

# 13

## Education

**W**e step into the world of an educational institution in this chapter, looking first at the applicant pipeline as an accumulating snapshot. When accumulating snapshot fact tables were introduced in Chapter 4: Inventory, a product movement pipeline illustrated the concept; order fulfillment workflows were captured in an accumulating snapshot in Chapter 6: Order Management. In this chapter, rather than watching products or orders move through various states, an accumulating snapshot is used to monitor prospective student applicants as they progress through admissions milestones.

The other primary concept discussed in this chapter is the factless fact table. We'll explore several case study illustrations drawn from higher education to further elaborate on these special fact tables and discuss the analysis of events that didn't occur.

Chapter 13 discusses the following concepts:

- Example bus matrix snippet for a university or college
- Applicant tracking and research grant proposals as accumulating snapshot fact tables
- Factless fact table for admission events, course registration facilities management, and student attendance
- Handling of nonexistent events

### University Case Study and Bus Matrix

In this chapter you're working for a university, college, or other type of educational institution. Someone at a higher education client once remarked that running a university is akin to operating all the businesses needed to support a small village. Universities are simultaneously a real estate property management company (residential student housing), restaurant with multiple outlets (dining halls), retailer (bookstore), events management and ticketing agency (athletics and speaker events),

police department (campus security), professional fundraiser (alumni development), consumer financial services company (financial aid), investment firm (endowment management), venture capitalist (research and development), job placement firm (career planning), construction company (buildings and facilities maintenance), and medical services provider (health clinic). In addition to these varied functions, higher education institutions are obviously also focused on attracting high caliber students and talented faculty to create a robust educational environment.

The bus matrix snippet in Figure 13-1 covers several core processes within an educational institution. Traditionally, there has been less focus on revenue and profit in higher education, but with ever-escalating costs and competition, universities and colleges cannot ignore these financial metrics. They want to attract and retain students who align with their academic and other institutional objectives. There's a strong interest in analyzing what students are "buying" in terms of courses each term and the associated academic outcomes. Colleges and universities want to understand many aspects of the student's experience, along with maintaining an ongoing relationship well beyond graduation.

## Accumulating Snapshot Fact Tables

Chapter 4 used an accumulating snapshot fact table to track products identified by serial or lot numbers as they move through various inventory stages in a warehouse. Take a moment to recall the distinguishing characteristics of an accumulating snapshot fact table:

- A single row represents the complete history of a workflow or pipeline instance.
- Multiple dates represent the standard pipeline milestone events.
- The accumulating snapshot facts often included metrics corresponding to each milestone, plus status counts and elapsed durations.
- Each row is revisited and updated whenever the pipeline instance changes; both foreign keys and measured facts may be changed during the fact row updates.

### Applicant Pipeline

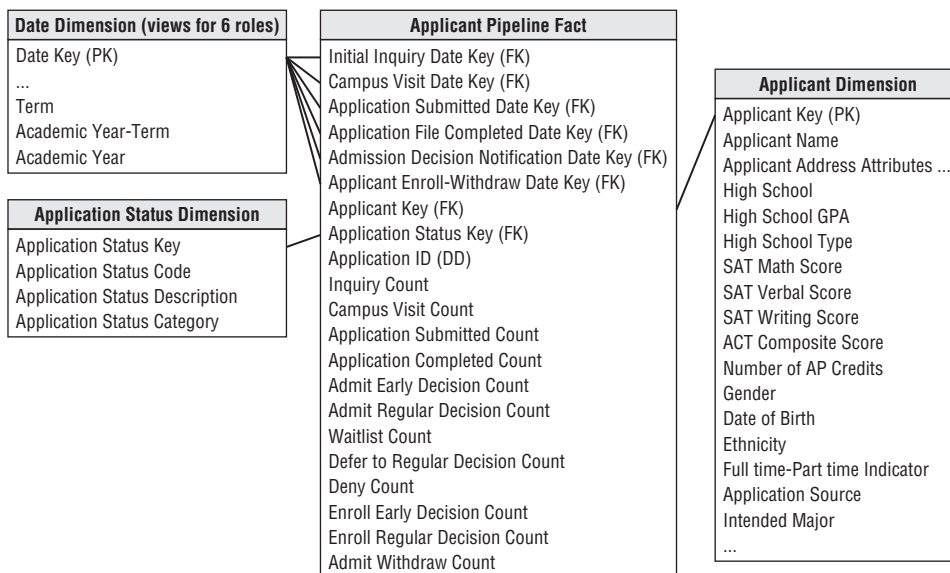
Now envision these same accumulating snapshot characteristics as applied to the prospective student admissions pipeline. For those who work in other industries, there are obvious similarities to tracking job applicants through the hiring process or sales prospects as they are qualified and become customers.

