Data Warehousing:
An Opportunity for Collaboration between Administrative Computing and Academic Departments

Robert Van Abel
Information Technology Services
University of Wisconsin - La Crosse
vanabel.robe@uwlax.edu

William Wehrs
Information Systems Department
University of Wisconsin – La Crosse
wehrs.will@uwlax.edu

Abstract

The campus administrative computing unit and an academic department at the University of Wisconsin – La Crosse (UWL) are engaged in a collaborative effort to develop a campus data warehouse. A data warehouse is a store of data created to support decision-making. It is separate from the operational systems of an organization, and holds time variant operational and other data that have been integrated by subject area. Data warehousing is the process of developing, maintaining, and employing a data warehouse. A bottom up approach to data warehouse development is being taken. A pilot data mart is being developed for use by the campus institutional research unit. This data mart will be focused on student enrollment as a single subject area.

This data mart will be used not only by campus decision makers, but also by students engaged in coursework. Both uses will employ the same data mart, but sensitive data will not be available for instructional use. There are a number of courses in the undergraduate and graduate curriculum where a data warehouse may be employed in instruction. These include courses requiring technical prerequisites focusing on the data warehousing process, as well as courses in the business curriculum focused on the use of a data warehouse in support of decision making.

The development of this data warehouse as a collaborative effort is based on a community of interest between administrative and academic units. The collaboration produces benefits for both parties. On the academic side one of the benefits is professional data server administration. A number of challenges must be faced in order to successfully complete the development process. These challenges relate to technical and budgetary issues in development, and to difficulties associated with dual administrative and instructional use of the resulting system.

Introduction

Computing-intensive academic departments such as Computer Science (C-S) and Information Systems (I-S) acquire hardware and software for specialized instruction in their discipline area. The technology is typically funded by outside grants, or on the basis of initiatives sponsored by the local campus or a state system such as the University of Wisconsin. However, the technology is usually not accompanied by personnel to support it. Computing-intensive departments at small to medium-size campuses typically have little or no staff to maintain and support their specialized technology. The support function is usually undertaken as service by faculty, through part-time student help, or some combination of the two. Another option is support from the campus IT unit. However, in the absence of other considerations, support from the IT unit is often contingent on its use by a broad cross-section of students.

This paper provides a prospective account of a collaborative effort between the campus Information Technology (IT) unit and an academic department at UWL, in which the campus IT unit provides significant support for the department's specialized instructional needs. A key characteristic of this collaboration is the development of a community of interest between the two units.

Information Technology Services (ITS) at UWL is in the process of migrating existing administrative systems from a proprietary, legacy system to an open systems environment based on relational database technology. Associated with this migration is a data warehouse whose purpose is to service data analysis and reporting needs of the campus.

Data warehouses and data marts emerged in the 1990's as a practical solution to the problem of drawing together data to support reporting requirements of managers. Universities, including the University of Wisconsin system, are increasingly investigating this as a solution to complex reporting requirements, often from external agencies, that bridge admissions, accounting, budget, personnel and student systems. Operational systems that are intended to primarily support transaction processing often do not preserve detailed historical data and may not support heavy query activity without impacting system performance. Universities have discovered that data marts can support a growing variety of uses. For example, the University of Wisconsin – Stout has developed a data warehouse project that provides data for the student portal, as well as more traditional roles to support management.

The development and use of a data warehouse is also a subject area of increasing importance within the academic area of Information Systems. But available instructional material provides little opportunity for students to have hands-on experience with a data warehouse. It would be very beneficial to have a functional data warehouse to employ for instructional purposes. Complex relationships that exist in a functional data warehouse are difficult to create in current instructional material. While some software employed in the process of data warehousing is available from vendors, a data warehouse of sufficient size, and that could be employed by students, does not appear to be available.

With this in mind, ITS personnel and faculty from the I-S department are collaborating in the design and construction of a campus data warehouse that will provide administrative decision support, and an instructional version that will be available for use by students.

This document is divided into five sections. The first section briefly describes the basic characteristics of a data warehouse. The second section presents the development process and features of the UWL data warehouse. The third section examines the instructional version of the UWL data warehouse. This includes how the instructional version will differ, and the way in which it will be employed within the curriculum. The fourth section provides a discussion of the mutually beneficial characteristics of the project for both ITS and for the I-S department. This discussion includes characteristics of the IT environment at UWL, which may be present at similar campuses, that provide the foundation for a community of interest between the respective units. The document concludes with a presentation of the challenges faced by the project.

