# Data structures and database systems concepts

## Prof. Marco Cremaschi

***COURSE AIMS AND INTENDED LEARNING OUTCOMES***

The course aims to acquire the knowledge for **modelling, designing, implementing and querying** a database.

The course will provide an updated overview of the main **storage systems** (e.g., Relational, NoSQL and RDF) and the currently available **architectures** (including commercial ones).

**Expected learning outcomes**

*Knowledge and understanding*

At the end of the course, the student will **know**:

– the concept of **information** and **data**;

– the concept of **data model** and **data structure**;

– the concept of **schema** and **instance**;

– the database design methodologies;

– the main DBMSs (e.g., MySQL/MariaDB, MongoDB, Neo4J, Virtuoso)

– the main query languages (e.g., SQL, MQL, Cypher, SPARQL);

– the data indexing methodologies;

– the creation of semantic databases;

– the use of Python libraries for querying databases (e.g., PyMongo, Py2neo).

*Ability to apply knowledge and understanding*

At the end of the course, the student will **be able to**:

– model real-world scenario data;

– design a database;

– select the technologies for the creation of a database concerning the types of data to be persisted;

– writing instructions to query databases.

“Making judgments”, “Communication skills” and “Learning skills”

The student will be able to identify and formally model a real scenario's entities and related relationships. The student will also manage data through a DBMS.

***COURSE CONTENT***

The course program is structured as follows:

Introduction to Databases and DBMSs

General concepts

Information and data

Schema and instance

Relational model

[MySQL](https://www.mysql.com/it/)/[MariaDB](https://mariadb.org/) and SQL

Document model

[MongoDB](https://www.mongodb.com/it-it) and MQL

Graph model

[Neo4J](https://neo4j.com/) and [Cypher](https://neo4j.com/developer/cypher/)

Knowledge Graph

[RDF](https://www.w3.org/RDF/) and [SPARQL](https://www.w3.org/TR/rdf-sparql-query/)

Data indexing

[Elasticsearch](https://www.elastic.co/)

***READING LIST***

P.Atzeni Ceri, Paraboschi, Torlone, Basi di Dati – Modelli e linguaggi di interrogazione 4/ED, McGraw-Hill, 2013.

Other readings will be discussed in the first lesson

***TEACHING METHOD***

The course includes **classroom** and **remote** lessons with **slides** for the theoretical aspects and **Jupyter notebooks** for the exercises.

The exercises and other practical activities will be performed individually and in groups.

Experts will organise seminars during the course to better understand the theoretical concepts.

***ASSESSMENT METHOD AND CRITERIA***

*Evaluation method*

Students will work on a small project as part of a group or individually based on a given brief. The project will be evaluated, and students must achieve a positive score to access the written text. The test will consist of four open-ended questions on topics discussed in the lessons, exercises, and bibliography.

The final evaluation will be obtained by taking the average of the project evaluation and the test evaluation, resulting in a single evaluation.

*Evaluation criteria*

As part of their project, students are expected to showcase their skills in creating a database that reflects real-world scenarios and models entities and relationships. This project will require them to apply the techniques learned in class.

The test will evaluate the student's understanding of the theoretical aspects of database creation and management. The relevance of the answers and the appropriate use of terminology will contribute to the overall evaluation.

***NOTES AND PREREQUISITES***

The teaching material produced during the course will be available online.

The teacher will provide more detailed information on the course, bibliographic and exam methods in the **first lesson**.

*Prerequisites*

The teaching does not require pre-requisites related to the contents. However, it assumes interest and intellectual curiosity for the topics of the course.

*Time and place of reception*

To be defined at the beginning of the course.