



Course Specification

DIPLOMA

Course Title: **Data Collection and Integration**

Course Code: **APDA2208**

Program: **Diploma in Data Analytics**

Department: **Diploma Department**

College: **The Applied college**

Institution: **Umm Al-Qura University**

Version: **1**

Last Revision Date: **05 May 2025**



Table of Contents

A. General information about the course:	3
B. Course Learning Outcomes (CLOs), Teaching Strategies and Assessment Methods	4
C. Course Content.....	5
D. Students Assessment Activities	6
E. Learning Resources and Facilities.....	6
F. Assessment of Course Quality	7
G. Specification Approval	7



A. General information about the course:

1. Course Identification

1. Credit hours: (3)

2. Course type

A. ☐ University ☐ College ☐ Department ☒ Track ☐ Others
B. ☒ Required ☐ Elective

3. Level/year at which this course is offered: (2nd leve/ 1st year)

4. Course General Description:

This course will introduce the basic principles of data collection and integration. Students will learn skills and methods for data collection, data cleaning, data wrangling and transformation. This course also covers the basic concepts of data Integration and how to resolve structural heterogeneity through schema matching and mapping and focuses on how to query different heterogeneous data sources at once and translate data between databases with different data representations to cover the interoduction of fundamental techniques and novel applications of data warehouse.

5. Pre-requirements for this course (if any):

None

6. Co-requisites for this course (if any):

None

7. Course Main Objective(s):

The objective of this course is to develop students' skills required to obtain data from existing sources and manage new data collection systems, with the focus on value-driven data collection. This course also aims at Learning basic concepts of data integration, how to resolve structural heterogeneity through schema matching and mapping, and concepts of data warehousing and different techniques.

2. Teaching mode (mark all that apply)

No	Mode of Instruction	Contact Hours	Percentage
1	Traditional classroom	60	100%





No	Mode of Instruction	Contact Hours	Percentage
2	E-learning		
3	Hybrid <ul style="list-style-type: none"> Traditional classroom E-learning 		
4	Distance learning		

3. Contact Hours (based on the academic semester)

No	Activity	Contact Hours
1.	Lectures	30
2.	Laboratory/Studio	30
3.	Field	
4.	Tutorial	
5.	Others (specify)	
Total		60

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

Code	Course Learning Outcomes	Code of PLOs aligned with the program	Teaching Strategies	Assessment Methods
1.0	Knowledge and understanding			
1.1	Identify basic methods and techniques used for data collection, cleaning, wrangling, and integration.	K1	Lectures Discussions	Quizzes, Exam and Group project
1.2	Discuss ethical and social considerations of data collection and integration.	K1	Lectures Discussions	Quizzes, Exam and Group project
2.0	Skills			
2.1	Collect, manipulate, wrangle, and integrate data.	S4	Lectures Discussions Labs	Quizzes, Exam and Group project
2.2	Inferring new data from existing data, and integrating data	S4	Lectures Discussions Labs	Quizzes, Exam and Group project



Code	Course Learning Outcomes	Code of PLOs aligned with the program	Teaching Strategies	Assessment Methods
	sources to build data warehouse.			
3.0	Values, autonomy, and responsibility			
3.1	Demonstrate ethical considerations of data collection and integration.	V2	Lectures Discussions Labs	Quizzes, Exam and Group project
3.2	Enhance teamwork and leadership skills by working in groups to achieve the solutions to addressed exercises/projects.	V3	Lectures Discussions Labs	Quizzes, Exam and Group project
3.3	Demonstrate self-learning abilities to acquire new methods and techniques to deal with data.	V3	Lectures Discussions Labs	Quizzes, Exam and Group project

C. Course Content

No	List of Topics	Contact Hours
1.	Introduction to data collection and preprocessing	4
2.	Ethical and legal considerations in data collection	4
3.	Exploratory data analysis and visualization	4
4.	Handling missing data, duplicate data, outliers	8
5.	Data transformation	4
6.	Text data preprocessing	4
7.	Data scraping (Web scraping + API scraping)	4
8.	Autonomous and distributed data sources	4
9.	Data Integration: Schema matching	4
10.	Data Integration: Schema mappings	4
11.	Data Exchange and transformations	4
12.	Basic Concepts of Data Warehousing	4
13.	Data warehousing- Online analytical processing (OLAP) and data mining	4
14.	Data Warehousing architecture, Design and modeling	4
Total		60



D. Students Assessment Activities

No	Assessment Activities *	Assessment timing (in week no)	Percentage of Total Assessment Score
1.	Quizzes	4-11	20%
2.	Midterm Exam	8	20 %
3.	Group Project	14	20%
4.	Final Exam	Exam week	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	The Definitive Guide to Data Integration: Unlock the power of data integration to efficiently manage, transform, and analyze data by Pierre-Yves BONNEFOY, Emeric CHAIZE, Raphaël MANSUY, Mehdi TAZI Released March 2024 publisher(s): Packt.
Supportive References	Gagolewski, M. (2022). Minimalist Data Wrangling with Python Rattenbury, T., Hellerstein, J. M., Heer, J., Kandel, S., & Carreras, C. (2017). Principles of data wrangling: Practical techniques for data preparation. " O'Reilly Media, Inc.".
Electronic Materials	Blackboard
Other Learning Materials	

2. Required Facilities and equipment

Items	Resources
facilities (Classrooms, laboratories, exhibition rooms, simulation rooms, etc.)	Classrooms and laboratories
Technology equipment (projector, smart board, software)	Projector
Other equipment (depending on the nature of the specialty)	

F. Assessment of Course Quality

Assessment Areas/Issues	Assessor	Assessment Methods
Effectiveness of teaching	Students	Indirect



Assessment Areas/Issues	Assessor	Assessment Methods
		Course survey and students' feedback.
Effectiveness of Students	Faculty Members, Peer Reviewers	Direct Report on the satisfaction of exam standards.
Quality of learning resources	Faculty Member, Course Coordinators	Direct Learning resources evaluation survey.
The extent to which CLOs have been achieved	Faculty Members, Program Leaders	Direct Course reports.
Other		

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

Assessment Methods (Direct, Indirect)

G. Specification Approval

COUNCIL /COMMITTEE	Umm Al-Qura University Council
REFERENCE NO.	851281214463/193664
DATE	1447/01/20

