



The impact of Green Database, Data Mining and Data Warehousing on the HealthCare System

Omar S. SalehMinistry Of higher Education and
Scientific Research- Baghdad- Iraq**Samar H. Muhammad**University of Diyala- Engineering
College- Diyala- Iraq**Marwah K. Farhan**Ministry Of higher Education and
Scientific Research -Baghdad- Iraq

Abstract - Currently the majority of companies have data centers for storing their information; high-end server machines and devices are what data centers required. This claim is true for healthcare system. However, applying green database techniques in Regional Health Authority (RHA) will helps to reduce the amount of energy that be consumed by data centers.

Nowadays, the healthcare system has more need for data mining although the differences in its approaches. Data mining helps the hospital to prevention the errors, hospitals can discover new, useful and potentially life-saving knowledge if they apply data mining on the existing data. For example, a study (Heath Grades Hospital Study 2007) in the United States found about 87% of the hospital death might have been prevented, the staff of hospitals had been more careful to avoiding the errors. By mining the records of hospital such safety issues could be addressed through the hospital management.

The implementation of data warehouses and data centers requires patients and costs. Without a careful planning implementing a data warehouse can be very costly; especially when it comes to electricity. RHA has to look at the available possibilities to implementing a cost effective environmental friendly solution.

In this research, a dive into the possibilities with green database, data mining and data warehousing on healthcare system.

Keywords- RHA, Green Database, Data mining and Data ware housing

I. INTRODUCTION

Currently the majority of companies have data centers for storing their information; high-end server machines and devices are what data centers required. There are two factors will be affected, a lot of power and deploying will be consumed and a lot of money will be consumed as well. [1]. However, applying green database techniques in Regional Health Authority (RHA) will helps to reduce the amount of energy that be consumed by data centers. Green database developing method is different with the developing method of normal database. The concepts of green database are derived from the green computing idea or green IT. Green computing is a set of approaches that be designed to make data centers more efficient through reducing the consuming of power and cooling required. When it comes to compares the difference between the daily power consumption of typical data centers with the monthly power consumption of thousands homes, it will be noticed they are equivalent. The size of data center and the number of systems in data centers are the factors that will be used to identify the number of homes. Green database is a type of database technique that used to manage power consumptions of data centers.

The implementation of data warehouses and data centers requires patients and costs. Without a careful planning implementing a data warehouse can be very costly; especially when it comes to electricity. RHA has to look at the available possibilities to implementing a cost effective environmental friendly solution.

In this research, a dive into the possibilities with green database would take place and the possibilities would presented.

II. RHA BASED SCENARIO

The Regional Health Authority (RHA) manages several local health authorities for each authority there are many hospitals associated with it; for each hospital there are patients. RHA is expected to handle all issues that arise and should be able to management the authorities smoothly. Patients should not face difficulties as they go to the hospitals.

Unfortunately, that is not the case; RHA is facing tremendous issues. It is obvious with the current data structure that the RHA cannot go any further. The current situation RHA is going through is very dangerous; because the RHA is dealing with human soul it should operate in the best possible and efficient manner. To propose a solution the problems have to be thoroughly examined and the solution should go very well with the issues.

The obvious solution for RHA is to build up a data warehouse to build up a data mining application on top of it. Without the need to go in any further this report will flow as to suffice this point.

The RHA is responsible for several local health authorities and each health authority is responsible for approximately 1 million patients and for up to 15 hospitals. The problems that might be raised are all the hospitals and authorities performed differently. Example of that, the duration of staying of the patient in the hospitals is different under the same condition; some of diseases need specific period of time as fixed range of days. But this procedure is not applied in the hospitals of RHA because each authority follows its policy; that's introduces the low satisfaction of the patient. The database includes a huge amount of information that would be useful for the research purpose. After depth analysis some problems will found; Complexity, lack of communication, lack management due to difficult to manage all the authorities, compatibility, proper resource utilization and redundancy are the problems that has been found. The following explanation was driving the researcher to address the mentioned problems.

