



UNIVERSITÀ DEGLI STUDI DI PALERMO

DEPARTMENT	Matematica e Informatica
ACADEMIC YEAR	2016/2017
MASTER'S DEGREE (MSC)	COMPUTER SCIENCE
SUBJECT	BIG DATA MANAGEMENT
TYPE OF EDUCATIONAL ACTIVITY	B
AMBIT	50341-Discipline Informatiche
CODE	17393
SCIENTIFIC SECTOR(S)	INF/01
HEAD PROFESSOR(S)	ROMBO SIMONA ESTER Professore Ordinario Univ. di PALERMO
OTHER PROFESSOR(S)	
CREDITS	6
INDIVIDUAL STUDY (Hrs)	102
COURSE ACTIVITY (Hrs)	48
PROPAEDEUTICAL SUBJECTS	
MUTUALIZATION	
YEAR	1
TERM (SEMESTER)	2° semester
ATTENDANCE	Not mandatory
EVALUATION	Out of 30
TEACHER OFFICE HOURS	ROMBO SIMONA ESTER Monday 9:30 13:30 In presenza: Via Archirafi 34, Piano II, Stanza 220 - Telematico: via Microsoft Teams o altro canale - In entrambi i casi, e' consigliabile prenotarsi tramite email alla docente

DOCENTE: Prof.ssa SIMONA ESTER ROMBO

PREREQUISITES	Basic knowledge on the design of databases. Advanced knowledge on programming.
LEARNING OUTCOMES	<p>Knowledge and ability to understand Acquisition of knowledge about the organization of business information systems, on the use and design of decision support systems, on problems related to big data management.</p> <p>Ability to apply knowledge and understanding Ability to analyze issues related to contexts characterized by large data sets and propose design solutions for the management and analysis of data in such contexts and decision support.</p> <p>Making judgments Ability to analyze and evaluate solutions for the management of large quantities of data. Ability to design decision support systems by analyzing the technical specifications provided.</p> <p>Enable communication Ability to describe designing solutions for complex information systems and decision support systems, and also to analyze their performance. Ability to cooperate for determining the appropriate design solutions in application contexts characterized by the presence of large amounts of data.</p> <p>Learning Abilities Ability to update by consulting advanced books and scientific publications related to the topics covered during the course. Ability to follow, using the knowledge acquired, both second master level and advanced courses and specialized seminars in the field of Business Intelligence.</p>
ASSESSMENT METHODS	<p>Project, oral examination. Project: Design and implementation of a project based on big data technologies. It will be mandatory that the project is carried out by a team consisting of a minimum of two to a maximum of five students, in order to encourage them to work in team. Two revisions of the project will be necessary, during which the teacher will verify the individual contribution of each team member and assign a rating to the different phases of design. An overall score will be given to the project, weighted for each student depending on his/her individual contribution. A positive evaluation will depend on the originality of the proposed solution, the methodological rigor, the acquisition of technical skills supplied by the course. The requirements to achieve the minimum score to access the second part of the examination consist of the ability to realize a project that, while so basic, satisfy the assigned requirements and it is correct.</p> <p>Oral test: it will serve to check the knowledge gained during the course, the autonomy in deepening even complex content and the ability to find individual solutions to the proposed problems. The oral test will begin with the presentation of a topic chosen from those proposed by the teacher during the course, on which the student will have produced a short essay. So, it will be checked for critical capacity and autonomy of the student's judgment through a sufficient number of questions. The assessment of the oral examination will complement the marks obtained by each student as a result of the project evaluation.</p>
EDUCATIONAL OBJECTIVES	The main goal of Big Data Management is to provide students with knowledge about the organization of business information systems, the use and the design of decision support systems, and also on issues related to the management of large amounts of data. The course will begin with an overview of Business Intelligence and it will focus on the design of data warehouse and OLAP, which will also cover part of the activities of laboratory. The course will also address the design of non-relational database and the use of technologies such as Apache Hadoop and Spark. Finally, real application scenarios will be illustrated such as social networks, smart cities and biological data.
TEACHING METHODS	Lectures.
SUGGESTED BIBLIOGRAPHY	<p>M. Golfarelli, S. Rizzi, "Data Warehouse – Teoria e pratica della progettazione", Seconda Edizione, McGraw Hill, 2005.</p> <p>J. Han, M. Kamber, "Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers, 2006.</p> <p>H. Garcia-Molina, J.D. Ullman, J. Widom, "Database Systems – The complete book", Second Edition, Pearson Education, 2008.</p>

SYLLABUS

Hrs	Frontal teaching
2	INTRODUCTION TO BIG DATA The three Vs. Needs. application contexts. An overview of the technologies.

SYLLABUS

Hrs	Frontal teaching
6	BUSINESS INTELLIGENCE AND DATA WAREHOUSING Data acquisition. Processing of data in order to produce information. Archiving of raw data and information. Transmission of data and information. Presentation of data and information. Decision support systems. Architectures for Data Warehousing. The metadata. Quality of a Data Warehouse. The multidimensional model and OLAP. The main OLAP operations. Logical models of Data Warehouse.
18	DESIGN OF A DATA WAREHOUSE Design methodologies, process selection, choice of granularity, identifying and bringing the size, selection of measures, precalculations in the fact table, complete the dimension table, choice of the duration of the database, track "slowly changing dimension", deciding the priorities queries, query mode, integration of Data Marts. Database and applications implementation. Using ETL tools for the design and implementation of data warehouse, OLAP, Data Mining.
2	REVIEW OF DATA MINING Description of Concepts. Market Basket Analysis. Classification and prediction. Decision trees. K-Nearest Neighbor. Clustering.
6	STRUCTURED STORAGE Non-relational database. NoSQL database types. Analysis of the advantages and disadvantages of not-relational databases. Implementation. Examples and exercises on NoSQL database.
8	TECHNOLOGIES FOR BIG DATA The MapReduce algorithm. Study of tools such as Apache Hadoop and Spark. Applications and exercises using such tools.
6	ANALYSIS OF REAL CONTEXTS Social networks. Biological networks. Next-Generation Sequencing (NGS). Smart Cities.