Journal of Information Technology Management

ISSN #1042-1319

A Publication of the Association of Management

THE APPLICATION OF BUSINESS INTELLIGENCE TO HIGHER EDUCATION: TECHNICAL AND MANAGERIAL PERSPECTIVES

DENNIS GUSTER
SAINT CLOUD STATE UNIVERSITY
dcguster@stcloudstate.edu

CHRISTOPHER G. BROWN
SAINT CLOUD STATE UNIVERSITY
chrisb@stcloudstate.edu

ABSTRACT

Business-intelligence has proven to be an effective means of improving the decision-making within an organization. While traditionally applied to profit making private companies BI has recently been applied to public institutions as well. The authors present a case study of the application of BI to a public university. The details of the technical considerations as well as the limitations that resulted from political and control issues are presented. In spite of the limitations the authors were able to devise a successful system structure. However, the limitations in regard to data control and data definition have prevented the BI system from reaching its full potential.

Keywords: BI, Business Intelligence, Higher Education, Data Architecture, Analytics, Metrics, Cube, Scorecard, Dashboard, Inman, Kimball

INTRODUCTION

Business Intelligence (BI) has been widely accepted as a means for effectively designing and managing the system life cycle to support intelligent decision making, Moss et al. [19]. Applying BI to any large organization in theory should be a relative straight forward procedure, however politics, differences in management styles and varying expectations sometimes result in unexpected consequences, Lupu et al. [18]. Many have turned to BI as a remedy for a poor organization within a business. However, BI is really a tool that can be used to make a business run more smoothly and profitably. In a way it's like you are building a house, you can build the house with hand tools, but you can build the same house much more efficiently with power tools and BI of course would be analogous to using power tools, Pucher [22]. Perhaps, one can relate the value of BI to an early analogy used in IT (back then it was called data processing): garbage in garbage out. BI provides a very efficient structure for designing and implementing a computer system, but it is dependent on the underlying business logic. A classic flaw in this logic is presented in the work of Schonberg et al. [25], in which they point out that for BI to be successful the organization needs to determine what behavior (which of course hopefully translates to metrics) indicates success. If that is known then BI can be effectively used to make intelligent decisions that will enhance the success rate.

While the planning process for small or medium sized businesses can at times seem overwhelming, Grabova et al. [11], the problem is often even more difficult when BI is applied to higher education, particularly if the institution is public in nature. On the surface one would think that higher education would be a prime target to employ BI. Most institutions have large amount of data and given the current economic environment there is a great need to run the institution more effectively, Engelbert [9]. However, often the data integration capability is non-existent , any new cost is prohibitive and many University executives simply do not understand the advantages that could be offered by BI. This situation often results in multiple data sources that contradict one another and data in raw form which is of limited value for decision making purposes. What BI does is integrate the data, provide built in analytics and well-designed user friendly interfaces for end users. Besides creating an easier to use integrated data system, BI becomes the sole source for that authoritative data. This concept is described by Kuster et al. [16], as "one source for the truth". They further state if properly deployed that BI can be very effective and even cite an example in which $1.1 million was saved by the State of Ohio.

However, the road for achieving success for many institutions, particularly small and medium sized institutions, may well be a bumpy one. It seems understanding the benefits and operational characteristics is not a problem limited to just University executives, but rather to IS professionals in training as well. Hart [13] reports holes in the training process and has proposed adding a meaningful hand-on project designed to help the students link the BI frame work to a an existing database project. The success stories within BI certainly indicate that BI can vastly improve efficiency. However, it appears that each application will in fact be a little different because no two schools operate in the same manner. There certainly are guidelines in the literature, for example: Cates et al. [6]. Further, operational plans for implementing BI also exist, for example: Rylee et al. [23]. However most would agree while BI can be quite valuable it is still an emerging process, Dell'Aquila et al. [7].

Therefore, the purpose of this paper is to delineate how the approach was applied to a medium sized Midwestern University. The discussion will center around how the common IT planning concepts such as: top down design, SLA (service level agreement), definition of data governance, hardware architecture design, data dictionary requirements, data warehousing and ODS (operational data store) were used to reach a BI solution. Throughout the discussion pitfalls such as data conversion problems, lack of understanding the needs of BI and political roadblocks will be delineated as well.

REVIEW OF LITERATURE

The Effectiveness of BI

The literature in general indicates that the implementation of BI can have dramatic effects on the improvement of the decision making process. Of course central to the BI strategy is the use of data warehousing. In designing a strategy, an agile rather than a waterfall approach is key and has the potential of providing improved quality, reduced risk, and responsive changes, Sandler [24]. Sandler, further states that improved quality is realized through close collaboration with business users and stake holders which reduces the risk of missed business expectations and increases overall quality.

Sandler also compares the costs/benefits associated with agile methodologies versus waterfall approaches and states that they provide a similar final delivery timeframe. However, the agile method would have more builds and appear faster, but in the end the final delivery time is usually the same. Agile methods often increase or at least retain the same cost as waterfall approaches.

Sandler cautions that agile approaches are not the silver bullet to fix failing BI initiatives. According to his data 77 percent of the organizations surveyed indicated they had been involved in a failed agile project. The usual causes of failure cited are lack of experience with the agile approach and conflict with the agile guidelines and the organizational culture. To be successful with the agile approach Sandler's keys to the agile approach of Data Warehousing are:

- Show customers early and continuous updates/deployments
- Encourage customers to offer recommended changes (even late in development)
- Show working models frequently with small time-frames between demonstrating build changes
- Business folks and developers must work together daily and preferably face-to-face
- Creating working code has more intrinsic value than verbose documentation
- Empower teams to self-organize, build requirements and improve the original design.

In summary, this approach could be viewed as increased collaborative testing while maximizing the interaction of all constituents involved.

Viewing BI as a service and structuring process around Service Oriented (SO) driven event triggers can increase BI driven analytical decision making, Gordon et al. [10]. An example of this SO perspective is the detection of a suspect credit card at the time of swipe based on historical mined data. The authors further state that SO also encourages the proponents of BI to secure data to improve decision-making capabilities across the organization. In other words, BI views SO as a collection of data sources and event sources. Therefore, SO views BI as a collection of services. The key for SO BI to add value is the identification of a middle grain service to consume appropriate data to fulfill a particular requirement. A summary of the ways SO BI might be implemented follows:

- Referential integrity events can be validated before releasing the data from desperate sources to the data warehouse and dependent reports.
- Single validation service - ETL validation process can be exposed to other parts of the business.
- Timely updates - A hook to allow real-time updates to the data warehouse and provision real time reporting (i.e. Log shipping, replication, etc.)
- Common business schema - common logical model and common definitions referenced heavily within a Master Data Management model.

Given that Integration of data is the largest cost in a BI project, Gordon et al. [10]. Successful Database environments require a balance between business metrics, understanding the DBA's role and challenges, and integrating the appropriate tools for DB administration. The three main business metrics for a DBA are: Cost, Service, & Risk, Pearson [20]. The major cost is usually the DBA's salary which to some extent can be reduced by automating tasks and providing the DBA a simple UI (user interface). This allows for an increase in the number of Databases that the DBA can manage while maintaining a satisfactory service level. Service level is usually measured in availability and response time metrics. Risk is associated with database changes and usually impacts service levels. Both Service Level and Risks can impact the DBA cost if mismanaged.

Pearson further states that DBAs face big challenges related to resource management and service level requirements. He suggests that it is important to focus on accuracy over speed when addressing problems, thereby creating a state of continuous improvement and enhanced service levels. Also according to Pearson, third party tools can increase productivity by twenty percent (assuming the DBA is not limited by a legacy system being not compatible with such tools). Finally, Pearson feels that a database health monitoring tool can be effective to improve productivity and reduce risks.

However, the literature cautions that greater investment in IT does not necessarily guarantee better information. The key emphasis appears to be related to the manner in which data is gathered, processed and presented. The old adage "garbage in garbage out" still holds true in that inaccurate BI reports can be a major problem in the decision making process, maybe even more risky then not having access to the data at all. Effective BI is attained when organizations view information as the most valuable asset and remove barriers to its use, University of Cambridge [26]. Further, most failed BI initiatives are linked to poor integration of the organization's processes. The Cambridge study also cites a lack of well-defined KPI's and recommends a top down approach where by reports are driven by decision-making strategic directives.

Data governance is also a critical component requiring: clear ownership of the data/processes that generate the data and a clear understanding of who uses the information and purpose of its use. These governance principles require the use of a quality prioritized funding model not a technology driven one. The Cambridge study states that 95% of employees are unaware or do not understand the organizations strategy. The main focus of successful BI is to provide less information and increase the strategic alignment of the information provided in attempt to keep its constituents better informed. In other words BI should serve as means of streamlining and increasing accuracy within the data so that planning, forecasting, and reporting mechanisms all link to the same information.

Unfortunately, fewer than 10% of organizations have simultaneously enhanced organizational and technical structure use, University of Cambridge [26]. Therefore, a data governance layering strategy is recommended featuring three layers: who owns and controls the information; who uses the information; and who produces it. It is also very important to have a manageable number of performance measures to make effective data driven decisions. In summary, the Business Intelligence platform should provide a unified and trusted view of the business, empowering all employees with insight and align with the organization's operational strategy."

