Course Description Form

1. Course Name:
Big Data
2. Course Code:
3. Semester / Year:
First Semester/MSc
4. Description Preparation Date:
1-9-2025
5. Available Attendance Forms:
Traditional Attendance (in-person) Blended Attendance
6. Number of Credit Hours (Total) / Number of Units (Total)
30/2
7. Course administrator's name (mention all, if more than one name)
Dr. Suhad A. Yousif Email:suhad.a.yousif@nahrainuniv.edu.iq
8. Course Objectives

Course Objectives	• Fundamental Understanding of Big Data Concepts: Ensure students grasp key concepts such as data volume, velocity, variety, and veracity. Discuss the evolution and significance of big data in modern computing. • Proficiency in Big Data Technologies: Teach students about various big data technologies and platforms like Hadoop, Spark, NoSQL databases, and cloud-based big data solutions. • Data Analytics and Processing Skills: Develop students' ability to perform data analytics, including the use of statistical methods, machine learning algorithms, and data mining techniques suitable for large datasets. Data Management and Storage: Educate on effective methods for storing, retrieving, and managing big data, including distributed file systems, data warehousing, and data lakes. • Practical Application and Problem Solving: Provide hands-on experience through projects or case studies that involve real-world big data challenges, encouraging students to apply their knowledge to solve practical problems.

9. Teaching and Learning Strategies

Strategy	• Flipped Classroom: Assign foundational material (readings, tutorials, or recorded lectures) for students to review before class. • Use in-class time for practical activities like coding, problem-solving, or case study discussions. • b. Hands-On Lab: Provide datasets and tasks requiring students to use tools like Apache Spark, Hadoop, SQL, Python (Pandas, NumPy). • c. Project-Based Learning • Assign semester-long group projects where students design, implement, and present a Big Data solution. • Include datasets and scenarios from domains like healthcare, finance, or social media to diversify learning. • e. Peer Learning: Encourage group activities, discussions, and code reviews to enhance collaborative skills and knowledge sharing.

10. Course Structure

Week	**Hours**	**Required Learning Outcomes**	**Unit or subject name**	**Learning method**	**Evaluation method**
Week	Hours	ILOs	Unit/Module or Topic Title	Teaching Method	Assessment Method
1	2	What is Big data?	✓ Characteristics of Big Data (Volume, Velocity, Variety, Veracity, Value) ✓ Importance and Applications of Big Data	**Lectures Presentation**	**Lab for coding**

2	2	Big Data Technologies and Ecosystem	✓ Hadoop, Spark, NoSQL databases, and other key technologies ✓ Overview of Big Data processing frameworks ✓ Understanding data storage options	=	=
3	2	Distributed File Systems and Data Storage	✓ HDFS (Hadoop Distributed File System) ✓ Data replication and fault tolerance ✓ Storage strategies for big data	=	=
4	2	MapReduce and Hadoop	✓ Introduction to MapReduce ✓ Hadoop ecosystem components ✓ Writing MapReduce programs	=	=
5	2	Apache Spark	✓ Introduction to Spark ✓ RDDs (Resilient Distributed Datasets) ✓ Spark transformations and actions	=	=

6	2	Introduction to Data Analytics	✓ Types of analytics (Descriptive, Diagnostic, Predictive, Prescriptive) ✓ Data mining and machine learning ✓ Data visualization for insights	=	=
7	2	Advanced Analytics with Spark	✓ Machine learning libraries in Spark	=	=
8	2	Advanced Analytics with Spark	✓ Building and evaluating predictive models' performance.	=	=
9	2	NoSQL Database Concepts	✓ Introduction to NoSQL databases ✓ Types of NoSQL databases (Document, Key-Value, Column-Family, Graph)	=	=
10	2	Hands-on with NoSQL Databases	✓ A comparison between common NoSQL databases.	=	=

11	2	Cloud Computing and Big Data	✓ Introduction to cloud platforms (AWS, Azure, GCP) ✓ Deploying big data solutions in the cloud ✓ Cost considerations and scalability	=	=
12	2	Mid Exam		=	=
13	2	Project Kickoff	✓ Explanation of the course project ✓ Choosing a dataset and problem statement ✓ Forming project teams	=	=
14	2	Project Work Sessions	✓ Guided project work sessions ✓ Project presentations and feedback	=	=

11. Course Evaluation

20% for the formal final writing exam
10 % for the practical exam
70 % for the formal final writing exam

12. Learning and Teaching Resources

Required textbooks (curricular books any)	• **"Big Data: Principles and Best Practices of Scalable Real-Time Data Systems"** • Author: Nathan Marz and James Warren

	• Focus: Core concepts and architecture of Big Data systems, real-time data processing. • **"Big Data Processing with Apache Spark"** • Author: Srini Penchikala • Focus: Hands-on guidance on using Apache Spark for processing large datasets.
Main references (sources)	• **"Mining of Massive Datasets"** • Authors: Jure Leskovec, Anand Rajaraman, and Jeffrey D. Ullman • Focus: Algorithms and techniques for analyzing large-scale data.
Recommended books and references (scientific journals, reports…)	**Big Data and Business Analytics"** • Editors: Jay Liebowitz • Focus: Applications of Big Data in business contexts. **Journal of Big Data** • Topics: Advances in Big Data analytics, machine learning, and applications. • Publisher: Springer. **IEEE Transactions on Big Data** • Topics: Research papers on Big Data systems, tools, and applications.
Electronic References, Websites	• **Coursera** Courses on Big Data by top universities (e.g., University of California San Diego's "Big Data Specialization"). • **edX** Big Data programs, such as MIT's "Data Science and Big Data Analytics.

	• **Apache Spark Documentation** URL: https://spark.apache.org/documentation.html • **Hadoop Documentation** URL: https://hadoop.apache.org/docs/ •