UTILIZING ANALYTICS TO DISCERN CAMPUS LIBRARY USE

Punith Etikala and Erich Rice
Masters of Science in Information Assurance
St. Cloud State University
St. Cloud, MN 56301
petikala@stcloudstate.edu
rier1201@stcloudstate

Laura Lebentritt
Cyber Security
University Of Maryland University College
Adelphi, MD 20783
lauraleigh4097@att.net

Dr. Dennis Guster
Information Systems
St. Cloud State University
St. Cloud, MN 56301
dguster@stcloudstate.edu

Abstract

Databases allow organizations the ability to process large amounts of data. These vast repositories have created new areas of study related to data analytics. This has allowed easy access and manipulation of data to determine useable business information or business intelligence (BI). The authors were given a dataset that encompassed student and staff use of physical library assets at St. Cloud State University over several semesters. The data was gathered from three very large-flat files without any relational structure. From these flat files a database in SQL Server 2014 was created. Then the authors were able to develop a multi-dimensional cube using the Visual Studio BI, which allowed for greater analysis of the data. The impetus for analyzing the data was to try and discern what resources students and faculty were using, when they were using them within a semester, and how their performance (by analyzing their grades) were effected.

1 Introduction

The efficacy of Business Intelligence (BI) as a means to support intelligent decision-making has been well-accepted [1]. There has also been much interest in its application to higher education, as shown in Guster and Brown [2]. Usually the application of a business intelligence initiative to any organization should be a relatively simple and easily formatted procedure. However, in some organizations differences in management styles, and varying expectations sometimes can limit the effectiveness of BI [3]. Guster and Brown observed the impact that office politics often play in higher education in their work, and they found that it had a major impact on BI components from the data dictionary definitions to the way that data was eventually stored [2]. Because of this, one might expect implementing BI in higher education might be more challenging than in a private enterprise where managers often have more say in how things are accomplished. However, there currently exists a balance between priorities and resource requirements for BI to be successful, and BI initiatives need to produce results to continue to be a valuable priority for any organization.

Challenges arise when short cuts are presented which may create long-term resource costs and distract from building out an enterprise architecture framework that would allow true BI utilization. An example would be in building out a statistical model based on static data sets without initially investing in the statistical software to link to a Data Warehouse, which would allow provisioning of dynamic data sources for modeling. Perhaps a good way to look at BI is to view it as a tool that can be used to make more informed decisions based on the available data, which has been massaged into useful information. Another useful analogy of BI might be the building of a house, you could build a house with hand tools, but you could build the exact same house far more efficiently with modern power tools. The use of BI (and other analytics) would be analogous to using the same power tools to construct the house [4]. However, the success of devising effective analytical tools within a BI structure is dependent on many things. First and foremost, there must be an accurate and readily available data source, which is often in the form of some type of a relational database.

The old adage: garbage in garbage out is most appropriate here and can be especially disconcerting when dealing with an Enterprise Resource Planning software such as SAP which requires very specific naming conventions and data types. Also, implementing BI provides a very efficient structure for designing and implementing an information system, but it is dependent on the underlying business logic of an organization, which in this case is an educational institution. The implementation of the wrong business logic can be fatal to the BI process. The work of Schonberg presents the essence of this concern, which is for BI to be successful the organization, needs to determine what behavior indicates success and then build on that to create a useful end product [5]. Furthermore, that behavior must be quantifiable and recordable in such a way that valuable data can be extracted in a timely manner to support the analytical strategy that has been developed by the BI platform. If all of these conditions are met then BI can be effectively used to make intelligent decisions that will enhance the success rate of the decision making process of the organization in question.

To create an effective decision making strategy one has to realize that true BI involves more than just generating a large number of reports. It must instead focus on evaluating competing priorities, as in the case of educational concerns maximizing student retention while at the same time shortening time to graduation [6]. In terms of what type of educational management is appropriate for analytical analysis or BI the literature states that it is appropriate in any facet related to ensuring operational success of the institution [7]. More over it has been stated that analytics or BI could help educational leaders cut costs and improve teaching/learning through more useful and usable information [8]. Also, that the use of predictive analytics within BI could improve efficiencies and save money while also enhancing student achievement. Some specific examples include: planning courses, recruiting/retaining college students, optimizing the scheduling of classrooms and courses, and also maximizing alumni donations [8]. Whatever the target management function for analytics or BI, for it to be successful the process needs to be understood and well documented. In the early days of "data gathering" a common way to describe the flow of logic was flow-charting or data-flow charting. While this technique has been revised and optimized the basic logic is still fairly sound, it shows the flow of data through an organizations business processes. Further, one often needs to coordinate the logic in regard to how the data is stored. Usually one would expect the data to be stored in some type of database format and thereby it could be categorized as some type of data mining procedure. Thus, it is important to select the data mining technique that will most optimize the analytical formula or BI process being utilized.

Another critical point in devising an effective analytics strategy is to understand the limitations of the silo approach. In other words, the BI algorithms must be implemented on an enterprise level so as to allow for the most useful range of knowledge distribution. Often models are devised, implemented and/or constructed on an individual, departmental or college level, which can severely limit the usefulness of the information gathered from the BI process. Often the silo-based approach does not take into account the complex model inter-relationships, possible model correlation and covariance, as well as other independencies that can confound the resulting analysis of the BI process [9].

