

Roll No.

TCS-501

B. TECH. (CSE) (FIFTH SEMESTER) END SEMESTER EXAMINATION, 2018

SYSTEM SOFTWARE

Time : Three Hours

Maximum Marks : 100

- Note : (i) This question paper contains two Sections.
(ii) Both Sections are compulsory.

Section—A

1. State True-False : (1×5=5 Marks)
 - (a) On the standard version of SIC, input and output are performed by transferring 1 byte at a time to or from the rightmost 8 bits of register A.
 - (b) Forward reference is a reference to a label defined later in the program.
 - (c) ESTAB is data structure used in Linking Loader.
 - (d) LEX is a Parser Generator.
 - (e) A text record can hold at most 60 bytes in SIC.

(2)

TCS-501

2. Attempt any five parts out of seven :

(3×5=15 Marks)

- Briefly explain two functions that are provided by debugging systems.
- Why does the assembler need to handle the source files in more than one passes ?
- Write the rules for writing the TEXT record in SIC.
- Illustrate the importance of static and dynamic linking with an example.
- Briefly explain the reason to generate the unique labels in Macro Processing.
- Write a LEX program to count the number of words in a given input.
- Explain the various instruction formats of SIC/XE.

Section—B

3. Attempt any two parts of choice from (a), (b) and (c). (10×2=20 Marks)

- What is program relocation ? Explain the different techniques that are used to make a program relocatable with their pros and cons.
- Consider the following program. Which technique will be efficient to write the relocatable object program for this

(3)

TCS-501

program ? Apply the answered technique and write the relocatable object program. (OPCODEs are given corresponding to mnemonics in the program)

0000	END1	START	0
0000	INLP	TD(E0)	INDEV E00018
0003		JEQ(30)	INLP 300000
0006		RD(D8)	INDEV D80018
0009		STCH	DATA 54001A (54)

000C	OUTLP	TD(E0)	OUTDEV E00019
000F		JEQ(30)	OUTLP 30000C
0012		LDCH(50)	DATA 50001A
0015		WD (DC)	OUTDEV DC0019
0018	INDEV	BYTE	X'F1' F1
0019	OUTDEV	BYTE	X'05' 05
001A	DATA	RESB	1

- Give the object program for the following SIC/XE program. (OPCODEs are given corresponding to mnemonics in the program, ASCII values are as S = 53, Y = 59, T = 54, E = 45, M = 4 D)

END2	START	1000
FIRST	LDT(74)	#6
	LDX(04)	#0
	LDCH(50)	STR1,X

	STCH(54)	STR2,X
STR1	BYTE	C'SYSTEM'
STR2	RESB	6
	END	FIRST

4. Attempt any two parts of choice from (a), (b) and (c). (10×2=20 Marks)

(a) Mention all the functions of Text Editors. Give a complete Block diagram of Text Editor and show its working after taking an example.

(b) Consider the following Macro :

```
RDBUFF MACRO &INDEV,
&BUFADR,
&RECLTH, &EOR,
&MAXLTH
IF (&EOR NE ' ')
&EORCK SET 1
ENDIF
CLEAR X
CLEAR A
IF (&EORCK EQ 1)
LDCH =X'&EOR'
RMO A,S
ENDIF
IF (&MAXLTH EQ ' ')
+LDT #4096
ELSE
+LDT #&MAXLTH
ENDIF
```

\$LOOP	TD	=X '&INDEV'
	JEQ	&LOOP
	RD	=X '&INDEV'
	IF	(&EORCK EQ 1)
	COMPR	A,S
	JEQ	#EXIT
	ENDIF	
	STCH	&BUFADR,X
	TIXR	T
	JLT	&LOOP
#EXIT	STX	&RECLTH
	MEND	

Invoke the above macro by the statement given as: RDBUFF F1, BUFF, RLENG, 04 and write the expanded code.

(c) Disassemble the following SIC object program (assume the meaningful name for labels) OPCODEs are JSUB-48, LDX-04, TD-E0, JEQ-30, RD-D8, STCH-54, TIX-2C, JLT-38, RSUB-4C.

HEND 0010000001AF

T0010001C481003041080E0101B301006

D8101B54901C2C10833810064C0000F1

T00108003000000

E001000

5. Attempt any *two* parts of choice from (a), (b) and (c). (10×2=20 Marks)

(a) Write the algorithm to show the working of one Pass Macro processor.

(b)

END3	START	0
CHARZ	BYTE	C'Z'
C1	RESB	1
FIRST	LDCH (50)	CHARZ
	STCH (54)	C1
END	FIRST	

- (i) Assemble the above SIC program and generate the relocatable object program, according to the file format using modification record. Assemble the same source program and generate the object program with relocation bit.
- (ii) When these two different formats of object file are stored in the hard disk, how many bytes do they occupy, respectively ? (Note: Do not forget to count the newline character).
- (c) Explain Expressions, Literals and Literal pools. Give the difference between immediate addressing and literals.

6. Attempt any *two* parts of choice from (a), (b) and (c). (10×2=20 Marks)

(a) What is parsing ? What are the different types of parsing ? Explain each operation of shift reduce parsing.

(b) Write a LEX/YACC specification file to check the validity of an arithmetic expression. Write the rules to evaluate the arithmetic expression.

(c) Differentiate the linkage editor and linking loader. Explain any *two* situations to show the importance of these. Describe the different data structures that are used to implement the Macroprocessor.

Roll No.

TCS-507

**B. TECH. (CS-BDA)
(FIFTH SEMESTER)**

END SEMESTER EXAMINATION, 2018

MACHINE LEARNING—I

Time : Three Hours

Maximum Marks : 100

- Note :** (i) This question paper contains two Sections.
(ii) Both Sections are compulsory.

