Cross-Platform Relational Data Extraction Utilizing SQL Server (X-PRESS)

Jimmy Z. Bantog
S&R Membership Shopping
32nd St. 5th Ave., Bonifacio
Global City,
Taguig, Philippines

Luisito Lolong Lacatan
Laguna University
Laguna Sports Complex
Bgy. Bubukal Sta Cruz,
Laguna, Philippines

Mary Ann F. Quioc
SST IV, Philippine Science
High
School, Central Luzon Campus,
Pampanga, Philippines

ABSTRACT

The use of data is vital to the advantage and success of every company and businesses. However, querying, extracting and analyzing data from cross-platform relational databases in just a single run without the need for programming is very challenging. While this process can be done manually, it requires more resources, very time consuming and prone to error. This also involves the use of different tools and lots of consolidation. Data are coming from multi-platform applications, from open source or proprietary software, from different operating systems, from different relational database management systems, from different servers and from any location within the organization. Data are complex that requires it to be processed, stored and managed in several ways. There is a need for a solution that can analyze through the company's different data sources and allowing users to perform queries. Analyzing data is best performed with the use of tools that facilitates efficient data exploration and better querying. This study finds its way to the development of a single tool that is capable of exploring, analyzing and transforming data into information regardless of its environment and configuration and makes it available to everyone. This all-in-one database querying tool provides the facility for accessing data from any cross-platform database across the network via SQL Server linked server connection through distributed and heterogeneous query execution. This tool commonly known as X-PRESS is built on top of a framework based on dynamic creation of entities and attributes and utilizes the power of Structured Query Language. Connection to data is more secure and took advantage of server processing power thus making this tool faster and more suitable for data analysis. X-PRESS is the acronym of Cross-Platform Relational Data Extraction Utilizing SQL Server. "X-P" derives from Cross-Platform while "RESS" is the initial of the keywords Relational Data, Extraction, SQL and Server, respectively.

General Terms

Linked Server, SQL, Database

Keywords

Cross-Platform, Relational Data Extraction, Linked Server, SQL, SQL Server, Database Querying Tool, Database Analyzer, RDBMS

1. INTRODUCTION

Commercial Enterprise relies on information that is coming from processed data. Having more data allows to do new things that were not possible before [1]. This is true that every organization uses data whether they belong to retail, manufacturing, banking and finance, real estate or any other forms of business. This data could be records of their sales, products, clients, customers, suppliers and even employees that are very useful in running their businesses. Using data to its full potential is by transforming it from insight to profit [2]. Data can be simple or very complex. It can be located anywhere in the entire organization, in a regular desktop computer or in extremely powerful servers. These computers and servers could be running in various Operating Systems (OS) like Microsoft Windows or Linux. Data are managed using different Relational Database Management System (RDBMS) across the network for instance Microsoft SQL Server, IBM DB2, Oracle, PostgreSQL, MySQL and among others.

It is common in a company setup to use multi-platform applications and databases depending on their requirements. Some are also using open source while others are using proprietary software or a combination. In short, data are processed, stored and managed anywhere in several ways.

In this digital age, data is being captured and accumulated at a rapid pace in a wide range of fields. A new generation of technologies is needed to help people extract usable information from fast rising volumes of digital data. In other words, data is not information – it is just data. Since information is required for decision-making, the data itself in its common form is frequently not particularly useful [3]. It is really desirable to process and transform the data into information that can be valuable once it's been analyzed. As quoted by Steve Haeckel, IBM's Advance Business Institute's former Director of Strategic Studies, "A company's intelligence quotient is determined by the degree to which its IT infrastructure connects, shares, and gives structure to information" [4].

Given the rising complexity of data as it grows bigger in size, changes frequently, and is distributed across many independent sources, accessibility itself is no longer sufficient. [5]. A solution that can analyze through the company's different data sources and allowing users to perform queries is [6] is needed. This study has focused on the nature of these data, the accessibility despite complexity, the extraction and its transformation into valuable information and makes it available for everyone. In general, this study is best applicable for all big companies that are using multiple applications that deal with big data from different data sources. But this is also applicable for other small to medium enterprises that are using computerized systems that produce data.

2. BACKGROUND OF THE STUDY

Maximizing data to its full potential involves the application of data-science strategies within the whole organization [7] with the combination of a reliable tool. But the problem is the

availability of a tool that is custom-fit to the organization's need. There may be software available that can be purchased but their features are excessive when it is only need to extract and analyze information from data in disparate data sources.

The main beneficiary of the study will be S&R Membership Shopping (S&R). S&R is a company which adopts a warehouse club concept where majority of the merchandise offered are imported brand mostly sourced from the United States. S&R is owned by Puregold, a publicly listed retail company in the Philippine Stock Exchange (PSE) with stock symbol PGOLD. As posted in the PSE website, PGOLD was operating a total of 355 stores all over the country by the end of 2017 which includes 171 hypermarkets, 103 supermarkets, 29 extras, 6 minimarts, 14 S&R warehouse clubs and 32 S&R Quick Service Restaurants.

