

Basavarajeswari Group of Institutions

2024 SCHEME

BALLARI INSTITUTE OF TECHNOLOGY & MANAGEMENT

(Autonomous Institute under Visvesvaraya Technological University, Belagavi)

USN

--	--	--	--	--	--	--	--	--

Course Code

M	M	C	4	1	1	2
----------	----------	----------	----------	----------	----------	----------

Fourth Semester MCA Degree Examinations, May/June 2026

DATA ENGINEERING

Duration: 3 hrs

Max. Marks: 100

Note: 1. Answer any FIVE full questions, choosing ONE full question from each module.
2. Missing data, if any, may be suitably assumed

<u>Q. No</u>	<u>Question</u>	<u>Marks</u>	<u>(RBTL:CO: PI)</u>
<u>MODULE – 1</u>			
1.	a. Explain the stages of the data engineering lifecycle with a neat diagram.	10	(2:1:1.4.1)
	b. Illustrate the role of data engineering in a real-world system and demonstrate the differences between data engineers and data scientists.	10	(3:1:2.5.2)
(OR)			
2.	a. Explain the concept of big data and describe its key characteristics (4Vs).	10	(2:1:1.4.1)
	b. Compare database, data lake, and data warehouse with suitable examples.	10	(3:1:2.5.2)
<u>MODULE – 2</u>			
3.	a. Explain the concept of data ingestion and describe the two primary approaches used in data ingestion.	10	(2:2:1.4.1)
	b. Compare the ETL and ELT approaches based on their features and applications.	10	(3:2:2.5.2)
(OR)			
4.	a. Explain the concept of data transformation and describe its major techniques.	10	(2:2:1.4.1)
	b. Apply the steps of data cleaning to a real-world dataset and illustrate the results.	10	(3:2:2.5.2)
<u>MODULE – 3</u>			
5.	a. Explain the architecture of a distributed system and how it supports big data processing.	10	(2:3:1.4.1)
	b. Apply Spark SQL to query a structured dataset loaded from a JSON file and explain the steps involved.	10	(3:3:2.5.2)
(OR)			
6.	a. Describe the components of the Hadoop ecosystem and explain their functions in data processing.	10	(2:3:1.4.1)
	b. Demonstrate how to read and process a CSV file using Spark DataFrames with appropriate options.	10	(3:3:2.5.2)

MODULE – 4

- | | | | | |
|----|----|---|----|-------------|
| 7. | a. | Explain with example Online Transaction Processing (OLTP). | 10 | (2:4:1.4.1) |
| | b. | Construct a star schema for a given business scenario and illustrate how fact and dimension tables are organized. | 10 | (3:4:2.5.2) |

(OR)

- | | | | | |
|----|----|--|----|-------------|
| 8. | a. | Explain with example Online Analytical Processing (OLAP). | 10 | (2:4:1.4.1) |
| | b. | Develop a snowflake schema for a given dataset and demonstrate how normalization is applied to dimension tables. | 10 | (3:4:2.5.2) |

MODULE – 5

- | | | | | |
|----|----|--|----|-------------|
| 9. | a. | Explain the core principles and key services of Amazon Web Services for data engineering. | 10 | (2:5:1.4.1) |
| | b. | Apply the concepts of AWS, Azure, and GCP to compare their suitability for different data engineering scenarios. | 10 | (3:5:2.5.2) |

(OR)

- | | | | | |
|-----|----|---|----|-------------|
| 10. | a. | Explain the core data engineering services and key responsibilities in google cloud platform. | 10 | (2:5:1.4.1) |
| | b. | Demonstrate how ETL processes can be implemented using tools available in google cloud platform | 10 | (3:5:2.5.2) |

**** ** ***