21/12/24 (FN)

Roll No.

ANNA UNIVERSITY (UNIVERSITY DEPARTMENTS)

B.E. /B.Tech / B. Arch (Full Time) - END SEMESTER EXAMINATIONS, NOV / DEC 2024

INFORMATION TECHNOLOGY
V Semester
ITM502 – BIG DATA ANALYTICS
(Regulation2019)

Time:3hrs Max.Marks: 100

CO1	Understand the basics of Big Data
CO2	Know about Hadoop and Map Reduce
CO3	Know about Big Data Technology, Tools and Algorithms
CO4	Analyze the Stream Data and Link Analysis
CO5	Know about the role of Big Data in Recommender Systems and Social Network Analysis
CO6	Design and Implementation of Basic data intensive Application

BL – Bloom's Taxonomy Levels
(L1-Remembering, L2-Understanding, L3-Applying, L4-Analysing, L5-Evaluating, L6-Creating)

PART- A(10x2=20Marks)
(Answer all Questions)

Q.No.	Questions	Marks	CO	BL
1	What is Big Data Analytics?	2	CO1	L1
2	Define Data Science.	2	CO1	L1
3	In which situations Distributed files system works better?	2	CO2	L2
4	Normally how many times chunks are replicated in distributed file system?	2	CO2	L2
5	Write the MongoDB command to create an index, where mobile_no is a field in the collection employee.	2	CO3	L1
6	Write the MongoDB command to find the document wherein the "StudName" has value "Kuttiappa".	2	CO3	L1
7	List the characteristics of Stream Data?	2	CO4	L1
8	What are the ways in which stream data arises naturally?	2	CO4	L2
9	What is meant by Social Network?	2	CO5	L1
10	Write the issues in on-Line Advertising.	2	CO6	L2

PART- B(5x 13=65 Marks)
(Restrict to a maximum of 2 subdivisions)

Q.No.	Questions	Marks	CO	BL
11 (a) i	Define big data and discuss the challenges in handling big data.	5	CO1	L3
ii	Discuss the various types of digital data, their sources, challenges in processing them, advantages and disadvantages of utilizing them.	8	CO1	L4
	OR			
11 (b) i	Define CAP theorem and discuss its implications on distributed system.	5	CO1	L3
ii	Explain the following technologies used in modern data processing and analytics: in-memory analytics, in-database processing, massively parallel processing and shared nothing	8	CO1	L4

	architecture.			
12 (a) i	Design MapReduce algorithm which take a very large file contains marks of the students in five subject and produce as output: a. The number students who scored more than 90 in all subjects. b. The number of students who scored less than 50 in a subject.	5	CO2	L4
ii	Explain in details the extensions to MapReduce to perform complex function.	8	CO2	L3
	OR			
12 (b) i	Design MapReduce algorithm which take a very large files contains the names of persons working in a company and produce as output: a. The number names starting with alphabet a. b. The number of same names.	5	CO2	L4
ii	Explain the idea behind implementing natural Join, grouping and Aggregation by MapReduce.	8	CO2	L3
13 (a)	Explain in detail how data is processed with Hadoop.	13	CO3	L3
	OR			
13 (b)	Explain in detail Hadoop Distributed file system.	13	CO3	L3
14 (a)	Write in detail about a technique to eliminate or filtering most of the tuples that do not meet the criterion in streaming data.	13	CO4	L3
	OR			
14 (b)	Write in detail about Frequent item sets, application of frequent item sets, and association rule in detail.	13	CO4	L3
15 (a)	Explain in details about advertising opportunities, Direct placement of Ads and Issues for display Ads.	13	CO5	L4
	OR			
15 (b)	With suitable example explain Girvan-Newman Algorithm and how communities are formed using betweenness.	13	CO5	L4

PART- C(1x 15=15 Marks)

(Q.No.16 is compulsory)

Q.No.	Questions	Marks	CO	BL
16.	Using PCA reduce the following two-dimension data into a single dimension data.	15	CO6	L5

Features	Ex1	Ex2	Ex3	Ex4
X	4	8	13	7
Y	-11	4	5	14