Save trees use e-Question Paper Go green
DOWNLOAD THIS FREE AT www.vturesource.com

CBCS SCHEME

USN | | | | | | | | | | 15CS82

Eighth Semester B.E. Degree Examination, November 2020
Big Data Analytics

Time: 3 hrs. Max. Marks: 80

Note: Answer any FIVE full questions irrespective of modules.

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 HDFS (Hadoop Distributed File System). **(08 Marks)**

b. Write and explain the mapper and reducer scripts for the MapReduce model. **(08 Marks)**

2 a. With a neat diagram, describe the steps of MapReduce parallel data flow. **(08 Marks)**

b. Explain the following roles in HDFS deployment with a diagram: (i) High availability (ii) Name Node Federation. **(08 Marks)**

Module-2

3 a. What is the significance of Apache pig in Hadoop context? Describe the main components and the working of Apache pig with a simple example. **(08 Marks)**

b. Explain the features and the benefits of Apache HIVE in Hadoop. **(08 Marks)**

4 a. With neat diagrams, explain the oozie DAG workflow and the types of nodes in the workflow. **(08 Marks)**

b. What is Apache Flume? Describe the features, components and the working of Apache Flume. **(08 Marks)**

Module-3

5 a. Draw the flow of BIDM cycle. Explain the strategic and operational decisions. **(08 Marks)**

b. Differentiate between datamart and datawarehouse based on the following with justifications:
(i) Scope (ii) Target organization (iii) Cost (iv) Approach
(v) Complexity (vi) Time **(08 Marks)**

6 a. Describe any 8 considerations for a data warehouse and explain the key elements with a diagrammatic representation. **(08 Marks)**

b. Explain the CRISP-DM cycle with a diagram. **(08 Marks)**

Module-4

7 a. Explain the steps and three differentiating criteria of a decision tree algorithm. Construct a decision tree for the following data set (table 1) and predict the outcome for the given question. **(10 Marks)**

Outlook	Temp	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No

www.vturesource.com
ALL BRANCHES | ALL SEMESTERS | NOTES | QUESTON PAPERS | LAB MANUALS
A Vturesource Go Green initiative

Save trees use e-Question Paper Go green
DOWNLOAD THIS FREE AT www.vturesource.com

Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No
Outlook	Temp	Humidity	Windy	Play
Sunny	Hot	Normal	True	?

Table 1 : Data Set

b. Differentiate between C4.5, CART and CHAID decision tree algorithms. **(06 Marks)**

8 a. Explain the design principles of ANN by constructing a model representation for a single and multilayer ANN. Describe the steps to build an ANN (Artificial Neural Network). **(10 Marks)**

b. For the dataset in table 2, find the affinities of the product-product which sell together. Consider S = 33%, C = 50% and 3 item-set level only. **(06 Marks)**

Transactions List				
1	Milk	Egg	Bread	Butter
2	Milk	Butter	Egg	Ketchup
3	Bread	Butter	Ketchup	
4	Milk	Bread	Butter	
5	Bread	Butter	Cookies	
6	Milk	Bread	Butter	Cookies
7	Milk	Cookies		
8	Milk	Bread	Butter	
9	Bread	Butter	Egg	Cookies
10	Milk	Butter	Bread	
11	Milk	Bread	Butter	
12	Milk	Bread	Cookies	Ketchup

Table Q8 (b)

Module-5

9 a. Compare text mining and data mining. **(08 Marks)**

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

10 a. Explain the text mining process and the architecture. **(08 Marks)**

b. Compute the rank values for the network in Fig. Q10 (b), when is the highest ranked node? **(08 Marks)**

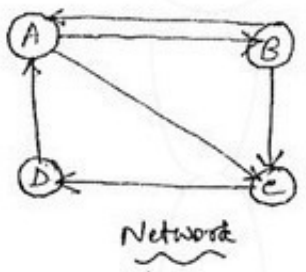

Fig. Q10 (b)

* * * * *

www.vturesource.com
ALL BRANCHES | ALL SEMESTERS | NOTES | QUESTON PAPERS | LAB MANUALS
A Vturesource Go Green initiative