Enhance Confidentiality by the usage of a Database Management Framework Features to Facilitate Applications Integration

Hena Iqbal[1], M. Habli[2] and Tawfik Zeki[3]
[1]Assistant Professor, Department of School of Engineering and Technology, Dubai, UAE.
[2]Professor, Department of School of Engineering and Technology, Dubai, UAE.
[3]Assistant Professor, Department of School of Engineering and Technology, Dubai, UAE.

(Corresponding author: Hena Iqbal)

ABSTRACT: Links to and integrate data gathered in different countries through international research collaborations. Data managers ought to track who seems to have the access to their servers and whether it is used for various reasons, e.g. Legislature. The data processing is conducted in mathematical applications, and is normally stacked on peak of the data handling framework, example: a Data Base Management System (DBMS), the data analysis is carried out. File Sharing is often managed by the DBMS, whereas statistical analysis is normally managed by another program. To improve efficiency, we suggest an innovative framework for the completion of quantitative study on information put in a DBMS. By calling statistical applications we implemented a system for the implementation of research groups, SAS out of a hierarchical DBMS IBM DB2 over data contained in DB2 database. The research explains the architect design suggested, and the implementation done. Data incorporated and operated by DBMS (Database Management System). The statistically oriented program is named out of the DBMS in the proposed architecture that controls the data that has been implemented. The DBMS 'built in privacy security strategies are used to track all access to integrated data as well as calls to computational application conducting data analysis. Through data acquisition, data is moved from multiple individual data holders to a common position where the acquisition exists. The data owners will be forced to monitor who is processing their data and how it is analyzed. It involves remotely analyzing integrated data on a remote analytics server in a managed manner, because monitoring analyzes which is inefficient to move consolidated data to researchers' computers. Additionally, common programs for illnesses have insufficient funding over integrated data to build their remote infrastructures. We've introduced the SAQeL protocol, SQL Statistical Analysis. SAQeL supports in learning how the statistical program from a Relational DBMS (RDBMS) is difficult to use. The system provides an app for calling the RDBMS IBM DB2 predictive program SAS [5][6]. This was based on an epidemiological analysis of cervical cancer. The core conclusion we can draw from those in its concept even though it's simple incorporate a request framework from an RDBMS, such as IBM DB2, to analytical applications, such as SAS. The second move is to implement a fully featured GUI to report study findings to participants, that are mostly various tables and photos. One of the most challenges which we faced is how to secure the application integration which has communication with the database which was not covered in the previous papers.

Keywords: DBMS, Database Server, infectious disease, SQL, Statistical Software, VPD.

Abbreviations: DBMS, Database Managements System; RDBMS, Relational Database Managements System; SAS, Statistical Analysis System; GUI, Graphical User Interface; VPD, Virtual Private Database.

I. INTRODUCTION

This is also essential so that you can to combine information from different sources for the research of Clinical Immunology. Throughout data migration, data across various separate data holders must be moved to one position where the convergence occurs. Data holders will be obligated to control who handles their data and how they interpret it.

This involves remotely analyzing integrated data on a remote analysis server in a managed manner, because monitoring analyzes with optimized data moved to devices with scholars impractical. In addition, Health-specific initiatives restricted control over interconnected data to build their remote infra structures. The selection of an existing utilities [1-4], facilitate data from databases with remote, secure statistical analyses.

Any of these are focused on systems, where access for computational as well as descriptive applications output are managed isolated with data connection to the databases. Only downside of infrastructures is that running them is expensive: build and operate both. We recommend a new framework for the execution of statistical programs.

Authorization is provided without fact making electronic or duplicates of all or portion behind this job for subjective or academic use, when this case version are not created or distributed for profit or business gain and that duplicates on the first section involve this mention and the perfect quotation. Plagiarism or otherwise resubmitting, submitting to internet sites or transferring them to lists generally requires straight forward approval or a payment.
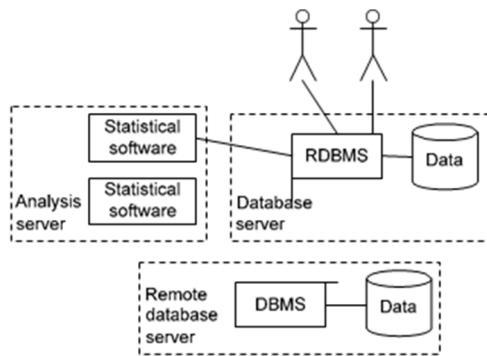
Fig. 1. Proposed architecture.