Section—A

1. State True-False : (1×5=5 Marks)
 - (a) Naive Bayes is a statistical learning technique. (True/False)
 - (b) Cluster analysis is used for market segmentation. (True/False)
 - (c) There are multiple ways of measuring the distance between data points. (True/False)
 - (d) Artificial neural network is a machine learning technique. (True/False)
 - (e) Decision trees are essentially a hierarchy of if then statements. (True/False)

2. Attempt any five parts out of seven :
 (Define/Short Numerical/Short Programming/
 Draw) $(3 \times 5 = 15 \text{ Marks})$
- (a) What is a decision tree ? Why are decision tree the most popular classification technique ?
 - (b) What is a scatter plot ? How does it help ?
 - (c) What is unsupervised learning ? When is it used ?
 - (d) What is a confusion matrix ?
 - (e) What is standard deviation ? How is it computed ?
 - (f) What is the difference between mean, mode and median ?
 - (g) What is support vector machine ? What are support vectors ?

Section—B

3. Attempt any two parts of choice from (a), (b) and (c). $(10 \times 2 = 20 \text{ Marks})$

- (a) What is Machine Learning ? What are the key tasks of Machine Learning ?
- (b) Explain the steps required for selecting the right machine learning algorithm ? Explain recall, precision and F measure.

- (c) Design a Hebb net to implement logical AND function using bipolar inputs and targets.
- 4. Attempt any two parts of choice from (a), (b) and (c). $(10 \times 2 = 20 \text{ Marks})$
 - (a) Show that the derivative of unipolar sigmoidal function is :
$$f'(x) = \lambda f(x)[1 - f(x)].$$
 - (b) Using Hebb rule find the weights required to perform the following classification. The vectors $(1, -1, 1, -1)$ and $(1, 1, 1, -1)$ belongs to class target value + 1 vectors $(-1, -1, 1, 1)$ and $(1, 1, -1, -1)$ do not belong to class target value-1. Also using each of training x vectors as input test the response of net.
 - (c) Implement XOR function using McCulloch-Pitts neuron consider binary data.
- 5. Attempt any two parts of choice from (a), (b) and (c). $(10 \times 2 = 20 \text{ Marks})$
 - (a) Explain classification using back propagation algorithm with a suitable example.

(b) In what ways Naive Bayes is better than other classification techniques ? Compare with decision tree.

(c) Data about height and weight for a few volunteers is available. Create a set of 3 clusters for the following data using K means clustering :

(71, 165), (68, 165), (72, 180), (67, 113),
 (72, 178), (62, 101), (70, 150), (69, 172),
 (72, 185), (63, 149), (69, 132), (61, 115)

6. Attempt any two parts of choice from (a), (b) and (c). (10×2=20 Marks)

(a) Compare and contrast biological neuron and artificial neuron.

(b) Write difference between classification and clustering.

(c) Apply agglomerative hierarchical clustering on given data and draw dendrogram.

	A	B	C	D	E
A	0				
B	9	0			
C	3	7	0		
D	6	5	9	0	
E	11	10	2	8	0

Roll No.

TCS-509

**B. TECH. (CSE-CC)
(FIFTH SEMESTER)**

END SEMESTER EXAMINATION, 2018

MACHINE LEARNING

Time : Three Hours

Maximum Marks : 100

Note : (i) This question paper contains two Sections.

(ii) Both Sections are compulsory.

Section—A

1. State True or False : (1×5=5 Marks)
 - (a) Median is sensitive to outliers.
 - (b) Support vectors are the points that are nearest from the decision boundary.
 - (c) Logistic Regression is used for solving regression problem.
 - (d) Clustering is a problem under reinforcement learning.
 - (e) Entropy is a measure of impurity.

(2)

TCS-509

2. Attempt any *five* parts out of seven :
(Define/Short Numerical/Short Programming/
Draw) (3×5=15 Marks)
- What is the difference between k-means clustering and fuzzy c-means clustering ?
 - Derive Naive Bayes formula.
 - Draw and describe perceptron model of neuron.
 - When should we use normal equation and when should we use gradient descent for solving a linear regression problem and why ?
 - Define Gain Ratio.
 - Describe kernel trick with an illustration.
 - Define information gain.

Section—B

3. Attempt any *two* parts of choice from (a), (b) and (c). (10×2=20 Marks)
- Describe supervised and unsupervised learning approaches using suitable diagrams.
 - Write a Matlab/C/Python program for performing kNN.

Q6-20T

(3)

TCS-509

- (c) Find the separating hyperplane equation of the data below using SVM. Note that X and Y are features, and Z is the label.

	X	Y	Z
01	0	0	1
02	1	0	1
03	0	1	1
04	5	0	-1
05	5	5	-1
06	0	5	-1

4. Attempt any *two* parts of choice from (a), (b) and (c). (10×2=20 Marks)

- Describe reinforcement learning approaches using suitable diagrams and definitions.
- Write a Matlab/C/Python program for performing Naive Bayes classification.
- Give and explain cost functions of linear and logistic regression algorithms.

5. Attempt any *two* parts of choice from (a), (b) and (c). (10×2=20 Marks)

- Describe the three approaches to reinforcement learning.
- Write a Matlab/C/Python program for performing k-means clustering.

