Total No. of Questions : 6]

SEAT No. :

P1448

[Total No. of Pages : 2

T.E./Insem/Apr-148
T.E. (I.T) (Semester-II)
DATA SCIENCE AND BIG DATA ANALYTICS
(2015 Pattern)

Time : 1 Hour] ***[Max. Marks : 30***

Instructions to the candidates:

1) ***Answer Q1 or Q2, Q3 or Q4, Q5 or Q6.***
2) ***Neat diagrams must be drawn wherever necessary.***
3) ***Figures to the right side indicate full marks.***
4) ***Assume suitable data if necessary.***

Q1) a) Explain shared-everything and shared-nothing architectures in detail with respect to Big Data. **[5]**

b) Explain 5V's for defining Big Data along with the factors responsible for data explosion? **[5]**

OR

Q2) a) List and explain choices for reengineering the data warehouse. **[5]**

b) Discuss the processing complexities associated with the big data. **[5]**

Q3) a) Assume that a man's profession can be classified as professional, skilled labourer or unskilled labourer. Assume that of the sons of professional men, 80 percent are professional, 10 percent are skilled labourers, and 10 percent are unskilled labourers. In the case of sons of skilled labourers, 60 percent are skilled labourers, 20 percent are professional and 20 percent are unskilled. Finally, in the case of unskilled labourers, 50 percent of the sons are unskilled labourers, and 25 percent each are in the other two categories. Assume that every man has at least one son, and form a Markov chain by following the profession of a randomly chosen son of a given family through several generations. Set up the matrix of transition probabilities. Find the probability that a randomly chosen grandson of an unskilled labourer is a professional man. **[6]**

b) Explain Flajolet Martin Algorithm. List the limitations of algorithm and how will you overcome these limitations? **[4]**

OR

P.T.O.

Q4) a) A computer system can operate in two different modes. Every hour, it remains in the same mode or switches to a different mode according to the transition probability matrix **[6]**

P = [0.4, 0.6
0.6 0.4]

i) Compute the 2-step transition probability matrix.

ii) If the system is in mode I at 5:30pm, what is the probability that it will be in mode I at 8 : 30 pm on the same day?

b) Explain following terms. **[4]**

i) Expectation

ii) Pair wise independence

Q5) a) Explain the following terms **[6]**

i) Google File System

ii) Heartbeat mechanism in HDFS

b) Differentiate between SQL and NoSQL databases with example. What is the need to develop Big Data applications using NoSQL databases? **[4]**

OR

Q6) a) Explain HDFS Read & Write operations in detail. **[6]**

b) What is the role of Sorter, Shuffler and Combiner in Map reduces Paradigm? **[4]**