University of Mumbai
Examinations: Summer 2022

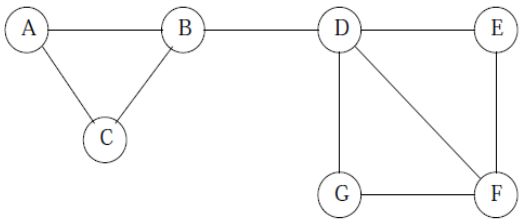
Time: **2 hour 30 minutes** Max. Marks: **80**

==========

Q1	**Choose the correct option for following questions. All the Questions carry equal marks. (2 Marks each) Total Marks - 20**
1	Vertical scaling means
Option A:	Adding computers serially
Option B:	Adding computers in parallel
Option C:	Adding computers serially as well as parallel
Option D:	Adding more processors, more memory and faster hardware typically within a single server
2	The original purpose of creation of Google implementation of MapReduce was to
Option A:	Perform matrix-vector multiplications for calculating PageRank
Option B:	Count number of keywords on webpages
Option C:	Count maximum number of keywords
Option D:	Find minimum keywords required on web page
3	Which of the following phases occur simultaneously?
Option A:	Map and combining
Option B:	Reduce and partitioning
Option C:	Shuffle and sort
Option D:	Map and Reduce
4	In DGIM algorithm, bucket cannot _____in size as we move to the left (back in time).
Option A:	increase
Option B:	decrease
Option C:	big
Option D:	small
5	Taxation assumes:
Option A:	A random surfer has a finite probability of leaving the Web at any step
Option B:	A random surfer has an infinite probability of leaving the Web at any step
Option C:	A random surfer has zero probability of leaving the Web at any step
Option D:	A random surfer has 50% probability of leaving the Web at any step
6	Following are the NoSql Business Drivers
Option A:	Data, Supply, Information, Idea
Option B:	Demand, Supply, Trends, Data
Option C:	Volume, Velocity, Agility, Variability
Option D:	Data, Information, Knowledge, Idea
7	Multistage algorithm uses
Option A:	1 hash functions in 2 different passes

Option B:	1 hash functions in 1pass
Option C:	2 hash functions in 1pass
Option D:	2 hash functions in 2 different passes
8	Two k-cliques are adjacent when they share
Option A:	2*k nodes
Option B:	k+1 nodes
Option C:	k-1 nodes
Option D:	k nodes
9	Assume that a text file contains the following text. In a map-reduce logic of finding frequency of occurrence of each word in this file, what is the output of map function? This is a exam Yes it is exam
Option A:	(This,1), (is, 2), (a, 1), (exam, 2), (Yes, 1), (it, 1)
Option B:	(This,1), (is, 1), (a, 1), (exam, 1), (Yes, 1), (it, 1), (is, 1), (exam,1)
Option C:	(This,1), (is, 1), (a, 1), (exam,1)
Option D:	(This,1), (is, 1), (a, 1), (exam, 2), (Yes, 1), (it, 1), (is, 1)
10	In a map-reduce logic of finding Matrix-Vector Multiplication, what is the output of the map function? 3 4 1 5 6 * 2 7 8
Option A:	(1,11), (2,17), (3,23)
Option B:	(1,1,3), (1,2,4), (2,1,5), (2,2,6), (3,1,7),(3,2,8)
Option C:	(1,1,3), (1,2,4), (2,1,5), (2,2,6), (3,1,7),(3,2,8), (1, 1), (2, 2)
Option D:	(1, 3), (1, 8), (2, 5), (2, 12), (3, 7), (3, 16)

Q 2.		**Attempt ANY TWO QUESTIONS out of THREE** **Each question is for 10 marks**
	A	Recall all NoSQL design patterns with examples. Justify CAP with suitable examples.
	B	Explain with example Collaborative based filtering in a recommendation system.
	C	Apply Matrix - Matrix Multiplication using MapReduce model and solve the following example 1 2 3 4 1 2 * 3 4
Q 3.		**Attempt ANY TWO QUESTIONS out of THREE** **Each question is for 10 marks**
	A	Apply PCY algorithm to find frequent itemset for the given dataset with minimum support 50% with hash function h(ij)= i*j % 8 T1-- 1,2,4,5 T2-- 2,4,5

		T3-- 1,2,4 T4-- 1,2,5
	B	Figure is an example of a social-network graph. Use the Girvan-Newman approach to find the between-ness of every edge. 
	C	Discuss all phases of the CURE algorithm for clustering with suitable example.
Q 4.		**Attempt any FOUR Questions out of SIX** **Each question is for 5 marks**
	A	What are the five Vs of Big Data? Explain.
	B	Recall Hadoop architecture with diagrams and give its advantages.
	C	Discuss any 5 different relational algebra operations with examples.
	D	Mention problems of PageRank along with its solution.
	E	State Bloom filter and explain with the help of an example.
	F	Explain KNN with proper example.