Reg. No.



A Constituent unit of MAHE. Manipal

VI SEMESTER B.TECH. (COMPUTER SCIENCE & ENGINEERING) END SEMESTER EXAMINATIONS, MAY 2024

SUBJECT: DATA WAREHOUSE AND DATA MINING [CSE 4060]

REVISED CREDIT SYSTEM (--/05/2024)

Time: 3 Hours

MAX. MARKS: 50

Instructions to Candidates:

✤ Answer ALL FIVE questions.

✤ Missing data may be suitably assumed.

QNo					CO/	AHEP	Blooms	
			Ouestion		Marks	CLO	LO	Taxonomy
			(Level
1A.	Formulate on the colu- family is a LastName 'karthik@g and (+) op the user_do with data beginning Operation user_id	4	2	2,5	5			
1B.	Using the sperform the command 'rajiv22'. (column 'do datatype o user_detail	4	2	2,5	5			
1C.	Formulate Employee: perform th Query Lan the value ' Search for Id_1 1	2	2	2,5	5			

2A.	Consider the o	collection named 'colours' shown in Table 2A. Each				
	document in	this collection is a 'colors' array. Formulate				
	MongoDB Qu	ieries to perform the following operations on this				
	array. Opera	tion 1 : Find the documents from the 'colours'				
	collection whi	ch have the element 'white' in the 1 st index position.				
	Ensure the out	put is formatted. Operation 2: Update the document				
	with id:5 a	nd push the colors 'black'.'white' into it using				
	addToSet. Or	eration 3: Update the document with 'id: 3' by				
	popping two e	lements from the list of elements present in the array				
	'colors' The el	lement popped are 'cyan' and 'burgundy' Operation				
	4: Find docum	pents in the colours collection with id:1 and display				
	the first two e	lements in the array 'colors'. Operation 5: Find all				
	documents fro	om the colours collection that have the elements	5	2	2.5	5
	'blue'.'purple'	in the array 'colors'		_	_,_	-
	, purpre					
		Table 2A				
	id:1	colors: ['black','white','grey','beige']				
	:10					
	_1d:2	_id:2 colors:['pink','blue','purple','orange']				
	_id:3	_id:3 colors:['green',' cyan','burgundy','				
	.1.4					
	_10:4	colors:['brown', magenta', purple', turquoise']				
	_id:5	colors:['maroon', 'peach', 'indigo']				
2B.	Describe the fe	3	1	1,4	6	
2C.	Outline the f	2	3	1,4	5	
2 4	Write o MonD		2	5	2	
3 A.	white a Mapk	5	3	5	3	
3R	Outline any 6	differences between Hive and RDBMS	3	3	14	5
3C		2	3	1,4	5	
50.	Summarize the functionalities of the NameNode in HDFS				1,7	5
4A.	Outline some of the challenges faced by researchers in the domain of Data Science			4	5	4
4B.	Describe the steps involved in creating a SPARK program			4	4	4
4C.	Outline 6 key advantages of Hadoon System			3	1,4	6
54	Elaborate on t	the different data formats supported by Spark With		Δ	4	5
571.	the help of an	example describe how the read and write operations	4	-	-	5
	are carried out	are carried out on a data frame in spark				
5B.	Describe the d	isadvantages of the RDD based countBvValue		4	4	6
	method in spar	k. With the help of an example illustrate how this	4	, r	•	
	disadvantage is overcome by the count method in DataFrame API					
5C.	With the help	•	4	4	6	
	deviation func	2		-	-	