Invited Editorial

Using Real Data to Invigorate Student Learning

Nenad Jukic

School of Business Administration
Loyola University Chicago
1 E. Pearson Avenue
Chicago, Illinois 60611-2196 USA
njukic@luc.edu

Paul Gray

School of Information Systems and
Technology
Claremont Graduate University
130 E. 9th St.
Claremont, California 91711
paul.gray@cgu.edu

Abstract: Students need hands-on-experience with industrial strength databases and large data sets so they can deal with them when once employed. Many schools find it too expensive because of the cost and faculty resources required. That excuse is no longer valid. Colleges and Universities throughout the world can avail themselves of an extraordinary gift from Teradata Corporation that allows introducing large, complex problems into computer science classes without cost. The gift includes software and enormous databases available through the Internet as Software as a Service. This editorial describes the Teradata University Network, and how faculty can obtain this resource for their classes.

Categories and Subject Descriptors: H.2.m [Database Management]: Miscellaneous. H.2.6 [Database Management]: Database Machines. K.3.2 [Computer Milieux]: Computer and Information Science Education—*computer science education, information systems education, curriculum* K.8.1 [Computer Milieux]: Application Packages -- *database processing*

General Terms: Management, Design

Keywords: Database, Data warehouse, Decision support, Teaching, Portal

1. INTRODUCTION

In too many Computer Science database courses, particularly first courses in the field, faculty members follow the easy road. Select a textbook and follow it, chapter by chapter. Use the ancillary material publishers provide to ease the teaching burden. Make sure that students understand every principle and can spout them all back to you on an examination. Use the relatively simple, small (some say, toy) problems at the end of the chapters to give the students finger exercises that illustrate the principles. But under no circumstance prepare the students

for the true feel and experience of the massive data sets and the database problems they will need to cope with once they graduate and work in the real world. Teaching in this way deprives our students and abrogates our responsibility to them.

We recognize that many readers of *inroads* do indeed use large data sets and databases as part of their courses. However, as discussed in Section 2, the resources required represent a significant cost to all but the wealthiest institutions and many departments are priced out. Schools are faced with a classic opportunity problem: shall we use

the limited available resources to pay for the database and related software or shall we use them for other, equally important purposes?

In this editorial, we present an approach which offers students the capabilities they need, yet provides the first truly-free comprehensive resource for database-related educational use. The reasons for shortchanging our students are gone. They will be able to compete as professionals.

2. THE PROBLEM

When teaching computer-related subjects in Colleges and Universities, the cost of software and teaching materials is a major issue. Even when the software is given away by a vendor and no annual fee is charged, additional cost and effort is required for other items such as:

- installing, updating, administering, and maintaining the software on the computer.
- purchasing add-ons and teaching materials.
- obtaining and maintaining the data sets in the appropriate format used by students.
- providing faculty the teaching aids and materials that support the course, such as student exercises and homework assignments.
- using significant storage for the software and the data base files.

The problem is compounded by the need to teach students how to work with the size and complexity of that they will encounter in industry when they graduate. The importance of such realistic examples is high when teaching topics such as database management, data warehousing, data-driven decision support, or software as a service. The customary small examples shown in the texts and the problems that come at the end of the chapters, while illustrating the concepts taught, simply do not reflect the size or complexity of real world problems.

Fortunately, academia received an extraordinary gift from one of the large database/data warehousing vendors, Teradata Corporation, that makes it possible to introduce large, complex problems into the classroom and the computer science laboratory without cost. Specifically, Teradata offers universities the use of its database and data warehouse software (and decision support/business intelligence software through its business alliance with MicroStrategy) free of charge in a software-as-a-service model available over the Internet.

To make use of this gift, a consortium of university faculty from around the world formed the Teradata University Network (TUN) designed for faculty use and the Teradata Student Network (TSN) designed for student use. The two networks differ only in that teaching aids and case solutions are available to faculty but not to students.

In the sections that follow, we report on what is available to universities, what has already been achieved by the faculty at 861 universities in 69 countries (and growing), and to tell computer science faculty about how

they can participate in using the cornucopia of capabilities that are available in this environment. Information on becoming a member of TUN and how to use its facilities and capabilities in the classroom is given at the end of the editorial. Table 1 shows the features available through TUN and through TSN.