(4)

TCS-509

- (c) Find optimal linear regression parameters for fitting the data below using normal equation :

Feature 1	Feature 2	Label
1	16	10
2	19	20
5	18	30
4	17	40
3	20	50

6. Attempt any *two* parts of choice from (a), (b) and (c). (10×2=20 Marks)

- (a) Give different approaches to the validation process along with the comparison of the time required for carrying out each of them.
- (b) Draw an orange workflow that can display confusion matrices computed in all the three processes of supervised learning: training, validation, and testing. Simply use a rectangular block with widget's name inside it. You can use any supervised learning algorithm.
- (c) Following is the table depicting h5-indices of top publication venues in the Artificial

(5)

TCS-509

Intelligence research area as per Google Scholar. Divide them into three clusters using k-means algorithm based on h5-index :

Publication venue	h5-index
1. Neural Information Processing Systems (NIPS)	134
2. International Conference on Machine Learning (ICML)	113
3. Expert Systems with Applications	92
4. IEEE Transaction on System, Man and Cybernetics Part B, Cybernetics	88
5. IEEE Transactions on Neural Networks and Learning systems	87
6. The Journal of Machine Learning Research	79
7. Applied Soft Computing	77
8. IEEE Transactions of Fuzzy systems	74
9. Knowledge-based Systems	74
10. Neurocomputing	71
11. AAAI Conference on Artificial Intelligence	69

12. International Joint Conference on Artificial Intelligence (IJCAI) 61
13. Neural Networks 56
14. Neural Computing and Applications 54
15. Engineering Applications of Artificial Intelligence 50
16. Robotics and Autonomous Systems 49
17. Artificial Intelligence 46
18. Conference on Learning Theory (COLT) 44
19. International Conference on Artificial Intelligence and Statistics 43
20. Journal of Intelligent and Robotic Systems 42

Roll No.

TCS-502

**B. TECH. (CSE/IT)
(FIFTH SEMESTER)**

END SEMESTER EXAMINATION, 2018

OPERATING SYSTEMS

Time : Three Hours

Maximum Marks : 100

Note : (i) This question paper contains two Sections.
(ii) Both Sections are compulsory.

Section—A

1. State True/False : (1×5=5 Marks)
 - (a) Size of Logical address is always smaller than Physical address.
 - (b) The medium term scheduler only runs in times of high resource contention, as when physical memory is full.
 - (c) One of the four conditions necessary for deadlock is that a process acquire a resource and hold onto it while it acquires other resources.

(2)

TCS-502

- (d) In multicore systems, symmetric multiprocessing uses one core to handle the scheduling for all the other cores.
- (e) In non-distributed systems. Deadlock is rare enough that most systems just ignore it.
2. Attempt any five parts out of seven :

(3×5=15 Marks)

- (a) What is scheduling ? What criteria affect the scheduler's performance ?
- (b) What are the differences between paging and segmentation ?
- (c) Explain, why spinlocks are not appropriate for single processor systems yet are often used in multiprocessor systems.
- (d) What are the advantages of a multiprocessor system ?
- (e) Explain the difference between internal and external fragmentation.
- (f) Compare MFT and MVT memory management schemes.
- (g) What do you understand by belady's anomaly ?

(3)

TCS-502

Section-B

3. Attempt any two parts of choice from (a), (b) and (c). (10×2=20 Marks)
- (a) Describe the differences among short-term, medium-term and long-term schedulers. Also explain, which of them is most frequently invoked ?
- (b) Suppose you try to enforce mutual exclusion as follows :

Process 1	Process 2
<pre>while (true) { while (p2_in_crit_sect == true) /* do nothing */ p1_in_crit_sect = true; ... critical section ... p1_in_crit_sect = false; }</pre>	<pre>while (true) { while (p1_in_crit_sect == true) /* do nothing */ p2_in_crit_sect = true; ... critical section ... p2_in_crit_sect = false; }</pre>

Does this code always enforce mutual exclusion ? Justify your answer.

- (c) When the following program executed, all together how many processes are created ? Your count should include the initial process that runs the main() function. Show the process tree and the output :
- ```
Void main()
{
 Int pid;
```

```

pid=fork();
if(pid==0)
{
 Fork();
 Printf("hello");
}
Else
{
 Fork();
 Printf("world");
}
}

```

4. Attempt any *two* parts of choice from (a), (b) and (c). (10×2=20 Marks)

- (a) What is a race condition in critical section ? Explain, how a critical section avoids this condition. Explain all conditions that satisfy the solution to critical section problem.
- (b) The GEU memory architecture design team has a dilemma. The team is considering several different memory configuration variations for an upcoming machine design. Consider the following designs. (All memory accesses are in

terms of bytes and all are using paging techniques) :

| Characteristic | Design 1 | Design 2 |
|----------------|----------|----------|
|----------------|----------|----------|

|                               |       |        |
|-------------------------------|-------|--------|
| Physical memory address width | 8 bit | 16 bit |
|-------------------------------|-------|--------|

|                       |        |        |
|-----------------------|--------|--------|
| Logical address width | 12 bit | 20 bit |
|-----------------------|--------|--------|

|                          |          |          |
|--------------------------|----------|----------|
| Page/Frame size in bytes | 16 bytes | 32 bytes |
|--------------------------|----------|----------|

|                 |        |        |
|-----------------|--------|--------|
| Page Table type | Single | Single |
|-----------------|--------|--------|

(i) For each design, list the maximum number of pages each process can access in logical address space.

(ii) For each design, list the maximum number of frames in physical memory.

- (c) Consider the following table :

| Process | Arrival Time | Burst Time | Priority |
|---------|--------------|------------|----------|
| 1       | 01           | 10         | 3        |
| 2       | 03           | 07         | 2        |
| 3       | 04           | 04         | 4        |
| 4       | 02           | 09         | 1        |

For each of the following scheduling algorithms, determine the turnaround time

for each process and the average turnaround time for all jobs :

- (i) SJF (pre-emptive)
  - (ii) Round Robin (Time Slice = 2)
5. Attempt any two parts of choice from (a), (b) and (c). (10 $\times$ 2=20 Marks)

- (a) What do you understand by paging scheme ? Why are segmentation and paging sometimes combined into one scheme ?
- (b) How many page-faults would occur for the following reference string, for three page frame using LRU and Optimal algorithms ?