In S&R, there are frequent and unexpected requirements for data extraction as well as request for data and reports on a daily, weekly and monthly basis. Examples are sales reports which include top selling items or slow moving items, membership data, members purchase report like top members purchases as well as items information and inventories. All these requirements are forwarded to Information Technology (IT) Department who have the control over the data and have the skills to facilitate user's requests. In this way, users are becoming very dependent to the help of IT. While on the other hand, the IT Department needs to allot or pull-out resources to process users' request. In this case, the task or project that is usually assigned to the pulled resource will have a tendency of getting delayed or the company will be forced to hire additional resource in response to increasing demands for data.

Data from different systems that are requested by S&R users are usually stored in several relational databases managed by different RDBMS that are installed in various OS. S&R users refer to anyone ranging from employees of different departments within the organization like Supervisors and Managers up to the top management like Directors, Vice Presidents and President. The data cannot be generated instantly because the process involves several steps and procedures. To be able to provide the data, IT will have to use several querying tools provided for each database to extract the data because the problem is that these tools are lack of the ability to perform cross-database queries (a unique feature that is not present in any of these tools). After extraction, the data is needed to be consolidated and merged using another application like Microsoft Access or Excel before getting the final output. This manual process is extremely time-consuming.

Information must be available anytime to anyone who needs it. S&R users have no freedom to subscribe or access the data that they needed because it is not available to them. They will instead request for particular data on a regular basis. If the request for these data is done routinely, IT will also have to extract these data over and over again. This becomes a recurring task that requires repetitive time and effort for users to request the data and for IT to provide the request. If there is no program that can automatically produce these data, the process can also be prone to error.

3. OBJECTIVES OF THE STUDY
The general objective of this study is to develop a relational data extraction tool that will utilize the Microsoft's linked server technology called Cross-Platform Relational Data Extraction Utilizing SQL Server (X-PRESS). The four specific objectives are:

- To develop a module that will provide data access solution for identifying data sources;

- To develop a module that will extract data from the cross-platform databases in the locale;

- To improve processing time as compared to manual in terms of extracting data according to user requirements;

- To distribute processed data through subscription wherein users can run anytime via self-service without the need for programming and dependency from IT.

4. SIGNIFICANCE OF THE STUDY
People shape their tools and their tools shape them [1]. There are multiple benefits for making business data accessible to everyone. A good starting point is by fostering an analytics-driven culture that begins with empowering the workforce across an organization on how to use data analytics tool [8]. Many companies today have large amounts of business data but little real-world business experience. It's important to use resources and technology that can transform data into useful and actionable information. One of the most important aspects of data is its influence on better analysis and decision-making [9].

The significance of having a tool for data analysis is to give business users with insights over the use of data such as gaining competitive edge through business intelligence with the combination of tools that can gather, access, and analyze business data [6]. Furthermore, analyzing customer's behavior and buying patterns give more advantage in gaining and retaining customers, improving services, marketing, planning, innovation and better business operation. Having the right tools is critically important [10]. This will turn a business into a data-driven organization, where anyone who can benefit from data has access to the information whenever they need it.

X-PRESS will provide benefits to people who have passion about data, to organizations driven by data and to those who are regular users of data.

In global context, there are technologies particularly in the area of relational databases that are powerful and useful but their potentials are often overlooked, unexplored and has very few application. This study will bring awareness in the IT industry to make use of the Linked Server technology to enable the implementation of distributed databases.

In the context of IT, business and enterprise are always backed by Information Technology Professionals. IT Managers, Business Analyst, Data Analyst, Reporting Analyst, Systems Analyst and Systems Developers are among IT Professionals who are the enabler of change and their job involves the use of data. They are primarily those who handle data and provide processed information to other stakeholders.

In the context of business operations, anyone from the business organization such as owner, managers and employees depends on information. Stakeholders are driven by data and it is vital to their performance and business existence. Information is the driving force that empowers them to gain competitive advantage over the use of data. The main beneficiary of X-PRESS is S&R Membership Shopping. It is the initiative of the researcher who also works for S&R to provide solution for the current needs of the company.

In the context of continuous learning, the researcher who is directly handling the company systems and data will be the

first to benefit from using X-PRESS. While future researchers can adapt the techniques that is use by X-PRESS as groundwork in conducting and producing more advanced research in the future.

5. SCOPE AND DELIMITATION

This study is focused on the development of a relational data extraction tool that will utilize the Microsoft's linked server technology. This specifically includes the development of a module that will provide data access solution for identifying data sources that creates and establish SQL Server Linked Server connection to the most widely-used RDBMS. For security reason, it should encrypt sensitive data like user password and Linked Server connection password using built-in encryption algorithm for enhanced security. This also includes the development of a module that will extract data from the cross-platform databases in the locale. This will have features for exploring and previewing data from a database, tables and fields. It can also perform filtering, grouping, sorting and summarizing of data. There will be an SQL Editor for advance users as well as interface for plugging custom queries. In particular, this study will also improve the data processing time through dynamic building of simple to complex queries and cross-database joins optimize for faster execution using its built-in hierarchical execution algorithm. Logging and viewing of executed queries are also essential for performance monitoring and auditing purposes. Lastly, the processed data will be distributed to users in the form of subscription wherein users can run anytime via self-service without the need for programming and dependency from IT. The result can be analyzed, viewed and exported to Microsoft Excel.