Table 1: Features Available Through TUN and TSN

	Available to faculty	Available to students
Teradata Database DBMS via Teradata SQL Assistant/Web Edition	x	x
OLAP Tool –MicroStrategy BI Version 8	x	x
Data modeling software tools (FatFree ERD, FatFree Star)	x	x
Trial copies of expert system, ETL, and data quality/ integration tools	x	x
Articles from journals and from magazines	x	x
Case studies	x	x
Data sets (of various sizes and complexity)	x	x
Teaching tips	x	
Tests and exercises	x	
Lecture notes, PowerPoint presentations	x	
Course syllabi	x	
Podcasts	x	x
Web-based training	x	

3. DATABASE

The list of topics included in a typical database course remained stable for many years. The consensus called for covering topics associated primarily with transaction processing and operational systems such as

conceptual database modeling	SQL
relational database modeling	data administration
data normalization	

However, new topics are emerging, including:

data warehousing	data marts
online analytic processing(OLAP)	data mining
customer data integration	

each of which are associated with large analytical databases, continually increasing in size, and their use in decision making. Our current graduating students are expected to demonstrate knowledge and understanding of

database concepts, data analysis, data management, and the role of each of them in decision making.

TUN makes the Teradata Database engine available to students. As the corporate name implies, the Teradata database platform is capable of efficient handling of terabytes of data. It is considered the fastest of the database engines available on the market. The Teradata Relational Database Management System (RDBMS) uses a standard Structured Query Language (SQL) and serves as a platform for high-end data warehouses.

The RDBMS is hosted and maintained on a massively-parallel-processing (MPP) server, which provides for extremely fast execution of SQL queries, even on large data-sets. Although users are welcome to create their own data sets, they are also given access to collections of already created data sets. These collections include:

- Corporate data sets from Sam's Club, Dillard's Department Stores, and Frozen Foods Inc. that total more than 195 million of rows of data, through gifts obtained and managed by The University of Arkansas.
- Data-sets associated with six of the most popular database management text books.

The Teradata Database can be used as to illustrate the RDBMS tool category on its own, and it can also be used in conjunction with Oracle, MySQL, and MSAccess RDBMS software. Incorporating Teradata into a database course expands the "software used" category in students' résumés. Furthermore, students can achieve formal basic Teradata certification through the site.

The corporate data sets hosted and managed by the University of Arkansas provide unique and valuable opportunity to expose students to assignments and scenarios based on the large scale, real-world data. University faculty who tried, on their own, to obtain access to such corporate transaction data for teaching use know how difficult (and often impossible) the task is. Fortunately, the University of Arkansas was able to acquire these valuable data sets and then generously share them with the TUN community. The structure and the size of these data sets provide a dimension of reality that can not be achieved by the textbook examples alone. The Teradata Database platform at the University of Arkansas, used to store this data, provides quick and user friendly access to it for TUN and TSN members.

4. DECISION SUPPORT THROUGH MICROSTRATEGY

Microstrategy 8 BI is the decision support software tool hosted by TUN. This tool expose students to the fundamental OLAP functionalities (e.g., pivoting, drill-down, drill up, slice-and-dice), to standard analytic techniques used in industry, and to display techniques ranging from simple graphs to complex dashboards. A number of data sets are supplied for use with this software

as are scripted demos and in-depth tutorials for faculty who only want to demonstrate decision support capabilities in data warehouses.

5. ER AND DIMENSIONAL MODELING

Entity-Relationship (ER) modeling is one of the principal techniques for conceptual modeling of both operational databases and data warehouses. Dimensional modeling is a principal modeling technique for databases to support end-user queries in data marts and data warehouses. As part of TUN, an ER modeling tool, FatFreeERD, and a dimensional modeling tool, FatFreeStar, are available for download. Neither tool requires installation since they can be downloaded, unzipped, and then instantly used to create conceptual database models for class exercises and projects. Both tools represent a free-tool alternative to commercially available case tools. Neither tool requires licensing.

6. DEMONSTRATIONS AND TEACHING AIDS

TUN offers access to classroom demonstrations and free trial copies of

- expert system tools,
- ETL tools, and
- data quality and integration tools.

Note that every tool available on TUN is offered as *an* example (rather than *the* example) of a tool category.

Combined with the collection of lectures, articles, case studies, and other teaching materials provided on TUN, these resources can be used on their own, or in conjunction with text books and other content, to provide a solid theoretical foundation of concepts and methodologies. Some texts integrate the use of TUN into their instructional materials.

Learning the concepts and theory as the foundation and the tool as a mean of implementation, positions students as prospective employees that understand how to use tools to facilitate processes and solve problems. With that understanding, students can switch from one platform to another and become competitive in the job market.

7. COURSES IN WHICH THE TERADATA UNIVERSITY NETWORK CAN BE USED

As shown in Table 2, TUN provides access to teaching materials that range from traditional database courses to electives.

The course materials can be accessed in three ways:

- by Course Type
- by Content Type
- by Teaching Tips

In the Course Type option, all of the TUN materials of use in a particular course (e.g. articles, assignments, PowerPoint presentations related to the database course) are listed in a structured and organized way for that course.

