

# 14D003

## Data Warehousing and Business Intelligence

### Overview and Objectives

The digital revolution implies abundant and rich data easily available as never before. As the volume of data grows, extracting information gets more and more challenging and the only way to deal with this complexity is through computing and expertise in programming environments. The recent explosion in data availability sometimes even challenges the classic technological and analytical paradigm and calls for a new approach to data treatment. On the other hand, helping decision-making with quantitative insight requires being able to transform data into useful information, and presenting the most relevant results to a non-technical audience. This course deals with the basics of database management systems, the relational model, and the classical environment of data storage and retrieval. The course will cover practical issues in Business Intelligence and it will provide the tools to manipulate, cleanse and transform data sets in a relational database environment. Students will also be provided with the fundamentals of non-relational databases.

### Course Outline

#### Introduction to Data Base Management Systems

- The relational model: properties and rules
- Introduction to relational algebra
- The Structured Query Language: syntax and semantics
- Query optimization methods

#### Limitations of relational databases model

- NoSQL databases.
- Basics of distributed processing and the Cloudera big data environment.
- Introduction to Hadoop storage and computing ecosystem
- Introduction to Spark core engine

#### Business intelligence

- The dimensional model
- Stored routines and OLAP analysis
- Reporting & ad hoc queries

14D003

## Data Warehousing and Business Intelligence

### Required Activities

Attendance at classes, and submission of homeworks.

### Evaluation

Homework during the course (40%)

Final exam at the end of the course (60%)

### Materials

Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer D. Widom, *Database Systems: The Complete Book*, Prentice Hall.

J.D. Ullman and A. Rajaraman, *Mining Massive Datasets*, (Cambridge University Press, UK 2012

Ralph Kimball and Margy Ross, *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*, 2nd Edition, Wiley Computer Publishing, 2002