Resource Optimization Using XML

Gaurav Kumar[1] and **Anu Suneja**[2]

Abstract - Extensible Markup Language [XML] database is one of the hot areas of research now days. Number of developers and research organizations are working on the capabilities and efficiency of using Extensible Markup Language [XML] database because of its unique structure and processing speed. In any web based database search, there is need of a server side script as well as back-end RDBMS [Relational Database Management System]. Generally SQL [Structured Query Language] is used for querying the database. Moreover, the web hosting cost is also very high in various plans using these technologies. XML has an exceptional feature to be used as a database as well as a document. XML document is generally formatted using CSS [Cascading Style Sheets] or XSL [Extensible Stylesheet Language.

In this research work, we have compared the capability and efficiency of XML as a database rather than a simple web document. This work is dedicated to the competence of PHP [Hypertext Preprocessor] and XML over PHP and MYSQL. In our results, we have proved that XML is giving good results rather than MYSQL.

Index Terms – Optimization, XML, Web Development, Database.

1. INTRODUCTION

In any web based database search, there is need of a server side script as well as back-end RDBMS [Relational Database Management System]. Generally SQL [Structured Query Language] is used for querying the database. Moreover, the web hosting cost is also very high in various plans using these technologies**.** XML has an exceptional feature to be used as a database as well as a document [1]. XML document is generally formatted using CSS [Cascading Style Sheets or XSL [Extensible Stylesheet Language] [2].

2. EXISTING THEORY

The problem in the existing architecture is the platform dependence of various scripts and languages. The hosting of Windows Platform support Microsoft products and obviously costlier than the other hosting plans [3]. Servers which host the website require operating systems and licenses. Windows 2003 and other related applications like SQL Server each cost a significant amount of money; on the other hand, Linux is a free operating system to download, install and operate.

The basic ideas underlying XML are very simple: tags on data elements identify the meaning of the data, rather than, e.g., specifying how the data should be formatted [as in HTML], and relationships between data elements are provided via simple nesting and references [4].

To host the website on Windows Server platform, there is restriction that we can use ASP [Active Server Pages] for database maintenance. At the back-end, there is restriction to use MS-ACCESS or SQL SERVER as RDBMS.

On other side, Linux hosting uses PHP or CGI [Common Gateway Interface]/PERL [Project Extraction and Report Language] with MYSQL as back-end database. ColdFusion and JSP do not fit in these categories of hosting. Java Server Pages [JSP] is a Java technology that allows the software developers to create dynamically web applications, with HTML, XML, or other document types. JSP needs the web server compatible with Java Technology and the commonly used Web Servers are JRun and Apache Tomcat. On the other hand, ColdFusion is a software language that is also used for Internet application development such as for dynamically-generated web sites. ColdFusion is a similar Active Server Pages, JavaServer Pages or PHP but it is platform independent. More specifically, Microsoft based technologies require additional tools including Antivirus utilities for security which is more costly to host in international web hosting servers [5].

Table 1 depicts the platform dependence of various scripts and respective back-end databases.

Server Side Script [HTTP Web Server required]	**Back End Database**
ASP [PWS, IIS]	MS Access, SQL Server
JSP [JRun, WebSphere, WebLogic, Apache Tomcat, JBoss, JWS]	MS Access, SQL Server, Oracle
PHP [Abyss, Apache http server]	MS Access, MySQL, SQL Server, Oracle
CGI with PERL [Abyss, Apache http server]	MS Access, MySQL, SQL Server, Oracle
JavaScript[No Need Of Web Server]	Data File [Not Secured]
XML **[No Need Of Web Server]**	It may be used as Database itself

Table 1: Different Server Side Scripts with their back-end databases

Note –The above specified table has been prepared after hands-on experience with all these technologies.

3. RESEARCH METHODOLOGY

We have taken two databases XML and MYSQL for storing records. The values from database are fetched using PHP. Entire exercise was performed to calculate the execution time of getting results from database. The script was executed and

[1]Sr. Lecturer, Chitkara Institute of Engineering and Technology, Rajpura, Punjab
[2]Lecturer, Chitkara Institute of Engineering and Technology, Rajpura, Punjab
E-Mail: [1]kumargaurav.in@gmail.com and [2]suneja.anu@gmail.com

tested on all prominent web browsers so that the actual performance can be obtained.

To prove the efficiency related to XML database, we have used LAMP [LINUX APACHE MYSQL PHP] stack on one side and LAXP [Linux, Apache, XML, PHP] on other side. These two different frameworks are used to get the same type of results in optimized manner.

First of all, I have used XML as back-end database and retrieved the records using PHP. Time of Execution of the query is recorded so that it can be compared with other stack LAXP.

At second attempt, I have used LAMP [Linux Apache MYSQL PHP], in which the same database structure and records are used. These records are retrieved using PHP from MYSQL database.