	<i>Date/Term</i>	<i>Applicant-Student-Alum</i>	<i>Employee (Faculty, Staff)</i>	<i>Course</i>	<i>Department</i>	<i>Facility</i>	<i>Account</i>
<b>Student Lifecycle Processes</b>							
Admission Events	X	X	X				
Applicant Pipeline	X	X	X		X		
Financial Aid Awards	X	X	X		X		
Student Enrollment/Profile Snapshot	X	X	X		X	X	
Student Residential Housing	X	X			X	X	
Student Course Registration & Outcomes	X	X	X	X	X	X	
Student Course Instructor Evaluations	X	X	X	X	X	X	
Student Activities	X	X			X		
Career Placement Activities	X	X			X		
Advancement Contacts	X	X	X				
Advancement Pledges & Gifts	X	X	X				X
<b>Financial Processes</b>							
Budgeting	X		X		X		X
Endowment Tracking	X				X		X
GL Transactions	X				X		X
Payroll	X		X		X		X
Procurement	X		X		X		X
<b>Employee Management Processes</b>							
Employee Headcount Snapshot	X		X		X	X	
Employee Hiring & Separations	X		X		X		
Employee Benefits & Compensation	X		X		X		
Staff Performance Management	X		X		X		
Faculty Appointment Management	X		X		X		
Research Proposal Pipeline	X		X		X		
Research Expenditures	X		X		X		X
Faculty Publications	X		X		X		
<b>Administrative Processes</b>							
Facilities Utilization	X		X		X	X	
Energy Consumption & Waste Management	X				X	X	
Work Orders	X	X	X		X	X	X

**Figure 13-1:** Subset of bus matrix rows for educational institution.

In the case of applicant tracking, prospective students progress through a standard set of admissions hurdles or milestones. Perhaps you're interested in tracking activities around key dates, such as initial inquiry, campus visit, application submitted, application file completed, admissions decision notification, and enrolled or withdrawn. At any point in time, admissions and enrollment management analysts are interested in how many applicants are at each stage in the pipeline. The process is much like a funnel, where many inquiries enter the pipeline, but far less progress through to the final stage. Admission personnel also would like to analyze the applicant pool by a variety of characteristics.

The grain of the applicant pipeline accumulating snapshot is one row per prospective student; this granularity represents the lowest level of detail captured when the prospect enters the pipeline. As more information is collected while the prospective student progresses toward application, acceptance, and enrollment, you continue to revisit and update the fact table row, as illustrated in Figure 13-2.



**Figure 13-2:** Student applicant pipeline as an accumulating snapshot.

Like earlier accumulating snapshots, there are multiple dates in the fact table corresponding to the standard milestone events. You want to analyze the prospect's progress by these dates to determine the pace of movement through the pipeline and spot bottlenecks. This is especially important if you see a significant lag involving a candidate whom you're interested in recruiting. Each of these dates is treated as a role-playing dimension, with a default surrogate key to handle the unknown dates for new and in-process rows.

The applicant dimension contains many interesting attributes about prospective students. Analysts are interested in slicing and dicing by applicant characteristics such as geography, incoming credentials (grade point average, college admissions test scores, advanced placement credits, and high school), gender, date of birth, ethnicity, preliminary major, application source, and a multitude of others. Analyzing these characteristics at various stages of the pipeline can help admissions personnel adjust their strategies to encourage more (or fewer) students to proceed to the next mile marker.

The facts in the applicant pipeline fact table include a variety of counts that are closely monitored by admissions personnel. If available, this table could include estimated probabilities that the prospect will apply and subsequently enroll if accepted to predict admission yields.

