
Student Graduation Paths: A Preliminary Data Warehouse Analysis

Mark Segall
segall@msudenver.edu
Department of Computer Information Systems
Metropolitan State University of Denver
Denver, CO 80217

Abstract

The ability to complete a college degree in a timely fashion is a concern for college administrators given the pressures to use resources efficiently and effectively. However, the traditional student graduation path of completing a bachelor's degree in four years on a residential campus does not apply to most students anymore. Given the difficulty of tracking a student's progress when many students go to school part time and transfer from one school to another, an alternate method is presented to accurately describe the graduation paths of college students. This paper proposes that students who complete a bachelor's degree be the basis for a descriptive system to inform stakeholders on the graduation paths for undergraduates. This system is not comprehensive in capturing the metric of graduation rates, but does address the difficult to measure metrics of transfer rates and time to get a degree for all students. A star schema dimensional model is proposed to capture the credits earned by a student and a preliminary analysis is discussed.

Keywords: Undergraduate Graduation Rate, Undergraduate Transfer Rate, Time to Degree, Dimensional Model, Star Schema, Data Warehouse

1. INTRODUCTION

The *Chronicle of Higher Education* has a micro site on the topic of College completion (Selingo 2012) which reports that of the 4.3 million students that entered college in the US in the Fall of 2004, only 1 million were reported to have graduated by six years later. Some students are not counted because they are part time (about 1.2 million) which leaves about half of the student, 2 million, as not officially graduating. Clearly a 25% graduation rate is alarming.

One problem with the data is that there is no mechanism to track students that transfer from one school to another. Another problem is that part time students are not counted toward graduation. Students going part time to school are a challenge to the traditional notion of going to a residential college and completing a degree

in 4 years. Given the increased cost of colleges, more students are working full or part time and completing their studies at a part time rate.

While it is still important to measure completion rates, this paper proposes that the graduation paths of student who do complete an undergraduate bachelor's degree be analyzed and reported by colleges in order to get an accurate picture of the time to graduation. This method will account for part time and transfer students.

2. DIMENSIONAL MODEL

Ideally we want to record information about every course in every semester from every college attended by a student. Having a single source of this information would allow for a complete analysis of graduation rates.

A data warehouse is ideally suited to hold this information. As defined by Inmon (1992), a data warehouse has 4 characteristics:

- Subject Oriented: not process oriented around transactions
- Integrated: from multiple operational data sources
- Time Variant: meaning it holds historical data
- Nonvolatile: in that data is not updated or deleted unless it is obsolete.

The data we are looking for is "Subject Oriented" around the college graduate, "Integrated" from operation databases in multiple schools, "Historical" since we want to see what trends there are in graduation paths, and "Nonvolatile" in that once a student graduates, there is no need to change the data.

A data warehouse dimensional model works well to implement a data warehouse (See Kimball & Ross, 2002). It contains a fact table in which each row contains facts about an individual or event, as well as dimensions (attributes). To conserve space in fact tables, which runs in the millions of rows, the dimensions are coded and linked to dimension tables. The single central fact table is surrounded by many smaller dimension tables leads to the description of this type of data warehouse as a 'Star Schema'. The fact table can be fine grained meaning it holds a detailed, atomic fact or coarsely grained meaning it holds summarized facts. Finely grained fact tables can be rolled up into coarsely grained tables. The dimensional model is an efficient way to store the data versus a standard database which is optimized for transactional processing. The facts can be thought of in statistical terms as quantitative dependent variables while the dimensions are the categorical independent variables. This conceptualization of facts and dimensions it easy to build systems that business analyst can use to create useful reports.

The dimensional design for this ideal "Enrollment" fact table, which holds every course taken, might have five dimensions and one fact (Figure 1).

This would allow us to determine the number of courses a student takes each semester, the number of courses the student takes at each college, and the dates the courses are taken.

Enrollment Fact Table
Term (FK)
Course (FK)
Student (FK)
Faculty (FK)
College (FK)
Grade

Figure 1

Once a student enrolls in the course the information will not change. The outcome of the course (pass, fail, or drop-out) will be part of the historical record. A student that takes the same course more than once will have two rows in the fact table, only the Term and Grade will be different.

However, there is no mandatory nationwide system to collect this detail of data. A student that transfers to a new college has no obligation to inform the previous school of the change in the enrollment status, but the student does have a motivation to tell the new college about all of the previous colleges attended. In order to get credit for previous courses taken, the student is motivated to submit transcripts from previous schools. This information allows the graduation school to construct the path the student took to complete their degree. The school the student graduates from will have the detailed information about the courses taken at the current school. Taken together the school the student graduates from can construct the graduation path of the student. Thus the model proposed in this paper relies on the graduating school to construct the path that students take to graduation.