In this attempt, the time of execution of the script and query execution time is recorded. Finally, the execution time of the queries is compared. After comparing the conclusion is found that XML is obviously the better option as a back-end database rather than other RDBMS packages.

Here is the source code in PHP to fetch the records –

PHP SCRIPT FOR CONNECTION AND QUERIES EXECUTION

```
<?php
list($usec, $sec) = explode(' ',microtime());
 $querytime_before = ((float)$usec + (float)$sec);
$q=$_GET["q"];
$xmlDoc = new DOMDocument();
$xmlDoc->load("EmployeeDatabase.xml");
$x=$xmlDoc->getElementsByTagName('NAME');
for ($i=0; $i<=$x->length-1; $i++)
{
//Process only element nodes
if ($x->item($i)->nodeType==1)
  {
  if ($x->item($i)->childNodes->item(0)->nodeValue == $q)
    {
    $y=($x->item($i)->parentNode);
    }
  }
}

$Employee=($y->childNodes);
for ($i=0;$i<$Employee->length;$i++)
{
//Process only element nodes
if ($Employee->item($i)->nodeType==1)
  {
  echo($Employee->item($i)->nodeName);
  echo(": ");
  echo($Employee->item($i)->childNodes->item(0)-
>nodeValue);
  echo("<br />");
  }
}
list($usec, $sec) = explode(' ',microtime());
 $querytime_after = ((float)$usec + (float)$sec);
 $querytime = $querytime_after - $querytime_before;
 $strQueryTime = 'Query took %01.4f sec';
 echo sprintf($strQueryTime, $querytime);
?>
```

3.1 OPTMIZATION ISSUES

1. Using XML as back-end database entirely rather than a costly RDBMS package hosting
2. Execution time of the queries
3. Optimization of Back-end Database
4. Optimization of processing load on the Server

3.2 DEVELOPMENT TOOLS TO BE USED:

XML with its various standards with PHP as server side script

1. XML-DOM [Document Object Model]
2. XQuery
3. XForm

3.3 TESTING TOOLS

To test the web scripts/code written in XML will be tested in the various prominent HTTP Clients [Web Browsers]

1. Internet Explorer 7
2. Firefox 1.0.2
3. Mozilla 1.7.8
4. Opera 8
5. Netscape 6

4. SIMULATION AND EXPERIMENT RESULTS FOUND

4.1 XML Database Structure Used

An XML document has a logical and a physical structure. More specifically, XML the document is composed of units called entities. An entity may refer the other entities which cause their inclusion in the document. An XML document begins with a root or document entity. The document comprises various declarations, elements, comments, character references, and processing instructions, all of which are indicated in the document by explicit markup. A software component called an XML processor is used to read XML documents and provides access to their content and structure. An XML processor is doing its work on behalf of another module which is called the application. This specification tells the required behavior of an XML processor in terms of how it should read the XML data and the information it must provide to the application [6].

Here is the XML Database which is used at back-end named as EmployeeDatabase.xml -

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<EMPLOYEES>
    <EMP>
        <ID>1</ID>
        <NAME>BOB DYLAN</NAME>
        <SALARY>9000</SALARY>
        <DEPARTMENT>RESEARCH</DEPARTMENT>
            </EMP>
    <EMP>
        <ID>2</ID>
        <NAME>BONNIE TYLER</NAME>
        <SALARY>10000</SALARY>
        <DEPARTMENT>TESTING</DEPARTMENT>
            </EMP>
    <EMP>
        <ID>3</ID>
        <NAME>DR. SMITH</NAME>
        <SALARY>19000</SALARY>
        <DEPARTMENT>TESTING</DEPARTMENT>
            </EMP>
    <EMP>
        <ID>4</ID>
        <NAME>GARY MOORE</NAME>
        <SALARY>23000</SALARY>

<DEPARTMENT>DEVELOPMENT</DEPARTMENT>
            </EMP>
    <EMP>
        <ID>5</ID>
        <NAME>EROS RAMAZZOTTI</NAME>
        <SALARY>30000</SALARY>

<DEPARTMENT>DEVELOPMENT</DEPARTMENT>
            </EMP>
    <EMP>
        <ID>6</ID>
        <NAME>BEE GEES</NAME>
        <SALARY>13000</SALARY>
        <DEPARTMENT>RESEARCH</DEPARTMENT>
            </EMP>
    <EMP>
        <ID>7</ID>
        <NAME>DR.HOOK</NAME>
        <SALARY>UK</SALARY>
        <DEPARTMENT>TESTING</DEPARTMENT>
            </EMP>
    <EMP>
        <ID>8</ID>
        <NAME>ROD STEWART</NAME>
        <SALARY>12000</SALARY>

