Developing Data Marts for Healthcare

Xiaoni Zhang  
*Northern Kentucky University, zhangx@nku.edu*

Kevin Gallagher  
*Northern Kentucky University, gallagherk2@nku.edu*

Samuel Goh  
*Northern Kentucky University, samuel.h.goh@gmail.com*

Follow this and additional works at: [http://aisel.aisnet.org/amcis2011_submissions](http://aisel.aisnet.org/amcis2011_submissions)

Recommended Citation

[http://aisel.aisnet.org/amcis2011_submissions/438](http://aisel.aisnet.org/amcis2011_submissions/438)
Developing Data Marts for Healthcare With SAP

Xiaoni Zhang
Northern Kentucky University
zhangx@nku.edu

Kevin Gallagher
Northern Kentucky University
gallagherk2@nku.edu

Samuel Goh
Northern Kentucky University
samuel.h.goh@gmail.com

ABSTRACT
Research shows that information technology adoption in healthcare continues to lag behind many other industries, in response to this challenge. This paper outlines a graduate course designed to give students a hands-on experience in building a data mart in a healthcare setting. Students will learn technical skills as well as teamwork, communication and problem solving skills. The structure of the course requires students to apply classroom concepts. For example, students use multiple methods to collect information for multidimensional modeling techniques as they identify various InfoObjects, such as characteristics, units, key figures and time characteristics. This course design presents an opportunity for students to gain valuable experience in building a healthcare-related system in SAP and thereby developing highly marketable skills.

Keywords
Data mart, InfoObject, key figure, dimensional modeling.

INTRODUCTION
According to The Office of the Actuary in the Centers for Medicare & Medicaid Services (http://www.cms.gov/NationalHealthExpendData/03_NationalHealthAccountsProjected.asp#TopOfPage), national health expenditures will greatly increase over the next decade. Faced with the pressure to reduce costs, the healthcare sector has turned to IT systems, through which better information can deliver benefits such as improving care quality, reducing cost and ensuring patients privacy (Baker et al. 2008). For example, technology has transformed areas such as care delivery, practices and communications in healthcare. Traditionally, healthcare has lagged behind other industries in technology adoption (Khoumbati et al. 2006, Menachemi et al. 2006) and represents tremendous growth opportunities (iHT2, 2010). Therefore, getting students involved in healthcare related projects is important to raise awareness of healthcare-related opportunities in the areas of operations, requirements, reporting and thereby help them in developing and attaining more marketable skills.

In this paper, we describe the pedagogical approach taken in a course titled “Data warehouses and data marts”. This course required students to work in teams to carry out a semester long project on a data mart development project. Students chose a project in healthcare, investigated a specific healthcare setting and its related IT systems. The students then develop a data mart based upon their team’s analysis. Projects are self-selected, so students can examine almost any field of healthcare: a family practice, hospital, lab, pharmacy, etc. The purpose is to encourage students to take ownership of the project, thereby hopefully dedicating more time and focusing more attention on the healthcare business operations and practices, since the topic is one of interest to them.

The prerequisite for this course is an introductory level SAP class. The learning objectives of the course include 1) familiarity with the area of healthcare operations that students choose; 2) understanding the IT applications used by the healthcare professionals; 3) developing business requirements for a data mart; 4) conducting the logical design of a data mart; 5) implementing a data mart in SAP. Students are asked to analyze the business problems within the context, examine its current reports and define data/information needs for reporting. Students perform the role of a data mart solution analyst to define business requirements. They are asked to identify business questions based on the case setting and to design a data mart to serve the needs of the business. Business analytic skills and technical skills are required for this project. Through the course, students will learn technical skills as well as teamwork, communication and problem solving skills. Students use multiple methods to collect information for multidimensional modeling. Students are required to identify various InfoObjects.
InfoObjects are business evaluation objects and are classified into several categories: characteristics, key figures, units, and time characteristics.

