



Jordan University of Science and Technology
Faculty of Computer & Information Technology
Artificial Intelligence Department

AI328 Big Data Processing
First Semester 2024-2025

Course Catalog
<p>3 Credit Hours. This course offers an introduction to large-scale data analytics, focusing on the methods and technologies that enable the extraction of actionable insights from massive datasets. Big Data analytics involves uncovering non-trivial knowledge and patterns within vast collections of data, using advanced computational tools and statistical models. The course covers two primary areas: (1) the programming techniques employed by data scientists for large-scale data processing, and (2) the models used for data analysis. On the technical side, students will explore fundamental systems and techniques for managing and storing large volumes of data. We will explore modern cluster computing systems, with a focus on MapReduce-based frameworks such as Hadoop and Apache Spark, and cover their role in distributed data processing. On the modeling side, the course will cover key supervised and unsupervised learning models, providing a solid foundation in data mining techniques.</p>
Teaching Method: Blended

Text Book	
Title	Data Mining: Concepts and Techniques
Author(s)	Jiawei Han, Micheline Kamber and Jian Pei
Edition	3rd Edition
Short Name	2
Other Information	The Morgan Kaufmann Series in Data Management Systems

Course References

Short name	Book name	Author(s)	Edition	Other Information
1	Huawei Training Material	Huawei	3rd Edition	
3	Hadoop-The Definitive Guide	Tom White	4th Edition	

Instructor	
Name	Dr. Farah AlShanik
Office Location	-

Office Hours	
Email	fmalsharik@just.edu.jo

Class Schedule & Room
Section 1: Lecture Time: Sun, Tue : 14:30 - 15:30 Room: M2008

Prerequisites		
Line Number	Course Name	Prerequisite Type
1742210	Cls221 Fundamentals Of Database Systems	Prerequisite / Study

Tentative List of Topics Covered		
Weeks	Topic	References
Weeks 1, 2	Introduction to Big Data Analytics	From 1
Weeks 3, 4	Getting to know your data	From 2
Weeks 5, 6	Data preprocessing	From 2
Weeks 7, 8	Mining association rules in large databases	From 2
Weeks 9, 10	HDFS- Hadoop Distributed File System	From 1, From 3
Week 11	MapReduce	From 1, From 3
Week 12	Spark	From 1, From 3
Weeks 13, 14	Classification and prediction	From 2

Relationship to Program Student Outcomes (Out of 100%)					
SO1	SO2	SO3	SO4	SO5	SO6

Policy	
Attendance Policy	It is very important to attend the class since the lecture contents may be drawn from various sources. Failure to attend the course most likely leads to missing some information that won't show in lecture notes or textbook.
Changes to Syllabus	Lecture topics and assignments are subject to change. The course syllabus is a general plan for the course; deviations to the class may be necessary and will be announced to class by the instructor.

