

Total No. of Printed Pages:03

**B.E (Computer) Semester- VII (Revised Course 2016-17)**  
**EXAMINATION NOV/DEC 2019**  
**Compiler Construction**

**[Time: Three Hours]****[Max. Marks: 100]****Instructions:**

- 1) Answer any two questions from Part-A.
- 2) Answer any two questions from Part-B
- 3) Answer any one question from Part-C.
- 4) Make suitable assumption only if required.

**Part A**

- |     |  |  |
|-----|--|--|
| Q.1 | a) Discuss briefly the compiler writing tools.<br>b) Define context free grammar for the following statements:-<br>i) Nested switch statement with two or more assignment statement in each case.<br>ii) Function definition with atleast two assignments, three if statements and two for loop statements inside function body.<br>c) Construct the operator precedence parser for the following grammar.<br>$S \rightarrow a \mid ^\wedge \mid (T)$<br>$T \rightarrow T, S \mid S$<br>Parse the input : (a, ^)         | <b>(4 marks)</b><br><b>(8 marks)</b><br><b>(8 marks)</b>                     |
| Q.2 | a) What do you mean by an ambiguous grammar? Explain with the help of an example how to eliminate ambiguity in the grammar.<br>b) Eliminate left recursion from the following grammar.<br>$A \rightarrow A B d \mid A a \mid a$<br>$B \rightarrow B e \mid B b \mid b$<br>$C \rightarrow C a \mid b \mid d$<br><br>c) Write a lex program to validate tokens such as identifier, keyword, numeric number and floating point number.<br>d) Explain the process of bootstrapping and porting with the help of a T diagram. | <b>(4 marks)</b><br><b>(5 marks)</b><br><b>(7 marks)</b><br><b>(4 marks)</b> |
| Q.3 | a) Write a YACC program to validate the syntax of following statements:<br>i) While loop with more than one assignment statement inside while loop.<br>b) Write the rules to compute the FOLLOW set. Compute the FIRST and FOLLOW sets for the grammar.<br>$S \rightarrow aAbB \mid aAaB \mid$<br>$A \rightarrow A S \mid a \mid b$<br>$B \rightarrow B S a \mid a \mid c$<br>c) Write an algorithm to compute leading and trailing of a grammar.  | <b>(10 marks)</b><br><b>(5 marks)</b><br><b>(5 marks)</b>                    |

**Part - B**

Q.4

- a) What is peephole optimisation? Discuss at least four peephole optimisation techniques with the help of an example. **(6 marks)**
- b) Consider the source code given below : **(9 marks)**
- ```
for(i=0;i<10;i++)
{
    k=0;
    for(j=0;j<20;j++)
    {
        Sum=Sum +i+j;
    }
    temp = sum;
    while(k < sum)
    {
        A[k]= temp % k;
        k=k+1;
    }
}
```

Using basic block partitioning algorithm optimize the above source code.

- c) Write a short note on run time storage management. **(5 marks)**

Q.5

- a) Describe the various storage allocation strategies. **(5 marks)**
- b) Construct a DAG for the following code: **(6 marks)**
- ```
while(i<10)
{
    a=a +i * a;
    i=i+1;
}
```
- c) Obtain a Quadruple, Triple, Syntax tree and indirect triple for the following expression: **(6 marks))**
- $$(a-b) * (c-d) - (a+b)$$
- d) Write a short note on Next Use Information. **(3 marks)**

Q.6

- a) Write a final code generated by the statement  $x=((a+b)-e-(c-d))$  show the contents of address descriptor and register Descriptor at each statement. How to improve the performance of code generator? Discuss. **(10 marks)**
- b) Explain the function preserving transformations of a program with respect to optimisation. **(10 marks)**

**ParT C**

- Q.7      a) Describe the different issues in the design of a code generator. How to improve the performance of Code generator? Discuss. **(8 marks)**
- b) Explain  
          i) Syntactic Errors  
          ii) Semantic Errors
- c) Write a Yacc program to validate the syntax of Nested If statement with two or more assignment statements. **(8 marks)**
- Q.8      a) Construct parsing table using Canonical LR parser for the below given context free grammar and validate the respective strings.  
 $S \rightarrow id \; E \; E$   
 $E \rightarrow (E) \mid E \; Num \mid id$   
         Input string : a a ( a a 10 )
- b) With help of a neat block diagram explain briefly the different phases of compiler construction. Discuss different types of error at each phase of compiler construction. **(10 marks)**

Total No. of Printed Pages: 03

**B. E. (Computer) Semester-VII (Revised Course 2016-17)****EXAMINATION AUGUST-2020****Compiler Construction**

[Duration : Two Hours]

[Total Marks : 60]

**Please check whether you have got the right question paper.****Instructions:-**

- 1) Attempt THREE FULL Question at least ONE question from EACH Part
- 2) Make suitable assumption only if required.

**Part-A**

- Q.1      a) What is context free grammar? Why CFG is called as Context free grammar? Discuss. (4 marks)  
           b) With the help of neat block diagram, explain different phases of compiler construction. (9 marks)  
           c) Define context free grammar for the following statements:- (7 marks)
  - i) Nested switch statement
  - ii) Function definition with atleast two assignments and for loop statement inside function definition.
- Q.2      a) Write a YACC program to validate the syntax of following statements: (10 marks)
  - i) For loop with more than one assignment statement inside for loop.
- b) Consider the following grammar: (10 marks)
   
S --> If1 | If 2  
  If1--> if ( C ) { Stmt }  
  If2--> if ( C ) { Stmt } else { Stmt }  
  C --> id <= id | id >= id | id > id | id < id  
  Stmt --> Stmt Stmt 1 | Stmt 1  
  Stmt 1 --> id = Exp | S  
  Exp --> Exp + Exp | Exp - Exp | id

Construct parsing table using predictive parser and validate following input string:

If (num1 &gt; 100) {a=a+1}

- Q.3      a) What is the effect of left recursion on bottom up parser? Eliminate left recursion of the following grammar:- (4 marks)
   
S --> A B  
  A --> A d | g  
  B --> B a | B c | h
- b) Construct operator precedence table and parse the given input string: (9 marks)
   
S --> while ( C ) { S1 } | do { S1 } while ( C )  
  C --> id Relop id  
  Relop - -> < | > | <= | > =  
  S1 - -> id = Exp  
  Exp - -> Exp + Exp | Exp - Exp | id
- Input string : While (i < 10 ) { total = total + i - 10} (7 marks)
- c) Write a lex program to identify valid tokens such as identifier, keywords, and floating point number.

**Part-B**

- Q.4 a) What is the role played by symbol table in compiler construction? Discuss. (4 marks)
- b) Consider the following code :- (8 marks)

```
temp = a*b +c
while (temp < 20 )
{
    result = result + temp * n1;
    temp = temp + 2;
}
```

Generate the intermediate code for above given code.

- c) Explain semantic action of intermediate code generation of assignment statement. Generate intermediate code for the following given assignment statement:- (8 marks)

Temp = a\*b \* (b-c) +d - f \* f +g

- Q.5 a) Explain how to design code generator? Design code generator for the given input expression: (8 marks)

Cost = a / b + c \* a - g \* f + a \* b

- b) What is peephole optimization? Using basic block partitioning algorithm, optimize the below given code :- (8 marks)

```
for (i = 0; i< 10; i++)
{
    for (j=0;j<10;j++)
    {
        total = total +A[i] * B[j] + A[j] * B[i];
    }
}
```

- c) Explain the different data structure used to implement symbol table. (4 marks)

- Q.6 a) Discuss different issues with Code generator? What is the significance of next use information in code generator? Discuss. (7 marks)

- b) What is Directed Acyclic Graph? Construct DAG for the give source code: (7 marks)

```
for(i=0;i<20;i++)
{
    temp = a* b* (c+d) - d* (f+g) * d
}
```

- c) Explain the algorithm to compute first and follow of a grammar. Also, explain the algorithm to construct parsing table of predicative parser using first and follow. (6 marks)

**Part C**

- Q.7 a) Explain different errors at each phase of compiler construction. (6 marks)
- b) Write a Yacc program to validate the syntax of Nested If statement with two or more assignment statements inside if loop. (8 marks)
- c) With the help of example, explain Recursive Descent Parser. (6 marks)
- Q.8 a) Explain how Intermediate code is generated for Switch statement. For the given code, generate Intermediate code. (6 marks)
- Switch (a)
- ```

{
    case 1 : temp =a*a+a*b+a*c;
              break;
    case 2 : temp = b*b+b*a+c*a;
              break;
    case 3 :
              temp=c*c+c*a+c*b;
              break;
    default : printf ("Do nothing");
}

```
- b) What is left factoring? Is there any impact of left factoring on parsers? Discuss. (4 marks)
- c) With help of example, explain syntax and parse tree? Where these trees are used in compiler construction? Discuss. (5 marks)
- d) Explain an algorithm to construct parsing table of SLR parser. (5 marks)

Total No. of Printed Pages: 3

**B. E. (Computer) Semester-VII (Revised Course 2016-17)**  
**EXAMINATION OCTOBER 2020**  
**Compiler Construction**

[Duration : Two Hours]

[Total Marks : 60]

**Instructions:**

- 1) Answer THREE FULL QUESTIONS with ONE QUESTION FROM EACH PART.
- 2) Make suitable assumption only if required.