**DATA MART PROJECT**

As the course encompasses collecting business requirements necessary for a data mart, analyzing the business requirements, developing a star schema, and building a data mart in SAP, this requires a whole semester to accomplish. Students form groups consisting of three or four students. The teams are formed in the first week of the class. Since this is a graduate level course and almost all of the students have professional experience, students are asked to find a real world project or a project that simulates real world healthcare settings. The teams make their own decisions on what kind of project to take on. Students may get project ideas from their own experiences, acquaintances, friends, or relatives. Students are advised to interview healthcare professionals, understand applications used by healthcare setting, define reporting parameters while examining current forms, and reports. In addition, students are required to read materials provided by the instructors and search the Internet for data mart related information. At the end of the semester, students present their healthcare business case and data mart development process to the whole class. Lastly, each student will be evaluated by teammates at the end of the semester.

To help students understand the technical aspects of data mart development, we provide students a simple business case and two reports documenting customers and products information. We show students several sample reports and walk through the logical design of star schema.

**Sample Business Scenario**

ABC company sells electronics. ABC uses a system called ETRAN to capture customer information and product information. Table 1 shows sample customer information stored in its database. Table 2 shows sample product information captured in its database. Employees and managers query their database to produce various reports.

<table>
<thead>
<tr>
<th>Customer ID</th>
<th>First name</th>
<th>Last name</th>
<th>Phone</th>
<th>Street</th>
<th>City</th>
<th>State</th>
<th>Zip</th>
</tr>
</thead>
<tbody>
<tr>
<td>10000</td>
<td>Mary</td>
<td>Lin</td>
<td>8596000000</td>
<td>Ash</td>
<td>New Port</td>
<td>KY</td>
<td>41099</td>
</tr>
<tr>
<td>10001</td>
<td>Jack</td>
<td>Chen</td>
<td>5133000000</td>
<td>Pearl</td>
<td>Cincinnati</td>
<td>OH</td>
<td>45242</td>
</tr>
<tr>
<td>10002</td>
<td>Wyne</td>
<td>Johnson</td>
<td>8722345000</td>
<td>Nunn</td>
<td>Denton</td>
<td>TX</td>
<td>76203</td>
</tr>
<tr>
<td>10003</td>
<td>Tom</td>
<td>Smith</td>
<td>9144003000</td>
<td>Reagon</td>
<td>Dallas</td>
<td>TX</td>
<td>75201</td>
</tr>
<tr>
<td>10004</td>
<td>Joseph</td>
<td>Brandon</td>
<td>3127809000</td>
<td>Lakeview</td>
<td>Chicago</td>
<td>IL</td>
<td>60613</td>
</tr>
</tbody>
</table>

**Table 1 Sample Customer Information**

<table>
<thead>
<tr>
<th>Product ID</th>
<th>Category</th>
<th>Brand</th>
<th>Description</th>
<th>Price</th>
<th>Quantity</th>
<th>Sales</th>
<th>Date</th>
<th>Customer ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>D100</td>
<td>Desktop</td>
<td>Dell</td>
<td>Computer</td>
<td>1200</td>
<td>4</td>
<td>4800</td>
<td>2/15/2009</td>
<td>10000</td>
</tr>
<tr>
<td>L100</td>
<td>Laptop</td>
<td>HP</td>
<td>Keyboard</td>
<td>60</td>
<td>5</td>
<td>300</td>
<td>2/23/2009</td>
<td>10001</td>
</tr>
<tr>
<td>D300</td>
<td>Desktop</td>
<td>Sonny</td>
<td>CPU</td>
<td>200</td>
<td>20</td>
<td>4000</td>
<td>3/6/2009</td>
<td>10003</td>
</tr>
<tr>
<td>L300</td>
<td>Laptop</td>
<td>Dell</td>
<td>Mother board</td>
<td>150</td>
<td>5</td>
<td>750</td>
<td>4/5/2009</td>
<td>10004</td>
</tr>
</tbody>
</table>

**Table 2. Sample Product Information**

**Design of the Data Mart – Logical Design**

The design process is technology neutral. The star schema is the commonly used design method for data marts. It is used to implement multidimensional modeling. A star schema consists of a fact table and a few dimensions which are linked to the fact table. The fact table consists of primary key of the dimensional tables and key figures of the business transactions. Key figures are additive values that measure business performance. Dimensional tables consist of a primary key and characteristics. Characteristics are attributes of dimensions.

