Seventh Semester B.E. Degree Examination, Dec.2024/Jan.2025

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. Describe the data, big data and web data. Explain the different classification of big data. **(08 Marks)**

　 b. Illustrate the 3V characteristics of big data? Explain the challenges faced from large growth in volume of data. **(06 Marks)**

　 c. Briefly explain the steps involved in data processing. **(06 Marks)**

OR

2. a. With the proper illustration, deduce how the data is managed for analysis. **(06 Marks)**

　 b. Explain how the Berkley's big data stack helps in analytics. **(06 Marks)**

　 c. Explain the importance of data quality in discovering new knowledge and decision making. **(08 Marks)**

Module-2

3. a. Explain with a neat diagram the different core components of Hadoop. **(07 Marks)**

　 b. List and explain the different features of Hadoop. **(07 Marks)**

　 c. Consider a data storage for university student each student data, stuData which is in a file of size less than 64 Mb, a data block stores the full file data for a student of stuData.idN where N = 1 to 500.
　　(i) How the files of each student will be distributed in a Hadoop cluster.
　　(ii) How many student data can be stored at one cluster.
　　(iii) What is the memory capacity of the cluster in TB and data nodes in each rack.
　　(iv) What shall be the changes when the stud file size ≤ 128 Mb.
　　Assume that each rack has 2 data nodes for processing each of 64 Mb memory. Assume the cluster consists of 120 racks. **(06 Marks)**

OR

4. a. With a neat diagram, explain the function of client, resource manager, node manager application master and containers in YARN based execution model. **(06 Marks)**

　 b. Explain the different co-ordination service provided by the zookeeper. **(06 Marks)**

　 c. Explain the features of the following :
　　i) Flume　　　　ii) Oozie **(08 Marks)**

Module-3

5. a. List and explain the different features of distributed computing architecture. **(08 Marks)**

　 b. Explain the different NoSQL data store and their characteristics features. **(06 Marks)**

　 c. With an inference made, explain how does CAP theorem hold in NoSQL database. **(06 Marks)**

OR

6. a. Explain the different characteristics of schema-less model. **(06 Marks)**

　 b. Explain the different ways for handling the big-data problems. **(06 Marks)**

　 c. Demonstrate the working of shared nothing architecture for big-data tasks, using master-slave distribution model. **(08 Marks)**

Module-4

7. a. With a neat sketch, explain the different execution steps in MapReduce. **(06 Marks)**

　 b. Deduce how the MapReduce handle the failure of Map TaskTracker, Reduce TaskTracker and Master JobTracker. **(06 Marks)**

　 c. Explain the following with respect to MapReduce:
　　i) Counting and summing
　　ii) Sorting
　　iii) Colating
　　iv) Filtering or parting **(08 Marks)**

OR

8. a. With a neat diagram, explain the architecture of Hive. **(08 Marks)**

　 b. With a neat illustration, deduce the creation of partitions and buckets in a Hive database. **(06 Marks)**

　 c. With a neat sketch, explain the Pig architecture for scripts dataflow and processing. **(06 Marks)**

Module-5

9. a. Explain the significance of performing t-Test and F-test. **(06 Marks)**

　 b. Explain the KNN regression analysis. **(06 Marks)**

　 c. What are the features of Apriori principle that enable frequent itemset mining? **(04 Marks)**

　 d. Explain any 2 fields where the application of association rule fits. **(04 Marks)**

OR

10. a. Explain the different phases for mining the text. **(10 Marks)**

　 b. With a neat sketch, explain the taxonomy of web mining. **(10 Marks)**

* * * * *