### *Alternative Applicant Pipeline Schemas*

Accumulating snapshots are appropriate for short-lived processes that have a defined beginning and end, with standard intermediate milestones. This type of fact table enables you to see an updated status and ultimately final disposition of each applicant. However, because accumulating snapshot rows are updated, they do not preserve applicant counts and statuses at critical points in the admissions calendar, such as the early decision notification date. Given the close scrutiny of these numbers, analysts might also want to retain snapshots at several important cut-off dates. Alternatively, you could build an admission transaction fact table with one row per transaction per applicant for counting and period-to-period comparisons.

## Research Grant Proposal Pipeline

The research proposal pipeline is another education-based example of an accumulating snapshot. Faculty and administration are interested in viewing the lifecycle of a grant proposal as it progresses through the pipeline from preliminary proposal to grant approval and award receipt. This would support analysis of the number of outstanding proposals in each stage of the pipeline by faculty, department, research topic area, or research funding source. Likewise, you could see success rates by various attributes. Having this information in a common repository would allow it to be leveraged by a broader university population.

## Factless Fact Tables

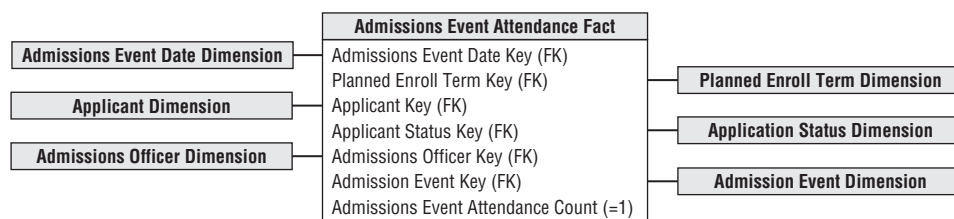
So far we've largely designed fact tables with very similar structures. Each fact table typically has 5 to approximately 20 foreign key columns, followed by one to potentially several dozen numeric, continuously valued, preferably additive facts. The facts can be regarded as measurements taken at the intersection of the dimension

key values. From this perspective, the facts are the justification for the fact table, and the key values are simply administrative structure to identify the facts.

There are, however, a number of business processes whose fact tables are similar to those we've been designing with one major distinction. There are no measured facts! We introduced factless fact tables while discussing promotion events in Chapter 3: Retail Sales, as well as in Chapter 6 to describe sales rep/customer assignments. There are numerous examples of factless events in higher education.

## Admissions Events

You can envision a factless fact table to track each prospective student's attendance at an admission event, such as a high school visit, college fair, alumni interview or campus overnight, as illustrated in Figure 13-3.



**Figure 13-3:** Admission event attendance as a factless fact table.

## Course Registrations

Similarly, you can track student course registrations by term using a factless fact table. The grain would be one row for each registered course by student and term, as illustrated in Figure 13-4.

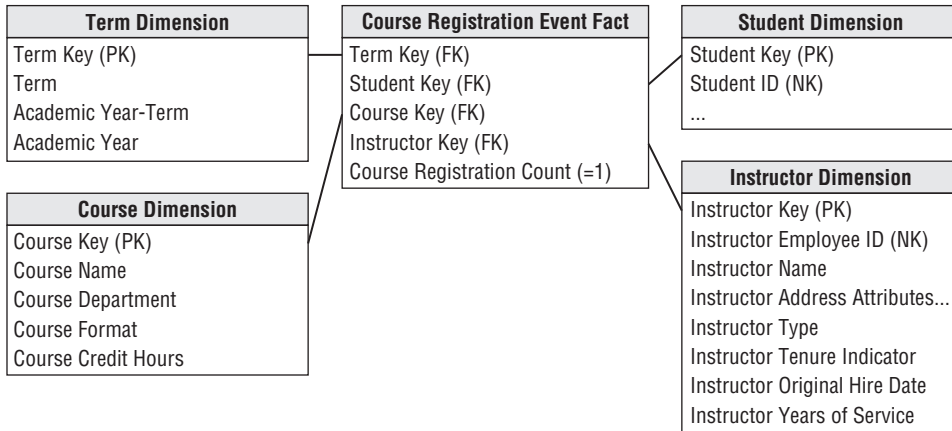
### *Term Dimension*

In this fact table, the data is at the term level rather than at the more typical calendar day, week, or month granularity. The term dimension still should conform to the calendar date dimension. In other words, each date in the daily calendar dimension should identify the term (for example, Fall), term and academic year (for example, Fall 2013), and academic year (for example, 2013-2014). The column labels and values must be identical for the attributes common to both the calendar date and term dimensions.