In the Content Type option (represented on the TUN web-site by Content Types and Software tabs), faculty interested in looking for a particular category of available resources (e.g. articles) can see the lists of all available resources of the same category (e.g. content type articles which give a structured and organized list of all articles available on TUN).

Table 2: Examples of Courses for Which TUN Can Be Used

traditional data base course	data warehousing
decision support and business intelligence	data mining
BPM (business performance Management)	data integration
CRM (customer relationship management)	data visualization

One of the content type options is the Teaching Tips link. This link provides the descriptions and guidelines on how materials can be used together, organized by various topics that are often covered in different courses. For example, for the topic of Dimensional Modeling and OLAP, an entry in the Teaching Tips link includes suggestions on how to introduce the topic to the students, which articles to assign as readings, what assignments to do, which tools to use, and how and which demos apply.

8. SOFTWARE AS A SERVICE

TUN works as software as a service (SaaS). That is, students use the Internet to interact with the software, hardware, and other materials located at and operated by Teradata from its facilities. As pointed out in the Introduction, the result is a cost-free arrangement for schools. By using TUN, faculty and students have the opportunity to learn about and experience the SaaS concept, a concept which is likely to become a significant way that information systems are operationalized in the years ahead.

9. COMPARISON TO OTHER SOURCES

Several other vendor-sponsored academic initiatives are available for classroom use, notably Oracle Academy and Microsoft Developer Network Academic Alliance. TUN differs from those initiatives in several ways:

- Unlike any of the similar initiatives, the content on TUN is completely free whereas similar initiatives incur significant costs even when the software is given to the school. The difference in cost comes from Teradata's use of the Software as a Service approach.
- Other initiatives are basically in-house departments of commercial vendors.

- TUN is steered and managed by the advisory board (see below) almost entirely composed of university professors. Teradata Corporation provides the budget as well as hosting for TUN, but the faculty board is in charge of the content and structure of the portal, giving authentic focus to the needs of educators and students.

10. THE TUN ADVISORY BOARD

The TUN advisory board (Table 3) consists of representatives from both academia and from Teradata. The Teradata members act in an advisory role and help decide what software and services are made available through the network.

Table 3: TUN Advisory Board

Executive Directors	
Michael Goul	Arizona State University
Jeff Hoffer	University of Dayton
Barbara Wixom	University of Virginia
Senior Board	
Scott Gnau	Teradata Executive Sponsor
Sanju K. Bansal	Teradata Corporation
Paul Gray	Claremont Graduate University
Darryl McDonald	Teradata Corporation
Hugh J. Watson	University of Georgia
Ron Swift	Teradata Corporation
Board Members	
Thilini Ayachandra	Xavier University
Paul Cronan	University of Arkansas
Haluk Demirkan	Arizona State University
Mary Gros	Teradata Corporation
Babita Gupta	California State University
Lakshmi S. Iyer	University of North Carolina
Nenad Jukić	Loyola University Chicago
Robert Winter	St. Gallen University, Switzerland

11. BECOMING A MEMBER OF TUN

The TUN web site is designed exclusively for university faculty and the membership is free. University faculty apply to become members of TUN by completing the "register" option form available on the TUN website: www.teradatauniversitynetwork.com

New members receive a welcome e-mail once the membership application is approved. The registration information provides TUN with contact information and allows TUN to authenticate that the application comes from a faculty member.

Students enrolled in a course taught by TUN members are given a password by their instructor so that they can access the Teradata Student Network (TSN). When students use the password for the first time to access the student website (www.teradatastudentnetwork.com) they

complete a short registration process after which they are ready to use this resource.

12. WHY BECOME INVOLVED

TUN offers a giant step for improving student learning. It also offers the student that extra line on their résumé that opens doors to jobs because they have experience dealing with large databases and, with just a little effort on their part, Teradata certification

For the faculty member, in addition to being cost free, TUN is now supported by some of the leading database textbooks. In addition, the web site contains problems, teaching hints, and large databases that are already keyed in. For the instructor and for the student, the learning curve is not steep. Yes, we recognize that adding material to a course requires removing some things of equal length. It is, like all change, a zero-sum game. However, our experience is that there are topics that are or are becoming

obsolete and should be replaced to keep courses current. Furthermore, TUN provides the support needed for starting new advanced courses that cannot be undertaken with existing resources. And you will even learn something new that will make your teaching life more interesting and personally rewarding.

13. CONCLUSION

Since the Teradata University Network was established in 2001 it has become a prime source for teaching and learning about database management, data warehousing, data-driven decision support, and many other contemporary data management analysis topics. Its board members and users continue to contribute new material. It is a capability of great value to the teaching community and it available at no cost. The Computer Science community is indebted to Teradata for making its capabilities available to us.

Check out the

AIS

Association for Information Systems

<<http://plone.aisnet.org/>>