COLLEGE OF ENGINEERING, PUNE

(An Autonomous Institute of Government of Maharashtra.)

END Semester Examination

Programme:	**B.Tech**	**Semester:**	**VI**
Course Code:	**CT-21012**	**Course Name:**	**Data Science**
Branch:	**Computer Engineering**	**Academic Year:**	**2020-21**
Duration:	**3 Hours**	**Max Marks:**	**60**

Student MIS No. | | | | | | | | | |

Instructions:

1. Figures to the right indicate the full marks.
2. Mobile phones and programmable calculators are strictly prohibited.
3. Writing anything on question paper is not allowed.
4. Exchange/Sharing of stationery, calculator etc. not allowed.
5. Write your MIS Number on Question Paper.

Q1 a

Sample	A1	A2	A3	Target Class
1	T	T	1.0	+
2	T	T	6.0	+
3	T	F	5.0	-
4	F	F	4.0	+
5	F	T	7.0	-
6	F	T	3.0	-
7	F	F	8.0	-
8	T	F	7.0	+
9	F	T	5.0	-

i) Calculate information gain for attributes A1, A2.
ii) Calculate information gain for every possible split of attribute A3
iii) Using answer for (i) and (ii) which is the best split among attribute A1, A2 and A3.

Marks: 7 | CO: 1,2 | PO: a,d

b Match the following:

a) Clustering	i) uses windowing concept to study data
b) Classification	ii) detects that certain queries follow certain other queries with a high probability
c) Sequential association analysis	iii) group results with a similar theme
d) Anomaly detection techniques	iv) discover unusual patterns

Marks: 3 | CO: 1,2 | PO: a,d

3 (5 2 4)

e) Regression model
f) Datastream
v) predicting real valued target variable
vi) assign results to pre-defined categories

Q2 **a** Normalize the given dataset to make the norm of each data point equal to 1. Use Euclidean distance to rank the other data points from point X_1. — 5 — 3 — d

	A1	A2
X_1	1.5	1.7
X_2	2	1.9
X_3	1.6	1.8
X_4	1.2	1.5
X_5	1.5	1.0

b Categories the following systems into structured and unstructured data use case: Sound recognition system, Railway reservation system, Image recognition system, Inventory control system — 2 — 3 — d

c Use a χ^2 test to find whether gender and their choice of pet (cat or dog) are co-related or not? Use the confusion matrix given below and alpha = 0.001. Chi-square table required for calculation is given at end of question paper. — 4 — 3 — d

	Cat	Dog
Male	207	282
Female	231	242

Q3 **a** Find all quartiles (Q1, Q2, Q3), range, inter quartiles range for the dataset given below and draw box-plot with whiskers using the obtained result and comment on its distribution. — 5 — 4 — f,d

59, 60, 61, 62, 62, 63, 63, 64, 64, 64,,65, 65, 65, 65, 65, 65, 65, 65, 65, 66, 66, 67, 67, 68, 68, 69, 70, 70, 70, 70,,70, 71, 71, 72, 72, 73, 74, 74, 75, 77.

b Plot a scatter plot for the dataset given below. Comment on the type of co-relation between two variable 'x' and 'y'. — 2 — 4 — f,d

x	100	90	80	80	70	65	60	60	55	45	35
y	8	10	13	10	12	18	15	20	23	25	27

Q4	a	Consider the data given below: Use linear regression to estimate the target variable 'Y' as a function of the input feature 'X'. i) Which of the following given θ parameter is best comparatively. Values: ($\theta_0 = 6, \theta_1 = 4$), ($\theta_0 = 4, \theta_1 = 4$). ii) Plot the hypothesis for best θ parameters of (i) also find the cost function using it. iii) Use answer of (i) to evaluate $h_\theta(x = 8)$	7	5	f

X	Y
1	12
2	12
3	20
4	21
5	25
6	28

	b	Compare and comment on Gradient Descent optimisation algorithm for: i) cost function minimization when applied in linear regression model ii) minimization of any other arbitrary function	3	5	f
Q5	a	Consider the following set of documents: **Doc1**: I am Ram **Doc2**: Ram I am **Doc3**: I am Ram I like sweet potato and milk For each document create 2-shingles (word level). Make use of shingles created to find the jaccard similarity between document Doc2-Doc3	2	5	f
	b	Consider the dataset given above. Use Apriori algorithm to find out frequent items	9	5	f

