Annihilation of Twofold Data from Dataset

C. Firza Afreen
Thiruvalluvar University
M.Phil (Research Scholar)
Marudhar Kesari Jain College for Women
Vaniyambadi, Tamil Nadu, India

G. Shobarani
Thiruvalluvar University
Head, PG Department of Computer Applications
Marudhar Kesari Jain College for Women
Vaniyambadi, Tamil Nadu, India

ABSTRACT

The main objective of this research work is to identify the twofold data from the dataset. The primary data collection method is used to collect the data. The virtual environment is created using Cloud Tool VMWare Workstation then the Apriori's frequent item mining method is used to find out the twofold data and the Hashing technique is used to search the data using keyword. The software is designed for the same using .NET with a user friendly GUI.

General Terms

Cloud Computing, Virtualization

Keywords

VMWare Workstation, Twofold Data, Virtual Machine

1. INTRODUCTION

1.1 Problem statement 1

The research problem which is presented here is how to annihilate the two fold data from the dataset.

1.2 Problem statement 2

The application software which is used to solve the above problem, if it is not compatible with the Operating System, then how the compatibility is created between the Operating System and Application Software by using the CLOUD TOOL.

1.3 Research outline

The purpose of this research work is to remove the twofold data from the dataset. If massive amount of data is present, then it will take more storage space and the processing time will also become more. In order to overcome from such problem, here the Apriori Algorithm is used in this research work, where by using the frequent item mining method [2], the repeated data is removed and the hashing technique is used to search the data using keyword. If the Operating System is not compatible with the application software, the CLOUD TOOL VMware Workstation is used to create the Virtual Environment [5]. In the Virtual Environment the new Operating System which is compatible with the Application Software can be installed, and then the Application Software is installed and used to annihilate the twofold data. Once the twofold data is annihilated, the processing time will become fast and the memory space is also saved. The advantage of using Cloud Tool is that multiple operating systems can be used in one system.

2. VMWARE WORKSTATION

VMWare Workstation is the virtualization platform. With the help of VMWare it is possible to run multiple operating systems at the same time. The original Operating System on which Virtual Machine is installed is known as Host Operating System. The operating system which is installed in Virtual Machine is known as Guest Operating System.

2.1 Options provided by virtual machine

VMWare provides various options, important options are listed below. [5]

2.1.1 Starting and shutting down virtual machine

Both of these operations can be performed simply by using Virtual Machine Menu Bar or by using Virtual Machine Toolbar. [5]

2.1.2 Suspending and resuming a virtual machine

The suspend option saves the current state and the resume option is used to restart the work from that state. [5]

2.1.3 Unity mode

The Unity Mode option displays all the applications present in the Guest Operating System on the Host Operating System. [5]

2.1.4 Using removable devices in virtual machine

While installing the Virtual Machine, if the relevant setting is done, then it is possible to use the removal devices in Virtual Machine. [5]

2.1.5 Drag and drop feature

As the name indicates, by using this feature it is possible to move Files, Folders, Images etc between Host and Guest Operating System. [5]

2.2 Virtual machine memory allocation and selecting virtual machine configuration

The maximum amount of memory which can be allotted for Virtual Machine which is present in 64 bit Host is 64 GB and the maximum amount of memory which can be allotted for 32 bit host is 8 GB.

While selecting configuration for Virtual Machine, VMWare provides two options that is typical or custom. If the Custom Configuration is selected then the user can customize the settings. In order to accept the default settings, Typical Configuration must be selected.

3. DEVELOPMENT ENVIRONMENT

3.1 Microsoft visual studio 2008

Microsoft Visual Studio is selected as front end for this research work. The purpose of selecting this application is with the help of this software it is possible to create both Windows and Web based application [1, 4].
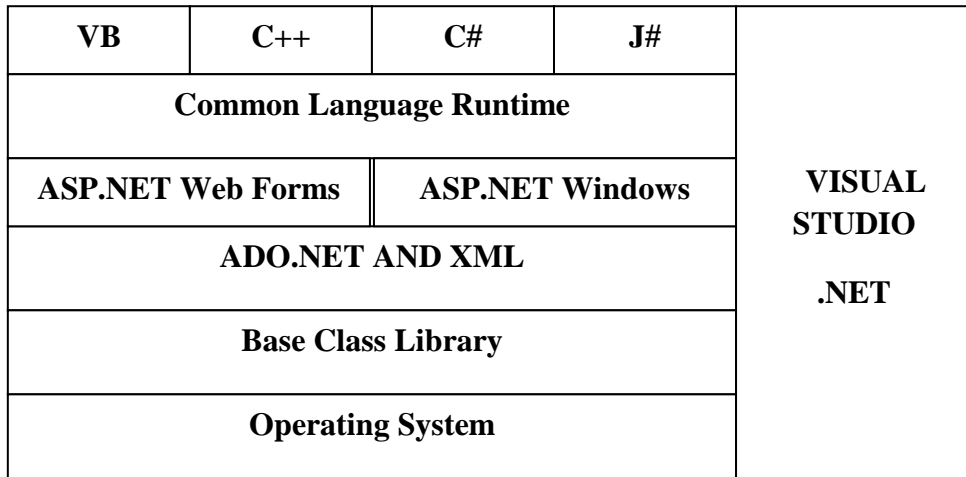
VB	C++	C#	J#	VISUAL STUDIO .NET
Common Language Runtime				
ASP.NET Web Forms		ASP.NET Windows		
ADO.NET AND XML				
Base Class Library				
Operating System				