Due to design constraints, X-PRESS is developed to be a Windows-based application, not a Web-based, and not intended to run in mobile devices. The system cannot connect to any database server if that is not added as linked server. Data access privilege and restriction on a linked server are based on the remote security context. Granular access control is defined only on the remote database server and not in the linked server connection itself. X-PRESS will not limit access to database on the linked server because all authorized users within the system by default are allowed to connect, explore, analyze and use all available data. Data integrity is also managed by RDBMS, not by X-PRESS. Even if it is equipped with built-in query optimizer, the speed of extracting data is not exclusively depends on X-PRESS but also rely in the database and server performance. X-PRESS is mainly for cross-platform relational data extraction tool, not intended for NoSQL databases.

6. CONCEPTUAL FRAMEWORK

In order to successfully achieve the desired outcome of this study, certain procedures, requirements and ideas were carefully discussed to conceptualize the project's design and development. The graphical representation of the extraction, processing and distribution of data in Fig. 1 is used to describe the conceptual framework of this study.
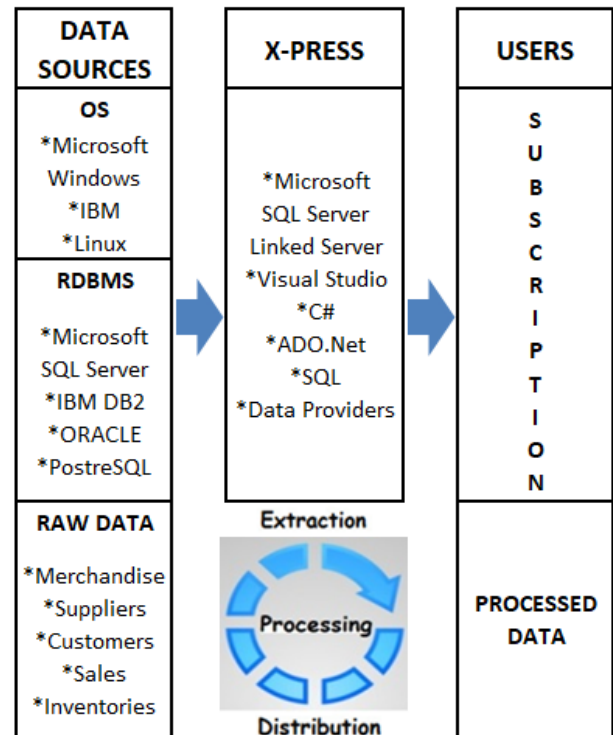

Figure 1: Conceptual Framework of the Study

Fig. 1 shows the system approach for extracting the raw data from disparate data sources that are managed in different RDBMS and installed in different OS. C# which is packaged with Microsoft Visual Studio is the programming language used in this research to developed X-PRESS as it blends well with Rapid Application Development (RAD) software development process. It enables rapid application development because of its innovative, powerful, simple and object-oriented design for building .Net Framework applications [11]. In order to process the data, X-PRESS must adapt several technologies and strategies. The researcher must have basic knowledge in Windows Forms, Multiple Document Interface (MDI), Multi-threading and Data Encryption. ADO.Net is the data access technology for accessing data from RDBMS including Microsoft SQL Server, IBM DB2, Oracle and PostgreSQL which are installed in Windows, Linux or IBM OS/400. Researcher must be also knowledgeable in Structured Query Language (SQL), Microsoft SQL Server Linked Server connection and SQL Server Management Studio (SSMS) as well as familiar with the concept of Data Providers and Drivers including OLEDB Provider for Oracle and SQL Server, OLEDB Provider for Postgres and iSeries Access ODBC Driver. Finally, processed data will be distributed to users via subscription. Users can be anyone from the organization, from the top management or employees from different departments who has given privilege to access X-PRESS.

7. REVIEW OF RELATED STUDIES AND LITERATURES

7.1 Possibilities

Analyzing data from database is one way to utilize its significance and many possibilities. Querying means that users need something out from data and they know exactly what they are looking for. Several studies have been conducted in different fields and have shown that data has the powerful potential towards various applications. In a study

from Japan, the identification of outcomes and clinical characteristics using the analyzed data extracted from a large scale database has been derived from patients with myxedema coma, a life-threatening and emergency presentation of hypothyroidism [12]. The Bill & Melinda Gates Foundation funded Global Pregnancy Collaboration (CoLab) has encouraged study harmonization and systematic data collection to make pregnancy and placental research easier. Its online database COLLECT, which is web-based and allows for online data input and viewing, has a global application in applied science studies [13]. The USDA National Nutrient Database has been supplemented with nutrient composition and ingredient information on branded and store-brand food items from a public-private collaboration for public health between the US Department of Agriculture (USDA), GS1 US, 1WorldSync, the North American Branch of the International Life Sciences Institute and Label Insight through open data sharing [14]. A database tool was used in a study for the cardiovascular function assessment to show the indexes accuracy for calculating pulse wave velocity, which is used for central blood pressure computation [15]. Rice, one of the world's most important agricultural crops, has been studied using bioinformatics tools and specialized databases for data processing, analysis, effective organization and visualization. The article addressed a collection of biological databases that store genomics data on sequence, genetic variation, gene expression, gene-interactomes and pathways to make data interpretation easier [16]. Querying is also an introductory tool in data mining. Though it is considered as an advanced separate study that is far beyond the scope of X-PRESS, it still uses data that may involve the use of querying tools. To mention a few, a data mining study to forecast the Stock Market Index through Sentiment Analysis was able to collect and analyze historical data from Twitter and the Philippine Stock Exchange's closing index. [17]. Some studies which involve the use of data include the assessment of building operational performance [18] and diabetes research using clinical datasets [19]. Improving the quality of education system by studying the instructor performance [20] as well as student performance [21] and enhancing the processes of making decisions to retain valued customers [22] are among studies that relies on data. There are still many data possibilities yet to be discovered in future studies using state of the art tools combined with algorithms and performance optimization.