III. GREEN DATABASE TECHNIQUES

If tuning is to take place it is important to consider the users. The users should not feel any lack in the performance for them to be able to carry out their task in the manner that is expected from them. Tuning is not the only possibility to utilize and control the power consumption over the datacenter, Regional Health Authority (RHA) can look into the cooling system in the data centers; RHA can save a lot with proper cooling system architecture in place. If the proper cooling architecture is in place the data center would not need to acquire more power for more cooling systems.

Tuning and creating a suitable cooling system are very much related to each other. The equipment can be tuned if they were producing much heat. That is why before having hands on any servers or similar equipments it is advisable to check the power specifications and the tuning possibilities that go along with the equipment. [2].

There are many benefits of implementing green database techniques in RHA, one of which is that the green database can decrease the amount of money that be invested in the database centre and it has big impact on power management. However, implementing green database in RHA will reduce the amount (cost) of maintain their IT equipment.

One of the most common technology for green computing is virtualization; in terms of resources, more than one resources will be merged together to work single resource. This reflects point to the RHA scenario which is merge their systems together; that will drive RHA to gain optimum performance. [3].

Virtualization creates a highly efficient environment. The organization would have total control over resources and energy. The RHA would not need multiple physical servers to run the different applications it has. This will result in less hardware equipments; hence less power consumption and less headache as to locate servers and cool them.

Recently, data virtualization gained some traction; virtualizing data is made huge fuzzi. It means the top layer of data storage and processing devices.

An easy way to hide the infrastructure of IT from the people who are not interested is data center virtualization. It cuts the costs of training new staff and lowers operating costs. The reason behind cutting the cost of training is because the staff used one system to learn instead of several systems. Data visualization has a role for saving the money of training new staff especially on the training members who are concerned with actual value of data. One of the methods reuse the old data is virtualization; the old data here does not necessarily mean outdated information but it will reuse in such cases of problems. [4].

Writing an optimized query is another way to go greener database especially the database queries will access to multiple tables. If the number of tables in database is big, accessing to these tables need more processing and more resources will be consumed. Applying these techniques will be depends on the experience and skills of quires writing.

Nowadays, green computing is getting important airtime; the main attention and an important part of green initiative is to reducing the consumption of computer power.

Server virtualization, cloud computing and power monitoring software are the recent trends in green computing. One of the commercial branches that realize to go green is IBM commercial, which started in black and white; the IBM claim is to save 40% on energy costs.

RHA can look at many aspects in order to reduce in energy waste and go green. The obvious point to think of is to tune the equipments for faster performance or lower performance. Sometime it is not necessary for the equipments to run at their full capacity; when there is a shortage in power the data center has to have a control until were control the power usage is possible. Many IT equipments manufacturers like in computer manufacturers the devices like the processor can be tuned for over clocking to achieve faster performance but that of course in the expense of more power consumption. IT who is driving the energy usage and that are an important benefits associated with reducing the consumption. \$ 4.5 billion is roughly cost for data centers account for 61 billion kWh this according to the Department of Energy. [5].

IV. THE BENEFITS OF GREEN DATA BASE FOR HEALTHCARE SYSTEM.

Database integration would be a first step the RHA should go with. Since authorities and hospitals have different operations, unify the policies should be considered as well. When the policies and hospitals are unified, the patient defeat can be reduced due to the patient does not have to go throughout these policies. Unifying the databases of authorities and hospitals would be possible if the policies of the authorities and hospitals are unified and improved.

The proposed solution tend to solve some of the problems mentioned if not most of them; however to make use of the data and aim for the better is to merge those solutions together to make use of data mining.

V. DATA MINING AND ITS IMPACT IN HEALTHCARE SYSTEM

There is a misconception of what data mining comprises in the medical community. However, Data mining is set of techniques that has been used to discovering and describing the patterns and trends in data. [6].