1, 2, 3, 4, 5, 5, 3, 4, 6, 7, 8, 7, 3, 9, 7, 6, 4, 5, 4, 5, 4, 2

- (c) Consider the following snap shot of system :

| Process | Allocation |   |   |   | Max |   |   |   |
|---------|------------|---|---|---|-----|---|---|---|
|         | A          | B | C | D | A   | B | C | D |
| P0      | 0          | 0 | 1 | 2 | 0   | 0 | 1 | 2 |
| P1      | 1          | 0 | 0 | 0 | 1   | 7 | 5 | 0 |
| P2      | 1          | 3 | 5 | 4 | 2   | 3 | 5 | 6 |
| P3      | 0          | 6 | 3 | 2 | 0   | 6 | 5 | 2 |
| P4      | 0          | 0 | 1 | 4 | 0   | 6 | 5 | 6 |

Let the available number of resources be given by AVAIL Vector as (1, 5, 2, 0). Use Banker's algorithm and answer the following :

- (i) What is the "Need Matrix" ?
- (ii) Determine whether system is in safe state or not ?

6. Attempt any two parts of choice from (a), (b) and (c). (10 $\times$ 2=20 Marks)

- (a) What is the cause of thrashing ? How does the system detect thrashing ? Once it detects thrashing, what can the system do to eliminate this problem ?
- (b) Consider a system with 8% hit ratio, 50 nano-seconds time to search the associative registers, 750 nano-seconds time to access memory. Find the time to access a page :

- (i) When the page number is in associative memory ?
  - (ii) When the time to access a page when not in associative memory ?
  - (iii) Find the effective memory access time.
- (c) Suppose that a disk drive has 5000 cylinders, numbered 0 to 4999. The drive

**TCS-502** (8)

**TCS-502**

is currently serving a request at cylinder 143 and the previous request was at cylinder 125. The queue of pending requests, in FIFO order, is :

86, 1470, 913, 1774, 948, 1509, 1022, 1750, 130

Starting from the current head position, what is the total distance (in cylinders) that the disk arm moves to satisfy all the pending requests, for each of the following disk-scheduling algorithms ?

- (i) SSTF
- (ii) SCAN

**TCS-502**

**570**

**TCS-503****B. TECH. (CSE) (FIFTH SEMESTER)  
END SEMESTER EXAMINATION, 2018****DATABASE MANAGEMENT SYSTEM****Time : Three Hours****Maximum Marks : 100**

Note : (i) This question paper contains two Sections.

(ii) Both Sections are compulsory.

**Section—A**

1. Fill in the blanks/True-False : ( $1 \times 5 = 5$  Marks)

(a) Create user is a system privilege.  
(True/False)

(b) A database is called "self-describing"  
because it contains a description of itself.  
(True/False)

(c) ..... is used in a query to replace null  
with an actual value.

(d) Database applications are seldom intended  
for use by a single user. (True/False)

(e) A database design may be based on  
existing data. (True/False)

(2)

TCS-503

2. Attempt any *five* parts out of seven :  
 $(3 \times 5 = 15 \text{ Marks})$

(a) What is Weak Entity ?

(b) R (A B C D E F)

$$AB \rightarrow C$$

$$BC \rightarrow AD$$

$$D \rightarrow E$$

$$CF \rightarrow B$$

Find  $(AB)^+$ .

(c) Which component of DBMS is responsible to maintain the ACID properties of transactions ?

(d) What is top-n analysis ?

(e) Why do we calculate closure of an attribute ?

(f) How can we represent *m:n* cardinality between two tables in E-R diagram ?

(g) What problems you will face without normalization in DBMS ?

### Section—B

3. Attempt any *two* parts of choice from (a), (b) and (c).  $(10 \times 2 = 20 \text{ Marks})$

(a) What is deadlock ? Give an example for deadlock considering any table.

(b) What is Functional Dependency ? Mention Armstrong Axiom with example.

TCS-503

(3)

TCS-503

- (c) Consider the following schema R = (A, B, C, D, E, F, G, H, I) and set of functional dependencies :

$$AB \rightarrow C$$

$$BD \rightarrow EF$$

$$AD \rightarrow GH$$

$$A \rightarrow I$$

Find the normal form of the relation.

4. Attempt any *two* parts of choice from (a), (b) and (c).  $(10 \times 2 = 20 \text{ Marks})$

(a) R (A B C D E)

$$F : \{AB \rightarrow CD, A \rightarrow E, C \rightarrow D\}$$

$$R_1(ABC) R_2(BCD) R_3(CDE).$$

Check whether it is lossy or lossless decomposition ?

(b) R (A B C D E F)

$$F = \{B \rightarrow CD, AD \rightarrow E, B \rightarrow A\}$$

$$G = \{B \rightarrow CDE, B \rightarrow ABC, AD \rightarrow E\}$$

Check the equivalence of these two set of functional dependencies.

- (c) Write query for the following problem statement :

Employee (emp\_id, emp\_name, salary, dob, Hire\_date, dep\_id, manager\_id)

(4)

TCS-503

Department (dep\_id, dep\_name, loc\_id, manager\_id)

- (i) Find emp\_id, dob, Dep\_name of all employees in dep\_id 60.
- (ii) Find the emp\_name and their manager name.
- (iii) Find dep\_id, and emp\_id, emp\_name of employees working in same department where 'ankit' works.
- (iv) Create view that contains dep\_id and number of employees in each department.
- (v) Find manager name of all the departments with department\_id.

5. Attempt any two parts of choice from (a), (b) and (c). (10×2=20 Marks)

(a) What is Transaction ? What are the advantages of concurrency of transaction ? Explain the ACID properties of transaction.