**PART A**

- Q.1      a) With the help of a neat diagram explain the phases of compiler. (10marks)
- b) Define porting. You have a Pascal to C translator running on Pascal machine and a working Pascal compiler. Create a Pascal to C translator for a C machine using T-diagrams. (5 marks)
- c) Write a LEX specification program to identify various tokens such as Identifier, Floating point number, and keyword. (5 marks)
- Q.2      a) Construct the operator precedence table for the following grammar and parse the string: (10marks)  
a+(a-a)+a  
  
$$\begin{array}{l} E \rightarrow +T|T \\ T \rightarrow -F|F \\ F \rightarrow (E)|a \end{array}$$
- b) Write a Yacc program to validate the syntax of nested while and for loop statement with at least two assignment statements inside for or while loop. (10marks)
- Q.3      a) Define Context free grammar for the following statements: (10marks)  
i) Function definition with atleast two assignment and if statements inside function definition.  
ii) Switch statement with at least two for and one while statement inside switch case.  
iii) Structure declaration with 'N' number of members of different data types.
- b) Explain the following with examples: (5marks)  
i) Left recursion  
ii) Left factoring
- c) What is the difference between SLR and Canonical parser? Which is the strongest LR (5marks)

parser? Discuss.

### **PART B**

- Q.4**    a) What is peephole optimization? Explain the program transformation characteristics of peephole optimization. (7marks)

- b) What are the applications of DAG in code optimization? Construct DAG for the following code (8marks)

```
while (i<10)
{
    a[i]=b[i] + c[i] + 20
    i=i+1
}
```

- c) Discuss error detection and recovery mechanism for following context free grammar?

S --> D V

D --> int | float | char

V --> id, V | id

(5marks)

- Q.5**    a) Explain the control flow translation scheme for Boolean expressions and generate intermediate code for following Boolean expression. (8 marks)

Result = A < B or C > D and A < E or E > F

- b) Explain the translation scheme for assignment statement. Generate Intermediate code for following assignment statement. (7 marks)

temp = a+b\*c+f\*(b+k)+h-f

- c) Explain different issues with code generator. (5 marks)

- Q.6**    a) Consider the following program: (8 marks)

```
i=0, j=0; sum=0; A[100][100], B[100][100];
do {
    do {
        sum=sum + A[i][j]*B[i][j];
        i=i+2;
    } while (i<=30);
    j=j+1;
} while(j<=30);
print(sum);
print(i);
```

Discuss the basic block partitioning algorithm. Optimize the above code using basic block partitioning algorithm.

- b) List and explain the data structures used to implement symbol table. (8 marks)  
c) Explain how intermediate code is generated for procedure call? (4 marks)

**PART C**

- Q.7 a) Explain the various errors and error recovery methods of all phases of compiler construction. (8 marks)
- b) Obtain the LALR parsing table for the following:  
 $S \rightarrow CCd$   
 $c \rightarrow e \quad C \mid f$  (10marks)
- c) Why CFG is called as Context free grammar? (2 marks)
- Q.8 a) Explain the code generation algorithm with an example. (5 marks)
- b) Explain significance of symbol table in the compiler construction. (7 marks)
- c) Write a YACC program to validate the syntax of following statements:  
i) Nested for loop and if statement with more than one assignment statement. (8 marks)

Total No. of Printed Pages:4

**B.E. (Computer) Semester- VII (Revised Course 2016-17)**  
**EXAMINATION NOV/DEC 2019**  
**Data Mining**

**[Duration : Three Hours]**

**[Total Marks :100]**

**Instructions:**

1. Assume suitable data
2. Answer any **02** questions from PART-A and PART-B and attempt 01 question from PART-C

**PART A**

- Q.1**
- A) Discuss whether or not each of the following activities is a data mining task. Justify (4)
    - i) Predicting the outcomes of tossing a (fair) pair of dice.
    - ii) Predicting the future stock price of a company using historical records
  - B) Explain using a flowchart the following procedures for attribute subset selection: (6)
    - i) Stepwise forward selection
    - ii) Stepwise backward elimination.
    - iii) A combination of forward selection and backward elimination.
  - C) Explain with a suitable example the following OLAP operations (6)
    - i) Slice
    - ii) Dice
    - iii) Roll-Up
  - D) For the following vectors X and Y, calculate the proximity measures (4)  
 $X=(0, 1, 0, 1, 1, 1, 0, 0)$   
 $Y=(0, 0, 1, 1, 1, 1, 1, 1)$ 
    - i) Euclidean
    - ii) Cosine
    - iii) Jaccard
    - iv) Simple Matching Coefficient
- Q.2**
- A) Classify the following attributes as binary, discrete or continuous. Also classify them as qualitative (nominal and ordinal) or quantitative (interval or ratio). Justify (3)
    - i) Number of patients in a hospital.
    - ii) Bag check number. (When you go to the mall, you often give your Bag to someone who, in turn, gives you a number that you can use to claim your Bag when you leave.)
    - iii) Height above sea level
  - B) Distinguish between noise and outliers. Is noise interesting or desirable? (2)
  - C) Consider the following age values.  
20, 20, 21, 22, 22, 25, 25, 25, 25, 13, 15, 16, 16, 19, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70 (6)  
Describe various data smoothing methods. Perform equal frequency smoothing with a bin size=9

- D) How a snowflake scheme does differs from a star schema? (2)
- E) Consider the following group of points:  
125, 45, 175, 95, 400, 60, 215, 30, 56, 140  
Normalize the data points by using:  
 i) Min-Max Normalization Method where min=0 and Max=1  
 ii) Z-Score Normalization Method (7)

- Q.3**
- A) Explain principal component analysis and state the significance of the same in data preprocessing. (5)
- B) Explain the different types of data sets used in data mining (6)
- C) What is OLAP? How is it different from OLAP? (5)
- D) What is curse of dimensionality? How does it affect classification and clustering? (4)

### PART B

- Q.4**
- A) Consider the following transaction data (6)

| TID | Items_bought  |
|-----|---------------|
| 100 | {A,B,N,E,M,Y} |
| 101 | {T,B,N,E,M,Y} |
| 108 | {A,V,E,M}     |
| 109 | {A,L,K,E,Y}   |
| 110 | {K,P,E,W,M}   |

Find the frequent itemsets using Apriori Algorithm with a min\_support=3 and find all the strong rules with a min\_confidence=70%

- B) Explain overfitting due to presence of noise. (3)
- C) Explain the two important properties of the rule set generated by a rule based classifier. How do we overcome the situation when these properties are not met? (4)
- D) Consider the data mining task to cluster the following points into 3 clusters. (7)

| Points | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 |
|--------|----|----|----|----|----|----|----|----|
| x      | 4  | 2  | 9  | 6  | 7  | 5  | 2  | 6  |
| y      | 6  | 5  | 3  | 9  | 5  | 7  | 2  | 6  |

Suppose the initial cluster centroids are P1, P4 and P7. Use the k-Means algorithm to show the three clusters after two rounds of execution.

- Q.5**
- A) Why the k-nearest neighbour classifier is called a lazy learner? (2)
  - B) Illustrate through examples the different measures for Selecting the Best Split (6)
  - C) Consider the following data set:

| Data points | X   | Y   |
|-------------|-----|-----|
| P1          | 1   | 1   |
| P2          | 1.5 | 1.5 |
| P3          | 5   | 5   |
| P4          | 3   | 4   |
| P5          | 4   | 4   |
| P6          | 3   | 3.5 |

Construct the dendrogram and draw the nested clusters using single linkage clustering

- D) Write the algorithm for basic k-means clustering. What are the advantages and disadvantages of k-means clustering? (4)

- Q.6**
- A) Explain indirect methods for rule extraction with an appropriate example. (4)
  - B) Explain and give examples for the following (6)
    - i) Maximal Frequent Itemset
    - ii) Closed Frequent Itemset
  - C) Explain how the clusters found by the agglomerative clustering algorithm differ to those found by the k-means clustering algorithm? (6)
  - D) Explain how outliers in a univariate normal distribution are detected (4)

### **PART C**

- Q.7**
- A) List the characteristics of three types of outliers and provide a method to detect outlier of each type. (6)
  - B) Explain using an appropriate example of each difference between supervised and unsupervised learning. (4)
  - C) Why are decision trees prone to overfitting in decision trees? Explain the approaches used to deal with overfitting? (7)
  - D) What is sampling? Explain different sampling methods. (3)

- Q.8**
- A) Write and explain the k-Nearest Neighbour algorithm (5)

- B) Define Outlier. Demonstrate using an example and explain the issues faced by statistical approach to (5) outlier detection.

- C) Construct the FP-Tree for the following dataset

| TID | ITEMS   |
|-----|---------|
| 1   | B,C,E,J |
| 2   | B,C,J   |
| 3   | B,M,Y   |
| 4   | B,J,M   |
| 5   | C,J,M   |

Consider the minimum support count =50%

- D) Illustrate the process of knowledge discovery in databases.

Total No. of Printed Pages:3

**B.E (Computer) Semester-VII (Revised Course 2016-17)**  
**EXAMINATION AUGUST-2020**  
**Data Mining**

**[Duration : Two Hours]**

**[Total Marks : 60]**

**Instructions :**      1) Attempt **THREE FULL** Question at least **ONE** question from **EACH Part**.  
 2) Assume suitable data if necessary.

