Partnering with a Vendor to Prototype a Data Warehouse in Ninety Days

by Robert G. May and Teresa Stankiewicz

Many organizations have strong motivation to weigh the benefits of a data warehouse against the investments required to build it. The University of Texas College of Business Administration/Graduate School of Business partnered with a vendor to develop a data warehouse within a ninety-day window to overcome the investment barrier by testing the value of data warehousing with a real program.

Unless specific measures have been taken to enable additional functionality, most campus legacy mainframe systems are being used for operational purposes. They are transaction-based and commonly rely on hierarchical databases. These systems are still functioning effectively for their original purposes. But times change, as do our work requirements and expectations of the computing tools we use. If, for example, you want to prepare a report on how data have changed over time and must query a production system with a hierarchical database for this information, you are likely to find this new task a challenge. Besides the fact that such a query usually must be performed by and accommodated into the workload of the central data processing staff, it is also the case that many legacy systems keep only current data, which makes it difficult to perform ad hoc queries for purposes of analysis and comparison.

To change this picture, many campuses are turning to a data warehouse solution. It is one new technology that the computer services unit within the College of Business Administration/Graduate School of Business (CBA/GSB) at the University of Texas explored last year. Computer Services envisioned that the data warehouse concept would enhance administrators’ effectiveness and ability to manage their areas of responsibility proactively, with the goal of working within their familiar microcomputing environment for such ad hoc queries.

A tangible need to fill, yet a leap of faith to develop

Our Educational Resources Data Warehouse (ERDW) project began last April, when the vendor (Software AG) who had created the database and extract languages used with the University’s mainframe approached us with the idea of accomplishing a working data warehouse. The proposal was to install a client/server system and model, on a limited scale, with genuinely useful capabilities, as the precursor of a new, comprehensive system—all in a ninety-day burst of work that would limit our investments of time and money and enable us to readily evaluate the potential of the solution. Quality metrics collected during the project were used to validate the usefulness and effectiveness of the data warehouse.

Our prototype ERDW encompassed the CBA/GSB need to support future course and instructor scheduling activities. From meetings facilitated by our vendor and interviews with executives and administrators, we designed the prototype with an efficient interface to the warehouse and the software tools administrators would need as they schedule “resources” (courses, classes, and instructors) to resolve scheduling conflicts and obtain historical information for trend analysis on class enrollment and faculty evaluations.

Even a test project of this nature required a high level of interdepartmental collaboration and cooperation, starting with evaluating the data being captured and maintained. The University’s data processing department keeps the course, instructor, and student data on the mainframe. The DP department served as consultants to the project and provided the staff member to administer the new database. DP also assumed responsibility for supporting the new client/server operating system and hardware underpinning the ERDW. The prototype was work-
ing successfully just three months after the start of the project.

Transaction data, which continue to be stored on the University mainframe in ADABAS C, are periodically transformed into business information and loaded into the ERDW—an ADABAS D relational database running on a separate application server (Hewlett-Packard Model 817 UNIX). This client/server environment, which will contain three to five years worth of historical data, is something that CBA/GSB administrators can freely access from client microcomputers in their offices. They use the query tool Esperant to readily access and extract information and import it directly into Excel, Lotus 1-2-3, QuattroPro, or Esperant as they construct their graphs, reports, slides, or other documents.

This is the improvement we had been looking for.

**En route solutions**

As with any proposal for data warehousing, the time required for downloading data from the mainframe to the server was an early concern. In Phase 1, the prototype contains over 375,000 records, using 200 MB disk storage. However, the issue of data transfer performance, which can potentially prevent “refreshing” the data warehouse in a timely fashion, proved not to be a concern in Phase 1 and subsequent phases, due to the particular design for the data warehouse. Our first experiences at downloading data from the mainframe (daily during student registration and at several other scheduled times) proved the ERDW concept was indeed feasible. Server performance was such that we quickly dispelled concerns over data transfer feasibility.

We were also very interested in ensuring fast query response times for users. When we saw response times from the ERDW were matters of seconds rather than days—due to a combination of hardware performance and design of the application itself—we were confident enough to move ahead with subsequent phases and expand the ERDW project by adding still more functionalities and quickly making it available to other colleges.

