

Data Warehousing: piloting a minimalist approach

SHEILA MUKERJEE, *La Trobe University, Australia*

Abstract

Value-added corporate and client information has fast become a highly sought-after commodity in sectors that strive to succeed in increasingly competitive markets. Companies and institutions around the globe are finding solutions for their corporate strategies through the extraction of useful knowledge from their data. Data warehousing provides the underlying infrastructure for further application of business tools and techniques such as data mining and knowledge discovery to complete the picture of value-added information.

This paper describes how an Australian university embarked on a journey into data warehousing and is now poised at the threshold of exciting possibilities in the world of business intelligence. A minimalist approach coupled with a quick implementation was found to be successful in convincing both management and clients of the tangible benefits of the project. Future plans and recommendations, to extend the basic model to utilise more sophisticated features of the tool, offer enhanced long-term benefits too irresistible to ignore. In view of such enhanced capability and proven benefits, on-going support for development is virtually guaranteed.

Introduction

Information is now regarded as the currency of the new economy (Liataud, 2001) and this realisation has led to the growing stampede for tools and techniques to transform raw data into more useable and intelligent forms to provide maximum benefit. The concept of data warehousing was introduced to cater for this demand. This process transforms raw data into a format that can be readily converted to information through the application of on-line multi-dimensional analysis. By further subjecting the information to data mining and interpretation processes, knowledge is finally obtained. (Fayyad et al, 1996a). This entire process of transformation of data into knowledge has become a highly researched field of study known as knowledge discovery in databases (KDD) (Fayyad et al, 1996b). Both researchers and business experts recognise that transforming raw data into useful knowledge offers tremendous value and potential. The business world is fast embracing the concept of adding intelligence to information to provide new insights into the business (Liataud, 2001). Regardless of ownership of data, one thing is clear. The value does not lie in the data itself or in the volume of stored data but in leveraging the data for critical business decisions and strategies. How relevant is this concept to today's world of academia – a world that finds itself caught in a dilemma described by Marginson (2000, p. 29) as one between traditional academic values and modern corporate cultures?

In the last decade or so, universities everywhere have experienced great changes brought about by a multitude of factors such as funding cuts, organisational change, technological breakthroughs, globalisation, increased competition and technology-driven movement away from the traditional methods. In order to survive and succeed, universities must equip themselves to ride these paradigm shifts. But how does a bureaucracy bogged down with sluggish traditional practices make this leap?

Peter Coaldrake (2000, p. 20-21) suggests that, in order for Australian universities to operate successfully in this rapidly changing environment, universities need to adapt to change as well as reassess old traditions and practices. In today's climate, this equates to quick decision-making processes achieved through effective communication strategies supported by superior decision support systems.

Recognising the critical role of effective and timely communication in the life of an organisation, La Trobe University embarked on a project to implement a decision support system in 1999. In many ways, this project was a reflection of the changing philosophy of the University to break away from the traditional methods of management and information governance. Not surprisingly, this move seems entirely consistent with the models of information governance that have evolved as a result of the change from the centralised style of management to decentralisation. In the 1980s, information dictatorship existed as part of the centralised management model where access to data concentrated in the hands of a few. Decentralisation in the 1990s saw the unleashing of information empowerment. This led to information rebellion and the beginning of an era of information anarchy resulting in the proliferation of numerous uncoordinated and unintegrated information systems, causing data conflict and chaos. In recent years, with the onset of a new management philosophy of empowerment and decentralisation of decision-making, it has become vital for all levels of an organisation to access, analyse and share information. This model of Information Democracy ensures that all employees are empowered with information they need to successfully fulfill their roles (Liataud, 2001).

The project

1999 saw the dawn of the era of Information Democracy at La Trobe University. Following a feasibility study, the University decided to adopt an institution-wide standard to data extraction, reporting and analysis for effective communication of information. It was envisaged that the decision support system would also provide 'virtual' integration of its administrative systems and other data sources. Formal acceptance testing procedures found that the Business Objects tool was able to fulfill the University's requirements to support an integrated decision support solution. Once the decision for the tool was made, the next step was to decide on the approach for implementation. Should the University devote time and resources to building a full data warehouse in one huge implementation or should it take a minimalist approach in a phased implementation?

Two very important factors influenced the decision that was made. The first was the availability of resources at a time when other resource-intensive projects such as the implementation of an Enterprise Resource Planning (ERP) system were being undertaken. The second was pressing demands from users who were disillusioned with their short-lived joys of information anarchy and were eager to embrace information democracy. Under such conflicting demands, the University made the obvious choice – provide a quick interim solution that would immediately empower users to meet their urgent reporting needs. Once the approach was established, the first phase of the implementation began in earnest. The most frequently requested student data over the past two years were analysed and translated into user requirements. This constituted approximately 90% of the known reporting needs for student data. Phase one of the implementation was intended to deliver access to the most requested and urgently needed data. Envisaged as a launching pad for the

eventual construction of a full data warehouse, this implementation also had to address several foundational issues. These included the type of deployment (full client or thin client), hardware requirements, software installation to user sites, selection of a pilot user group, change management, training, and the establishment of a structure for security and access levels that would accommodate a rapidly expanding and diverse user population.