**PART- A**

- |     |                                                                                                                                                                                                                                                      |
|-----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Q.1 | a) Draw a neat labeled diagram and explain the process of knowledge discovery in databases (KDD) <span style="float: right;">5</span>                                                                                                                |
|     | b) Given a list of coordinates ‘cords’ and a point T. Find the item in the list that is closest to the point T.<br>Cords=[(455,12), (188,90),(74,366), (10,10)]<br>$T=(18,448)$ <span style="float: right;">4</span>                                 |
|     | c) Define binning. Perform data smoothing by bin means, bin boundaries and bin median for the following sorted data of Age attribute. 4,8,9,15,21,21,24,25,26,28,29,34 with bins of depth =4 <span style="float: right;">7</span>                    |
|     | d) Find mode, median, mean and variance for the following data series:<br>5,10,13,35,50,50,99 <span style="float: right;">4</span>                                                                                                                   |
| Q.2 | a) What is attribute oriented induction (AOI)? Explain with a suitable example. <span style="float: right;">5</span>                                                                                                                                 |
|     | b) Define Data Cube. Explain with neat diagram different types of OLAP Schemas. “The snowflake schema saves storage space compared to the star schema. “Justify. <span style="float: right;">9</span>                                                |
|     | c) Compute Cosine Similarity and Extended Jaccard coefficient for following two document vectors:A=(3,6,0,3,6) and B= (1,2,0,1,2) <span style="float: right;">4</span>                                                                               |
|     | d) State the general characteristics of Data Sets that have significant impact on data mining techniques. <span style="float: right;">2</span>                                                                                                       |
| Q.3 | a) Explain principal component analysis and state the significance of the same in data preprocessing <span style="float: right;">7</span>                                                                                                            |
|     | b) Suppose the given data is : 300,440,700,990,1100<br>i) Use Z-score normalization to transform values 440,700 and 990<br>ii) Use Min- Max normalization to transform all given values into range of [0.0,1.0] <span style="float: right;">8</span> |
|     | c) Discuss whether or not each of the following activity is Data Mining task<br>i) Dividing the customers of a company according to their gender<br>ii) Monitoring seismic waves for earthquake activities <span style="float: right;">2</span>      |
|     | d) Using following set of stock prices:<br>40,50,70,80,90,100,120,120,140,150,180,200<br>Find 20 <sup>th</sup> Percentile and 50 <sup>th</sup> Percentile. <span style="float: right;">3</span>                                                      |

**PART- B**

Q.4

- a) Write K- Nearest Neighbor Classifier Algorithm  
 b) Draw Decision Tree for following Data set. Explain Steps.

**6**  
**12**

| TID | Home Owner | Marital Status | Annual Income | Class: Loan Defaulter |
|-----|------------|----------------|---------------|-----------------------|
| 1.  | Y          | S              | 125           | N                     |
| 2.  | N          | M              | 100           | N                     |
| 3.  | N          | S              | 70            | N                     |
| 4.  | Y          | M              | 120           | N                     |
| 5.  | N          | D              | 95            | Y                     |
| 6.  | N          | M              | 60            | N                     |
| 7.  | Y          | D              | 220           | N                     |
| 8.  | N          | S              | 85            | Y                     |
| 9.  | N          | M              | 75            | N                     |
| 10. | N          | S              | 90            | Y                     |

- c) Define Rule based Classifier

**2**

Q.5

- a) Consider the data set shown below:

**8**

| Transaction | Items Bought                |
|-------------|-----------------------------|
| T1          | Pasta, Lemon, Bread, Orange |
| T2          | Pasta, Lemon                |
| T3          | Pasta, Orange, Cake         |
| T4          | Pasta, Lemon, Orange , Cake |

Generate Association Rules using Apriori Algorithm.

Consider values of support= 60 % and Confidence =80%

- b) Explain with neat figures k- means and different types of clusters  
 c) Generate FP Tree for the following transaction dataset ( Consider Support count =3)

**6**  
**6**

| TID | ITEM SET                              |
|-----|---------------------------------------|
| 1   | Fan, Axe, Cake, Doll, Gun, Mat, Pan   |
| 2   | Axe, Bat , Cake, Fan, Lock, Mat ,Oven |
| 3   | Bat, Fan, Hat, Oven                   |
| 4   | Bat, Key, Cake , Pan                  |
| 5   | Axe, Fan, Cake , Lock , Pan, Mat, Net |

Q.6

- a) Differentiate between supervised and Unsupervised Learning  
 b) Using Agglomerative Hierarchical Clustering Algorithm, Generate dendrogram for the following Proximity Matrix given below.

**5**

|   | A    | B    | C    | D    | E    | F |
|---|------|------|------|------|------|---|
| A | 0    |      |      |      |      |   |
| B | 0.12 | 0    |      |      |      |   |
| C | 0.51 | 0.25 | 0    |      |      |   |
| D | 0.84 | 0.16 | 0.14 | 0    |      |   |
| E | 0.28 | 0.77 | 0.70 | 0.45 | 0    |   |
| F | 0.34 | 0.61 | 0.93 | 0.20 | 0.67 | 0 |

- c) Explain in brief Rule Ordering Schemes.

### PART - C

- Q.7      a) Explain Challenges that motivated development of Data Mining. **5**  
 b) Suppose that a data warehouse consist of the 4- Dimensions i.e. date, spectator, location and game & the measures were count and charge (where charge is the fare that spectator pays when he/ she is watching a game on given date). Spectators may be students, adults or seniors with each category having its own charge rate. Draw a STAR Schema diagram for the data warehouse. **8**  
 c) Explain Feature Subset Selection process with a flowchart. **5**  
 d) What is the difference between Predictive and Descriptive Data Mining Tasks? **2**

- Q.8      a) Apply KNN(K-Nearest Neighbor ) Classification Algorithm on following data -set & **6**  
 Predict the class for testing data: X(I1=3,I2=7). (Consider k=3)

| Sr. No | I1 | I2 | Class |
|--------|----|----|-------|
| 1      | 7  | 7  | False |
| 2      | 7  | 4  | False |
| 3      | 3  | 4  | True  |
| 4      | 1  | 4  | True  |

- b) Define Outlier. Explain the Issues faced by Statistical Approach to Outlier detection. **4**  
 c) Consider the following data of eight objects for clustering **10**

| Objects | X | Y  |
|---------|---|----|
| A1      | 2 | 10 |
| A2      | 2 | 5  |
| A3      | 8 | 4  |
| A4      | 5 | 8  |
| A5      | 7 | 5  |
| A6      | 6 | 4  |
| A7      | 1 | 2  |
| A8      | 4 | 9  |

Use the k-means Algorithm to cluster the above data into three clusters

Total No. of Printed Pages:4

**B.E.(Computer) Semester-VII (Revised Course 2016-17)**  
**EXAMINATION OCTOBER 2020**  
**Data Mining**

**[Duration : Two Hours]**

**[Total Marks :60]**

- Instructions:-**
1. Answer THREE FULL QUESTIONS with ONE QUESTION FROM EACH PART.
  2. Assume suitable data if necessary.

**Part – A**

- Q.1**
- a) Describe how data mining techniques could be used in an E-commerce setting to decide which customers should the company contact by email with an offer to buy a new book. Justify. (5)
  - b) For the following vectors, x and y:  
x: (2,1,0,2,0,3), y: (1,1,1,0,0,1)  
Compute distance between x and y using SMC, Jaccard, Cosine and Euclidean measures. (4)
  - c) What is principal component analysis (PCA)? Explain its application to the following case: The people in marketing would like a better understanding of their different customers. They want to know what distinguishes customer – What are the key attributes that make a customer unique? The idea is not to group similar customers, but to identify the attributes that set customers apart. (7)
    - i) Justify your answer.
    - ii) Also suggest how you would relate one of the other data mining technique in this regard.
  - d) Find mode, median, mean and variance for the following data series: (4)  
5, 10, 13, 35, 50, 50, 99
- Q.2**
- a) What is binarization in data mining? Explain its significance. (3)
  - b) Suppose that we have sales data given by address and the address fields include house number, street name, city, state, pincode and country. Generate concept hierarchy and explain its significance. (3)
  - b) Suppose the population of a particular breed of cats having weights is normally distributed. Furthermore, suppose we know that the mean of the distribution is 10 pounds and the standard deviation is 2 pounds. What is the z-score for 13 pounds? (2)
  - c) In data mining context, what is curse of dimensionality? Suggest three effective methods used for dimensionality reduction. (4)

- d) The college wants to record the marks for the courses completed by students using the dimensions course, student, time and a measure aggregate \_marks. Design a snowflake schema for the data warehouse, clearly depicting the fact tables, dimension tables, their attributes and measures. Explain why a snowflake schema saves storage space compared to the star schema.

Q.3 a) What is data mining? Describe the steps involved in data mining when viewed as a process of knowledge discovery.

- b) The age values for the data tuples are 20, 20, 21, 22, 22, 25, 25, 25, 13, 15, 16, 19, 30, 33, 33, 35, 35, 36, 40, 45, 46, 52, 70.
- Use smoothing by bin means to smooth the above data using a bin depth of 5.
  - Compute the transformed values for 13, 35, 70 using min-max normalization by setting min=0 and max=1.

- c) Draw a schematic to explain the following typical operations on multidimensional data:

- Roll-up
- Dice
- Slice
- Drill-down
- Pivot

### Part – B

Q.4 a) In hierarchical clustering, state the characteristics of each of the following linkages between objects:

- Single linkage
- Complete linkage
- Average linkage.

