Client/Server Based Data Warehouse at the University of Arizona

Cheryl Bach-Scott
Htay Hla
Michele Singleton
Elizabeth Taylor
Paul Teitelbaum

The University of Arizona
Tucson, Arizona

Abstract

The University of Arizona is implementing a data warehouse on a relational database to be accessed by desktop tools using client/server technology. The warehouse is a repository of data extracted from the "legacy" systems and loaded into a Digital Equipment Corporation's Rdb database. The database contains data from University Student Information, Payroll/Personnel, and Financial Account systems.

This presentation will provide an overview of how information is acquired currently for decision support and other management needs. The client/server model being implemented will be described including the warehouse system configuration and the client desktop tools being used. Management issues related to this project will also be discussed along with the associated data administration activities. The presentation will include a live demonstration of the system if Internet connectivity is available in the presentation area.

The targeted audience is information services personnel at all levels of institutions with basic to intermediate knowledge of relational database and client/server computing.
Tucson, located 50 miles north of the Mexican border in the beautiful southwest, is the home of the University of Arizona. Rich in Spanish, Mexican and American Indian heritage and now populated by a melting pot of people from all over the world, it provides a perfect backdrop for an institution of higher learning.

The University of Arizona campus is large in both acreage and population. The campus covers 339 acres and has over 150 buildings. Serving over 34,000 students each year, the University offers some 150 majors in 12 different colleges. A world leader in astronomy, space science, and systems engineering, the Institution is ranked 19th in the nation for research and development. Last year over 50 National Merit finalists chose to attend last year. Of the 4000 faculty members, 85% have earned doctoral degrees and some are noted Nobel and Pulitzer prize winners. Over 5,000 staff members provide supporting services to this large and diverse campus environment.

The University's computing services are provided by the Center for Computing and Information Technology, commonly referred to as CCIT. Led by the Associate Vice President for Computing and Information Technology, CCIT has four major areas of responsibility: Decision Support, User Support, Telecommunications, and Operations. Each group plays a critical role in supporting the University's administrative systems, but it is the Decision Support analysts and technicians who are responsible for the maintenance and enhancement of the major administrative computing systems. Decision Support is comprised of five groups: Distributed Data Management, Human Resources, Financial Records, Student Information, and Production Services.

Decision Support provides the support for the three major administrative systems: the Student Information System (SIS), the Personnel Services Operating System (PSOS), and the Financial Records System (FRS). The Student Information System provides detailed information about student admissions, financial aid, registration, orientation, transfer evaluation, and billing. The Personnel Services Operating System is responsible for payroll, applicant processing, budget and employee records maintenance. The Financial Records System maintains all accounting, purchasing, accounts payable, accounts receivable, and fixed assets data for the University. Each of these legacy systems is extremely large and contains millions of lines of COBOL code, hundreds of reporting programs, and dozens of screens, combining to provide a support various processes on campus.

The application environment is also very diverse. SIS and FRS reside on an IBM 309.0 with SIS data stored in VSAM files and FRS data stored in IDMS/R files. PSOS executes on a PRIME minicomputer with its data in ISAM files. This configuration of multiple hardware platforms and separate, incompatible database management systems complicates the data processing environment and adds a level of difficulty when combining data items that span the systems.

These three administrative computing systems serve a myriad of customers located throughout the large campus. The growing information requirements of these customers are as diverse as the customers themselves. To support their decision making process many of these customers have gained expertise in the report writing tools available to them (e.g., SAS, Z-writer, OLQ, Focus). While these tools satisfy many of their reporting needs, there are many occasions when a customer determines a need for
information that they cannot provide for themselves. It is the Decision Support staff that they typically rely on to provide them with this information.

The interaction between the customer and the Decision Support analyst is a paper process known as a Service Request. The Service Request Form allows the customer to identify problems, enhancements, or information requirements. The Service Request is routed to Decision Support where it is assigned to an analyst and then passes through the typical analysis, prioritization, coding, testing, and customer approval steps. The coding, testing, and customer approval phases are often repetitive and time consuming. These phases may include mailing output from analyst to customer and back again or scheduling meetings where the analyst and customer can review the output and determine modifications. Once the customer approves the output and signs the Service Request Completion Form, the job is scheduled and formally documented. Finally, the job can be placed into the production environment. This entire process of gaining information to make a decision can take anywhere from a few hours to several weeks depending on the complexity of the request, the mainframe loads, and the workload of the analysts. Often the lengthy Service Request process results in customers making decisions without information that would be available to them in a more dynamic computing environment.