Given that the analysis team is often prepared to undertake complex data analysis on the enterprise architecture level, decisions need to be made about the modeling techniques before hand. One of the first decisions is "how dependent will the BI team be on existing software packages"? There are certainly some viable options available that have proved effective. The work of Bao-sheng and Xin-quan is a good example, they used Statistical Analysis System (SAS) Enterprise Miner to generate and operationalize predictive models for a Tele-Com Company [10]. Starting with a software standard such as SAS Enterprise Miner has the advantages of having many well-proven algorithms already pre-programmed within it, while also still having a framework from which to devise and implement your own algorithms. It is clear that a well-organized underlying database framework is crucial, without which most BI initiatives will fail in short order. Typically, the volume of data renders basic human visual analysis or traditional tools ineffective [11]. Thus, more specialized BI algorithms and the ability to customize those algorithms are often needed when devising successful BI analytics in any organization and especially within an educational environment.

Often it is the proprietary design factor of the primary source data that precludes mainstream analytic vendor options, especially when trying to utilize SAP, thus often predisposing a homegrown solution or requiring significant work to “cleanse” the data to make it acceptable to use in an off the shelf software application. Another key factor to consider when selecting a build vs. buy decision relates to how well the business logic and the analytics developers understand data definitions. Well-defined data definitions and ensuring business practices follow a standard definition that suggests a significant buy-in by management should be required when developing the BI process. Extreme caution should be used when implementing a build solution to ensure designs are following current best industry practices to allow for sustainability of solutions and potential future vendor options as they may become available.

Because of the need to utilize underlying database frameworks the traditional approach to data analysis for decision support in which domain expertise is coupled with statistical modeling techniques to generate handcrafted solution needs must be expanded [4]. Specifically, a Knowledge Discovery Database or (KDD) is often recommended [4]. This approach is designed to increase the availability of large volumes of data, facilitate the rapid deployment of data-driven analytics and deliver the results in a format easily understandable to end users.

Given this backdrop, the task for the authors was to try and develop a useable BI framework to make sense of the use of St. Cloud State University Library physical assets, such as books, videos, journals, and rooms. While the dataset that was provided to the authors only provides a small snapshot of the overall utilization of the Library’s resources, as the majority of resources are now utilized through electronic means, it is none the less a good test case for the use of BI in this area. Through the use of this dataset and related BI analysis it was hoped that statistical correlations could be made between Library use and student academic performance, as well as how the utilization of Library resources changed throughout the time a student was enrolled and progressed from a freshman to a senior.

2 REVIEW OF LITERATURE

2.1 BI Background And Databases

Business intelligence is a very complex and resource intensive process, both due to hardware and software constraints but also human work hours as well. Algorithms need to be developed and data needs to be placed into a useable format, dependent on what application will be utilized to do the BI analysis. William and Williams stated that key vendor BI initiatives include: (1) providing pre-integrated BI product offerings generally known as packaged analytical applications; (2) advocating expansion of the BI footprint within organizations, often referred to as BI for the masses and/ or enterprise analytics; and (3) positioning the use of their products as reflective of BI best practices [12]. Thus the need to tailor BI software to the backend systems that are running, such as databases or data warehouses preferably, which will hopefully allow for a more smooth and integrated approach to utilizing the data that has been gathered. The need to utilize BI for most medium to large enterprises has grown exponentially as the ability to house ever larger amounts of data has grown. Matching the BI software to the enterprise is crucial because today it is difficult to find a successful enterprise of moderate size that has not leveraged BI technology for its business [13]. While the use of BI in commercial and government enterprises has become well entrenched, its use in academic and educational enterprises is less universal. Attempts have been made to better utilize BI applications, even within academic library settings such as described in Zucca [14], though the potential pitfalls and difficulties in utilizing the "data: by the complex and ornery problem of harvesting, structuring, and storing the vast troves of activity data resting dormant in the systems libraries all use to conduct business".

After the development of a team of personnel is in place and the appropriate BI software has been identified it is important to devise an effective strategy of data collection and storage. Ramakrishnan, et al., have investigated the factors influencing such strategies [15]. Specifically, they found that the strategies could be classified as either comprehensive in nature or problem driven. They categorized the business intelligence purpose as adding insight, consistency or to be transformational in nature. This of course illustrates the potential problems with BI, which is a tradeoff between how quickly the system needs to come on line and how well it is planned out, and the issues regarding the backend system planning and front-end applications that are available. Much too often an enterprise's goal is to get the BI system developed quickly, at the potential cost of being less cognizant and relevant to future applications and able to scale to future requirements. A data warehouse design methodology would be the best methodology to pursue, as it would allow for the greatest access to an enterprise's data, though it depends on the commitment, priority level, and the skill level of the people developing the system. So again the entire team needs to be involved in the process of identifying, storing, extracting and analyzing the data as well as the end result of utilizing the data to provide worthwhile business intelligence.

