

Course on Database Design

Carlo Batini

University of Milano Bicocca, Italy

Course on Database Design

The course is made of six parts:

Part 0 – What you will learn in this course

Part 1 - Introduction to Databases

Part 2 - Entity Relationship Model

Part 3 - Relational Model

Part 4 - Conceptual Data Base Design

Part 5 - Logical Data Base Design



© Carlo Batini, 2015

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Course on Database Design

Part 0 – What you will learn in this course

**In the appendix: The data management
learning project**



© Carlo Batini, 2015

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Part 0 – What you will learn in the Course on Database Design

In this course, you will achieve expertise in two topics, which I would like to introduce by means of an example. Assume that you are a student in a Chinese University and you want for some reason to organize information of interest for you and your colleagues on courses passed. Such information is represented in Figure 1.

Assume we are in a Chinese University (where grades are between 0 and 100). You want to represent:

- three exams passed by a student with Student Id = 12127 referring to:
 1. Course with Course Id = 37 and exam passed with grade 80.
 2. Course with Course Id = 29 and exam passed with grade 90.
 3. Course with Course Id 18 and exam passed with grade 85.
- one exam passed by a student with Student Id = 23344 referring to:
 1. Course with Course Id = 29 and Exam passed with grade 95.

Figure 1: Data that you may need to organize.....

A first competence you will achieve concerns database models, that provide you modeling constructs to organize data. In software systems for databases, called Data Base Management Systems, data are organized in tables (also called relations). Figure 2 shows you how the above data are organized in the relational model of data adopted in Data Base Management Systems.

Student

Student Id	Course Id	Grade
12127	37	80
12127	29	95
12127	18	85
23344	29	95

Figure 2: The same data in the relational model

So, you will be able to use data models for your needs, transforming requirements as shown in Figure 1 in terms of tables as illustrated in Figure 2, through a process called modeling, see Figure 3.