TID	Items
1	A, B, C
2	A, B
3	A, D, E
4	D, E
5	C, E
6	A, D, E

		and generate rules using those frequent items. Required threshold for support = 33.34% and confidence = 60%			
Q6	a	Consider bloom filter of size 11 with two hash functions h_1 and h_2 as given below. $h_1(x)$: take odd numbered bits from right in the binary representation of 'x'. Add all the bits and consider it as an integer 'i' and final value returned by hash function will be i % 11. $h_2(x)$: take even numbered bits from right in the binary representation of 'x'. Add all the bits and consider it as an integer 'i' and final value returned by hash function will be i % 11. Input stream consists of integers. i) Show how bloom filter is updated step by step as input stream 25, 159, 585 is received. ii) Check whether number 118 is passed through the bloom filter or not.	5	3	d
	b	Explain the data stream model with a neat diagram. List and explain different types of queries that are used in streaming system	6	3	d

Chi- Square table required for Q2) c) is given below

	P										
DF	0.995	0.975	0.2	0.1	0.05	0.025	0.02	0.01	0.005	0.002	0.001
1	.0004	.00016	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.55	10.828
2	0.01	0.0506	3.219	4.605	5.991	7.378	7.824	9.21	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.86	16.924	18.467
5	0.412	0.831	7.289	9.236	11.07	12.833	13.388	15.086	16.75	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.69	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.18	11.03	13.362	15.507	17.535	18.168	20.09	21.955	24.352	26.124
9	1.735	2.7	12.242	14.684	16.919	19.023	19.679	21.666	23.589	26.056	27.877
10	2.156	3.247	13.442	15.987	18.307	20.483	21.161	23.209	25.188	27.722	29.588
11	2.603	3.816	14.631	17.275	19.675	21.92	22.618	24.725	26.757	29.354	31.264
12	3.074	4.404	15.812	18.549	21.026	23.337	24.054	26.217	28.3	30.957	32.909
13	3.565	5.009	16.985	19.812	22.362	24.736	25.472	27.688	29.819	32.535	34.528
14	4.075	5.629	18.151	21.064	23.685	26.119	26.873	29.141	31.319	34.091	36.123
15	4.601	6.262	19.311	22.307	24.996	27.488	28.259	30.578	32.801	35.628	37.697
16	5.142	6.908	20.465	23.542	26.296	28.845	29.633	32	34.267	37.146	39.252
17	5.697	7.564	21.615	24.769	27.587	30.191	30.995	33.409	35.718	38.648	40.79
18	6.265	8.231	22.76	25.989	28.869	31.526	32.346	34.805	37.156	40.136	42.312
19	6.844	8.907	23.9	27.204	30.144	32.852	33.687	36.191	38.582	41.61	43.82
20	7.434	9.591	25.038	28.412	31.41	34.17	35.02	37.566	39.997	43.072	45.315

COLLEGE OF ENGINEERING, PUNE

(An Autonomous Institute of Government of Maharashtra.)

END Semester Examination

Programme: B.Tech **Semester: VI**

Course Code: (CT-21014) **Course Name: Design and Analysis of Algorithms**

Branch: Computer Engineering **Academic Year:** : 2021-2022

Duration: 3 Hours. **Max Marks: 60**

Student PRN No.

Instructions:

1. Figures to the right indicate the full marks.
2. Mobile phones and programmable calculators are strictly prohibited.
3. Writing anything on question paper is not allowed.
4. Exchange/Sharing of stationery, calculator etc. not allowed.
5. Write your PRN Number on Question Paper.