Figure 1: .NET Architecture

3.2 Microsoft sql server 2005

Microsoft SQL Server is selected as back end for this research. The purpose of selecting this software is it allows a database server to be exposed over web services. The second major purpose of selecting this software is, it is supported only in 32 bit host Operating System, Windows XP is 32 bit host which is installed as a Guest Operating System in VMWare and then the SQL Server is installed in the guest OS, and the compatibility is created between Operating System and Application Software.
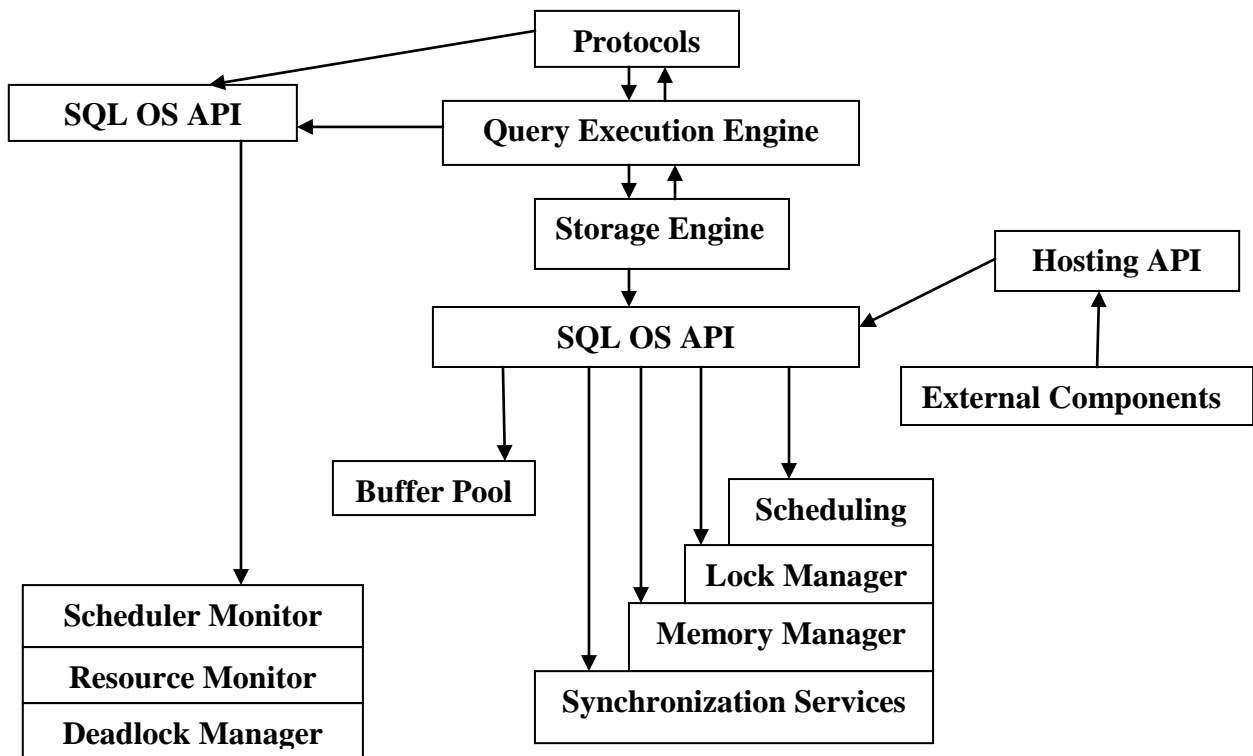
Figure 2: SQL Server 2005 Architecture

4. IMPLEMENTATION AND METHOD DESCRIPTION

Implementation is the realization of an application, execution of a plan, idea etc.

4.1 Method description: Apriori (frequent item mining method)

The method which is used to remove the twofold data from the dataset is described. The data which occurs repeatedly in database is removed by using the same above mentioned method. How it is achieved through SQL SERVER 2005 is represented below diagrammatically. Hashing technique which serves as an address to retrieve the data from database. Here the key is assigned manually to the database for one field, and the remaining data is retrieved by using the field for which the key is assigned.