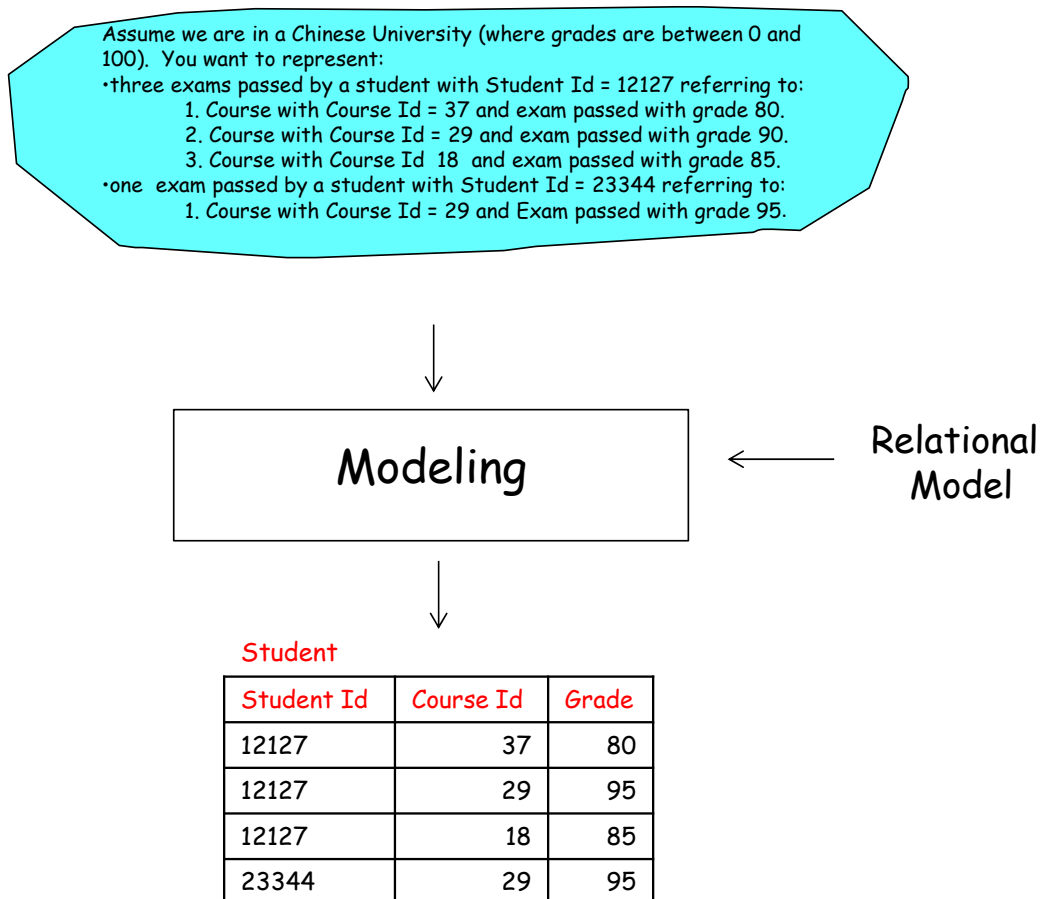


Figure 3: The modeling issue

Look with more attention at the table of Figure 3. We can identify (see Figure 4) two parts with completely different roles, that are called:

1. the *schema*, which represents *concepts*, such as Grade, and
2. *data instances*, that represent *values of concepts*, e.g. for grades, 80 and 95.

We can also represent the schema of the relation in the following way:

Student (Student Id, Course Id, Grade).

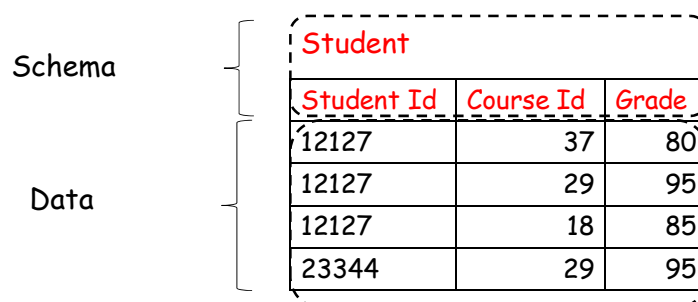


Figure 4: Schema and data instances

Frequently it happens that requirements are not expressed in terms of long lists of data, but, instead, refer to concepts, such as the following requirements in Figure 5.

Students of a University attend courses, and pass them. Courses have a last name, a first name and a year of enrollment. Students have an ID, a Last Name and First Name. For Chinese Students we want to represent City of Birth and Region of Birth. For Foreign Students we want to represent the Country of Birth and the Continent of Birth. A Course passed by a Student has a grade and a date. Courses are taught by professors. Professors have a last name and a first name. They may be Associate Professors or Full professors, only for Full Professors we are interested to the City of Birth, with name and region. Every professor belongs to a Department. Departments have a Name and an address.

Figure 5: Requirements of a database application

In this course, you will achieve a second important skill; you will be able to *design* a database, that is to transform requirements into a database schema, as illustrated in the following Figure 6; the relational database schema, made of eleven relation schemas, is represented using the notation introduced above.