Figure 1 shows an example of a star schema. In the center of three dimensional tables (product, customer, and time), there is a fact table. The product dimension contains a key and characteristics: category, description and brand. Characteristics shown in the product dimension are based on the sample product report. The customer dimension contains a key and these characteristics: first name, last name, phone, street, city, state and zip. Characteristics shown in the customer dimensions are...
based on the sample customer report. The fact table consists of the primary keys of the dimension tables and key figures – additive values.

**Figure 1. Star Schema**

**InfoObjects**

In SAP Business Intelligence (BI), business objects are called InfoObjects. InfoObjects describe business related information for dimensions. They are the smallest unit in BI. They may be characteristics (for example, customers, products, patients or drugs). Students need to examine the sample forms/reports used by the employees of the business setting they select and determine the number and name of the characteristics they need. During the design stage, it is recommended that students use a word file to document the name of each characteristic, description, data type, field length etc. Table 4 shows an example of characteristics. Students need to use Table 4 as a sample to develop their own characteristics.

<table>
<thead>
<tr>
<th>Name of characteristics</th>
<th>Data type</th>
<th>Field length</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient/Customer First Name</td>
<td>Text</td>
<td>20</td>
<td>First name of a patient</td>
</tr>
<tr>
<td>Patient/Customer Last Name</td>
<td>Text</td>
<td>20</td>
<td>Patient last Name</td>
</tr>
<tr>
<td>Patient Date of Birth</td>
<td>Date</td>
<td>mm\dd\yyyy</td>
<td>Patient date of birth. The date format is mm\dd\yyyy.</td>
</tr>
</tbody>
</table>

**Table 4 Characteristics Documentation.**

**Key Figures**

Key figures are facts related to business operations. They can be aggregated for query and reporting purposes. For example, key figures can be amounts, quantities, or the number of items. In the role of a BI analyst, students have to decide how many key figures are needed in order to implement the data marts. In a laboratory setting, examples of measures include the number of tests and type of tests being performed each day, the number of sample label errors, laboratory testing errors, laboratory reporting errors, corrections of errors on ordered tests, order accuracy, and duplicate orders, the appropriateness of test orders, turnaround time, etc. Students are encouraged to elicit key figures from interviews with staff and/or managers.

**Dimensions**

Dimensions are tables that are used to describe business related information. Dimensions consist of characteristics. For example, customer and product are commonly used dimensions. In the healthcare environment, patient is a commonly used dimension. Another dimension may be lab test. Playing the role of a BI solution developer, students should examine the lab forms, decide on the number of dimensions required, and name these dimensions. Students also need to add a time dimension for reporting purposes so that reporting can be produced based on week, quarter, year or multiple years at ease. For example,
date, week, quarter, fiscal year are possible characteristics for the time dimension. As with key figures, reporting requirements may be elicited via interviews with the lab technicians and the manager.

IMPLEMENTATION OF THE DATA MART IN SAP

Implementation of data mart consists of the following steps: installation of SAP GUI client, creation of InfoObject, key figure and dimensions. For this course, we used SAP to implement a data mart, and the procedure is detailed below.

Installation of SAP GUI Client

In order to use the SAP software, universities must first join the SAP University Alliance. After becoming a member of the SAP University Alliance, user IDs and passwords will be issued by a hosting university and the SAP GUI software can be downloaded at the SAP community network site (https://www.sdn.sap.com/irj/scn/uac). The SAP University Alliance site contains step-by-step instructions on how to install and configure SAP GUI. For example, Northern Kentucky University uses the California State University at Chico as our SAP hosting center, which help set up course modules and issue user ID and password information.