The first step will involve whether data will come from disparate sources (individual databases) or be organized into a data warehouse structure. The way in which data is organized will have a dramatic effect in the success of the BI system [16]. It is clear that the data has to be accurate, consistent, complete, valid and timely [17]. Furthermore it is important to note that a data warehouse often exposes data quality issues inherent in the original data systems. Thus the great importance of data cleansing throughout the entire Extract Transform Load (ETL) processes cannot be over stated, and must be done throughout the development process. If this cleansing doesn't take place prior to the data being made available to the end-users for the business intelligence system, then the data will be inaccurate and outdated and confidence in the system could potentially wane. Once confidence begins to be lost it is often very difficult and impossible to regain.

When potential data quality issues appear during the ETL phase, it is recommended that these steps be taken: define the data quality requirements, profile, analyze and assess the data quality, define data metrics, define the data quality business rules, clean and correct the data defects, and implement process improvements so that the data defects can be avoided in the future [19]. Often the data will be coming from disparate sources, in which case it is important to be aware of any interdependencies within the data as it may ultimately influence data quality. Also, it may be useful in providing a quality check on the data and allow for compliance and auditing after the fact. For example, Guster and Brown [2] state that referential integrity events can be used as a validation mechanism before releasing the data from disparate sources to the data warehouse and dependent reports, which might provide better quality data.

Prior to a data collection strategy beginning in earnest serious thought needs to take place in regard to the refresh rate of the data provided to the database or data warehouse and if it is not sufficient then what needs to be done to increase the transfer rate. Because it depends on the business, or in this case a university library, the need for useable data may be somewhat lengthier in nature. However, for a business in e-commerce it is critical that the data is as close to real time as possible to avoid inventory shortages resulting in the loss of valuable potential sales and the potential loss of consumer confidence given the nature of the company's business model. For higher education institutions, the granularity can be much larger, though the potential needs can be far finer or granular depending on the need of insight that is necessary. In normal working circumstances records need to be current only to the previous day and it is common to select a daily refresh strategy for the data warehouse at SCSU. Execution of MiH [20] states the reasons for this common refresh rate as:

- Integrating disparate sources conforming to differing reconciliation schedules.
- Avoid ETL processing during peak business usage.
- Common refresh rates are desirable to ensure uniform scheduling of updates to relationally constrained entities.

Due to the fact, some businesses practice requirements are more stringent than these refresh rates, sometimes within these instances it is important to recognize the impact of increasing the refresh rate and to ensure system performance (i.e. ETL cycle time needs to be increased) which then can meet the increased service level demands on the system. Often a good example of scheduling data refresh cycles would be to process employee expenditure reports at an evening or after-hours time so as to limit the impact on other data manipulation events. However, even with an indexed dedicated reporting database, running ETL during peak load can often restrict business practices. Thus, the need to leverage differential ETL processing can alleviate some of the processing performance constraints within data warehouses. Thus, determining the number of refresh cycles per day or per hour should be tied to the strategic objective associated with how the data will be used and for what purpose.

Thus requiring a student to wait until the following day to access course readings could have an adverse effect on student success and the reputation of the institution, though in all likelihood it would not. Alternatively, data supporting budgetary planning reports would more than likely have a slower than desired refresh rate of once per week. This would then allow adequate time to for review and revision and potential changes that would need to be made. Another important concern would address where the data should be housed both short and long term, given the security concerns and business needs of the institution. This decision would have major ramifications regarding data extraction with regards to accessibility and performance.

2.2 Cleansing The Data

The literature clearly indicates that there is a great need to clean the data, and it is an integral part of the BI process if it is to work properly. It has been reported that there are three commonly dealt with problems: first, the absence of universal keys across different databases or the object identity problem; second, the existence of data entry errors in the dataset; and third, the inconsistencies in data being inputted from multiple sources [19]. To try and combat these problems Florescuand suggested a framework for data cleansing as a directed graph of data transformations whereby, the transformations are placed in four classes: mapping, matching, clustering and merging. Although this basic logic is implemented via something like a Structured Query Language (SQL) script, there still needs to be some form of human interaction on some level [19]. Lastly, the issue of performance needs to be considered, for as the size of the data increases it can become much more difficult to keep up. If the data comes from multiple sources and be used by multiple applications, then data errors not properly cleansed can affect the validity of analytical evaluation across an entire enterprise [23]. The concept of a “data quality broker” is offered as a way of tracking data stored on more than one source, this “data quality broker” would then evaluate the validity of each data source and ensure that the most accurate source was used [23]. Once again the old adage “garbage in – garbage out” applies, many times this is due to upper level decision makers not understanding the BI process and thereby placing greater resources in report generation. Because this offers a tangible element they understand it can often mask the short changing of the data warehousing and cleansing processes.

2.3 Modeling And Reporting

BI and the goal of taking all of an enterprise’s data and deeming it “big data” greatly increased the need for adopting a sound analytics strategy [24]. Usually big data is understood to be unstructured, unlike most of the warehousing frameworks, which are based on relational database structures. This can also be true in Online Analytic Processing (OLAP) type models. An example would be an OLAP cube requires that every dimensional element in an existing fact table exist in the associated dimension table. For instance, when a referenced student ID does not exist, or is not recognized, then the cube will fail to process unless adjustments are performed which would be ill advised. It is also important to zero in on the pertinent variables to what you are trying to get at from your BI analysis, rather than using a “kitchen sink” methodology whereby every potential variable is included [25]. The “kitchen sink” method tends to cloud the modeling process and adds noise when undertaking regression analysis testing. Over time, as modeling techniques prove to be effective, it is then possible to add additional variables or modify existing variables within the models.