Basic Characteristics of a Data Warehouse

A data warehouse is a collection of data created to support decision making. According to Inmon [1], “A data warehouse is a subject-oriented, integrated, nonvolatile, time-variant collection of data in support of management decisions.” The characteristics Inmon identifies distinguish the warehouse data from its primary source – the operational data of a business organization. While data in operational systems tend to be organized by business process (e.g. order entry, accounts receivable), warehouse data is organized around major subject areas for the organization (e.g. sales). In organizations, data exists in these independent business process systems. In a warehouse, data is consolidated and integrated. Typically, data in the warehouse is read-only. In operational systems, users are continually changing or updating system data. Lastly, while operational data is usually only the most current, a warehouse maintains historical data.

A data warehouse is a store of data. Data warehousing is the process of developing, maintaining, and employing a data warehouse. This process includes extracting the data from source systems, transforming and loading the data into the warehouse, and access to the warehouse by applications and end users. A tutorial on data warehousing is included in [7].

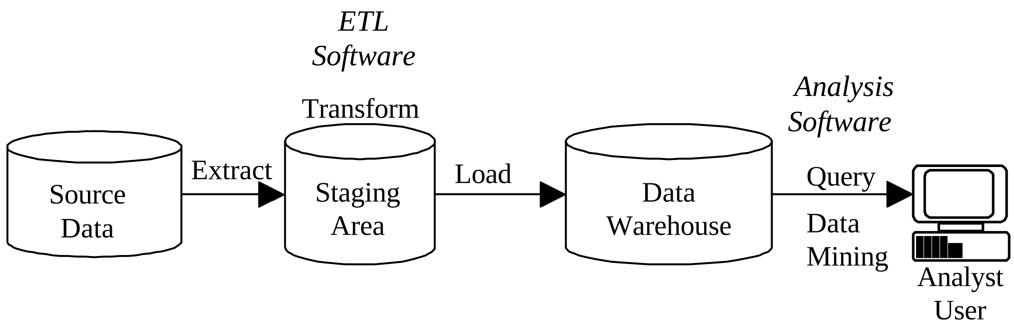

Figure 1: Data Warehousing Architecture

An architecture for data warehousing (Figure 1) shows the components and relationships between the components. The bulk of the source data comes from operational systems. It may also come from external sources such as government organizations. In a specific business context, each of which has unique characteristics, the ETL (Extract, Transform, Load) process is difficult and time consuming. This process involves data extraction from diverse systems, transformation of that data in order to integrate it into subject areas, and loading the data into the warehouse. The software to accomplish this process can be custom-developed or purchased on the commercial market. A staging area provides a middle-ground for transformation processing to occur. Transformed data is loaded into the warehouse. A first load provides initial data that typically covers a number of time periods. Subsequent loads serve to refresh the data by adding data pertinent to the time period since the prior load. The length of the refresh cycle varies between organizations.

The technology employed to store source data is as diverse as the systems from which these data are extracted. The dominant storage model for the staging area and for the warehouse is a relational database. However, in the warehouse, in order to facilitate efficient retrieval of large data sets, and to provide a dimensional view of the data, a non-normalized data model known as a “star schema” is employed. For a given subject area, this schema takes the form of a normalized fact table related to multiple non-normalized dimension tables. The fact table usually contains numeric data regarding the actual transactions or values in the subject domain (e.g. sales). Dimension tables contain categorical data pertinent to those transactions or values (e.g. time, customer, part, etc.). Because a data warehouse contains historical data, time is always one of the dimensions. An example of a simple star schema based on the Northwind database included with Microsoft Access can be found at [4].

End users access warehouse data using a wide variety of analysis software. These tools vary widely in terms of their sophistication and the skills required of users. Users who understand relational database technology are able to write their own SQL queries. Many users employ a tool that presents a query-by-example (QBE) interface and subsequently translates the completed QBE grid into SQL code. Query tools such as these are often provided with a capability to do both multidimensional analysis and reporting. Data mining is a set of activities used to find new, hidden, or unexpected patterns in the data. This involves the use of specialized algorithms (e.g. neural networks) many of which were created by academic researchers and are now finding their way into data mining software products. Consequently, effective use of data mining software requires an analyst with substantial skills.