7.2 Algorithm
Big data generated by organizations are considered of high business value, and algorithms can be applied to extract information from this data. In Aboalsamh's study on a join algorithm for large database, he developed a new join algorithm that uses quadtrees to improve the efficiency of joining two database tables that are very large. Both tables are interpreted by a storage-efficient quadtree in the proposed algorithm, which is designed to facilitate One-Dimensional arrays (1-D arrays). To perform join operations, the algorithm uses the two 1-D arrays of the two tables. The proposed algorithm has been shown to be fast and effective when applied to two very large tables that are too big to fit in main memory [23]. Also, by means of algorithm, Cohen and Narayanaswamy made it possible to write first-order logic queries over data from a variety of sources spread around a network. The data sources are viewed as if they all came from a single database. Web servers and REST APIs, web pages, XML/JSON repositories, flat files, and full-featured databases are examples of data sources that use stored or computed methods to provide data. End users can use their approach to create ad-hoc queries in logic to link data sources. When opposed to manually coding the equivalent correlation algorithms in procedural languages, using a logic-based language to determine correlation parameters shortens the development life cycle and lowers the cost of maintenance for programmers developing applications [24].

7.3 Performance
Querying data from different data sources and platform requires intensive performance that every querying tool should address. Since this is a computationally intensive task, parallel computing techniques are gaining popularity as a way to boost performance [25]. In the study of Mariuta, he identified three methods for improving search efficiency of encrypted data. In his paper, he performed a case study on Microsoft SQL Server's technologies and tools. His first approach encrypts the text without the use of a salt value. The second method is to use the internal encryption function called EncryptByKey, and the third method is to add an additional hash column that is not salt, where the hash values are truncated or even limited in numerical precision, similar to the previous one. [26]. In the same manner to speed up the processes of analyzing huge amount of data as seen in the study of Prajapati, Garg and Chauhan, they have discovered an interesting link between rule mining with consistent and inconsistent detection rule from large sales data in a distributed system. By applying computations in multi-node cluster, their study overcomes the bottleneck in main memory and the processing overhead in a single computing environment. They have created two algorithms, the Hadoop MapReduce based Consistent and Inconsistent Rule Detection (MRCIRD) and Fast Distributed Mining (FDM), which divides the transactional dataset into partitions and distributes the workload to all participating nodes to reduce message forwarding between nodes in the cluster, making it more efficient, flexible, and scalable for mining large amounts of data [27].

7.4 Optimization
The querying of large datasets of data repositories in distributed databases on a worldwide scale has become a demanding research subject since the introduction of distributed computing, particularly since the rise of grids and clouds as well as other service-oriented computing paradigms. In response, Beran, Mach, Schikuta and Vigne conducted a study on how to optimize these database queries. They also took into account a wide range of homogeneous and heterogeneous infrastructures that span multiple virtual organizations with little to no centralized authority. Their study provided a novel global heuristic framework for query execution plan optimization based on a multi-staged blackboard method for determining which available data, operations and resources must be considered to optimize query performance. Furthermore, an evaluation scenario demonstrates their findings, proving that even little adjustments, such as sort operations for a query execution tree can result in considerable performance gains [28]. For instance, Robert Taylor in his thesis about query optimization for distributed database systems stated that the query optimizer is often regarded as a database management system's most significant component. Taylor provided a cost model that can be used to identify inter-operator parallelism opportunities in query execution plans. This enables for a more accurate estimation of a query's response time. He proposed the novel multilevel optimization algorithm framework namely Distributed Multilevel Optimization algorithm in order to maximize the system resources

availability by spreading the optimization step among many optimization sites [29]. Query optimization in distributed database system is also studied by Raipurkar and Bamnote using optimal algorithms as the foundation for the creation of their own algorithm for processing general query called Semi Join Based Query Optimization and SDD-1. For assessing the response time of a query across global relations fragmentation, the query processing subsystem has been integrated into a distributed database management system. The result of their study shows that network traffic in distributed databases is reduced because local queries have run faster. Query optimization lowers communication costs and hence improves distributed query response time [30]. Mishra and Singh also work on designing optimized query plans for processing distributed query. In their study, they attempt to build such optimal query strategy using the Teaching Learner Based Optimization (TLBO) approach that will not require parameters. For the multi-objective constrained and unconstrained benchmark problems, the TLBO algorithm outperformed the other optimization techniques. The experimental comparisons of this approach for generating distributed query plan with a multi-objective genetic algorithm reveal that the TLBO-based algorithm can generate substantially better quality query plans for a larger number of relations [31].

