Diana Wang

CMSC 23500

Homework 1

This database contains the medalists of all summer Olympic Games.

I found the dataset through this site: http://www.theguardian.com/sport/datablog/2012/jun/25/olympic-medal-winner-list-data#data

The original data set was created by the IOC Research and Reference Service. I reformatted the original .xls file into 8 separate .csv files to .import into the tables.

While I was designing my database, I started from three different objects: the Athletes, each of the Olympic Games and the different Events. Certain cities have hosted the Olympics more than once, so Games and Cities had a many to one relationship. Events had a many to one relationship with its Discipline which had a many to one relationship with its Sport. The relationship between Games and Events was many to many, so I created a join table named Awards (which additionally contained medal type). Finally, since athletes can win many awards and one awards can be awarded to many people (team events), I created another join table called Winners for Athletes and Awards. Originally, in the diagram I sketched out, I had the Awards join table join Games, Event, and Athlete. I wasn’t sure if I could create a join table that joined more than 2 other tables, so I ended up deciding to follow the structure outlined above.

I assumed that different medalists of the same gender cannot have the same name, because I could not differentiate them from the information given in the dataset. I deleted one entry which looked like a duplicate, but theoretically there could have been two people competing in the same event, on the same team, in the same game.