

Reg. No. :

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

Question Paper Code: 49217

B.E./B.Tech. DEGREE EXAMINATION, AUGUST 2021

Elective

Computer Science and Engineering

14UCS917 - MASSIVE DATASET ANALYTICS

(Regulation 2014)

Duration: 1:45 hour

Maximum: 50 Marks

PART A - (10 x 2 = 20 Marks)

(Answer any ten of the following questions)

1. List the characteristics of big data and challenges in handling big data.
2. Write any two possible web data from which effective analysis can be carried out.
3. Highlight the uses of regression modeling.
4. Define principal component analysis.
5. Give any two examples for stream data.
6. State how to count the distinct elements in a stream.
7. List the different hierarchical clustering techniques.
8. Define K-Means clustering algorithm.
9. State the significances of Map Reduce.
10. List the components of Hadoop framework.
11. Write down the specification for Bloom's filter.
12. What are realtime analytics platform?
13. Highlight the uses of regression modeling.
14. What is "market-basket" model?

15. List out any four NoSQL databases.

PART – B (3 x 10= 30 Marks)

(Answer any three of the following questions)

16. Discuss the challenges with Big Data. (10)
17. Write short notes on Bayesian modeling. (10)
18. Explain in detail about Alon-Matias-Szegedy algorithm for second moments (10)
19. How does the Multistage algorithm take more than two passes to find the frequent pairs? Explain. (10)
20. Describe the various visualization techniques that can be used for visualizing data. (10)