Nowadays, the healthcare system has more need for data mining although the differences in its approaches. There are many arguments that could be supported to using the data mining in healthcare system. [7].

Data mining helps the hospital to prevention the errors, hospitals can discover new, useful and potentially life-saving knowledge if they apply data mining on the existing data. For example, a study (Heath Grades Hospital Study 2007) in the United States found about 87% of the hospital death might have been prevented, the staff of hospitals had been more careful to avoiding the errors. By mining the records of hospital, such safety issues could be addressed through the hospital management.[7].

This indicates that apply data mining on the RHA will helps to prevention the errors, discovering the useful data of the local authorities and hospitals would be possible. There are many benefits of apply data mining in the healthcare systems some of them will be explained. According to the event that has happened in Philippines in October 2006, the hospital discovered death of several new born babies due to bacterial infection, no one knew what happened until the rate of death increased, the department of health examined the hospital records, it found that 12 babies out 28 babies born in 4 October.[8]

With the application of data mining the department of health could detect the unusual events before they worsen. This reflects the role of data mining to predict the unusual events before happened. However, RHA could predict the unusual disease in the hospital by applying data mining techniques. The experts of Health begun to search and examined how to apply data mining for early detection and management of pandemics. Combining spatial modeling, spatial data mining and simulation are techniques outlined by Waltham (2011) to find interesting characteristics of disease outbreak. [8]

VI. DATA MINING FOR HEALTHCARE SERVICES

Data mining technology would be used in the hospitals to improve the quality of services and it helps to give more understanding on the status of patient illness. An example might be raised, RHA is responsible for up to 15 hospitals; it means different database in use such that database of patients' records, database of physical check-up results, and database for radiology, all these databases to analyze patient status. Therefore, to select different types of patient categories for treatment service or another service, data mining and knowledge management tools will be used and applied. For best service in these regards, many constructs can be explored such as patient segmentations, chronic illness, different assurance reimbursement, physical checkup services and self-pay. For the patient segmentation on different kind of services point is to examine different kinds of patient illness, different services will be provided by the hospitals with treatment service customization, notification to the patients from the hospitals to return for planning the most excellent services. For the physical checkup, data mining will be applied by the hospitals to retrieve the status of patient illness to inform the patient for the physical checkup. The self pay service mining the patients' need by hospital such as cancer checkup and skin disease for the treatment of skin beauty. Data mining will be applied to provide a suitable service to earn the maximum reimbursement. [9]

VII. DATA MINING TECHNIQUES IN HEALTHCARE SYSTEM

Data mining techniques can be classified into description and visualization; association and clustering, classification and estimation. Each one of them has roles, to understand the data set especially a big one, detecting the hidden pattern in data especially nonlinear interactions and data containing complex, they can be contributed by Description and visualization. The aim of association model is to decide the jointly of variables while the objective of clustering model is to group the objects like patients, an example of clustering the objects are similar when they belong to the same cluster and dissimilar when the objects belong to different clusters. [10]

Probably one of the most important applications in data mining that involves predictive modeling. The classification model can refer the prediction of the target variable. Decision tree is one of the techniques that can be used in the healthcare, it is an approach that be used to represent sequence of rules that drive to a set or value. Decision trees can be used for direct data mining. Since it takes a special form of explicit rules, it is quite reasonable; the decision trees consist of nodes and branches organized as a tree form; it includes trees of classification and regression. However, the decision trees can be used the data.

Classification is another technique can be used in healthcare system; one of its objectives is to assign a class to locate the previously hidden records as precisely as possible especially if there is a collection of records and each record has set of attributes, a class contain one of these attributes. The reason behind finding classification model for class attributes is to determine the accuracy of the mode.[11].

VIII. DATA WAREHOUSE ARCHITECTURE

Data warehouse is a subject-oriented, integrated, time-variant, and nonvolatile collection of data in support of management's decision-making process. [12].

