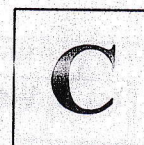


MSc. III(DS)/01.22.003

Reg. No

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



M.Sc. COMPUTER SCIENCE WITH SPECIALISATION IN DATA SCIENCE
THIRD SEMESTER EXAMINATION, JANUARY 2022
20-359-0316 BIG DATA ANALYTICS
(Regular)

Time : 3 Hours

Maximum Marks:50

(Answer ANY FIVE questions)
 Each question carries EQUAL Marks

		QUESTIONS	MARKS	CO	BL	PI
1	(a)	Which are the different types of financial frauds and how big data is used for Visually Analyzing Fraud in insurance sector?	5	CO1	L3	2.5.1
	(b)	Explain how distributed data models and computing models differ from each other.	5	CO4	L2	1.6.1
2.	(a)	Assume you have multiple files and each files contains two columns that represent the city and the corresponding temperature recorded in that city for the various measurement days. Write a map reduce program to find the maximum temperature for each city across all of the data files.	7	CO2	L3	1.7.1
	(b)	What is the purpose of the Map Reduce framework's shuffle and sort components? Is it possible to use this component in the case described above? If so, explain how.	3	CO2	L3	2.5.1
3.	(a)	Explain the different failure management techniques used in HDFS.	5	CO2	L2	1.6.1
	(b)	How resource monitoring is done in YARN?	5	CO5	L2	1.6.1

4.	(a)	Write a spark program to create a DStream using queueStream.	5	CO3	L3	1.7.1
	(b)	Which are the two types of transformations on Dstreams? Explain with an example.	5	CO3	L3	1.6.1
5.	(a)	Which are the different types of RDD based distributed matrix representations?	5	CO3	L2	1.6.1
	(b)	Write a spark program to perform the following operations i. Create an RDD with any two python collections. ii. Find the number of elements in the RDD iii. Create a new RDD which gives the square of the elements of the existing RDD.	5	CO3	L3	1.7.1
6.	(a)	Explain which are the different ways in which data can be distributed in No SQL databases?	4	CO6	L2	1.6.1
	(b)	Explain how HIVE outperforms traditional SQL in terms of query response time?	6	CO6	L2	2.5.1
7.	(a)	Write a Spark MLlib code for text classification problem using ML pipelines?	5	CO7	L3	1.7.1
	(b)	List and discuss the tools and techniques used for Big data visualization.	5	CO1	L2	1.6.1