However, there are limits to the efficiency gains an organization can achieve by bringing information together. Data gathering processes can have a diminishing returns effect when optimized and streamlined. Therefore it is logical to investigate business process enhancement (product development and service delivery innovations) which appear to have more growth opportunities. Basically BI is not an open ended prescription it is what you do with BI to shape the organization that enhances the bottom line. BI processes tend to be in a state of flux as they are continually providing feedback to organizational leaders who in turn change the business process requiring the BI process and tools to change accordingly.

Most effective BI initiatives leverage prototyping and an AGILE method of software development. It is recommended that six key indicators of BI competency and pervasiveness be used to measure BI effectiveness in an organization:

- Degree of use of BI solutions within the organization across all departments and work areas.
- Degree of external use of BI solution by stakeholders
- Percent of BI power users who use BI systems regularly to provide decision support.
- Number of domains or subject areas in the primary data warehouse
- Data update frequency (refresh rate)
- Analytical orientation - how does BI influence employees actions

BI in Higher Education

Specifically within higher education BI is viewed as a solution with much promise in regard to adding much needed efficiency on an operational level. However, there appears to be some confusion as to what actually constitutes a BI system. Angelo [1] reports that apparently everyone is labeling any kind of Higher Ed reporting systems a business intelligence solution. She further states that one large impetus for BI in Higher Ed is the amount of disparate data sources and the time required to process integrated reports. To illustrate the point Angelo [1], provides the following example. The National University of Health Sciences (NUHS) needed to improve their enrollment reporting process and opted for a BI solution. Of course implementing such a solution is not easy. The business objects solution adopted costly $200,000 and took two years of planning before implementation was possible. This solution is not aligned with BI best practices which normally would involve a custom solution and use a top down holistic approach, rather than focusing on one particular facet of the business. However, focusing on one facet at a time might provide a means of breaking the problem into modules. A risk in this approach is that the interaction among those modules might not be properly assessed. In contrast Florida State University (FSU) is integrating BI metrics and thresholds in conjunction with workflow processes. Further, (FSU) has been able to adopt BI to sift and filter out at risk students that end up being "dead ends" and then focus attention on students with adequate academic skills, Durso [8].

One could sum the current challenges of implementing BI in higher education as: multiple versions of the truth; significant time spent on data gathering then analyzing it and the high IT development cost associated with BI.

The result is a situation in which many campuses are in a "Data fog" where they don't know where to get data or if it is even accurate. Accordingly, Durso offers the following recommendations:

- Ensure BI has a vocal advocate in the administration
- Use seasoned BI Vendors
- Tools should integrate with other strategic initiatives
- Perform Data cleansing early on
- Research the right tool for the institutions requirements. (often home built or custom legacy systems restrict the use of most out-of-box BI solutions)
- Establish an effective data governance structure involving administrative officers
- Identify specific goals for the BI initiative
- Do not focus only on the technical aspects of the BI solution, recognize the business value and ensure it is adequately provisioned.

There have also been successes related to the use of BI technology within higher education. Specifically, data mining and analytics have been used to analyze student data, guide course redesign and for retooling assessment as well as to encourage new communication models between instructors and student learning, Baepler et al. [2]. In other words, the authors investigated how analytics and data mining can shape the effectiveness of teaching and learning.

IMPLEMENTATION METHODOLOGY

Designing the Service Level Agreement (SLA)

Characteristics Required for the Warehouse and BI Customers: One of the first parameters that needed to be addressed was uptime requirements. Much discussion occurred regarding the granularity of the data. For ease of design and management coupled with the fact that all data driven decisions were made on a “daily” basis resulted in a proposed granularity of one day. From an update perspective this was easy to support because an update could be done in the middle of the night when usage was almost non-existent. In the event users were to attempt to access a MOLAP cube driven report during the DataMart ETL process, they would be unaffected. However, if the cube is being updated the user would then have to wait until processing is complete to access the cube. Future options to improve accessibility are being examined including clustering the cubes and providing cube shadow copies. Further, disruption of service in most instances could be easily corrected within the one day time frame.

This one day granularity also made it easy to support changes in reporting requirements. Usually a requested report modification simply required tweaking a few fields in an existing report that could often be accomplished in an hour or less. Further, if the report request was largely new and there was no existing structure to build from then it was still possible using existing tools such as Microsoft® Report Builder to easily accomplish the task in three to four hours. In both cases the goal of daily granularity was easily supported. Last, errors in the code or data once detected could also be resolved for the most part in the one day time frame. The one day granularity also impacted historical data. The current ODS source allows the changing of historical data. Due to the importance of being able to reproduce reports from a given point in time, specifically a day in time, it was necessary to provide historical data at a day grain. The historical algorithm can reproduce reports based on the latest data updates for that day, Brown et al. [3]. Current discussions around data governance and leveraging the warehouse to service dynamic applications have forced a reexamination of using a set refresh interval for the entire data warehouse. The goal is to identify specific tables to have higher refresh intervals of every fifteen minutes. The design supporting history tracking would allow for more frequent updates; however only the latest update would be retained for any given day.

Characteristics Required for the Service Level Agreement on the Warehouse and the Operational Data Store Level: All data stored that can be useful for decision making is seldom stored in the same location. This can typically be traced to the fact that historically individual departments tend to devise their own data collection strategy and later learn that the data is useful for decision making outside of that department. For BI to function effectively it is critical to consolidate the desperate data into a single structure, hence the need for the warehouse logic. In the case of the author's application the data did need to be collected from several diverse sources and organized into a data warehouse structure designed and indexed in such a way to ensure adequate retrieval performance. Of course when linking together various data sources it is important to address differences in granularity. In the author's case the primary data set was maintained and downloaded from their parent organization because their BI system was being devised for one campus (the largest) of an organization that contained numerous campuses. This data set was considered up to date to the last day which met SLA goals and hence was a good fit in terms of desired granularity. However, if granularity of less than a day was desired that primary data source would present a huge obstacle in meeting that goal particularly because control of this data source, regarding many of the revealed tables and views, was outside the purview of the authors. As a means of keeping the end-user informed in regard to the concurrency of data displayed in the various reports the authors found it useful to place a "data freshness stamp" at the beginning of each report so that if the report was run at 4pm they would understand that it did not contain any change that took place since the night before when the last update was run.

To ensure concurrency of data even at the daily level requires a well thought out download procedure. The original method called for a complete copy of the data to be downloaded due to the unavailability of deltas or incremental updates from the parent ODS. However, over time it has been recognized as a less than effective method. The authors are in the process of converting to an incremental approach in which only the changes from the previous day are downloaded and then reconciled with the previous day's copy. This methodology significantly reduces the data transition overhead which can greatly enhance efficiency. While the authors recognized this as the preferred method it took much discussion and political pressure to convince all the people responsible for managing the diverse data sources to support this superior methodology. It was especially difficult to achieve this goal when working with data managers that were not located on the author's home campus. Perhaps, the short term object was that it would require extra work to devise the procedure. However, if one looks at the long term savings in resources (reduced data transfer) coupled with the perceived ability of the new BI to ensure better (and more cost effective) decision making certainly from a long term return on investment the request was easily justified.

From a security perspective overcoming the clumsiness of having to deal with several diversely managed data source was challenging. For example, forced password changes to the account that supports the downloading process take place every 90 days. The password policy also enforces a three strike rule, whereby if three consecutive incorrect password attempts are made on an account, the account is locked until an administrator enables it. Currently the hours of operation for the Administrators follows normal business hours, excluding weekends and holidays. If the account is locked on a Friday evening, the account will remain disabled until a contact can be made the following Monday, resulting in a substantial loss of ETL updates from the ODS. The security policy in place did not allow BI personnel to devise that new password and enter it, so the change needed to be made while a system administrator for the data store in question was available. An automated process is currently being designed to facilitate service account password changes. There are still risks with account lockout and satisfying the 90 day change policy; however, the automation reduces the risk of manual errors.

Situations such as this illustrate the problem of not giving BI personnel (particularly during the development phase) enough control of the supporting computer system infrastructure. This is an important point that should be weight carefully when planning a BI implementation. If the BI teams does not have sufficient control or at least a very high level of system support the project is likely doomed to failure. Last, the authors felt that there needed to be some type of auditing mechanism that would check threshold variance during the download procedure to ensure adequate reliability of the data. One simple example would involve a comparison of the record count of the previous day's data to the current day's data. It was expected that a variance of not greater than 5% would occur. Therefore, in cases where the 5% threshold was exceeded additional manual validation of the data would be applied in an attempt to explain the source of the variation. In the event an out of variance ETL error occurs the previous extracted table would remain intact and the new data would be stored in a separate table with the appropriate date stamp appended to the name. The DBA would then replace the old data with the new data upon validating the source of the variance. The authors have plans to improve the variance checking process by adding checks at the field level.

