

Reg. No. :

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

Question Paper Code: 49217

B.E./B.Tech. DEGREE EXAMINATION, NOV 2018

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)

1. Which is the largest phase in data analytics lifecycle CO1- R
(a) Model Planning (b) Model Building (c) Data Preparation (d) Operationalize
2. Near real time processing deals with _____ characteristics of data. CO1- R
(a) velocity (b) value (c) storage (d) volatility
3. 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
4. A fuzzy set is associated with a _____ CO2- R
(a) Linguistic variable (b) Certainty factor
(c) Hypothesis to be tested (d) Linguistic value
5. 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

6. Stream Sources are use in _____ data. CO3- R
 (a) Image Data (b) Sensor Data
 (c) Huge Data (d) Both A & B
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. _____ has the world’s largest Hadoop cluster. CO5- R
 (a) Apple. (b) Datamatics.
 (c) Facebook. (d) None of the mentioned

PART – B (5 x 2= 10Marks)

11. What is analytic scalability? CO1- R
12. List out the usages of Regression Modeling CO2- R
13. Define Frequent Item set. CO3- R
14. Define Clustering. CO4- R
15. What is the use of Hive in Hadoop CO5- R

PART – C (5 x 16= 80Marks)

16. (a) (i) Discuss the challenges with Big Data. CO1- U (8)
 (ii) What are the characteristics of Big Data? Discuss CO1- U (8)
- 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) Describe the procedure to build a fuzzy decision tree with an example CO2- U (16)
18. (a) Explain the architecture of stream data model. CO3- U (16)
Or
(b) Explain in detail about Alon-Matias-Szegedy algorithm for second moments CO3-U (16)
19. (a) Examine how the data is processed in BFR Algorithm are generated from frequent itemsets. 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) Describe the various visualization techniques that can be used for visualizing data. CO5- U (16)

