

## **Data Warehousing Assignment**

### **Scenario I**

#### **Question 1. What are the strengths and weaknesses of each option?**

Strength:

- 5 courses there will be 5 rows for that student in the fact table.
- Instructor dimension.
- The difficulty is that a few courses (less than 5%) have multiple instructors.
- There will be an Instructor row representing "Manning/Raghavan" (as well as separate rows for Manning and Raghavan, assuming that they sometimes teach courses as sole instructors.
- Each of the two rows will have a value of 0.5 in the Enrollment Count field instead of
- Value of 1, in order to allow the fact to aggregate properly
- Create two fact tables.

Weakness:

- FirstName, Last Name, Title (e.g. Assistant Professor), Department, and Tenured Flag.
- 1 fact table.
- CS 276a.

#### **Question 2. Which option would you choose and why?**

If you consider the 2-fact table it needs less storage compared to the 1 fact table but the considering the execution time one with the 2-fact table will be low compared to the other, so that I will go with the 2 fact.

#### **Question 3. Would your answer to Question 2 be different if the majority of classes had multiple instructors? How about if only one or two classes had multiple instructors?**

If we had a multiple instructor we would choose only 1 fact table and row value would be 1 directly rather than 0.5.

### **Scenario II**

#### **Question 5. What are the strengths and weaknesses of each option?**

Strengths:

- Date, Customer, Account, Security.
- Each customer is placed into one of nine Customer Activity Segments based on their frequency of
- Transactions, average transaction size, and recency of transactions
- The score can be either 1,2,3,4, or 5, with 5 being the most profitable.
- Total commission earned in each quarter of 2003 on trades of IBM stock by customers with a profitability score of 4 or 5
- There is a total of 100,000 customers, and scores are recalculated every three months
- 45 rows.
- The Trades fact table includes a foreign key to the Customer Scores dimension.

**Weaknesses:**

The activity level or profitability level of some customers changes over time, and users are very interested in understanding how and why this occurs.

The scores are attributes of the Customer dimension. When scores change, the old score is over written with the new score.

**Question 6. Which option would you choose and why?**

Slow changing update with the FK is the good option, but if we plan to increase the value from 40 to the larger number that would be great to calculate it a easiest way.

**Question 7. Would your answer to Question 6 be different if the number of customers and/or the time interval between score recalculations was much larger or much smaller?**

We can do without the FK by using 2 fact table and more dimensional table so that we can able to store the data systematically rather than waiting another table with the FK.