1.	Course title	Data Engineering and Analytics
2.	Course number	1915231
3.	Credit hours (theory, practical)	3
	Contact hours (theory, practical)	3
4.	Prerequisites/corequisites	Principles of Data Science (1915121) AI Programming (1915111)
5.	Program title	Data Science Program
6.	Year of study and semester (s)	Second Year, Fall
7.	Final Qualification	Bachelor degree
8.	Other department (s) involved in teaching the course	None
9.	Language of Instruction	English
10.	Date of production/revision	October 9, 2022
11.	Required/ Elective	Required

12. Course Coordinator:

Prof. Ibrahim Aljarah

Office numbers: KASIT 117
Office hours:
Monday: 11:30 – 12:30
Sunday, Tuesday: 11:30 – 12:30

Email addresses: i.aljarah@ju.edu.jo

13. Other instructors:

14. Course Description:

This course will examine the typical Data Engineering pipeline includes architecting data platforms, designing data stores, ETL, data collection, importing, wrangling, querying, and analysing data. It also includes performance monitoring and finetuning to ensure systems are performing at optimal levels. In addition, Data modelling and design techniques, Data storage and warehousing. Also, the course will discuss the popular data engineering tools such as Airflow. Furthermore, this course introduces you to the core concepts, processes, and tools you need to know to get a foundational knowledge of data engineering. You will gain an understanding of the modern data ecosystem and the role Data Engineers, Data Scientists, and

Data Analysts play in this ecosystem. This course will discuss the exploratory data analysis, feature generation and extraction. The course also includes hands-on labs and assignments that guide you to create and load data into the different types of databases, and perform some basic querying operations that help you understand your dataset. Lectures will be given in the lab for practical application. This course is assessed through exams, practical tests and assignments.

15. Course aims and outcomes:

A- Aims:

- The student will become familiar with data pipelines and how to build data systems to deploy and manage the data pipelines.
- Will become competent in managing extracting, transforming and loading (ETL) data from structured, semi-structured and unstructured data types and perform exploratory data analysis (EDA) techniques.
- Identify and transform data using different data wrangling techniques and build, train, and optimize predictive models
- The student will be aware of how to perform the Data Engineer role ethically and protect user's privacy and comply with opensource data licenses.

B- Intended Learning Outcomes (ILOs): Upon successful completion of this course students will be able to …

A-Knowledge and understanding with the ability to …

A1) Provide students with the typical Data Engineering pipeline includes architecting data platforms, designing data stores.

A2) The course helps to understand the ETL, data collection, importing, wrangling, querying, and analyzing data.

A3) Differentiate between the data engineer responsibilities, and data scientist responsibilities

A4) Study the data modelling, data manipulation and data warehousing 2

A5) The course helps to apply the exploratory data analysis process, feature generation and extraction. 2a.

16. Topic Outline and Schedule:

Topic	Week	ILOs	Student Outcomes	TLA (teaching, learning and Assessment)
Introduction to Data Engineering Data engineering and big data	1	A1, A2	1	T: Lectures and discussion L: [1] CH 01 A: Project, Midterm
Go with the flow Data engineers vs. data scientists The data pipeline	2	A1, A2	1	T: Lectures and discussion L: [1] CH 02 A: : Project, Midterm
Storing data Data structures SQL databases Data warehouses and data lakes	3	A2, A3	1	T: Lectures and discussion L: [1] CH 03 A: Project, Midterm
Moving and processing	4	A3, A4	1	T: Lectures and discussion

data Processing data Scheduling data Cloud computing				L: [1] CH 04 A: Project, Midterm
Data Loading: airflow Importing-data-from-csv-file Importing-data-from-excel-fil	5	A3	1	T: Lectures and discussion L: [1] CH 05 A: Project, Midterm
Data Loading: Importing-data-from-json-file Importing-data-from-mysql-sqlserver-databases	6	A4	1	T: Lecture and discussion L: [1] CH 06 and CH07 A: Project, Midterm
Midterm	7			
Data Manipulation	8	A4	1, 2	T: Lectures and discussion L: [2] CH 1 A: Practical Assessment, Final
Data Exploration, Joins, and Visualization	9	A4, A5	2	T: Lectures and discussion L: [2] CH 2 A: Practical Assessment, Final
Data Staging, Profiling, and Cleansing	10	A4, A5	2	T: Lectures and discussion L: [2] CH3 A: Practical Assessment,, Final
Feature Engineering	11	A4, A5	2	T: Lecture and discussion L: [2] CH 3 A: Practical Assessment, Final
Modelling 1	12	A4, A5	2	T: Lectures and discussion L: [2] CH 03 and CH09 A: Practical Assessment, Final
Modelling 2	13	A4, A5	2	T: Lecture and discussion L: [2] CH03 and CH09 A: Practical Assessment, Final
Use Cases	14	A1,A2,A3,A4, A5	1, 2	T: Lecture and discussion L: [2] CH10 A: Practical Assessment, Final
Revision	15			
Final	16			

17. Evaluation Methods and Course Requirements (Optional):

Opportunities to demonstrate achievement of the ILOs are provided through the following assessment methods and requirements:

There will be several assessment methods of evaluation the performance of the students such as project design, Practical Assessment, midterm and final exams.

18. Course Policies:

A- Attendance policies:

Deliberate abstention from attending 1911231 *classes and any other similar acts will lead to student deprivation from the course according to the UJ regulations*

B- Absences from exams and handing in assignments on time:

If you miss the midterm, then a makeup exam will not be provided unless you submit a valid absence excuse, within three days from the midterm, to your lecturer. This excuse must be signed and stamped from the UJ hospital in order to be valid. If your lecturer accepts the excuse, then you will be able to take the makeup. You need to follow up the departmental announcements regarding the makeup date and time. Please note that the lecturer may either accept or reject your excuse based on UJ regulations

C- Health and safety procedures:

N/A

D- Honesty policy regarding cheating, plagiarism, misbehavior:

All students in this course must read the University policies on plagiarism and academic honesty

http://registration.ju.edu.jo/RegRegulations/Forms/All_Regulations.aspx

E- Grading policy:

- *Midterm Exam:*	*30%*
- *HomeWorks*	*10%*
- *Practical Assessment*	*10%*
- *Final Exam:*	*50%*

F- Available university services that support achievement in the course:

N/A

G- Statement on Students with disabilities

Students with Disabilities: Students with disabilities who need special accommodations for this class are encouraged to meet with the instructor and/or their academic advisor as soon as possible. In order to receive accommodations for academic work in this course, students must inform the course instructor and/or their academic advisor, preferably in a written format, about their needs no later than the 4th week of classes.

19. Required equipment:

Python 2.7

Airflow

20. References:

A- Required book (s), assigned reading and audio-visuals:

1. *Data Engineering with Python, Paul Crickard, 2020*
2. *Python Data Cleaning Cookbook, Michael Walker 2020*

21. Additional information:

Course website:
elearning.ju.edu.jo

Date: ----------

Name of Course Coordinator: Dr. Khair Eddin Sabri Signature: ----------

Head of curriculum committee/Department: ---------- Signature: ----------

Head of Department: ---------- Signature: ----------

Head of curriculum committee/Faculty: ---------- Signature: ----------

Dean: ---------- -Signature: ----------

Copy to:
Head of Department
Assistant Dean for Quality Assurance
Course File