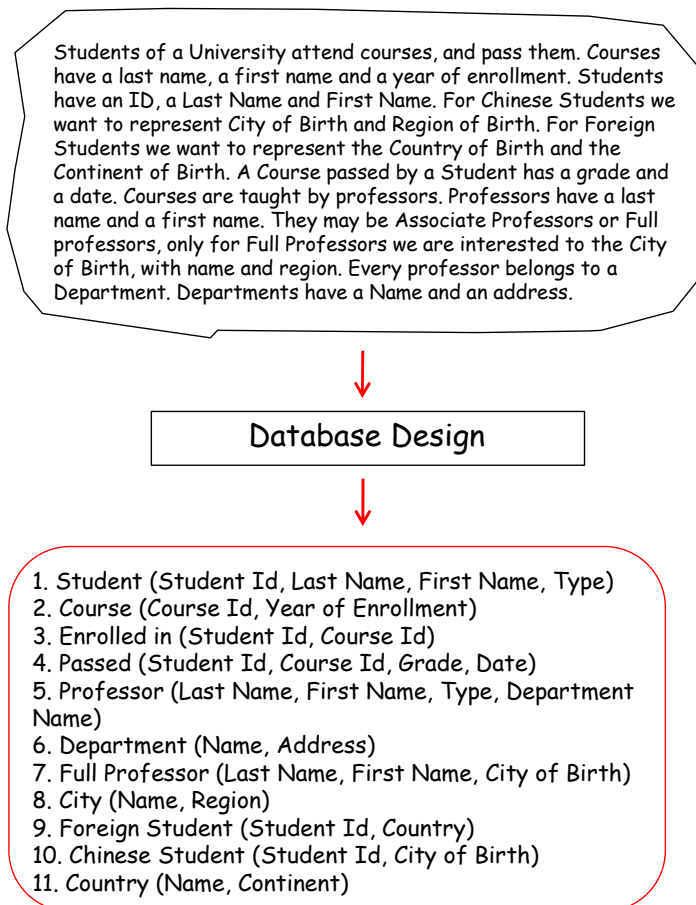


Figure 6: Design of requirements in terms of a relational schema

The relational model, adopted in Database Management Systems, is a bit far from the way human beings mentally model a reality of interest; so, we introduce two models for the design process, the *Entity Relationship model* for the first phase called *conceptual design*, and the *relational model* for the second phase called *logical design*, see Figure 7.

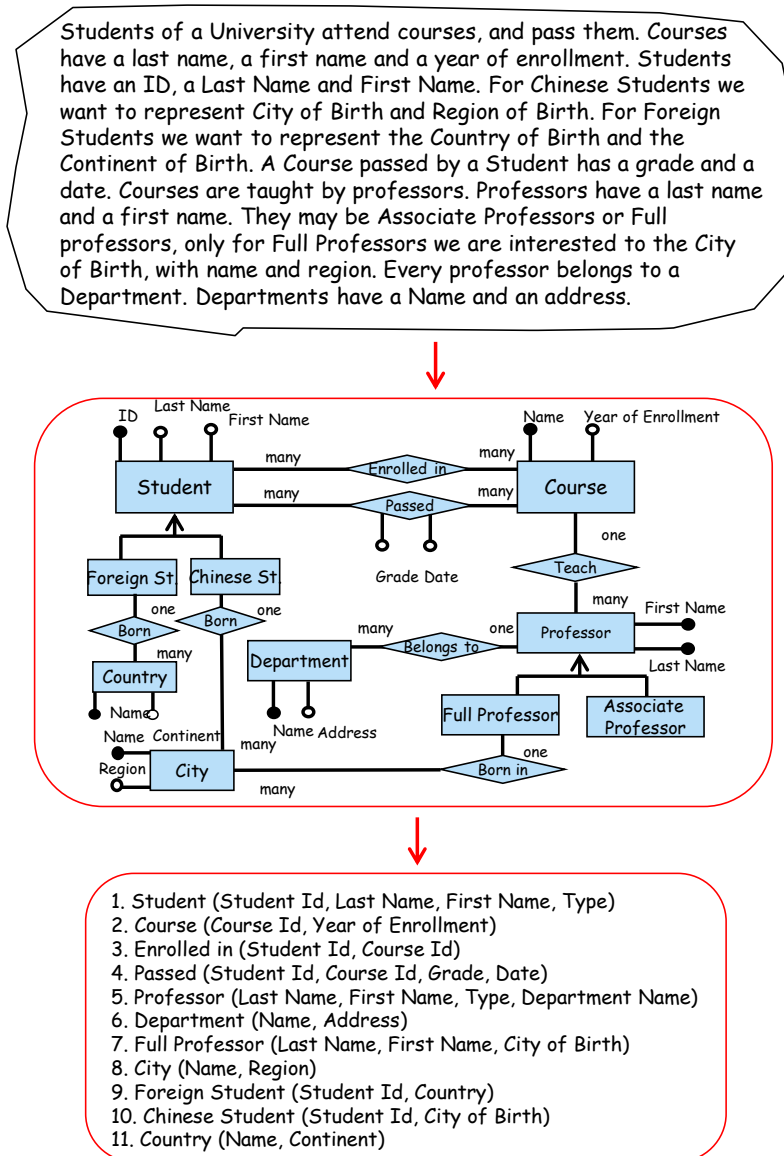


Figure 7 - Two steps design to ease the design activity

The Entity Relationship model is nice, as it has an easy to understand diagrammatic representation of concepts. For instance, the box Student represents the set of students enrolled in the University, and Id, Last Name and First Name are the properties to which data values will be assigned.