One of the latest and advanced database technologies is data warehousing; it can be defined as relational database management system. When it comes to the normal data base system; its data can be altered, updated or removed based on its needs while the data warehouse just for load information and arrange the data in subject oriented manner.

Data warehouse usages are all interrelated; they can be used to enhance hospital operations and for data mining. It also can be used for reporting and to integrate data.

The data warehouse consists of various components which are source systems, ETL (extract, transform and load) and DDS (dimensional data store). The source systems are the data that will be loaded into the data warehousing after some operations; in the scenario that has been given the hospital databases are the source system. The hospitals' data is examined using a data profiler; the purpose of using data profiler is to understand the characteristics of the hospitals' data. The data profiler is a tool which has the ability to analyze the hospitals' data such as checking how many rows are in each table, checking if there is a null value in the tables and so on. The (ETL) is mean extract, transform and load system; this system has a vital role in the data warehousing building. Here the (ETL) systems' role is to connect the hospital databases, read the data, transform the data and load it into a target system. After that, the (ETL) integrates, transforms and loads the hospitals' data into dimensional data store (DDS). The dimensional data store is a database that stores the data warehouse data in different format. There is a reason from getting the data from hospitals' data into the dimensional data store is the data in the dimensional data store arranged in a dimensional format. There is another reason from that which is the dimensional data store contains integrated data from several hospitals' databases. The data quality rules do different data quality checks when the (ETL) system loads the data into the dimensional data store. Based on the sequence, rules and logic stored in metadata the (ETL) system is managed by the control system. The metadata is database, which contains the information about the data meaning, the data structure, the data usage and the data quality rules. [13].

Therefore, the (ETL) system mechanism is to clean the data from hospitals' data before putting it into the data warehouse. Here, data-cleansing process is necessary to identify and correct the dirty data. The data quality rules define the dirty data means; it means after the data is extracted from the hospitals' data but before the data loaded into the warehouse. The quality rules will be examined the data whether the data is correct or not and how to deal with them. If the data is correct, load it in to the warehouse. If the data is incorrect there are three options rejected data, corrected data or allowed to be loaded into the warehouse; in that case choosing appropriate one will be depends on the situation. [14].

As mentioned earlier data warehouse has the ability to extract, transform and transform the data from different database management system to integrate them in a single system but in organized method. Since RHA has different database system. It recommends keeping everything in one place to be load easier into data warehousing. This indicates a main point, which is the integration process of data, will be easier. RHA will be responsible to maintain the data of data warehousing that will be drive RHA to monitor the activities of the local authorities and the hospitals. This introduces good points for RHA, which are an easy to making decision process and save time to audit each authority. Once data warehouse placed. [15].

IX. THE BENEFITS OF DATA WAREHOUSE FOR THE HEALTHCARE SYSTEM

There are many reasons make the data warehouse is a good solution for healthcare System. One of which is the integration which means combine the data from different sources. The data warehousing has the ability to handle all the processing data. Data warehousing has consistent processes for loading data from different applications; it has the ability to make standard schema. [16].

Rainardi (2009) has been asked a question of data warehouse implementation the companies are thirsty for data warehouse?

The reasons behind implementation of data warehousing by the companies are the managers of the companies want more information to support their decision-making. In addition to that, some of competitors are implementing data warehouse that encourage the other companies.

