Persistent architecture for optimizing web service for e-government implementation

Dr. Badr Almutairi
College of Computer and Information Sciences
Majmaah University
Kingdom of Saudi Arabia
b.algoian@mu.edu.sa

Dr. Ahmad Raza Khan
College of Computer and Information Sciences
Majmaah University
Kingdom of Saudi Arabia
ar.khan@mu.edu.sa

Abstract— **Information in government sector is increasing day by day. The government is providing its citizens e-services which are getting cluttered and difficult to use and consume by the citizens of the country. Information redundancy is becoming a critical topic for all the governmental transactions and thus overloading the databases and the information pool. In this research we focus on creating an architecture for optimizing the information flow within the governmental database schemas which will pull and push information as and when required by the services. We will be designing web services which will act as tuners and transmitters that pull and push data from the data warehouse and transmit the required data to the requesting web services. Thus reducing the amount of information redundancy in the warehouse. We will be creating optimization web services which will inform us about information redundancy in the e-government services. Thus by doing so we improve the throughput and efficiency of the e-government system.**

Keywords—Service Oriented Architecture; Web Service; E-Governance; Optimization ; Datawarehousing; Service Providers and Service Consumers

I. INTRODUCTION

As internet is flooded with information its important to keep track of relevant information which is useful to the government and its stake holders. E- government is driven by various other aspects surrounding the information technology which as economics, cultural and technological. As most of the countries and their governments to around the globe are striving to provide electronic services to the stake holders its becoming difficult to manage huge information for least cost. As the services provided by the government to the stake holders are creating transparency between the citizen and the governmental procedures and policies they are also creating chaos and misleading information dissemination the stake holders at large. Riley in his definition of E-Governance state that it provides better relation among government and its people, providing much more opportunity for expression [1].

E-government has been introducing lots of services to its stake holder some of which include e-democracy, e-voting, e-justice, e-education, e-healthcare, e-reservations, e-market and many more. But as the number of services are increasing the data warehouse for managing these services is also increasing and the complexity of the data bases are also increasing. Thus as information is plaguing the governmental services its important to manage and maintain some kind of information redundancy services which can cater to information mapping with other services as and when they request for the required information from the data warehouse. Information security is a key concern and it needs to be addressed as the data warehouse hold huge amount of e-government data to address this issues prioritization can be used as a tool for better management of these risks using Analytic Hirarchy Process (AHP) [2]. The e-government system needs to be more agile and all the services which cater to most of the developments need to talk to each other using SOAP (Simple Object Access) protocol. As most of the stakeholder's issues are solved using the e-services provided by the government information security becomes a critical factor for the government in this insecure world if information is growing many folds its important to have information redundancy so that all the information is centralized and secured from the hackers and the attackers. Citizens data is crucial to the country today as many e-services are requesting data from different data warehouses its difficult to keep track for various transactions being performed by the citizens of the country which lead to frauds and security birches which can affect the economical strength of the banking sectors and others too. As information grows infrastructure becomes a problem we need to have bigger space to house many servers which in turn needs cooling system like air conditioned rooms which leads to global worming problem and much more. All these problems arise because of information duplication and mismanagement of the inter process communication between web services which are employed in the e-government sectors. Cloud computing has also taken a big leap in e-governance but the security issues still remain not only with the e-government service but also the cloud computing security issues. Duplicate Information is flooded across the government sectors as they do not share information as and when required by the other departments of the government.

II. SYSTEM CONFIGURATION AND MANAGEMENT

As e-government revolves across giving better and more services to the stakeholders our focus in this research is to centralize he data storage and connect and redesign the web services for inter-process communication so that information duplication can be reduced and the efficiency of the e-

978-1-5090-0793-6/16/$31.00 ©2016 IEEE

governments services will improve and as well as the security issues related to information storage and retrieval will also be addressed. There are various e-government models which have been proposed in various research articles but they do not specify the outcomes and meet the desired needs of E-gov. ICT governance consists of internal processes, structures, and mechanisms that organize the decision making process and guiding the government actions [3]. In our research work we are using some optimization algorithms some of which are Randomized Algorithms, Randomized Low-rank Approximation, Randomized K-means Clustering, Randomized Least-squares regression, Randomized Classification (Regression), Randomized Kernel methods and parallel selective algorithm for data optimization. The data warehouse architecture which we are proposing is a centralized data store where most relevant information required by the services will be stored and accessed as and when requested by the e-services in various governmental domains. The architecture which we propose is as follows:

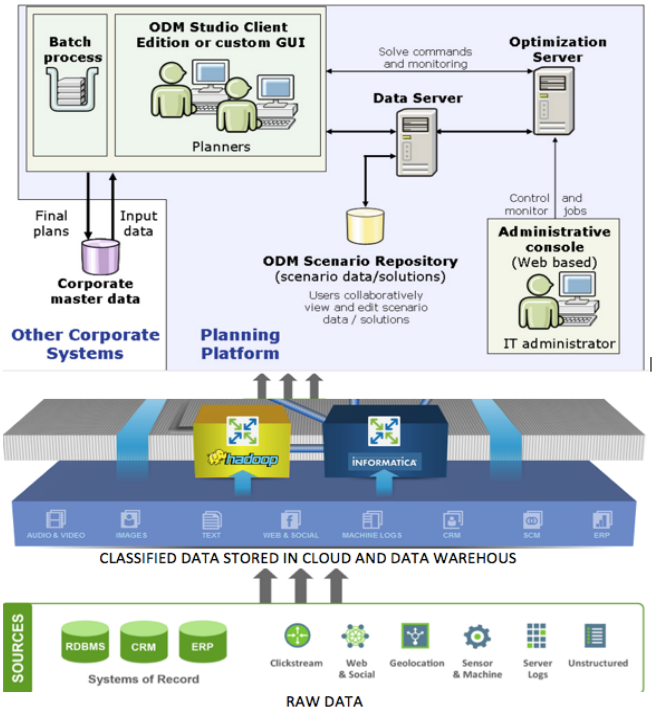

Fig. 1. Framework for data optimization.

The above figure shows the framework which we have designed for optimizing the data storage on the servers. The framework includes three main layers for optimizing and storing the data in the database servers.

A. Raw data

The raw data is also known as primary data which is collected from various sources. The stakeholders have many e-government services running currently and data is collected and stored in different servers which are some times located geographically apart. These servers collect huge amount of data which is just stored on the servers and not processed. Also the amount of data which is collected on different servers can be similar or apart of the data can be similar. The raw data from various sources like Text, Audio, Video, Images, Social, Web, CRM, ERP etc. are all treated as RAW data in our case. We will be collecting the RAW data from all the servers which host different e-government services or we will connect with the servers via a static link of the server and further categorize these data sets to find out similarity in the type of data which is stored.

B. Classified Data stored on Cloud Storage or Data warehouse

In this phase we will be classifying the data into different types based on the file types and their sizes. In this process we will be creating our own data ware house which will now host all the data which is collected from different e-government services which are hosted online. gradable development or implementation of electronic governance using cloud environment, providing a transparent indication of each short-range and long-term designing and policy making [4]. The classification will help us in identifying the type of data which are being consumed by the stakeholders and also we will rank the data according to the throughput of the system. This will improve the efficiency of accessing the data and also improving the efficiency of the optimizing algorithms which will process the data in the upcoming phases.

C. Optimization of calssified data and its storage

In this section we are deploying the Optimization Server which will run the optimization algorithms on various data sets. The administrator will monitor the activity of optimization and will also query the optimization algorithms to run on the data sets based on the classification of the data. These algorithms will give the administrator an overview of the amount of data which is recurring and can be reduced or normalized. It will be the administrator call to optimize or to keep the data as it is in the data warehouse created by us. The e-government services which are requesting for the data will need to be reconnected to the optimized data warehouse which hosts the collection of new data thus by doing so the throughput of the system will improve and the stakeholders will get faster access to the e-government services which are some times running on a non efficient server due to lack of budget and maintenance costs.

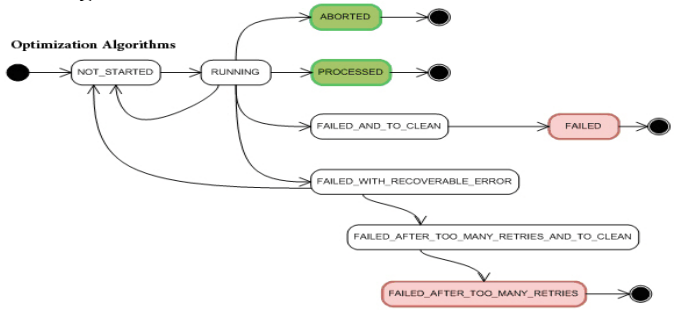

Fig. 2. Optimization Algorithms and their states.