Learning resources

1.Web Video MOOC Course

Soon accessible

2.This book

Transcription of lessons of the Web MOOC Course

Course on Database Design

Carlo Batini

University of Milano Bicocca, Italy



© Carlo Batini, 2015

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

This text is a transcription of the video lessons. It can be useful for those of you that, after watching at the video, need for a moment of further in depth analysis of the concepts presented. It is also useful for those of you that do not possess entirely spoken English language, and need for a written version of lessons to be examined comfortably. It can also be used independently form the video version. Enjoy!

3. Map of concepts, reproduced at the beginning of every part

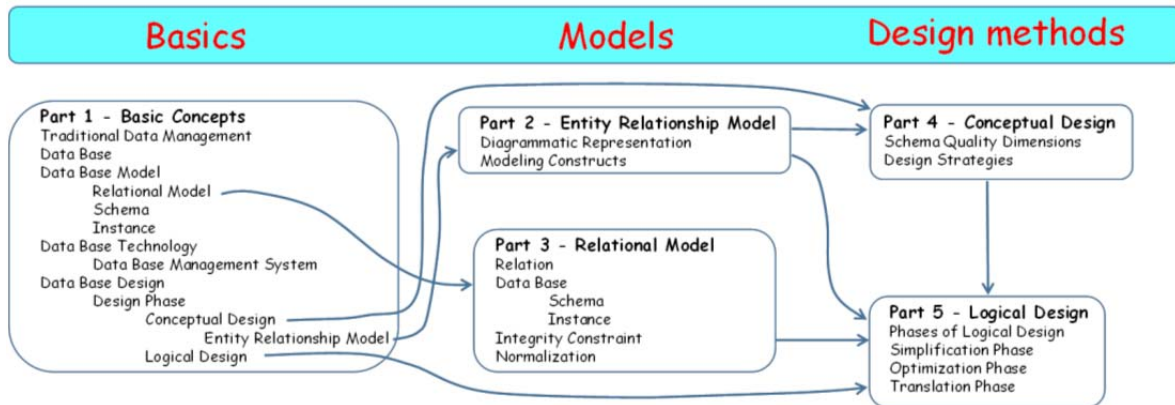


Figure 8: High level map of concepts

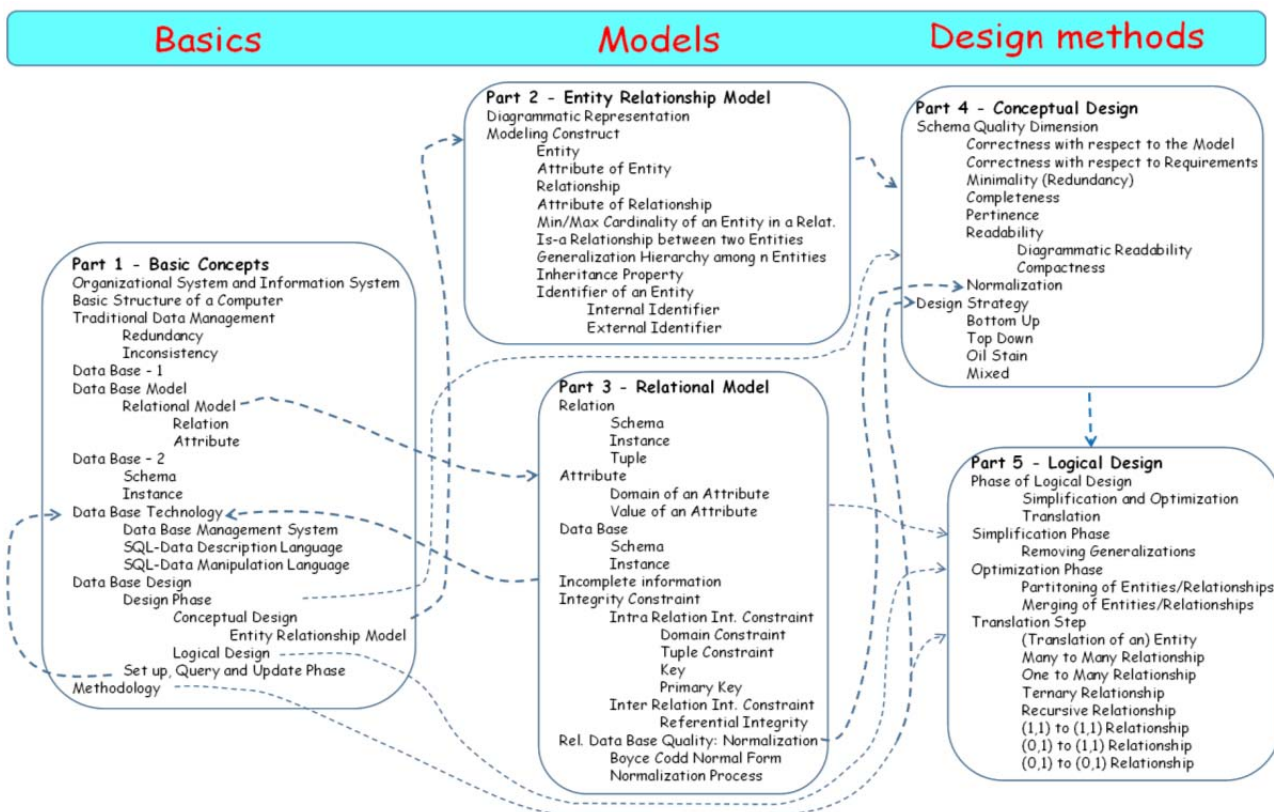


Figure 9: Low-level map of concepts

4. Teaching & Learning method

In this course teaching & learning are not seen as separate, rather as two faces of the same coin.