(b) Let R = (W X Y Z) be a relational schema with the following dependencies :

$$X \rightarrow W$$

$$WZ \rightarrow XY$$

$$Y \rightarrow WXZ$$

Find the canonical form/irreducible set of FD.

(5)

TCS-503

- (c) Explain all locking protocols with diagram.
6. Attempt any two parts of choice from (a), (b) and (c). (10×2=20 Marks)
- (a) Consider the following relation and write SQL query for the problem statement :  
Student (roll\_number, name, dob, course)
    - (i) Create sequence for column roll\_number.
    - (ii) Verify the sequence.
    - (iii) Insert one row in relation using sequence.
    - (iv) Create a view containing roll\_number, name and age of student.
    - (v) Verify the view.
    - (vi) Show that your view is updateable, insertable.
    - (vii) Give select, insert privilege on student table to user cse01.
    - (viii) Write a command to create new user with some and password.
    - (ix) Give sufficient privileges and role to user so that user can create table.
    - (x) Alter your view and make it read only.

(b) Explain different state of transaction with diagram. Explain the role of locks in concurrent execution of transaction.

(c) Consider the following two transactions :

T1 : read (A);

    read (B);

    if A = 0, then B := B + 1;

    write (B).

T2 : read (B);

    read (A);

    if B = 0, then A := A + 1;

    write (A).

Let the consistency requirement be  $A = 0$

$\vee B = 0$ , with  $A = B = 0$  the initial values.

(i) Show that every serial execution involving these two transactions preserves the consistency of the database.

(ii) Show a concurrent execution of T1 and T2 that produces a no serializable schedule.

(iii) Is there a concurrent execution of T1 and T2 that produces a serializable Schedule ?

Roll No. ....

# TCS-505

## B. TECH. (CSE/IT) (FIFTH SEMESTER) END SEMESTER EXAMINATION, 2018

### DESIGN AND ANALYSIS OF ALGORITHMS

Time : Three Hours

Maximum Marks : 100

- Note : (i) This question paper contains two Sections.  
(ii) Both Sections are compulsory.

#### Section—A

1. Fill in the blanks :  $(1 \times 5 = 5 \text{ Marks})$ 
  - (a) Time complexity of Kruskal's algorithm is ..... .
  - (b) Time complexity of extracting minimum element from Min-Heap is ..... .
  - (c)  $f(n) = \Omega(g(n))$  iff  $f(n) = \dots g(n)$ .
  - (d) Tower of Hanoi takes ..... moves for five disks and three pegs.
  - (e) Array  $\{23, 17, 14, 6, 13, 10, 1, 5, 7, 12\}$  is a ..... (Max-Heap/Min-Heap)

(2)

TCS-505

2. Attempt any five parts : (3×5=15 Marks)  
(Define/Short Numerical/Short Programming/Draw)
- What do you mean by sparse and dense graphs ? Which representation of graph is better and in which case ?
  - "Every tree is a graph but every graph is not a tree." What do you understand by this statement ? Explain using a suitable example.
  - What is the difference between Min-Heap and Max-Heap ? Apply and show stepwise implementation of Max-Heapify (A, 2) on array A = {16, 4, 10, 14, 7, 9, 3, 2, 8, 1}, where array index starts at 1.
  - How can you find out cycle in an undirected graph using disjoint set data structure ?
  - What do you understand by minimum spanning tree ? Is it possible for a graph to have multiple spanning trees ? If yes, provide a suitable example. Write an application of MST.
  - What do you mean by asymptotic analysis ? Explain all the asymptotic notations ( $O$ ,  $\Omega$ ,  $\Theta$ ,  $o$ ,  $\omega$ ) with example.

202-203  
TCS-505

(3)

TCS-505

- (g) Solve the following recurrence using master theorem :

$$T(n) = 2T\left(\frac{n}{4}\right) + \sqrt{4}, \text{ if } n > 1$$

$$T(n) = 1, \text{ otherwise}$$

### Section—B

3. Attempt any two parts of choice from (a), (b) and (c). (10×2=20 Marks)

- Briefly explain the difference between fractional and 0/1 knapsack problem. Write algorithm of fractional knapsack and apply it on the following set of weights and their profits to maximise the profit where Knapsack capacity is 60. Weights = {40, 10, 20, 24} and Profits = {280, 100, 120, 120}.
- Write pseudocode and apply all pair shortest path algorithm (Floyd Warshall) on given graph (Figure 1). Write time and space complexity of algorithm.

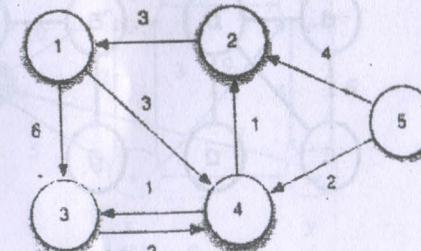


Fig. 1

- (c) Apply Kruskal and Prim's algorithm on given graph (Figure 2) to find out MST. Do you get same MST by applying both the algorithms ?

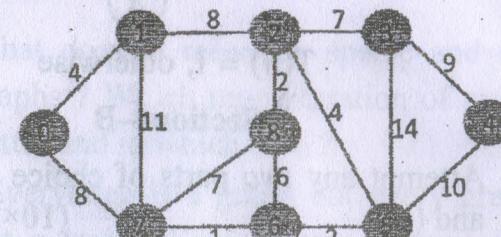


Fig. 2

4. Attempt any two parts of choice from (a), (b) and (c). (10×2=20 Marks)

- (a) Explain disjoint set data structure and operations that can be performed on it. Find out the number of connected components and vertices in each component of graph (Figure 3) using disjoint set data structure.