4.2 How the sql server 2005 is used to remove the redundant data?

For example if we take the student information their Personal, Attendance and Mark data the fields which occurs repeatedly is the Register Number, Name, Department and Year. Normally the primary key is assigned for Register Number in one table, and the reference key is created in the other table for the same field so that if the register number given in the main table does not occurs then the system will display an error message.

In such case what happens are repeatedly the fields like Register Number, Name, Department and Year is created in three different tables like Alumni Personal, Alumni Attendance and Alumni Mark Table. To avoid the repeated creation of data here common one table is created named Alumniinfo where all the fields related to their Personal, Attendance and Marks are stored in one table where only the Primary key is created for Student Register Number.

The needed functionality is achieved if the table is created in this way. The SQL SERVER binds the student needed data alone in the grid even though the various data is present in the same table. The following steps are used to achieve the result. Place DataGrid and SqlServer data binder tool in the form.

Select the Server Name and Database Name. After selecting this test condition must be checked. If the condition is satisfied the tables which are present inside the selected database will be displayed. Select the table name, the alumniinfo is the table name which is selected as shown in the below figure. Once the table name is selected the fields which are present inside the table will be displayed. Select the required field which you want to bind in the grid. In the below given Figure 3 the fields related to personal information of the student alone is selected.
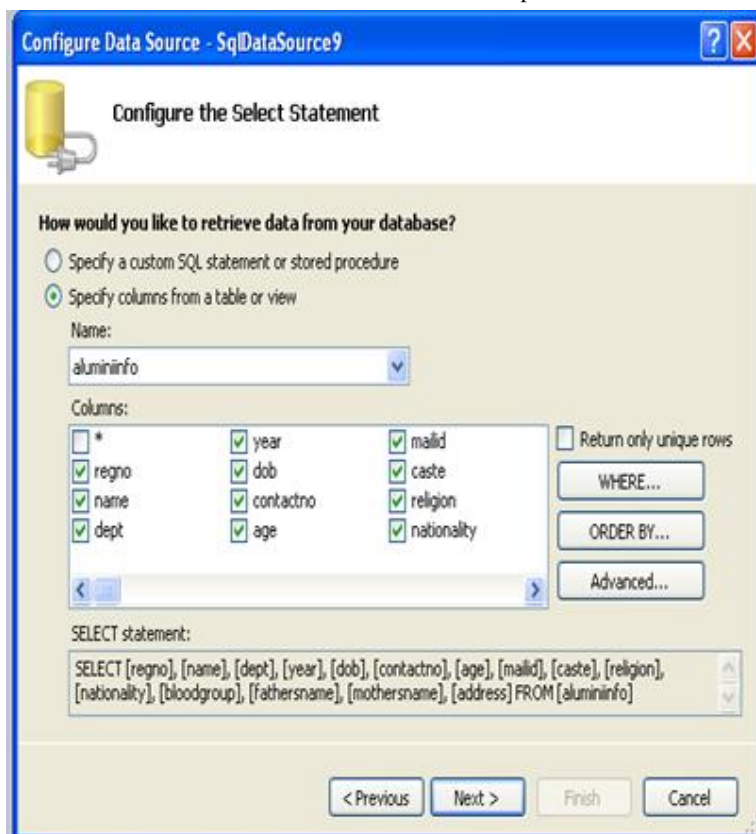

Figure 3: Represents selected table name and various fields present inside the selected table

Select next and finish button, then the selected data will be binded in the grid. The same method will be followed to display the student's attendance and mark data, just select the needed fields from the table and bind it to the grid. In such way the data is mined in one table and the required field alone is accessed and assigned to grids.

Basic operations like insertion, deletion, updation is done through the front end with the help of coding. Whatever data the user enters, deletes and modifies through coding it is automatically updated in the Grid.

To view the entire data of one student based on their Register Number then follow the same procedure explained above the difference is instead of clicking next and finish button select WHERE condition, select Column name, Operator, Source, Control ID then finally select ADD button as shown in Figure 2.