Data Governance Policy Building

Define and Align Data Rules: A major question that arises when building a data warehouse is what data can any given user gain access to? The resulting access control list can be logically fragmented and result in security holes within procedures for granting those rights is not well thought out and implemented correctly. Therefore, the development of policy and rules is critical to this process. Certainly, it is important to use a proven structure in developing these rules and of course the rules are dependent on meeting the needs of a potential diverse group of users. Therefore, tools such as the role descriptor templates as proposed by Lemme [17] can be invaluable in defining end-user's roles and the rights sets they will require. Specifically in the author's BI structure it was determined that a four layer security structure in which each level can grant access to direct subordinates (or below). While this plan may not have enough sophistication once the BI process matures the method has proven effective and elegant while the development and acceptance process is taking place. Table 1 below describes the characteristics of each layer.

Table 1: Access Control Layers

Layer	Description	Job Titles
1	Top tier/strategic root	Vice President Office of Strategy Planning and Effectiveness (VP OSPE)
2	Executive Group	Deans, Vice Presidents' and President
3	Middle Management	Department Chairs and Department Directors
4	End users	Faculty and Staff

Within this model access is typically granted by using the following logic set. First, an access request is submitted to requestor's supervisor. The supervisor in turn submits the request to the Office of Strategy Planning and Effectiveness (OSPE). If granted the request is forwarded to the DB architects who in turn add the account to the Strategic Analytics Management System (SAMS) access group. More specifically the security model in terms of AD groups is delineated in more detail in Figure 1. Security model (AD Groups). In this model a user has to be a member of all vertically aligned groups to have access to a SAMS report whether it exists on either the development or production side.

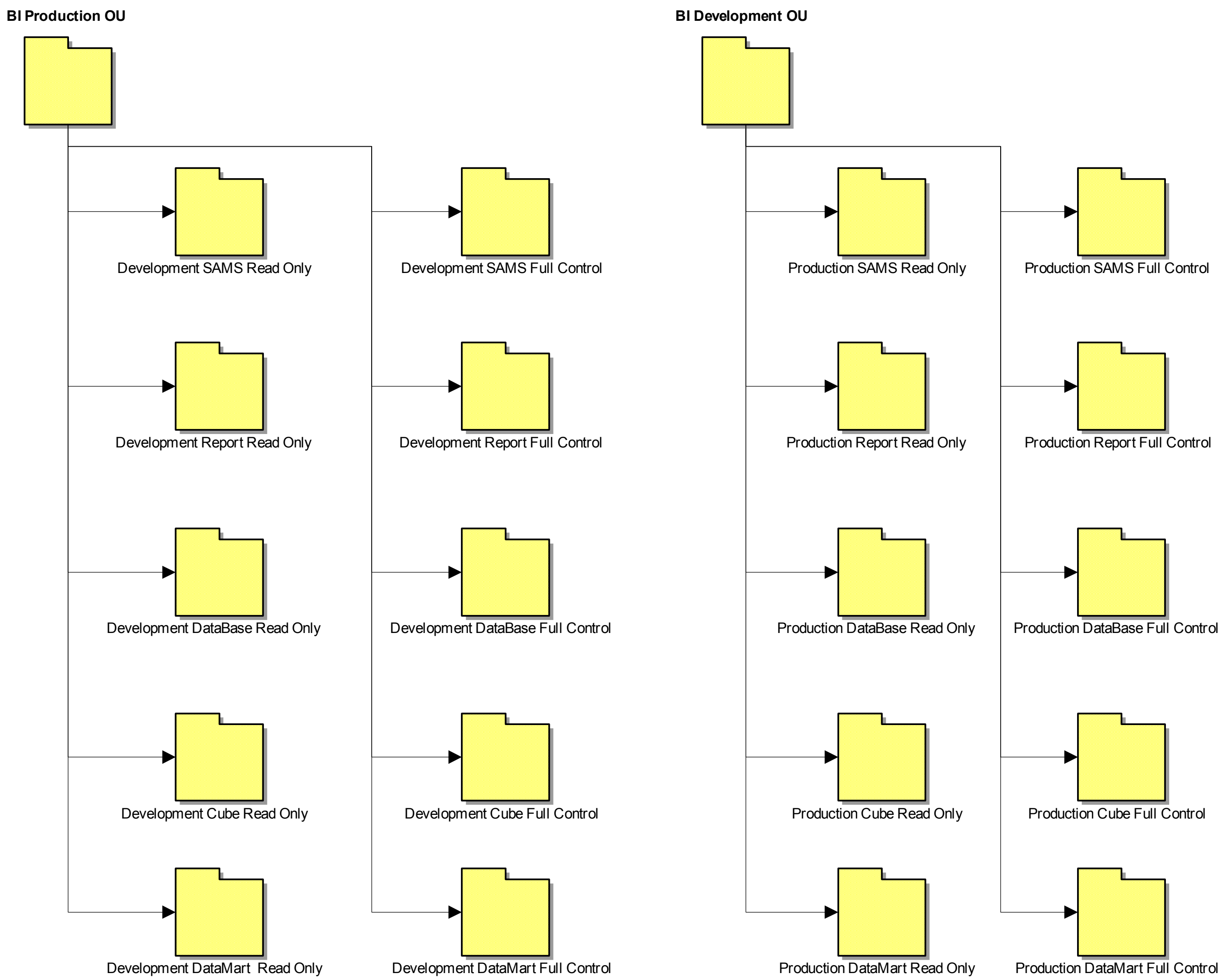


Figure 1: Security Model for AD Groups

Protect Data Stake Holders: It is clear that within higher education that some data can be very sensitive, such as grade and financial related data. Therefore, the prime directive within this concept is that the existing data governance rules do not get violated and sensitive data is released to the wrong individual. A good rule of thumb is to evaluate and ascertain if the data to be released is really pertinent to the decision making process of the individual in question. Currently the VP of OSPE determines if the accessing the BI system is pertinent to the job duties of the requestor. The system is currently exclusive with the users given access due to the lack of restriction once they are in the system. All users are allowed to see all reports, unless a specific report is provisioned with exclusive access, which is rare.

Resolve Non-compliance: Short term a risk assessment approach is being used to provide some idea of the likelihood of noncompliance occurring within each access control group. The BI System Engineer tracks usage patterns of users and reports that are being accessed. A user response or notification is the current communication method used to determine and remedy users unable to access the system or any part of the system. Monitoring the data grain security is even more challenging as it is difficult to ascertain all of the combinations of ways a report could be filtered to prevent users from filtering down to an identifiable subset of otherwise anonymous data. Certainly, a much more efficient methodology would be to rely on log file analysis. However, the system administration group has retained ownership of the log files and the BI team is unable to use the logs at this time. This is a critical component in devising an ethical BI system and a point that needs to be dealt with as the author's BI system matures.

Security Strategy

Hardware Layout concerns: Once the BI system goes into full-scale production adequate performance will be critical. To assure such performance a distributed server layout will be selected. This design provides adequate and scaling CPU resources as well as fault tolerance through host replication while still permitting isolation of given applications, Guster [12]. Although the aforementioned design is critical, it is still a political issue in that the BI staff does not have control of its hardware and can only recommend the structure. Currently a hardware refresh has offered opportunities for a server topology design change. The new design would move the stand alone SQL server layout to a SQL cluster for the Data Warehouse and analytics. The pros of this decision include: Increased availability and fault tolerance and ability to add log shipping to a disaster recovery site or second datacenter.

Security Model: The first and fundamental question that needs to be answered is will users sign into to the BI system or use some kind of pass through authentication such as Kerberos. It is important to provide single sign-on as a service to the users so perhaps the BI system could be attached to the existing global authentication system. Because the campus is currently committed to the Microsoft platform so a form of LDAP (light-weight directory access protocol) hardened with Kerberos is probably the most robust option available. Some challenges exist with using Kerberos such as access proximity and platform dependency. Leveraging Kerberos precludes off campus pass through access without tunneling or using a VPN connection. Non windows users are challenged with access and are currently required to utilize a Citrix client through a virtual lab environment to access the SAMS reports.

Once user authentication is established there is need to provide access control to the data. Because there will be large numbers of users with varying data needs a centralized model would be difficult to maintain. A more logical solution would be a tiered approach featuring the creation of active directory groups and access to the various reports could be provided by a group level manager which would mean that a supervisor would manage the access of his/her subordinates. However, this decentralized model still needs to be compliant with the data governance rules. So in effect the owner of the data has to be aware and approve all user access to the reports using their data. This has proved to be a cumbersome process. Getting everyone to agree and automating the process so it is practical has been a major political problem. So in effect by not effectively dealing with this issue, the political non-decision forces non-compliance.

In terms of addressing the reliability of the data itself the current logic in place is an automated script that compares each day's data with the previous day. If there is a change of greater than 5% in the number of records then the validity of the data is considered questionable. This situation requires the data warehouse staff to perform a mini-audit to ascertain if the data has really been compromised. Because much of the original data come from the parent organization and is downloaded into the warehouse this can be a cumbersome and potentially unsecure process because the data non-compliance information is received via email including the table name and date of occurrence. Further, the whole process is administered by the DBA and there is no security officer involved which in effect is a conflict of interest.

