deatils needed to be refined progressively and developed .They should then be reviewed before documenting them. There are two perspectives from which a system design can be viewed. They are technical perspective or project management perspective.

1. **Software Architecture**

Software architecture defines the high level structures of the software and then helps in constructing the process of those structure creation and then the documentation.These structures are essential to explain about the software. There are software elements in each structure.It is used to define the relations among those elements and to explain their properties.It also helps in describing the connectivity between elements and relations. The software architecture is shown in Figure 4.1.1



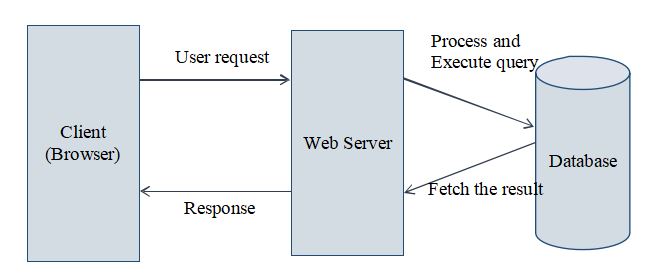
**FIG 4.1.1 SOFTWARE ARCHITECTURE**

This Software architecture and its documentation facilitates communication between stakeholders, and also helps in capturing the early decisions regarding the high level design. This facilitates the reusage of these components design in another projects.

1. **Technical Architecture**

Technical Architecture of an application helps to design and document teh software application. When developers want to build the system from the scratch or modify the existing computer system,this architecture can be used to create a schematic blueprint which helps in achieving the target very easily.

The layer of the system architecture of the computer which helps in defining and specifying the protocols, interfaces and parameters used in the system and product architecture layers are included in this technical architecture. Before building a software application, it is important that the computer architecture to be defined.The communication network used by the application is typically defined using the technical architecture. The technical architecture is given in the Figure 4.1.2



**FIG 4.1.2 TECHNICAL ARCHITECTURE**

1. **Modules**
2. **Database Selection Module**

This module is responsible for selecting a particular source database and destination database for migration. There are a set of databases listed under “select source databases” drop down button and similarly a set of destination database under “select destination button”. After choosing the destination database, the migration starts and enters the next module. For additional confirmation, a pop up can be added with the dialogue with source and destination databases.

1. **Migration Module**

This module is responsible for migration of tables from a source database to a destination database. This module is initiated once the user confirms the source and destination databases. A request is sent to the server for the auto execution of migration JSP code process. This module converts the entire data of source Db into XML format and then inserts XML format generated by source DB into destination DB in its format.

1. **Database Module**

This module is responsible for storing the tables in various databases after or before migration. This module handles the key constraints, schema specifications and storing the migrated information. This module helps in solving the duplicate table problems if migration occurs more than once.

1. **Login Module**

This is an authentication module for entering into the migration software. The login details are also stored in the xml format and no backend database is used to store these login details.

1. **Change Password**

This is for providing better security for the admin password, so that the password is changed regularly for better authentication. This module is restricted for other users as the information in the databases may be confidential and can be misused and hence only admin can register new people and change their passwords and give permissions to another registered in case the admin wants to quit.

**Chapter 5**

**LANGUAGES AND TECHNOLOGIES**

1. **CODE AND LANGUAGES USED**
2. **Java**

This gives a brief of what and why Java language is used in this project. It is intended for the developers where the code which runs on one platform need not be compiled again to run on the other. Java applications can typically run on any Java virtual machine despite the architecture of the computer by compiling to byte code. Especially for client- serverweb applications, java is one of the most popular programming languages in use. Here, for the application it is needed to retrieve the required tables and data from My SQL DB which is the source database.

1. **HTML**

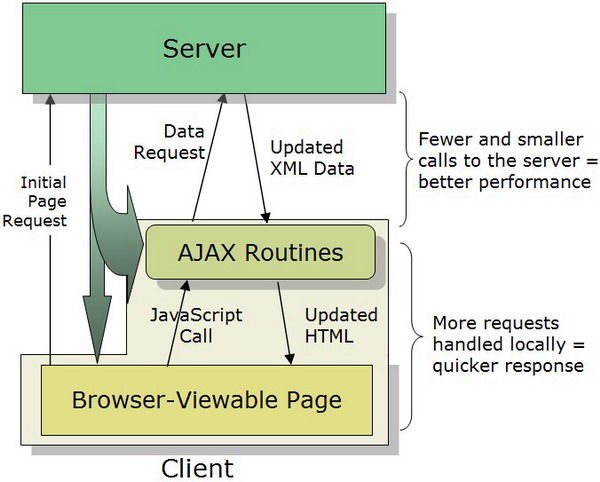
The main purpose of using Hypertext Mark-up Language (HTML), allow users in producing the web pages which includes text, pointer and graphics which redirect to other webpages usinhg hyperlinks. The main point of using hyperlinks is to link one point into the another one.