In addition to the data itself, a data warehouse must also store and manage "metadata" about the data in the warehouse. Metadata must include details related to the ETL process that would be employed by personnel who service the data warehouse. It must also include details required by end users in order to access and analyze data pertinent to their information needs.

There are two general approaches to systems development for a data warehouse; top-down and bottom-up. A top-down approach proceeds from the point of view that if the organization's goal is an integrated data warehouse containing multiple subject areas, then a large development effort is called for from the very start in order to address issues of data volume, integration, and governance associated with a warehouse of that scope [1]. This approach is subject to the risks associated with large-scale development projects. On the other hand, the bottom-up approach seeks to limit that risk while still maintaining the ultimate goal of a warehouse with broad scope [2]. This involves the incremental development of data marts. A data mart is similar to a warehouse, but only stores data for a limited (perhaps one) number of subject areas. Due to its smaller scope, a data mart can be developed more rapidly and with less risk. If successful, the data mart may be expanded to include additional subject areas. However, it is difficult to successfully expand into an integrated data warehouse. Successful expansion requires foresight in several arenas, including a data model with the potential for integration, and a scalable architecture.

The UWL data warehouse

UWL has started a data warehouse pilot project. The Institutional Research unit is the client for the pilot. Data from operational administration databases will be extracted, transformed and batch loaded to form an initial data mart. Since an important management area under the purview of Institutional Research is student enrollment, the subject focus of the data mart will be a student enrolled in a particular course section taught by a faculty member.

Development process

Data warehouse modeling presents a range of challenges and continues to be the subject of debate between database designers and modelers. "Forget everything you know about entity relationship data modeling… using that model with a real-world decision support system almost guarantees failure." [3].

The general approaches to building a data warehouse are 'top down' and 'bottom up.' Our approach is bottom up. We are creating a single data mart as a pilot to provide decision support for a key organizational need. In this case, Institutional Research expressed an interest in working to create a pilot data mart for student enrollment decision support. Our goal is to create a data mart that can be extended over time to support other organizational needs such as admissions, budget and course schedule management.

The development of the pilot data mart is constrained by a migration project between operational systems. At UWL, operational data currently resides in a Unisys DMSII legacy database that is more than 17 years old. This database is not relational. Data relationships are created within the COBOL programs that utilize it. The migration from DMSII to Oracle 9i is part of an architecture migration project that is expected to go-live on July 22, 2003. In order to eliminate the cost of rewriting the COBOL programs that process legacy operational data, the Oracle database structure had to mirror the DMSII structure. The Data Mart will be fed from this legacy Oracle database as depicted in Figure 2.

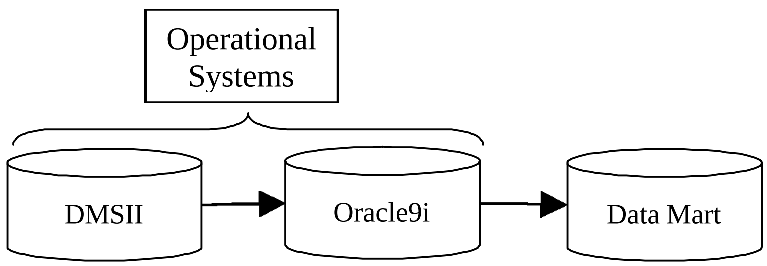

Figure 2: Source Data for the Pilot Data Mart

In the DMSII database there is no primary key definition. Key structures are maintained in the COBOL code. A sequence number was created as a primary key for each Oracle table as the DMSII data was migrated. This sequence number serves to preserve COBOL data sorts that vary from one program to another.

Referential integrity is not enforced within the operational database. Defining table relationships as part of the DMSII to Oracle migration may seem sensible on the surface, but creates unintended consequences for the legacy transaction programs. For example the COBOL structure to sort duplicate ascending or descending records is maintained in our migrated Oracle database by ordering the primary key sequence number. Creating primary keys using logical elements such as social security number or student ID number would provide for referential integrity between tables and would restrict the table from containing duplicate entries. This may seem attractive from a relational design perspective, yet, there is good reason that the legacy system allows duplicate social security numbers. In this case, foreign students are assigned a 111-11-1111 social security number until the actual number is received.

