Institutional strategy and information support: the role of data warehousing in higher education

Jeff Guan
William Nunez and
John F. Welsh

The authors

Jeff Guan is Associate Professor of Computer Science, College of Business and Public Administration, **William Nunez** is Assistant Director of Planning and Research, Office of Planning and Budget, and **John F. Welsh** is Associate Professor/Acting Dean, College of Education and Human Development, all at the University of Louisville, Louisville, Kentucky, USA.

Keywords

Data, Warehousing, Higher education

Abstract

This paper examines the challenges that colleges and universities confront in the management of information necessary for strategic planning and decision making and explores data warehousing as an approach to knowledge management in academia.

Electronic access

The research register for this journal is available at
http://www.emeraldinsight.com/researchregisters

The current issue and full text archive of this journal is available at
http://www.emeraldinsight.com/1065-0741.htm

Campus-Wide Information Systems
Volume 19 · Number 5 · 2002 · pp. 168–174

· ISSN 1065-0741
DOI 10.1108/10650740210452274

A context for change

Colleges and universities today face three common challenges that are prompting a critical reexamination of the architecture, management and use of their information systems: first, administrators and policy makers at all levels demand improved data management strategies to support resource management and strategic planning. Second, due to a changed competitive environment for higher education as a whole, faculty and administrators are hungry for information that can assist institutions in the recruitment and retention of students. Third, external oversight agencies, such as the federal government, state governments and accrediting agencies, demand information about the performance of institutions and programs on a wide array of indicators. At issue is the extent to which existing information systems and knowledge management strategies optimize the ability of institutions to respond to a new environment.

Advances in technology have dramatically enhanced the level of productivity and efficiency of universities and colleges. They have forever altered the way universities operate. However, despite the availability of powerful computers, advanced network and communication infrastructures, and sophisticated software applications, university decision makers, including planning and budget administrators, still lack access to the critical information necessary for informed decision making. Academic deans and provosts often lament lack of access to valid and reliable information about their finances, staffing and students. Ironically, the information they require frequently exists in the organization. However, only a fraction of the data that are captured, processed and stored in a college or university's information system is actually available to decision makers in an organized manner.

This article examines the root causes of information problems from the perspective of decision makers. It assesses data warehousing as a solution to decision support and knowledge management in higher education. The paper first looks at the need for data-driven decision making in higher education and then explains why typical information technology

architectures found in higher education fail to meet the needs for decision making. The paper next discusses how data warehousing functions as knowledge management through the organization mountains of data residing in campus information systems.

Institutional issues

Ideally, organizational decision making begins with an in-depth analysis of available management information. Decision makers then use information to weigh alternatives, analyze options, anticipate implications and project outcomes within the context of their organization and environment. If, however, decision makers lack relevant, reliable and timely information, the quality and credibility of their decisions are suspect.

Higher education has always been a dynamic environment requiring strong leadership and solid decision making, but the current pace of change in higher education forces leaders to seek rapid access to information to expedite decision making. They are frequently required to consider many variables simultaneously as well as the impact of their decisions on internal and external constituencies. These variables and constituencies fall into four broad categories to consider:

(1) complexities of the organization;
(2) vast number of stakeholders and appraisers;
(3) competitive marketplace; and
(4) resource limitations.

Complexities of the organization

Higher education institutions vary widely in their organizational structure and administrative and governance systems. Generally, however, colleges and universities share governance with a plethora of constituents. Birnbaum (1988) classifies these structures as "loosely coupled", "tightly coupled", or some variant between. Birnbaum (1988, p. 43) suggests that governance and management procedures should exist that address the specific "economic, social value, political, informational and physical characteristics" of the institution and its environment. The decision-making subsystems within institutions require vast amounts of quality information to address the requirements of the environment. Frequently, demands for information about the programs and services of an institution go unanswered because information management systems cannot respond.

Stakeholders

Depending on the type and size of the institution, stakeholders evaluate the quality of the institution, rate the programs, or impact institutional policy. These stakeholders include agencies from state coordinating boards, to accrediting bodies, to institutional rankings by *US News and World Report*. These agencies all require detailed information regarding admissions, student demographics, program effectiveness, student outcomes, and many other measures. Additionally, decision makers must consider the impact of decisions on internal stakeholders at their institutions. While the number of the stakeholders is inevitably linked to the size and type of institution, they generally include students, parents, faculty, staff, board members, legislators and community leaders. Information systems must provide institutional decision makers with access to quality data for, and about, each of these constituent groups.