### *Student Dimension and Change Tracking*

The student dimension is an expanded version of the applicant dimension discussed in the first scenario. You still want to retain some information garnered from the application process (for example, geography, credentials, and intended major) but

supplement it with on-campus information, such as part-time or full-time status, residence, athletic involvement indicator, declared major, and class level status (for example, sophomore).



**Figure 13-4:** Course registration events as a factless fact table.

As discussed in Chapter 5: Procurement, you could imagine placing some of these attributes in a type 4 mini-dimension because factions throughout the university are interested in tracking changes to them, especially for declared major, class level, and graduation attainment. People in administration and academia are keenly interested in academic progress and retention rates by class, school, department, and major. Alternatively, if there's a strong demand to preserve the students' profiles at the time of course registration, plus filter and group by the students' current characteristics, you should consider handling the student information as a slowly changing dimension type 7 with dual student dimension keys in the fact table, as also described in Chapter 5. The surrogate student key would link to a dimension table with type 2 attributes; the student's durable identifier would link to a view of the complete student dimension containing only the current row for each student.

### *Artificial Count Metric*

A fact table represents the robust set of many-to-many relationships among dimensions; it records the collision of dimensions at a point in time and space. This course registration fact table could be queried to answer a number of interesting questions regarding registration for the college's academic offerings, such as which students registered for which courses? How many declared engineering majors are taking an out-of-major finance course? How many students have registered for a given faculty member's courses during the last three years? How many students have registered

for more than one course from a given faculty member? The only peculiarity in these examples is that you don't have a numeric fact tied to this registration data. As such, analyses of this data will be based largely on counts.

**NOTE** Events are modeled as fact tables containing a series of keys, each representing a participating dimension in the event. Event tables sometimes have no variable measurement facts associated with them and hence are called factless fact tables.

The SQL for performing counts in this factless fact is asymmetric because of the absence of any facts. When counting the number of registrations for a faculty member, any key can be used as the argument to the `COUNT` function. For example:

```
select faculty, count(term_key)... group by faculty
```

This gives the simple count of the number of student registrations by faculty, subject to any constraints that may exist in the `WHERE` clause. An oddity of SQL is that you can count any key and still get the same answer because you are counting the number of keys that fly by the query, not their distinct values. You would need to use a `COUNT DISTINCT` if you want to count the unique instances of a key rather than the number of keys encountered.

The inevitable confusion surrounding the SQL statement, although not a serious semantic problem, causes some designers to create an artificial implied fact, perhaps called course registration count (as opposed to “dummy”), that is always populated by the value 1. Although this fact does not add any information to the fact table, it makes the SQL more readable, such as:

```
select faculty, sum(registration_count)... group by faculty
```

At this point the table is no longer strictly factless, but the “1” is nothing more than an artifact. The SQL will be a bit cleaner and more expressive with the registration count. Some BI query tools have an easier time constructing this query with a few simple user gestures. More important, if you build a summarized aggregate table above this fact table, you need a real column to roll up to meaningful aggregate registration counts. And finally, if deploying to an OLAP cube, you typically include an explicit count column (always equal to 1) for complex counts because the dimension join keys are not explicitly revealed in a cube.

If a measurable fact does surface during the design, it can be added to the schema, assuming it is consistent with the grain of student course registrations by term. For example, you could add tuition revenue, earned credit hours, and grade scores to this fact table, but then it's no longer a factless fact table.