			Marks	CO	PO
Q 1	a	Solve following using Substitution method. Validate complexity using Master's theorem $T(n) = 3T\left(\frac{n}{4}\right) + n$	3	1,2	1
	b	Solve following using Recursive tree method $T(n) = T\left(\frac{n}{3}\right) + T\left(\frac{2n}{3}\right) + n$	3	1,2	1
	c	What is the time complexity of the following code write in detail :	4	1,2	1

i. Assume value of i greater than 1

```
while (i > 0) {
        a += i;
        i /= 2;
}
```

ii.

```
function(int n){
int i, j, count=0;
for (i=n/2;i<=n;i++)
  for (j=1;j<=n;j=j*2)
    count++;
}
```

Q 2	A	If we perform merge sort on the array of integer [5,3,8,9,1,7,0,2,6,4]. What is the number at the 6th position of partially sorted array [array start with index as 0] after the outermost two recursive calls have completed? [i.e. just before the very last merge step]	2	2,3	2

b Draw the schedule for summer camp in such way that you can Schedule max activities in a single day. 3 2,3 1,3

Activities	Timing [Starting time – Finish time]	Cost Rs:
Swimming	8.00 am – 9.30 am	1000/-
Horse Ridding	10.00 am – 11.00 am	1200/-
Canoeing	9.30 am – 11.30 am	1100/-
Kayaking	11.00 am – 1.30 pm	600/-
Lunch	1pm - 2.30pm	500/-
Drawing	12.00 noon – 1pm	1300/-
Pottery	2.00 pm - 3.30 pm	1500/-
Tea Break	11.30 am - 12.00 noon	50/-

c Solve the following job scheduling problem (maximizing the profit by completing jobs before their deadlines) using greedy algorithm: 3 2,3 5

Number of jobs (with execution time 1 hour each): n = 6

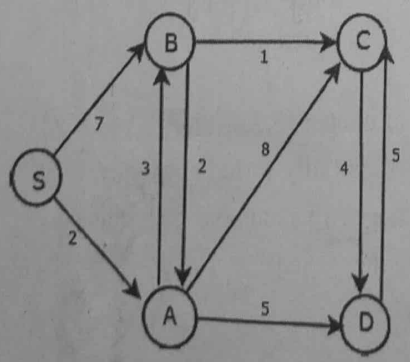
Profits associated with jobs: $(p_1, p_2, p_3, p_4, p_5, p_6) = (20,15,10,7,5,3)$

Deadlines (in hours) associated with jobs:$(d_1, d_2, d_3, d_4, d_5, d_6) = (3,1,1,3,1,3)$

Q3 A Compute and draw optimal parenthesization using Matrix Chain procedure for n = 4 matrices A1, A2, A3, A4 and their dimensions 2, 3, 5 and 4 A1 → 2 x 3, A2 → 3 x 5, A3 → 5 x 4, A4 → 4 x 2 5 3,4 2,5

b Write out the table for these inputs, and find the longest common subsequence: S1= GCCCT S2= GCGCAAT 3 3,4 2,5

c Find the all pair shortest path for the below graph, using dynamic programming approach 4 3,4 2,5



Q4	a	For Following Suppose T = 1011101110 , P = 111. Find all the Valid Shift/occurrences of pattern using naïve shift algorithm	2	4,5	2

OR

	a	Compute the prefix function Π for the pattern ababbabbabbababbabb when the alphabet is ∑ = {a, b}	2	4,5	2
	b	Calculate GCD(~~1124,84~~ 2322, 654) and Find integers p and q such that 2322p+654q= GCD(1124,84)	3	4,5	2
	c	Calculate X using Chinease remainder theorem. X=4mod 8 X=3 mod 17 X= 1 mod 13	5	4,5	1,5
Q5	a	Apply the branch and bound algorithm to solve TSP for following cost	8	4	1,5

	V1	V2	V3	V4	V5	V6
V1	∞	5	∞	6	5	4
V2	5	∞	2	4	3	∞
V3	∞	2	∞	1	∞	∞
V4	6	4	1	∞	7	∞
V5	5	3	∞	7	∞	3
V6	4	∞	∞	∞	3	∞

	b	Consider weights of items are w_i={5,10,15,20,25} and W=30 what are the different combinations possible? Solve problem by using backtracking approach.	2	4	1,5
Q6	a	Suppose we perform a sequence of n operations on a data structure in which the i^{th} operation costs i, if i is an exact power of 2 and 1 otherwise. Use aggregate analysis and accounting method to determine the amortized cost per operation.	2	4,5	1,5
	b	Inventory operator maintain a collection of lists that support the following operations. **i)** insert (item, list): insert item into list (cost = 1).	2	4,5	1

