Question Paper

Exam Date & Time: 13-Jul-2022 (09:00 AM - 12:00 PM)



SIXTH SEMESTER B.TECH END SEMESTER MAKE UP EXAMINATIONS, JULY 2022

BIG DATA ANALYTICS [ICT 4034]

Α

Marks: 50

Duration: 180 mins.

Answer all the questions.

Instructions to Candidates: Answer ALL questions Missing data may be suitably assumed

1)		Explain the architecture of Spark.	(5)
	A)		
B)		Explain the following with respect to Spark: i) RDD, ii) dataframes iii) Spark context	(3)
	C)	Netflix is the most loved American entertainment company specializing in online ondemand streaming video for its customers. Can you provide one suitable example each for various types of analysis, done on the Netflix data.?	(2)
2)		With an example explain the working of Spark stream.	(5)
	A)		
	B)	Explain in detail the main stages in Hadoop MapReduce data processing?	(3)
	C)	With atleast two suitable examples each, differentiate between Spark transformations and actions.	(2)
3)		Explain in detail MLlib Pipeline with an example of text document processing.	(5)
	A)		
	B)	When we visit the profile page of our friend we get list of common friends. Using the map reduce paradigm devise an algorithm map() and reduce() to give the list of such common friends.	(3)
	C)	List the benefits of processing Big data in a business environment.	(2)
4)		Following documents are inserted into inventory collection via suitable MogoDB query.	(5)
	A)	{ item: "journal", qty: 25, size: { h: 14, w: 21, uom: "cm" }, status: "A" },	
	,	{ item: "notebook", qty: 50, size: { h: 8.5, w: 11, uom: "in" }, status: "A" },	
		{ item: "paper", qty: 100, size: { h: 8.5, w: 11, uom: "in" }, status: "D" },	
		{ item: "planner", qty: 75, size: { h: 22.85, w: 30, uom: "cm" }, status: "D" },	
		{ item: "postcard", qty: 45, size: { h: 10, w: 15.25, uom: "cm" }, status: "A" }	
		{ item: "paper", qty: 100, size: { h: 8.5, w: 11, uom: "in" }, status: "D" },	
		{ item: "paper", qty: 75, size: { h: 22.85, w: 30, uom: "cm" }, status: "D" },	

	{ item: "paper", qty: 45, size: { h: 10, w: 15.25, uom: "cm" }, status: "A" }	
	Write the mongoDB query for the following tasks	
	To remove all items having status "D"	
	Toretrieve all documents in the collection where thestatusequals"A"orqtyis less than (\$lt)30:	
	Update the items whose name is "paper" and status is "A" to status as "D"	
	iv)to create aggregation pipeline to calculate the total order quantity of items whose status is "A"	
	v) find the total quantity of paper item.	
B)	What are the features of NOSQL devices?	(3)
C)	List the various Spark deployment modes	(2)
A)	A data file named student.json is available with details of students. Write the spark script to perform the following tasks: a) start a spark session b) read the contents of the given file c) display the contents of the file	(5)

B) Explain with suitable example various characteristics of Big data.

C) MLlib supports sparse matrices, in column-major order. Write the scala/ python code using MLlib to (2) get the following matrices.

				0	0	44	
				22	0	0	
	[11	0	0	0	0	33	
i.	lo	12	<mark>0</mark>] ii)	Lo	55	0	
.:							

5)

-----End-----

(3)