The answer was to provide more timely information for decision making while some of managers claim is to serve the current customers. The expansion of the business intelligence indicates needs to access to the data fast. Hence, integrate the data from different resources; perform new type of analysis; reduce the cost to access historical data; the data across the organization will be standardized; turnaround time for analysis and reporting will be improved; and to remove the information load from transaction-oriented databases are the major reason that improve the data warehouse is good for RHA

X. CONCLUSION

Green database, data mining and data ware housing would be an appropriate choice for the healthcare system due to their characteristics. As demonstrated from the techniques of green database. Implement a database without considering the green aspects of it would be just a waste. Green database helps to save energy and save the environment. The green includes the infrastructure, the optimization techniques and the virtualization solutions. The implementation of Data mining and data warehousing would benefit the healthcare system and it was a good solution to RHA. Several points has to consider within the implementation such as proper management and proper coordination. . RHA has to keep track for any changes that can happen after the implementation takes place and make the necessary adjustments accordingly.

REFERENCES

- [1] SD Times. (2010). *Cloud computing is green computing* [Online]. Available from: http://www.sdtimes.com/CLOUD_COMPUTING_IS_GREEN_COMPUTING/By_JOHN_BASSO/About_CLOUD_COMPUTING_and_GREENCOMPUTING/35070 [Accessed: 12 Dec 2013]

- [2] Bouley, D. 2010 *.Impact Of Virtualization On Data Center Physical I infrastructure* [pdf] Available at: <http://www.thegreengrid.org/~media/WhitePapers/White Paper 27 Impact of Virtualization Data On Center Physical Infrastructure 020210.pdf?lang=en> [Accessed 17 Dec 2013].
- [3] VMware. (2006). *Reducing Server Total Cost of Ownership with VMware Virtualization Software*. [Online]. Available from: <http://www.vmware.com/pdf/TCO.pdf>. [Accessed: 10 Dec 2013]
- [4] Hubpages, (2011) *Data Virtualization*. [Online]. Available from: <<http://doups3.hubpages.com/hub/Data-Virtualization>> [Accessed: 10 Dec 2013]
- [5] Rick, A. (2009). *The Green Data Warehouse*. [Online]. Available from: < <http://www.b-eye-network.com/view/9409>> [Accessed: 7 March 2014]
- [6] TecTarget. 1998. *Datamining* [online] Available at: <http://searchsqlserver.techtarget.com/definition/data-mining> [Accessed 6 May 2014].
- [7]. Sybase, n.d. *Replication Server* [online] Available at: <http://www.sybase.ca/products/businesscontinuity/replactionserver> [Accessed 1 May 2014].
- [8] Waltham, M. 2011. *Novell Announces Industry's Unified Compliances and Security monitoring Solution for Healthcare*. [Online]. Available at: <<http://www.darkreading.com/database-security/167901020/security/application-security/229219444/novell-announces-industry-s-unified-compliance-and-security-monitoring-solution-for-healthcare.html>> [Accessed on 11 January 2014].
- [9] Trujillo, G. 2008. *MySQL versus Oracle Features/Functionality* [online] Available at: http://blogs.oracle.com/GeorgeTrujillo/entry/mysql_versus_oracle_features_functionality [Accessed 18 January 2014].
- [10] Bach, M., J. and Cosic, D. (2007) *Data mining usage in health care management: literature survey and decision tree application*, Med glas. 5(1), pp.:57-64.
- [11] Ponniah, P. (2001). *Data Warehousing Fundamentals: A Comprehensive guide for IT Professionals*, JOHN WILEY & SONS, INC.
- [12] Inmon, W. (2005). *Building the Data Warehouse*, 4/e Wiley Publishing, Inc.
- [13] Shelly, S., Saphina, M., Subhash, B., 2010. *Web Services Security Issues in Healthcare Applications*. Ieee.
- [14] Rick V., 2010. *The Advantages of Data Warehouse Appliances Revisited*. [Online]. Available at: <<http://www.b-eye-network.com/view/13506>> [Accessed on 7 March 2014].
- [15]. Rainardi, V. (2009). *Building a Data Warehouse With Examples in SQL Server*, Apress.
- [16] Dan, P., *what are the advantages and disadvantages of data warehouse*. [Online]. Available at: <<http://dssresources.com/faq/index.php?action=artikel&id=180>> [Accessed on 18 February 2014].