No mobile phones

15MCSEE06/15MCNEE08

USN | 1 | M | S | | | | | | | | |

M S RAMAIAH INSTITUTE OF TECHNOLOGY

(AUTONOMOUS INSTITUTE, AFFILIATED TO VTU)
BANGALORE – 560 054

SEMESTER END EXAMINATIONS – JANUARY 2016

M.Tech- Computer Science & Engg./ Computer Network Engg.

Course & Branch : **Semester :** **I**

Subject : **Big Data and Data Science** **Max. Marks :** **100**

Subject Code : **15MCSEE06/15MCNEE08** **Duration :** **3 Hrs**

Instructions to the Candidates:

- Answer one full question from each unit.

UNIT – I

1. a) Under the topic on Big Science, elucidate the features of "The Large Hader Collider (LHC) Experiments" – by CERN. (10)
   b) Give a brief description of the following Big Data Areas of Application. (10)
   i. In Education **ii** in urban planning

2 a) Justify that "Privacy" is one of the Challenges in Big Data. (12)
   b) Data scientists' knowledge on machine learning has three broad classes: Identify these three classes and explain each. (08)

UNIT –II

3. a) What do you understand by Structured Data? Bringing out the salient points in this. (10)
   b) R1,R2, R3,...Rm and S1,S2,S3,...Sn are fragments of Relations R & S respectively. JP: is joint predicate and Output:T1,T2,T3,....Tn: Result fragments. Write **PAJ** algorithm to create $\mathbf{R_j} \leftarrow \mathbf{U_1^m}\ \mathbf{[R_{ij}]}$ and also **Tj** is represented by $\mathbf{T_j} \leftarrow$ Join $\mathbf{[R_j, S_j, JP]}$. (10)

4. a) Identify the five types of data under Master Data Management in Big Data and briefly describe each. (10)
   b) From the table given below evaluate the following: (05)

$\pi_{Id, Name}\ (\sigma_{Hobby='Drawing'\ OR\ Hobby='Tv'}\ (Person))$

Person ID	Name	Address	Hobby
2020	Smith	72 Maple	Drawing
2020	Smith	72 Maple	TV
7008	Chrles	100 third	Riding
8352	Marco	20 Ring	Drawing

   c Illustrate the concept of Parallel Associative Join between four nodes through a diagram and associated equation. (05)

UNIT – III

5. a) Describe the four steps involved in k-means algorithm (08)
   b) Bring out the important aspects of Google's Page Ranking and write a typical equation involving the damping factor "d", and outbound links (12)

6. a) i. Draw the block diagram of Text Processing and explain the function of each block. (12)
ii. Write the equation representing each element of a document vector under ontology

b) i. What are the steps involved in designing a learning system under Machine Learning (08)
ii. Explain Teacher vs Learner Controlled experience

UNIT – IV

7. a) Distinguish between Name nodes and Data nodes in HDFS and draw the diagram of HDFS architecture (12)

b) Under the R-language evaluate the following: (08)

i. > x <- c(1:10)>
> x[(x>8) | (x<5)]

ii. > a = (1+1 ==3)
>

iii. > m <- matrix(1:12,nrow=3)

8. a) Using R Language Solve the following: (10)
For the following Character String evaluate:
i. > as.character(a) and ii. > as. Integer(a)
> a

[1] Kolon(Rektum)	Magen	Magen
[4] Magen	Magen	Retroperitoneal
[7] Magen	Magen(retrogastral)	Magen

b) i. Explain the following under MapReduce: Map Step, Reduce Step (10)
ii. Draw the diagram of Hadoop Distributed File System

UNIT – V

9. a) Under the case studies of Big Data give an account of N-Gram and how it works. (10)

b) Identify and explain IBM solutions available for use with IBM Netezza Analytics (10)

10. a) Explain the following under the benefits of Cassandra: (10)
Elastic Scalability, Always on architecture, Fast Linear Scale Performance, Flexible data storage, Easy Data Distribution

b) What solutions does RainStor Big Data database provides for business organizations? (10)

****