Evaluation of the Data Dictionary Requirements

Current Institutional Data Dictionary: While a data dictionary can be a powerful tool it needs to be well designed and accurate to be effective. The first problem encountered was that there was no recognized campus data dictionary. To deal with this problem the BI staff in effect had to create their own mini data dictionary which in effect was equivalent to putting a band-aid on a gaping wound. The solution was to simply define the terms used in the report and hence BI created their own dictionary which was not officially recognized as authoritative beyond the BI systems. However in an attempt to ensure consistency across non-reporting based automated applications, such as web services and other IT processes, comments embedded in the SQL scripts have been developed and retained to provide operational documentation. Future plans include migrating operational documentation into a central data dictionary.

In regard to future operations how will the BI data dictionary be made available to clients? It is critical that clients gain access because the data dictionary is the central depository of data definitions. This is needed not only to provide clients access to accurate definitions and metadata, but to ensure the endorsement of the data dictionary as the definitive defining standard for data within the institution. That being said it is not a matter of deciding which reports should contain the definitions, but rather the data definitions need to be pervasive throughout all reports. In other words, the definitions need to remain constant across all reports. Gaining the authority to make data definitions available campus wide has been a difficult political issue and an ongoing battle. Because of this lack of authority there is a need to deal with situations in which there are multiple definitions for the same term. The strategy is to retain multiple definitions for each term and index them with the currently accepted definition within the BI data dictionary. This is critical in that there are often multiple definitions for the same term. For example, the term student is viewed by some as anyone enrolled in at least one credit. However, from a financial aid perspective a "student" must be enrolled in at least 12 credits. Further, the email system views a graduate as a student for 6 months after graduation because their email account is kept active for that duration. All of this confusion could be eliminated by a well thought out data dictionary strategy.

Building the Data Warehouse

Inman versus Kimball: There are two prevalent schools of thought in regard to how a data warehouse should be constructed. First, there is the approach devised by Inmon [14], in which a comprehensive plan is developed. This plan features allocating time and money to plan, requires a data architect, hopes to create a holistic understanding of a vision/scope for BI, generates an understanding of how BI will be used in decision making and provides the ability to make quick data definition decisions. Whereas the alternative approach offered by Kimball [15], is more of a mini-mart approach. This approach require less vision and planning, features a low budget with fast results requirements, the work can easily be distributed across the organizations and allows the postponement of attaining a single version of the truth/definition requirements.

Although, the advantages of the Inmon model were apparent the need to show results in a timely manner necessitated a hybrid approach. Interestingly, the first phase of Inmon detailed planning was adopted. The purpose of the first phase is to attain a 5 year vision of how BI will be used to facilitate the decision making process. Funding was provided and a partnership established with another University in the same system to leverage the planning process. This process was further supported by Microsoft Corporation as well. However, while the planning process continues the need to prove the effectiveness of the data warehouse strategy led to the development of some mini-mart logic so as to provide tangible evidence that the project was in fact working. Once again political hurdles slowed down the planning process as a result of issues such as: who owns the data, what is true meaning of a term that should be placed in a data dictionary, what data granularity is needed, how to resolve data entry errors previously undetected, and how should the data be secured. Not only were the road blocks themselves a problem, but often in seeking a resolution to the problem additional problems arose. In other words, the multiple layers of bureaucracy clouded who was responsible to make the important decisions related to supporting a BI system and created a risk adverse culture whereby no one was willing to make such decisions. In summary, while the Inmon approach was probably the better solution considering the starting point, the political structure made the approach almost untenable at least over the 5 year visionary time period.

Connecting to an ODS (operational data store): This foundation procedure turned out to be problematic for several reasons. First, the data itself is entered on campus but stored in a system wide data base. This requires an extraction process so that the necessary data gets downloaded and stored in the campus data warehouse needed to support BI activities.

Granularity and Historical Data: Because customized scripts needed to be developed to facilitate the extraction of data from the centralized data base granularity was a real concern. While there are some decisions that need an hourly granularity the majority of the decision making could be supported with daily granularity. This fit well with the majority of what the centralized database was capable of providing. The pervasive logic is that the extract process should take place at night during a slack usage time. So therefore the previous day is extracted during the middle of the night and hence available at 7am the next morning to support BI related decision making. If hourly or even more frequent updates are to be supported then the data will need to be updated on the campus data warehouse and near real time access provided. This could be problematic in the sense that the campus and centralized records could in fact get out of syn. The other alternatives would be to run the BI directly from the centralized database or to extract from that centralized database on an hourly level. Central IT is hesitant to allow such activity because the system is already hard pressed to meet it reliability and performance goals. While the current extraction process is cumbersome the data retention policy is even less clear.

There appears to be no agreement in regard to how long data should be retained. Because the system is in its infancy the current methodology is to retain all data indefinitely. Of course as time progresses it will not be cost effective much less possible to retain all data (at least within the data warehouse). How archived data will be stored will be a function of how often it needs to be accessed. Due to the inability to follow the Inmon model coupled with the inability to get timely decisions made in regard to custody of the data at this point it is not clear how often archival data may need to be accessed. Perhaps, the next step would be to at least get it off the high performance SAN (storage attached network) to a lower performance and less costly NAS (network attached storage). Certainly as time progresses some form of off-line storage such as tape will need to be implemented. Several factors that buy more time in regard to migrating archived data include: The inherent compression in the model used for storing history and the current small footprint required for the current data warehouse.

Assessing Data Quality: The minimum amount of data required to support the BI process is downloaded through an ETL process in which the agreed upon granularity is a day. Currently that data is spread across more than 100 tables and requires about 30 minutes to download which takes place during a slack period at night. Unfortunately there are integrity issues the first is that the centralized database is not normalized, Codd [5]. So as one might expect there is redundancy in the data, the table structure is not optimized and updates can be cumbersome from a performance/efficiency perspective. In designing the campus data warehouse care has been taken to minimize the effects of downloading from a non-normalized database. Therefore, several process steps have been developed that attempt to solidify the validity of the campus warehouse. The first method is in place to deal with multiple versions of the truth because no data dictionary exists on campus much less across the entire university system. A second process identifies and isolates duplication, misaligned, or data entry errant data and in effect parses the suspect data into a separate category to be evaluated and corrected by the data owners. This method of quarantining or isolating suspect data is what authors refer to as the "bit bucket" as it is analogous to a rework bin used in manufacturing environments. The bit bucket allows data of questionable validity to still be considered. For example, suppose the central database uses a different definition for student than the BI system the bit bucket method allows the data to be retained as an exception and a report can still be generated. The primary concern is that the decision makers realize that the suspect data is in a different category and may want to consider it in the decision making process or wait until the data owners resolve the discrepancies. Last, processes have also been devised to verify (on a day to day basis) whether the variance in the number of table records between updates are within the expected range. The details of how the cube interacts with the database structure are depicted below in Figure 2.