(3+7)

Consider the following input distance matrix T of size 6 by 6 which was created based on the object A, B, C, D, E and F. Perform single linkage agglomerative hierarchical clustering on T and display the generated dendrogram.

|   | A    | B    | C    | D    | E    | F    |
|---|------|------|------|------|------|------|
| A | 0.00 | 0.71 | 5.66 | 3.61 | 4.24 | 3.20 |
| B | 0.71 | 0.00 | 4.95 | 2.92 | 3.54 | 2.50 |
| C | 5.66 | 4.95 | 0.00 | 2.24 | 1.41 | 2.50 |
| D | 3.61 | 2.92 | 2.24 | 0.00 | 1.00 | 0.50 |
| E | 4.24 | 3.54 | 1.41 | 1.00 | 0.00 | 1.12 |
| F | 3.20 | 2.50 | 2.50 | 0.50 | 1.12 | 0.00 |

- b) With the help of an appropriate example demonstrate building a rule-based classifier. Explain different rule ordering schemes.

(8)

- c) How does bisecting k-means work?

Q.5

- a) Consider the following transaction dataset D:

| TID | Items      |
|-----|------------|
| 100 | 1, 3, 4, 6 |
| 200 | 2, 3, 5    |
| 300 | 1, 2, 3, 5 |
| 400 | 1, 5, 6    |

- i) Consider minimum support minSup=2 Find all frequent itemsets from D using Apriori algorithm.
- ii) Write apriori algorithm for finding frequent itemsets.
- b) Is clustering an example of supervised or unsupervised learning? Give an example.
- c) Using a neat labeled graph demonstrate the condition of overfitting in decision trees. What are the different reasons for cause of overfitting? What are different approaches which can be used to deal with overfitting?

Q.6

- a) Discuss the difference between noise and outlier? Demonstrate one approach to deal with each of them and state the implications of the said approach.

- b) Consider the training data as given in Table 1, having three variables: hair, height and country.

**Table 1:**

| Hair  | Height | Country  |
|-------|--------|----------|
| Blond | Tall   | Polvia   |
| Blond | Tall   | Polvia   |
| Blond | Short  | Gromland |
| Dark  | Short  | Gromland |
| Dark  | Tall   | Gromland |
| Blond | Short  | Gromland |

Construct decision tree classifier. Show the split test used at each node. For each leaf node, show the class and the records associated with it. Explain how you derived the split node using the information gain.

- c) Write KNN (K-Nearest Neighbor) classification algorithm and apply the same on following data-set Predict the class for testing data: X(I1=3, I2=7). (Consider k=3).

(7)

| Sr.No. | I1 | I2 | Class |
|--------|----|----|-------|
| 1      | 7  | 7  | false |
| 2      | 7  | 4  | false |
| 3      | 3  | 4  | true  |
| 4      | 1  | 4  | true  |

**PART – C**

- Q.7      a) Draw a neat labeled diagram to depict the process of knowledge discovery and pattern finding from large data. Discuss whether or not each of the following activities is a data mining task. If “Yes” mention the associated DM task. (5)
- i)      Sorting a student database based on student identification numbers
  - ii)     Monitoring the heart rate of a patient for abnormalities.
  - iii)    In an internet search engine company, there is a need to find potential users who will click a particular advertisement on the webpage.
- b) Consider a two dimensional database D with the records: R1(2, 2), R2(2, 4), R3(4, 2), R4(4, 4), R5(3, 6), R6(7, 6), R7(9, 6), R8(5, 10), R9(8, 10), R10(10, 10). Show the results of the k-means algorithm at each step, assuming that you start with two clusters (k=2) with centers C1=(6,6) and C2=(9, 7). (8)
- c) Why is entity-relationship modeling technique not suitable for the data warehouse? How is dimensional modeling different? Demonstrate. (5)
- d) What is the difference between Predictive and Descriptive Data Mining Tasks? (2)
- Q.8      a) List the characteristics of three types of outliers and provide a method to detect outlier of each type. (6)
- b) Define outlier. Explain the Issues faced by statistical approach to outlier detection. (4)
- c) Consider the following transactional database D. Assume that min support = 40%. (10)

| TID   | Items bought          |
|-------|-----------------------|
| T100  | { 16, 11, 13}         |
| T200  | { 11, 12, 14, 15, 13} |
| T300  | {13, 12, 15}          |
| T400  | {16, 17}              |
| T500  | {11, 13, 12, 14, 15}  |
| T600  | {11, 13, 16}          |
| T700  | {11, 12, 15, 17}      |
| T800  | {12, 18, 15, 11}      |
| T900  | {14, 16}              |
| T1000 | {11, 12, 15}          |

- i) Apply the FP-growth algorithm to generate the frequent itemsets of D. Show the FP –tree and the header table. For each frequent item, show how to generate the conditional pattern bases and conditional FP-tress, and the frequent itemsets generated by them.
- ii) From the frequent itemsets you have mined, generate all association rules of confidence 100%

Total No. of Printed Pages:4

**B.E. (Computer) Semester- VII (Revised Course 2016-17)**  
**EXAMINATION NOV/DEC 2019**  
**Image Processing**

**[Duration : Three Hours]**

**[Total Marks:100]**

**Instructions:**

1. Answer ANY 2 questions from Part-A and Part-B and ANY 1 question from Part-C.
2. Assume Suitable Data if required.

**Part A**

**Q1**      a. Explain the need and conversion process of analog to digital signals in Digital Image Processing. (4)

b. Convert following RGB color pixel (255, 128, 64) to grayscale color. How many grey levels does this image require? (4)

c. What will be the result of convolution operation given image ‘a’ and convolution kernel ‘h’? Use suitable padding for the image. (4)

$$a = \begin{bmatrix} 200 & 201 & 202 \\ 202 & 203 & 205 \\ 205 & 210 & 211 \end{bmatrix} \text{ And } h = \begin{bmatrix} 1/8 & 1/8 & 1/8 \\ 1/8 & 0 & 1/8 \\ 1/8 & 1/8 & 1/8 \end{bmatrix}$$

d. What is a color model? Explain the HIS and RGB color models. (8)

**Q2**      a. Explain the various image storage formats used in digital image processing. A true color image has a resolution of 800 x 600, calculate the size of image file when the image is stored as a BMP. (6)

b. Calculate the point of highest correlation for following 1-dimensional signals.  $A=[1.0, 0.4, -1.0, 0.4]$ , and  $B=[0.4, 1.0, 0.4, -1.0]$ . (4)

c. List out the properties of Discrete Fourier Transform. Compare the complexities of DFT and FFT. (6)

d. Explain 2D Discrete Fourier transform. What will be the Fourier Transform of the image (4)

$$f(m, n) = \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix}$$

- Q3** a. Consider a gray level histogram of a  $64 \times 64$  image, Apply Histogram equalization technique to enhance the image (6)

| Gray level | 0   | 1  | 2   | 3   | 4   | 5    | 6   | 7   |
|------------|-----|----|-----|-----|-----|------|-----|-----|
| Frequency  | 140 | 54 | 295 | 400 | 643 | 1050 | 814 | 700 |

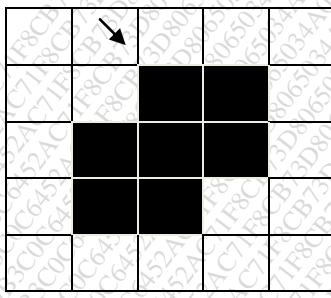
- b. With the help of transfer function explain Ideal and Butterworth low-pass filtering. (4)  
 c. Explain the linear and shift invariant property of Image Processing system. (4)  
 d. Apply suitable Sobel operator to highlight the edge in the image (6)

$$f = \begin{bmatrix} 3 & 3 & 3 & 3 \\ 3 & 3 & 3 & 3 \\ 7 & 7 & 7 & 7 \\ 3 & 3 & 3 & 3 \end{bmatrix}$$

### Part B

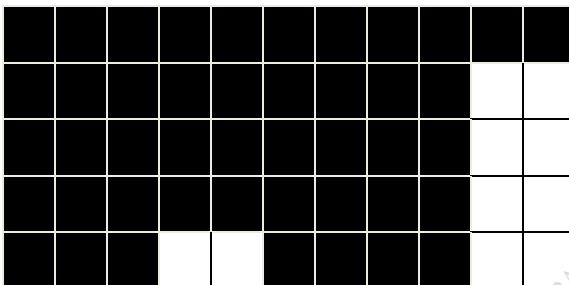
- Q4** a. Explain the different types of Thresholding techniques used in image segmentation (4)  
 b. The histogram of a 4-bit grey-scale image having a size  $16 \times 16$  is described as the following (6) array. Determine an optimal threshold to segment the foreground from the background.  
 $(0, 5, 5, 10, 16, 10, 5, 10, 20, 35, 55, 35, 25, 15, 10, 0)$   
 c. Explain Hough transform for edge detection. (6)  
 d. Explain the method of border tracing w.r.t. image A. (4)

A

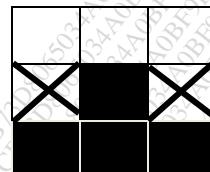


- Q5** a. Explain the technique of Edge Relaxation used for image segmentation. Compute the Edge type, given the confidences for vertices on either side of the crack edge as  $(a_u, b_u, c_u) = (0.9, 0.9, 0.01)$  and  $(a_v, b_v, c_v) = (0.01, 0.01, 0.01)$ . (8)  
 b. Given structuring element B, Define suitable structuring elements and perform morphological thinning operation on image A. (8)