Research indicates that statistical analytics is becoming an integral function within the BI process and that the sophistication of these analytical functions continues to improve [26]. It is also important to address both individual and group trends within the datasets under analysis, as well as consider the granularity of the data [27]. In the context of higher education, especially in regards to library data, BI means that the characteristics of the individual students need to be evaluated as well as the characteristics of the cohorts (groups of students who begin their studies together, i.e. as freshmen) to which they belong. No matter of how well thought through the analytics, it is critical to present the results in a user friendly format that can be easily digested by end-users, thus the necessity for a well thought out reporting apparatus.

Although this is often viewed as the last step in the BI process, and sometimes overlooked, it may be the most important step in the process. Although, the large-scale nature of many BI implementations can make this a difficult step to actually achieve, it is fundamentally imperative that it is done well to provide usable insight to those who have sponsored the program [28]. Further, it is critical that the reports are in a format that the end-user can understand, this means the BI staff needs to be flexible in meeting individual user needs or all prior work could be in vain. This attention to detail in matching the reports to the end-user "personalized reporting" was noted in Ulanov, et al. [28]. Certainly there are numerous options for generating these reports, as there is even a wealth of open-source reporting tools, such as the many packages that have been built within the R statistical environment [29].

3 Methodology

The data that the authors were given to analyze was generated from a log file, which contained three very large and flat files, they included files named StudentRetention, CirculationLog, and Majors. These files were very flat, and held an inordinate number of attributes, over 750 between the three files. As there were no relations or relational constraints between the files it made the process of culling useable business intelligence data from them nigh on impossible. Attempts were made to work with the data in spreadsheet applications, though the ability to tie the files together and massage useable intelligence from them proved very difficult and laborious. Thus, the authors were tasked with taking the large and cumbersome flat files containing hundreds of attributes and breaking them up into more logical relational tables that could be placed into a relational database which would allow for far easier manipulation and analysis of the data. For this purpose it was decided to utilize Microsoft SQL Server 2014 with the Business Intelligence platform to aid in the analysis of the data once it was organized and cleansed.

The first step was to take the flat file data and try and break it up more logically into a relational structure that would be better suited to the relational constraints of the SQL Server database. The relations were brought up to a second normal form, as any greater normalization would have greatly degraded the processing of a dimensional cube, and the processing would have slowed especially when dealing with large amounts of data. Once the data was housed in a relational database format it allowed it to be managed and manipulated with the Microsoft Visual Studio Business Intelligence application. Creating a multi-dimensional cube within the BI package requires four main things: (1) Data Sources, this creates the connection to the relational database that has been created from the data provided; (2) Data Source Views, this makes use of the source creation that was created in the first step, while included the required tables within the database that will be needed to create the data cube, please see Figure 1 below;