**Benefits earned, lessons learned**

Computer Services analysts and developers worked from a very detailed, methodical approach supplied by our vendor’s consultants. The outside perspective proved invaluable in applying numerous best practices for successful transitions such as basic project preparations, business drivers and needs identification, model definition and refinement, prototype development and incremental refinement, hardware/software infrastructure construction, procedures for operations, and user training and ERDW support planning. Key to these results was the teamwork between Computer Services, Data Processing, and vendor personnel; we proved to skeptics that it was possible to meet our goal of building a working data warehouse in ninety days.

One of the very first real uses of the ERDW is to support faculty review and analysis of student evaluations of teaching performance in the CBA. Preparation for evaluation meetings using the ERDW is far simpler now than in the past, and can be accomplished in a handful of steps. These steps can then be saved as a macro and icon, so the entire process can be repeated in the future in just one step by clicking on the icon that represents that query.

Those reviewing these reports will have as-

---

“The outside perspective proved invaluable in applying numerous best practices for successful transitions...”
surance that the presentation accurately reflects trends in a real-time fashion. For example, previously run reports—such as those that reveal student satisfaction with CBA/GSB and student ratings of teaching performance—can be compared to current results to track the school’s trajectory toward its goal of consistently high satisfaction. This kind of analysis and reporting is important, because faculty rewards and tenure are tied to merit and demonstrated performance. Furthermore, news that can be readily understood, in pie-chart form for example, is ideally suitable for public dissemination via student and alumni newsletters, the local news media, and so forth.

Another representative use of the ERDW is helping to take the once-a-semester frustration out of course planning. As departments construct schedules of courses, the dean’s office must evaluate the entire schedule. A review by the dean’s office assures students that they will not find required classes offered by the different departments scheduled at conflicting times. Previously, these conflicts were identified and resolved by a tedious manual process that involved office staff posting notes for each class to the office wall and checking for same-day, same-time conflicts. This process is now completed quickly and accurately using the new system.

Early results like these demonstrate the value of continuing our investment in the data warehouse concept and additional phases of implementation. As one of the next steps, we will again collaborate with administrators, who have now become familiar with the potential of the ERDW, in order to develop pre-built objects that will handle as much as 75 percent of most administrators’ querying needs and greatly increase the convenience of the system. The balance of ad hoc queries will be done by users themselves using the query tools. In Phase 2, we will add to the library of queries and bring in all relevant student data, followed by the addition of financial data in Phase 3.

Conclusion

One of the key lessons learned is that the system should be designed in such a way that it can be expanded or enhanced later in increments. In addition, the vendor’s commitment at the very beginning of the process is an important key to a successful, ninety-day launching. We were gratified that our software vendor shared our commitment to spending the time necessary to clearly research and understand the University’s needs before beginning to devise new hardware/software recommendations. Following a standard but flexible approach thereafter enabled us to consistently meet project milestones and the ninety-day completion target. The collaborative approach and focus on the original goals, phase by phase, helped us avoid distractions that can easily sabotage projects of this nature.

The authors would like to thank Sara Gill, Lead Systems Analyst for the CBA/GSB Computer Services, who was team leader for the ERDW project.

Electronic Messenger...

(continued from page 43)

However, increased levels of protection of information is analogous to protection of your home. It’s a balance, an attempt to find an appropriate comfort level. Using bars on your home windows (like using high level/complex encryption algorithms on data transmission) provides security, but limits the number and types of communication exchanges. Institutions will make determinations as to their comfort level with new technology just as they currently do with phone, e-mail, FAX, etc. The greatest threats to privacy/confidentiality are similar to the greatest threats to security of any correspondence or information. That is, the greatest threat is not in the transmission or interception of the information/correspondence, but rather with the particular level of integrity and commitment of staff in the originating and receiving offices.

Unfortunately, some people are demanding that new and emerging technology interpret and establish new and more expansive policies and procedures surrounding privacy and confidentiality. We seem to be ready to hold technology hostage while we try to come to consensus on matters of privacy and confidentiality. The application of technology within privacy and confidentiality context is somewhat analogous to writing instructions for the computer: It will do your bidding; you only need be clear as to what exactly it is that you think you’re asking for.

Educational institutions, quite often in the locus of the Registrar’s Office, have routinely made decisions and evaluations regarding the collection, maintenance, and release of information within the context of privacy and confidentiality. New and emerging technology, particularly EDI, does not change the way we’ve been “doing business” for the past twenty years. As long as the stewards of protected information continue to act in their heretofore responsible manner, new and emerging technology should be considered an opportunity and not a threat.