CBCS SCHEME

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

15CS82

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

Time: 3 hrs. Max. Marks: 80

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

Module-1

1 a. What are the various system roles in an HDFS development? Explain with the neat diagram. **(08 Marks)**

b. Illustrate any 8 HDFS commands and briefly explain. **(08 Marks)**

OR

2 a. Explain with neat diagram the parallel MapReduce data flow. **(08 Marks)**

b. Write the code for simple mapper script and simple reducer script. **(08 Marks)**

Module-2

3 a. Demonstrate 2 step Apache sqoop data import and export method with neat diagram. **(08 Marks)**

b. With a neat diagram, the Apache oozie workflow for Hadoop architecture. **(08 Marks)**

OR

4 a. What is Apache flume? Describe the features components and the working of Apache flume. **(08 Marks)**

b. Discuss the different views supported by Apache Ambari. **(08 Marks)**

Module-3

5 a. What is BI? List the different BI applications and explain in detail any five applications. **(08 Marks)**

b. Describe any 4 design considerations of Data warehousing. **(08 Marks)**

OR

6 a. Explain with diagram CRISP-DM Data mining cycle. **(08 Marks)**

b. Explain with neat diagram, different types of graphs. **(08 Marks)**

Module-4

7 a. Explain with a data set how to construct the decision tree. **(08 Marks)**

b. Using the data given in Data set shown in Table Q7(b), create a regression model to predict the Test 2 from Test 1 score. Then predict the score for the one who got 46 in Test 1.

Table Q7(b)

Test 1	Test 2
59	56
52	63
44	55
51	50
42	66
42	48
41	58

Test 1	Test 2
45	36
27	13
63	50
54	81
44	56
50	64
47	50

(08 Marks)

OR

8 a. What are the different design principles of artificial neural networks? **(08 Marks)**
b. Explain the K means clustering algorithm state its advantages and disadvantages. **(08 Marks)**

Module-5

9 a. Explain the test mining process. **(08 Marks)**
b. Explain the 3 types of web mining. Use appropriate flow diagrams to represent the same. **(08 Marks)**

OR

10 a. What is support vector machine? Explain its model. **(08 Marks)**
b. Describe Naïve Baye's model to classify the text data into class using following dataset Table Q10(b).

Training set	Document ID	Keyword in the Document	Class = h (Healthy)
	1	Love Happy Joy Joy Love	Yes
	2	Happy Love Kick Joy Happy	Yes
	3	Love More Joy Good	Yes
	4	Love Happy Joy Pain Love	Yes
	5	Joy Love Pain Kick Pain	No
	6	Pain Pain Love Kick	No
Test data	7	Love Pain Joy Love Kick	?

Table Q10(b) **(08 Marks)**

* * * * *