The dimensional table that combines data from multiple schools is of a courser grain. The "School Enrollment" table proposed has five dimensions and five facts (Figure 2). The Student dimension records a primary key for each student and important demographic attributes. The School dimension allows an analysis of where students transfer from and how many credits are transferred in. The Enrollment Start Date and Enrollment End Date dimensions use the same Time dimension table allowing for the calculation of how much time it takes for a student to complete a degree over multiple schools. The Degree Major dimension

captures the degree and the division/school within the college the student is enrolled in.

The facts for this table are the number of Courses Attempted, Credits Attempted, Courses Completed, and Credits Completed. This helps determine how often a student takes courses multiple times. The granularity of this table does not pinpoint the courses that are repeated. The last fact for this table is if the student Graduated. This will be critical if data from multiple schools are combined into a single table. Also students can start at college A, transfer to college B, and then transfer back to college A to finish the degree. The Graduated fact along with the Enrollment End Date identifies the final semester for the student.

School Enrollment Fact Table
Student (FK)
Enrollment Start Date (FK)
Enrollment End Date (FK)
School (FK)
Degree major (FK)
Courses enrolled
Credits enrolled
Courses Completed
Credits Completed
Graduated (Y/N)

Figure 2

3. RESULTS

MSCD is an urban institution with no official residential facilities. Data was collected for 42 CIS majors from the school of business graduating in the 2011/2012 academic year.

What Percent of students start as freshmen at MSCD over the last 5 years?

In the 2007/2008 academic year 41% of students started as freshmen, while 38% were freshmen in 2011/2012. Most students (59% in 2007/2008 and 62% in 2011/2012) transfer to MSCD after taking courses at other institutions (MSCD Common Data, 2012). This increase could be due to new articulation agreements with local community colleges that allow

students to continue at MSCD after completing an associate's degree (Table 1).

Academic Year	Transfer	High School
2011/2012	62%	38%
2010/2011	62%	38%
2009/2010	60%	40%
2008/2009	58%	42%
2007/2008	59%	41%

Table 1

How many schools did graduates transfer credits from?

Fewer graduates came in with a High School degree, about 33%, than the recent incoming class of students while 67% of students transferred in credits (Table 2). Of the 28 transfer students, 75% transferred in credit from only one or two schools. One student transferred credit from six schools.

Number of schools	Freq.	Percent
0	14	33.3
1	12	28.6
2	9	21.4
3	3	7.1
4	3	7.1
5	0	0
6	1	2.3

Table 2

Where are students transferring from?

The 28 transfer students transferred in credit from 32 unique institutions, Eighteen (56%) of the institutions where from the same state of Colorado. The top five schools where students transferred from were local community colleges.

How many credits are transferred in?

The average number of credits transferred in is 43.2 with a standard deviation of 22.9. The minimum number of credits is three and the maximum is 80. Transfer students come in with about three semesters of credit.

How long does it take to earn a degree?

For all graduates in 2011/2012 the average number of years to complete a degree was 8.6 ± 4 SD. Students who only attended MSCD graduated in an average of 6.9 years ± 2.6 SD, while transfer students graduated in an average of 9.5 years ± 4.2 SD.

How many credits are earned each semester?

While all graduates were at MSCD they took an average of 8.5 credits per semester or about three courses per semester. This indicates that they were taking courses on a part time basis.

4. CONCLUSIONS

MSCD is a school that does not have a traditional student population. Many students come to the school with credit from other colleges and do take less than the typical 15 credits per semester.

This paper illustrated a method to describe the graduation path that students take and can be used at any school.

There are a number of programs to improve the measurement of student graduation rates. One is the Common College Completion Metrics (Reyna 2010). They promote that all states use 4 standard Outcome Metrics (Degrees Awarded,

Graduation Rates, Transfer Rates, and Time & Credits to Degree). The graduation rates metric assumes that the normal program time of a bachelor's degree is 4 years and the extended time is 6 years. The data from this study suggests that this is not a reasonable time frame. Not all students complete their degrees in 4 years taking 15 credits per semester.

An accurate measurement should be taken of graduation rates to accommodate the realities that students face.

5. REFERENCES

- Inmon, W.H. (1992), "Building the Data Warehouse." John Wiley & Sons.
- Kimball, R., & Ross, M. (2002) The Data Warehouse Toolkit. Wiley Computer Publishing, New York.
- Reyna, R.. (2010). Common College Completion Metrics. Retrieved June 10th, 2012 from <http://www.nga.org/files/live/sites/NGA/files/pdf/1007COMMONCOLLEGEMETRICS.PDF>
- MSCD Common Data Sets. (2012). Retrieved June 28th, 2012 from <http://www.mscd.edu/oir/commondatasets/>
- Selingo, J. (2012). College Completion. Retrieved May 15th, 2012 from <http://collegecompletion.chronicle.com>