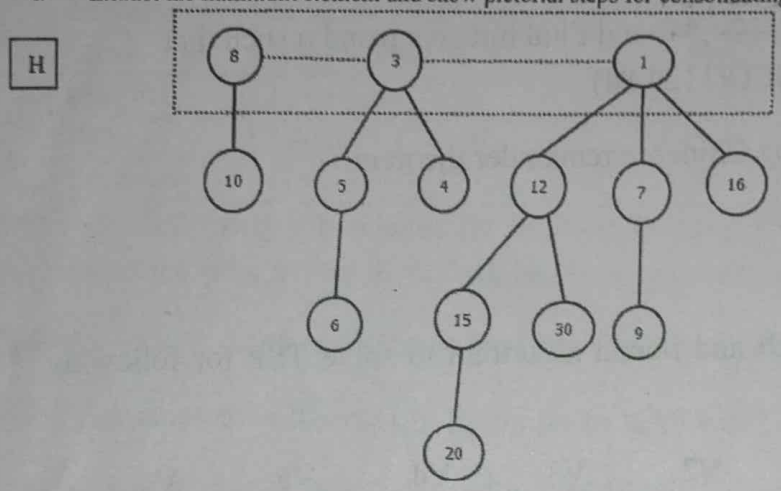
(ii) Sum (list): sum the items in list, and replace the list with a list containing one item that is the sum (cost = length of list).

Use the Potential Method to find that the amortized cost of an insert operation and sum operation

OR for complete Q6

Q6	a	For the following Fibonacci heap H,	4	4,5	1

i. Extract the minimum element and show pictorial steps for consolidating a new fib heap.



Q7	a	State whether this is true or false with justification	4	5	1,2

i. The problems 3-SAT and 2-SAT are NP-complete and in P, respectively.
ii. If a new problem is in NP and we can reduce a known NP-Complete Problem to it, then the new problem is NP – Complete
iii. The problem of determining whether there exists a cycle in an undirected graph is in NP.
iv. Hamiltonian Cycle problem belongs to P set.

	b	State which complexity can be executed in polynomial and non-polynomial time.	2	5	8

i. $O(2^{n./2})$ **ii.** $O(n)$ **iii.** $O(n^2 2^n)$ **iv.** $O(n \log n)$

----All The Best----

COLLEGE OF ENGINEERING, PUNE

(An Autonomous Institute of Government of Maharashtra.)

END Semester Examination

Programme: B.Tech | **Semester:** VI

Course Code: (CT-21015) | **Course Name:** Software Engineering Mini Project Stage- II

Branch: T.Y. B.Tech(Computer) | **Academic Year:** 2021-22

Duration: 2hrs | **Max Marks: 40**

Student PRN No.

1	1	1	9	0	3	0	9	3

Instructions:

1. Figures to the right indicate the full marks.
2. Mobile phones and programmable calculators are strictly prohibited.
3. Writing anything on question paper is not allowed.
4. Exchange/Sharing of stationery, calculator etc. not allowed.
5. Write your PRN Number on Question Paper.

SECTION-A

			Marks	CO	PO
Q 1	a	Describe the waterfall model of software development and discuss its strengths and weaknesses. For which of the following projects would it be suitable, and why? (a) An incremental compiler for Java (b) A clinical-record-keeping system for dentists (c) A word-processing package	5	1,4 ,5	a,d
	b	What are SRS document and design documents? How do they differ from each other?	5	1,2 ,3	c,d ,e
Q2	a	Define Software Architecture. Explain following architectural styles with examples. (a) Data-centered architecture (b) Call and return architecture	5	3,4	c,d
	b	What are the elements of Analysis model? Explain significance of each element with example.	5	1,2	a,b ,c,f