A data warehouse or mart must work with existing source data. The student enrollment data mart must load data in a manner that provides data relationships that will support queries generated by decision support tools that campus users are planning to use. In view of the lack of logically meaningful relationships in the tables of legacy data in the

Oracle database, these relationships will need to be carefully created in the data mart during the data extract, transform and load (ETL) processing. Currently we are defining these relationships for the data mart model presented here.

Data detective work to determine data definitions is needed. Much of the legacy data is in the form of numeric codes. Documentation that defines the meaning of these codes is limited and some definitions have changed over time. It is important that the information contained within these codes be defined consistently in order for campus users to create queries that are meaningful. Such specifications and definitions are a critical component of the metadata. Metadata definitions can be derived from COBOL programs used to generate reports from the legacy system. However, there are more than one million lines of COBOL code contained in 783 batch programs to be analyzed. We will also need to provide metadata in a format that is easily understood by query users. In light of these issues, building from the bottom up allows a focused effort on metadata development that can be completed within a reasonable timeline.

The student enrollment management data mart pilot has several milestones. The data mart model is currently undergoing review and refinement. Legacy data is being mapped to the model as the data mart model is created. This first step includes metadata definitions. The completion date milestone for this current step is May 2003. The next milestone will be to create the extract, transformation and loading (ETL) process for the data mart so that legacy operational data can be loaded and refreshed. This process will be designed to use the migrated Oracle 9i operational database as the source data for refreshing the data mart pilot on a daily basis. We will evaluate several tools to perform this step, including Informatica and Oracle's Data Warehouse Builder tool. This project milestone is to have periodic ETL processing in support of the student enrollment data mart by Aug. 2003. Queries against the data mart must be developed to support Institutional Research. One goal of this project is to provide end users with data access for decision support. Toward this goal, we will support a variety of query tools including Microsoft Excel, Access, Brio Intelligence and native SQL code. Many end users may want to use predefined queries rather than creating their own. As part of providing data access, we will form a query library. A Dec. 2003 milestone includes query development and the creation of a web interface for authenticated users to access a query library. The data mart may require data that is contained in other source systems. The final milestone is to develop ETL processes that can incorporate data from multiple sources into an expanded student enrollment data mart. This milestone should be completed by winter of 2004.

Features

The initial model for the pilot data mart is illustrated in Figure 3. This model is based upon a star schema that supports the dimensions identified by Institutional Research as required for student enrollment management. These are the Student, Course, Faculty, Facility and the Date dimensions. These dimensions are related through a central fact table.

