USN | | | | | | | | | |

15CS82

Eighth Semester B.E. Degree Examination, Feb./Mar. 2022
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. Explain with a neat diagram the workflow of a Map reduce model. Also write a mapper and a reducer code to find the number of occurrence of the word "data" in the file "File.txt". **(09 Marks)**

   b. Explain the following :
      (i) Name mode high availability
      (ii) Speculation execution
      (iii) Difference between two version of hadoop. **(07 Marks)**

OR

2 a. An enterprise has 500 GB of unstructured data to be processed. Suggest a suitable architecture for this work. Explain with a diagram all the components of the file system supported by your solution. **(10 Marks)**

   b. Write HDFS commands for the following operations :
      (i) to find the version of HDFS
      (ii) download and install Hadoop
      (iii) to run postmap service
      (iv) to run zookeeper controller failover daemon
      (v) to start and stop namenode and datanodes
      (vi) to run mapper and reducer with input file **(06 Marks)**

Module-2

3 a. Explain any Nine different YARN application frameworks. **(09 Marks)**
   b. Explain the Hadoop tool to acquire data streams. **(07 Marks)**

OR

4 a. With a neat diagram, explain the role of Apache Oozie in Hadoop ecosystem. **(09 Marks)**
   b. Explain Apache Sqoop Import and Export methods using diagram. **(07 Marks)**

Module-3

5 a. What is data warehousing? Explain the data warehouse architecture. **(10 Marks)**
   b. How can raw data be cleansed and transformed before it can be used for data mining. **(06 Marks)**

OR

6 a. Explain various Business Intelligence applications. **(10 Marks)**
   b. How can one evaluate the results of data mining? **(06 Marks)**

Module-4

7 a. Develop a decision tree to evaluate a new store option:

City-Size	Income	Local Investors	Lohas Awarness	Decision
Big	High	Yes	High	Yes
Medium	Medium	No	Medium	No
Small	Low	Yes	Low	No
Big	High	No	High	Yes
Small	Medium	Yes	High	No
Medium	High	Yes	Medium	Yes
Medium	Medium	Yes	Medium	No
Big	Medium	No	Medium	No
Medium	High	Yes	Low	No
Small	High	No	High	Yes
Small	Medium	No	High	No
Medium	High	No	Medium	No

(10 Marks)

b. List down the advantages and disadvantages of artificial neural networks. (06 Marks)

OR

8 a. Given that house price depends on the size develop the regression model for the following data. Also find the house price when size is 2000 sq. ft.

House Price	Size (Sq.ft.)
229500	1850
273300	2190
247000	2100
195100	1930
261000	2300
179700	1710
168500	1550
234400	1920
168800	1840
180400	1720
156200	1660
288350	2405
186750	1525
202100	2030
256800	2240

(07 Marks)

b. Find the affinity among the following items. Suggest which items can be marketed as a single package for higher sales? Assume support level of 33% and confidence level of 50%.

Milk	Egg	Bread	Butter
Milk	Butter	Egg	Ketchup
Bread	Butter	Ketchup	
Milk	Bread	Butter	
Bread	Butter	Cookies	
Milk	Bread	Butter	Cookies
Milk	Cookies		
Milk	Bread	Butter	
Bread	Butter	Egg	Cookies
Milk	Butter	Bread	
Milk	Bread	Butter	
Milk	Bread	Cookies	Ketchup

(09 Marks)

Module-5

9 a. What is text mining? Explain the architecture. **(07 Marks)**

b. What is web mining? What are the various types of web mining? **(09 Marks)**

OR

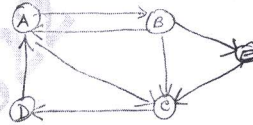**10** a. Compute the rank values of the modes for the following network shown in Fig.Q10(a). Which is the highest rank mode?

Fig.Q10(a) **(10 Marks)**

b. Discuss various topologies of social networks. **(06 Marks)**

* * * * *