Competitive marketplace

College students have many choices available to them today. The new competition in the educational marketplace has forced institutions to be more focused and competitive as they attempt to recruit these potential students. In an effort to capture the "best and brightest" of these students, proactive institutions have implemented sophisticated marketing campaigns and focused recruitment plans to define their niche in the educational marketplace. Building a successful marketing plan or recruitment campaign requires detailed understanding of the pool of qualified applicants. Questions about the potential pool of students are answered by data available in most institutional databases that, once transformed into management information, become a powerful resource decision makers can use to combat market pressures and increased competition. If this information is not readily available to decision makers, the

Campus-Wide Information Systems
Volume 19 · Number 5 · 2002 · 168–174

institution will struggle in defining its customer base, service area, core competencies, and other aspects necessary for an integrated marketing plan.

Resource limitations

As operating costs increase, strain on the budget follows as institutional leaders struggle to balance the academic and human resources needs of colleges and universities. As state support stays at levels of inflation and increased emphasis is placed on student tuition and fees and external funding, decision makers are forced to determine which programs are critical, in high demand, and/or operating at a lower cost per student. Institutions also concerned about program efficiencies and resource management. An analysis of how to optimize resource management can only be achieved through an integration of data from both financial and student systems to evaluate the costs of operating versus the benefit received from those investments.

The characteristics of information needs of today's institutional decision makers can be summarized as follows:

- the information required often has to be derived from a vast amount of data;
- the information required is often integrated from diverse internal and external data sources;
- historical data are often necessary to understand trends;
- rapid response is necessary for decision makers; and
- information requests are often *ad hoc* in nature.

The structure and scope of data warehouses in higher education

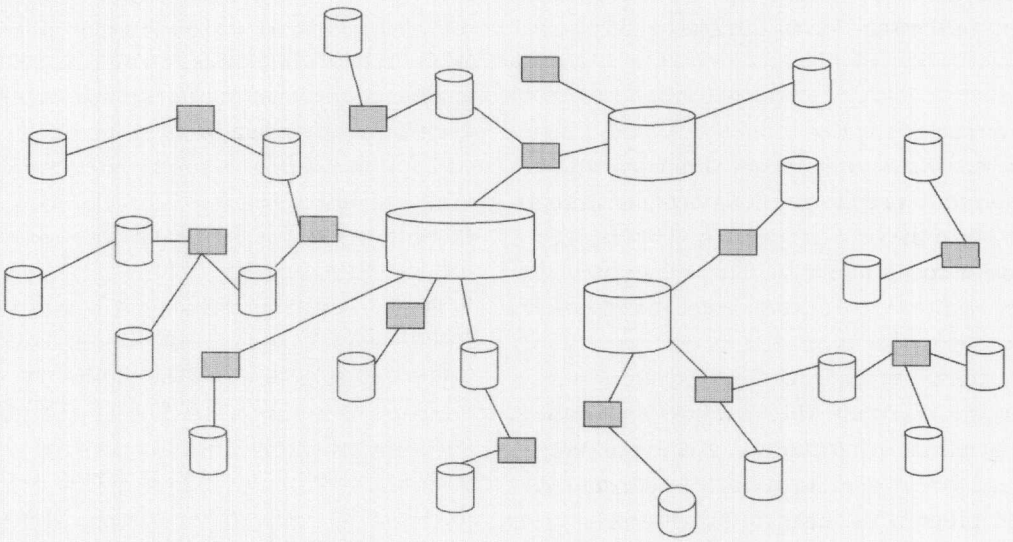
The existing information technology (IT) infrastructures at many organizations are inadequate to meet the information needs of institutional decision makers. There are two main causes of this lack of adequate decision support. The first has to do with the nature of many of the existing IT systems. The second cause lies in the often poorly integrated IT architecture.

Colleges and universities have invested heavily in IT over the past few decades, including more recent investments in Web-based technologies. Most of these IT systems have one thing in common: they have been acquired to support the day-to-day operations. Such systems are often referred to as online transaction processing (OLTP) systems. An OLTP environment is not suitable for decision support as they have been designed to support short transactions affecting a few records at a time. This type of data tends to reflect only the current state of the system and seldom keeps historical snapshots, which are critical for planning purposes.