I always start from examples, students are invited to solve a problem, or understand hidden or tacit features of the example, so that concepts are learnt with a free, autonomous and unconstrained learning process. Let us see an example!

Part 2 – Lesson 2 – Entity and Attribute of entity

Entity

Let us start with an example.

Exercise 2.1 - Look at these pictures. Now, how do you refer to these pictures?

1. This is a set of persons
2. This is a set of students
3. This is set of Chinese students.

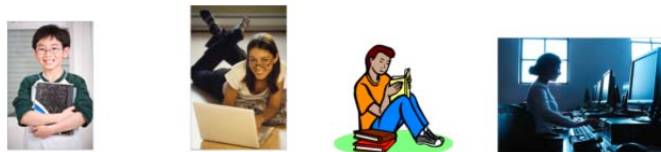


Figure 10: A set of pictures

The student is invited to think autonomously, or in collaboration with other students. After that, a discussion is proposed (in the book, in the next page, in the video in the next video frame). In this case:

Discussion on Exercise 2.1

Probably your answer is the second one. The first answer is too generic, while the third answer is wrong, since some of the pictures do not show Chinese people. We say also that this set of pictures represents a class, and we say that such class is an *entity* in the Entity Relationship model. We can assign to the entity the name **Student** (see next Figure). More precisely.

Definition - An *entity* is a class of real world objects that:

1. Are of interest for a database application.
2. Have common properties.
3. Have autonomous existence

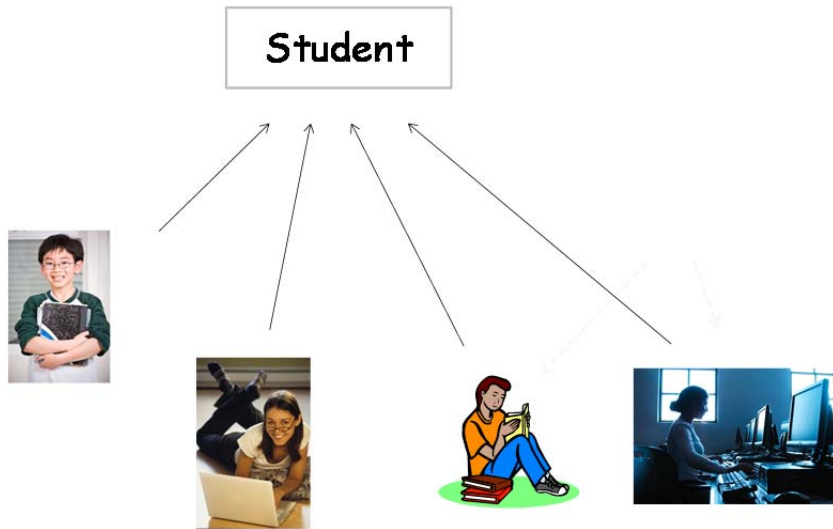


Figure 11: The four student instances and the entity Student

5. Atzeni's et al. book and solutions to exercises in Atzeni's et al. book

Paolo Atzeni, Stefano Ceri, Stefano Paraboschi and Riccardo Torlone
Database Systems - Concepts, Languages and Architectures
McGraw-Hill Book Company 1999, ISBN 0-07-709500-6
Freely available at <http://dbbook.dia.uniroma3.it/>

For

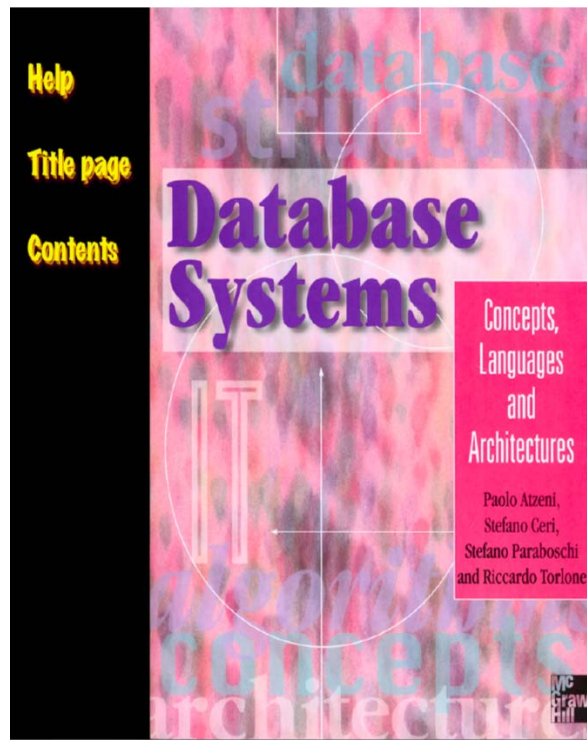
Part 1 - Introduction to Databases

Part 2 - Entity Relationship Model

Part 3 - Relational Model

Part 4 - Conceptual Data Base Design

Part 5 - Logical Data Base Design



6. Batini's et al. book

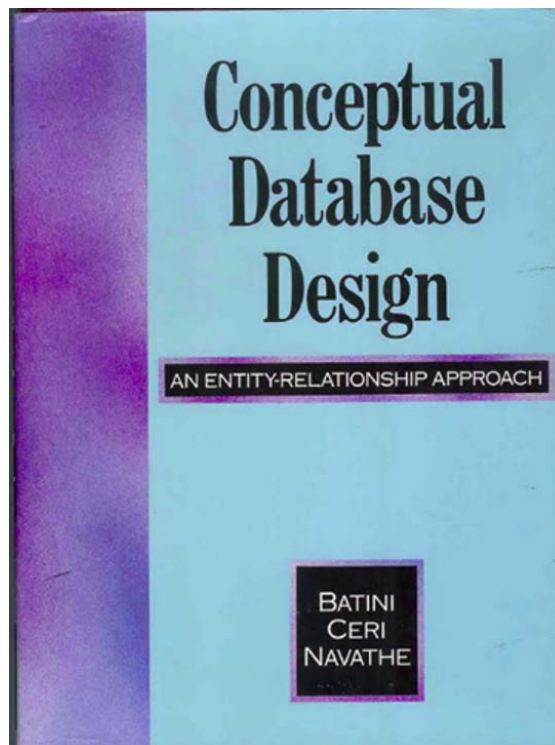
Carlo Batini, Stefano Ceri, Sham B. Navathe - Conceptual Database Design: an Entity Relationship Approach - ISBN-10: 0805302441 • ISBN-13: 9780805302448 ©1992, Addison-Wesley, 470 pp, available in the e-market