1. **Apache**

Apache and MySQL modules are started as server application and database. Web server acts as a proper tool as middleman between clients and server communication .

1. **JQuery**

Using JavaScript on websites is made easy by JQuery. It is a set of JavaScript methods and functions. It is light weighted and helps in wrapping many javascript lines into methods which can be called by single line code. It simplifies many complex JavaScripts, CSS manipulations, AJAX calls and HTML/DOM manipulations. JQueries has many plugins for all the complex tasks.



**FIG 5.1.4 IMPLEMENTATION OF JQUERY**

1. **AJAX**

AJAX is the used for server data exchange and to update webpage modules without reloading the whole page. It loads data in the background to display it on the webpage. It doesn’t require to reload the entire page again. JQueries provide various methods for AJAX functionality. Together with JQuery, AJAX is used in requesting text, HTML,XML from remote web servers using HTTP Get or Post to in loading the external data directly to the web page HTML elements.

1. **UTF-8 Encoding**

**Unicode Text Format – 8 :**

Unicode format is a standard in computing industry for representing and to handle the text which is expressed in most of the writing systems in the world and for consistent encoding. The latest version of Unicode contains the repertory of over 110,000 characters which covers a range of 100 scripts and many other symbols are being developed in the conjuction with the Universal Character Set standard. There is an encoding method in each standard with set of reference data computer files, a standand encoding set and for the reference of vision it includes a set of charts with code. It also contains a huge number of items related to the standards like rules for normalization, properties of the characters,collation,rendering,decomposition, a display order which is bidirectional. The success of the Unicode’s character sets which are unifying is the main reason why the computer software iation and in internationalization. In many recent technologies, including many modern operating systems XML, Java , and the Microsoft has implemented this standard.

By using various types of character encoding, unicode can be implemented. UTF-8, UTF-16 are the most widely and predominently used encodings and the now-obsolete UCS-2. For any ASCII character, one byte is used by UTF-8 where all have the same values for the code either in ASCII or UTF-8 and for the other characters, it takes up to four bytes. For each character 16-bit code unit (two 8-bit bytes) is used by the UCS-2 but in the current Uniode standard, not every character can be encoded. UCS- 2 is extended by UTF-16 , for the characters which can be represented in UCS-2, one 16-bit unit is used and to handle each of the other characters which are additional, two 16-bit units are used to handle them.

1. **DATABASES**
2. **My SQLServer**

MySQL server is a database management system which acts like a client and server system. My SQL server provides capabilities of querying and connectivity, along with the ability of excellent structuring of data and for many different platforms integration. This MySql server is very suitable for accessing databases as it provides very rich functionalities like the speed, connectivity and also the security. For highly demanding environments of production, it helps a lot in handling very large databases in a quick and reliable manner. For supporting different administrative tools, varied back ends, client programs and libraries, and also many other API’s there is a SQL server in the system which is multi-threaded.

1. **SQLite**

The programming library of C contains the SQLite. Unlike any other database management systems, SQLite doesn’t work like a client–server database engine. Rather, it is embedded into the end program.

There are no standalone processes with which the application program communicates in the SQLite engine unlike any other client–server database management systems. The SQLite library which can also be called dynamically, instead is linked in and turns out to be application program’s integral part.Through simple function calls, the SQLite’s functionality is used by the application program, which helps in reducing database access latency. There is more efficiency in a single process function calls rather than that of in inter-process communication. On the host machine, the entire database (definitions, tables, indices, and the data itself) is stored in the SQLite as a single cross-platform file. Though SQLite writes can be only sequentially performed read operations can be multitasked. By locking the entire database during the write operation, this simple design is implemented.

1. **PostGRE SQL server**

PostgreSQL or simply Postgres is an open-source software and an object-relational database management system (ORDBMS) with focuses on standards- compliance and also extensibility . Many workloads with the range starting from very small single-machine applications to very large Internet-facing applications which consists of many concurrent users can be handled. Working as a database server, the main functionality is to store data, securely also for supporting the best practices,. It also allows the retrieval on the request of other software applications. Being a cross-platform, PostgreSQL can run on many different operating systems including Linux ,FreeBSD ,Solaris , and Microsoft Windows. The vast majority of Linux distributions have it available in supplied packages.

All the databases which are newly created will by default have a schema named “public”. Also the public schema isn’t mandatory and any additional new schemas can be added. The system check order for the unqualified objects schemas (those without a prefixed schema) is determined by a "search\_path"; Search paths can be configured on a database or role level. The search path by default contains the special schema “$user” which first looks for schema named after the user in the connected database. If such schema isn't found, then the next schema in the list is proceeded to. Which ever valid schema is found first in the search path, new objects are created in it.