II. MATERIALS AND METHODS

A. Research Work

There area range of technologies that enable secure data processing. Explanations of that include LISSY from the Luxembourg Income Study [1], Statistics Denmark's Danish System[2], Melbourne Health's BioGrid Australia[3], Statistics Sweden's MONA [4], and CSIRO's PPA[7]. Increasing integrates a digital applications framework where quantitative data processing is conducted on app queries. Scientists on a few of these [2-4] link remote analytics databases immediately and communicate with statistical resources. A distributed analytics databases implement security of scientists, enable researchers to reach on-site statistical tools and similar services, as well as inspect activities for scholars. Such systems [1] enable investigators to use their study services via input mechanisms and don't have clear online connection to study servers. The operation of analytical processes is then conducted out in group. Most of the [1, 2, 4, 7] systems evaluate data the initial files are removed and processed in archives. Hence, remote analytics server file systems conduct authorization of data file accesses.

Many frameworks require DBMS links to mathematical applications, and request that data be extracted directly from Revealed by the database Server [3]. The quantitative technicians developed earlier are represented in forms of perspective. The researchers are calling for the implementation of their mathematical software. Once the investigators are allowed to name the analytical services, the RDBMS delivers the initiatives to the applicable statistical applications for operation on the analytical server. Statistical software access details from either the perceptions set out throughout the RDBMS during implementation. The information will be transferred on to the statistical system until the investigators are authorized. Following the execution of the statistical projects, the findings quantitative research shall be carried out to the investigators. All permissions are built-in and automatically performed in RDBMSs. Work will provide adequate information to enable repetition of most operations. While many techniques are defined it can be separated in to the subcategories.

B. Proposal Design

We suggest to name the RDBMS analytical system that handles the results. The design that we suggest as seen in Fig. 1. Utilizing the same RDBMS GUI, investigators conduct distributed searches, construct displays, or

conduct methodology. Throughout the figure on the application table, the RDBMS combines user data with remote data. For example, federated repositories have been used for information incorporation in CODIR [11] or BioGrid Australia [3]. Until investigators may enter the system, the panel must accept them and their ventures. Database application queries via the RBDMS database GUI.

The inquiries either automate repetitive analyses that produce raw data as a consequence of sheets or the storing of embedded database opinions, or the inquiries produce beliefs which are defined from the embedded database in aspects of other opinions or tables. The queries can only be executed if the prosecutors are permitted to carry them out. The developers then set up their computational programs that are ideally adapted to us among other potential existing protocols and formulations of application because Java SQL queries.
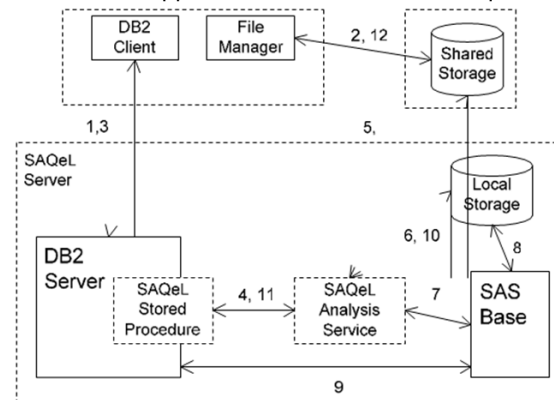

Fig. 2. SAQEL architecture.

C. The Prototype

By implementing a prototype, SAQeL, we examined our proposed architecture. With the ability to call SAS [5], SAQeL Extends DB2 [6] and executes SAS software as for the records contained in DB2. This section explains SAQeL's present application: specifications concerning the architecture, as well as an overview of the mathematical research in it.