In keeping with the growing philosophy of empowering users without delay, the decision was made to proceed with the client-server application in the first instance. Options for a thin-client deployment would be explored in the next phase when more information and technical expertise were available. Since the client-server application is platform dependent, adapting non-PC users to a new working environment became a challenging change management issue.

Recruiting the right user group for the pilot implementation was crucial to the success and acceptance of the project. To maintain equity across the University and accommodate the varying needs of users, Managers of the relevant work areas were invited to submit nominations for the pilot group. Staff members were carefully selected on the basis of their reporting needs and knowledge of the student data.

The importance of training was recognised right from the outset of the project and priority was given to developing a comprehensive and well-defined program. Since the full data warehouse solution was not implemented, the data had not undergone the necessary transformation and integration into a readily useable form. Hence, the need to understand complex underlying data structures was not totally eliminated. It was necessary for users to understand some of the data relationships in order to design the right queries. A two-day program was designed to provide separate training sessions on the tool and student data. The acceptance of the pilot user group was so enthusiastic that interest for this new support system immediately took on greater proportions. Additional training had to be conducted in rapid succession to cater for the growing demand. Interest and enthusiasm were so high that it proved to be a challenge to keep up with the demands for training with existing resources.

Outcomes and achievements

The quick implementation, made possible by the minimalist approach of delivering to restricted specifications, had a number of benefits. Undoubtedly, the success of the project was due mainly to a pre-existing lack of timely access to vital data and the absence of a user-friendly, versatile tool capable of extracting, formatting and analysing data to meet needs as diverse as the users. In this respect, the pilot implementation met the expectations of the users. The interim solution provided access to the most frequently requested data and was immediately accepted as a quick, reliable and efficient reporting tool, capable of supporting decision-making processes in teaching, administration and management. This ready acceptance, however, rapidly turned into demands for further enhancements. In order to keep pace with growing user expectations and needs, and, more importantly, share ownership of further development with users, a User Group was quickly formed. This group was responsible for identifying and prioritising developments.

This delivery was also a cost-effective means of meeting user demand at a time when resources were limited and stretched to capacity across several concurrent projects.

The User group provided additional valuable resource to assist in further developments. Some of the skilled users were recruited as testers to facilitate timely delivery of further enhancements. Because of their positive experience, users were enthusiastic to assist in whatever way possible to maintain the momentum of the project.

One of the most valuable outcomes was that both management and users were able to see tangible results and benefits immediately. A long drawn-out process bent on delivering an all-inclusive warehouse may have seen enthusiasm dwindle and interest in the project overshadowed by possibly late delivery and inability to meet user expectations. The staged implementation provided a natural vehicle for the management of user expectations. The user-driven focus ensured that delivery was always in step with user expectations.

The iterative method of development proved to be a consultative and inclusive process that effectively shared ownership and participation with the users. This user-driven approach assured both users and the project team that further developments or enhancements were, in fact, meeting the needs of users and ultimately, the institution.

Gaining independence to design and run reports was a liberating experience for both users and the systems staff. Users were able to produce reports in formats of their choice on their own desktop in a matter of seconds. The reports could be printed on their own printers rather than a centrally located networked printer. This led to enhanced operational efficiencies and enabled administrative staff to streamline labour-intensive processes. Systems staff were freed up to undertake more systems development work and, generally, pursue a more pro-active approach to systems development and support. Needless to say, staff morale was given a welcome boost in the process.

Another positive outcome that ranked highly with staff was the staff development opportunities presented by this project. Both technical and general staff were able to upgrade their skills in 'state-of-the-art' technology and broaden the scope of their daily endeavours. Administrative and management staff are able to exploit knowledge gained from the information to explore opportunities such as innovative marketing strategies, new ventures based on informed decisions and modelling of information to assist in planning, strategy development and decision-making processes.

An unexpected outcome was the identification of anomalies and inconsistencies in the format of the stored data. More often than not, assumptions are made on the format that are not necessarily right. It is in retrieving or manipulating data that these assumptions are tested and proven. Several cases of inconsistencies led to the identification of loopholes in the Student Information System application software that were quickly fixed. The systems area also scored a further bonus. Having a quick and efficient tool to retrieve data provided time-saving shortcuts in the process of identifying and setting up of test cases for systems testing work.

From a technical perspective, the pilot project provided a framework for the deployment of future client-server applications across multiple campuses. Issues such as hardware requirements, servers, secure access to servers, authentication, software installation, connection to different databases and technical expertise were discussed

and explored. These efforts will be invaluable for future enhanced deployments. The project also established a methodology upon which future developments, ultimately leading to a data warehouse incorporating the University's suite of corporate databases, could be built. Currently, data from two separate corporate databases (RDB and Oracle) are accessible through Business Objects.