Use SAP to Create InfoObject, Key Figures and Dimensions

The SAP site contains detailed instructions on how to create InfoObjects, key figures and infocubes (data mart). To view instructions on how to create InfoObjects, click http://help.sap.com/saphelp_nw04/Helpdata/EN/80/1a637fe07211d2acb80000e829f8fe/content.htm.

To view instructions on how to create key figures, click http://help.sap.com/saphelp_nw04s/helpdata/en/46/8e1d551870616ae10000000a1553f7/content.htm.

To view instructions on how to create infocubes, click http://help.sap.com/saphelp_nw70/helpdata/EN/80/1a643fe07211d2acb80000e829f8fe/content.htm.

In addition to the instructions provided on SAP web site, students are provided step-by-step PowerPoint slides showing how to implement a data mart. Due to space limitations, please contact the primary author for further details.

TEAM FORMATION

At the beginning of the semester, the instructor hands out a short survey to get basic information on students. This survey asks questions about prior courses taken in healthcare and system analysis and design, and prior experience with healthcare or labs and systems. Students with prior experience in healthcare or have taken courses in healthcare provide good working knowledge about workflows or healthcare specific issues. Students with prior experience in system analysis and design or have taken courses in system analysis and design are equipped with knowledge of the system design process. Therefore, matching context-specific knowledge with systems building knowledge is essential. Suggested team composition should generally include at least one student in healthcare and at least one in system analysis and design. Each team should elect a coordinator who is responsible for the management of the project and keep track of the progress of each task. Each team should meet at least once each week.

REQUIREMENTS GATHERING

Students are expected to be familiar with the concepts of data mart and related technologies from classroom lectures. Applying this knowledge requires students to be able to use fact-finding techniques, in combination with data mart concepts to define requirements. Students will interview relevant people and examine healthcare business operations, data captured by its IT systems, the current forms/reports used etc. Students will interview people with different responsibilities to get at a complete picture of healthcare business operations.

Students are encouraged to develop their own questions for interview. Several sample questions that students may ask during interviews are shown below. Students are encouraged to hold brainstorming sessions in order to develop questions specific to their project context.

- What is the nature of the organization’s tasks and data needs?
- What kind of data are currently stored and available in your systems?
- What information is presently used in generating the most valued reports?
- What database technologies are available for use in developing a data mart in your organization?
REQUIREMENTS ANALYSIS

Upon completion of requirements gathering, students are advised to design the data mart from a decision support perspective, starting on a smaller scale, and later expanding on the InfoObject. This allows for scope creep, a most common error, to be kept under control given the time constraints of one semester to complete the implementation of a data mart. It is important to encourage students to examine reports/forms carefully when deciding on the number of dimensions, the name of each dimension table, the characteristics in each dimension table, the name of these characteristics, and the characteristics in key figures as the results of the requirements analysis is what will be later implemented. For example, the name of the dimension table and the characteristics must be meaningful and self-explanatory. Also, characteristics in key figures are used in computations and are the most commonly used characteristics for reports. Students are advised to rely on the findings from their interviews with subject matter experts, for example examining the reports used by healthcare professionals provide hints on the appropriate characteristics for key figures, which allows the development of a system which more closely meets users’ requirements.

PROJECT DELIVERABLES

Students are required to provide a one-inch three-ring binder containing the following:

1. Description of a business scenario, provide background information on the project such as business environment, business operations, the number of people, products/services, forms/reports etc.).
2. Interview protocol
3. Name of the dimension tables and characteristics of each dimension table
4. Sample transaction data
5. Implementation of data mart in SAP.
6. Team presentation slides

CONCLUSION

This paper provides a tutorial on how to develop a data mart in a healthcare setting. Students form teams and choose a healthcare related project to develop a data mart. They interview healthcare professionals to collect business requirements and transaction forms/reports. Based upon the team analysis of business requirements, an actual data mart is implemented in SAP. Through this course, students gain deep knowledge about healthcare operations by contacting, observing or interviewing healthcare professionals. Given the growth opportunities in the healthcare field, we hope that students are able to gain applicable knowledge and build marketable job skills.

REFERENCES