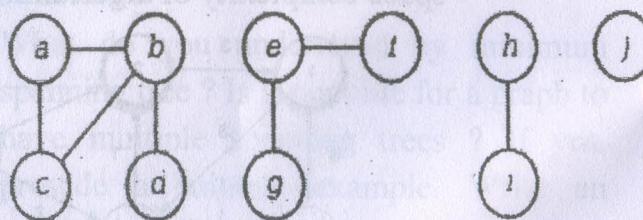


Fig. 3

- (b) Which data structures are used to implement DFS and BFS ? Write and run DFS algorithm (start from A) on given graph (Figure 4) and draw the depth first tree. Also mark all the edges as tree, back, forward and cross edges in the graph.

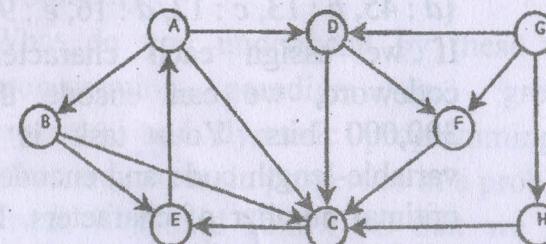


Fig. 4

- (c) Write pseduocode of Bellman algorithm and apply in on graph if Figure 5, the compute shortest path to all nodes from node S. Edges should be relaxed in the following order :

$su, uv, sx, ux, xu, xy, xv, vy, yv, ys$

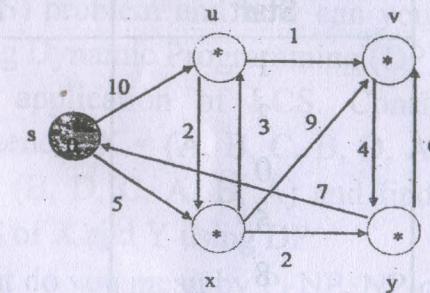


Fig. 5

5. Attempt any two parts of choice from (a), (b) and (c).  $(10 \times 2 = 20 \text{ Marks})$

(a) Why do we use Huffman codes ? A data file of 100,000 characters contains only the characters  $a-f$ , with the frequencies (in thousands) indicated as follows :

$$\{a : 45, b : 13, c : 12, d : 16, e : 9, f : 5\}$$

If we assign each character a 3-bit codeword, we can encode the file in 300,000 bits. Your task is to assign variable-length code and encode the file in optimal number of characters. How much space are you able to save with variable length encoding ?

(b) Write a recurrence relation when Quick sort repeatedly divides the array into two parts of 99% and 1%. Derive the time complexity in this case. Show the recursion tree while deriving time complexity :

| Start | Finish |
|-------|--------|
| 1     | 2      |
| 3     | 4      |
| 0     | 6      |
| 5     | 7      |
| 8     | 9      |
| 5     | 9      |

and also find out the difference between heights of 2 extreme branches of the tree. What insights do you get about quick sort after the division and derivation of its time complexity ?

(c) What do you understand by these two programming paradigm i.e., greedy approach and dynamic programming ? What are two main properties of a problem which it should follow so that we can apply dynamic programming to solve it ? Consider example of Fibonacci numbers and show how these two properties are followed by the problem.

6. Attempt any two parts of choice from (a), (b) and (c).  $(10 \times 2 = 20 \text{ Marks})$

(a) What is Longest Common Subsequence (LCS) problem and how can you solve it using Dynamic Programming (DP) ? Write one application of LCS. Consider two sequences  $X = (A, B, C, B, D, A, B)$  and  $Y = (B, D, C, A, B, A)$  and find out the LCS of X and Y using DP.

(b) What do you mean by P, NP, NP-complete and NP-hard problems ? Explain Travelling Salesman problem. Does there exist any polynomial time solution that

exists for the problem ? Apply approximation algorithm to solve Travelling Salesman problem for graph given in Figure 6. Does approximation algorithm produce optimal solution in this cases ? If, yes will it provide optimal solution in all other cases.

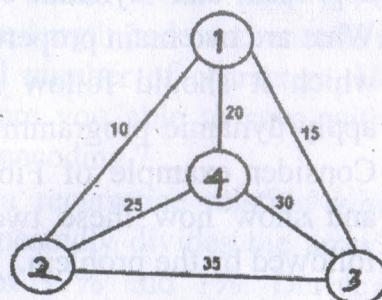


Fig. 6

- (c) Write pseduocode of activity selection problem. You are given 6 activities with their start and finish times. Implement the algorithm to select the maximum number of activities that can be performed by a single person, assuming that a person can only work on a single activity at a time.

Roll No. ....

**TCS-551**

**B. TECH. (CSE) (FIFTH SEMESTER)**

**END SEMESTER EXAMINATION, 2018**

**BIG DATA STORAGE AND PROCESSING**

**Time : Three Hours**

**Maximum Marks : 100**

**Note :** (i) This question paper contains two Sections.

(ii) Both Sections are compulsory.

**Section—A**

1. Fill in the blanks/True-False : (1×5=5 Marks)

(d) (a) Hadoop follows default replication factor “4”. (True/False)

(b) The default size of each block is 128 MB in Apache Hadoop 1.x. (True/False)

(c) HIVE is a project developed by Google. (True/False)

(d) The expansion of CAP is \_\_\_\_\_, \_\_\_\_\_ and \_\_\_\_\_.

(2)

TCS-551

- (e) The receipt of heart beat implies that data node is functioning properly. (True/False)
2. Attempt any *five* parts :  $(3 \times 5 = 15 \text{ Marks})$   
(Define/Short Numerical/Short Programming/Draw)
- What is Job Tracker in Hadoop ?
  - Explain the main components of Hadoop.
  - Draw the HDFS architecture.
  - Define Rack Awareness Algorithm and its importance.
  - Write a short note on Yarn components.
  - Why are writable introduced in Hadoop ?
  - What is Hadoop ecosystem ?

