



**AMERICAN UNIVERSITY
OF CYPRUS**

Course Code MSE 481	Course Name BIG DATA ANALYSIS	ECTS Credit 7.5
Pre-Requisite	Course Type Major Elective	Language of Instruction English
Year of Study 4 th /8 th	Level of Course BSc/1st Cycle	Mode of Delivery On Campus

Course Objectives:

This course will provide an understanding of what big data can offer through real life examples and hands-on exercises. Students will learn the various data modeling techniques. Attendees will be guided through the basics of using Spark and by the end of the course they should be able to analyze unstructured data and develop small ETL project using spark to refine those datasets.

Learning Outcomes

6. Understand the Fundamentals of Big Data.
7. Categorize and define the various big data modeling techniques.
8. Understand the fundamentals of Cloud Security
9. Analyze and recommend batch or stream data ingestion mechanisms for specific case studies
10. Implement basic Spark software that processes unstructured data, converting them to structured datasets.

Teaching Methodology:

In the Classroom: Lecturers make use of whiteboards, flipcharts, overhead projector, VR glasses, video material and power point presentations. Students are supplied with handouts on extra or relevant material. Two Personal Computer Labs equipped with Multimedia PCs of the latest technology with the required software, scanners, printers and LCD-Projectors, satisfy the classes' requirements. All PCs are connected to the Internet, through a Broad Band High speed permanent connection using cable technology.

Web Supported Learning: All the teaching material and the Lecturer's presentations are uploaded on the electronic learning platform of the college as a supporting studying tool.

Guest Speakers / Visits: External visits to agencies or relevant industry/subject related organizations are arranged. Guest speakers that are experts in their field are invited to address the students. Students are also encouraged to visit industry players and familiarize themselves with the profession they have chosen.

Teaching Methods: Lectures, presentations, videos, problem and case study discussion, discussion on relevant articles, independent and private study, preparation of projects, fieldwork and group work.

Course Content:

- 8. Introduction to Big Data**
 - a. Facts about Big Data

- b. What makes Big Data Valuable
 - c. Types of Big Data
 - I. People Generated
 - II. Machine Generated
 - III. Organization Generated
 - d. Examples of Big Data
- 9. The 5 V's of Big Data**
 - a. Volume
 - b. Variety
 - c. Velocity
 - d. Veracity
 - e. Valance
- 10. Getting Value out of Big Data**
 - a. Acquiring Data
 - b. Exploring Data
 - c. Pre-processing Data
 - d. Analyzing Data
- 11. Big Data Modeling and Management**
 - a. Introduction
 - I. Data Ingestion
 - II. Data Storage
 - III. Data Quality
 - IV. Data Operations
 - V. Data Scalability and Security
 - b. Data Models
 - I. Data Model Structures
 - II. Data Model Operations
 - III. Data Model Constraints
 - IV. Vector Space Model
 - V. Graph Data Model
- 12. Big Data Battle: Batch Processing vs Stream Processing**
 - a. What is Batch Processing
 - b. What is Stream Processing
 - c. When do we choose Batch and when Stream Processing?
- 13. Introducing Apache Spark**
 - a. Data-Parallel to Distributed Data-Parallel
 - b. Latency
 - c. Resilient Distributed Dataset (RDD)
 - d. Evaluation in Spark
 - e. Cluster Topology
- 14. Data partitioning and Shuffling**
 - a. Reduction Operations
 - b. Distributed Key-Value Pairs
 - c. Shuffling: Why it is important
 - d. Partitioning
 - e. Wide vs Narrow Dependencies

Assessment Methods:

The final course grade is made up of:

Coursework

Final Examination

Final Examination marks constitute 60% of the final semester mark, while Midterm Examinations and assignments (when applicable) constitute the 40%.

The pass mark is set at 50%

The Midterm examinations are based on material covered during a given period and are set towards the end of November in the case of Fall semesters, the end of March in the case of Spring semesters, and the end of July in the case of summer sessions.

These take place during lesson time and take no more than two study periods to complete.

Final examinations are based on material covered throughout the semester. The dates for these are set down on the academic calendar. The Final Examinations have duration of three hours for Diplomas, Bachelor Degrees and Master Degrees.

Final examination marks are combined with the marks from the Midterm examinations, participation and assignments (when applicable) to produce the final mark for the semester.

Required Textbooks/Reading:

Required Bibliography:

Title	Author(s)	Publisher	Year
Data: Principles and best practices of scalable realtime data systems	Nathan Marz and James Warren	Manning Publications;(10 May 2015)	2015

Recommended Further Bibliography:

Title	Author(s)	Publisher	Year
Spark: The Definitive Guide: Big Data Processing Made Simple	Bill Chambers; Matei Zaharia	O'Reilly Media	2018