CBCS SCHEME

USN ☐☐☐☐☐☐☐☐☐☐ 17CS82

Eighth Semester B.E. Degree Examination, Jan./Feb. 2023
Big Data Analytics

Time: 3 hrs. Max. Marks: 100

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

Module-1

1   a. What is HDFS? List and explain the components of HDFS. **(10 Marks)**
    b. Write the Java code for MAP and Reduce of word count problem. Describe the steps of compiling and Removing the Map Reduce program. **(10 Marks)**

OR

2   a. With example, explain the different general HDFC commands. **(08 Marks)**
    b. Elaborate the steps required for Map Reduce parallel data flow. **(06 Marks)**
    c. How the problem of Name Node as a single point of failure is overcome in Hadoop installations? **(06 Marks)**

Module-2

3   a. With a neat diagram describe the Two-Step Apache Sqoop and Export and import method. **(10 Marks)**
    b. Define Apache pig. List and explain the usage modes of Apache pig. **(10 Marks)**

OR

4   a. With a neat diagram, explain 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)**

Module-3

5   a. Illustrate and explain the Architecture of Data-warehouse with the help of a neat diagram. **(08 Marks)**
    b. Why should organizations invest in business intelligence solutions? **(06 Marks)**
    c. Draw the flow of BIDM cycle. Explain the strategic and operational decisions. **(06 Marks)**

OR

6   a. What are the objectives for graphical excellence in data visualization suggested by Tufle? Illustrate the same with an example. **(08 Marks)**
    b. Explain CRISP-DM Data Mining cycle. **(06 Marks)**
    c. Describe the process of selecting and cleansing of data for mining. **(06 Marks)**

Module-4

7   a. Explain the design principles of Artificial Neural Network (ANN) by constructing a model representation for a single and multilayer ANN. Describe the steps to build an ANN. **(10 Marks)**
    b. Construct a decision tree to predict the play decision given the atmospheric conditions. (Data set – Table Q7(b)).

Outlook	Temperature	Humidity	Windy	Play
Sunny	Hot	Normal	True	?

Outlook	Temperature	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
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

(10 Marks)

OR

8 a. Discuss the three key elements which differentiate the algorithms for decision making.
   (08 Marks)
   b. Explain the advantages and disadvantages of Regression models. **(06 Marks)**
   c. Explain K-Means algorithm for clustering. **(06 Marks)**

Module-5

9 a. Explain three types of web mining. Use appropriate flow diagram to represent the same.
   (08 Marks)
   b. What is Naïve Bayes technique? Explain its model. **(06 Marks)**
   c. Discuss the application and practical consideration of social network analysis. **(06 Marks)**

OR

10 a. Explain the Text Mining process and the Architecture. **(10 Marks)**
   b. Compute the rank values for the network in the Fig.Q10(b). When is the highest ranked code? Solve the same with 8 iterations.
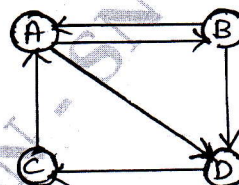

Fig.Q10(b)

(10 Marks)

* * * * *