SECTION-B

Q 3	a	Calculate FPA(Function Point Analysis) if all CAF(Complexity Adjustment Factor) are average for the following count values EI=10, EO=30, EQ=50, EIF=10, ILF=20	5	4,5	b,c ,g
	b	Identify the classes for ATM system. What are the criteria to select right classes? Explain.	5	3,4 ,5	a,d ,g
Q 4	a	A project size of 200KLOC is to be developed. Software development team has nominal experience on similar type of projects and the project schedule is on high priority. Calculate the efforts development time of the project using intermediate COCOMO model.	5	4,5	a,d ,g
	b	Differentiate white box testing and black box testing	5	1	a,f

$0.65 + 0.01 \times 3 \times 14$

Total = 600

EI = 4	40
EO = 5	150
EQ = 4	200
ILF = 10	70
EIF = 7	200

3.2 1.05 3.0 1.12

COLLEGE OF ENGINEERING, PUNE

(An Autonomous Institute of Government of Maharashtra.)

END Semester Examination

Programme: B.Tech | **Semester: VI**

Course Code: CT(DE)-21001 | **Course Name: Advanced Data Structures**

Branch: Computer Engineering | **Academic Year: 2021 -2022**

Duration: 3 hours | **Max Marks: 60**

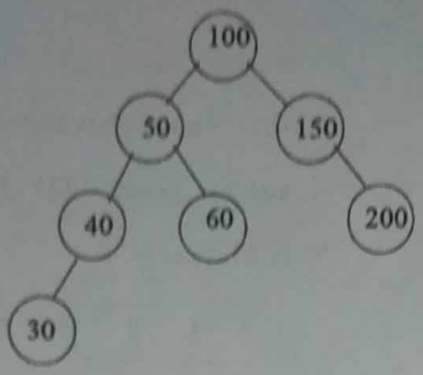
Student PRN No.

1	1	1	9	0	3	0	9	3

Instructions:

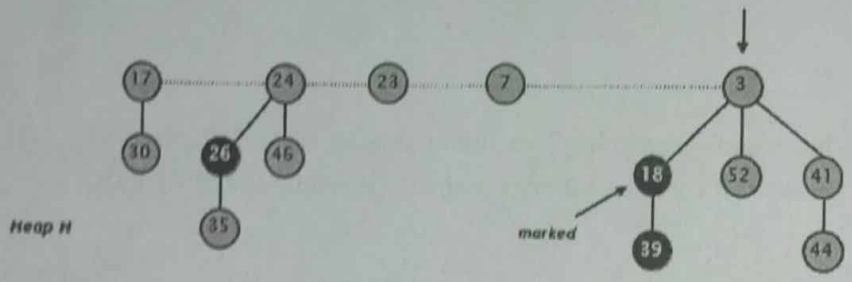
1. Figures to the right indicate the full marks.
2. Mobile phones and programmable calculators are strictly prohibited.
3. Writing anything on question paper is not allowed.
4. Exchange/Sharing of stationery, calculator etc. not allowed.
5. Write your PRN Number on Question Paper.
6. Draw neat and clear diagrams wherever required.
7. Answer all sub-questions together.
8. Write brief to the point answers.

			Marks	CO	PO
Q 1	a	Consider a binary max-heap implemented using an array. Which of the following arrays represents a binary maxheap? Assume the array index starts at 1. (a) 25,12,16,13,10,8,14 (b) 25,14,13,16,10,8,12 (c) 25,14,16,13,10,8,12 (d) 25,14,12,13,10,8,16	2	CO-1	PO-1
	b	Given the BST below, show the BST that would result after inserting the key of value 55 if splaying is performed starting at the node that was inserted.	2	CO-1	PO-1



Write the sequence of zig/zag operations performed during splaying.

c	Given a hashing table of 10 entries, the hash function h(x) = x mod (10) and the input sequence 71, 23, 73, 99, 44, 79, 89 Draw the hash table using Linear Probing. What is the load factor of the table?		2	CO-1	PO-1
d	What is the formula to calculate Potential of a Fibonacci Heap?		2	CO-2	PO-1



What is potential of heap H given in the figure above?

e	What are persistent data structures? Give any two real life applications of same.		2	CO-5	PO-1
Q 2	a	Solve the following recurrence relation: $T(n) = 2T(n/4) + \sqrt{n}$.	4	CO-2	PO-1
	b	Imagine a stack S with three operations: (i) PUSH(S, x) – pushes object x onto the stack in constant time (ii) POP(S) – pops and returns the top object of S in constant time (iii)MULTIPOP(S,k) – pops the top k items from the stack (or until empty) and returns the last item (complexity is $\Theta(\min\{\|S\|,k\})$	6	CO-2	PO-1

```
/* Internal implementation of Multipop is as follows */
while S is not empty and k > 0 do {
    x ← POP(S) ;
    k ← k − 1 ;
}
return x;
```

What is the worst-case complexity of a sequence of n operations?