For

Part 2 - Entity Relationship Model

Part 4 - Conceptual Data Base Design

Part 5 - Logical Data Base Design



Appendix: The Data Management Learning Project

The course on Database Design is part of a larger, more ambitious project on Data Management. In 1975 M. Porat produced a figure on the dynamics of jobs in the last 2 millions years and in subsequent 75 years, as long as 2050. Such forecast, that has revealed correct so far, says that in the next 30 years the percentage of jobs in informative services, namely services based on information will increase up to 60% and more.

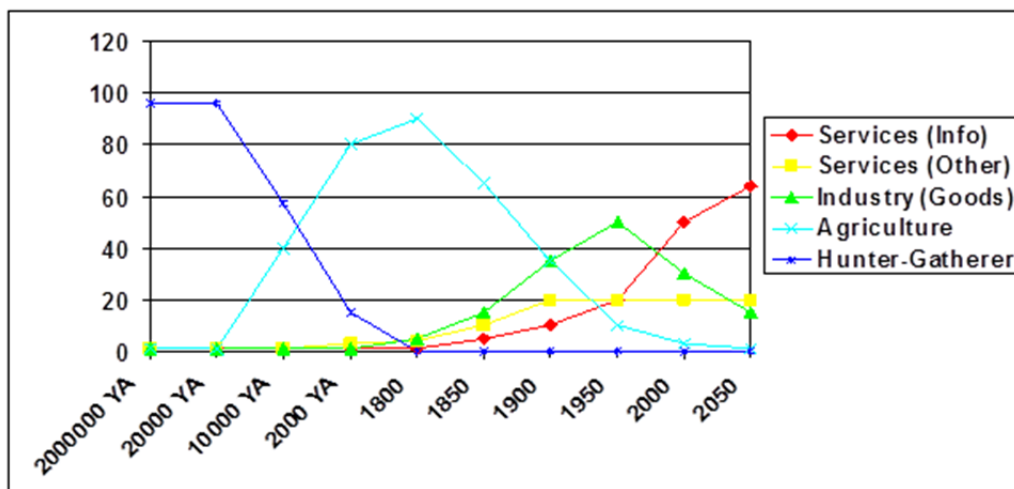


Figure 12: The figure from Porat (M. Porat, The information Economy: Definition and Measurement, Office of Telecommunications, Washington DC, 1977)

This is not surprising, if we consider that every year and a half the amount of information exchanged in the Web doubles. The term “Big Data” has been coined to describe the phenomenon of large quantities of data that, due to their volume, velocity and heterogeneity exceed the processing capacity of conventional data management systems and need for innovative techniques and alternative ways to process it.

At the same time, all over the world the MOOC phenomenon is rapidly growing. MOOC means Massively Open On line Courses, that is courses delivered in the Web and accessible by a wide audience of students and professionals.

This course on Database Design is the first milestone of a wider and more ambitious learning project on Data Management. The building blocks of the Data Management learning project are:

1. Data Management – basics

- a. Data Base Design
- b. Data Base Querying

2. Data Management – advances

- a. Data and Text Mining
- b. Data Analytics
- c. Business Intelligence
- d. Data Visualization
- e. Data and Information Quality & Value
- f. Semantic Web and Linked Data
- g. Open source SW platforms and languages for data analytics.
- h. Techniques and Tools for Data Management

3. Data Management – domain specific

Examples

- a. Business Intelligence in Finance
- b. Big data in Physics
- c. Open Data

Both the **Data Management – advances** and the **Data Management – domain specific** courses have to be seen as open-ended lists of topics.



© Carlo Batini, 2015

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.