COMPUTER ENGINEERING (LM55)

(Lecce - Università degli Studi)

Teaching BIG DATA MANAGEMENT

GenCod A005793

Owner professor ANTONELLA LONGO

Teaching in italian BIG DATA MANAGEMENT

Teaching BIG DATA MANAGEMENT

SSD code ING-INF/05

Reference course COMPUTER ENGINEERING

Course type Laurea Magistrale

Credits 9.0

Teaching hours Front activity hours: 81.0

For enrolled in 2020/2021

Taught in 2021/2022

Course year 2

Language ENGLISH

Curriculum PERCORSO COMUNE

Location Lecce

Semester First Semester

Exam type Oral

Assessment Final grade

Course timetable https://easyroom.unisalento.it/Orario

BRIEF COURSE DESCRIPTION

The aim is to provide the basics about the main database theories, techniques and tools to design / implement databases and database applications.

Topics:

- Database, relational databases, NoSQL and NewSQL;
- DataBase Management Systems;
- Relational Model and Relational Algebra;
- SQL: data definition and manipulation;
- Basics of Human-Computer Interaction and interface design for DB;
- Architectural aspects: Clients, Servers, Peers, Devices, IoT, ...
- Big data, data lakes, data analytics, machine learning, AI;

REQUIREMENTS

Good knowledge of Object Oriented Languages (at least 1), techniques and tools. Elements of computer networks and Web technologies.

COURSE AIMS

Knowledge and understanding. Students must have a solid background related to the basics of big data management and information systems:

- They must have the basis to think analytically, creativelly and critically and being able to create abstraction and problem solving skills to cope with complex systems
- They must have a basic knowledge of design and implementation of big data management systems
- They must have the tools to design transactional and analytical databases applied to different contexts
- They must have the skills to argument data in different scenario, the tools for managing them, together with its impact.

Applying knowledge and understanding. After the course the student should be able to:

- Describe the model and frameworks of an Information System; illustrate the main components of an information system from the technical and application perspective.
- Distinguish conceptual, logical and physical models in big data management.
- Model Online Transaction processing systems from a big data perspective, distinguishing among conceptual models, relational models and physical models
- Model Online Analytical processing systems form a data perspective, distinguishing among conceptual, logical and physical models, being able to describe the relationships among them and the processes

Making judgements. Students are guided to critically approach the topics treated during the class, to compare different solutions to a problem, to identify and propose the most effective or efficient solution in an autonomous way.

Communication. Students must learn to communicate with heterogeneous audiences, explaining their position, in logical, coherent and effective way. During the course students will be provided with domain specific vocabulary and the proper scientific knowledge and methods to expose and argument in precise and formal way the main topics related to big data management and information systems

Learning skills. Students must acquire the critical ability to autonomously relate to the typical problems of data and information management and, in general, cultural issues related to information systems and their management. They should be able to develop an approach to independently structure knowledge and methods learnt with a view to possible continuation of studies at higher (doctoral) level or in the broader perspective of cultural and professional self-improvement of lifelong learning. Therefore, students should be able to switch their learning approach according to different learning sources and the objectives they must achieve in terms of results and audience

TEACHING METHODOLOGY

The course aims to provide students with tools and knowledge for data management in business organizations. The course consists of frontal lessons and classroom hands on exercises. The frontal lessons are aimed at improving students' knowledge and understanding through the presentation of theories, models and methods; students are invited to participate in the lesson with autonomy of judgement, by asking questions and presenting examples. The exercises are aimed at using tools which supports the models and approaches presented

ASSESSMENT TYPE	The exam is an interview made up of both practical and descriptive aspects The practical part aims at evaluating to what extent the student has: 1) the ability to design data models according to the methodologies presented during the call, 2) reasoning about his/her choices and the capacity to integrate different concepts and tools. The descriptive part follows the practical part and is aimed to verify to what extent the student has gained knowledge and understanding of selected topics and he is able to communicate them.
OTHER USEFUL INFORMATION	**Office Hours** By appointment; contact the instructor by email or at the end of class meetings.

FULL SYLLABUS

- Databases and Database Users
- Database System Concepts and Architecture
- Data Modeling Using the Entity–Relationship (ER) Model
- The Enhanced Entity–Relationship (EER) Model
- The Relational Data Model and Relational Database Constraints
- Basic SQL
- More SQL: Complex Queries, Triggers, Views, and Schema Modification
- Relational Database Design by ER- and EER-to-Relational Mapping
- Introduction to SQL Programming Techniques: Basics of Functional Dependencies and Normalization for Relational Databases, Informal Design Guidelines for Relation Schemas, Functional Dependencies, Normal Forms Based on Primary Keys, General Definitions of Second and Third Normal Forms, Boyce-Codd Normal Form
- Concurrency Control Techniques
- Main topics of Informations Systems: Anthony pyramid, Enterprise applications, OLTP and OLAP
- Data Warehouse Definition and lifecycle
- Modelling of analytical data processing systems: DFM, Snow Flakes schema
- Tools for data Visualization
- Introduction at big data and big data processing pipeline
- No SQL Databases and CAP Theorem
- Different kinds of no SQL Databases
- Distributed systems
- Big Data quality
- Hands on Data Exploratory analysis
- Machine learning tecniques and big data

- **Teaching material:** more concepts on requirement elicitation and database application design and implementation, multidimensional analisys, datawarehouse, big data, big data management, database security, database administration, NoSQL, NewSQL, distributed databases.

REFERENCE TEXT BOOKS	R. Elmasri, S. Navathe, Fundamental of Database Systems, 7a edizione, Pearson ed. Balamurugan Balusamy, Nandhini Abirami R, Amir H. Gandomi, Big Data: Concepts, Technology, and Architecture,John Wiley & Sons Inc; 1. edizione Material provided during the class