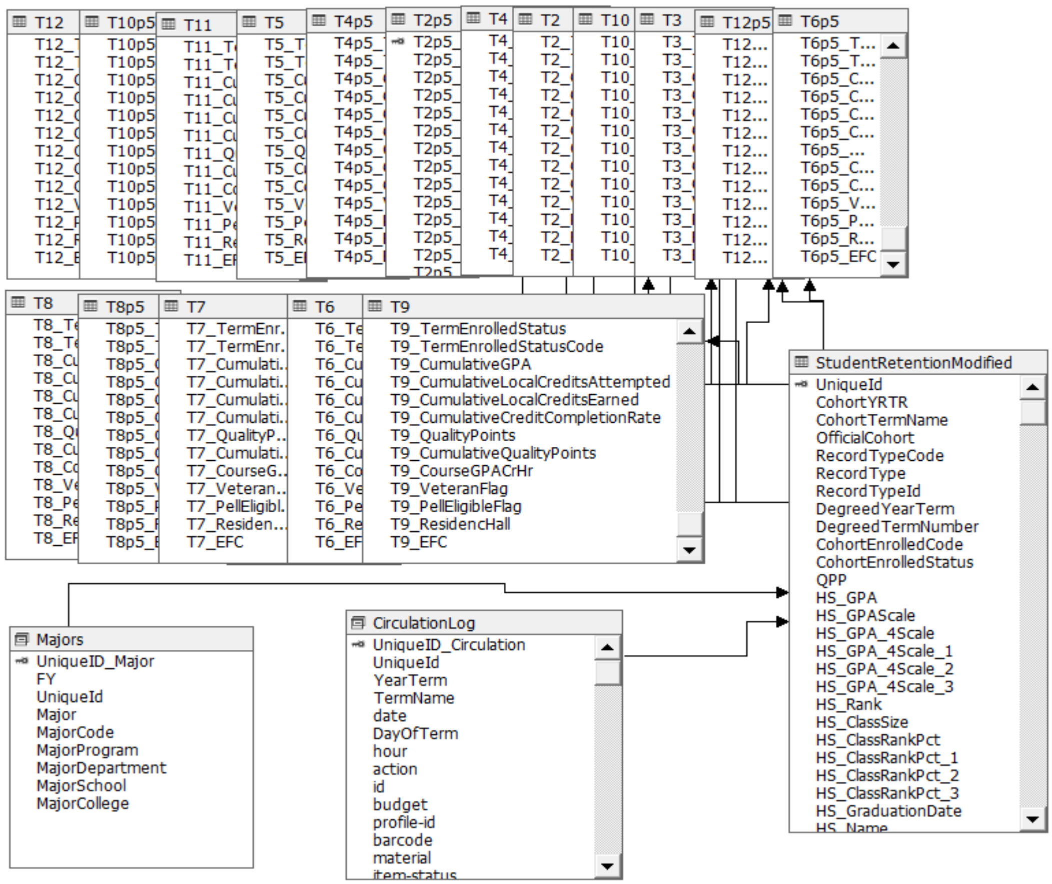

Figure 1: Data Source View for LRS Circulation

(3) Data Cubes, again the database connection needed to be in place, then the tables that were to be used need to be selected and loaded into the cube, this is done through the Measure group tables option, once the group tables are selected then it is necessary to identify and select the required dimensions of the cube, after this is done the multi-dimensional cube can be created, please see figure 2 below; (4) If the prior processes are successful, then it should be possible to deploy the cube and process the results in the cube browser view, please see Figure 3 below for an example of the cube browser view.