8. SYSTEM DESIGN

With the suitability of RAD methodology and the use of Use Case technique in system analysis, X-PRESS has been designed based on the Use Case Diagram in Fig. 2.
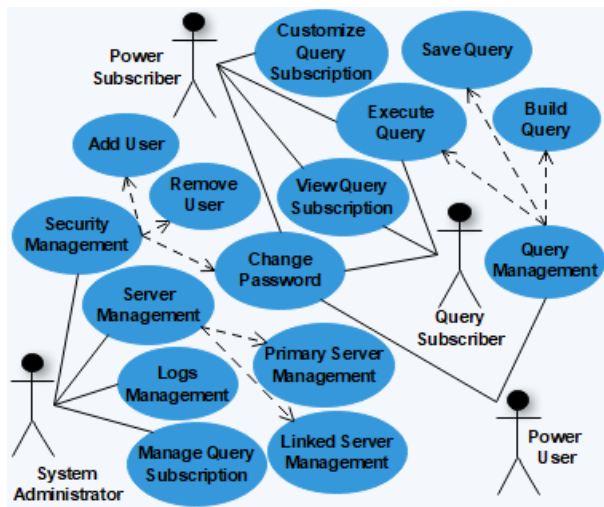

Figure 2: Use Case Diagram

There are four types of user in X-PRESS, the System Administrator, the Power User, Power Subscriber and the Query Subscriber. The System Administrator from IT Department performs system configuration and entry of information such as user, linked server and subscription information as well as manages system logs. Power User from IT Department creates query. Query Subscriber access their query subscription, run it and get the query result. Power Subscriber can do the same thing as the Query Subscriber but with added privilege to customize the subscribed query like dynamic picking of fields to be included in the query result.
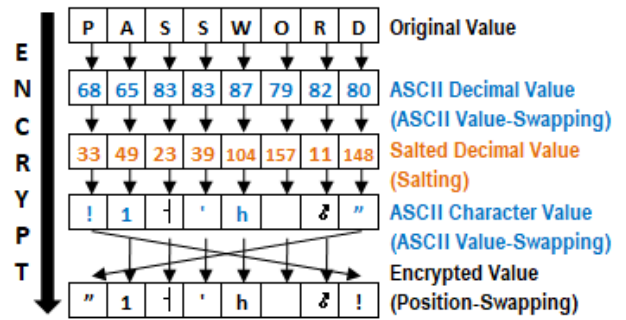

Figure 3: Encryption Algorithm

The vital part of the system design is to secure sensitive information using encryption algorithm and to optimize query execution using the hierarchical execution algorithm.

Three-way algorithm which involves ASCII Value Swapping, Salting (using key and exponents) and Position Swapping is combined to create the encryption algorithm in Fig. 3. This is an improvement to the common ASCII Value Swapping algorithm to create a strong custom-made cryptographic algorithm for X-PRESS. The same algorithm still applies to the decryption process in reverse direction. This encryption algorithm is used for encrypting sensitive information like password for connecting to the primary server, user password and linked server password [32].
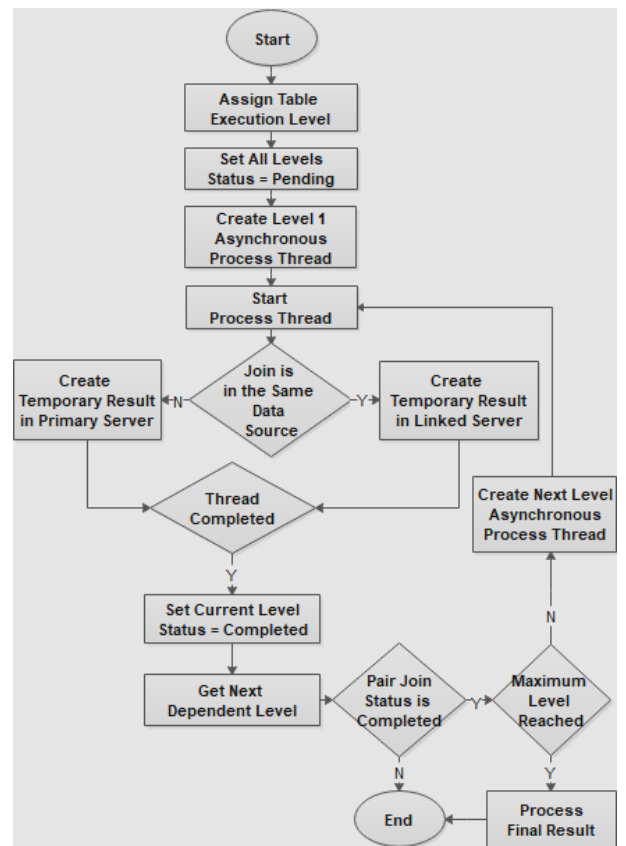

Figure 4: Query Execution and Optimization

The hierarchical execution algorithm for query execution and optimization is used in dynamic query formation and faster data access as represented by Flowchart in Fig. 4. This algorithm performs calculation to assign the table execution level and generate a sequence of asynchronous execution arranged in order of the assigned level. This will also perform data processing to create temporary and final query result and

decide as to where the query result will be created either in Linked Server or Primary Server.

