

Reg. No. :

--	--	--	--	--	--	--	--	--	--

Question Paper Code: 49217

B.E./B.Tech. DEGREE EXAMINATION, APRIL 2019

Elective

Computer Science and Engineering

14UCS917 - MASSIVE DATASET ANALYTICS

(Regulation 2014)

Duration: Three hours

Maximum: 100 Marks

PART A - (10 x 1 = 10 Marks)

(Answer all Questions)

- Which of the following is a characteristics of Big Data ? CO1- R
(a) Huge volume of data (b) complexity of data types and structures
(c) Speed of data creation and growth (d) all of the above
- Near real time processing deals with _____ characteristics of data. CO1- R
(a) velocity (b) value (c) storage (d) volatility
- The effectiveness of an SVM depends upon CO2- R
(a) Selection of Kernel (b) Kernel Parameters
(c) Soft Margin Parameter C (d) All of the above
- Fuzzy logic is in the form of _____ CO2- R
(a) Two-valued Logic (b) Crispest logic
(c) Many – valued Logic (d) Binary set Logic
- Using of the main memory as a bit array is called CO3- R
(a) Bloom filter (b) Window filter (c) Blur filter (d) Drop filter
- Bloom filter consists of _____. CO3- R
(a) Array (b) Vector
(c) Key values (d) Both A & C

7. Market-basket problem was formulated by _____. CO4- R
 (a) Agrawal et al (b) Steve et al. (c) Toda et al (d) Simon et al
8. The best known family of clustering algorithms is CO4- R
 (a) A-priori (b) Limited pass (c) K-means (d) Multihash
9. _____ was the first to publicize MapReduce – a system they had used to scale their data processing needs. CO5- R
 (a) Yahoo (b) Google (c) Microsoft (d) Linux
10. The Hadoop MapReduce framework spawns one map task for each _____ generated by the InputFormat for the job. CO5- R
 (a) OutputSplit (b) InputSplit
 (c) InputSplitstream (d) All of the mentioned

PART – B (5 x 2= 10Marks)

11. How will you apply the methods of re-sampling in Big Data? CO1- R
12. List out the usages of Regression Modeling CO2- R
13. Define Frequent Item set. CO3- R
14. Explain Clustering Using Map-Reduce? CO4- R
15. What is the use of Hive in Hadoop CO5- R

PART – C (5 x 16= 80Marks)

16. (a) Consider a sample of size 10 from a population with fraction of successes $\pi=0.8$. What is the sampling distribution of Fr , the sample fraction of successes? CO1- App (16)
 Or
 (b) Briefly describe some important resampling techniques. CO1- App (16)
17. (a) Explain with an example support vector and kernel methods. CO2- U (16)
 Or
 (b) How do use the generalization techniques needed to illustrate neural networks? CO2- U (16)
18. (a) Explain how to count ones in a window using DGIM algorithm. Also explain how query answering is done using DGIM algorithm. Use the stream 1 0 1 0 1 1 0 0 0 1 0 1 1 1 0 1 1 0 0 1 0 1 1 0. CO3- U (16)

Or

- | | | | |
|-----|---|---------|------|
| | (b) Explain in detail about Alon-Matias-Szegedy algorithm for second moments | CO3-U | (16) |
| 19. | (a) Explain how to run the A-Priori Algorithm using the data mining library Patterns. | CO4-Ana | (16) |
| | Or | | |
| | (b) Explain the algorithm for clustering in non-euclidean spaces. | CO4- U | (16) |
| 20. | (a) Explain in detail the Ecosystem of the Hadoop Framework | CO5- U | (16) |
| | Or | | |
| | (b) (i) Describe in detail about Interaction techniques | CO5- U | (8) |
| | (ii) Briefly explain about Visualization techniques | CO5- U | (8) |

