Course Specification
(Bachelor)

Course Title: Data Engineering	
Course Code: SE1505	
Program: Software Engineering	
Department: Software Engineering	
College: Faculty of Computer Science and Information Technology	
Institution: Al-Baha University	
Version: V1.0	
Last Revision Date: 24-April-2024	

Table of Contents

A. General information about the course:

1. Course Identification

1. Credit hours: (3)

3 Credit Hours (2, 2, 0) (Lecture, Lab, Tutorial)
(4 Contact Hours)

2. Course type

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

3. Level/year at which this course is offered: (8th Level/3rd Year)

4. Course General Description:

This course provides students with a 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 and how to tackle challenges commonly faced in different aspects of data engineering. Students start with an introduction to data modeling. ETL Processes, Database Systems, Big Data Technologies, and Data Warehousing.
 Students will 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):   Database 1 (SE1003)

6. Pre-requirements for this course (if any):   None

7. Course Main Objective(s):

The main objectives of this course are

- **Students will have gained a foundation for data science concepts, data science thinking and analytics.**
- **Students 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.**
- **Students will have gained a clear understanding of data modeling techniques.**
- **Students 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	22	66%
2	E-learning		
3	Hybrid • Traditional classroom • E-learning		
4	Distance learning		
5	Other (Lab)	22	34%

3. Contact Hours (based on the academic semester)

No	Activity	Contact Hours
1.	**Lectures**	22
2.	**Laboratory/Studio**	22
3.	**Field**	
4.	**Tutorial**	
5.	**Others (specify)**	
Total		44

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	Students will be able to Understand how data engineering supports data science workflows	K1	- Lectures	***Direct Assessment Tool*** Midterm Exam Final exam ***Indirect Assessment Tool*** Course Exit Survey
1.2	Students will be able to understand staging and validation to check data before loading in the data warehouse.	K2	- Lectures - Lab Work	***Direct Assessment Tool*** Midterm exam Final exam ***Indirect Assessment Tool*** Course Exit Survey
1.3	Students will be able to understand how to deploy pipelines in the production environment	K3	- Lectures -Lab Work	***Direct Assessment Tool*** Midterm exam Final exam ***Indirect Assessment Tool*** Course Exit Survey
2.0	**Skills**			
2.1	Demonstrate how to implement a data pipeline and dashboard to visualize results. .	S1	- Lectures - Lab work - Project	***Direct Assessment Tool*** Project (rubric) Final exam ***Indirect Assessment Tool*** Course Exit Survey
2.2	Discover how to extract data from files and databases and then clean, transform, and enrich it	S2	- Lectures - Lab work - Project	***Direct Assessment Tool*** Project (rubric) Final exam ***Indirect Assessment Tool*** Course Exit Survey
2.3				
3.0	**Values, autonomy, and responsibility**			
3.1	Recognize the importance of teamwork, collaboration, and	V1	-Small Groups	***Direct Assessment Tool***

Code	Course Learning Outcomes	Code of CLOs aligned with program	Teaching Strategies	Assessment Methods
	communication in the design and development of secure software systems. Furthermore the importance of giving and receiving constructive feedback.			Project Presentation (rubric) *Indirect Assessment Tool* Course Exit Survey

C. Course Content

No	List of Topics (Lectures)	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**	3
2.	**Data Pipelines: Understanding the design and construction of data pipelines**	2
3.	**Data Flow: Grasping the flow of data through different stages of processing, transformation, and storage within a data engineering ecosystem**	3
4.	**Data Storage and Retrieval: Knowledge of various data storage technologies**	3
5.	**Data Processing and Transformation:**	3
6.	**Data Integration and ETL (Extract, Transform, Load):**	2
7.	**Data Warehousing:**	3
8.	**Big Data Technologies:**	3
9	**Cloud Computing for Data Engineering:**	3
Total		**25**

| No	Lab Topics	Contact Hours
1	Reading and Writing Files and Working with Databases	3
2	Lebesgue Cleaning, Transforming, and Enriching Data	3
3	Building a 311 Data Pipeline	3
4	Features of a Production Pipeline	3
5	Version Control Using the NiFi Registry	3
6	Monitoring and Logging Pipelines. Deploying your Pipelines	4
7	Building a Production Data Pipeline, Building a Kafka Cluster	3
Total		22

D. Students Assessment Activities

No	Assessment Activities *	Assessment timing (in week no)	Percentage of Total Assessment Score
1.	Lab Work	Weekly	20%
2.	Midterm Exam	6	20%
3.	Final Project and Presentation	11	20%
4.	Final Exam	12	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 • 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.
Supportive References	• ACM (Association for Computer Machinery) Curricula Recommendations ⬚ - http://www.acm.org/education/curricula-recommendations
Electronic Materials	• Access to the Saudi Digital Library (SDL). • Using the learning management system of the university – Rafid System (https://lms.bu.edu.sa/).
Other Learning Materials	• Nielsen Norman Group: https://www.nngroup.com/

2. Required Facilities and equipment

Items	Resources
facilities (Classrooms, laboratories, exhibition rooms, simulation rooms, etc.)	• A classroom or lecture hall with a whiteboard for 25 students.

Items	Resources
Technology equipment (projector, smart board, software)	• A digital image projection system with a connection to a desktop computer and laptop computer. • High speed Internet connection
Other equipment (depending on the nature of the specialty)	**None**

F. Assessment of Course Quality

Assessment Areas/Issues	Assessor	Assessment Methods
Effectiveness of teaching	• Students • Faculty • Peer Reviewers • Program Leader • Course Coordinator	• Surveys (indirect). • Direct feedback from students. • Course evaluation by Peer Reviewers (indirect). • Class visit by Program Leader (indirect). • Comprehensive Course report (where we can find information about teaching difficulties and action plan)
Effectiveness of Students assessment	• Students • Faculty • Peer Reviewers • Program Leader • Exam Evaluation Committee • Course Coordinator	• Surveys (indirect). • Direct feedback from students. • Course evaluation by Peer Reviewers (indirect). • Class visit by Program Leader (indirect) • Exam evaluation by the Exam Evaluation Committee (indirect)
Quality of learning resources	• Students • Faculty • Peer Reviewers • Course Coordinator	• Surveys (indirect) • Course evaluation by Peer Reviewers (indirect). • Comprehensive Course report (where we can find information about difficulties and challenges about learning resources as well as consequences and action plan)
The extent to which CLOs have been achieved	• Faculty • Program Leader • Course Coordinator	• Student Results (direct) • Comprehensive Course report (where we can find the CLO assessment results)
Other		

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

Assessment Methods (Direct, Indirect)

G. Specification Approval

COUNCIL /COMMITTEE	Curriculum Committee
REFERENCE NO.	
DATE	28 April 2024