Testi del Syllabus

Resp. Did. **DELLA MONICA DARIO** **Matricola: 007120**

Anno offerta: **2021/2022**
Insegnamento: **MA0739 - ADVANCED DATABASE SYSTEMS FOR BIG DATA**
Corso di studio: **818 - ARTIFICIAL INTELLIGENCE & CYBERSECURITY**
Anno regolamento: **2021**
CFU: **6**
Settore: **INF/01**
Tipo Attività: **B - Caratterizzante**
Anno corso: **1**
Periodo: **Secondo Periodo**

Testi in italiano

Contenuti	La lingua del corso è l'inglese, pertanto si rimanda alla versione in inglese del syllabus.
Testi di riferimento	
Obiettivi formativi	
Prerequisiti	
Metodi didattici	
Altre informazioni	
Modalità di verifica dell'apprendimento	
Obiettivi per lo sviluppo sostenibile - Agenda 2030 [max 3]	

Testi in inglese

Contents	Advanced database models, languages, and systems. The students will learn, and practice, advanced query processing techniques for relational databases. They will also be introduced to the basic elements of distributed and parallel database management systems that play a fundamental role in the management of big data. Moreover, alternative data models and languages (e.g., XML databases) are introduced. Data analysis and big data. The students will learn, and practice, the main techniques and tools for data analysis and big data management. A special attention will be given to practical use cases, data warehousing, and methods and tools for big data. A number of key topics will be addressed, ranging from the MapReduce paradigm to time series and text analytics.
Texts	- Fundamentals of Database Systems (7th Edition), Elmasri and Navathe, Pearson, 2016 - Database System Concepts (7th Edition), Silberschatz, Korth, and Sudarshan, McGraw-Hill, 2020 - Readings in Database Systems (online, http://www.redbook.io) - Principles of Distributed Database Systems (3rd Edition), Özsu and Valduriez, Springer, 2011 - Data Warehouse Systems - Design and Implementation, A. Vaisman, E. Zimányi, Springer, 2014 - Business Analytics: A Contemporary Approach, Thomas Jackson, Steven Lockwood, WHSmith, 2018 - SQL & NoSQL Databases - Models, Languages, Consistency Options and Architectures for Big Data Management, Andreas Meier, Michael Kaufmann, Springer, 2019 - Text Mining: Concepts, Implementation, and Big Data Challenge (1st Edition), Taeho Jo, Springer,

2019
- Temporal Data Mining, Theophano Mitsa, CRC Press, 2010.
- Hadoop: The Definitive Guide (4th Edition), Tom White, O'Reilly, 2015.
- The MongoDB 4.2 Manual, MongoDB, Inc., https://docs.mongodb.com/manual/

Objectives

The overall aim of the course is to acquire an in-depth knowledge on advanced topics in data management within the relational paradigm (advanced query processing and optimization techniques, physical design, and distributed database systems), as well as alternative data models and languages (e.g., XML databases).

In addition, the course aims at providing competences about techniques and tools for big data management and analysis. A special attention will be given to data warehousing, data mining, and other methods and tools specific for big data. A number of key topics will be addressed, ranging from the MapReduce paradigm to blockchain and its applications.

At the end of the course, the student will be able to evaluate and tune the performance of a database, will have learned the concepts and methodologies for the configuration of distributed databases, and for the analysis of small and big data.

Sector-specific skills

1.1. Knowledge and understanding
- Parallel and distributed database system architectures.
- Data partitioning and replication in parallel and distributed systems.
- Centralized and distributed query processing and optimization.
- Alternative data model (with respect to the relational paradigm) for semi-structured and

unstructured data.
- Features of new generation (NoSQL, NewSQL) systems.

1.2. Applying knowledge and understanding
- Techniques and tools for small and big data analysis and visualization (e.g., R and RStudio).
- Optimization techniques for performance improvement in relational systems.
- Data processing in non-relational systems (e.g. XML and MapReduce).

Cross-sectoral skills/soft skills

2.1. Making judgments
- Choose the correct techniques and the appropriate tools to carry out data analyses.
- Interpret the experimental results of the analysis and draw effective conclusions relevant to the domain of discourse.
- Determine the most suitable (centralized, parallel, distributed, relational or non-relational) architecture for a specific data management problem.
- Implement the best strategies to improve the query performance.

2.2. Communication skills
- Communicate using the technical lexicon of database systems.
- Communicate using the terminology of parallel and distributed systems.
- Communicate with the (technical and non-technical) stakeholders involved in the process of design, implementation, and use of a database system (e.g., communicate effectively the results of the analysis).

2.3. Learning skills
- Learn to optimize a (possibly parallel or distributed) data management system.
- Learn to choose a sufficiently rich row data set, to analyze the data to extract meaningful information, to draw and to communicate conclusions.

Prerequisites	Knowledge about centralized relational database systems is required; basic knowledge about programming, algorithms and data structures, logic, and statistics are also desirable.
Teaching Methods	Classes mainly consist in lectures given by the teacher. Students are also introduced to software resources to download, install, and run for the first time: the teacher will give a brief practical introduction to them. Some classes are given by invited speakers, experts in some specific fields.
More Information	Additional suggested books: - PostgreSQL: Up and Running (3rd Edition), Regina Obe and Leo Hsu, O'Reilly Media, 2017 - An Introduction to XML and Web Technologies, Anders Møller and Michael I. Schwartzbach, Addison-Wesley, 2006 - Building the Data Warehouse (4th Edition), W. I. Immon, Wiley Publishing, 2005 - Big Data: A Very Short Introduction, Dawn Holmes, Oxford, 2017 - The Design and Implementation of Modern Colum-Oriented Database Systems, Daniel Abadi, Peter Boncz, Stavros Harizopoulos, Stratos Idreos, Samuel Madden, 2013 - What's Really New with NewSQL?, A. Pavlo and M. Aslett, ACM SIGMOD Record, Vol. 45, No. 2, pages 45-55, June 2016 - Column-Oriented Database Systems (slides), Stavros Harizopoulos, Daniel Abadi, and Peter Boncz, VLDB 2009 Tutorial, http://nms.csail.mit.edu/~stavros/pubs/tutorial2009-column_stores.pdf - Graph Databases (2nd Edition), Ian Robinson, Jim Webber, and Emil Eifrem, O'Reilly Media, 2015 - Big Data Management and NoSQL Databases - Lecture 7. Column-family stores (slides), Irena

	cz/~svoboda/courses/2015-1-NDBI040/lectures/Lecture-07-Column.pdf - Tutorial by Jeffrey Heer on Text Visualization (CSR 512 - Data Visualization), University of Washington - Introduction to Time Series Mining (slides), Keogh Eamonn - Temporal Data Mining, Theophano Mitsa, Taylor & Francis Ltd, 2010 - Apache Hadoop Online Documentation, Pig Latin Basics, https://pig.apache.org/docs/latest/basic.html - Hadoop Platform and Application Framework - Tutorial offered on Coursera by the University of California San Diego - MongoDB 4 Quick Start Guide, Doug Bierer, Packt Publishing Ltd, 2018 - Mastering MongoDB 3.x, Alex Giamas, Packt Publishing, 2017 - MongoDB Architecture Guide, MongoDB, Inc., http://s3.amazonaws.com/info-mongodb-com/MongoDB_Architecture_Guide.pdf - MongoDB Data Modeling, Wilson da Rocha França, Packt Publishing Ltd, 2015
Verification of learning	The exam consists of a written test and, possibly, an additional oral examination.