Hierarchical Execution Algorithm is highly dependent with the execution level which is set during query creation because it determines the query execution sequence at run-time.
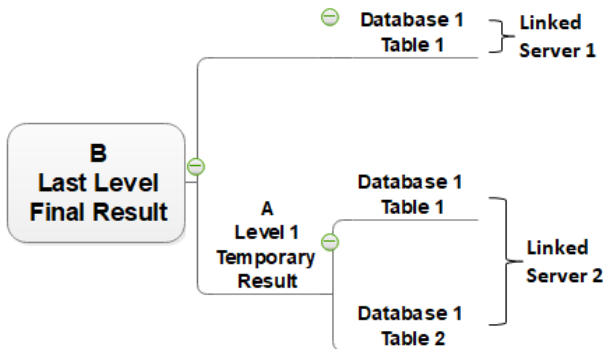

Figure 5: Query Execution Hierarchy

A sample view of query execution hierarchy is shown in Fig. 5. In order to perform cross-database queries, levels are assigned to the temporary results based on the table dependencies originated from multiple Linked Servers. These levels will determine the path and sequence of execution going to the last level to obtain the final result.

9. RESULTS AND DISCUSSIONS

X-PRESS has been developed into an efficient system for identifying data sources and extracting data from the cross-platform databases that saves time and effort as well as for distribution of data through user subscription. The system is tested for dynamic data querying from cross-platform databases and results shown that data can be generated from different relational database systems at once. Tests are made against multiple data sources and have shown that it is capable of extracting data at a satisfying speed. Compare to manual extraction that uses several tools; it has been shown that XPRESS became effective all-in-one tools for extracting data. X-PRESS is also tested for accuracy at the Query Subscriber side and shown that the query result is the same as the result generated from the Power User side.

9.1 Data Source Identification

To evaluate the capability of X-PRESS in identifying data sources, two major databases in totally different environment, operating system and server location is used to generate the top 100 members purchase data during the five days sales from September 25 to 29, 2019 in Bonifacio Global City (BGC) branch. Members information is coming from IBM DB2 database installed in IBM OS/400 operating system running in AS/400 machine located physically in Makati data center while members purchases is coming from Microsoft SQL Server database installed in Windows Server operating system located physically in Manila data center.

Test made shows that members information data sources is identified through MEMBERS Linked Server whereas members purchases data sources is identified through SALES Linked Server as shown in Fig. 6. Each Linked Servers were able to provide the database that contains the data as well as the underlying tables and fields. The raw data preview in Fig. 7 signifies that data sources have been configured and identified successfully.

Figure 6: Data Sources

Figure 7: Data Preview

9.2 Data Extraction

To evaluate the capability of X-PRESS in extracting data, the Query Analyzer module is used to design the data extraction query for the top 100 members with outstanding purchases during the five days sales in selected location using the previously configured data source. The query is ready to be called from subscription. Fig. 8 shows the result after extracting the data. Note that query masking feature is applied in end part of the Member Name for data protection.

Figure 8: Query Result

9.3 Data Subscription

To evaluate the capability of X-PRESS in distributing data, a Subscription module is developed for data subscription. Test made in Fig. 9 shows that users can view their subscription and analyzes the data anytime via self-service.
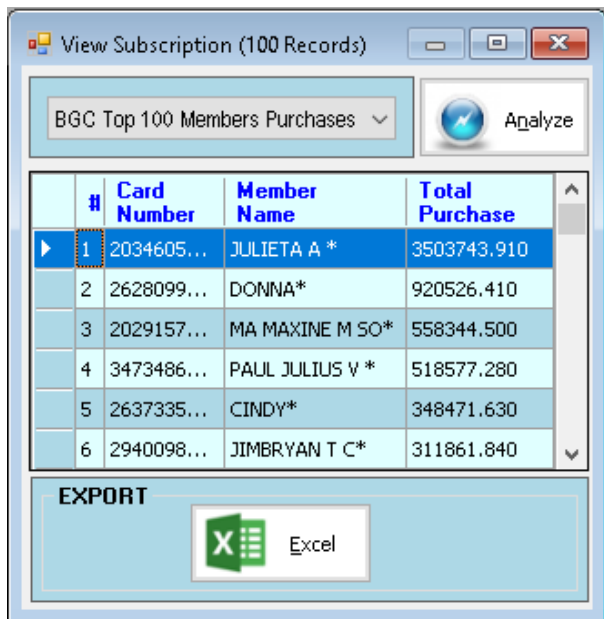
Figure 9: Viewing and Analyzing Subscription

9.4 Efficiency

To measure the X-PRESS efficiency in terms of processing time, the top 100 members purchases is extracted manually through different steps to compare the time savings. These steps involve querying the data separately, save the initial results in Excel individually and merge again in Excel one by one to produce the final data. This manual method on actual preparation took 40 minutes to complete. By using X-PRESS, the one-time data source setup is done in approximately 19 minutes and the query consistently runs for only 1 minute to generate the result as shown in system logs in Fig. 10.
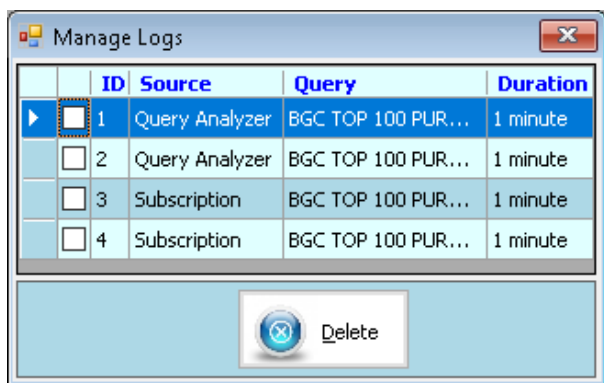

Figure 10: System Execution Logs

X-PRESS gives 50% process improvement by cutting the time in half from 40 minutes of manual process to 20 minutes based on the comparison in Table 1.