(i) **Project Recommendations:** The SAQEL Design is focused on efficiency usability usability. One thing that needs to be understood is whether DB2 will contact SAS. Another way is utilizing SAS Compatibility Innovations to build an SAS system and afterwards link to the SAS system from DB2. The database method involves SAS components from the SAS Investment remedy, in contrast to SAS Framework components for example SAS Base. It also adds up costs and makes this construction of projects more complex. Some other disadvantage of the SAS Server technique is the incapacity to submit the remedy to different quantitative implementations that don't enforce the own web server alternatives, such as Stata [12]. Some other way to call SAS is through device calls to an SAS client, SAS Foundation, and SAS system batch execution. This involves the preparation together with the interface variables in the SAS program files. The solution is more cost-effective and cheaper than using the SAS server. Moreover, you might name other mathematical programs, such as R [13] and Stata, in batch phases. So, SAQeL calls an SAS client in batch phase.

The issue that needs to be addressed is whether to carry information from DB2 to SAS. Mostly, we have 2 key methods to do the same: by scraping details into file and afterwards entry also to records from an SAS program, or through providing forward junction between SAS and DB2 via, for eg, the ODBC interface [14]. Extraction of the information within reports have smaller efficient, and is worse than linking DB2 directly from SAS.

Additionally, many mathematical programs, such as Stata and R, often incorporate frameworks that link directly that RDBMS. After which data is moved to SAS with the operation in SAQeL by developing a direct link between SAS and DB2 at runtime. External routines are successfully implemented as a traditional way to increase RDBMS functionality, e.g. DB2. The following restrictions are imposed by DB2 on external routines: external routines cannot generate different channels as well as techniques and procedures operating external routines cannot establish new links SAQeL then follows the SAQeL Analysis System listening protocol for accepting queries from an external system and scheduling SAS in group format.

Therefore, SAS must create new links to DB2 for batch processing. The final architecture decision is to set up an additional Java storage procedure to access the SAQeL Review Facility of DB2 to determine the legitimacy for an specification through verifying compatibility with the internal configuration. In the event of a successful response, it collects details from the obtained document, (5) retrieves this current File Server SAS software of the customer, and generates an SAS execution system with the introduction of additional configuration code into the actual SAS software. The created SAS software (6) is transferred to Local Storage in a temporary file. The SAQeL Analysis Service (7) then sends a device request in this development strategy to operate SAS Base in batch mode. SAS Base (8) checks and performs a Local Data file utilizing SAS tools. It (9) creates a link to the DB2 database during the implementation of the data and whether the usage is authorized to access the data. After implementation of the peer reviewed study, the power is transferred to the SAQeL Study Unit.

SAQeL Analysis Service (10) must forward the final output of the study from the Local Storage to the File Server if the implementation is effective. The SAQeL Analysis System then generates feedback report with general information of the result of the performance or error; and (11) sends a notification to SAQeL Stored Procedure. Throughout the DB2 program, SAQeL Stored Procedure submits a notification to the consumer regarding success. Finally, the user (12) uses the File Manager to see the findings of the test.

(ii) **SAQEL Design Form:** Fig. 2 demonstrates Design for SAQeL. The same database, SAQeL Server, operates on both the DB2 and SAS Application systems. DB2 database is expanded by SAQeL Stored Process, an additional storage procedure which is deployed in Java. Enabled on DB2 Folder, SAQeL Stored Method can be invoked with SQL commands. Primary objective of SAQeL Stored Procedure is to pass variables of the operation of data analysis by a customer to SAQeL Application Facility and restore the state of the implementation of statistical data to the consumer. SAQeL Analytics System is a Java process which runs steadily. SAQeL Research System accepts SAQeL Stored Process demands for data analysis through a TCP/IP connection port. Customers connect with DB2 Database and File Administrator to send them evaluating to SAQeL Server for implementation. File Administrator is used to store SAS applications between the User's Machine and SAQeL Database on a shared database and view analytical data. DB2 Developer is used to problem SQL queries performing data analysis and to obtain for the results of constructive evaluation or the incidence of errors. Process of Analysis is given below.