**A**



**B**

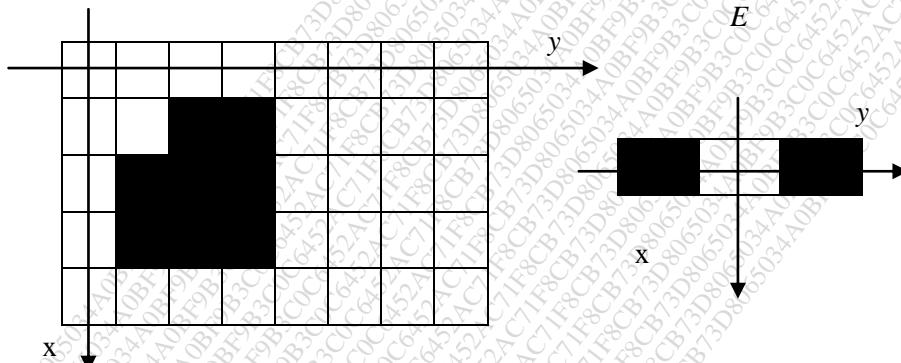


- c. Explain Hit-or-Miss transformation in mathematical morphology. (4)

- Q6**
- a. Explain the various Region-Based segmentation techniques. (6)
- b. Perform Dilation and Erosion operations to the binary image A given with the structuring element E={(-1,0),(0,1)} (6)

element E={(-1,0),(0,1)}

**A**



- c. Compute the new signal generated by the dilation  $f \oplus h$  given one dimensional signal as follows (6)

$$\begin{aligned} f(t) &= \{f(0), f(1), f(2), f(3), f(4), f(5), f(6), f(7), f(8), f(9)\} \\ &= \{3, 5, 8, 4, 2, 6, 8, 10, 5, 4\} \end{aligned}$$

With the structuring element  $h(t) = \{h(-1), h(0), h(1)\}$

- d. List the various applications of binary morphological operations. (2)

### Part C

- Q7**
- a. Explain various Spatial domain smoothing methods. (4)
- b. Compute Discrete Cosine Transform of the following image (6)

$$f(m, n) = \begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix}$$

- c. With the help of transformation graph explain Limiting Linear Transform. Apply the transformation of the image segment given  $(r_1, s_1) = (2, 2)$  and  $(r_2, s_2) = (5, 7)$ . (6)

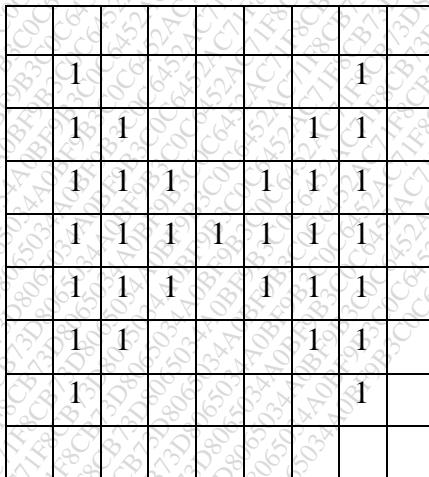
|   |   |   |   |   |
|---|---|---|---|---|
| 2 | 7 | 4 | 5 | 3 |
| 5 | 6 | 7 | 1 | 2 |
| 0 | 1 | 2 | 3 | 4 |
| 7 | 5 | 3 | 1 | 0 |

- d. Explain image sharpening. (4)

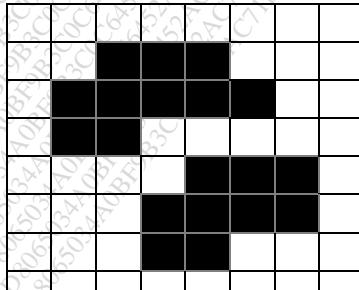
- Q8** a. Explain contrast stretching in digital image processing. (4)
- b. Using the  $3 \times 3$  median filtering method, smooth the image (6)

$$f = \begin{bmatrix} 0 & 2 & 2 & 2 & 3 \\ 2 & 3 & 5 & 4 & 4 \\ 5 & 6 & 6 & 6 & 7 \\ 5 & 6 & 7 & 7 & 6 \\ 3 & 3 & 5 & 7 & 7 \end{bmatrix}$$

- c. Perform Opening and Closing operations to the binary image given in figure with the structure element  $\{(0, -1), (0, 1), (0, 0), (1, 0), (-1, 0)\}$ . Element  $(0, 0)$  is the origin of the structuring element. (6)



- d. Apply the region-growing method to find the objects given in figure below (4)



Total No. of Printed Pages:4

**B.E (Computer) Semester-VII (Revised Course 2016-17)**

**EXAMINATION AUGUST 2020**

**Image Processing**

**[Duration : Two Hours]**

**[Total Marks :60]**

**Instructions:-**

1. Answer THREE FULL QUESTIONS with ONE QUESTION FROM EACH PART
2. Assume suitable Data if required.

**Part-A**

Q1. a) Explain sampling and quantization process applied on an image. Explain the effects of inadequate sampling and quantization on the quality of the sampling. (6)

b) Define the Fourier and Inverse Fourier transform of an image. Explain the importance of the Fourier transform in image processing. State any two applications of the Fourier transform. Compute the DFT of the following matrix. (8)

$$f = \begin{bmatrix} 3 & 2 \\ 4 & 6 \end{bmatrix}$$

c) An image matrix  $f$  is given  $f(m, n) = \begin{bmatrix} 5 & 6 & 4 & 3 \\ 3 & 3 & 3 & 8 \\ 3 & 3 & 3 & 3 \\ 3 & 3 & 3 & 3 \end{bmatrix}$  (6)

Write matrices for the processed image when the following techniques are applied to the image

- i. Contrast stretching using point processing when  $(r_1, s_1) = (2, 1)$  and  $(r_2, s_2) = (4, 7)$ .
- ii. Low pass filtering using  $3 \times 3$  mask

Q2. a) Consider the  $4 \times 4$  image given below:

$$f = \begin{bmatrix} 10 & 12 & 9 & 5 \\ 3 & 1 & 8 & 6 \\ 9 & 9 & 8 & 5 \\ 4 & 5 & 13 & 10 \end{bmatrix}$$

Write the processed image when the following techniques are applied to the image with padding bit 0: (3)

- i. Median filtering (3)
- ii. Average filtering (3)

b) Explain the use of second order derivatives for image enhancement. (6)

c) Explain the basic steps of filtering an image in the frequency domain. (5)

- d) Find the number of bits required to store a  $512 \times 128$  image with 128 gray levels. (3)
- Q3. a) Gray level histogram of a  $64 \times 64$  image is given below. Compute the gray level histogram of the output image obtained by enhancing the input by the histogram equalization technique. (6)
- | Gray level | 0    | 1   | 2   | 3   | 4   | 5   | 6   | 7  |
|------------|------|-----|-----|-----|-----|-----|-----|----|
| Frequency  | 1023 | 850 | 790 | 856 | 122 | 245 | 329 | 81 |
- b) Explain the CMY color model. Red, Green and Blue values of a pixel are 108, 110 and 98 respectively. Maximum intensity in RGB is 255 then what are its corresponding values in CMY model in the range 0 to 1. (6)
- c) Explain any two properties of the two dimensional Fourier Transform. (4)
- d) Compare and contrast the Ideal and Butterworth low pass filters. (4)

### **Part-B**

- Q4. a) Explain the region split and merge method for segmentation. Segment the image shown below by using region split and merge method for segmentation. Use same gray scale intensity as the homogeneity principle. Indicate the number of regions obtained after segmentation. (7)
- |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| 1 | 1 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| 1 | 1 | 7 | 7 | 3 | 3 | 3 | 3 | 3 |
| 1 | 1 | 7 | 7 | 3 | 3 | 3 | 3 | 3 |
| 6 | 6 | 6 | 6 | 6 | 6 | 2 | 2 | 2 |
| 6 | 6 | 6 | 6 | 6 | 6 | 2 | 2 | 2 |
| 6 | 6 | 6 | 6 | 6 | 6 | 0 | 0 | 0 |
| 6 | 6 | 6 | 6 | 6 | 6 | 0 | 0 | 0 |
- b) Explain the detection of straight lines using the Hough Transformation with suitable examples. (6)
- c) Find the skeleton of the following image using mathematical morphology. Use the structuring element  $E = \{(-1,0), (1,0), (0,0), (0,1), (0,-1)\}$  (7)
- |   |   |   |   |   |   |  |
|---|---|---|---|---|---|--|
|   |   |   |   |   |   |  |
| 1 | 1 | 1 | 1 | 1 | 1 |  |
| 1 | 1 | 1 | 1 | 1 | 1 |  |
|   | 1 | 1 | 1 | 1 |   |  |
|   | 1 | 1 | 1 | 1 |   |  |
| 1 | 1 | 1 | 1 | 1 | 1 |  |
| 1 | 1 | 1 | 1 | 1 | 1 |  |

- Q5. a) Describe region growing approach. What are the problems associated with this method? (6)
- b) Perform erosion, opening and closing operations on the image given below using the structuring element  $\{(0, -1), (0, 1), (0, 0), (1, 0), (-1, 0)\}$ . Element (0,0) is the origin of the structuring element. (8)

|   |   |   |   |   |   |   |  |
|---|---|---|---|---|---|---|--|
|   |   |   |   |   |   |   |  |
| 1 | 1 | 1 | 1 |   |   |   |  |
|   | 1 | 1 | 1 | 1 |   |   |  |
|   |   | 1 | 1 | 1 | 1 |   |  |
|   |   |   | 1 | 1 | 1 | 1 |  |
|   |   |   | 1 | 1 | 1 | 1 |  |
|   | 1 | 1 | 1 | 1 |   |   |  |
| 1 | 1 | 1 | 1 |   |   |   |  |

