Reg. No. :

Question Paper Code: 49217

B.E./B.Tech. DEGREE EXAMINATION, APRIL 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. _____ is an index plot of the principal component variances.   CO2- R

(a) Vector diagram   (b) Decision Tree   (c) Binary Tree   (d) Scree diagram

4. _____ denotes the process of cooling a molten substance   CO2- R

(a) Inference   (b) Regression   (c) Induction   (d) Annealing

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. RTAP stands for   CO3- R

    (a) Real Time Analytics Platform   (b) Run Time Analytics Process

    (c) Real Time Analytics Platform   (d) None of the above

7. _____ model of data is used to describe a common form of   CO4- R
    many to many relationship between two kinds of objects.

    (a) Hierarchical   (b) Market-basket   (c) Structured   (d) Vertical

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   CO5- R
    had used to scale their data processing needs.

    (a) Yahoo   (b) Google   (c) Microsoft   (d) Linux

10. A single point of failure of Hadoop cluster is _____.   CO5- R

    (a) NameNode   (b) DataNode   (c) EditLog   (d) DataLog

PART – B (5 x 2= 10Marks)

11. What is analytic scalability?   CO1- R

12. Specify the main idea of Kernel Methods   CO2- R

13. Give examples of stream sources.   CO3- R

14. State the importance of A-Priori Algorithm.   CO4- R

15. What do you mean by sharding?   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 the Bayesian Inference with suitable example.   CO2- U   (16)

                              Or

    (b) Describe the procedure to build a fuzzy decision tree with an CO2- U   (16)
        example

18. (a) Explain the architecture of stream data model.   CO2- U   (16)

                              Or

    (b) With an example, explain the Datar-Gionis-Indyk-Motwani   CO3-U   (16)
        Algorithm.

19. (a) How does the Multistage algorithm take more than two passes to CO4-Ana   (16)
        find the frequent pairs? Explain.

                              Or

    (b) Explain the algorithm for clustering in non-euclidean spaces.   CO4- U   (16)

20. (a) With neat diagram, explain the architecture of Hive.   CO5- U   (16)

                              Or

    (b) Describe the various visualization techniques that can be used for CO5- U   (16)
        visualizing data.

49217