Kevin Nolasco

MCIS 510

Milestone 5

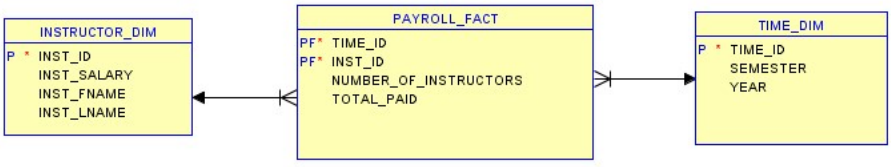
**Star Schema for Data Warehouse**

The purpose of this milestone is to build a data warehouse for this database. To build the data warehouse, we must define the *facts* and *dimensions* for the star schema. A *fact* is a numeric measurement that represents a specific business aspect or activity. A *dimension* is a qualifying characteristic of a fact.

***Building Fact and Dimension Tables***

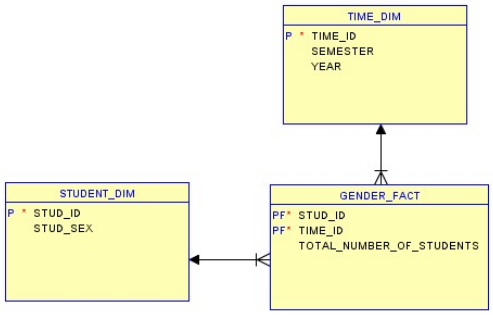
To build fact and dimension tables, we must have questions in mind about our data that we would like answered. In this section, some interesting questions are asked. Also, the fact and dimension tables that correspond to those questions are created and explained.

1) How much does the school spend on instructor salary? This question is important to a school so they can know how much of their budget is being allocated to their instructors. To answer this question, there are a few things to consider. First, it is possible that some instructors are paid the same, but most of the instructors have different salaries. Therefore, it is important to consider the instructor and their salary. Next, the year or semester that this question is asked also determines the total cost. Not every instructor is active every year, so time must be considered in this relationship.

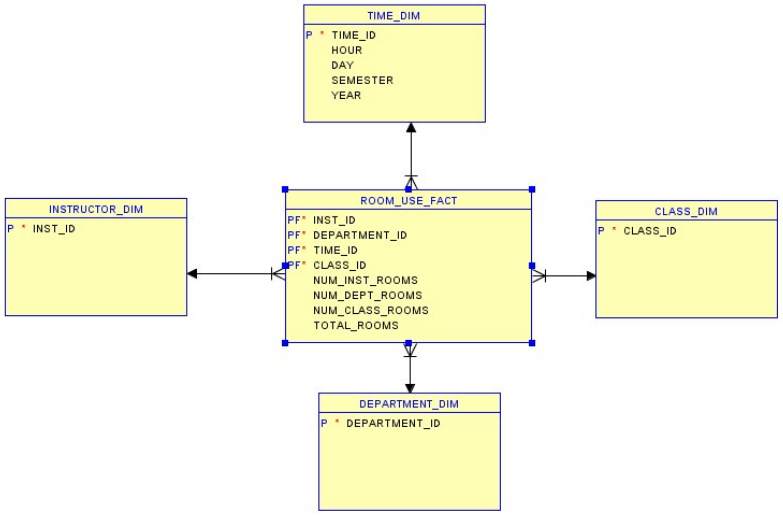



2) What is the gender distribution among students? This question is important because schools should attract an equal number of males and females. If there is a big variation between male

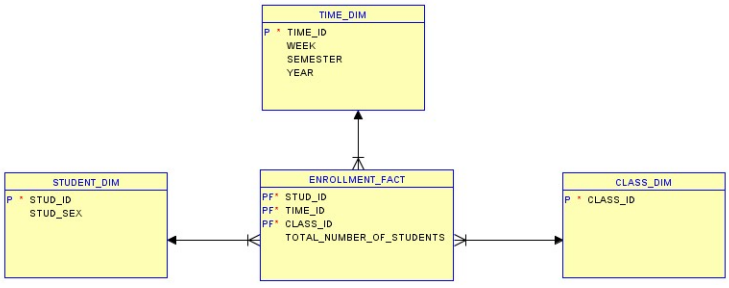
and female, the school should investigate why this is the case. To answer this question, we must have student information. This question varies depending on when it is asked, so time will be a dimension as well. The student dimension table will contain sex as an attribute. This will allow slice and dicing to be performed by student sex.