<DEPARTMENT>DEVELOPMENT</DEPARTMENT>
         </EMP>
    <EMP>
        <ID>9</ID>
        <NAME>ANDREA BOCELLI</NAME>
        <SALARY>17000</SALARY>

<DEPARTMENT>DEVELOPMENT</DEPARTMENT>

    </EMP>
    <EMP>
        <ID>10</ID>
        <NAME>PERCY SLEDGE</NAME>
        <SALARY>80000</SALARY>
        <DEPARTMENT>RESEARCH</DEPARTMENT>
    </EMP>
</EMPLOYEES>
```

4.2 Experiment Results on Execution of PHP Script

ATTEMPTS OF RETRIEVAL OF RECORDS FROM DATABASE	**XML/PHPv**	**PHP/MYSQL**
ATTEMPT 1	Query took 0.0435 sec	Query took 0.0885 sec
ATTEMPT 2	Query took 0.0430 sec	Query took 0.0825 sec
ATTEMPT 3	Query took 0.0235 sec	Query took 0.0715 sec
ATTEMPT 4	Query took 0.0245 sec	Query took 0.0825 sec
ATTEMPT 5	Query took 0.0435 sec	Query took 0.0845 sec
ATTEMPT 6	Query took 0.0425 sec	Query took 0.0755 sec
ATTEMPT 7	Query took 0.0435 sec	Query took 0.0765 sec
ATTEMPT 8	Query took 0.0425 sec	Query took 0.0875 sec
ATTEMPT 9	Query took 0.0335 sec	Query took 0.0685 sec
ATTEMPT 10	Query took 0.0435 sec	Query took 0.0885 sec
ATTEMPT 11	Query took 0.0535 sec	Query took 0.0885 sec
ATTEMPT 12	Query took 0.0630 sec	Query took 0.0825 sec
ATTEMPT 13	Query took 0.0435 sec	Query took 0.0755 sec
ATTEMPT 14	Query took 0.0645 sec	Query took 0.0865 sec
ATTEMPT 15	Query took 0.0435 sec	Query took 0.0875 sec
ATTEMPT 16	Query took 0.0325 sec	Query took 0.0785 sec

ATTEMPTS OF RETRIEVAL OF RECORDS FROM DATABASE	**XML/PHPv**	**PHP/MYSQL**
ATTEMPT 17	Query took 0.0535 sec	Query took 0.0755 sec
ATTEMPT 18	Query took 0.0625 sec	Query took 0.0835 sec
ATTEMPT 19	Query took 0.0735 sec	Query took 0.0625 sec
ATTEMPT 20	Query took 0.0235 sec	Query took 0.0815 sec

4.3 Input Output:

Input

Select the Department: DEVELOPMENT

Output

ID: 9
NAME: Andrea Bocelli
SALARY: 17000
DEPARTMENT: Development
Query took 0.0435 sec

5. CONCLUSION

Generally, to implement the web based database maintenance or standalone, there must be a Web Server [known as HTTP Server], which cater the requests from Web Browser [known as HTTP Client]. The addition resources [Back-End Database] for web based database searching will be avoided using the plan; rather we can use XML as database and retrieval, search. I have achieved the goal of optimization of resources using XML code with PHP.

The existing structure of the web based database maintenance includes the ADO which builds connection between the various web applications. A range of server side scripts are available for structuring the ADO connection. The well-known server side scripts are ASP [Active Server Pages], PHP [Personal Home Pages], CGI [Common Gateway Interface], PERL [Project Extraction and Report Language], JSP [Java Server Pages], Macromedia ColdFusion. The web administrator is required to use ADO connection for creation, updations and maintenance of the back-end database. XML overcomes all these complications by its use as a database structure.

REFERENCES

[1] Salminen and F. W. Tompa, "Requirements for XML Document Database Systems," In *Proceedings of the ACM Symposium on Document Engineering*, Atlanta, Georgia, USA 2001, pages 85-94.

[2] D. DeHaan, D. Toman, M. P. Consens, and M. T. Özsu. "A Comprehensive XQuery to SQL Translation using Dynamic Interval Encoding," In *Proceedings of the ACM SIGMOD Conference on Management of Data,* San Diego, California, 2003, pages 623-634.

[3] Boston Computing Network, http://webhostinginformation.net/disadvantages-of-windows-hosting/ Date Accessed August 10th, 2008

[4] Jennifer Widom, IEEE Data Engineering Bulletin, Special Issue on XML, 22[3]:44-52, September 1999.

[5] Rain Host, http://www.webhostinginfo.in/Disadvantages-of-windows-hosting.html Date Accessed August 10th, 2008

[6] Ronald Bourret, Declaring Elements and Attributes in an XML DTD http://www.rpbourret.com/xml/xmldtd.htm Accessed August 10th, 2008