In spite of the benefits outlined above, this approach also presented some limitations and challenges. The full potential of the tool could not be exploited when the aim was to deliver quickly to minimum specifications. Utilising the reporting functionality of the tool without the full data warehouse and broadcast capabilities limited the scope of the project and relegated the achievements to a one-dimensional paradigm shift in information reporting. A full data warehouse implementation would have provided the multi-dimensional thrust into the world of information management.

Keeping pace with user expectations and evolving needs was a huge challenge. Moissis (1997, p. 1) reported that a survey of the Data Warehouse Institute found that "most organisations underestimate the speed at which extra users request the new decision support system tool ... the number of users actually double every three months." This was indeed the case at La Trobe University with users doubling as fast as training programs could be offered. Following the pilot implementation, the project team was inundated with requests for training and enhancements. Although this overwhelming acceptance and enthusiasm was exhilarating, the demands placed tremendous pressure on the project team. The project team comprised of staff members who were involved in other concurrent projects and hence were not dedicated to this project. Subsequent developments, although planned, had to be undertaken in a series of stops and starts due to the pressures of other priorities and commitments on the shared resources.

Future Plans and recommendations

Having taken the significant step of implementing an interim solution for decision support, La Trobe University is now strategically positioned to proceed to the next major delivery of a fully integrated data warehouse. Data integration of its major corporate systems, namely the Student Information System, SAP Finance and Human Resources, and Research Master, would offer an institution-wide standard for decision support. This prospect of a standard across the institution becomes even more appealing when one considers the implications of already acquired transferable skills in a climate of change with new systems being introduced. User expectation is also another significant factor in maintaining this concept of a standard. Any implementation of a new corporate system with reporting requirements should include parallel development of the information management strategy within the established data warehouse structure. To do otherwise would confuse the concept of an institution-wide standard and may adversely affect user acceptance of the new system being introduced.

Given that the resource constraints that influenced the approach adopted for the initial delivery remains unchanged, it is likely that on-going development will continue to be undertaken in a series of staged implementations. Since this process is likely to be resource-intensive, it is recommended that the University implement options designed to optimise the efficiency and use of technical staff time. Options include moving to the cost-effective thin client application and addressing issues relating to the

distribution, sharing and publishing of documents. These options, once automated or implemented, will eliminate the inefficient use of technical staff time taken up in redundant support roles and allow them to engage in productive development work. Users will also gain a welcome measure of independence through the automation of processes that, in the past, required technical staff intervention. The thin client solution offers a web deployment that, unlike the client-server application, does not require the application to be installed in the user's PC. Only a web browser is needed to access the data. This effectively removes the platform dependency and allows more users to access data in a cost-effective, "zero client administration" environment (Moissis, 1997). The Document Agent server and Broadcast agent in Business Objects provide capabilities for scheduling, refreshing, distributing and publishing reports.

The future, however, cannot and should not be seen solely in terms of technology and the never-ending quest for enhanced technological capabilities. Implementing technology is by no means the complete solution to the challenge of adapting to the new economy. The world seems to be experiencing change at such a rapid pace that in order to survive, one must be ready and equipped to move with the times. No longer is storing voluminous quantities of raw data without extracting value a viable proposition. Equally inadequate and outdated in a fast-changing environment is the use of data processed into information through the application of static rules and technologies. To be effective, information needs to be turned into a dynamic force that recognises change immediately and adapts accordingly. This refinement of information into knowledge is described as the emerging force with capabilities to provide the much sought-after competitive edge. Knowledge management is receiving growing acknowledgment as the key to surviving in an environment where rules are constantly changing. Knowledge management seeks to combine the factual representation of information with human creativity, experience and innovation into a powerful and dynamic resource for the gathering, interpreting and utilisation of information in order to anticipate and initiate change in response to changes in the environment (Radding, 1998). In this equation, the human element brings with it a dimension that cannot be replicated or quantified completely by rules or programmed applications. No matter what technology delivers, the human resource remains a priceless asset that must be recognised and cherished as the heart and soul of an organisation.

References

COALDRAKE, P. (2000) Reflections on the Repositioning of the Government's Approach to Higher Education, or I'm Dreaming of a White Paper, **Journal of Higher Education Policy and Management**, 22(1), pp 20 –21.

FAYYAD, U., AND UTHURUSAMY, R. (1996a) Data Mining and Knowledge Discovery in Databases, **Communications of the ACM**, vol 39,no.11, pp 24-26.

FAYYAD, U.,PIATETSKY-SHAPRIO,G. AND SMYTH, P. (1996b) From Data Mining to Knowledge Discovery in Databases, **AI Magazine**, vol. 17, no.3, pp 37 – 54.

LIAUTAUD, B. (with Hammond, M.) (2001) **e-Business Intelligence: turning Information into Knowledge into Profit**. McGraw-Hill.

MARGINSON, S. (2000) Rethinking Academic Work in the Global Era, **Journal of Higher Education Policy and Management**, 22(1), pp 29.

MOISSIS, A. (1997) From Pilot Project to Corporate Standard, **Technical White Paper**, Business Objects Collateral Production.

RADDING, A. (1998) **Knowledge Management: Succeeding in the Information-based Global Economy**. Computer Technology Research Corp.