The above figure 2 shows the optimization algorithms which are running and their current states. This is important from the administrative point of view as he will have control over the optimization processes which are taking place on the data.

III. System Design and Development

The deign of the system includes many components out of which one of the most important component is the optimization server which is responsible to collect the classified data from the data warehouse and then optimize it with the algorithms which are best suited for the classification.

Following steps are performed on the data base.

a) Collecting Raw Data from various sources.

b) Classification of the raw data collected.

c) Indexing the classification.

d) Running Optimization Algorithms.

e) Collecting Optimized data on a separate data warehouse.

f) Conecting the e-government web services to the optimized data warehouse.

These taks are important to optimize the huge redundant data which is currenlty flooding the government servers and also the amount of resources required to maintain these servers is huge.

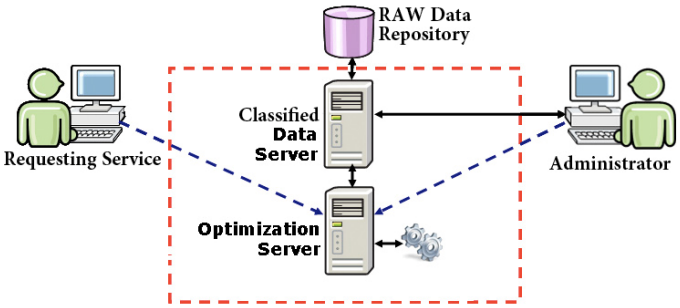

Fig. 3. Communication with the optimization servers.

As shown in the above figure 3. the communication between different servers and the data repository. The client or the requesting service is connected to the optimization server which will have non redundant data thus the throughput of the request given by the requesting service will be served without any delay in information search and dissemination to the web service requesting for the data. On the other end we have the administrator who will be connected to the classified data server and the optimization server the administrator is responsible to run the optimization algorithms and then check the reports which are generated from the optimization algorithms finally if optimizations are correct the administrator will commit the changes to the optimization server thus maintaining the final copy of data which is not redundant and is also optimized.

The RAW data repository will collect all the data from different servers which are hosted by the e-government web services these databases can be of any type and size. The administrators job is to connect to such data sources and then classify the data accordingly and store them in the classified data servers. Data classification will reduce the over load on the optimization algorithms as indexes will be assigned to each classified data source. This classification will result in efficient optimization of the database and also accuracy will be improved in data warehouse.

IV. Report Generation System

Information generated during the optimization process is important so its important to have reporting web services which will generated reports based of different criteria and algorithms used in the process of optimization and information redundancy reduction.

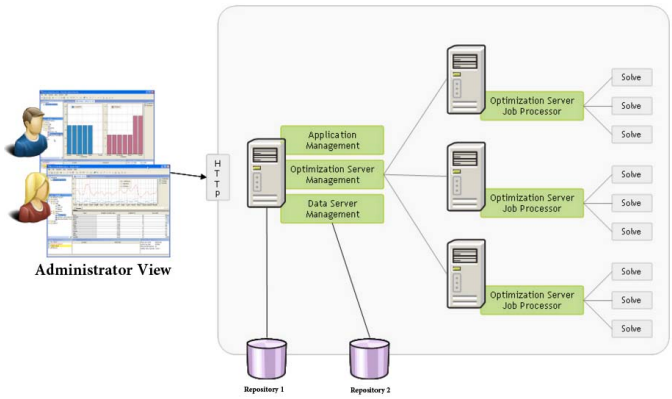

Fig. 4. Repository and Optimization Server's Administrative View.

The above figure 4. shows that the administrator is connected to the data warehuse server which has the following tasks to perform.

- Application Management

There are many applications e-government web services and applications running online which solve many issues related to the stakeholders. Its important to manage a list of web services and the applications which are running and requesting data from various data sources. The task of application management services is to keep track of the web services which are active and running and also the amount of data being requested by the web services.

- Optimization Server Management

Many optimization servers are running online to make the data warehouse redundant free so that the requesting e-government service can find the data fast and easily. Thus it becomes important to keep track of the optimization servers which are online and running to optimize the classified data from the data warehouse. The administrator has full control over the optimization servers and the also get the report from the optimization server about the discrepancies in the databases if any.

- Data Server management

This server is required to connect to the RAW data sources where all the information is initially stored. These sources are the currently running servers which are catering to the e-government web services which have huge amount of redundant data. The function of this server via the administrative control is to connect and collect data from the currently running e-government servers and then create a data warehouse of RAW data do that it can be further processed and the redundancy can be reduced.