What is the amortized complexity of a sequence of n operations? Use any method of your choice. State and explain the method you have used.

Q 3	a	Create an AVL tree by inserting the following keys one by one into an initial empty tree: 342, 206, 444, 523, 607, 301, 142, 183, 102, 157 and 149	4	CO-1	PO-1
	b	Into an empty min pairing heap, insert elements with priorities 20, 10, 5, 18, 6, 12, 14, 9, 8 and 22. From the resultant pairing heap, delete the minimum element using two-pass algorithm.	6	CO-1 CO-2	PO-1
	c	Joe, Bob and Sue are employees of the same corporation. The inventory of the products they sell is a database of over 1,000 entries where each entry has:	6	CO-2 CO-5	PO-4 PO-1

- A catalogue number which is a string of 6 characters: 3 letters followed by 3 digits.
- The number of products in stock for that catalogue number: an integer.
- The cost of the product: a real number.
- A description of the product: a variable length string of at most 20 characters.

Joe, Bob and Sue discuss a number of data structures that could be best applied to their individual needs. The choices are:

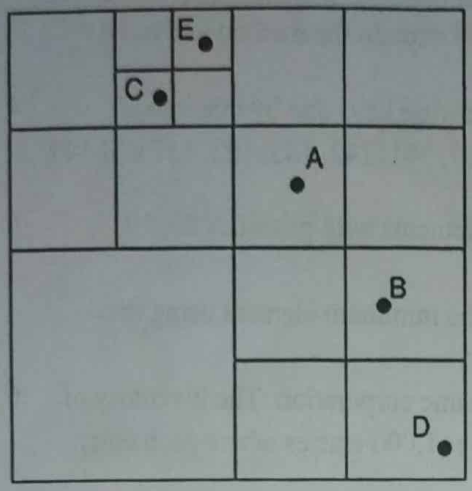
Arrays	Heaps	AVL Trees
Linked Lists	Splay Trees	Hash Tables
Binary Search Trees		

Which data structure will each one of them choose if their job description is as below? Explain in one or two sentences, the reason for his/her choice which will lead to the most efficient way to handle their task.

a) Joe is a salesperson. His job is to convince the customer to buy a product. Once the customer has chosen a product by looking at the printed catalogue, the only thing he needs is to find its availability and decrease the number in stock (if available).

b) Bob handles emergency situations. As soon as a product has its stock below a certain threshold he should reorder some from the manufacturer.

c) Sue has more of a managerial position. Her job is to look at the overall picture. She has already noted that 82% of the orders are for 17% of the products. She would like to periodically peek at a list ordered by product number of recently purchased items.

Q4	a	Draw a quadtree of the given image: 	4	CO-5	PO-1
	b	Imagine you are maintaining a quadtree of points, such that the points are being entered interactively. In order to make a quadtree, one has to assume the area covered by its root (as some square). How will you update the quadtree later when a point is provided that lies outside the square represented by the root?	4	CO-5	PO1-
	c	Explain the consequences of the following strategies of choosing the dividing line in a KD-tree (on its structure and the efficiency of searching whether a query point is present in the tree): (a) average of all the points (x-coordinates if vertical line, y-coordinates if horizontal line) (b) median of all the points.	6	CO-2	PO-1
Q5	a	Explain Van Emde Boas Layout and the motive behind it.	4	CO-2	PO-1

	b	Draw a partially persistent AVL tree using fat node method for the following operations performed in succession on an initially empty tree.
		
		insert(5)
		insert(7)
		insert(8)
		insert(6)
		traverse(v2) //inorder traversal
		traverse(v3)
			6	CO-1	PO-1