

ACS 1803 DATA MODELING ILLUSTRATION

Suppose we are to design a new database for a college information system that will use the “raw material” stored in tables on disk so as to produce a variety of management information reports that will be useful to different managers.

You are a data analyst on a team that will produce a design for the new database (i.e., what fields will be in what table and where will be the common columns to link the data in the tables).

To start, you want to find out what information the managers will need from this database. This could be the information they are getting now, with the current (old) system or information that they would like to get that they are not getting now.

Manager 1 tells you she needs:

1. A report showing : StNumber, StName, Street, City for all student in the college, sorted by student number.
2. A report showing: StName, StPhone, StMajor ordered by StName

Manager 2 needs:

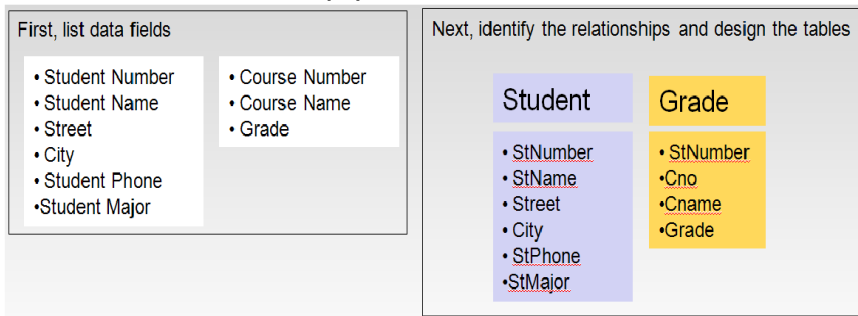
1. StName and then CName and Grade for all courses the student has taken last term
2. Major and then StName and StPhone for all students in the Major, ordered by Major and then by StName

Manager 3 needs:

1. CName, CNo and then StName and Grade for each student that took the course last term orderd by CName and then by StName.
2. CName, CNo for each course offered in the term

How will the team of analysts design a relational database such that each manager can get the computer to produce the reports he / she needs when the computer takes its raw data from the same, one, centralized database?

First, the analyst asks: “what data fields will be needed by anyone?”
He lists them on a piece of paper as below:



The designer then *analyzes the relationships between these fields* and decides on i) which fields will go in which tables and ii) where will be the common columns of data to link data between 2 tables?

Since the number of fields on his paper is small, the analyst can readily decide that

A STUDENTS table will have the following fields:

StNumber, StName, Street, City, StPhone, StMajor

(these fields all have a 1-1 relationship with StNumber)

A GRADES table will have

StNumber, CNo, CName, Grade

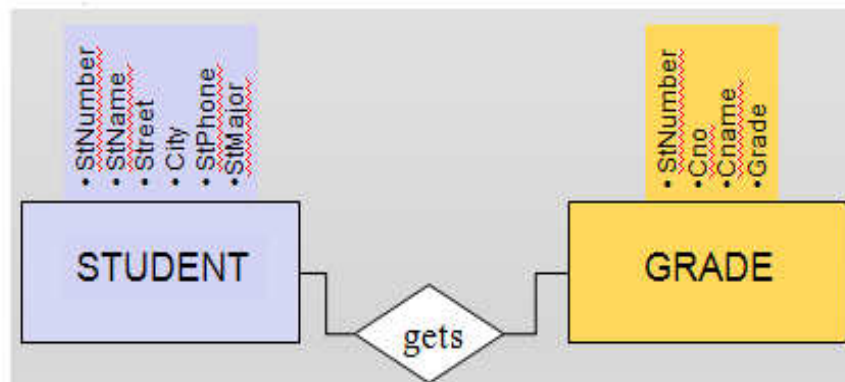
(these fields have a 1-many relationship with StNumber)

StNumber is the common column linking data in STUDENT table with data in the GRADES in the database on disk. The analyst will call the database itself (the container) by the name SCHOOL.

Since the fields on the "piece of paper" were small in number and clear to understand, it was not hard for the team of analysts to come up with the design above "in their heads".

But they could have done this more formally by drawing an entity-relationship diagram after they listed their fields on the paper.

The diagram might look as below:



From this diagram the analysts would extract their database design (ie. What tables should we have and what columns (fields) will be in what tables, on disk).

However, say this team now moves to a business such as a womens' clothing enterprise. They are to design a database. They talk to 123 managers and get a total of 437 reports that these managers will need.

They can still list the fields from all these reports on paper, but it is clear that they will soon be very lost!

So, they must find BASIC UNITS OF INFORMATION (entities) and then attributes of these entities, which will be the fields on the paper.

Then, they MUST draw an entity-relationship diagram [data model] for this situation. It may have 259 entities and then each entity will have several attributes. Entities will have relationships identified between them.

Once they decide upon their ERD, they will have to follow "rules of normalization" to "trim down" this ERD to "proper" form. This proper ERD will be such that *every entity [rectangle] will be a table in their database and every attribute of an entity will be a column in that table.*

The designers might come up with a database [container] with over 250 tables. They will then have to use a relational database management system (maybe Access or Oracle) to set this database up for them on disk and to fill each of the 250+ tables with current data.

Thus, a DATA MODEL is a diagram identifying major ENTITIES of data that we need to represent in the database and their ATTRIBUTES as well as RELATIONSHIPS between the entities that exist in the organization.

It is a diagram drawn after data analysts have talked with all the managers to see what kind of information they will need from the new, centralized database. This diagram, after several additional steps, will help the analysts to decide what tables they will need in the database and what columns they will need to have in each table.