QP Code : 721200

(3 Hours) [Total Marks : 80

N.B. : (1) Question no. **1** is **compulsory**.
      (2) Attempt **any three** of remaining questions.
      (3) Assume suitable data wherever necessary.
      (4) Figure indicates marks.

1. (a) Calculate the storage efficiency of RAID level 1 ,3,5 and 6. The number of disk   **10**
drives available are 6 each having capacity of 200 GB. Also find the usable capacity
of each RAID level.

   (b) Consider a disk I/0 system in which an I/0 request arrives at a rate of 1001 IOPS. The
service time Rs=8ms. Calculate the following measures of disk performance:
        (1)   Utilisation of I/O controller (U)
        (2)   Total Response Time (R).
        (3)   Average Queue Size.
        (4)   Total time spent by request in the Queue.

   If the service time is half i.e. Rs=4ms and I/0 request arrives at a rate of 100 OPS,
then calculate the following measures of disk performance:
        (1)   Utilisation of I/0 controller (U)
        (2)   Total Response Time (R).
        (3)   Average Queue Size.
        (4)   Total time spent by request in the Queue.

2. (a) Explain in detail Fibre Channel Protocol Stack with neat diagram.   **10**
   (b) Explain ILM for Loan Management System with neat diagram.   **10**

3. (a) Explain in detail the different RAID levels with neat diagram. Also list out the   **10**
advantages and disadvantages for all RAID levels.
   (b) Explain the architecture of ISS in detail with neat diagram.   **10**

4. (a) Explain Block-level and File-level storage virtualization in detail with neat diagram.   **10**
   (b) Explain BC planning Lifecycle in detail with neat diagram. Give comparison between   **10**
RPO and RTO.

5. (a) List out the components of information system and its types. Explain different data   **10**
compression techniques in detail.
   (b) Explain Boolean Queries and Vector Queries in detail.   **10**

6. Write Short note (Any **four**)
      (a)   NDMP   (b) DAFS   (c) VIA   **20**
      (d)   Backup operations   (e) FC-SAN (f) Vector-Based Matching

----

— COMP. SIM. & MODELLING.

(3 Hours)   [Total Marks 80]

    i.   Q. 1. is Compulsory.
    ii.   Attempt any three from the remaining.
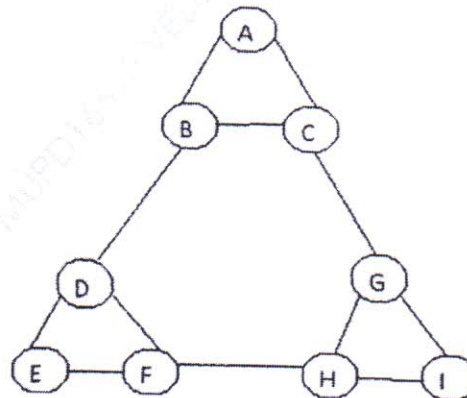   iii.   Assume suitable data.

Q.1 (a) What are the three Vs of Big Data? Give two examples of big data case   (5)
          studies. Indicate which Vs are satisfied by these case studies.

     (b) What is the role of a "combiner " in the Map reduce framework? Explain   (5)
          with the help of one example.

     (c) Through an example illustrate how the triangular array can be used to   (5)
          optimally store and count pairs in a frequent itemset mining algorithm.

     (d) List the different issues and challenges in data stream query processing. (5)

Q.2 (a) What are the different data architecture patterns in NOSQL? Explain "key (10)
          value" store and "Document" store patterns with relevant examples.

     (b) Show Map Reduce implementation for the following two tasks using   (10)
          pseudocode.
              i.   Multiplication of two matrices
              ii.   Computing Group-by and aggregation of a relational table.

Q.3 (a) Give a formal definition of the Nearest Neighbor problem. Show how   (10)
          finding plagiarism in documents is Nearest Neighbor problem. What
          similarity measures can be used.

     (b) Clearly explain the concept of a Bloom Filter with the help of an example. (10)

Q.4 (a) Suppose a data stream consists of the integers 3, 1, 4, 1, 5, 9, 2, 6, 5. Let (10)
          the hash function being used is $h(x) = 3x + 1 \mod 5$; Show how the Flajolet-
          Martin Algorithm will estimate the number of distinct element in this
          stream.

     (b) Clearly explain how the CURE algorithm can be used to cluster big data   (10)
          sets.

                                                    [Turn Over

Q. 5   (a)   Define Collaborative filtering. Using an example of an e-commerce site like   **(10)**
               Flipkart or Amazon describe how it can be used to provide
               recommendations to users.

        (b)   Define PageRank. Using the web graph shown below compute the   **(10)**
               PageRank at every node at the end of the second iteration. Use teleport
               factor = 0.8.
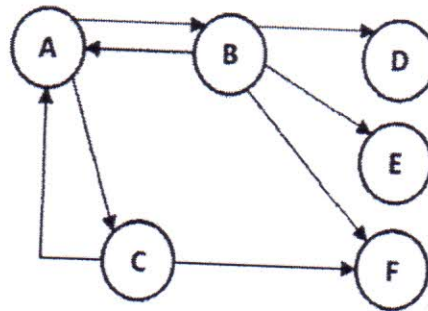
Q. 6   (a)   Explain clearly with diagrams how the PCY algorithm helps to perform   **(10)**
               frequent itemset mining for large datasets.