V. Optimization Center Console

The optimization control center is the dashboard which will display the overall view of the optimization web services running online and the processed data by the optimization web services.

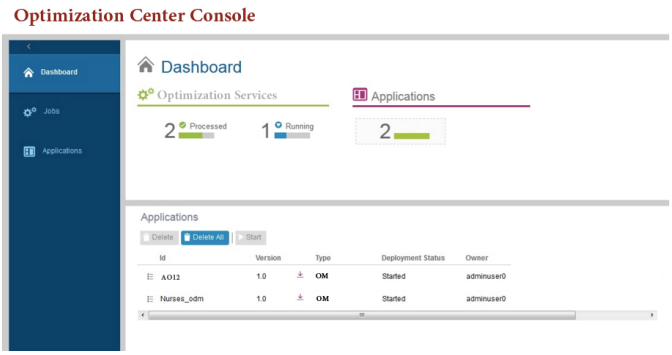

Fig. 5. Optimization Center Console.

The console is the hart of the system where the administrator will get updates on which server and web services are running for optimization and data redundancy and the processed and completed task by the services the Application view will show the administrator the number of applications which are running and are requesting for data from the optimization server.

With the help of this console the administrator will be able to control the optimization process and also the applications which are requesting for the optimized data from the warehouse. The number of jobs completed by the service will be listed in the jobs view of the console and the applications which are available will be listed under the Applications tab of the console. Currently running applications and the services which it is requesting will be displayed on the dashboard of the system. Which helps the administrator to have a birds I view of the entire system. As this system is an agile system its expandability to accept more modules and features is open to the developers. The application of the "information city" framework is called information systems urbanization. Nevertheless, the integration of urbanized information systems is among the major problems to solve in order to maximize and sustain the benefits of information systems urbanization. The researcher analyze the contribution of services to urbanized information systems integration [5]. The system can be enhanced with the latest developments in the optimization algorithms in the databases and query optimization can also be applied. The number of servers which the system can monitor and manage can also be increased and thus the administrator will have full view of what services are running and requesting for information from the data warehouse which is optimized and with less or no redundancy of information which in turn will improve the efficiency of the system .

VI. Conclusion

Last but not the least I would like to conclude by saying that this research project will help solve data redundancy and optimization problems which are plaguing the data warehouses today. Most of the e-government services running today are data intensive and the amount of data flowing with in these requesting web services is huge. Its important to optimize the data warehouse as well as to reduce the clutter of data by using data redundancy and optimization algorithms. One of the key challenges is to collect the RAW data from all the data warehouses, classify them and then apply optimization on these data sources this is a CPU intensive job which requires huge amount of time but its been never late to apply parallel processing on the data warehouses to optimize the databases so that the throughput of the entire system can be improved.

Acknowledgment

We have been working with different sectors of the government and private sector companies this research would have never been successful with out the corporation and collaboration of different people and resource. We thank department of computer and information sciences Majmaah University for giving us this opportunity to work on this research project and providing adequate resources that helped us in collecting information about the problems and also for providing labs for carrying out our research activity. This research work would never complete without the tremendous effort and guidance from our Dean Dr. Hisham Al Saghier who has been motivating us to work on this topic of research and making the resources available in the lab and else wherever required.

References

[1] Riley T.B, "Electronic governance in context, Electronic Governance and Electronic Democracy: Living and Working in the Connected World", 2001, The Commonwealth Centre for Electronic Governance, Ottawa.

[2] Debendra K.M, "Information Security Risk Assessment and Prioritization in E-Governance", The IUP Journal of Information Technology, Vol. X, No. 4, December 2014, pp. 52-58.

[3] Gabriel U.B, "A Development Model For E-Government Based On Ict Governance Concepts", 14th International Conference on WWW/INTERNET 2015 (ICWI 2015), October 2015, pp. 27-37.

[4] Manjunatha Rao, "Efficient and Ubiquitous Software Architecture of e-Governance for Indian Administrative Services", International Journal of Advanced Research in Computer Science, Volume 4, No. 11, Nov-Dec 2013, pp. 66-74.

[5] Sana Bent, "The Governance of Services-Based Integration of Urbanized Information Systems", CENTERIS 2013 - Conference on ENTERprise Information Systems / PRojMAN 2013 - International Conference on Project MANagement / HCist 2013 - International Conference on Health and Social Care Information Systems and Technologies, Procedia Technology 9 (2013), pp 596 – 607.