In the below given Figure 4 the Column name selected is Regno, Operator is =, source is Control, in Control ID select the Textbox number i.e. the ID of textbox for eg, Textbox1 or Textbox2 in which the register number is going to be entered.
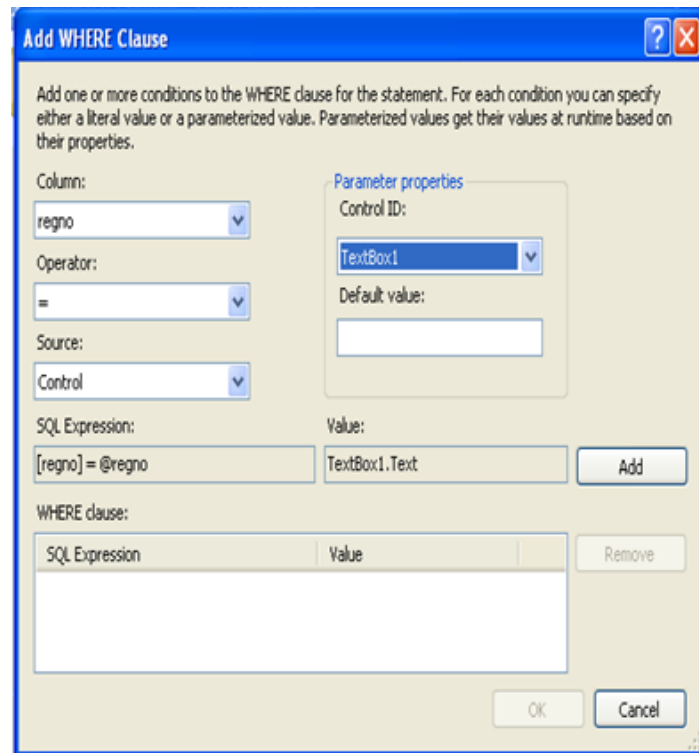
Figure 4: Represents how control id is assigned to server

Click Add button. Then the below given Figure 5 will be displayed, which shows that the Control ID is assigned to SQL Expression. Click OK.
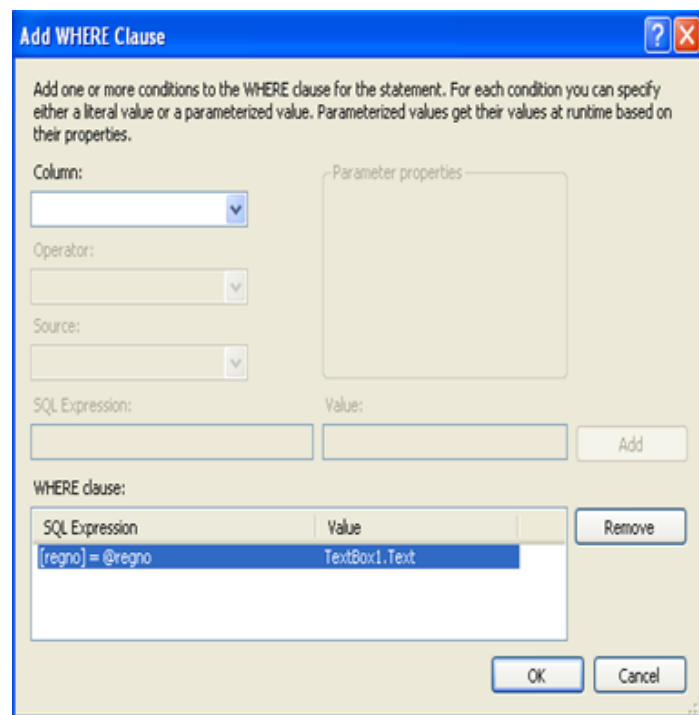
Figure 5: Represents which field is assigned for selected Control ID

This above given Figure 5 indicates that whenever the user enters the Register Number the relevant data of that particular Number is assigned to grid. In such way with the help of one single table the data can be accessed in many possible ways according to the need.

The item which occurs frequently is detected manually and without redundant data the table is created in SQL SERVER. The problem with SQL SERVER is, it is not compatible with all the OS, the Cloud Tool VMware Workstation plays an important role which creates a Virtual Environment, where the OS which is compatible with the SQL SERVER is installed and used.

In this dissertation it is shown that the Cloud Tool brings compatibility between the OS and the Application Software. The frequent item mining method which is used in many areas to mine the data can also be used in Cloud Environment to mine the data with the help of an Application Software.

5. CONCLUSION AND FUTURE WORK

The Compatibility is created between the Operating System and Application Software using Cloud Tool VMWare Workstation. The Software is designed as per the requirements using .NET and SQL SERVER in Virtual Environment. The twofold data is annihilated from the dataset which is achieved using SQL Server software. If the domain registration is done for the application then it is possible to access the software from any location.

6. REFERENCES

[1] Greg Buczek, 2001. ASP.NET Developer's Guide.

[2] Jiawei Han., Micheline Kamber, 2006. Data Mining: Concepts and Techniques.

[3] Kothari, C.R. 2004. Research Methodology Methods and Techniques.

[4] Mathew MacDonald, 2002. ASP.NET the Complete Reference.

[5] Rajkumar Buyya., James Broberg., and Andrzej Goscinski, 2011. Cloud Computing Principles and Paradigms.