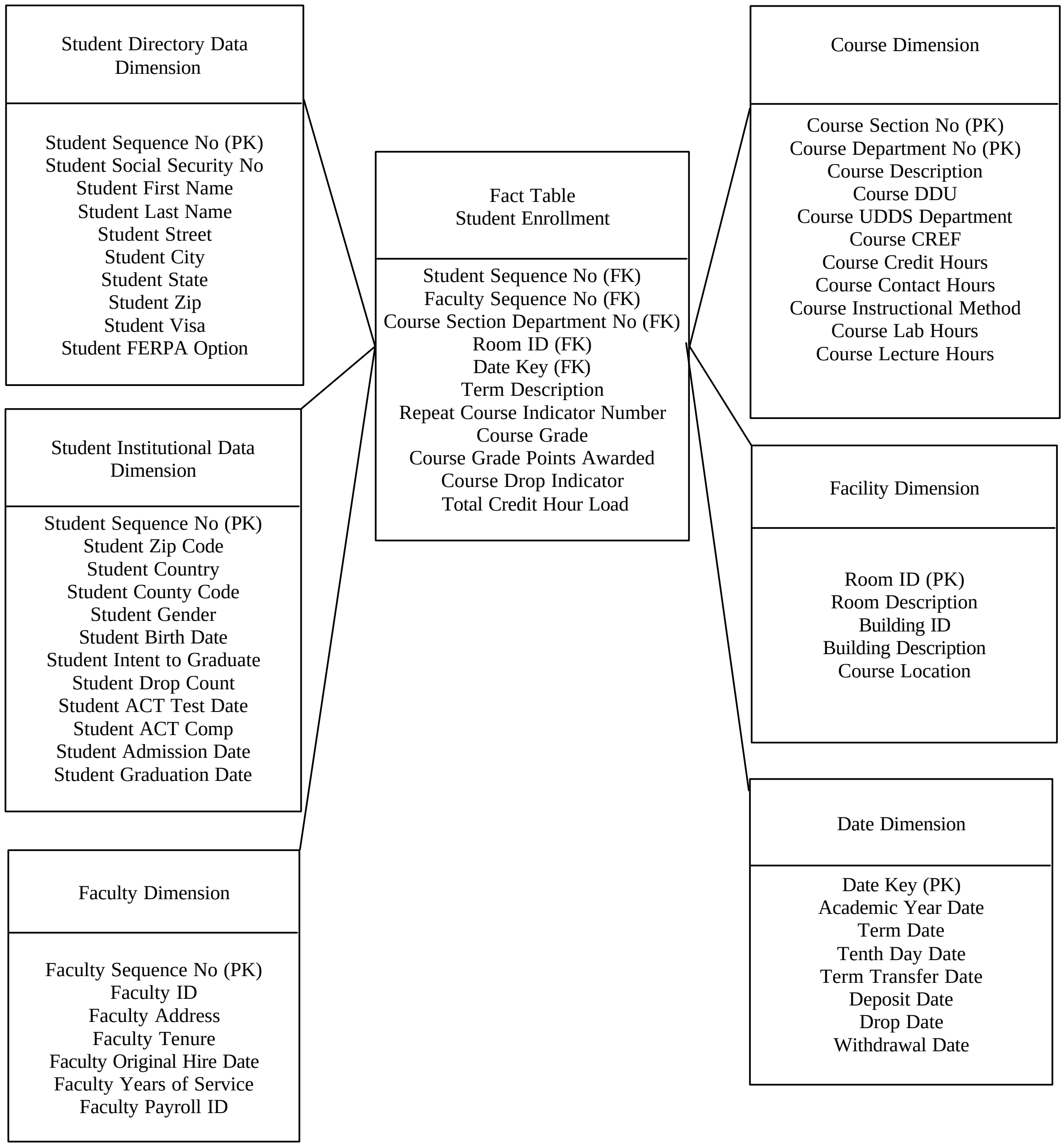

Figure 3: A Model for the Pilot Data Mart

A typical star schema model has a fact table that holds transaction data in atomic or summarized form, and dimension tables that categorize the data in the fact table. A normalized operational database schema is composed of related tables at various levels of detail. In such an operational schema, multiple joins must be undertaken to produce a query result with a given level of detail. As opposed to this, since each dimension table in effect contains redundant (pre-joined) data, a pure star schema supports only one level or dimension. This structure makes it easy for users to formulate queries based upon dimension categories that aggregate data from the fact table. Furthermore, even if such queries produce result sets with a large number of records, they run more rapidly than equivalent queries in a normalized schema. Some data marts do support multiple levels. These variants are known as snowflake schemas.

In the student enrollment process, events need to be recorded as the simultaneous occurrence of dimensional entities. The fact table contains one row for each course by student, by faculty, by room (facility), by term event (start or tenth day of term). This design is termed a *factless fact table*. "Events are often modeled as a fact table containing a series of keys, each representing a participating dimension in the event. Event tables often have no obvious numeric facts associated with them and hence are called factless fact tables." [2 pp. 248] The fact table, as modeled, represents a blend of event and numeric facts.

Maintaining a repository of historical information is one of the most important features of the data warehouse. Institutional users interested in student enrollment are often interested in trends. The current operational database reliably stores the current position, but often over-writes data as terms proceed from one to the next. Trend analysis requires information that is built over time and retained after it is no longer required in the transactional system. The challenge of preserving time-dependent data is greater for the data warehouse model than for the operational database model.

One example of how historical data from this data mart could be employed follows: Do student GPA scores parallel their incoming comprehensive ACT score? This question could be answered without exposing FERPA information by providing table access to the Student Institutional Data, the Course Dimension and the Date Dimension as well as the central Fact Table of the pilot student enrollment data mart. A query could be designed to return the ACT comprehensive score for each student sequence number by year admitted. Each student's cumulative GPA is essentially calculated by dividing the cumulative Course Grade Points Awarded by cumulative Course Credit Hours attempted. The cumulative GPA could be listed by year admitted for each student sequence number and compared to the incoming ACT comprehensive score. A subquery could easily provide a mean GPA and ACT for all student sequence numbers in a year admitted, or a median if the GPA or ACT scores did not follow a normal distribution. This query could also be extended to determine if there was a difference between students who graduated and students who have not graduated and students who failed to graduate within four years from the admissions date. Perhaps further analysis based on data resulting from this query could provide some insight into whether grade inflation was real or imagined over the period in question.