### *Multiple Course Instructors*

If courses are taught by a single instructor, you can associate an instructor key to the course registration events, as shown in Figure 13-4. However, if some courses are co-taught, then it is a dimension attribute that takes on multiple values for the fact table's declared grain. You have several options:

- Alter the grain of the fact table to be one row per instructor per course registration per student per term. Although this would address the multiple instructors associated with a course, it's an unnatural granularity that would be extremely prone to overstated registration count errors.
- Add a bridge table with an instructor group key in either the fact table or as an outrigger on the course dimension, as introduced in Chapter 8: Customer Relationship Management. There would be one row in this table for each instructor who teaches courses on his own. In addition, there would be two rows for each instructor team; these rows would associate the same group key with individual instructor keys. The concatenation of the group key and instructor key would uniquely identify each bridge table row. As described in Chapter 10: Financial Services, you could assign a weighting factor to each row in the bridge if the teaching workload allocation is clearly defined. This approach would be susceptible to the potential overstatement issues surrounding the bridge table usage described in Chapter 10.
- Concatenate the instructor names into a single, delimited attribute on the course dimension, as discussed in Chapter 9: Human Resources Management. This option enables users to easily label reports with a single dimension attribute, but it would not support analysis of registration events by instructor characteristics.
- If one of the instructors is identified as the primary instructor, then her instructor key could be handled as a single foreign key in the fact table, joined to a dimension where the attributes were prefaced with "primary" for differentiation.

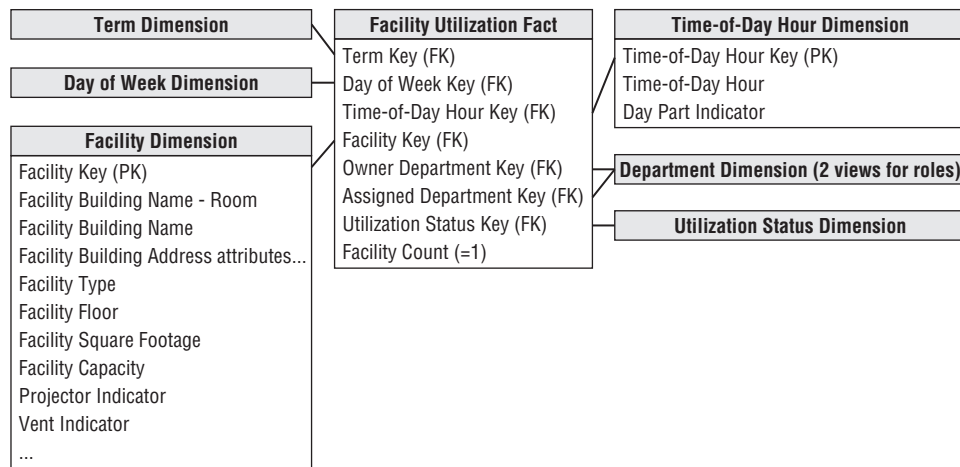
### *Course Registration Periodic Snapshots*

The grain of the fact table illustrated in Figure 13-4 is one row for each registered course by student and term. Some users at the college or university might be interested in periodic snapshots of the course registration events at key academic calendar dates, such as preregistration, start of the term, course drop/add deadline, and end of the term. In this case, the fact table's grain would be one row for each student's registered courses for a term per snapshot date.

## Facility Utilization

The second type of factless fact table deals with coverage, which can be illustrated with a facilities management scenario. Universities invest a tremendous amount of capital in their physical plant and facilities. It would be helpful to understand which facilities were being used for what purpose during every hour of the day during each term. For example, which facilities were used most heavily? What was the average occupancy rate of the facilities as a function of time of day? Does utilization drop off significantly on Fridays when no one wants to attend (or teach) classes?

Again, the factless fact table comes to the rescue. In this case you'd insert one row in the fact table for each facility for standard hourly time blocks during each day of the week during a term regardless of whether the facility is being used. Figure 13-5 illustrates the schema.

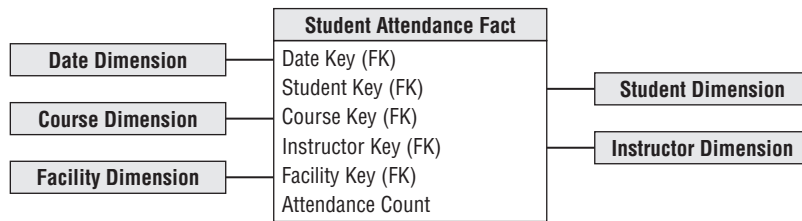


**Figure 13-5:** Facilities utilization as a coverage factless fact table.

The facility dimension would include all types of descriptive attributes about the facility, such as the building, facility type (for example, classroom, lab, or office), square footage, capacity, and amenities (for example, whiteboard or built-in projector). The utilization status dimension would include a text descriptor with values of Available or Utilized. Meanwhile, multiple organizations may be involved in facilities utilization. For example, one organization might own the facility during a time block, but the same or a different organization might be assigned as the facility user.

