Reg. No. : | | | | | | | | | | |

Question Paper Code: 49217

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

1. Near real time processing deals with _____ characteristics of data. CO1- R

   (a) velocity (b) value (c) storage (d) volatility

2. Near real time processing deals with _____ characteristics of data. CO1- R

   (a) velocity (b) value (c) storage (d) volatility

3. How the bayesian network can be used to answer any query? CO2- R

   (a) Full distribution (b) Joint distribution

   (c) Partial distribution (d) All of the mentioned

4. Fuzzy logic is in the form of _____ CO2- R

   (a) Two-valued Logic (b) Crispest logic

   (c) Many – valued Logic (d) Binary set Logic

5. The times by which stimuli must be processed and some response produced by the system is known as CO3- R

   (a) Compile time (b) Frequency (c) Deadlines (d) Execution time

6. 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. Which of the following clustering require merging approach? CO4- R

(a) Partitional (b) Hierarchical

(c) Naïve Bayes (d) None of the Mentioned

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. "Sharding" a database across many server instances can be achieved with CO5- R

(a) LAN (b) SAN (c) MAN (d) All of the above

PART – B (5 x 2= 10Marks)

11. Explain Analysis vs Reporting CO1- R

12. Define Bayesian Modeling. CO2- R

13. What are the applications of Realtime Analytics Platform(RTAP)? 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 linear, quadratic and cubic model as the possible candidates to be selected as the model with lowest prediction error and use leave- one-out cross validation to compute prediction error and select appropriate model. 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) (i) How to use the principal component analysis for the feature reduction? CO2- U (8)

(ii) Explain the process of extracting fuzzy models from data? CO2- U (8)

18.	(a)	Explain how to estimate moments using AMS algorithm. Suppose the stream is c, e, c, f, a, e, g, f, f, b, b, c, g, b, a, a, f, d, a, e. The length of the stream is n = 20 and consider the random positions 5, 9, 13. Illustrate the working of the AMS algorithm for this stream.	CO3- U	(16)
	(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)	Describe in detail about Hierarchical Clustering methods.	CO4- U	(16)
20.	(a)	Describe the various visualization techniques that can be used for visualizing data.	CO5- U	(16)
		Or		
	(b)	(i) Write Short notes on MapReduce and Sharding	CO5- U	(8)
		(ii) Explain in detail about Hadoop Distributed file systems?	CO5- U	(8)