Modeling of students' profile and learning chronicle with data cubes

Ade G. Ola
Virginia State University

Xue Bai
Virginia State University

Emmanuel E. Omojokun
Virginia State University

ABSTRACT

Over the years, companies have relied on On-Line Analytical Processing (OLAP) to answer complex questions relating to issues in business environments such as identifying profitability, trends, correlations, and patterns. This paper addresses the application of OLAP in education and learning. The objective of the research presented in the paper is to depict learning data with multidimensional data cube model so as to enable tools capable of answering multifarious questions regarding instruction and learning. Educational data abound in various formats and platforms. There are also interoperability standards which enable diverse applications to interact and share data. The challenge is to provide educators and teachers with interactive tools to identify the needs of individual students and to enable personalization of instruction and learning to meet the needs of the learners. Such tools will allow teachers to analyze profiles and performance history of individual students across multiple dimensions, calculate individual students' mastery of the curriculum, and recommend necessary changes to their learning plans. The ultimate goal of the research is to improve student learning by providing the ability to establish the acquired knowledge and skills of individual students through portable learning management cubes, which would maintain the profile and learning chronicle of the learners.

Keywords: Data cube, individualized instruction and learning, multidimensional data model, learning and educational data, OLAP

INTRODUCTION

The overriding goal of the research reported in this paper is to improve student learning and the success rate in K-16 courses through modeling of learning and educational data with multidimensional data cubes. In colleges, prerequisites are conditions of enrollment that students are required to meet prior to enrollment in particular courses and programs. However, students who have successfully completed prerequisite courses are not always ready for the corresponding target courses. It is also not uncommon for a K-12 student who has been promoted to a particular grade to possess skills well below that grade level in particular courses. Thus, it is becoming increasingly apparent that teachers and educators at all levels need superior analysis tools to meet the learning needs of their students:

1. Teachers in the classrooms should have ready answers to such questions as what proportion of the students has mastered a particular topic or curriculum item.
2. School administrators should have a means of comparing one cohort of students to another group of students
3. Local Education Authorities may need to know the difference in the performance of students who received one type of intervention as opposed to another type or compare how long it took the average student to master a particular skill using a particular curriculum
4. College program coordinators may want to know what percentage of cohorts of students possess the prerequisite knowledge and skills to enroll in a gateway course or determine the topics that particular students need to master to be admitted to a program.

Data analysis tools and decision support systems have been utilized widely in schools. The work reported in (Breiter and Light, 2006) identifies various data needs, at different levels of aggregation, for decision-making in the school environment. (NFES, 2006) also classifies the questions that a robust decision support system could potentially address as: questions about the classroom, operational questions that extend beyond the classroom, and policy questions at the state level. Moreover, there are several software that allow schools to gather and analyze data to improve student learning and school administration; a list of data analysis and assessment tools is presented in (Tech & Learning n.d).

The focus of this paper is on depicting learning data with multidimensional data cube model so as to enable tools capable of answering multifarious questions regarding instruction and learning and facilitate implementation of personalized learning systems. Educational data abound in various formats and platforms. There are also a number of interoperability standards which enables diverse applications to interact and share data, including Schools Interoperability Framework and inBloom technology (SIF, n.d; inBloom n.d.). What is needed is a sound logical model upon which useful tools for schools and learning can be implemented. Such tools will allow teachers to analyze profiles and performance history of individual students across multiple dimensions, calculate individual students' mastery of the curriculum, and recommend necessary changes to their learning plans. The research presented in this paper began with the premise that dimensional data model would be as effective in modeling the learning environment at it has been in the business environment. Dimensional model of data has proven to be an effective tool in the business environment. Over the years, companies have relied on OLAP and the underlying dimensional model of data to answer complex questions relating to issues in business environments such as identifying profitability, trends, correlations, and patterns. The ultimate goal of the research is to provide the ability to establish the acquired knowledge and skills of the individual student through portable learning management cubes, which would maintain the profile and learning chronicle of the student.

Overview of the Dimensional Data Model

This section presents an introduction to data cubes and dimensional data model; it is based on the overview described in Oracle® OLAP Application Developer's Guide (Oracle OLAP, n.d.; Oracle Docs, n.d.) and Data Cube Vocabulary (RDF Data Cube Vocabulary, 2013). The central structure of the dimensional model is the cube—hypercube to be exact. The edges of a cube consist of dimensions, where each dimension represents a real-world entity in the environment being modeled. The cells in the data cube represent the measure of interest. For example, a Sales measure may have four dimensions: Time, Customer, Product, and Store. Any given Sales measure has meaning only if the time, the customer, the unique product, and the sales store are known. Figure 1 depicts a set of business processes where business measures (such as sales, cost, item count) are represented as facts (in a Measure table) within the context of certain business entities (Product, Customer, Store and Time dimensions). The model allows Sales, Cost, Profit to be analyzed with respect to Product, Store, Time, Customer, or any combination of the dimensions. While the information presented in the dimensional cube model do exist in the source systems (such as a relational database), the Cube representation depicts information in a form that is easily understood because it uses the semantics of the business directly.