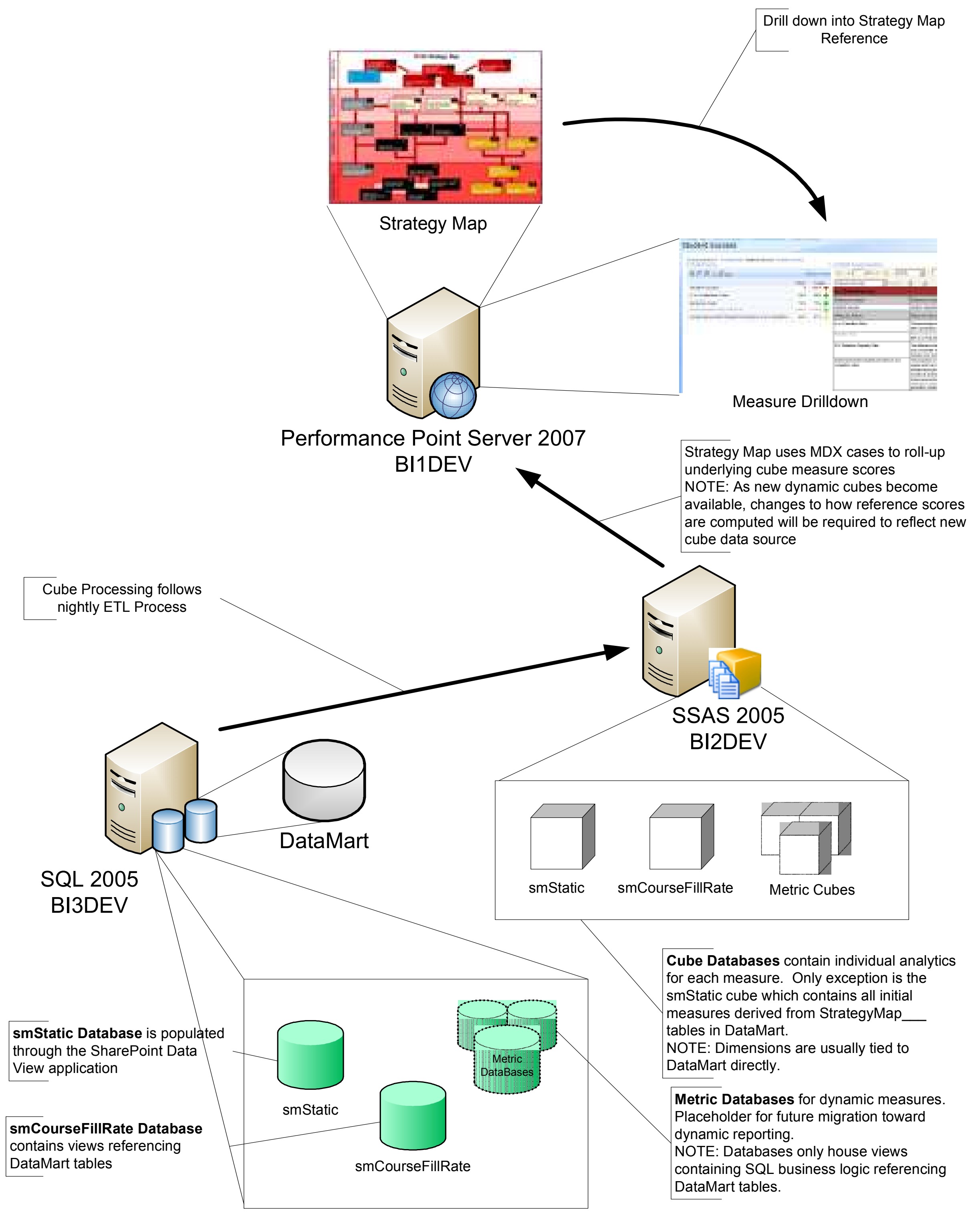


Figure 2: N-Tier Design Model

The ETL Process

Identify Desperate Sources: the extract transform load (ETL) process is critical to the success of any BI project. For the project herein this process is crucial in securing timely, accurate and complete data. As stated earlier the majority of the data is extracted from a system-wide centralized database. In fact, currently about 95% of the campus warehouse data is obtained in this way. This leaves about 5% which come from local sources in many forms, from spreadsheets to online data entry forms. Further, the information is entered with a variety of frequency and in the case of some departments using the online data forms the granularity (or refresh rate) maybe less than hour. However in some departments data is entered in batch based on when the office manager finds it convenient to enter which may result in a granularity of a week or more. It would be preferable to match the granularity of the centralized data which is up to date to the last day. However, for some applications it would be preferable to have an hourly granularity and to do so will require more local data and devising a methodology to ensure the target granularity meets the one hour target. It is expected that as the BI system matures that more and more data will come from local sources and reach a point in which 25-30% of data is extracted locally.

Build Staging Database: Because data is being extracted from desperate sources it must be ultimately integrated into the data warehouse through the database staging process. As stated in the previous section the extract/load process is generally done on a daily interval, but by using more local data sources it is hoped that some date and hourly granularity can be achieved. The goal would be to use the staging database to pull data into the data warehouse. Once the data is loaded into the data warehouse, it is hoped that for the majority of applications direct service can be provided to the data warehouse or a separate synchronized database linked to the data warehouse. There are however, particularly related to archived data when direct service will not be possible. Therefore, the trick will be ascertain if this is an occasional request or one that will be on-going and the data needs to be integrated into the staging process or separate application specific database. Last for security and performance reason it was important to devise a proper logging strategy for the ETL process.

The ETL logging process consists of two tables, an ETL Process table and an ETL Process Message table. The ETL Process table indicates the process that added, removed or changed data in a table and the date time stamp of when the event occurred. Each table has a foreign key constrain tied back to the ETL Process table. The ETL Message table has a foreign key constraint tied to the ETL Process table whereby each ETL Process record can have many ETL Message records further detailing activities performed on the data warehouse tables.

Data Modeling: It is important to realize that not every application is going to use the same data or search for what data it needs in the same way. Of course data within a database is stored within a physical architecture. However, the data can be logically viewed in a number of different ways. To meet this goal a Fact Qualifier Matrix (Table 2) was used for each pertinent measure. A sample is provided in Appendix A, which describes the instructional workload/department productivity cube. This tool allows the developers to dimension and organize the cube for the topic above in a structured manner. Once again the process of building cubes allows one to look at the data from multiple arrangements.

Transform Load: Once the disparate sources are extracted into the staging database the transform load process can be used to pull the data into the data warehouse. During the transform load process the dimensions are imported into the DataMart first, then the fact tables. This is due to the relational constraints inherent in the normalized relational design of the DataMart. Importing a fact table without its foreign key constraints being met would produce an error and possibly halt the entire transform load process. Another side note is the cube processing. If the relational constraints were relaxed (meaning the foreign key constraints were eliminated) and data from a fact table were loaded without matching lookup table data the cube processing would fail as cubes require a dimension record for each fact table key value.

Tuning the data warehouse and the data transformation can result in much better organized warehouse with vastly improved performance. Of course the final step is to place the transformed data into the data warehouse and devise an appropriate schedule for updating that warehouse. To illustrate the detail of the ETL process a detail chart is included in Appendix B at the end of the paper.

Building the MOLAP Databases

Implementing MOLAP: MOLAP (multidimensional online analytical processing) and can optimize storage through multidimensional indexing and caching Peterson [21]. Based on the size and scalability requirements herein MOLAP appeared to be an excellent solution in meeting the required performance goals. Of course to guide the logic of how MOLAP will be used a strategy needs to be devised first so that the sublevel metrics can be properly addressed.

Planning the Strategy Map: The strategy map besides adding structure to the query process was designed to provide immediate feedback using a granularity of one day. To automate this process the refresh was scheduled within the MS service broker, which is basically a thread management tool. For a representative example of how the scheduler works see Callahan [4]. The strategy map itself is included below as Figure 3. The strategy map is organized into four perspectives: student/stakeholders, resource stewardship, organizational processes and organizational readiness. The sublevels within each perspective have been devised as well as their relationship with other sublevel objectives within other perspectives as indicated by the arrows.

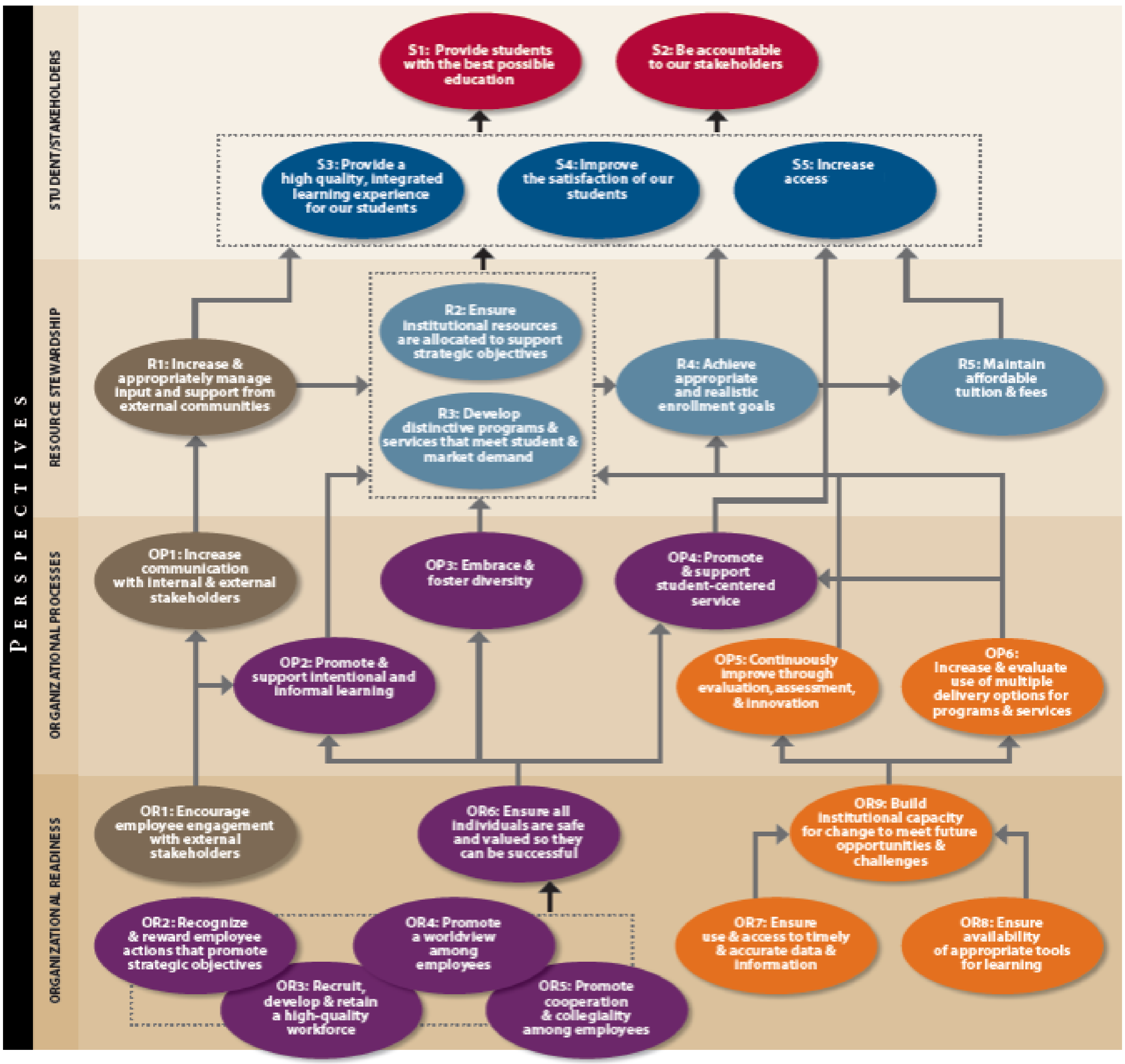


Figure 3: Strategy Map

Interfacing the MOLAP Databases to the Strategy Map

To provide effective access to the users of the BI system the strategy map has been implemented as a "dashboard" user front end and is depicted below as Figure 4. In addition to putting the strategy map into usable form the "dashboard" also provides immediate feedback in regard to how well each objective is being meet by using a "stop light logic". In other words, the small box in the upper right hand corner of each objective is color coded as follows:

- Green = Goal Achieved
- Yellow = Goal Not Yet Achieved
- Red = Not Achieved
- Black = No Data (to be implemented at a later date).