**Chapter 6**

**IMPLEMENTATION**

1. **DATA EXCHANGE PROBLEM**

This algorithm is constructed in a way to solve data exchange problem. Migrating the database contents to the target schema instance that reflects the data of the source base as accurately as possible. This problem of data exchange is several decades old, and now once again become very important when XML has been created , a data format which was created to solve the data exchange problem.

The basic problem here is to recognize two things : a) Which data from the source should be transferred and b) To where in the target must the source data go into. In the sense, with two major differences this partly has the problem of schema integration. Firstly, There is no need to create a new schema, the target schema is already present. Secondly, on contrary to the case of the schema integration problem, there persists a set of own constraints on the target schema that has be dealt with during this data migration.

1. **COMPLEXITIES AND CHALLENGES**

There are different databases types that are actually used in the Real-time. When there is the need for interoperation between databases, that’s where the main complexity arises. The entire database or some part it has be migrated to a common database at this stage. There will be a overhead for sure if the databases are too large. Some part of the memory is occupied by these databases. So inorder to maintain the databses ,it requires the users to pay some amount to the site owner.

The data exchange problem is because of the two main reasons. Firstly, in contrast with the XML files, there is very less structural data in relational database, which means that in many cases, in the XML files, the structural data is huge than the actual data. To find the associations between data items, this huge amount of structural information can be exploited. Secondly,by using namespaces, RDF and ontologies, data migration in the future might be fast, error-free and effortless.

1. **ALGORITHM :**

Login

* Select source and destination databases

Auto Execution of Migration JSP code process

* converts the entire data of source Db into XML format
* Inserts XML format generated by source DB into destination DB in its format

Generation of XML Format from SourceDBProcess:

* Connect to the jdbc driver
* If table=sourceDB,get column names,constraints and data

Create\_tags\_xml process:

* Insert table name as the root tag in xml
* Get the table details to create xml tags
* Get the schema of table to insert into xml

Insert XML format data into DestDB process:

* Connect to jdbc driver
* Parse the xml file of source database
* Get the table name from the root tag
* All the column names are obtained by considering the child tags
* Insert the values into table

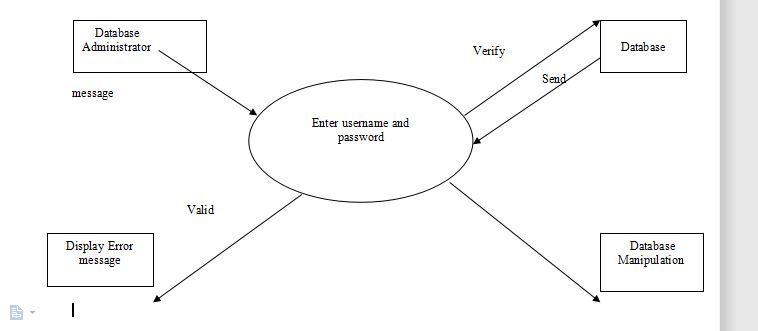
There are two assumptions made in the implementation of the proposed method. Firstly, each element of the XML file must be related semantically to both of its parent and as well as the children, hence related to its ancestors and descendents by transitivity. Secondly, in an XML file, since any two elements will share a common ancestor and so any two elements are infact semantically related, even if one of them doesn’t fall in the ancestors-descendants path of the other one.

The algorithm inputs is the XML source file, the data source schema of the XML, the target data source schema of the XML, and also includes file which specifies the mappings between the names of the source and target data source schema elements. Hence, to migrate this data, at each step the algorithm has to be constructed with the knowledge of the elements it will have to import from the source file so as to finally insert that data into the target file. In order to do that, the algorithm has to traverse the target Schema and for every node that will be encountere, create an appropriate XPath expression and fetch the appropriate node from the source. The enhanced database migration process has the following steps :

* Accessing the Server
* Retrieval of tables
* XML Code generation
* Preview of XML file

1. **Accessing the Server :**

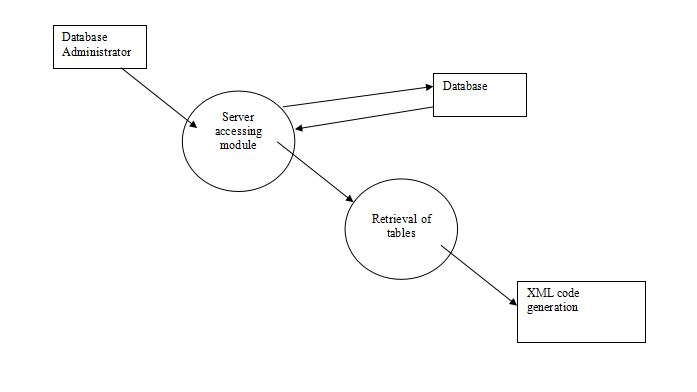
SQL server consists of its own username(root access) and password, for accessing the database file. It has high security which means wiith improper user name and password, the database files in SQL server can’t be accessed. During the conversion of databases into XML file format, the database like MS-Access gets connected in the network which then generates the XML code. Authentication must be provided for user the and also should also protect by restricting the unauthorized persons accessing the server database. MS-Access database can also be connected through the local drivers, floppy or compact disk. The function of accessing the server is shown in Figure 6.3.1.