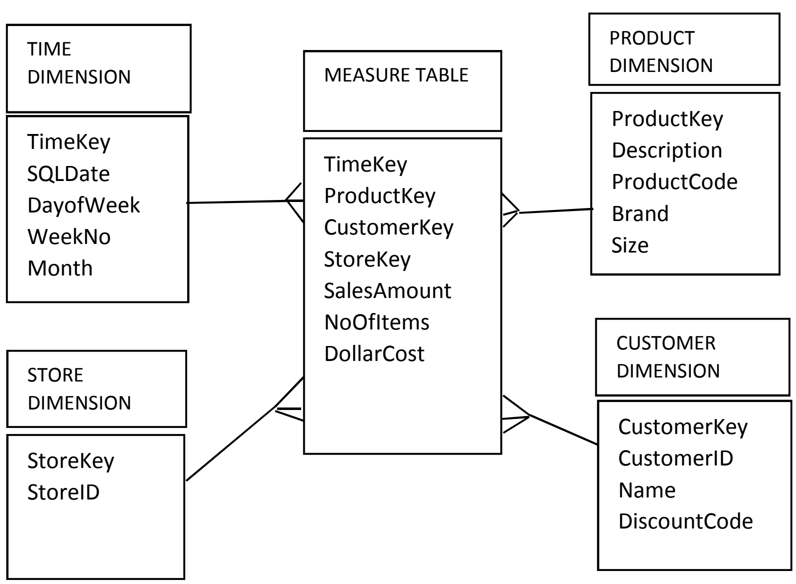

Figure 1: A 4-Dimensional Data Cube Model of a Business Environment

Thus, a data cube representing *m* dimensions is an *m*-dimensional array in which each element of the array contains the measure value, such as sales amount, a count, or a performance score.

In addition to the dimensions, *attributes* provide additional information about the data. For instance, a Product entity/dimension may have *size* or *color* as attributes. (RDF Data Cube Vocabulary, 2013) describes attributes as structural metadata that can be attached to individual observations/measures or to higher levels. Metadata give meaning to the measure, providing answers to questions such as whether the value is measured or estimated. Dimensions may also have hierarchies, which would allow data/measures to be analyzed at different levels. A level is a named position in a hierarchy; it provides an easy way to reference a group of dimension members that are at the same distance from the base. A Time dimension might have a hierarchy that represents data at the month, quarter, and year levels, and thus a hierarchy of the dimension would consist of Month, Quarter, and Year. A hierarchy such as the Month-Quarter-Year of Time is said to be level-based. Time attributes can also provide information about the Time dimension that may be useful in some types of analysis, such as identifying the last day or the number of days in each time period (Oracle OLAP, n.d.). Some dimension, such as Employee, may not have natural hierarchical levels. So, for the Employee dimension, a hierarchy such as the relationship between the reports and a manager is referred to as value-based hierarchy.

Thus, to use the language in Data Cube Vocabulary (RDF Data Cube Vocabulary, 2013), a data cube is organized according to a set of these components: *dimensions*, *attributes* and *measures*. The *attribute* components allow the observed value(s) to be qualified and interpreted. They enable specification of the units of measure, any scaling factors and metadata such as the status of the observation (e.g. *estimated*, *provisional*). The *dimension* components are used to identify the observations, and the *measure* components represent the phenomenon being observed. It is possible to have multiple measures as in the example in Figure 1 (sales, cost, item count), or as in a data set on student performance which might provide multiple different performance indicators for each curriculum item. There are two approaches to representing multiple measures supported by the Data Cube vocabulary (RDF Data Cube Vocabulary, 2013). In the first approach, an additional dimension is introduced and each observation records a single observed value for one measure. In the second approach a single observation can provide values for multiple different measures. This *multi-measure* approach is commonly used in applications such as Business Intelligence and OLAP.

Modeling of Learning and Educational data

This section presents a dimensional data cube model of a learning environment. In this case, the edges of the cube depict four dimensions: Student, Learning Objective, School, and Time. In a typical Local School Authority, the administration wants to track students' performance over time with respect to various learning objectives. Over the course of the school year students are evaluated against the learning objectives. Students move through the curriculum as they satisfy the learning objectives. In cases where a student does not meet the performance requirements for particular learning objectives, supplementary instruction and learning activities are prescribed. A representation of the cube is presented in Figure 2; the performance indicators stored in the Measure table can be analyzed with respect to the dimensions (Student, Learning Objective, School, and Time dimensions).