### Section—B

3. Attempt any *two* parts of choice from (a), (b) and (c).  $(10 \times 2 = 20 \text{ Marks})$
- Explain the challenges associated with big data analysis.
  - What are the main file input formats used in Hadoop ?
  - Explain combiner and write a java code to implement a combiner.

(3)

TCS-551

4. Attempt any *two* parts of choice from (a), (b) and (c).  $(10 \times 2 = 20 \text{ Marks})$
- Explain Mapper, Reducer and Driver Class using the word count program.
  - Differentiate between data warehouse and data marts. Which dimensional modelling is used for the creation of data marts ?
  - Write difference between internal and external table in Hive. Write the syntax for creating the creating the internal and external table in Hive.
5. Attempt any *two* parts of choice from (a), (b) and (c).  $(10 \times 2 = 20 \text{ Marks})$
- Differentiate between ACID properties and CAP theorem.
  - Write difference between SQL and NoSQL database.
  - What is Reducer ? Write the difference between combiner and reducer.
6. Attempt any *two* parts of choice from (a), (b) and (c).  $(10 \times 2 = 20 \text{ Marks})$
- Explain the different data types used in HIVE and write down the steps for creating external table in HIVE.

(d) (a) most suited to store large amount of data  
(b) Differentiate between Hbase and HIVE.

Write a steps for creation of a table "employee" in Hbase with some column families and feed some data. (Assume the data yourself).

(c) Draw the architecture of Hbase and explain each component in detail.

(d) (a) most suited to store large amount of data  
Why are writable Introspect in Hbase? (b) Difference between ACID properties and CAP theorem  
What is HDFS?

(d) What is difference between HDFS and Hadoop?

What is Redundancy in HDFS? (a) What is Replication factor in HDFS?

Permissible components and their effect (a) What is the difference between HDFS and Hadoop?

(d) (a) most suited to store large amount of data  
What are the main file input formats used in hadoop? (b) Explain the differences between HDFS and Hadoop?

HDFS and Hadoop and write a java code to

create a file in HDFS and read it from HDFS

Roll No. ....

**TCS-552**

**B. TECH. (CSE) (FIFTH SEMESTER)**  
**END SEMESTER EXAMINATION, 2018**

**CLOUD BASED APPLICATION  
DEVELOPMENT AND MANAGEMENT**

**Time : Three Hours**

**Maximum Marks : 100**

**Note :**(i) This question paper contains two Sections.  
(ii) Both Sections are compulsory.

**Section—A**

1. Fill in the blanks/True-False :  $(1 \times 5 = 5 \text{ Marks})$ 
  - (a) Communication between Cloud services is done widely using REST protocol.  
(True/False)
  - (b) A Cloud can be created within an organization's own infrastructure or outsourced to another data center.  
(True/False)
  - (c) ..... allows different OS to run in their own memory space.

(2)

TCS-552

- (d) EUCLYPTUS stands for .....
- (e) From the standpoint of an PAAS, it makes no sense to offer non-standard machine instances to customers. (True/False)
2. Attempt any *five* parts : (3×5=15 Marks)
- Define SLA with suitable diagram.
  - Define Monthly Uptime Percentage, Uptime and Maximum Available Minutes for Cloud Services.
  - Why is cloud essential to the success of IoT ?
  - Define ALM and CRON with suitable example.
  - Explain the differences between grid computing and cluster computing.
  - Explain the challenges of Cloud Governance.
  - Write a short note on Cloud Service model and Cloud Accounting.

### Section—B

3. Attempt any *two* parts of choice from (a), (b) and (c). (10×2=20 Marks)
- Discuss Inter Cloud SLA broker roles with suitable diagram.

(3)

TCS-552

- (b) Discuss the cloud risk principles and framework. What is the difference between API and SDK ?
- (c) Discuss the architecture of Pivotal Cloud Foundry on AWS with suitable diagram.
4. Attempt any *two* parts of choice from (a), (b) and (c). (10×2=20 Marks)
- Discuss the IBM Cloud deployment model. Briefly explain Salesforce.com, Heroku and Apex.
  - Define Cloud Computing as per NIST with characteristics. Briefly discuss the IoT security model.
  - Discuss SLA service performance metrics with limitations and advantages.
5. Attempt any *two* parts of choice from (a), (b) and (c). (10×2=20 Marks)
- Explain the differences between Cloud Computing and IoT. What are the challenges posed by Cloud and IoT together ?
  - Explain the Cloud Management Goals. Discuss the Cloud Service Management services by any CSP with suitable diagram.

- (c) Discuss the AWS Cloud architecture with suitable diagram. Explain the AWS Management Console features.
6. Attempt any *two* parts of choice from (a), (b) and (c). (10×2=20 Marks)
- Briefly explain the differences between Traditional Accounting and Cloud Accounting. Explain in which scenarios cloud accounting is more useful than traditional accounting.
  - Explain the types of Cloud Computing. Discuss the benefits of Cloud managed services providers.
  - Discuss the four stage IoT solutions architecture. Explain the machine learning requirements on Cloud Computing.

172-2017

TCS-571

Roll No. ....

**TCS-571**

**B. TECH. (CSE) (BDA)**  
**(FIFTH SEMESTER)**

**END SEMESTER EXAMINATION, 2018**

**BIGDATA VISUALIZATION**

**Time : Three Hours**

**Maximum Marks : 100**

**Note :** (i) This question paper contains two Sections.

(ii) Both Sections are compulsory.

**Section—A**

1. Fill in the blanks/True-False : (1×5=5 Marks)

(a) D3 is based on .....

