

Roll No.

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

Total No. of Pages : 02

Total No. of Questions : 09

B.Tech. (CSE / DS / ECE / BCA) (Sem.-5,7,8)

INTRODUCTION TO BIG DATA

Subject Code : BTEC-909A-18

M.Code : 90683

Date of Examination : 07-12-2024

Time : 3 Hrs.

Max. Marks : 60

INSTRUCTIONS TO CANDIDATES :

1. SECTION-A is COMPULSORY consisting of TEN questions carrying TWO marks each.
2. SECTION-B contains FIVE questions carrying FIVE marks each and students have to attempt any FOUR questions.
3. SECTION-C contains THREE questions carrying TEN marks each and students have to attempt any TWO questions.

SECTION-A

1. Answer briefly :

- a) Mention one example of a technology or application that contributed to the evolution of big data.
- b) Explain the concept of data deluge and its impact on businesses and organizations.
- c) How does Hadoop's HDFS contribute to creating a scalable architecture for handling big data?
- d) What is the purpose of data partitioning in a distributed architecture for big data?
- e) Explain the concept of data replication and fault tolerance in high-performance big data systems.
- f) Explain the difference between supervised and unsupervised machine learning algorithms.
- g) How does exploratory data analysis contribute to uncovering insights from large datasets?
- h) Explain the importance of low-latency processing in real-time analytics applications.

- i) What are the challenges of ensuring data consistency and accuracy in real-time analytics?
- j) Mention one technology or tool used for real-time data visualization in analytics.

SECTION-B

- 2. Define predictive analytics and provide an example of its application.
- 3. Discuss the challenges of selecting an appropriate number of clusters in unsupervised clustering algorithms and propose strategies to address this issue effectively.
- 4. How is data streaming different from traditional batch processing in analytics? Explain the importance of low-latency processing in real-time analytics applications.
- 5. Explain the process of data replication in HDFS, including the significance of replication factor and how data reliability is maintained?
- 6. What factors contributed to the emergence of big data as a significant field in the tech industry?

SECTION-C

- 7. Describe the advantages and limitations of using JAQL for indexing and querying large-scale data compared to traditional indexing methods.
- 8. What is the purpose of a distributed file system like Hadoop's HDFS in big data architect?
- 9. Real time analytics platforms need to handle both structured and unstructured data sources. How do these diverse data formats introduce challenges in terms of data integration, storage, and analysis?

NOTE : Disclosure of Identity by writing Mobile No. or Making of passing request on any page of Answer Sheet will lead to UMC against the Student.