Owing to the types of transactions an OLTP system supports, different kinds of data tend to be scattered in different tables. For example, in the PeopleSoft software a simple class list report with class data, meeting times, building, instructor, student data may require the integration of data from over 30 different tables. Extracting and integrating information from all these tables are not only time-consuming, they require very complicated programming. Additionally, the pertinent data reside in different systems or databases. Finally, the day-to-day operations performed by an OLTP system often are already constraining the hardware and software resources on the system. Adding a decision support function on top of an OLTP system severely impacts its performance. Decision support queries involve thousands and often millions of records. They are extremely resource-intensive.

A less obvious reason why many IT infrastructures are inadequate for decision support has to do with the naturally evolved information technology (IT) architecture found in many organizations today (Inmon, 1996; Linthicum, 2000). IT investments often come in the form of one application at a time, with little integration or interoperability in mind. New applications and new technologies are often introduced into the existing IT architecture in a hurry to meet an urgent business need or in an effort to catch up with competitors. This kind of natural evolution of an IT architecture eventually results in a set of extremely complex and incompatible systems, leading to enterprise-wide data and information chaos (see Figure 1).

Figure 1 Information chaos



From a decision support perspective, this kind of poorly integrated architecture leads to several problems that affect any data-driven decision making. The most common problems are data extraction costs and lack of data consistency. Customized extraction programs frequently have to be written to extract, cleanse, reconcile and integrate the data. There is often more than one path to obtaining the same type of report. Hence, different offices produce different results to the same inquiry.

Data warehousing

A separate environment is needed for decision support. Although decision support systems have been around for many years, it was not until the early 1990s that they were labeled "data warehouses". The most commonly used definition of data warehouse is:

> . . . an integrated, subject-oriented, time-variant, non-volatile database that provides support for decision making (Inmon, 1996).

The meaning of each of these terms for decision support is pertinent to this discussion:

- *Integrated.* The data warehouse is a centralized, consolidated database that integrates data derived from the entire organization. Thus the data warehouse consolidates data from multiple and diverse formats. Data integration implies a well-organized effort to define and standardize all data elements. This integration effort can be time-consuming but, once accomplished, it provides a unified view of the overall organization.
- *Subject-oriented.* Data warehouse data are arranged and optimized to provide answers to questions coming from diverse functional areas within an organization. Therefore, the data warehouse contains data organized and summarized by topic, such as student demographics and human resources. This type of data organization is very different from that found in a typical OLTP system.
- *Time-variant.* In contrast with OLTP systems that focus on current transactions, the data warehouse data represent the flow of data through time. It is also time-variant in the sense that, once the data are periodically loaded to the data warehouse, all time-dependent aggregations are recomputed.
- *Non-volatile.* Once data enter the data warehouse, they are never removed. Because the data in the data warehouse represent the company's entire history, the operational data, representing the near-term history, are always added to it. Because data are never deleted and new data are always added, the data warehouse is always expanding.

A data warehouse provides a solution to meeting the information needs of institutional decision makers. Figure 2 presents the architecture of a typical data warehouse system.

The data source component represents all of the sources from which the raw data originate. To identify the data sources, candidate systems and files are scrutinized and evaluated in terms of integrity and quality of data. The data acquisition component captures data from the source systems and directs them into the data warehouse. This is often the most complicated and critical component of data warehousing. Rules must be defined that guide the cleansing, enhancement, restructuring, integration and aggregation of source data. Data cleansing may involve the restructuring of records or fields, removal of operational-only data, supply of missing field values, and checking of data integrity and consistency. Data enhancement may involve the decoding and translating of field values, addition of a time attribute to reflect the currency of data, summarization of data or calculation of derived values. The definitions created for the data during the data acquisition process will be stored as metadata, which will be used later in the data warehouse system for generating reports and analytics.

The data warehouse component is often one or several databases that contain the data warehouse data. The data warehouse contains a large data repository, called the enterprise data warehouse, and several smaller data warehouses, called data marts, as indicated in Figure 3. The enterprise data warehouse usually contains enterprise-wide data that provide an enterprise view of the organization. The data marts are often data repositories that focus on data collections and analysis of interest to specific user communities, such as budget and planning, human resource, students, etc. In addition to holding cleansed, integrated, historical and aggregated data, the data warehouse also structures data in ways that make analysis easier.