## Student Attendance

You can visualize a similar schema to track student attendance in a course. In this case, the grain would be one row for each student who walks through the course's classroom door each day. This factless fact table would share a number of the same dimensions discussed with registration events. The primary difference would be the granularity is by calendar date in this schema rather than merely term. This dimensional model, illustrated in Figure 13-6, allows business users to answer questions concerning which courses were the most heavily attended. Which courses suffered the least attendance attrition over the term? Which students attended which courses? Which faculty member taught the most students?



**Figure 13-6:** Student attendance fact table.

### *Explicit Rows for What Didn't Happen*

Perhaps people are interested in monitoring students who were registered for a course but didn't show up. In this example you can envision adding explicit rows to the fact table for attendance events that didn't occur. The fact table would no longer be factless as there is an attendance metric equal to either 1 or 0.

Adding rows is viable in this scenario because the non-attendance events have the same exact dimensionality as the attendance events. Likewise, the fact table won't grow at an alarming rate, presuming (or perhaps hoping) the no-shows are a small percentage of the total students registered for a course. Although this approach is reasonable in this scenario, creating rows for events that didn't happen is ridiculous in many other situations, such as adding rows to a customer's sales transaction for promoted products that weren't purchased by the customer.

### *What Didn't Happen with Multidimensional OLAP*

Multidimensional OLAP databases do an excellent job of helping users understand what didn't happen. When the cube is constructed, multidimensional databases handle the sparsity of the transaction data while minimizing the overhead burden of storing explicit zeroes. As such, at least for fact cubes that are not too sparse, the

event and nonevent data is available for user analysis while reducing some of the complexities just discussed in the relational star schema world.

## More Educational Analytic Opportunities

Many of the business processes described in earlier chapters, such as procurement and human resources, are obviously applicable to the university environment given the desire to better monitor and manage costs. Research grants and alumni contributions are key sources of revenue, in addition to the tuition revenue.

Research grant analysis is often a variation of financial analysis, as discussed in Chapter 7: Accounting, but at a lower level of detail, much like a subledger. The grain would include additional dimensions to further describe the research grant, such as the corporate or governmental funding source, research topic, grant duration, and faculty investigator. There is a strong need to better understand and manage the budgeted and actual spending associated with each research project. The objective is to optimize the spending so a surplus or deficit situation is avoided, and funds are deployed where they will be most productive. Likewise, understanding research spending rolled up by various dimensions is necessary to ensure proper institutional control of such monies.

Better understanding the university's alumni is much like better understanding a customer base, as described in Chapter 8. Obviously, there are many interesting characteristics that would be helpful in maintaining a relationship with your alumni, such as geographic, demographic, employment, interests, and behavioral information, in addition to the data you collected about them as students (for example, affiliations, residential housing, school, major, length of time to graduate, and honors designations). Improved access to a broad range of attributes about the alumni population would allow the institution to better target messages and allocate resources. In addition to alumni contributions, alumni relationships can be leveraged for potential recruiting, job placement, and research opportunities. To this end, a robust CRM operational system should track all the touch points with alumni to capture meaningful data for the DW/BI analytic platform.

## Summary

In this chapter we focused on two primary concepts. First, we looked at the accumulating snapshot fact table to track application or research grant pipelines. Even though the accumulating snapshot is used much less frequently than the more common transaction and periodic snapshot fact tables, it is very useful for tracking the current status of a short-lived process with standard milestones. As we described,

accumulating snapshots are often complemented with transactional or periodic snapshot tables.

Second, we explored several examples of factless fact tables. These fact tables capture the relationship between dimensions in the case of an event or coverage, but are unique in that no measurements are collected to serve as actual facts. We also discussed the handling of situations in which you want to track events that didn't occur.