An instructional data warehouse

Data warehousing and the analysis of warehouse data are topics of increasing importance to students involved with information technology, such as students in I-S and C-S curricula. There is a demand for expertise in this rapidly growing area. Software, especially software for analysis, is available for instructional or demonstration use at little or no cost from vendors who provide that software to commercial markets. However, large data warehouses that provide a wealth of exploration opportunities in areas such as design, maintenance, or analysis, do not appear to be available for instructional use. This observation is largely based on unsuccessful efforts to locate an instructional data warehouse associated with a textbook that covers this topic. In order to verify this observation, several potential sources for an instructional warehouse were investigated. The on-line materials of a number of major data warehousing vendors (IBM, Microsoft, Oracle) were reviewed and each vendor was contacted via email. These search and associated contact activities did not turn up any evidence of a demonstration data warehouse that was available for instructional use. A search of on-line resources for a number of universities was also unsuccessful. Lastly, contact with the data warehousing institute produced no information regarding a demonstration or instructional data warehouse.

In view of these results, a data warehouse that could be employed as an instructional tool would need to be developed or acquired. One alternative would be to develop an artificial warehouse via simulation. Business models are available that could provide a structure for generating data sets. One source for such models are the gaming simulators often employed in strategic management courses. However, duplicating the quantity and quality of data for instructional purposes that is generated in a legacy database would be a tremendous task. A second alternative is to encourage a private sector organization that has achieved notable success with a data warehouse, to make the warehouse available for instructional use. However, such organizations view successful warehouses as strategic assets. Consequently, they are unlikely to expose them to public view. Furthermore, they are also sensitive to issues of confidentiality with respect to warehouse data.

Yet another alternative is the approach presented in this paper; to collaboratively develop a data warehouse for the university and also employ the warehouse for instruction. Aside from the collaborative benefits, an additional advantage of such a warehouse is that it is a familiar problem domain for students attending the university.

Differences between the two data warehouses

The differences between the administration and instructional uses for the data mart are primarily derived from a requirement to restrict certain confidential and sensitive data access to users demonstrating a legitimate need.   Security is provided by physical separation of the student directory data from the student institutional data and by database views that provide access to sensitive data to authenticated users with a legitimate need.

This pilot data mart will contain confidential information that falls within the Family Educational Right to Privacy (FERPA) Act. This federal law protects the privacy of student information records. Access to FERPA data is controlled and the “institution is required to maintain records that indicate all agencies, individuals or organizations that have requested or obtained access to a student’s education records … which will indicate specifically the legitimate interest that such agency, individual or organization has in obtaining the information"[6]. This restriction creates a high barrier for access to much of the student enrollment data. Our strategy to provide broad end user access while maintaining student privacy, is to create a dimension with student directory information that is distinct from student institutional data. There is a one to one correspondence between these two dimensions based on the student sequence number. Control is focused on limiting access to both the Student Directory Data Dimension and the Student Institutional Data Dimension to only users that meet the institutional legitimate interest test. Access to the Student Directory Data Dimension without the Student Institutional Data Dimension is primarily for institutional directory information. There is a FERPA switch to indicate students that choose to opt out of this disclosure. Access to the Student Institutional Data Dimension and not the Student Directory Data Dimension is provided to a variety of users since the individual is masked from any query structured against this data. This last group includes instructional and research oriented users.

A database view is an object that presents a customized presentation of the database. It is an SQL query that retrieves data from the physical database and presents it like a table. Another approach would be to create database views of a *combined* student data dimension that would present the data to preserve access control. This approach is valid, but the data owners responsible for managing FERPA access are not comfortable with it. Rather, they preferred a more fundamental separation in the form of distinct dimension tables. We will use database views to add additional security to the data mart design and to further simplify the presentation of the data mart for end users. For example, the Faculty Dimension contains sensitive information, such as salary data, that will be restricted to users with a legitimate need. The instructional users will have access to a Faculty Dimension view that will not contain salary, or other sensitive data.