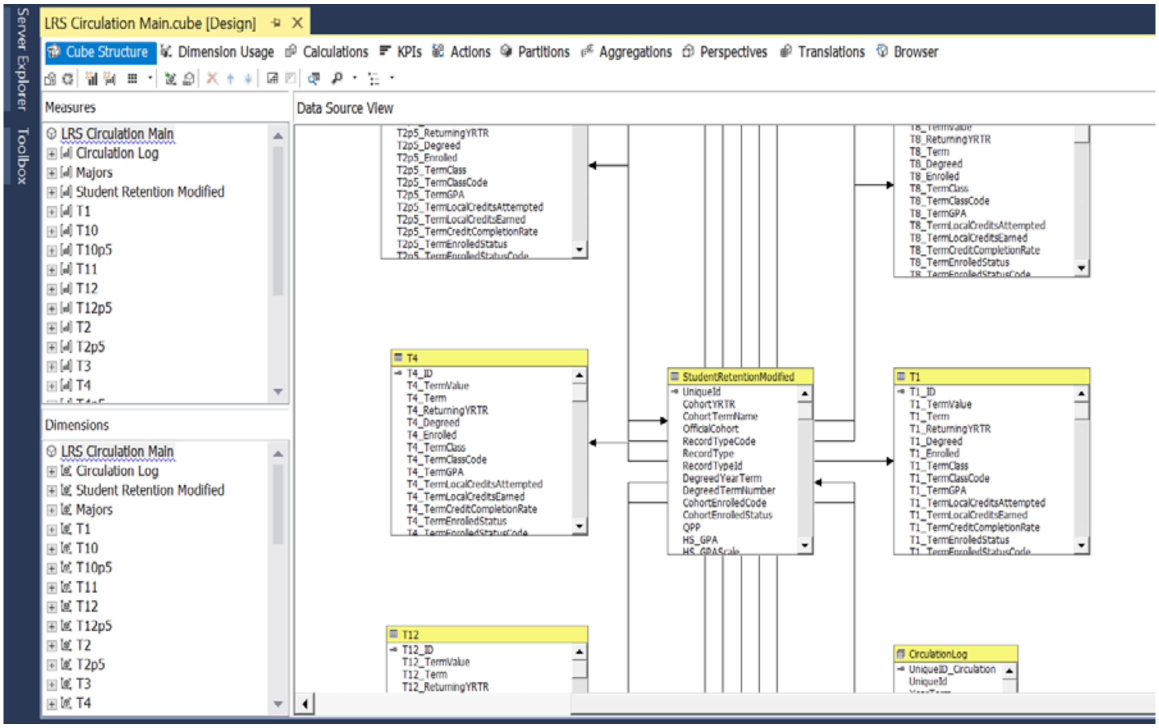

Figure 2: Library Resource Circulation Cube Design

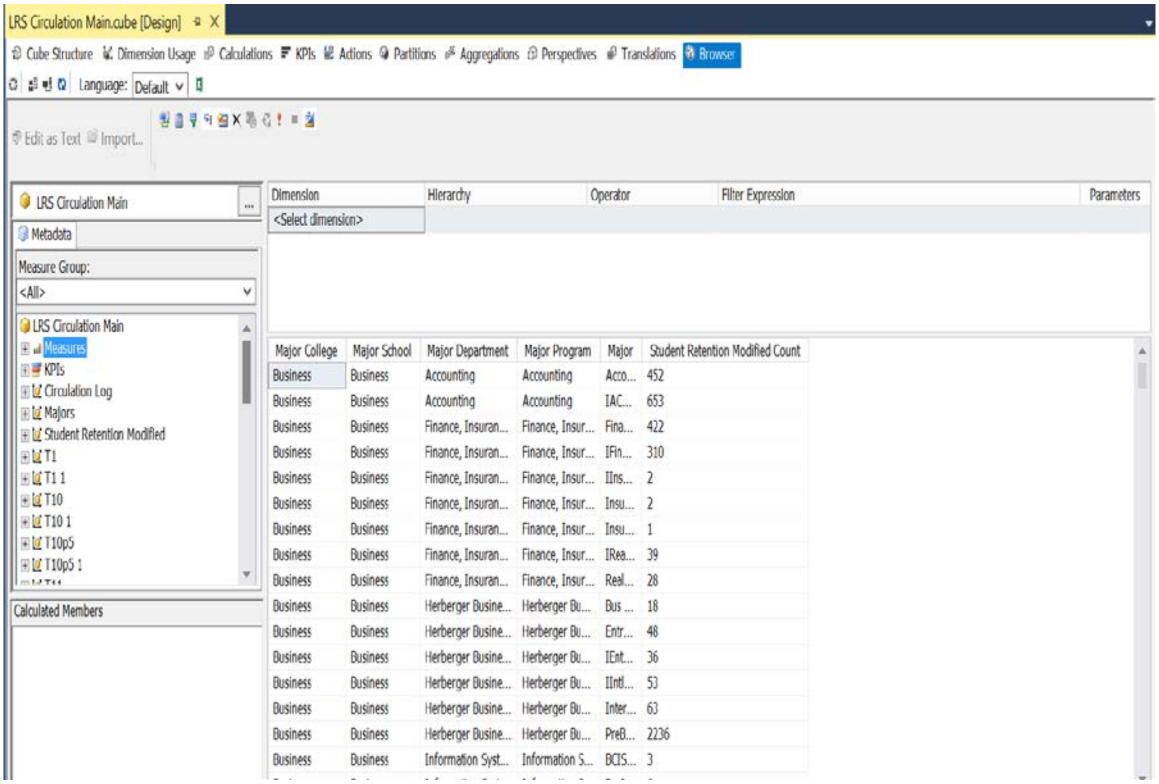

Figure 3: Cube Browser View Number of Students by Major

It is also possible to create dimensions and measures separately as well, this allows for the possibility to create hierarchies within the cube structure. For example, within the dataset that the authors were given there was a Majors file, this held the data regarding the students declared or intended major of study at St. Cloud State University. As can be seen in Figure 4 below, the Majors dimension table allows for the ability to create hierarchies between the various attributes housed within the Majors table that was created from the cumbersome flat file that previously housed the data.

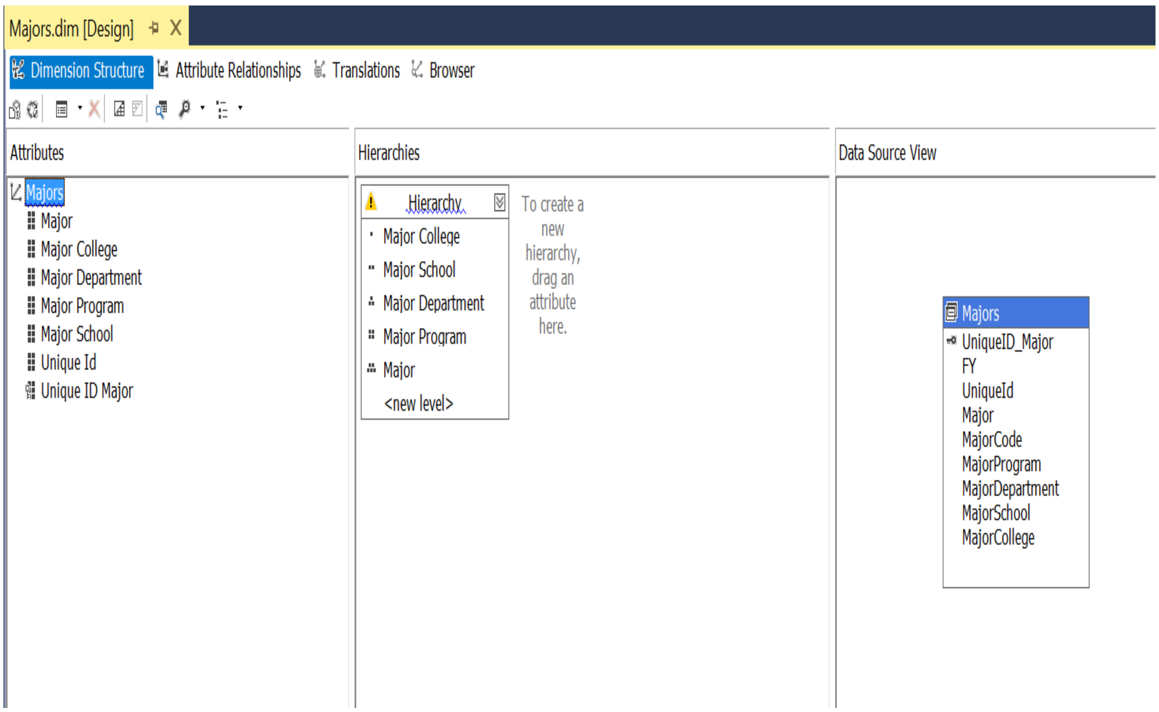

Figure 4: Majors Dimension Table

The next step was to develop queries within the BI application to begin to develop a sense of how students who used various physical library resources were performing academically. As this type of analysis is the ultimate goal of developing out databases and other data manipulation tools, the work in this area is ongoing for the authors and the project sponsors. Some of the queries that have been developed thus far within the Microsoft BI application look at simplistic usage of resources within declared majors across the time period that data is available, which equates to roughly 13 terms, or roughly four calendar years as a term encompasses fall and spring semester as well as summer term. The data continues to be collected so the database that has been developed by the authors will continue to grow in size and hopefully will be able to provide more actionable information as the picture provided by the data becomes more clear. As can be seen in Figure 5 below, the ability to query the database and provide visually clear and understandable information to high-level decision makers is a valuable resource for any enterprise to have in its toolkit.