        (b)   For the graph given below use betweenness factor and find all communities   **(10)**

(3 Hours)   Total Marks :80

1. Question No. 1 is **compulsory**.
2. Out of remaining 5 questions, attempt any **three** questions.
3. Assume **suitable** data wherever required but **justify** the same.
4. **All** questions carry **equal** marks.
5. Answer to each new question to be started on a fresh page.
6. **Figure** to the **right** in brackets indicate **full** marks.
7. Use of statistical table is allowed.

Q1. Solve the followings.

(a) Define simulation. What are the various steps in simulation study? Explain each of them. Draw the flowchart of the same.   **(10)**

(b) Consider a single server system. Let the arrival distribution be uniformly distributed between 1 and 10 minutes and the service time distribution is as follows -

Service Time (min)	1	2	3	4	5	6
Probability	0.04	0.20	0.10	0.26	0.35	0.05

Develop the simulation table and analyze the system by simulating the arrival and service of 10 customers. Random digits for interarrival time and service time are as follows.

Customer	1	2	3	4	5	6	7	8	9	10
R.D. for Interarrival Time	---	853	340	205	99	669	742	301	888	444
R. D. for Service Time	71	59	12	88	97	66	81	35	29	91

Also calculate server utilization and maximum queue length.   **(10)**

Q2. Solve the followings.

(a) Explain the dump trucks problem in detail.   **(10)**

(b) Explain the replication method for steady state simulation.   **(10)**

Q3. Solve the followings.

(a) The interarrival times as well as service time at a single-chair unisex barbershop have been shown to be exponentially distributed. The values of λ and μ are 4 per hour and 6 per hour, respectively. Compute the steady-state parameters and the probabilities for zero, one, two, three, and four or more customers in the shop. **(10)**

(b) Explain Poisson process and state its properties.   **(10)**

[Turn Over]

Q4. Solve the followings.

    (a) Design a generator for weibull distribution. Using this generator get a weibull variate for $\alpha = 8$, $\beta = 0.75$, $v = 0$, and $R = 0.612$.   **(10)**

    (b) Explain in detail the tree step approach of Naylor and Finger in the validation process.   **(10)**

Q5. Solve the followings.

    (a) State the properties of random numbers. What are the problems or errors than can occur while generating pseudo random numbers? Use the mixed congruential method to generate a sequence of three two-digit random integers between 0 and 24 with $X_0 = 13$, $a = 9$, and $c = 35$.   **(10)**

    (b) Discuss the various issues in manufacturing and material handling system's simulation.   **(10)**

Q6. Solve the followings.

    (a) Explain the time series input models.   **(08)**

    (b) The highway between Mumbai, Delhi, and Calcutta, Delhi, has a high incidence of accidents along its 100 kilometers. Public safety officers say that the occurrence of accidents along the highway is randomly (uniformly) distributed, but the news media say otherwise. The Delhi Department of Public Safety published records for month of June. These records indicated the point at which 30 accidents involving an injury or death occurred as follows (the data points represent the distance from the city limits of Mumbai):

88.3	40.7	36.3	27.3	36.8	91.7	67.3	7.0	45.2	23.3
98.8	90.1	17.2	23.7	97.4	32.4	87.8	69.8	62.6	99.7
20.6	73.1	21.6	6.0	45.3	76.6	73.2	27.3	87.6	87.2

Use the Kolmogorov-Smirnov goodness of fit test to determine whether the distribution of location of accidents is uniformly distributed for the month of June. Use a level of significance of $\alpha = 0.05$.   **(12)**

(3 Hours) [Total Marks : 80

N.B. : (1) Question No. 1 is **Compulsory.**
 (2) Solve any **Three** questions out of remaining **Five** questions.

1. (a) What is activation function? Discuss the role of Sigmoidal activation function 5
 in backpropagation.
 (b) What is Fuzzy logic? Give an example. 5
 (c) Give the application scope of Neural Networks. 5
 (d) Mention two 'classical' dissimilarities of 'GA' and 'traditional' search algorithms. 5

2. (a) Explain typical application of the error back-propagation algorithm for 10
 handwritten character recognition in detail.
 (b) Two fuzzy relations are given by 10

R=		y_1	y_2
	x_1	0.6	0.3
	x_2	0.2	0.9

S=		z_1	z_2	z_3
	y_1	1	0.5	0.3
	y_2	0.8	0.4	0.7

Obtain fuzzy relation T as a max-min composition and max-product
composition between the fuzzy relations.

3. (a) Explain in detail the Crossover operator in Genetic Algorithm. 10
 (b) Explain McCulloch Pitts neuron model with example. 10

4. (a) Implement OR function using Perceptron networks for bipolar inputs and 10
 targets.
 (b) Explain Projection and Cylindrical Extension of Fuzzy Relations in detail. 10

5. (a) Explain Bidirectional Associative Memory in detail. 10
 (b) Explain the architecture of ANFIS with the help of a diagram 10

6. Write short note on:
 (a) Define Fuzzy Set. How it is different from Crisp Set? 5
 (b) Applications of Hybrid system. 5
 (c) Radial Basis Function 5
 (d) Delta Learning Rule 5

_____