Pokhara University

Spring - 2025

Level: Bachelor		Year :	2025
Programme: BE		Full Marks:	100
		Pass Marks:	45
Course: Big Data (Elective)		Time :	3 hrs.

Candidates are required to give their answers in their own words as far as practicable.
The figures in the margin indicate full marks.

Attempt all the questions.

1a. Define the term Big Data. Discuss its key characteristics using the 5V model, supported by real-world examples. (7 marks)

1b. Define and differentiate between structured, semi-structured and unstructured data. Provide relevant examples to illustrate each category. (8 Marks)

2a. Identify and explain the major challenges associated with Big Data storage, processing and analysis.

2b. Define a distributed system. Illustrate how it differs from a centralized system in terms of architecture and performance. (8 Marks)

OR,

Discuss the role and significance of Big Data in contemporary industries such as healthcare, banking, logistics and e-commerce.
(8 Marks)

3a. Compare NOSQL systems with traditional relational databases in terms of scalability, schema design and consistency. (7 Marks)

3b. What is Apache Hadoop? Describe its architectural components and ecosystem. (8 Marks)

4a. Describe the MapReduce programming incdel. Illustrate its working with a relevant example. (7 Marks)

4b. Explain the architecture and working mechanism of the Hadoop distributed File System (HDFS). (8 Marks)

5a. Explain the limitations of Hadoop 1.x and describe how Hadoop 2.x addresses these issues. Highlight major architectural changes. (7 Marks)

5b. Define batch processing and real-time processing. Contrast their core principles and operational differences. (8 Marks)

OR,

Define data blocks in HDFS. Discuss how block size and replication contribute to fault tolerance and reliability. (8 Marks)

6a. Define data cleaning in the context of data preprocessing. Why is it considered foundational to data quality and analytics integrity? (7 Marks)

6b. Describe the architectural components of YARN, including the ResourceManager, NodeManager and Application Master. (8 Marks)

7. Write Short Notes (Any two). (2 * 5 = 10 Marks)

A. Data Visualization in Big Data Analytics

B. MapReduce vs Apache Spark

C. Vertical Scaling vs Horizontal Scaling.