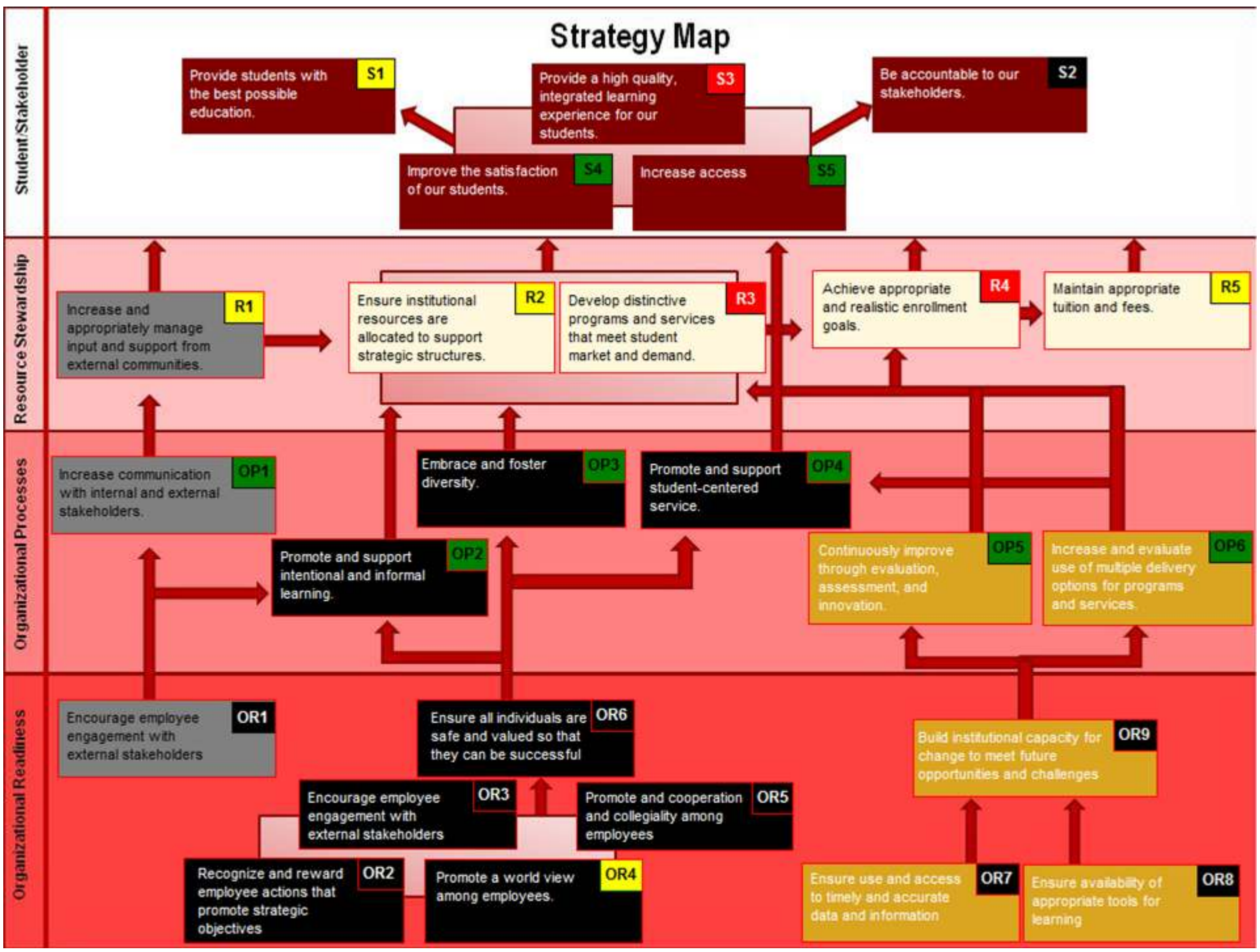
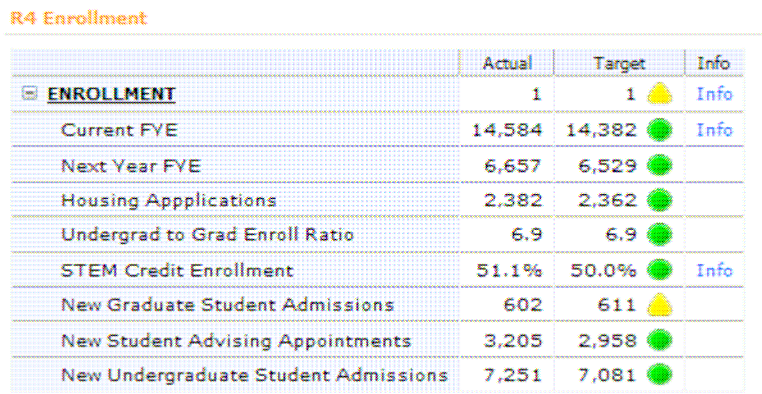


Figure 4: Strategy Map Implemented as a GUI

Further, detailed reports can be obtained by clicking on the appropriate objective box. For example, by clicking on the R4 Enrollment objective box a report depicted as Figure 5 below appears.

In addition, the BI system provides more detailed reports and graphics on topics within any objective. In Figure 6 below enrollment data is broken down by ethnicity.


R4 Enrollment

	Actual	Target	Info
ENROLLMENT	1	1	Info
Current FYE	14,584	14,382	Info
Next Year FYE	6,657	6,529	
Housing Appplications	2,382	2,362	
Undergrad to Grad Enroll Ratio	6.9	6.9	
STEM Credit Enrollment	51.1%	50.0%	Info
New Graduate Student Admissions	602	611	
New Student Advising Appointments	3,205	2,958	
New Undergraduate Student Admissions	7,251	7,081	

Figure 5: Report on Objective R4

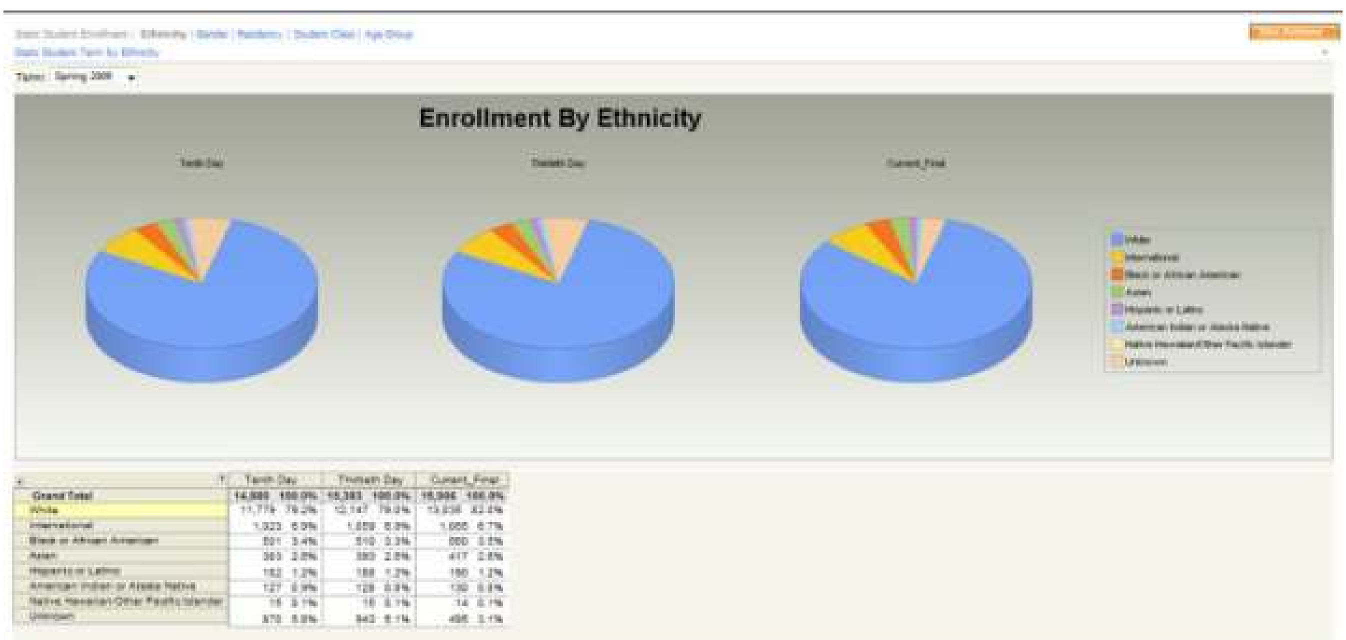


Figure 6: A Detailed Report on Enrollment by Ethnicity

So that the campus decision makers can obtain the most value from the data warehouse the "dashboard" also allows them (assuming they are authorized and have been given permission) to export data for their own use which they can use for customized decision making. Figure 7 below depicts the extraction screen for course enrollment. Note the comma delimited option is selected which is very popular when excel and SAS are the target software.