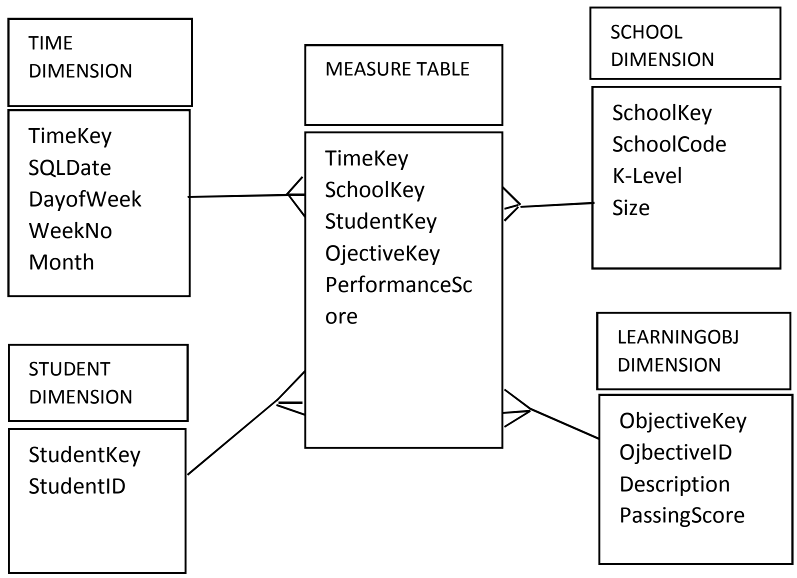

Figure 2: A 4-Dimensional Data Cube of a Learning Environment

The Cube representation depicts the learning data in a form that is easily understood by teachers and educators because it uses the semantics of the learning environment directly. With the model, an analysis can provide answers to such questions as:

1. What proportion of the students has mastered particular learning objectives?
2. How much improvement has a student made over a period of time?
3. Which learning objectives have not been met by particular students?
4. What are the ranks of schools by the percentage of students that met a preset performance level?

A major part of the appeal of dimensional data cube model is that it provides the stakeholders a depiction of their business in the language of that environment. In the example presented in figure 2, if the stakeholders are interested in comparing students' performance by instructional mode or by teacher, introducing two dimensions (e.g., Teacher, Mode) is all that is required.

CONCLUSION

In this paper, it has been shown that dimensional data cube along with OLAP, which has been used successfully in the business environment, is a good candidate for modeling educational and learning data. Just as multidimensional data cube and OLAP have been used in answering complex questions relating to issues in business environments, the data cube model would enable tools capable of answering multifarious questions regarding instruction and learning. The need for a model to integrate school data is evidenced by the inBloom Technology (inBloom, 2012) and the interest it has generated. The inBloom technology and its logical data model provide rich source of information on the entities, dimensions, and data items in the K-12 environment. The ultimate goal of the research presented in this paper is to improve student learning by providing the ability to establish the acquired knowledge and skills of individual students through portable learning management cubes, which would maintain the profile and learning chronicle of the student.

REFERENCES

Breiter, A., & Light, D. (2006). Data for School Improvement: Factors for designing effective information systems to support decision-making in schools. Educational Technology & Society, 9 (3), 206-217.http://www.ifets.info/journals/9_3/18.pdf

inBloom (2012) inBloom Logical Model. https://www.inbloom.org/sites/default/files/docs-developer/data_model-intro.html

NFES 2006–807 (2006). National Forum on Education Statistics. Forum Guide to Decision Support Systems: A Resource for Educators

(NFES 2006–807). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved December 2014 from http://files.eric.ed.gov/fulltext/ED493913.pdf

Oracle Docs (n.d.). Oracle Database Online Documentation 10g Release 2 (10.2) http://www.oracle.com/pls/db102/portal.portal_db?selected=5

Oracle OLAP (n.d.). Oracle® OLAP Application Developer's Guide,10g Release 2 (10.2) B14349-05 http://docs.oracle.com/cd/B19306_01/olap.102/b14349/toc.htm

RDF Data Cube Vocabulary (2013). The RDF Data Cube Vocabulary. W3C Editor's Draft 17 December 2013. https://dvcs.w3.org/hg/gld/raw-file/default/data-cube/index.html#data-cubes

SIF (n.d.) Schools Interoperability Framework (SIF) http://specification.sifinfo.org/Implementation/2.0/

Techlearning (n.d.). Analysis and Assessment Tools. http://www.techlearning.com/section/products/0038