Process improvement is computed by getting the Percentage Decrease which is equal to the Absolute Value of (Final Value-Initial Value)/Initial Value using the formula **ABS((FV-IV)/IV))**, wherein Initial Value (IV) is 40 minutes and Final Value (FV) is 20 minutes.

Table 1. Manual Process Versus X-PRESS

BGC TOP 100 MEMBERS PURCHASES			
MANUAL	**Duration (minutes)**	**X-PRESS**	**Duration (minutes)**
Extract data via multiple database tools one by one	20	Data Source Setup (one-time only)	19
Export/Save the extracted data in Excel one by one	10	Analyze	1
Merge data in Excel	10	Export to Excel	0
TOTAL TIME	40		20
IMPROVEMENT	**% Decrease = ABS((20-40)/40))** **% Decrease = 50**		

10. CONCLUSIONS

The use of different technologies, systems and databases and interconnect them in a single application is a challenge. Extracting and analyzing data was previously possible with the use of different tools manually and sometimes unproductive. It is more challenging to perform a cross-database query at once that is efficient in saving time, effort and resources. It is also hard to make data available to users at anytime. After developing X-PRESS, it is therefore concluded that these are achievable and those limitations can be overcome. With Query Analyzer that can seamlessly build simple to complex queries from cross-platform databases using hierarchical execution algorithm, and with X-PRESS all-in-one data access solution that is optimize for dynamic data querying, the system is able eliminate the use of several tools. The system also provides accurate data querying facility for users through subscription.

X-PRESS can play an important part in any organization and is proven useful in company like S&R Membership Shopping in dealing with data. Human are no match with data but with X-PRESS it brings a perfect match. It doesn't matter where the data are located and how big it is. With this single tool that can connect and use different data easily, anyone can be empowered to utilize the full potential of data. X-PRESS ability to merge different data sources at the maximum speed by simplifying any complexities including its other features is what the system has to offer.

11. REFERENCES

[1] K. Cukier, "Big Data and the Future of Business," 2015. https://www.technologyreview.com/s/538916/big-data-and-the future -of-business/.

[2] T. C. Redman, "Does Your Company Know What to Do with All Its Data?," 2017. https://hbr.org/2017/06/does-your-companyknow-what-to-do-with-all-its-data.

[3] M. Theriault, "6 Ways to Transform Data into Real Information That Drives Decision-Making," 2017. https://www.allbusiness.com/transform-data-real-information drives-decision-making-16096-1.html.

[4] S. Repiso, "Case study: The process of transforming data into useful information," 2015. https://nae.global/en/case-study-the-process-oftrans forming-data-into-useful-information/.

[5] D. Mitzner, "What is a data-driven company?," 2016. http://www.infoworld.com/article/3074322/big-data/what-is-a-datadriven-company.html.

[6] R. Shaw, "What is Business Intelligence?," 2011. https://www.dbta.com/Editorial/Trends-and-Applications/What-isBusiness-Intelligence-73502.aspx.

[7] K. M. Cresswell, D. W. Bates, and A. Sheikh, "Why Every Health Care Organization Needs a Data Science Strategy," 2017. http://catalyst.nejm.org/healthcare-needs-data-science-strategy/.

[8] S. Leung, "Fostering an Analytics-Driven Culture," 2014. https://www.tibco.com/blog/2014/07/29/fostering-an-analytics driven-culture/.

[9] A. McAfee and E. Brynjolfsson, "Big Data: The Management Revolution," 2012. https://hbr.org/2012/10/big-data-themanagement-revolution.

[10] C. Carande, P. Lipinski, and T. Gusher, "How to Integrate Data and Analytics into Every Part of Your Organization," 2017. https://hbr.org/2017/06/how-to-integrate-data-and-analytics-intoevery-part-of-your-organization.

[11] B. Wagner, L. Latham, and M. Wenzel, "C#," 2017. https://docs.microsoft.com/en-us/dotnet/csharp/csharp.

[12] Y. Ono, S. Ono, H. Yasunaga, H. Matsui, K. Fushimi, and Y. Tanaka, "Clinical characteristics and outcomes of myxedema coma: Analysis of a national inpatient database in Japan," J. Epidemiol., vol. 27, no. 3, pp. 117–122, 2017, doi: 10.1016/j.je.2016.04.002.

[13] L. Myatt, J. M. Roberts, and C. W. G. Redman, "Availability of COLLECT, a database for pregnancy and placental research studies worldwide," Placenta, vol. 57, pp. 223–224, 2017, doi: 10.1016/j.placenta. 2017.07.014.

[14] A. Kretser, D. Murphy, and P. Starke-Reed, "A partnership for public health: USDA branded food products database," J. Food Compos. Anal., vol. 64, no. July, pp. 10–12, 2017, doi: 10.1016/j.jfca.2017.07.019.