The combination of the time it takes to complete a Service Request and the increasing information demands by customers has resulted in the Service Request queue stretching far into the future. Additionally, the increasing demand for information has resulted in an increased mainframe workload. The mainframes are rapidly becoming saturated. There are not enough computing cycles available to service all computing needs. This situation still exists today, however, some time ago, CCIT began to explore alternatives in order to provide better computing services. The available options seemed to be either a costly hardware upgrade to the mainframes or moving applications toward a technology which would provide a more dynamic environment by taking advantage of the computing resources residing on the customer's desktops.

Decision Support began pursuing client/server technology as a means of providing the dynamic computing environment that was needed. Client/Server was pursued for several reasons. First, the model would make the most effective use of existing computing resources by cooperatively executing a process or transaction on two or more computers. In addition client/server technology would allow for the migration of computing processes away from the mainframe. The desktop would provide the computing resources for data presentation and manipulation. The resources associated with data management would be provided by a separate machine fine-tuned for that purpose. Finally, the desktop tools would provide a much more friendly interface for the customers, and allow them to have control over their own data needs.

An implementation team was formed and a pilot project to develop a data warehouse utilizing client/server technology was launched. This project was to be a prototype and its primary objectives were to prove the technology and to provide reporting and inquiry access to the data currently housed in the three legacy systems. The project was named the University Information System (UIS).

Early in the project, the UIS implementation team determined that the data warehouse had to be stored in a relational database. The relational model offered many advantages
over the older hierarchical and network databases: easier data access, more flexible database design, and changes to a relational database structure could be affected in minutes, as compared to hours or sometimes days for the network or hierarchical models. Lastly, Structured Query Language (SQL), recognized as the standard for accessing relational databases, would provide an effective data access mechanism. Digital’s Rdb product was chosen as the relational database and a DECvax-4000 chosen as the UIS server. The Digital Educational Software Library/Campus-wide Software License Grant provided the University with the ability to purchase the software at a significant discount.

The next step in developing the prototype was determining the data elements to reside in the warehouse. The server did not have enough disk storage to duplicate the volume of data in the three administrative systems. To effectively determine which elements to select for populating the new UIS database, an inventory was taken of data elements used to process ad-hoc reports. Any element referenced more than once was selected. Today these selected elements, which represent approximately 20% of the total data elements in the mainframe administrative systems, comprise the UIS database.

With the contents of the data warehouse resolved, the implementation team began experimenting with the many tools available to access the Rdb data warehouse. Today, customers who wish to access the data in the UIS data warehouse may choose from a variety of client tools depending on their needs or familiarity with various products. For example, many of the more popular desktop applications, such as LOTUS or EXCEL, provide add-on tools that allow the application to access data in remote databases. Other tools, such as DATAPRISM and DECQUERY, provide data access to the remote databases and the ability to export query results for use by other data manipulation packages. Thus, a customer can create a query in a familiar desktop environment, ship the query across the network to the UIS server, retrieve the results and manipulate the query results at their desktop. At this point, eight university departments were strategically chosen to become beta users in the use of this prototype.

As it exists today, the data warehouse provides the beta users with timely, accurate data. To achieve this, the appropriate data is extracted from the legacy system databases and then transferred to the UIS machine and loaded into the Rdb database. This process occurs in batch mode with different data being downloaded on varying schedules. For example, SIS student demographic data is propagated weekly, while Financial system transactions are loaded daily, and payroll data is updated after each bi-weekly payroll. In addition to this operational data, metadata is also available. The metadata provides a definition for each data element, a list of the valid values for an element, and the datetime stamp of the last data load for each table.

