Data Management for Big Data (DMBD)

and

Advanced Database Systems for Big Data (ADBS4BD)

Dario Della Monica

March 29, 2021

Teacher's information

Dario Della Monica

Dipartimento di Scienze Matematiche, Informatiche e Fisiche (DMIF)
University of Udine

- **email:** `dario.dellamonica@uniud.it`

- **web:** `http://users.dimi.uniud.it/~dario.dellamonica/`

    - [Home] ≫ Teaching ≫ Advanced Database Systems For Big Data
    - [Home] ≫ Teaching ≫ Data Management for Big Data

- **office hours:** Tuesday, 16:30-18:30 (or by appointment)

    - appointment is always preferable

- more info (**cell**, **skype**, ...) on my webpage

Course structure

These lectures are part of 2 courses

- Advanced Database Systems for Big Data (ADBS4BD)
    - Master Degree in Artificial Intelligence & Cybersecurity, University of Udine
    - Master Degree in Computer Science, University of Udine

- Data Management for Big Data (DMBD)
    - Master Degree in Data Science and Scientific Computing
      University of Udine and University of Trieste, Italy

Overview of the lectures

- Part 0: Fundamentals of database systems (24h)
    - already delivered by Andrea Brunello
    - only for DMBD course

- Part 1: Advanced database models, languages, and systems ($\sim$24h);

- Part 2: Data analysis and big data ($\sim$24h)
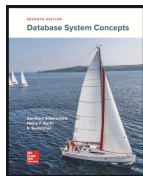
Overview of part 1
Advanced database models, languages, and systems

- ▶ Query processing and optimization

  - ▶ Query processing
    - ▶ algorithms for evaluating operations (e.g., selection, join)

  - ▶ Cost-based optimization and heuristics
    - ▶ choice of "best" evaluation plan

- ▶ Distributed database architectures

  - ▶ An introduction to distributed DBMS
  - ▶ Distributed database design (fragmentation and allocation)
  - ▶ Distributed query processing and optimization
  - ▶ Transaction management in distributed databases
    - ▶ distributed concurrency control (two-phase locking – 2PL)
    - ▶ distributed DBMS reliability (two-phase commit – 2PC)

- ▶ Semistructured Data and XML

  - ▶ Definition of semistructured data in XML
  - ▶ Querying XML data (XPath and XQuery)
  - ▶ XML and relational DBMS
 - ▶ Native XML databases

Material about part 1

- Database System Concepts
  Silberschatz, Korth, Sudarshan
  7th Edition (or 6th Edition)
  McGraw-Hill, 2020
  https://www.db-book.com/db7/index.html
    (extra contents: original slides, solutions, etc)

- Principles of Distributed Database Systems
  Özsu and Valduriez
  3rd Edition
  Springer, 2011
  ISBN: 978-1-4939-4174-2/978-1-4419-8834-8 (e-book)
  https://extras.springer.com/2011/978-1-4419-8833-1.zip
    (original slides)
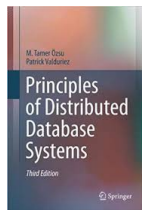
- other material (e.g., slides) will be made available on my webpage

Overview of part 2
Data analysis and big data

Forecast
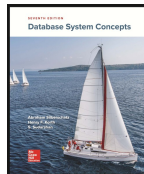
- Data warehousing and Business Analytics
- Introduction to big data
- NoSQL databases
- Text mining approaches and applications
- Time series modeling and forecasting
- Hadoop framework and MapReduce
- MongoDB

Actual

- Data warehousing and Business Analytics
- NoSQL databases
- Big data and the Hadoop framework
- Text mining approaches and applications
- Time series modeling and forecasting
- MongoDB   (by Paolo Gallo)

Material about part 2

- Database System Concepts
  Silberschatz, Korth, Sudarshan
  7th Edition (or 6th Edition)
  McGraw-Hill, 2020
  `https://www.db-book.com/db7/index.html`
    (extra contents: original slides, solutions, etc)

- Big Data – Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati
  Rezzani
  Maggioli Editore, 2013

- other material (e.g., slides) will be made available on my webpage