

Total No. of printed pages = 3

MCA 18250 E 34

Roll No. of candidate

2021

M.C.A. 5th Semester End-Term Examination

ELECTIVE III — BIG DATA ANALYTICS

(New Syllabus)

(New Regulation w.e.f 2018-19)

Full Marks - 70

Time - Three hours

The figures in the margin indicate full marks for the questions.

Answer question No. 1 and any *four* from the rest.

1. Choice the appropriate answer :

[Turn over

- (v) Which of the following are Benefits of Big Data Processing?
- (a) Business can utilize outside intelligence while taking decisions
 - (b) Improve customer service
 - (c) Better operational efficiency
 - (d) All of the above
- (vi) Which of the following are incorrect Big Data Technologies?
- (a) Apache Hadoop
 - (b) Apache spark
 - (c) Apache Kafka
 - (d) Apache Pytarch
- (vii) Apache kafka is an open-source platform that was created by?
- (a) Linkedin
 - (b) Facebook
 - (c) Google
 - (d) IBM
- (viii) What are the main components of Big Data?
- (a) MapReduce
 - (b) HDFS
 - (c) YARN
 - (d) All of the above
- (ix) Which of the following are the Goals of HDFS?
- (a) Fault detection and recovery
 - (b) Huge datasets
 - (c) Hardware at data
 - (d) All of the above
- (x) Face book tackles big data with _____ based on hadoop.
- (a) 'Project Prism'
 - (b) 'Prism'
 - (c) 'Project Big'
 - (d) 'Project Data'
2. (a) What is the difference between Structured and Unstructured data? (5)
- (b) What are the disadvantages of RDBMS over the switching techniques to Big Data? (3)
- (c) What is the formal definition of Big Data? Explain 5v's of the Big Data components? (2 + 5)
3. (a) What is hadoop? (2)
- (b) Why Hadoop is choose to be best for Big Data Technology? (3)
- (c) Explain with diagram the components of Hadoop Ecosystem and its Architecture? (10)

4. (a) What are the main components of Hadoop? (2)
(b) Explain the three core components of HDFS along with their functions? (8)
(c) "HDFS is file system design for storing very large files with streaming data access, patterns running on clusters of commodity hardware"-Justify the sentence with proper points. (5)
5. (a) What is block in HDFS structure? (2)
(b) State any three challenges of Big Data. (3)
(c) (i) Explain the reading and writing mechanism of HDFS with diagram? (10)

Or

- (ii) Explain the summary work of WALMART in implementing Big Data Technology.
6. (a) What do you mean by data ingestions? (2)
(b) Explain Apache Flume and its working mechanism? (5)
(c) Explain I/O notation of Hadoop? (8)
7. (a) MapReduce do not support data locality – Justify the statement. (3)
(b) How Hadoop runs MapReduce program explain with proper structure? (6)
(c) Using Map Reduce Word Count Algorithm count the words from below sentences.
(i) Dear Bear River Car, Car River Dear, Car Bear Dear
(ii) Welcome to Hadoop, Hadoop is supported over JVM, Hadoop is good programming Language. (6)

Total No. of printed pages = 2

MCA 202303

Roll No. of candidate

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

2021

M.C.A. 3rd Semester End-Term Examination INTRODUCTION TO DATA SCIENCE

Full Marks – 70

Time – Three hours

The figures in the margin indicate full marks
for the questions.

Answer question No. 1 and any *four* from the rest.

1. Answer the following (MCQ/ Fill in the blanks) : $(10 \times 1 = 10)$
- (i) What are probability and non-probability sampling techniques?
 - (ii) Define Range and Interquartile Range.
 - (iii) What is high dimensional data?
 - (iv) Name the two types of supervised learning techniques.
 - (v) How rule-based algorithm works?
 - (vi) Why clustering is done in Data Science?
 - (vii) What is block in deep learning?
 - (viii) Write the concept of layers and blocks in Deep Learning.
 - (ix) Why Neural Networks are used in Data Science?
 - (x) How important is Natural Language Processing in Data Science?
2. (a) Write the different techniques used for measuring central tendency and how they help in dealing with data. $(3+3 = 6)$
- (b) How CLIQUE algorithm works? (5)
 - (c) Write the basic functionalities of each block in a Custom Block. (4)

[Turn over

3. (a) What is supervised and unsupervised learning? Write the differences between them. $(2+2 = 4)$
- (b) Explain the architecture of Artificial Neural Network. Write the advantages of it. $(3+2 = 5)$
- (c) Why Natural Language Processing (NLP) is used? Write the advantage of NLP. $(2+2 = 4)$
- (d) Explain Binomial Distribution. (2)
4. (a) What is Sampling? Why it is used in Data Science? Write down the steps involved in Sampling. $(1+1+3 = 5)$
- (b) Describe and compare Euclidean, Manhattan and Mankowski distance in measuring similarity and dissimilarity. $(3+3 = 6)$
- (c) Write the differences between Semantic and Lexical Analysis. What is Pragmatic analysis in Natural Language Processing (NLP)? $(2+2=4)$
5. (a) Give the brief of different levels of Natural Language Processing (NLP). (5)
- (b) What is KNN? Write the intuitive steps for KNN. $(1+4 = 5)$
- (c) What is clustering? How K Mean clustering algorithm works? $(1+4 = 5)$
6. (a) Write the importance of descriptive statistics in Data Science? (4)
- (b) List the major tasks in data preprocessing. Explain briefly each of these tasks. $(2+4=6)$
- (c) How Feed forward Neural Network works? Explain the two phases of this network. $(3+2 = 5)$
7. (a) Why ensemble methods are required? Explain how Bagging works. $(2+3=5)$
- (b) What is MADALINE neural network? How MADALINE network is formed? $(2+3 =5)$
- (c) Write the importance of Correlation and Regression in Data Science. (5)