

B.Tech. SEM -VII (Computer) 2014 Course (CBCS) : WINTER - 2018
SUBJECT: BIG DATA ANALYTICS & ARCHITECTURE

Day: Wednesday
Date: 28/11/2018

W-2018-2538

Time: 02.30 PM TO 05.30 PM
Max Marks: 60

N.B.:

- 1) All Questions are **COMPULSORY**
- 2) Use of non-Programmable calculator is **ALLOWED**
- 3) Figures to the right indicate **FULL MARKS**
- 4) Assume Suitable data, if necessary.

-
- Q.1** What is Big data? Describe any five characteristics of big data. With neat diagram explain in detail the big data management lifecycle. (10)
- OR**
- Q.1** List and explain in detail the various sources of structured data and unstructured data in big data environment. Also illustrate the role of relational databases in big data. (10)
- Q.2** Elaborate on Bayes Rule/Theorem. Show using an example how Bayes rule is used for classification. (10)
- OR**
- Q.2** A bag has three coins in it. One of them is a fair coin, but the others are biased trick coins. When flipped, the three coins come up heads, with probability 0.5, 0.6, and 0.1 respectively. Suppose that you pick one of these three coins uniformly at random and flip it three times:
a) What is $P(HTT)$? (That is, it comes up heads(H) on the first flip and tails(T) on the second and third flips).
b) Assuming that the three flips, in order, are HTT, what is the probability that the coin you picked was a fair coin? (10)
- Q.3** Describe in detail the Big data analytics lifecycle with neat diagram. Explain any one tool used in each phases of the big data analytics lifecycle. (10)
- OR**
- Q.3** What do you understand by clustering? Illustrate with an example the various steps and commands involved for performing the k-means clustering in R. (10)
- Q.4** Is there a necessity of distributed computing infrastructure for big data? Explain. Also elaborate on how big data influences the economics of computing platforms? (10)
- OR**
- Q.4** Describe in detail the big data technology stack with neat diagram. (10)
- Q.5** Explain the concepts of Map reduce using an example. Write a Map reduce pseudocode for "Group By" aggregation in a database. (10)
- OR**
- Q.5** Describe in detail the Hadoop ecosystem with its core components. Explain its physical architecture. State the limitations of Hadoop. (10)
- Q.6** Explain in detail with example Collaborative and Content based filtering in a recommendation system. (10)
- OR**
- Q.6** What is sentiment analysis? Illustrate the steps involved in performing 'text analysis process' using sentiment analysis. (10)