The instructional warehouse in the curriculum

In the university curriculum, courses that could employ an instructional data warehouse fall into two groups. One group would be technical courses that required prior student experience (i.e. prerequisite courses) with the technologies employed in data warehouse design, development and operation. These courses would focus on data warehousing as depicted in Figure 1.

One entry in this group would be an elective course devoted to Data Warehousing at the junior or senior level in an undergraduate I-S program. With appropriate background in database management, programming, and the business core, all components of the data warehousing process could be explored. In particular this would include the design and development process of the data warehouse to include the development and implementation of the ETL processing.

A second entry could be an undergraduate course in database management or advanced database management. This type of course would focus on the design of a data warehouse relational database with the instructional data warehouse as a working example. An advanced database course might also examine data warehouse administration issues. For example, indexes are critical to speed a query that is processed in a data warehouse. A warehouse contains many more indexes than an operational system. Determining the correct index strategy for a warehouse is an important ingredient in successful warehouse performance.

A second group would be courses requiring less technical background on the part of the students. These courses would focus on data retrieval and analytic use of the data warehouse.

An entry in this group would be an upper level course in decision support within an I-S curriculum that focused on data oriented decision support. Such a course would provide an overview of the data warehousing process but would put primary emphasis on skill development in retrieval, reporting and analytics such as multidimensional analysis and elementary data mining.

A second entry in this group would be an MBA level course in Information Systems. Many MBA programs now have a core or elective course in I-S. An MBA course often provides a broad overview of the management of information technology. As part of such a course, a module would appropriately include an introduction to, and experience with, the instructional data warehouse. This would typically take place in a decision support or strategic management context.

The last entry in this group would be a variety of undergraduate courses in the functional areas of business (Accounting, Finance, Production, Marketing, Human Resources, etc.). Which of these functional areas could include a module on the use of a data warehouse would depend on the subject coverage of the university/instructional data warehouse. While the university warehouse will focus on subjects of interest to university decision makers (e.g. student admissions), some of these problem contexts are very similar to contexts faced by traditional business firms. For example, tracking student admissions with the purpose of attracting and admitting the best students to the university, is comparable to a business firm tracking job applicants in the interest of hiring a well-qualified and productive employee. Consequently, data warehouse analytical techniques focused on the admissions area can be made very relevant to a student who has not only personally gone through the admissions process, but is currently a student majoring or taking a course in human resource management.

A community of interest between ITS and academic departments

ITS at UWL has focused on education as the primary mission of the institution. While their main responsibility has been the operational administrative systems, in view of the campus mission they have been responsive to the needs of the computationally intensive academic units such as C-S and I-S. In that regard, in addition to indirectly supporting that mission via the administrative computing infrastructure, they have been directly supporting instructional needs. The instructional data warehouse is viewed as a mutually beneficial opportunity for such direct support.

As a component in the enterprise cluster of servers to be employed for the new administrative systems, ITS has provided a server dedicated to instructional use. This server provides database services based on an Oracle 9i database management system. The use of Oracle database software for this purpose was made possible by UW system and UWL contractual relationships with Oracle, and participation by academic departments in the Oracle Academic Initiative. Oracle 9i client software has been installed in campus academic computing labs, and faculty using the database software for instruction have management software as well as a set of management privileges relative to the classes in which the software is employed. Perhaps most importantly, ITS provides an experienced Oracle database administrator to administer this instructional data server. Lack of adequate technical support for installing and administering an Oracle data server is an oft-cited issue by faculty using it for instructional purposes [5 pp. 27]. Currently this data server is employed by the C-S department in both the introductory and advanced database courses. It is employed by the I-S department to support a senior level class in systems design and implementation. This server is also one of the candidates from which a host for the pilot data mart will be selected.

The collaboration also provides benefits for administrative computing. First of all, in a period of budget crisis, support services that are not considered critical to the organizations' mission may be expendable. By allying themselves with the educational mission of the institution, ITS may be able to provide itself with some protection from a reduced flow of resources.

Administrative computing by nature tends to be conservative in terms of the adoption of new technology. This largely results from the need for availability of those systems critical to institutional operations. Engaging in activities on the academic side allows ITS staff to become familiar with "bleeding edge" technology and to test the use of that technology in a context that is somewhat more tolerant. For example, the Linux operating system will be tested out on the instructional data server prior to use on other servers in the enterprise cluster. Correspondingly, the instructional data server will be "opened up" to use by students from off-campus in order to shake down security issues associated with external access.

