# Database Normalization

- Overview
  - Definition of database normalization
  - Why normalize?
  - First normal form
  - Second normal form
  - Third normal form

Information for this presentation borrowed from
http://www.devshed.com/c/a/MySQL/An-Introduction-to-Database-Normalization/

# Database Normalization

- Definition
  - Optimizing table structures
  - Removing duplicate data entries
  - Accomplished by thouroughly investigating the various data types and their relationships with one another
  - Follows a series of normalization “forms” or states

# Database Normalization

- Why Normalize?
  - Improved speed
  - More efficient use of space
  - Increased data integrity
    - (decreased chance that data can get messed up due to maintenance)

# Database Normalization

- A sad, sad database:
  - Refer to the following poor database design:

| student_id | class_name | time | location | professor_id |
|---|---|---|---|---|
| 999-40-9876 | Math 148 | MWF 11:30 | Rm. 432 | prof145 |
| 999-43-0987 | Physics 113 | TR 1:30 | Rm. 12 | prof143 |
| 999-42-9842 | Botany 42 | F 12:45 | Rm. 9 | prof167 |
| 999-41-9832 | Matj 148 | MWF 11:30 | Rm. 432 | prof145 |

  - Problems
    - no need to repeatedly store the class time and professor ID
    - redundancy introduces the possibility for error (Matj 148)

# Database Normalization

- First Normal Form
  - calls for the elimination of repeated *groups* of data by creating separate tables of related data
  - Student information:

| StudentID | StudentName | Major | college | collegeLocation |
|---|---|---|---|---|

  - Class information:

| StudentID | ClassID | ClassName |
|---|---|---|

  - Professor Information:

| ProfessorID | ProfessorName |
|---|---|

# Database Normalization

- Second Normal Form
  - Elimination of redundant data
    - Example data in Class Information:

| studentID | classID | className |
|---|---|---|
| 134-56-7890 | M148 | Math 148 |
| 123-45-7894 | P113 | Physics 113 |
| 534-98-9009 | H151 | History 151 |
| 134-56-7890 | H151 | History 151 |

Use:

| ClassID | ClassName |
|---|---|
| M148 | Math 148 |
| P113 | Physics 113 |
| H151 | History 151 |

To get Class Information:

| studentID | classID |
|---|---|
| 134-56-7890 | M148 |
| 123-45-7894 | P113 |
| 534-98-9009 | H151 |
| 134-56-7890 | H151 |

# Database Normalization

- Third Normal Form
  - eliminate all attributes(column headers) from a table that are not directly dependent upon the primary key
    - college and collegeLocation attributes are less dependent upon the studentID than they are on the major attribute
    - New college table:

| major | college | collegeLocation |
|---|---|---|

    - Revised student table:

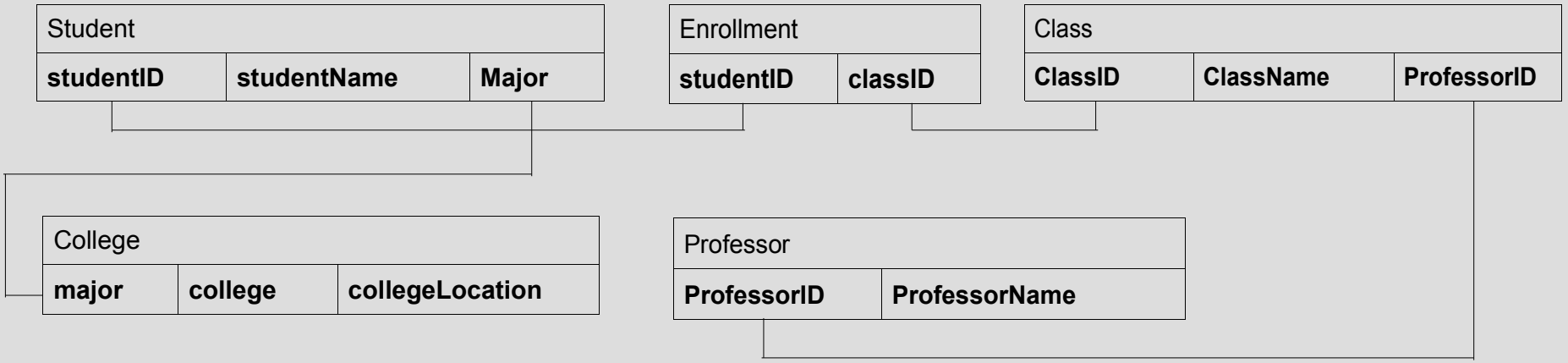
| studentID | studentName | Major |
|---|---|---|

# Database Normalization

- Old Design:

| student_id | class_name | time | location | professor_id |
|---|---|---|---|---|
| 999-40-9876 | Math 148 | MWF 11:30 | Rm. 432 | prof145 |
| 999-43-0987 | Physics 113 | TR 1:30 | Rm. 12 | prof143 |
| 999-42-9842 | Botany 42 | F 12:45 | Rm. 9 | prof167 |
| 999-41-9832 | Matj 148 | MWF 11:30 | Rm. 432 | prof145 |

New Design:



# Database Normalization

- Questions?

# Normalization Assignment

- For your tool:
  - Compile list of all data items used
  - Place all data into one table
  - Complete 1NF and describe reasons why this is better (or if data is already in 1NF continue)
  - Complete 2NF in the same manner
  - Complete 3NF in the same manner