A consumer (1) poses a query that DB2 Client to the DB2 server to build views to be included in the SAS softwares. DB2 Api allows app requests and creates images. The SAS software is user-created then (2) secured to Fileserver. The consumer creates a demand (3) by requesting SAQeL Stored Procedure to do a quantitative review, and the application is accepted by DB2 Server. The customer shall determine the rank and position of the SAS software, the credit points as well as the correct place to store the effects of the study as methodological data variables. First, we are improving the SAQeLSql query guiand delete that client identity requirement or key, because it includes further details on the operation performance. We are also researching when to go for the implementation information instantly via DB2 client, and as well as to treat usage of synchronous replication to do so. Equally we are aiming to install SAS programs from DB2 software.

III. RESULTS AND DISCUSSION

A. An overview of how to conduct analysis

This section is an explanation of a survival research with an infectious disease inquiry into cervical cancer. A scientist then produces views which can be presented by an SAS system. For example display cohort overvmig.pc is generated by:

CREATE VIEW GET_DATA AS SELECT lopnr, diagyr FROM (SELECT lopnr,MIN(cancer_yr) AS diagyr FROM cerv_db.cancerWHERE icd_7='171' AND transaction_DT IS NOT NULL GROUP BY lopnr) WHERE year BETWEEN 1960 AND 2005

First, we are developing the SAQeLSql query interface to remove the need in the name and key of the customer, and to provide additional details about the efficiency of operations. We are also exploring how to return details about the implementation directly through DB2 application, and how to stop using an automatic failover to do so. Similarly, we plan to install DB2 applications SAS applications.The author therefore develops an SAS software from a perspective GET_DATA as following:

PROCEDURE GET_LIST_INFO DATA=survmig.pc_cohort_duration METHOD=km PLOTS=(s) NOCENS; TIME years*censor(1); STRATA birth_place;
RUN;

Linked details was contained in the SAS. Thus, client appealsSAQeL Stored Procedure with the corresponding variables: SAS DB2 connection style user name and authentication, SAS system name, input and output routes:
– CALL SAQEL ('john', 'abc123', 'mysasprogram', 'S:\john\analysis\Programs\', 'S:\john\analysis\Results\');

Once research has been performed, the consumer generates a notification in the DB2 application that the study has been performed and so the findings are in the stated directory.

Paper Name	Our Paper
"Study on Database Management System Security Issues"Mohd Amin MohdYunus #, Sonniaa K.V. Gopala Krishnan #, NazriMohdNawi #, Ely Salwana Mat Surin*	Our paper title "Enhance Confidentiality by the Usage of a Database Management Framework Features to Facilitate Applications Integration" will propose:
– They talk about database encryption only. –They talked about watermarking technique for database encryption. – They are focusing on an algorithm which relies on encrypting the data and putting it into image and generating a secret key and then decrypt it again while retrieving it which causes an overhead while retrieving the data.	– We proposed a new technique in the database which more recent technique which virtual private database (VPD). – We proposed a new technique for data isolation which database materialized views to isolate the no authorized users or access. – What we proposed is more seamless and efficient in the sense of the security and data manipulation because it's not causing any overhead because we use a built in algorithms which will be faster for retrieving the data and more efficient

(i) Using VPD (virtual private database) to secure and limit the accessibility to the data according to the authorized people only, Virtual Private Database (VPD), a feature of RDBMS, Among the most famous functionalities in your directory has been presented. VPD was included where default user rights or even related client functions were inadequate that satisfy protection criteria for programs. Based upon this protection criteria VPD strategies may be easy but complicated. To implement advanced row and/or column level protection criteria regarding protection including legal enforcement, VPD may become included throughout conjunction to that "application context" function. A small explanation of VPD could constrain access to information all through office hours, as well as a more contrasting views of VPD could perhaps publish a higher conversion throughout an authentication button as well as maintaining policies of the consistent focus against the main document.

Regardless of whether clients link to the safe server (via an email, a proxy server or SQL*Plus), the effects are like this. There's never any "application protection problem," as the session key is connected to the screen, and can never be moved through.

Example: Also if a client is identified on a 'customer' list (above) of 'Column Importance' will VPD be programmed to apply the rule even if any crucial column was picked.

CUST_LAST_NAME	CUST_FIRST_NAME	CREDIT_LIMIT	ACCOUNT_MGR_ID
Ecwards	Guillaume		145
Mahoney	Maurice		145
Warden	Maria		147
Landis	Marilou		147
Dvrrie	Rufus		148
Belush	Rufus		148
Seignier	Blake	1200	149
Pcwell	Claude	1200	149

Example: each finance department with any "149" account mgr id could get all entries first from figure of clients, and not just the monthly payment. She might see her of others clients as quickly as she questions this same 'credit limit' section.