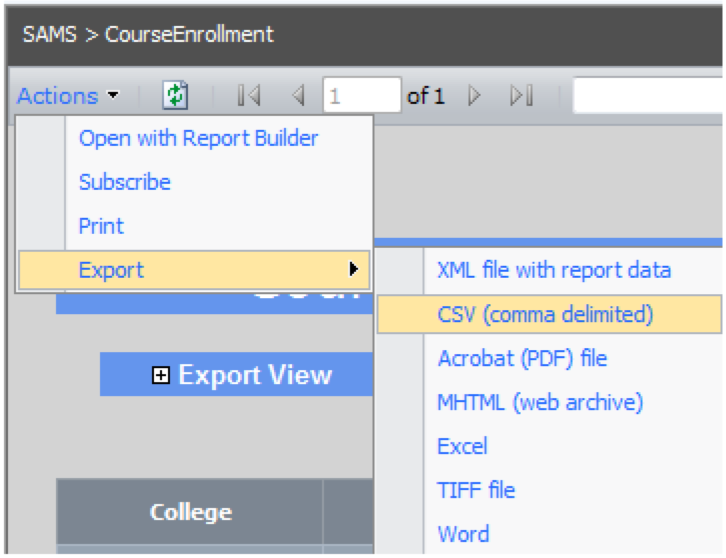


Figure 7: Data Export Window

DISCUSSION AND CONCLUSIONS

The literature and the author's experiences to date do in fact indicate that BI can be a powerful decisions making tool. However, implementing BI successfully can be problematic especially in regard to overcoming political roadblocks. While the author's experience began in a promising manner with initial commitment to the Inmon approach political consideration forced a redirect to a mini-mart approach. The complexities of BI particularly the many interrelationships within the data make the Inmon approach the best proactive solution. Having to deviate from this approach actually caused delays and limited validity in later parts of the project. Imagine how much simpler the project herein might have been if a data dictionary could have been devised and implemented at the beginning of the project.

Another, major consideration was the degree of control afforded to the BI team. In the example herein they were dependent on a centralized system-wide database to create and maintain the data warehouse and were not initially permitted to create and maintain their own service account password needed to support the extraction process. The BI team is crucial to the success of the project and especially during development needs to be given much latitude. However, as the system matures policy must be devised that limits their control and provides security oversight.

In spite of the political limitations the team was able to build a credible BI system structure. The "dashboard" developed has proven to be a simple, well received interface and the "stoplight" logic makes it easy for end users to monitor the health of the target subject areas. Although not entirely the original intent of the Inmon approach the development of the structure has put a face on the project and allowed some end users to employ a what if logic. That feedback to the BI team has resulted in additions/changes to the existing structure. As time progresses the sophistication of the system is progressing. However, that sophistication is hampered by lack of control in data extraction, business definitions, operational source data errors, disparate data source integration and politics that prevent data related decisions from being made. For the BI system described herein to truly reach its full potential these obstacles will need to be overcome.

REFERENCES

[1] Angelo, Jean M., "Business Intelligence: A new technology can analyze data at amazing speeds. So why is higher ed slow to adopt?", UniveristyBusiness.com, January 2007, 63-64.

[2] Baepler, Paul & Murdoch, Cynthia J., "Academic Analytics and Data Mining in Higher Education", *International Journal for the Scholarship of Teaching and Learning*, July 2010, Vol. 4, No. 2.

[3] Brown, C., Guster, D. C., and Jansen, B. "An algorithm to restore database content to past dates in

real time", *Presentation at the Midwest Instruction and Computing Symposium*, La Crosse, WI. April 11, 2008.

[4] Callahan, E., "Use of Service Broker Architecture for Downloading ISRS Data from 10g", *Winona State University*, 2007.

[5] Codd, E.F., "A Relational Model of Data for Large Shared Data Banks", *Communications of the ACM*, 1970, 13 (6): pp. 377–387. doi:10.1145/362384.362685.

[6] Cates, J., Gill, S., and Zeituny, N., "The Ladder of Business Intelligence (LOBI): a Framework for Enterprise IT Planning and Architecture", *International Journal of Business Information Systems*. 1(1/2).

[7] Dell'Aquila, C., Di Tria, F., Lefons, E. and Tangorra, F., "Business Intelligence Applications for University Decision Makers", *WSEAS Transactions on Computers*, 2008, 7(7).

[8] Durso, Thomas W., "From Data to Information: Business intelligence and its role in higher education today", *UniversityBusiness.com*, January 2009, pp.24-27

[9] Engelbert, N., "From Transactions to Transformation", http://www.edtechmag.com/higher/may-june-2007/business-intelligence.html, June 2007.

[10] Gordon, S., Grigg, R., Horne, M. and Thurman, S., "Service-Oriented Business Intelligence" *MSDN*, http://msdn.microsoft.com/en-us/library/bb245659.aspx, January, 2006.

[11] Grabova, O., Darmont, J., Chauchat, J. and Zolotarova I., "Business Intelligence for Small and Middle-sized Enterprises", *ACM SIGMOD Record*, 39(2). 2010.

[12] Guster, Dennis C., and Lee, Olivia, F., "Effective Infrastructure Protection Through Virtualization", *ICT Ethics and Security in the 21st Century: New Developments and Applications*. IGI Global Publishing, 2011.

[13] Hart, M., "Business Intelligence Projects in Second Year Information Systems Courses", *Proceedings of the South African Computer Lecturer' Association Conference*, ACM, New York, 2009.

[14] Inmon, W., "Building the Data Warehouse", *John Wiley & Sons*, New York, 1992.

[15] Kimball, Ralph; et al., "The Data Warehouse Lifecycle Toolkit", *New York: Wiley*, 1998.

[16] Kuster, J. and Rouse, C., "A Business Intelligence Primer for Higher Education: Applying BI in Higher Education.", http://www.b-eye-network.com/print/9704, 2009.

[17] Lemme, S., "Minding the Data and the Databases", http://searchdatamanagement.techtarget.com/news/1201961/Minding-the-data-and-databases, 2006.

[18] Lupu,A., Bologa, R., Lungu, I. and Bara, A., "The Impact of Organizational Changes on Business Intelligence Projects", *Proceedings of the 7th WSEAS International Conference on Simulation, Modeling and Optimization*, pp.415-419, 2007.

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

[20] Pearson, D., "Key Methods for Managing Complex Database Environments." *Quest Software, Inc. White Paper*, http://viewer.media.bitpipe.com/985899539_355/1296670467_138/Quest_Key_Methods_Managing_Complex_Database_Environments.pdf, 2010.

[21] Pedersen, Torben Bach, and Jensen, Christian S., "Multidimensional Database Technology", *Distributed Systems Online (IEEE)*: December 2001, pp.40–46.

[22] Pucher, M., "Management is not a science: It is an art. Welcome to Real World IT", http://isismjpucher.wordpress.com/category/business-intelligence-2/, 2010.

[23] Rylee, C., Rosen, R. and Tinkler, B., "Selecting a Business Intelligence Standard for Higher Education", *A Presentation at the Mid Atlantic Educause Conference*, January 10, 2006, Baltimore, MD.

[24] Sandler, Dan, "Best Practices for Applying Agile Techniques to Data Warehouses", *BUSINESS INTELLIGENCE Journal*, vol. 15, No. 4, 2010, pp.17-27.

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

[26] University of Cambridge commissioned by KPMG, "BI: Does your business intelligence tell you the whole story?", http://www.kpmg.com/EU/en/Documents/does_business_intelligence_whole_story.pdf, 2009.

AUTHOR BIOGRAPHIES

Dr. Dennis Guster is a Professor of Computer Information Systems and Director of the Business Computing Research Laboratory at St. Cloud State University, MN, USA. His interests include network design, network performance analysis and computer network security. Dennis has 25+ years of teaching experience in higher education and has served as a consultant and provided

industry training to organizations such as Compaq, NASA, DISA, USAF, Motorola, and ATT. He has published numerous works in computer networking/security and has undertaken various sponsored research projects.

Christopher G. Brown is the Lead Business Intelligence Data Architect for Saint Cloud State University. His work focuses on the modeling relational and multidimensional databases for higher education administrative decision making. Current projects include modeling enrollment and retention predictive analytics.