- c) Explain the thinning operation in mathematical morphology. (6)
- Q6. a) What is thresholding? Consider the image given below and obtain the threshold value using optimal thresholding and perform basic thresholding and semi-thresholding operation. (8)

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

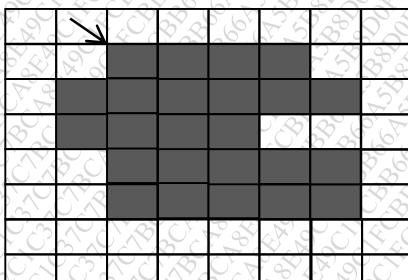
- b) Explain the basic grey scale dilation algorithm. Perform grey scale dilation on the given image using the structuring element  $E = \{(-1, 0), (1, 0), (0, 0), (0, 1), (0, -1)\}$  (8)

$$f = \begin{bmatrix} 11 & 2 & 6 & 5 \\ 13 & 5 & 8 & 6 \\ 8 & 5 & 8 & 5 \\ 4 & 5 & 3 & 10 \end{bmatrix}$$

- c) Explain thresholding with hysteresis. (4)

**Part-C**

- Q7. a) Explain the degradation / restoration model of an image. Explain how the problem of image restoration can be viewed as a problem of inverse filtering. (8)
- b) Briefly explain how sharpening filters works in frequency domain and explain the Butterworth filter used for sharpening an image. (6)
- c) Discuss the RGB color model. How many different shades of gray are there in a color RGB system in which each RGB component is (6)
- Q8. a) Differentiate between the following (6)
- Semi-Thresholding and Band-Thresholding
  - Adaptive and optimal thresholding
- b) Explain the opening and closing operations in mathematical morphology. What is the difference between the two operations? Specify situations where you would prefer an opening operation over a closing one. (6)
- c) Detect the border of the picture given below using border tracing detection algorithm in 8-neighborhood. (8)



Total No. of Printed Pages:5

**B.E. (Computer ) Semester- VII (Revised Course 2016-17)  
EXAMINATION OCTOBER 2020  
Image Processing**

**[Duration : Two Hours]**

**[Total Marks : 60]**

**Instructions:**

1. Answer THREE FULL QUESTIONS with ONE QUESTION FROM EACH PART.
2. Assume Suitable Data if required.

**Part A**

- Q.1      a) Write and explain 1D Discrete Fourier transform and its inverse. What will be the Discrete Fourier Transform of the one dimensional function  $f = \{1, 4, 5, 4\}$  (5)
- b) Find the number of bits required to store a  $128*128$  image with 64 gray levels. (3)
- c) Explain the convolution and the correlation operation. What is the difference between the convolution and correlation operations? (5)
- d) Explain the terms sampling and quantization. (4)
- e) Red, Green and Blue values of a pixel are 158, 120 and 98 respectively. Maximum intensity in RGB is 255 then what are its corresponding values in CMY model in the range 0 to 1. (3)
- Q.2      a) Differentiate between the following and provide circumstances wherein you would prefer one over the other.  
a) Median filtering and average filtering  
b) Laplacian and gradient filters (6)
- b) Explain any 3 properties of the Discrete Fourier Transform. (6)
- c) Consider the following image

$$a = \begin{bmatrix} 7 & 5 & 2 & 9 \\ 2 & 3 & 5 & 8 \\ 6 & 4 & 3 & 8 \\ 9 & 2 & 4 & 3 \end{bmatrix}$$

Apply the following operations on the above image and give the corresponding result. Use suitable padding.

- i) Median filtering
- ii) Average filtering

- iii) Weighted median filtering using the template  $t = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 1 \end{bmatrix}$

- Q.3 a) Consider the following  $4*4$  matrix of a 3 bit image. Perform histogram equalization and show the equalized histogram and the equalized image. (6)

$$a = \begin{bmatrix} 0 & 3 & 4 & 2 \\ 4 & 7 & 5 & 4 \\ 7 & 5 & 6 & 6 \\ 1 & 7 & 4 & 3 \end{bmatrix}$$

- b) With the help of transfer function explain Ideal and Butterworth high-pass filtering. (4)
- c) What filter would you apply to restore an image corrupted by additive Gaussian noise and motion blur if your goal is to minimize the expected mean squared error between the original image and the restored one? Explain the filter used. (4)
- d) Consider the image given below: (6)

$$f = \begin{bmatrix} 3 & 3 & 3 & 3 \\ 3 & 3 & 3 & 3 \\ 9 & 9 & 9 & 9 \\ 9 & 9 & 9 & 9 \end{bmatrix}$$

Apply the following sharpening filters to detect the edges.

- i) Laplacian filtering
- ii) Sobel filter

## **Part B**

- Q.4 a) An image matrix is given below: (4)

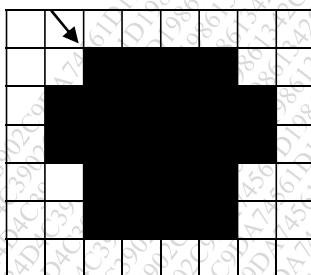
$$f = \begin{bmatrix} 4 & 3 & 7 & 1 \\ 2 & 0 & 7 & 6 \\ 1 & 5 & 2 & 1 \\ 2 & 3 & 4 & 1 \end{bmatrix}$$

Perform optimal or iterative thresholding on the above image.

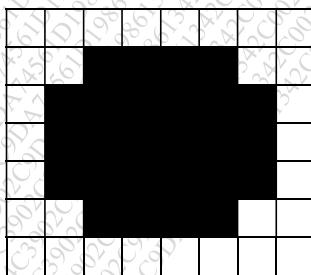
- b) Explain the image split and merge segmentation method. Segment the image shown below by using image split and merge segmentation method. Use same gray scale intensity as the homogeneity principal. (4)

$$f = \begin{bmatrix} 2 & 4 & 4 & 4 \\ 2 & 4 & 4 & 4 \\ 2 & 5 & 4 & 1 \\ 2 & 5 & 4 & 1 \end{bmatrix}$$

- c) What are the various steps in canny's edge detection algorithm? (8)
- d) Detect the border of the image given below using the border tracing detection algorithm in 8-neighbourhood. (4)



- Q.5
- a) Briefly explain thresholding. Differentiate between basic, semi and band thresholding. (6)
- b) Define suitable structuring elements and perform morphological thinning operation on below given image. (8)



- c) Explain the Hough transform. (6)

- Q.6
- a) Differentiate between the following (6)
- i) Opening and closing morphological operation
  - ii) Region growing and region merging method

- b) Explain the basic grey scale erosion algorithm. Perform grey scale erosion on the given image using the structuring element  $E=\{(-1,0), (1,0),(0,0),(0,1),(0,-1)\}$  (8)

$$f = \begin{bmatrix} 10 & 12 & 9 & 5 \\ 13 & 15 & 18 & 6 \\ 9 & 9 & 8 & 5 \\ 14 & 15 & 13 & 10 \end{bmatrix}$$

- c) What are crack edges? Where are they used? Which are the four crack edges associated with pixel  $(i, j)$ . Give its magnitude and direction. (6)

### Part C

- Q.7      a) Explain trapezoidal low pass filtering. (4)
- b) Explain the HSI color model. (6)
- c) Perform Discrete Cosine Transform on the following matrix. (6)

$$f = \begin{bmatrix} 5 & 1 \\ 1 & 5 \end{bmatrix}$$

- d) What is contract stretching? Consider the following image (4)

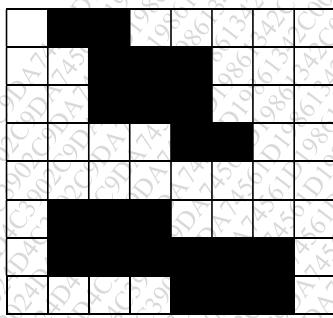
$$a = \begin{bmatrix} 4 & 5 & 2 & 7 \\ 2 & 3 & 5 & 6 \\ 6 & 4 & 3 & 8 \\ 7 & 2 & 4 & 3 \end{bmatrix}$$

Perform Contrast stretching to stretch the intensity in original image 2 to 0 in the stretched image and intensity 8 in original image to 15 in the stretched image.  $(r_1, s_1)=(2,0)$  and  $(r_2, s_2)=(9,15)$ .

- Q.8      a) What is non-maximal suppression? (6)
- b) Discuss the application of grey-scale morphological operations. (2)
- c) Perform erosion, dilation, opening and closing operations on the image given below using the structuring element  $\{(0,-1), (0,1), (0,0), (1,0), (-1,0)\}$ . Element  $(0,0)$  is the origin of the structuring element. (8)

|   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|
|   |   |   |   |   |   |   |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 |
|   | 1 | 1 | 1 | 1 | 1 |   |
|   |   | 1 | 1 | 1 |   |   |
|   |   |   | 1 |   |   |   |
|   |   |   |   | 1 | 1 |   |
|   |   |   |   | 1 | 1 | 1 |
|   |   |   |   | 1 | 1 | 1 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 |
|   |   |   |   |   |   |   |

- d) Apply the region-growing method to find the objects given in figure below **(4)**



Total No. of Printed Pages:3

**B.E. (Computer) Semester- VII (Revised Course 2016-17)**  
**EXAMINATION NOV/DEC 2019**  
**Elective - I**  
**Web Technologies**

**[Duration : Three Hours]**

**[Total Marks :100]**

**Instructions:**

1. Answer **any 5** questions by selection **two** questions from Part – A, **two** questions from Part – B and **one** question from Part – C.
2. Make suitable assumptions **if required.**

**PART- A**

Answer any **TWO** questions from the following:

**Q1.a) Compare the following:**

- i) URL and DNS
- ii) Website and Web service

**[6]**

b) Explain IIS 7 and the services supported by it.