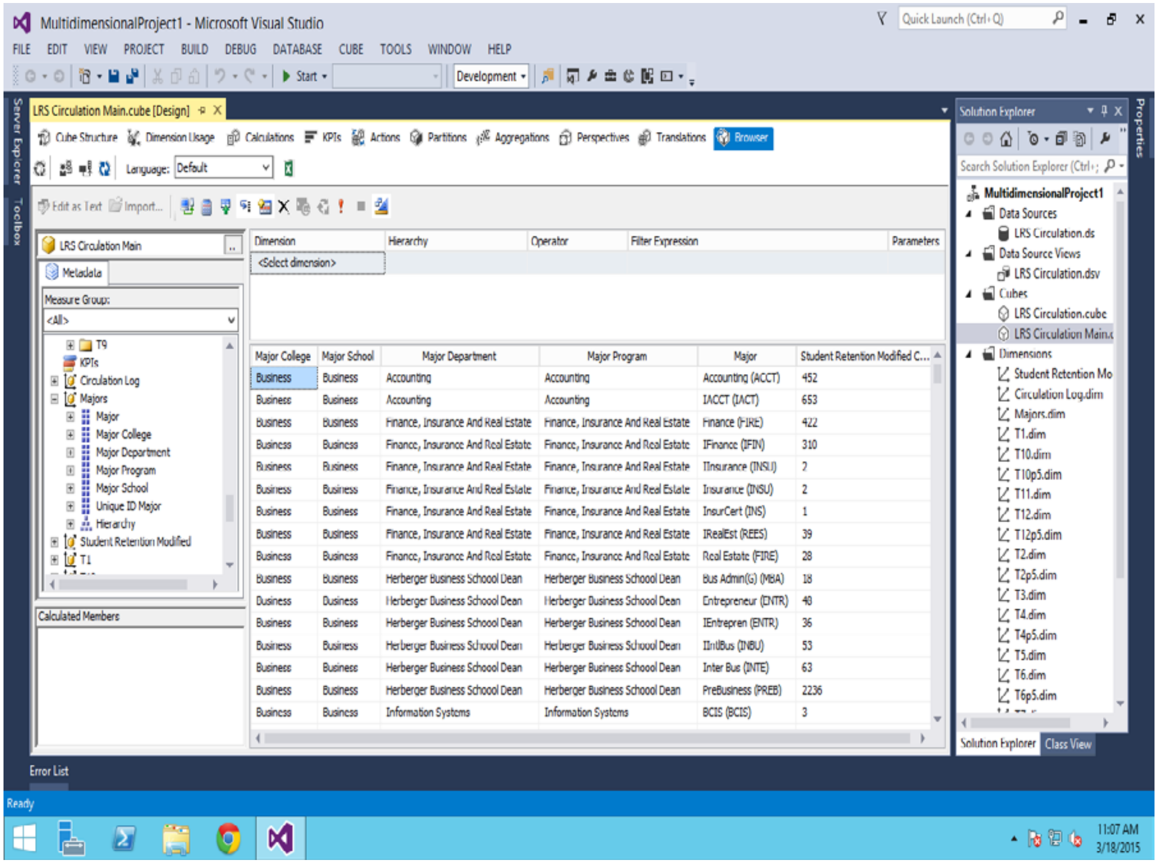

Figure 5: Microsoft Visual Studio BI Query

4 Conclusions

The use of a relational database to organize the disparate data into a more useable format has made the accessibility of the library use data at St. Cloud State University much easier to use. This has allowed for the creation of multi-dimensional data cubes with the Microsoft Visual Studio Business Intelligence application, which in turn has allowed for a much more in depth view of the raw data that was provided, and has already lead to some insights into the library's physical asset utilization. Hopefully, the authors will be given access to a greater amount of the library's usage records, as the physical usage has become a relatively insignificant amount of the total. This would allow for greater and more actionable BI moving forward. The next major hurdle would be to look at the utilization of the library's electronic resources, such as journals, books, and digital papers, and perform the same types of processes to organize the data into a useable structure and then perform various business intelligence analyses against it to hopefully make more informed decisions about how library resources are utilized going forward. Hopefully this paper has shown how organizing data into a useable form, such as a relational database, allows for much easier use and manipulation of the data and as an ancillary benefit greatly increases the ability to provide useable data analytics or business intelligence.

References

[1] Moss, L. and Atre, S., Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications, Addison-Wesley Longman, Boston, 2003.

[2] Guster, D. C., & Brown, C. (2012). The Application of Business Intelligence to Higher Education: Technical and Managerial Perspectives. Journal of Information Technology Management, 23(2), 42-62.

[3] Lupu,A., Bologa, R., Lungu, I. and Bara, A., “The Impact of Organizational Changes on Business Intelligence Pro-jects”, Proceedings of the 7th WSEAS International Conference on Simulation, Modeling and Optimization, pp.415-419, 2007.