**FIGURE 6.3.1 ACCESSING THE SERVER**

1. **Tables Retrievals :**

The structure of the table can be seen by selecting the database table each. In particular the displayed structural data that are data type, size, fieldname, scale and precision. By collecting these are used to veriify the cascading style shee(CSS). If any need to modify the database is identified then select the table, manipulate it using the operations like insert, update and delete.

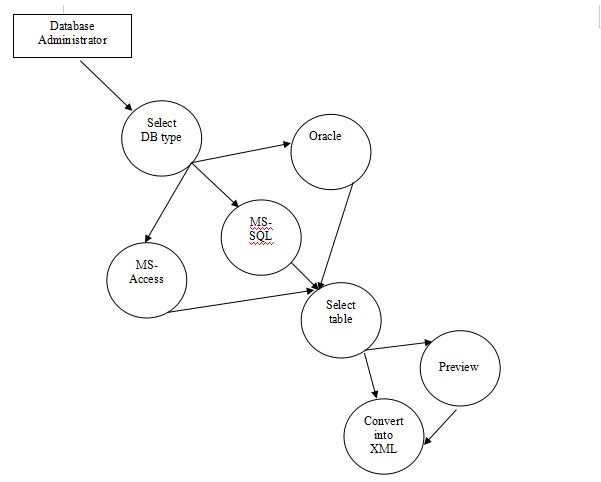


**FIGURE 6.3.2 TABLE RETRIEVAL**

After the user authentication, selection of the database type is allowed by the user and hence prompted for choosing the database name. The User also has the choice of selecting the table which needs to be converted and also be preview it to reduce malselections. The process of retrieving the tables is shown in Figure 6.3.2.

1. **Generation of XML code :**

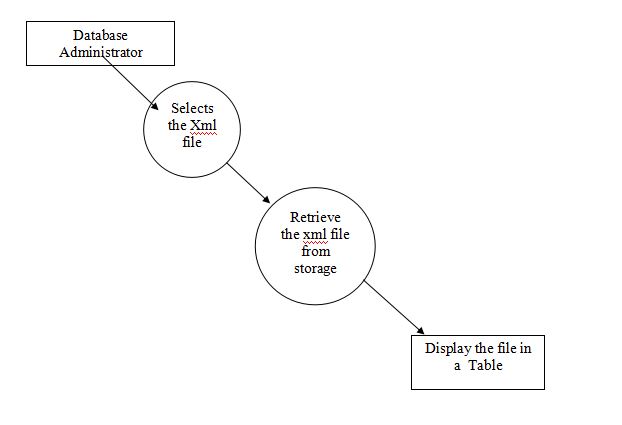
The table data is first selected and then is proceeded for the conversion into XML file format. For the process of conversion, the selected table has to be assigned implicitly to a temporary file. By using the appropriate code, that file which is assigned is then converted into a XML file. Even after the conversion, the original database’s data will remain unchanged. The tables are retrieved from the database, for each and every field individual tags are created as just by the user specifications. This step is used to generate the XML code for the database and it is shown in Figure 6.3.3



**FIGURE 6.3.3 GENERATION OF XML CODE**

1. **Preview of XML file:**

To view in XML format, the XML file name is prompted to the user and by doing this it can check if all the respective table fields are converted as requested by the client. If any error occurs, the XML file which is generated can be modified and the of XML file preview is shown in Figure 6.3.4 and Still there wont be any change in the source file.



**FIGURE 6.3.4 PREVIEW OF XML FILE**

**Chapter 8**

**CONCLUSION AND FUTURE SCOPE**

1. **CONCLUSION**

By making use of this project we can convert any database format into one single format called XML, as it is a cross-platform, machine independent file format . When the database formats like .mdb, .sql, .dmp, .dbf etc.. Cannot interact with each other directly, xml comes for rescue. Moreover, it can easily communicate with front- end languages like C#.NET,VB.NET or Java etc.., by serving as a database for these coding languages. It could be secure and can include constraints also. In our project, we used XML as a database after converting various database table formats from Oracle, My-SQL Server, Derby DB, MS-Access to one single XML Format. In this project there are only two users , The database Administrator and the .NET programmer.

1. **FUTURE ENHANCEMENT**

The Project could be extended further to include XSD/XS schema related conversions for further enhancement of the XML file format. Using this schema-based XML Format would greatly improve the security of XML File at file level.