3) How many rooms are being used? This question is important to determine if the school should expand the number of classrooms available or not. To answer this question, we must have the information for instructor, department, and classes. Each instructor uses a room for their office. Each department is assigned a room for their headquarters. Each class is assigned a room where lectures are held. Also, this question is dependent on time, so time will be a dimension. For this schema, the number of rooms used varies on the semester and year, but also day and hour. Some classes can be held on different days and at different times. Therefore, the hour, day, semester, and year attributes will be included in the time dimension. The fact table will include a breakdown of the number of rooms used for each dimension.

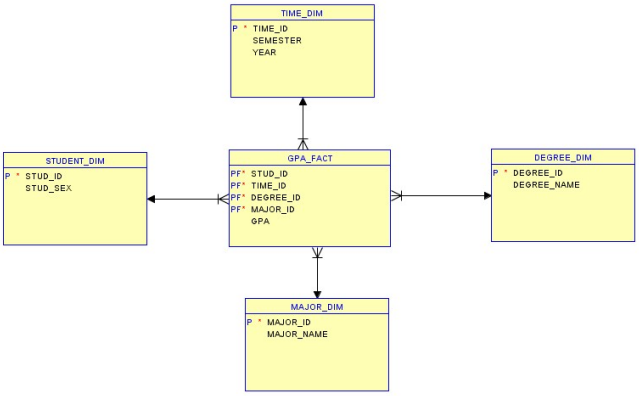



4) What is the class enrollment? This question is important because it can determine whether a class should be offered or not. This question can also be used to forecast or predict future enrollment for some classes. To answer this question, we need to look at all the classes that are offered at the school. We also need to consider the students to see which classes they are enrolled in. Finally, we must consider time. The attributes for the time dimension will be week, semester, year. For this schema, we consider week an attribute of time because some students may drop the class at a certain week. The attribute for student will include sex, so that it is possible to view the amount of male vs female enrollment for classes.

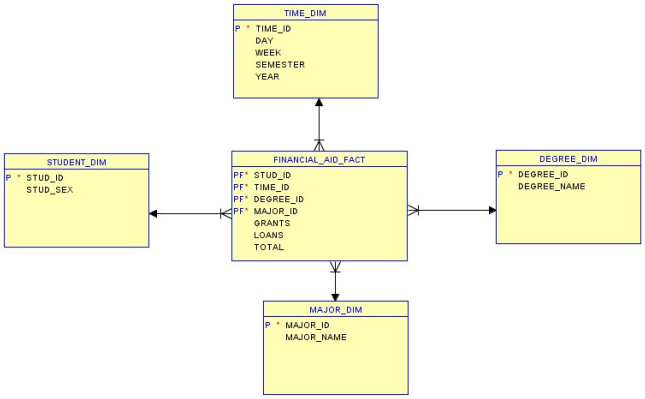



5) What is the student GPA? This question is important because it provides insight into how a student is performing overall. To answer this question, we must consider each individual student. We must also consider time, major, and degree type. There may be some variability in GPA depending on which major or degree type the student is pursuing. Also, the variability may

come from the time that the question is asked. The student dimension will contain a sex attribute, so that GPA by sex can be viewed. The time dimension will contain semester and year attributes. The time dimension does not need to include any more specific time attributes because a student's GPA changes after classes are finished per semester. The major dimension will include a major name attribute. The degree dimension will include a degree name attribute.




6) How much financial aid is offered to students? This question is important because it gives the school an idea of how much of their budget is allocated to student's financial aid. To answer this question, we could consider students, major, degree type, and time. The student dimension will include sex as an attribute. The major and degree dimensions will include major_name and degree_name as their respective attributes. Finally, the time dimension will include day, week, semester, and year attributes. The fact table will include a grant, loan, and total amount.



7) Which instructors produce the best grades? This question is important because it will show the effectiveness of an instructor and can be a measure of their success. To answer this question, we could consider instructor, student, class, and time. The instructor will include INST_SALARY as an attribute. By including instructor salary, it is possible to see if there is a correlation between salary and the student's grade in their class. The student dimension will include sex as an attribute. The class dimension will include number of units as an attribute. The time dimension will include semester and year as its attributes.