[4] Apte, C., Liu, B., Pednault, E and Smyth, P., Business Applications for Data Mining, Communications of the ACM, 45(8), 2002, pp. 49-53.

[5] Schonberg, E., Cofino, T., Hoch, R., Podlaseck and Spraragen, S., Measuring Success, Communications of the ACM. 43(8), 2000, pp. 53-57.

[6] Goldstein, P., “Academic Analytics: the Uses of Management Information and Technology in Higher Education”, ECAR Key Findings, EDUCAUSE, December, 2005.

[7] Van Barneveld, A., Arnold, K. and Campbell, J. (2012). Analytics in Higher Education: Establishing a Common Language. EDUCAUSE Learning Initiative Paper, Number 1.

[8] Natsu, J., “Advanced Analytics: Helping Educators Approach the Ideal”, eSN Special Report, eSchool News, 2010, pp. 17-23.

[9] Balkan,S. and Goul, M. A Portfolio Theoretic Approach to Administering Advanced Analytics: The Case of Multi-Stage Campaign Management, Proceedings of the 44th Hawaii International Conference on System Sciences (HICSS), Jan, 2011.

[10] Bao-sheng, L. and Xin-quan, G. Study on Predictive Model of Customer Churn of a Mobile Telecommunications Company, Proceedings of the 4th International Conference on Business Intelligence and Financial Engineering, Oct, 17-18, 2011.

[11] Argotte, l., Mejia-Lavalle, M and Sosa, R., Business Inteligence and Energy Markets: A Survey, Proceedings of the 15th International Conference on Intelligent System Applications to Power Systems (ISAP), Nov., 8-12, 2009.

[12] Williams, S. and Williams, N., The Business Value of Business Intelligence, Business Intelligence Journal, 8(4), 2003.

[13] Chaudhuri, S., Dayal, U. and Narasayya, V., An Overview of Business Intelligence Technology, Communications of the ACM, 54(8), pp. 88-98, 2011.

[14] Zucca, J., Evidence Based Library & Information Practice; 2013, Vol. 8 Issue 2, p172-182, 2013.

[15] Ramakrishnan, T., Jones, M. and Sidorova, A. (2012). 'Factors influencing business intelligence (BI) data collection strategies: An empirical investigation'. Decision Support Systems, 52 (2), pp. 486-496.

[16] Ariyachandra, T. and Watson, H. 'Key organizational factors in data warehouse architecture solutions'. Decision Sup-port Systems, 49 (2), pp. 200-212.

[17] Cupoli, P., Devlin, B., Ng, R. and Petschulat, S. (2012). ACM Tech Pack on Business Intelligence/Data Management, ACM. p. 15.

[18] Execution MiH. Refresh Frequency of a Data Warehouse: What should be the refresh rate of a data warehouse? http://www.executionmih.com/q/refresh-frequency-of-a-data-warehouse.php, 2012.

[19] Florescuand, D. (2000) 'An extensible framework for data cleaning', Proceedings of the 16th International Conference on Data Engineering, p312.

[20] Goldstein, P., "Academic Analytics: the Uses of Management Information and Technology in Higher Education", ECAR Key Findings, EDUCAUSE, December, 2005.

[21] Mehta, A., Gupta, C. and Dayal, U. 'BI batch manager: a system for managing batch workloads on enterprise data-warehouses', Proceedings of the 11th International Conference on Extending Database Technology: Advances in Database Technology, pp. 640-651.

[22] Duggin, J., Cetintemel, U., Papaemmanouil, O. and Upfal, E. (2011). 'Performance prediction for concurrent database workloads', Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data, pp. 337-348.

[23] Scannapieco, M., Mirabella, V., Mecella, M. and Batini, C. 'Data quality in e-business applications', Lecture Notes in Computer Science, 2512, pp. 121-138.

[24] Chiang, R., Goes, P. and Stohr, E. (2012). 'Business intelligence and analytics education, and program development: a unique opportunity for the information systems discipline', ACM Transactions on Management Information Systems, 3(3) 12.

[25] Krupa, M. (2011). 'HR business intelligence Apps-don't do the kitchen sink method', InfoBox: a Tidy Package of HR and Technology Information, Insight , Wit and Wisdom. http://infoboxinc.com/hr-business-intelligence-apps-dont-do-the-kitchen-sink-method/.

[26] Segev, A., Shao, S. and Zhao, J. (2003). ‘A data analysis model for business intellignence’, International Journal of Internet and Enterprise Management, 1(1), pp. 7-30.

[27]. Ferranti, J., Lnagman, M., Tanaka, D., McCall, J. and Ahmad, A. ‘Bridging the gap: Leveraging business intelligence tools in support of patient safety and financial effectiveness’, Journal of the American Medical Informatics Association, 17(2), pp. 136-143.

[28] Ulanov, A., Simanovsky, A. and Krohn, M. (2012). ‘Personalized report creation for business intelligence’, http://persdb2012.cs.umn.edu/papers/8.Ulanov-PersDB12.pdf.

[29] Grolemund, Garrett. (2013), https://support.rstudio.com/hc/en-us/articles/201057987-Quick-list-of-useful-R-packages.