

Big Data Analytics and Management

(1) GENERAL

SCHOOL	SCHOOL OF INFORMATION SCIENCES & TECHNOLOGY		
ACADEMIC UNIT	DEPARTMENT OF STATISTICS		
LEVEL OF STUDIES	POSTGRADUATE		
COURSE CODE		TRIMESTER	2
COURSE TITLE	Big Data Analytics and Management		
INDEPENDENT TEACHING ACTIVITIES		WEEKLY TEACHING HOURS	CREDITS
		3	5
		30	5
COURSE TYPE	Specialization Course in Analytics with Applications in Sports Science		
PREREQUISITE COURSES:			
LANGUAGE OF INSTRUCTION and EXAMINATIONS:	ENGLISH		
COURSE DELIVERY METHOD	Distance		
IS THE COURSE OFFERED TO ERASMUS STUDENTS	NO		
COURSE WEBSITE (URL)	https://cloud.aueb.gr/index.php/s/zLp3HGgD7bGAZRy		

(2) LEARNING OUTCOMES

Learning outcomes
• Understanding the fundamental principles of analyzing and managing big data. • Familiarity with the challenges and issues of the Big Data era, such as scalability and information complexity. • Understanding modern technologies and trends in the field of Big Data, including large databases and processing platforms. • Introduction to the Map-Reduce paradigm and the basic principles of distributed data processing. • Understanding Big Data Mining and its applications. • Familiarity with fundamental database concepts and basic design schemas. • Development of basic SQL skills for data retrieval and management. • Introduction to Data Engineering and data cleaning processes. • Handling missing values and applying data preprocessing techniques. • Familiarity with web scraping techniques for building datasets from websites.
General Competences
• Analysis and synthesis of data and information, using the necessary data technologies. • Decision-making. • Working in an interdisciplinary environment.

- Adaptability to new situations.

(3) SYLLABUS

The course examines the analysis and management of large-scale data. It also explores the problems and challenges that arise in the era of Big Data. Technologies and trends in the field of Big Data are analyzed, including large databases, the Map-Reduce paradigm, Big Data mining, and Big Data platforms.

Definitions and explanations of fundamental database concepts and their basic design schemas are provided. A brief introduction to the SQL language is also included.

In addition, the course covers data engineering and data cleaning, with emphasis on handling missing values and collecting data from websites (web scraping).

(4) TEACHING and LEARNING METHODS - EVALUATION

DELIVERY <i>Face-to-face, Distance learning, etc.</i>	Distance	
METHOD AND FREQUENCY OF COMMUNICATING WITH THE STUDENTS	Remotely via email and through weekly office hours (and whenever necessary) via teleconferencing or in-person office visits.	
ENSURING COMMUNICATION AMONG STUDENTS	Teleconference, Chat via eclass and/ or TEAMS, and QA sessions	
USE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY	MS Teams, e-class, MS Outlook, R, Python, WayGround or Kahoot educational games and quizzes	
REQUIRED EQUIPMENT AND TECHNOLOGY KNOWLEDGE	Camera, mic, PC, MS Office, and TEAMS	
COURSE POLICY ON PLAGIARISM AND PLAGIARISM DETECTION TOOLS	Turnitin	
COURSE POLICY ON USE OF AI TOOLS	The use of Artificial Intelligence is permitted with explicit reference to the bibliography (2) for the verbal correction of assignments and as long as students have understood the basic principles and methods of the course.	
TEACHING METHODS	Activity	Semester workload
	Lectures	30
	Assignment writing	40
	Lab Exercise	20
	Self-Study hours	40
	Course total	130
STUDENT PERFORMANCE EVALUATION	Evaluation language is English. Evaluation methods include lab exercises, written assignments, and/or written examinations. Oral examinations of the assignments may be conducted if clarifications are needed or if there	

	are suspicions of plagiarism or unauthorized use of Artificial Intelligence tools. Assessment criteria are provided on the course eClass platform and in the course materials.
--	--

(5) RECOMMENDED BIBLIOGRAPHY

• Chen, M., Mao, S., & Liu, Y. (2014). Big data: Related technologies, challenges and future prospects. Springer. • Marz, N., & Warren, J. (2015). Big data: Principles and best practices of scalable realtime data systems. Manning Publications.
--