One more sophisticated design ("Column Hiding") of VPD makes with a most efficient balance to ease-of-use and security: They often retains exposure for certain personal data throughout a 'customers' row, yet sensitive data is secret.

IV. CONCLUSION

So what we have done as a contribution in this paper in a way to secure these integration point through the database security layers by using the VPD (virtual private database) security which we proposed it in the paper. Through our method, we do not restrict the investigators to what SAS techniques they use. Now we are going to test SAQeL for various consequences samples of SAS systems. We'll include various software services.

The other mission occurring is to develop the customer experience. Firstly, we are enhancing the SAQeL Stored Procedure GUI to remove the demand in the name and password of the customer, and to provide further detail on the effects of the implementation. In addition, we are exploring how to specifically return operation results via DB2 client as well as how to stop using a data stores for this. Likewise, we are searching for DB2 software to install SAS systems.

V. FUTURE SCOPE

We also introduced new software to conduct quantitative examines on combined data in a secure manner. Our program allows full use of the privacy protection granted by RDBMS. This would allow for the development of secure, interconnected infrastructures at affordable expense.

The initial SAQeL implementation is showing strong progress on the framework among DB2 and SAS. It was submitted within 10 weeks by a Bachelor of Computer Science undergraduate. SAQeL was given to our coworkers, who do infectious disease research. Our potential plan would be to merge SAQeL with our organisation approach work to integrate data through Sweden's national population databases, as proposed in the Cross-Organizational Analysis Network (CODIR) project [11]. CODIR strives to prevent publication of all personal details. Furthermore, in order to investigate only historical data, we have to implement SQL query

control, evaluate program instructions and implementation results.

REFERENCES

[1]. Barry, S., & Marc, C. (2003). Remote access systems for statistical analysis of microdata. *Statistics and Computing, 13*, 381-389.

[2]. Borchsenius, L. (2005). New developments in the Danish system for access to micro data. Monographs of official statistics (2005) 13-20.

[3]. Hibbert, M., Gibbs, P., O'Brien, T., Colman, P., Merriel, R., Rafael, N., & Georgeff, M. (2007). The Molecular Medicine Informatics Model (MMIM). *Stud. Health Technol. Inform., 126*, 77-86.

[4]. Hjelm, C. G. (2005). MONA-Microdata ON-Line access at Statistics Sweden. *Monographs of official statistics*, 21-28.

[5]. SAS Software, http://www.sas.com.

[6]. IBM DB2 Software, http://www.ibm.com/software/data/db2/.

[7]. Sparks, R., Carter, C., Donnelly, J.B., O'Keefe, C.M., Duncan, J., Keighley, T., & McAullay, D. (2008). Remote access methods for exploratory data analysis and statistical modelling: Privacy-Preserving Analytics. *Comput Methods Programs Biomed, 91* 208-222.

[8]. Haas, L.M., Lin, E.T., & Roth, M. A. (2002). Data integration through database federation. *IBM Syst. J., 41,* 578-596.

[9]. Luc, B., Fran, oise, F., Fabio, P., & Patrick, V. (2001). Processing Queries with Expensive Functions and Large Objects in Distributed Mediator Systems. Proceedings of the 17th International Conference on Data Engineering. *IEEE Computer Society*, 91-98.

[10]. Tisell, C., & Orsborn, K. (2000). A system for multibody analysis based on object-relational database technology. *Advances in Engineering Software,* 971-984.

[11]. Ruslan, F., Magnus, S., & Jan-Eric, L. (2009). Federated Databases as a Basis for Infrastructure Supporting Epidemiological Research. Proceedings of the 2009 20th International Workshop on Database and Expert Systems Application. *IEEE Computer Society,* 313-317.

[12]. Stata: Data Analysis and Statistical Software, http://www.stata.com.

[13]. The R Project for Statistical Computing, http://www.r- project.org.

[14]. Open Database Connectivity Overview, http://support.microsoft.com/kb/110093.