The presentation component is the most visible component as the users interact with the data warehouse system through the presentation component. This component is often implemented with tools referred to as business intelligence (BI) tools. The presentation component represents different levels of services from simple reporting to roll-up and drill-down functionality to multidimensional analysis to true *ad hoc* query capability. What is worth mentioning in particular is the multidimensional analysis made possible by the integrated data in the data warehouse and the data analysis and presentation capability offered by BI tools. Multidimensional analysis allows the user to see the data the way he or she uses them.

The end user component refers to the various user communities that use the data warehouse, as demonstrated in Figure 2. The majority of users will be using the static reports most of the time. There is usually a sizeable group of users that will need to look beyond the reports to get the answers they want. They need to drill down or roll up. They need to look at the number

Figure 2 Architecture of a data warehouse system

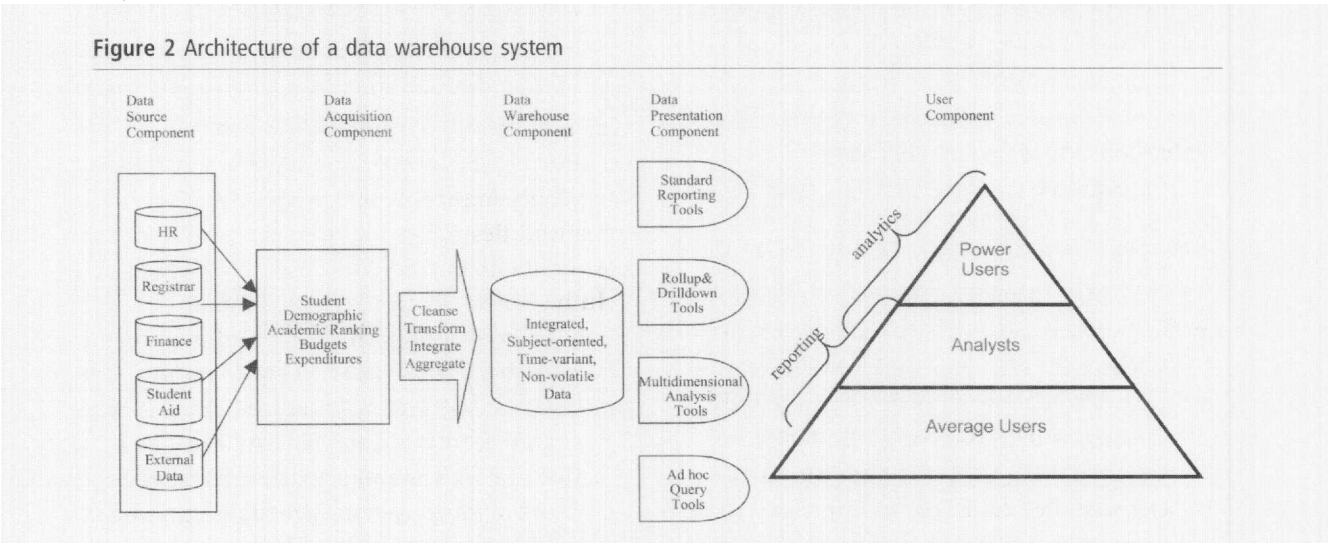
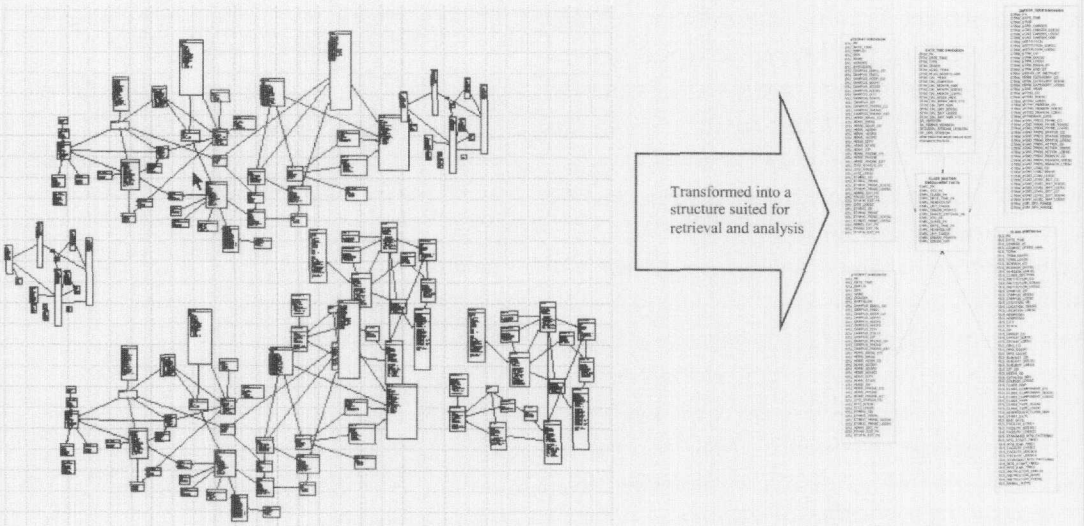


