هيئة تقويم التعليم والتدريب
Education & Training Evaluation Commission

2023
TP-153

Course Specification
(Bachelor)

Course Title:	**Data Engineering**
Course Code:	**DSC 232**
Program:	**Applied Statistics & Data Management**
Department:	**Mathematics**
College:	**College of Science**
Institution:	**Majmaah University, Saudi Arabia**
Version:	*Course Specification Version Number*
Last Revision Date:	*Pick Revision Date.*

Table of Contents

A. General information about the course:

1. Course Identification

1. Credit hours: (2+2) 3					
2. Course type					
A.	☐University	☐College	☒Department	☐Track	Others
B.	☒Required		☐Elective		
3. Level/year at which this course is offered: (4)					
4. Course general Description:					
This course provides the foundation for data science concepts, data science thinking and analytics, and forms an important part of all businesses. This course will help students to explore various tools and methods that are used for understanding the data engineering process using Python and how to tackle challenges commonly faced in different aspects of data engineering. You'll start with an introduction to the basics of data engineering, along with the technologies and frameworks required to build data pipelines to work with large datasets. You'll learn how to transform and clean data and perform analytics to get the most out of your data. As you advance, you'll discover how to work with big data of varying complexity and production databases and build data pipelines. Using real-world examples, you'll build architectures on which you'll learn how to deploy data pipelines.					
5. Pre-requirements for this course (if any)**:**					
MTHS 212 Mathematical Programming 1					
6. Co-requisites for this course (if any)**:**					
N/A					
7. Course Main Objective(s):					
• **Students will have gained a foundation for data science concepts, data science thinking and analytics.** • **Student will be able to identify the problems and tasks involved in the life cycle of a data science project, including data collection, data preprocessing and data analysis.** • **Student will have gained a clear understanding of data modeling techniques.** • **Student will be able to confidently build data engineering pipelines for tracking data, running quality checks, and making necessary changes in production.**					

2. Teaching mode (mark all that apply)

No	Mode of Instruction	Contact Hours	Percentage
1	Traditional classroom	25	55.56%
2	E-learning		

No	Mode of Instruction	Contact Hours	Percentage
3	Hybrid • Traditional classroom • E-learning		
4	Distance learning		
5	Blended	20	44.44%

3. Contact Hours (based on the academic semester)

No	Activity	Contact Hours
1.	**Lectures**	25
2.	**Laboratory/Studio**	20
3.	**Field**	
4.	**Tutorial**	
5.	**Others (specify)**	
Total		45

B. Course Learning Outcomes (CLOs), Teaching Strategies and Assessment Methods

Code	Course Learning Outcomes	Code of CLOs aligned with program	Teaching Strategies	Assessment Methods
1.0	**Knowledge and understanding**			
1.1	Understand how data engineering supports data science workflows	K3	Lectures Presentations Media Lectures Tutorials	Homework Quiz Mid Exam Final Exam E-exam Oral Exam
1.2				
...				
2.0	**Skills**			
2.1	Find out how to implement a data pipeline and dashboard to visualize results.	S3	Lectures Presentations Media Lectures	Homework Quiz Mid Exam

Code	Course Learning Outcomes	Code of CLOs aligned with program	Teaching Strategies	Assessment Methods
			Tutorials	Final Exam E-exam
2.2	Discover how to extract data from files and databases and then clean, transform, and enrich it.	S3	Lectures Presentations Media Lectures Tutorials	Homework Quiz Mid Exam Final Exam E-exam
...				
3.0	**Values, autonomy, and responsibility**			
3.1	Use staging and validation to check data before landing in the warehouse.	V3	Lectures Presentations Media Lectures Tutorials	Homework Quiz Mid Exam Final Exam E-exam
3.2	Get to grips with deploying pipelines in the production environment.	V3	Lectures Presentations Media Lectures Tutorials	Homework Quiz Mid Exam Final Exam E-exam
...				

C. Course Content

No	List of Topics	Contact Hours
1.	Introduction to Fundamentals of Data Science :Definition ,History ,Data, Information, Knowledge, Intelligence, Big Data Versus Small Data, applications of Data Science and Process-Based Data Science	4
2.	Introduction to data Engineering and Building Our Data Engineering Infrastructure	4
3.	Reading and Writing Files and Working with Databases	4
4.	Lebesgue Cleaning, Transforming, and Enriching Data	4
5.	Building a 311 Data Pipeline	4
6.	Features of a Production Pipeline	4
7.	Version Control Using the NiFi Registry	5
8.	Monitoring and Logging Pipelines Deploying your Pipelines	8

9.	Building a Production Data Pipeline Building a Kafka Cluster	8
Total		**45**

D. Students Assessment Activities

No	Assessment Activities *	Assessment timing (in week no)	Percentage of Total Assessment Score
1.	Mid exam - 1	6 th	15
2.	Mid exam - 2	12 th	15
3.	Homework	Through of semester	10
4.	Quiz	Through of semester	10
5.	E.exam	14th	10
6.	Final exam	End of semester	40
...			

*Assessment Activities (i.e., Written test, oral test, oral presentation, group project, essay, etc.).

E. Learning Resources and Facilities

1. References and Learning Resources

Essential References	Data Engineering with Python: Work with massive datasets to design data models and automate data pipelines using Python, Paul Crickard, Packt publishing Birmingham –MUMBAI, 2020. ISBN-13: 978-1839214189
Supportive References	Data Engineering with Apache Spark, Delta Lake, and Lakehouse: Create scalable pipelines that ingest, curate, and aggregate complex data in a timely and secure way, Manoj Kukreja, Packt , 2021.
Electronic Materials	
Other Learning Materials	

2. Required Facilities and equipment

Items	Resources
facilities (Classrooms, laboratories, exhibition rooms, simulation rooms, etc.)	Classroom with capacity of 30-students. Computer Lab of Mathematics Department
Technology equipment (projector, smart board, software)	Mathematical software packages like MINITAB and SPSS ، Python

Items	Resources
Other equipment (depending on the nature of the specialty)	

F. Assessment of Course Quality

Assessment Areas/Issues	Assessor	Assessment Methods
Effectiveness of teaching	Students/ internal committee	Direct (Students evaluation electronically organized by Deanship of registration and admission)/ Verification of students' papers
Effectiveness of Students assessment	Staff members (Peer Reviewer)	Indirect (Frequent meetings consultation among the teaching staffs)
Quality of learning resources	Staff members (course coordinators)	Direct (Meeting between course coordinators and the tutors)
The extent to which CLOs have been achieved		
Other		

Assessors (Students, Faculty, Program Leaders, Peer Reviewer, Others (specify)

Assessment Methods (Direct, Indirect)

G. Specification Approval

COUNCIL /COMMITTEE	
REFERENCE NO.	
DATE	