**[6]**

c) Design the following form using HTML.

**[8]**

**User info form:**

Name:

Password:

Party:

- Democrat
- Republican

State:

California  V

Comments:

Your comments would be appreciated

Send me more information

**Q2.** a) Explain 3 – Tier Web System Architecture with the help a neat diagram. [5]

b) Demonstrate a JavaScript program to calculate the percentage of a student using onclick event. [5]

c) Explain with the help of example, Absolute positioning and Relative positioning. [6]

d) Differentiate between DTD and XML Schema. [4]

**Q3.** a) Explain Inline Style sheets, Embedded style sheets and Contextual Selectors. [9]

b) List and explain the building blocks of XML documents. [5]

c) Consider the following XML document. [6]

```
<? Xml version="1.0"?>
<PAYROLL>
    <EMPLOYEE>
        <ID> 1011 </ID>
        <NAME> John Smith </NAME>
        <DEPARTMENT> Projects </DEPARTMENT>
        <BASIC PAY>20000</BASICPAY>
        <MANAGER> Ron Martin </MANAGER>
    </EMPLOYEE>
</PAYROLL>
```

Write XSL program to convert the above XML document into HTML output as shown below.

ID	NAME	DEPARTMENT	BASICPAY	MANAGER
1011	John Smith	Projects	20000	Ron Martin

### PART B

Answer any **TWO** questions from the following:

**Q4.** a) Explain MVC in reference to Angular JS. [6]

b) What is dependency injection and benefits of it? [5]

c) What are the directives? Mention any four directives used in Angular JS application with an example? [6]

d) Explain the use of Cookie class in JSP. [3]

- Q5.**
- a) Comment on the advantages of JSP over other languages. [4]
  - b) Write a code to retrieve the details of a Student (Roll Number, Name and Age) posted from a HTML file to a JSP file. [6]
  - c) Write a PHP program to check if a number is Palindrome using loop statement. [5]
  - d) List and explain any five features of PHP. [5]
- Q6.**
- a) What are the advantages of PHP over other scripting languages? [4]
  - b) Write a PHP program to illustrate the use of foreach looping statement. [4]
  - c) Mention any six String manipulation functions in PHP. [6]
  - d) Explain the fopen() function and the various modes of opening a file in PHP [6]

### **PART C**

Answer any **ONE** questions from the following:

- Q7.**
- a) Explain the hierarchy of HTML document. [5]
  - b) Compare Class as Selector and ID as Selector with help of an example in CSS. [5]
  - c) What is two-way binding in AngularJS? [4]
  - d) Explain the following function in PHP. [6]
    - i)    `$_GET[]`
    - ii)   `$_POST[]`
    - iii)   `$_REQUEST[]`
- Q8.**
- a) What is Image Map in HTML? Explain it with the help of an example. [5]
  - b) What are XML Namespaces? Write an example to explain its use. [5]
  - c) List and explain components of JSP. [6]
  - d) Write short notes on the following directory manipulation functions. [4]
    - i)    `mkdir()`
    - ii)   `chdir()`

Total No. of Printed Pages: 02

**B. E. [COMUTER] (Semester-VII) (RC 2016-17)**  
**Examination August 2020**  
**Web Technologies**

**[Duration : Two Hours]**

**[Total Marks : 60]**

**Instructions:**

1. Answer THREE full questions with ONE question from each part.
2. Make suitable assumptions if required.

**PART A**

Answer any **TWO** questions from the following:

- Q. NO 1.     A) Create a small XML file designed to contain information about a list of movies. Each **8**  
movies has: name, producer, duration. Consider the information about 2 movies.  
Create a DTD capturing the above document type.  
Using XSL, translate the above XML file into HTML. (Display output in tabular  
format)
- B) Explain the OSI Reference Model. **8**  
              C) What is a Web Service? Differentiate between Website and Web Service. **4**
- Q. NO 2.     A) Create an HTML file which will display the quantity sold of laptops, mobiles and **5**  
printers for January, February and March in tabular format.  
B) With the help of examples explain:  
    i)     Inline styles  
    ii)    Embedded styles sheets  
C) Write a short note on XML namespaces. **4**  
D) Differentiate between XML and HTML. **5**
- Q. NO 3.     A) Why do we use a DTD? Compare a Well formed XML document with a Valid XML **7**  
document.  
B) What is CSS? List the advantages and disadvantages of CSS.  
C) Write HTML code to create a form for displaying user details: **6**  
Name and Phone number should be accepted as textbox. Address should be accepted  
as textarea. For hobbies use options (“Reading”, “Cooking”, “Painting”) **7**

**PART B**

Answer any **TWO** questions from the following:

- Q. NO 4     A. With an example demonstrate the use of session object in JSP. **7**  
              B. Explain the benefits of AngularJS. **6**  
              C. Explain with an example, working and displaying array's in AngularJS **7**

- Q. NO 5. A. Explain what Model- View-Controller represents in an MVC application. **5**  
B. Explain the “ng-repeat” directive to show keys and values of an object in AngularJS. **5**  
C. Explain different types of arrays in PHP with examples. **5**  
D. Explain the string operator in PHP with examples. **5**
- Q. NO 6. A. Describe various features of PHP. Explain how PHP is better than other scripting languages. **5**  
B. Explain following form controls in AngularJS.  
i. Checkboxes  
ii. Radio Buttons  
iii. Combo Boxes/ Drop downs **10**  
C. Write a PHP program to print the following table. **5**

1	2	3	4	5	6	7	8	9	10
2	4	6	8	10	12	14	16	18	20
3	6	9	12	15	18	21	24	27	30
4	8	12	16	20	24	28	32	36	40
5	10	15	20	25	30	35	40	45	50
6	12	18	24	30	36	42	48	54	60
7	14	21	28	35	42	49	56	63	70
8	16	24	32	40	48	56	64	72	80
9	18	27	36	45	54	63	72	81	90
10	20	30	40	50	60	70	80	90	100

## **PART C**

Answer any **ONE** question from the following:

- Q. NO 7. A. With examples explain backgrounds, element dimensions and contextual selectors in **8** CSS.  
B. Write a JavaScript code to add, subtract, multiply and divide any two integer numbers. **6**  
C. Explain with an explain controllers in Angular JS. **6**
- Q. NO 8. A. Write a short note on using cookies in PHP. **5**  
B. Explain the following in JSP:  
i. Directives  
ii. Declaratives  
iii. Expression  
iv. Scriptlet  
C. Give an overview of HTTP. **5**

**Total Number of printed pages : 3**

**B.E. (Computer) Semester VII (Revised Course 2016-17)**

**EXAMINATION OCTOBER 2020**

**Web Technologies**

**[Duration : Two Hours ]**

**Instructions :**

**Total Marks 60]**

(a) Answer THREE FULL QUESTIONS with ONE QUESTION FROM EACH PART

2 Make suitable assumptions if required

**PART - A**

- Q. 1      (a) Describe the services supported by IID7. (5)  
(b) Demonstrate a JavaScript program to calculate the average of three numbers entered by the user. An alert message box should display the average on clicking of the Submit button. (5)  
(c) What are the advantages and disadvantages of CSS. (6)  
(d) Compare Internal DTD and External DTD with an example. (4)
- Q. 2      (a) Explain the Hierarchy of HTML documents. (6)  
(b) Write notes on  
    (i) History of the Web  
    (ii) types of browsers (6)  
(c) Design the below form using HTML. (8)

- Q. 3      (a) Compare XML elements and XML attributes. List the disadvantages of using attributes. (6)  
(b) Compare Class as Selector and ID as Selector with an example in CSS. (5)  
(c) Create an XNL to represent the items in a restaurant menu. each item record will have item

(2)

---

name, price and calories. Also create a DTD for the above XML file. Transform the above XML file to HTML using XSL (9)

**PART - B**

- Q. 4 (a) Explain Form validation and states in AugularIS (7)  
(b) Write notes on the following directives in AngularJS (6)  
    (i) ng-show (ii) ng-bind (iii) ng-model  
(c) Write a JSP cod which will retrieve Employee data (Employee id, Employee nname, department and Phone number) posted form a HTML file. (7)
- Q. 5 (a) Discuss the advantages of JSP over other languages (4)  
(b) Explain disabling sessions in JSP with an example. Is disabling sessions beneficial ? (6)  
(c) Write a PHOP program with a function to find the factorial of a number. Use a loop statement to calculate the factorial (5)  
(d) Describe five mathematical functions in PHP. (5)
- Q. 6 (a) Discuss the features of PHP (6)  
(b) Write notes on the following in PHP (8)  
    (i) associative arrays  
    (ii) Array lterator  
    (iii) Passing arguments by reference  
    (iv) Destorying variables  
(c) Write a PHP script to display th following pattern (6)
- \*  
\*\*  
\*\*\*  
\*\*\*\*

**PART - C**

- Q. 7 (a) Explain 3-Tier Web Architecture with a diagram. (5)  
(b) What are entities ? Compare internal DTD entities and External DTD entitles with syntax. (5)  
(c) What is dependency injection in AngularJS ? (4)  
(d) Write notes on the following file functions in PHP. (6)  
    (i) is link () (ii) flock () (iii) file ()
- Q. 8 (a) Write a HTML code which has a hyperlink image. The image should be aligned to the center. (5)  
Also set a border for the image.  
(b) Compare the different ways of defining the position of elements in CSS. (5)

(3)

---

(c) Describe the components of JSP .