(Java/Python/JavaScript)

(b) “Isarithmic map can be used to visualize Geo-Spatial Data”. Is the given statement : True/False

(c) To connect Tableau to a CSV data source you will use ..... type of connection.

(Excel/Text file/Access)

(2)

TCS-571

- (d) "JASON Lookup is a type of Quick Filter available in Tableau". Is the given statement True or False ?
- (e) The two main groups that all fields are broken up into in Tableau are ..... .  
(Items and Regions/Dimensions and Measures/Labels and Values/Details and Quantities)
2. Attempt any five parts out of seven :  
(Define/Short Numerical/Short Programming/Draw) (3×5=15 Marks)
- (a) What are calculated fields in Tableau ?
- (b) A dataset is available which gives the 100 popular baby names across the world with their ranks (the most popular name first rank and so on). What type of visualization diagram would you prefer to represent this and why ?
- (c) List seven different types of Annotation.
- (d) What are Exploratory Visualizations ?  
What are Exhibitory Visualizations ?
- (e) What is a Radial chart ? Give a disadvantage of a Radial chart.

(3)

TCS-571

- (f) What are the key factors surrounding a visualization project ?
- (g) What is the difference between heat map and tree map ?

### Section—B

3. Attempt any two parts of choice from (a), (b) and (c). (10×2=20 Marks)

(a) Discuss the different channels for obtaining feedback post launch of your visualization project.

(b) You are asked to create a visualization that depicts earthquake data ranging from 1900 to 2017 in the Himalayan states of India. Of the three—Explanatory, Exploratory and Exhibitory—which style will you choose ? Give the reasons for your choice.

(c) What are the three products of Tableau ? Discuss the advantages and disadvantages of Tableau.

4. Attempt any two parts of choice from (a), (b) and (c). (10×2=20 Marks)

(a) What are the five different method classification and what are the corresponding communication purpose for selecting the appropriate visualization method ?

- (b) What is seaborn ? Write code using seaborn to create two visualization (of your choice) on the popular iris dataset. (You should be aware of the iris dataset as a student of BDA, the details will not be provided)
- (c) Differentiate between choropleth map and a cartogram.
5. Attempt any *two* parts of choice from (a), (b) and (c). (10×2=20 Marks)
- (a) Discuss the different representations of data according to Jacques Bertin.
- (b) A dataset is available for different continents in the following format :

| Continent                  | Age<br>(in years) | Population    |
|----------------------------|-------------------|---------------|
| Africa                     | 1                 | 97, 036, 290  |
| Africa                     | 2                 | 84, 153, 168  |
| —                          | —                 | —             |
| Asia                       | 1                 | 558, 423, 716 |
| —                          | —                 | —             |
| Asia                       | 21                | 610, 164, 824 |
| —                          | —                 | —             |
| and so on for 7 continents |                   |               |

i. e., the name of the continent, ages (from 1 to 100) and corresponding population of the age. Median age is the age that divides a population into two numerically equal groups—that is, half the people are younger than this age and half are older. It is a single index that summarizes the age distribution of a population. For example, Canada has a median age of 40, which means approx. half of its population are less than 40 years of age and approx. half of its population are more than 40 years.

Given the above dataset what transformations would you perform to find the median age of the continents. After that provide a suitable method of visualization to show the median age and the distribution of population of different ages in one single visualization.

- (c) A dataset is available with the following attributes in the following order :
- Application\_ID (an alphanumeric uniform value like LP201123, all values present)
  - Gender (Male or Female, all values present)
  - Married (Yes or No, all values present)

- Dependent (Number of dependents, like 0, 1, 2, 3 etc, all values present)
- Education (Graduate or Not Graduate, all values present)
- Self Employed (Yes or No, all values present)
- Applicant Income (In Rupees, Range : 9000 to 95000. Some values missing)
- Co-Applicant Income (In Rupees, Range: 0 to 65000, some values missing)
- Credit History (1 or 0, all values present)
- Residence (Rural, Urban or Semiurban, all values present)
- Loan Amount Requested (In Rupees, 200000 to 1500000, all values present)
- Loan Granted (Yes or No, all values present)

Write a program in Python to perform all the necessary pre-processing tasks which you think are required for the given dataset.

6. Attempt any *two* parts of choice from (a), (b) and (c). (10×2=20 Marks)
- (a) What are the different relationships and connections that can be identified from a dataset ?

- (b) Design three different ways of visualizing Voter Sentiments for the upcoming Indian Lok Sabha Elections in 2019. (You should consider all aspects of visualization and think of rational ways of visualizing the data. Half-baked visualization with simple units will fetch no marks. Note : There are 29 states in India)
- (c) A dataset is available with a sample data as below :

| Pharmaceutical Products Exports by Country in 2017 |             |                                            |               |
|----------------------------------------------------|-------------|--------------------------------------------|---------------|
|                                                    | Exporters   | 2017 Pharmaceutical products (\$ millions) | % World Total |
| 1.                                                 | Germany     | 48,602.9                                   | 15.2          |
| 2.                                                 | Switzerland | 39,949.9                                   | 12.5          |
| 3.                                                 | Belgium     | 26,469.0                                   | 8.3           |
| 4.                                                 | France      | 22,778.4                                   | 7.1           |
| 5.                                                 | USA         | 22,514.6                                   | 7.0           |
| —                                                  | —           | —                                          | —             |
| —                                                  | —           | —                                          | —             |

Note: The above is a sample. The actual dataset contains a total of 194 countries.

Discuss how would you proceed with visualization starting from initial stage and culminating with an appropriate visualization (You should consider all aspects of visualization and think of rational ways of visualizing the data. Half-baked visualization with simple units will fetch no marks).