Sub Code: CST-320 **ROLL NO……..……..**

II SEMESTER EXAMINATION, 2022 – 23

I[st] Year, M.Tech. – Computer Science & Engineering

BIG DATA ANALYTICS

Duration: 3:00 hrs **Max Marks: 100**

Note: - Attempt all questions. All Questions carry equal marks. In case of any ambiguity or missing data, the same may be assumed and state the assumption made in the answer.

Q 1.	**Answer any four parts of the following.** a) Define Big Data. Explain the Evolution of Big Data and their characteristics. b) Explain the difference between analysis vs reporting. c) Describe the main features of a big data platform. List out any five platform. d) Explain the difference between structure and unstructured data. e) Discuss Big Data in Healthcare, Transportation & Medicine. f) Explain how big data processing differs from distributed processing.	**5x4=20**
Q 2.	**Answer any four parts of the following.** a) What is Real Time Analytics Platform (RTAP)? b) Explain the Data streaming concept in detail. c) Write CASE STUDY on STOCK MARKET PREDICTION. d) Explain with a neat diagram about Stream data model and its Architecture. e) Define the concept of Estimating Moments. f) Explain in detail on Counting ones in a Window.	**5x4=20**
Q 3.	**Answer any two parts of the following.** a) What are the advantages of Hadoop? Explain Hadoop Architecture and its Components with proper diagram. b) What is Map Reduce? Explain working of various phases of Map Reduce with appropriate example and diagram. c) Explain Job Scheduling in Map Reduce. How it is done in case of (i) The Fair Scheduler (ii) The Capacity Scheduler.	**10x2= 20**
Q 4.	**Answer any two parts of the following.** a) What do you mean by HiveQL Data Definition Language? Explain any three HiveQL DDL command with its syntax and example. b) What is Zookeeper? List the benefits of it. Differentiate: Apache pig Vs Map Reduce. c) Explain HBase architecture & Storage mechanism in HBase.	**10x2= 20**
Q 5.	**Answer any two parts of the following.** a) What is simple linear regression? Write a difference between simple linear & multiple linear regression. b) What is data visualization? Explain any four data visualization techniques. c) Describe the regression analysis. Predict the value of the dependent variable in case of linear regression	**10x2= 20**

****