(d) Describe the following PHP array functions

- (i) `array_multisort()`
- (ii) `array_shift()`
- (iii) `array_unique()`
- (iv) `array_merge()`

(6)

(4)

Total No. of Printed Pages:3

**B.E (Computer) Semester- VII (Revised Course 2016-17)**  
**EXAMINATION NOV/DEC 2019**  
**Elective - II**  
**Big Data Analytics**

[Duration : Three Hours]

[Total Marks :100]

**Instructions:**

- 1) Answer any two questions from each Part A & B & one question from Part C
- 2) Make necessary assumptions wherever necessary.

**Part- A**

- Q.1**
- a) Consider the following set of points  $\{(-2,-1), (1,1), (3,2)\}$  10
    - a) Find the least square regression line for the given data points
    - b) Plot the given points and the regression line in the same rectangular system of axis.
  - b) Discuss in detail about the importance of the Big Data Platform in Modern Data Analytics. 5
  - c) Describe any five characteristics of Big Data. 5

- Q.2**
- a) Life of bulbs produced by two factories A and B are given below. 10

Length of Life (in hours )	Factory A ( number of bulbs )	Factory B (number of bulbs)
550-650	10	8
650-750	22	60
750-850	52	24
850-950	20	16
950-1050	16	12
<b>Total →</b>	<b>120</b>	<b>120</b>

The bulbs of which factory are more consistent from the view point of length of life?

- b) “Massive Parallel Processing Systems” (MPP) are faster in processing big data” Justify the statement given and also show how Data Scoring is performed. 5
- c) Compare and contrast public and private cloud with the help of appropriate architecture. 5

- Q.3**
- a) A random sample of size 5 is taken from population containing 100 units. If sample observations are 10,12,13,07,18 find estimate of standard error. 10
  - b) Given Mean = 12 , standard deviation = 4 , Z (from 0 to 1 )=0.3413 and Z ( from 0 to 2 )= 0.4772. Calculate probability when value of “x” is greater than 20. Highlight the required area under bell curve to show the normal distribution. 5
  - c) Distinguish between probabilistic and non- probabilistic sampling 5

### Part- B

- Q.4**
- a) What is the purpose of “Flajolet Martin” algorithm? Given a stream of ‘n’ elements with **10** ‘m’ of them unique, what are the time and space constraints of this algorithm? Determine distinct elements in a stream 1,3,3,1,2,3,4,3,4,2,1,3 using Flajolet Martin algorithm. Given hash function as  $h(x) = (6x+1) \bmod 5$ . **5**
  - b) A set consists of some elements 9 & 11..... given, size (m) =5 and hash functions  $h_1(x)=x \bmod 5$  and  $h_2(x)=(2x+3)\bmod 5$  check whether 15 & 16 lie in this set or not using Bloom filter Technique. What are the semantics of false positive result in bloom filter? **5**
  - c) Discuss various issues related to stream processing. **5**
- Q.5**
- a) What is a page rank? Determine all the steps used to compute rank of a web page? Using **10** web graph given below, compute the required page rank considering damping factor as 0.80. Consider page rank for all pages as 1 at start.  
Note : A,B,& C different web pages.
- 
- Q.6**
- b) With suitable example explain K means algorithm. **5**
  - c) Discuss how frequent pattern – based clustering algorithms can be used in recommendation systems. **5**
- Q.6**
- a) Consider the following transactional database, with set of items:  
 $I=\{I_1, I_2, I_3, I_4, I_5\}$ . Let minimum support be 40% and confidence 60 %. Find all frequent item sets using Apriori Algorithm. **10**

TID	List of items
T1	I1, I2, I5
T2	I2, I4
T3	I2, I3
T4	I1, I2, I4
T5	I1, I3
T6	I2, I3
T7	I1, I3
T8	I1, I2, I3, I5
T9	I1, I2, I3
T10	I2, I4, I5

- b) Distinguish between PCY, Multistage & Multi – hash Algorithm.  
 c) Explain clearly how CURE algorithm can be used cluster Big Data Sets.

**5**

### **Part – C**

- |            |                                                                                                                                          |           |
|------------|------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>Q.7</b> | a) Using Hadoop architecture explains File write operations in detail.                                                                   | <b>10</b> |
|            | b) With suitable example, explain Hadoop MapReduce, illustrating role of combiners and partitioners in addition to Mappers and Reducers. | <b>5</b>  |
|            | c) Distinguish between Hadoop 1. X and Hadoop 2.X                                                                                        | <b>5</b>  |
| <b>Q.8</b> | a) Write Hive Queries to perform following task<br>a) Creation of a table<br>b) Joining of table                                         | <b>10</b> |
|            | b) Discuss advantages of MongoDB over RDBMS.                                                                                             | <b>5</b>  |
|            | c) Write short notes on visual data analysis techniques.                                                                                 | <b>5</b>  |

Total No. of Printed Pages:3

**B.E. (Computer ) Semester- VII (Revised Course 2016-17)**  
**EXAMINATION AUGUST 2020**  
**Big Data Analytics**

[Duration : Two Hours]

[Total Marks : 60]

**Instructions:**

- 1) Answer THREE FULL QUESTIONS with ONE QUESTION FROM EACH PART.
- 2) Figures to the right indicate full marks.
- 3) Draw neat diagram wherever necessary.
- 4) Make suitable assumptions wherever necessary.

**PART A**

Q.1	A. Define Big Data. Describe the characteristic of Big Data.	5
	B. Differentiate between Data Analysis & Data Reporting.	4
	C. Explain RTAP. Deliberate the importance of timely decision making.	4
	D. Given are four hash functions for a n = 20 bit Bloom Filter. $h_1(x) = ((3x + 2) \bmod 30) \bmod n$ $h_2(x) = ((5x + 7) \bmod 20) \bmod n$ $h_3(x) = ((6x + 4) \bmod 25) \bmod n$ $h_4(x) = ((4x + 3) \bmod 40) \bmod n$ 1. Work out the bit vectors for a Bloom Filter with the following input 20, 17, 100, 7, 3. 2. Analyse how the following inputs will be filtered by the above Bloom Filter 14, 8, 22, 9, 6	7
Q.2	A) Summarize the most commonly used data analytical tools used by Big Data Analyst.	6
	B) Differentiate between Big Data Analytics & Conventional Data Analytics.	4
	C) Justify why web data has become the most popular & important source of Big Data.	5
	D) Apply the FM Algorithm to approximate the number of unique elements in the given input stream. $X = \{1, 4, 2, 1, 4, 4, 4, 4, 1, 2, 4, 1, 7\}$ Hash Function $h(x) = 2x + 1 \bmod 32$ .	5
Q.3	A) Explain Analytic Sandbox & discuss the types & benefits of analytic sandbox.	10
	B) Define a DataStream. Model. Explain the characteristics of a generic model that attempts to store and retrieve data streams.	6
	C) Explain the pros & cons of statistical data analysis.	4
Q.4	A) Define a frequency item set with an example. Explain its applications.	6
	B) Distinguish between PCY, Multistage & Multihash Algorithm.	6
	C) Consider the following transactional database, with set of items $I = \{I_1, I_2, I_3, I_4, I_5\}$ . Let minimum support be 40% and confidence 60%. Find all frequent item sets using Apriori Algorithm.	8

**PART B**

TID	List of items
t1	{I1, I3, I4}
t2	{I2, I3, I5}
t3	{I1, I2, I3, I5}
t4	{I2, I5}
t5	{I1,I2,I3,I5}

- Q.5      a. Explain the following terms with respect to frequent item set mining with examples. (4x3)=12  
 1. Association Rules  
 2. Support & Support Threshold  
 3. Confidence  
 b. Describe the working of Map reduce with a relevant example. 8

- Q.6      A) Cluster the given sample data into two clusters by applying K- Means Algorithm.  
 Choose P2 & P10 as the initial centroids. 10

Choose	Height	weight
P1	185	72
P2	170	56
P3	168	60
P4	179	68
P5	182	72
P6	188	77
P7	180	71
P8	180	70
P9	183	84
P10	180	88
P11	180	67
P12	177	76

- B) Differentiate between MapReduce & Apache PIG. 5  
 C) Explain the decaying widow algorithm with an example. 5

### **PART C**

- Q.7      A) Define HDFS? Explain in brief about the basic building blocks of Hadoop? 8  
 B) Why is NoSQL essential in handling big data? 2  
 C) Explain clearly how CURE algorithm can be used cluster Big Data sets. How is it better from K Means Algorithm? 10
- Q.8      A) Apply PCY algorithm on following transaction to find the frequent item candidate sets. 8  
 Let minimum support be 30%. Hash function  $h(x) = (i * j) \bmod 5$

Transaction No	Itemset
1	1,3,4,5
2	2,3,4,
3	1,2,3,5
4	2,4
5	4,5,6
6	1,3,5
7	2,4,6
8	2,4,5

B) Explain the two environments of cloud computing. How is grid computing different from cloud computing? 8

C) Describe some of the risks associated with Big Data. 4