Given the rapid expansion in the use of information technology and a relatively constrained resource environment, campus computing units, including administrative computing, tend to be understaffed and under funded from a personnel perspective. Collaboration with academic units provides them with direct access to relatively inexpensive human capital in the form of student workers. Students in C-S and I-S academic programs have academic experience with modern information technologies. Some of them are very highly skilled. The C-S department has recently developed and begun offering a graduate program in software engineering. This program requires students to undertake a significant software development project. Both the undergraduate C-S and I-S programs offer internships for academic credit. This provides well-prepared undergraduates with an opportunity for jobs with ITS. Some of these students continue on as permanent employees.

Lastly, at UWL this collaboration has also been fostered by the location of the units. The campus IT unit (ITS) and the C-S and I-S departments are housed in adjacent areas of the same building – Wing Technology Center. The physical proximity facilitates interaction between the people from these units.

Challenges for this project

Successful completion of this project within the specified timeframe requires us to address a number of challenges. Two of these have been discussed earlier. First, the transformation component of the ETL process is relatively complex since the tables in the operational Oracle database do not contain logically meaningful relationships. Furthermore, most legacy data is represented by numeric codes. Occasionally the meaning of the codes has changed over time. Within the warehouse these numeric codes should be translated into text in order to make the information meaningful for end users. Second, in some cases a data definition for the operational source data must be inferred from the logic of COBOL code that processes the data. This will make development of quality metadata a time-consuming process. Approaching warehouse development from the bottom-up will serve to make both these challenges somewhat less overwhelming. Furthermore, with the pilot data mart we hope to gain experience in these areas that will prove fruitful when expansion to additional subject areas is undertaken.

Given the state of the economy and state government budgets, resources to undertake the project are unusually constrained. In general the hardware necessary to implement a data warehouse is in place. And, the software for use by campus decision-makers is available locally or via the University of Wisconsin System. The binding constraint is people time to supervise and carry out development activities, as well as the administration of the warehouse once it has been implemented. In general, personnel lines are contracting rather than expanding. An experienced Oracle database administrator is in place within ITS. However, this person's time is fully committed to administering the cluster of Oracle data servers.

A host site for the data warehouse needs to be specified. There are two broad alternatives; local hosting or remote hosting. Local hosting provides local control and the most flexibility with respect to the instructional use of the data warehouse. On the other hand, it also requires the commitment of scarce local human resources and requires local installation and management of ETL and Analysis software. Remote hosting is available via an application service provider sponsored by the University of Wisconsin System. There are significant advantages to this option. At the present time this service is provided to system campuses at no direct cost. The service also provides for the use of ETL and analysis software. However, licensing arrangements between UW System and certain software vendors appear to limit the availability of that software for instructional use.

Finally, dependent on the use of the data warehouse in the curriculum, supporting instructional material must be acquired or developed. While data warehouse development and implementation is proliferating in the business world, it appears to be in a very early stage as an instructional topic in higher education. One can find discussions of data warehousing in textbooks focused on decision support. Usually such discussion will be contained in one chapter, or at most, several chapters. Although there are a great many practitioner-oriented books on data warehousing, there are hardly any textbooks. Consequently, instructional material in support of courses focused primarily on data warehousing is not available off the shelf. At the present time these supporting materials will have to be generated by instructors who intend to provide instruction in data warehousing.

References

1. Inmon, W.H. (1992) *Building the data warehouse.* New York: Wiley.

2. Kimball, R. (2002) *The data warehouse toolkit*, 2nd edition. New York: Wiley.

3. Kimball, R. & Strehlo, K. (1994, June 1). Why Decision Support Fails and How to Fix It. *Datamation*, 41-45.

4. Moore, B. (2002). *NorthStar Data Cube Demonstration Contents.* Retrieved March 15, 2003, from http://www.pinetreedata.com/data_warehouse_demo.htm.

5. Morrison, J. & Morrison M. (2001). Using ORACLE to augment the Information Systems curriculum. *Communications of the Association for Information Systems,* 7, 1-36.

6. U. S. Congress. The Family Educational Rights and Privacy Act (FERPA) (20 U.S.C. § 1232g; 34 CFR Part 99).

7. Watson, H.J. (2001). Recent developments in data warehousing. *Communications of the Association for Information Systems*, 8, 1-25.