[15] M. Willemet, S. Vennin, and J. Alastruey, "Computational assessment of hemodynamics-based diagnostic tools using a database of virtual subjects: Application to three case studies," J. Biomech., vol. 49, no. 16, pp. 3908–3914, 2016, doi: 10.1016/j.jbiomech. 2016.11.001.

[16] P. Garg and P. Jaiswal, "Databases and bioinformatics tools for rice research," Curr. Plant Biol., vol. 7–8, pp. 39–52, 2016, doi: 10.1016/j.cpb.2016.12.006.

[17] A. R. Caliñgo, A. M. Sison, and B. T. Tanguilig III, "Prediction Model of the Stock Market Index Using Twitter Sentiment Analysis," Int. J. Inf. Technol. Comput. Sci., vol. 8, no. 10, pp. 11– 21, 2016, doi: 10.5815/ijitcs.2016.10.02.

[18] C. Fan and F. Xiao, "Assessment of Building Operational Performance Using Data Mining Techniques: A Case Study," Energy Procedia, vol. 111, no. September 2016, pp. 1070–1078, 2017, doi: 10.1016/j.egypro. 2017.03.270.

[19] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine Learning and Data Mining Methods in Diabetes Research," Comput. Struct. Biotechnol. J., vol. 15, pp. 104–116, 2017, doi: 10.1016/ j.csbj.2016.12.005.

[20] A. M. Ahmed, A. Rizaner, and A. H. Ulusoy, "Using data Mining to Predict Instructor Performance," Procedia Comput. Sci., vol. 102, no. August, pp. 137–142, 2016, doi: 10.1016/j.procs.2016.09.380.

[21] O. K. Oyedotun, S. N. Tackie, E. O. Olaniyi, and A. Khashman, "Data Mining of Students' Performance:

Turkish Students as a Case Study," Int. J. Intell. Syst. Appl., vol. 7, no. 9, pp. 20–27, 2015, doi: 10.5815/ ijisa.2015.09.03.

[22] B. T. Femina and E. M. Sudheep, "An efficient CRM-data mining framework for the prediction of customer behaviour," Procedia Comput. Sci., vol. 46, no. Icict 2014, pp. 725–731, 2015, doi: 10.1016/j.procs. 2015.02.136.

[23] H. A. Aboalsamh, "A Join Algorithm for Large Databases: A Quadtrees Structure Approach," J. King Saud Univ. - Comput. Inf. Sci., vol. 22, pp. 1–11, 2010, doi: 10.1016/s1319-1578(10)80001-1.

[24] D. Cohen and K. Narayanaswamy, "Using first-order logic to query heterogeneous internet data sources," Procedia Comput. Sci., vol. 62, no. Scse, pp. 170–177, 2015, doi: 10.1016/ j.procs.2015.08.431.

[25] T. A. Engel, A. S. Charão, M. Kirsch-Pinheiro, and L. A. Steffenel, "Performance improvement of data mining in weka through GPU acceleration," Procedia Comput. Sci., vol. 32, pp. 93–100, 2014, doi: 10.1016/j.procs. 2014.05.402.

[26] M. Şerban, "Methods to Increase Search Performance for Encrypted Databases," Procedia Econ. Financ., vol. 3, no. 12, pp. 1063–1068, 2012, doi: 10.1016/s2212-5671 (12)00274-2.

[27] D. J. Prajapati, S. Garg, and N. C. Chauhan, "Interesting association rule mining with consistent and inconsistent rule detection from big sales data in distributed environment," Futur. Comput. Informatics J., vol. 2, no. 1, pp. 19–30, 2017, doi: 10.1016/j.fcij.2017.04.003.

[28] P. P. Beran, W. Mach, E. Schikuta, and R. Vigne, "A multi-staged blackboard query optimization framework for world-spanning distributed database resources," Procedia Comput. Sci., vol. 4, pp. 156–165, 2011, doi: 10.1016/ j.procs.2011.04.017.

[29] R. Taylor, "Query Optimization for Distributed Database Systems Robert Taylor Candidate Number : 933597 Hertford College Supervisor : Dr . Dan Olteanu Submitted as part of Master of Computer Science Computing Laboratory University of Oxford August 2010," no. August, 2010.

[30] K. Huang, "Query Optimization in Distributed Databases.," vol. 6, no. 2, pp. 319–322, 1983, [Online]. Available: http://oai.dtic.mil/oai/oai?verb=getRecord& metadataPrefix=html&i dentifier=ADA124921.

[31] V. Mishra and V. Singh, "Generating Optimal Query Plans for Distributed Query Processing using Teacher-Learner Based Optimization," Procedia Comput. Sci., vol. 54, pp. 281–290, 2015, doi: 10.1016/j.procs. 2015.06.033.

[32] J. N. Mindoro, N. U. Pilueta, Y. D. Austria, L. Lolong Lacatan, and R. M. Dellosa, "Capturing Students' Attention through Visible Behavior: A Prediction Utilizing YOLOv3 Approach," 2020 11th IEEE Control Syst. Grad. Res. Colloquium, ICSGRC 2020 - Proc., no. August, pp. 328–333, 2020, doi: 10.1109/ ICSGRC49013.2020.9232659.