Figure 3 Transforming OLTP data into DW data



from different dimensions. Toward the top of the user pyramid, there are users that need to use advanced data analysis techniques, such as data mining and longitudinal analysis.

These needs can be addressed by a data warehouse system because of its data content and the special data structure it employs. The data warehouse system not only has to provide different types of access to meet the different information needs, but it also has to provide correct data. This implies that end users have a very critical role to play in the construction of data warehouses. They should help in selecting source systems. They should certainly be involved in designing rules in the data acquisition process that transfers the source data into the data warehouse.

Institutional support for data management

Massy and Wilger suggest that there are three levels of information technology adoption in the teaching and learning process. These stages may have considerable relevance for the adoption of technology to support institutional decision making. Each stage represents "different levels or degrees of innovation" (Massy and Wilger, 1998, p. 50). The levels include:

- *Personal productivity aids*. This level denotes the use of a standard mode of instruction but "allows teachers and learners to perform familiar tasks faster and more effectively using productivity software for work processing, spreadsheets, and graphing programs".
- *Enrichment add-ins*. This level adds "new materials to the teaching and learning mix without changing the basic mode of instruction". This includes the use of e-mail communication and listservs, Web pages, Web search engines, video, and the use of multimedia.
- *Paradigm shift*. This level is reached when "faculty and their institutions reconfigure teaching and learning activities to take full advantage of the new technology".

Massy and Wilger (1998) further suggest that the use of technology and the various levels of technology implementation have implications for quality improvement, productivity gains, enhanced flexibility, and may be a source of cost containment.

In higher education, this model has implications for the various levels of investment and commitment to data management and decision-support systems made by institutions. For example, institutions that embrace the practice of data warehousing or institutional data management systems probably have done so in response to internal and external

requirements for improved, timely, accurate and shared information. The commitment to invest in such technology and practice requires tremendous financial, human resources and management investment and would require a true "paradigm shift" in the way the institution captures, collects, stores and reports information. Conversely, those institutions unwilling to support an enterprise-wide knowledge management solution may simply be content to invest in desktop or "personal productivity" solutions for information and analyses. Consequently, organizational leaders may receive information needed to "get by", while a tremendous amount of flexibility and data richness will be lost. What is needed in developing effective information support for institutional decision making is a "paradigm shift" that reconceptualizes information systems as critical to decision making. Data warehousing provides an effective means for managing massive amounts of information intended for decision support.

References

Birnbaum, R. (1988), *How Colleges Work: The Cybernetics of Academic Organizations and Leadership*, Jossey-Bass, San Francisco, CA.

Inmon, W.H. (1996), *Building the Data Warehouse*, John Wiley & Sons Inc., New York, NY.

Linthicum, D. (2000), *Enterprise Application Integration*, Addison-Wesley, Reading, MA.

Massy, W.F. and Wilger, A.K. (1998), "Technology's contribution to higher education productivity", *New Directions for Higher Education*, Vol. 26 No. 3, pp. 49-59.

Further reading

Gray, P. and Watson, H. (1998), *Decision Support in the Data Warehouse*, Prentice-Hall, Upper Saddle River, NJ.

Hey, D.V. (2000), "Having it all", *Personnel Management*, Vol. 45 No. 3, March, pp. 127-31.

Kimball, R. (1996), *The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses*, John Wiley, New York, NY.

Soh, C., Kien, S.S. and Tay-Yap, J. (2000), "Cultural fits and misfits: is ERP a universal solution?", *Communications of the ACM*, Vol. 43 No. 4, pp. 47-51.