APPENDIX A

Fact/Qualifier Analysis
Instructional Workload/Department Productivity Cube

Business Questions

1. What is the faculty's teaching assignments?

2. What other assignments does a faculty member have?

3. What is the number of student credit hours generated?

4. What is the revenue generated by the faculty member and department?

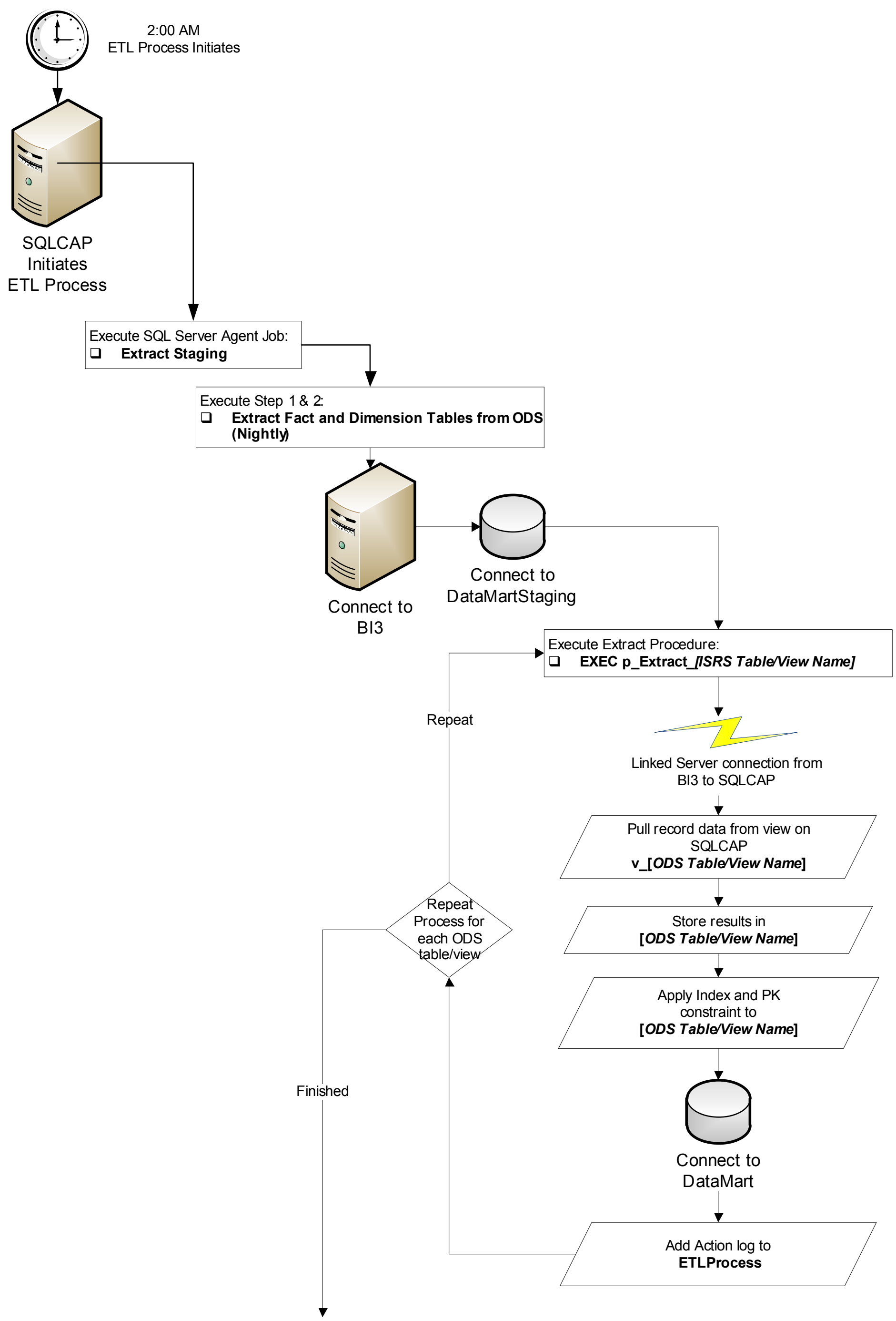
5. What are the expenses associated with the faculty member and department?

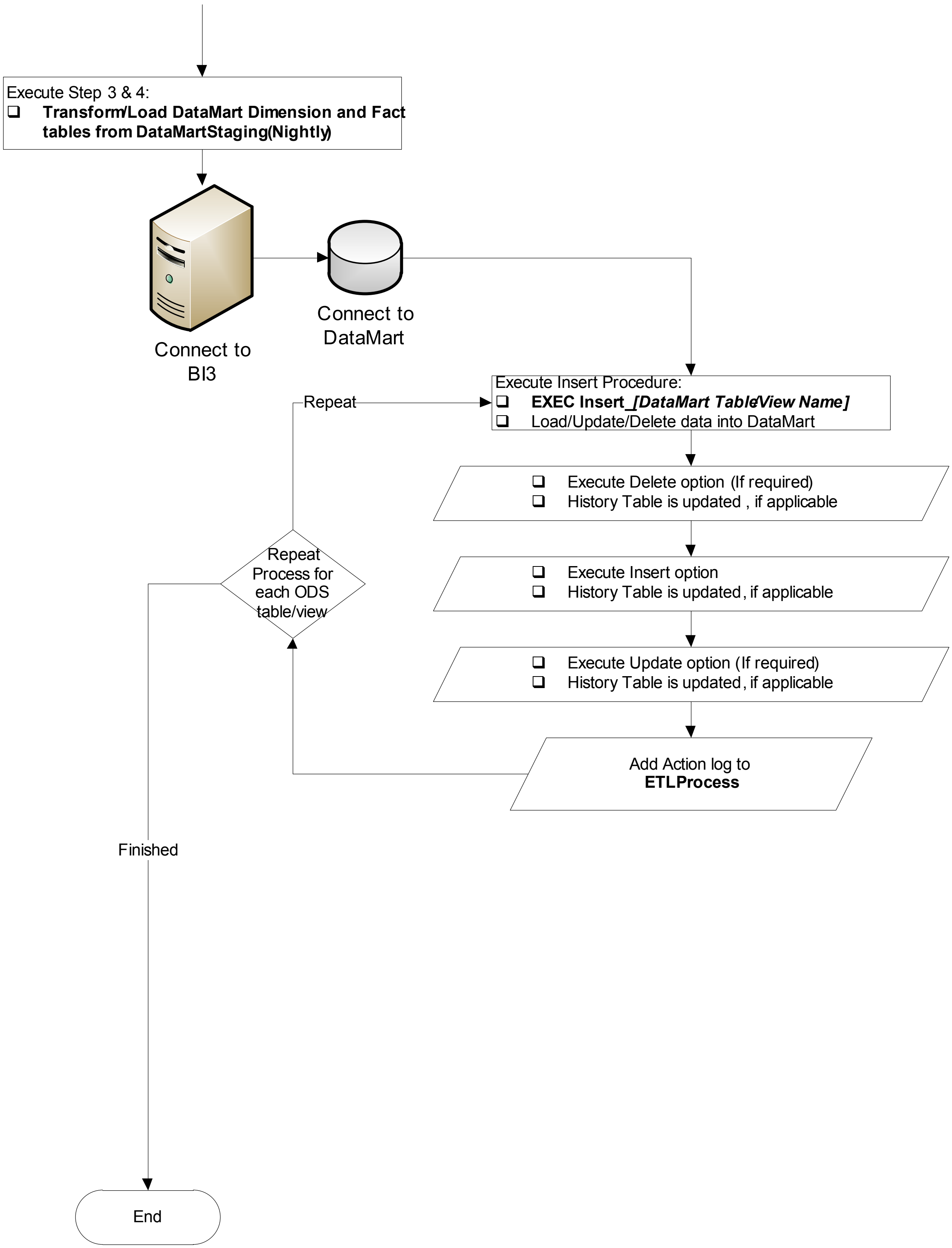
Table 2: Fact/Qualifier Matrix

	Teaching assignments	Other assignments	SCH generation	Revenue generated	Related expenses
Employee's name	X	X	X	X	X
Rostered department	X	X	X	X	X
College	X	X	X	X	X
Faculty Rank	X	X	X	X	X
Faculty Status	X	X	X	X	X
Course title	X				
Course number	X				
Section number			X		
On-load credits for course	X				
Student credits for course			X	X	
# of students enrolled in course			X	X	
Term	X	X	X	X	
Fiscal year	X	X	X		

APPENDIX B

ETL Process




Execute Step 3 & 4:
Transform/Load DataMart Dimension and Fact tables from DataMartStaging(Nightly)
Connect to BI3
Connect to DataMart
Execute Insert Procedure:
EXEC Insert_[DataMart Table/View Name]
Load/Update/Delete data into DataMart
Repeat
Execute Delete option (If required)
History Table is updated , if applicable
Repeat Process for each ODS table/view
Execute Insert option
History Table is updated, if applicable
Execute Update option (If required)
History Table is updated, if applicable
Add Action log to
ETLProcess
Finished
End