



## Data science and database technology

### Introduction to the course

# Transaction processing

## ➤ On Line Transaction Processing (OLTP)

- Traditional DBMS usage

## ➤ Characterized by

- snapshot of current data values
- detailed data, relational representation
- structured, repetitive operations
- read/write access to few records
- short transactions
- isolation, reliability, and integrity are critical (ACID)
- database size  $\approx$  100MB-GB

# Analytical processing

## ➤ On Line Analytical Processing (OLAP)

- Decision support applications

## ➤ Characterized by

- "historical" data
- consolidated, integrated data
- ad hoc applications
- read access to millions of records
- complex queries
- consistency before and after periodical loads
- database size  $\approx$  100GB-TB

# Course content

## ➤ First part (weeks 1-7)

- Data warehouse design
- OLAP analysis
- Data science and data mining

## ➤ Second part (weeks 8-14)

- DBMS server technology
- SQL Triggers
- Distributed databases

## ➤ Course books

- Golfarelli, Rizzi, *Data warehouse: teoria e pratica della progettazione*, McGraw-Hill, 2006
- Tan, Steinbach, Kumar, *Introduction to data mining*, Pearson, 2006
- Atzeni, Ceri, Fraternali, Paraboschi, Torlone, *Basi di dati – Architetture e linee di evoluzione*, McGraw Hill, 2007
- Atzeni, Ceri, Paraboschi, Torlone, *Database systems*, McGraw Hill, 1999

## ➤ Other books

- Ramakrishnan, Gehrke, *Database Management Systems*, McGraw-Hill, 2004
- Kimball e altri, *several books and white papers on data warehouse design methodologies and case studies*, Wiley
- Han, Kamber, *Data mining: concepts and techniques*, Morgan Kaufmann, 2006