The UIS data warehouse represents a break from the operational application systems that CCIT has deployed in the past. Traditionally, CCIT develops application systems which perform administrative functions such as registering students, paying employees, and posting financial transaction. UIS, in contrast, does not perform any administrative processes, it provides read-only access to data from the three major administrative systems in a single data warehouse.

With the prototype developed and the client/server tools adequately tested, it became evident that a method for introducing this new technology to the campus community was
needed. CCIT knew that acceptance by the campus community would be imperative for a successful implementation of any new technology. Campus departments would have to be sold on the new technology, new ideas, and new ways of doing business. Many people felt that a live demonstration of UIS would put the project in the limelight and generate the most interest. Therefore, a team was established that would be responsible for demonstrating the new technology and the prototype. This team consisted of three application analysts, one from each of the application support groups within Decision Support, along with two members of the original implementation team. All members of "The UIS Demo Team" were very familiar with the user community and understood their concerns and needs.

The demo team began developing the format of the presentation. The team decided that the demonstration to the customers had to be state-of-the-art and could not consist merely of simple overheads. They wanted to use software on a workstation to show the use of technology. Presentation software was evaluated and selected and a software slide show was developed. This slide show consisted of 15 slides which explained in a logical sequence how business was currently done at the University, and why a new way of accessing and sharing information was needed. This was followed by an introduction to the client/server environment, an overview of the UIS prototype, and the benefits derived from implementation of this new technology. Finally, an introduction to the terminology associated with relational databases and client/server technology and a live demonstration accessing the UIS database was provided. The presentation was designed so that each team member participated in the demonstration by discussing a different slide. During the live demonstration, teamwork was again used with one person speaking and another running the software.

The team presentation method proved to be successful for a variety of reasons. First, it made each member more effective by allowing him/her to focus on particular portions of the presentation. Second, shifting the audience attention from one person to another made the presentation more interesting. Third, one or more persons on the team is familiar and respected by the audience. Having familiar and trusted applications analysts has helped build confidence. Finally, the cross-sectional team within Decision Support helps to communicate a joint effort by all three administrative systems groups.

Many people have been involved in the evolution of the presentation. Initially, the team presented to the managers of Decision Support. After this presentation, a critique session was held. This feedback was incorporated, the slide show was modified, and the live presentation was altered. The next group was a peer group presentation within Decision Support. This played a dual role by introducing the 40+ staff members to client/server concepts and allowing them to give feedback to help the team further refine the presentation. Following these presentations, each department in CCIT was given a chance to see the presentation and offer suggestions for improvement. After the CCIT presentations were completed, the team began giving demos outside of CCIT.

The demo has been presented over 30 times with audiences ranging from Vice Presidents and the Provost to Departmental Systems Analysts. Because the audiences are so diverse, the live portion of the demonstration is customized. For example, if the audience was a group of high level administrators, the live demonstration was tailored to answer general University questions such as "What is the undergraduate enrollment
by residency and class level?. If the audience was a group of business managers
the demonstration might contain a budget report for a particular department.

After each presentation, a question and answer period followed. Often the same
issues surfaced. Many of the concerns were anticipated, but required policy
changes for the University. For example:

What server should we move to?
Who will pay for the server?
How will we use other servers?
Who pays for Ethernet Connections?
What client software should be purchased?
Who pays for the client software?
What desktop hardware should be purchased?
Who pays for the desktop hardware?
Who will install the Software and Hardware?
Who will provide training on the Software and Hardware?
Who will troubleshoot problems?
Who will provide documentation?
Who will develop queries?
How will the queries be made public?
Who will provide consulting for query development?
Who needs the information?
Should there be data access/security policies?
Who administers the data access/security policies?
Who determines the data to be housed?
Who interprets the data?

Today these issues and concerns are being addressed by several committees and
top administrators. In addition, CCIT has formed a new cross-departmental team
whose
mission is to identify the next steps in fully implementing UIS at the University of Arizona. Inputs and recommendations from these groups will determine the future of client/server computing on our campus and specifically the future of our prototype project. In the meantime, eight beta users continue to use the new prototype to answer their daily business questions. The prototype has been successful in proving the technology, however, it has opened a pandora's box regarding management issues.