CBCS SCHEME

USN | | | | | | | | | | |

17CS82

Eighth Semester B.E. Degree Examination, June/July 2023
Big Data Analytics

Time: 3 hrs. Max. Marks: 100

Note: Answer any FIVE full questions, choosing ONE full question from each module.

Important Note : 1. On completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages.
2. Any revealing of identification, appeal to evaluator and /or equations written eg, 42+8 = 50, will be treated as malpractice.

Module-1

1 a. With a neat diagram explain the components of the Hadoop Distributed File System (HDFS). **(10 Marks)**

b. With a neat diagram, describe the steps in the MapReduce parallel flow data model. **(10 Marks)**

OR

2 a. Write the Java code for MAP and REDUCE of word count problem. Describe the steps of compiling and removing the MapReduce program. **(10 Marks)**

b. Briefly explain HDFS Name Node federation, NFS Gateway, Snapshots, Checkpoint and backups. **(10 Marks)**

Module-2

3 a. With neat diagrams, explain the Oozie DAG workflow and the types of nodes in the workflow. **(10 Marks)**

b. Explain the features and benefits of apache HIVE in Hadoop. **(10 Marks)**

OR

4 a. How do you run MapReduce and Message Passing Interface (MPI) on YARN architecture? **(08 Marks)**

b. With neat diagram discuss the various frameworks that run under YARN. **(08 Marks)**

c. Discuss the various features of Hadoop YARN administration. **(04 Marks)**

Module-3

5 a. Write any five Business Intelligence (BI) applications for various sectors. **(10 Marks)**

b. Explain the star schema of design of Data Ware Housing with an example. **(07 Marks)**

c. What is a confusion matrix? Explain. **(03 Marks)**

OR

6 a. Explain with diagram CRISP-DM data mining cycle. **(10 Marks)**

b. What do you understand by the term Data visualization? How is it important in Big Data Analytics? **(05 Marks)**

c. Differentiate between Data Mining and Data Warehousing. **(05 Marks)**

Module-4

7 a. Explain the design principles of an artificial neural network. **(08 Marks)**

b. List the advantages and disadvantages of a regression model. **(06 Marks)**

c. What is a splitting variable? Describe three criteria for choosing a splitting variable. **(06 Marks)**

OR

8 a. Explain the design principles of an Artificial Neural Network. **(10 Marks)**

b. How does the apriori algorithm work? Apply the same for the following example.

T_{ID}	List of Item – IDs
T_{100}	I_1 , I_2 , I_5
T_{200}	I_2 , I_4
T_{300}	I_2 , I_3
T_{400}	I_1 , I_2 , I_4
T_{500}	I_1 , I_3
T_{600}	I_2 , I_3
T_{700}	I_1 , I_3
T_{800}	I_1 , I_2 , I_3 , I_5
T_{900}	I_1 , I_2 , I_3

Assume the support count = 2. **(10 Marks)**

Module-5

9 a. Compare text mining with data mining. **(06 Marks)**

b. What is Naïve Baye's technique? Explain its model. **(06 Marks)**

c. Explain steps in the text mining process and architecture. **(08 Marks)**

OR

10 a. What is Web mining? Explain the different types of Web mining. **(08 Marks)**

b. Write a short note on Social Network Analysis (SNA). Numerical examples on Naïve Baye's model, SYM and SNA (Rank calculation). **(06 Marks)**

c. Explain three types of Web mining. Use an appropriate flow diagram to represent the same. **(06 Marks)**

* * * * *