WEST BENGAL STATE UNIVERSITY
B.Sc. Honours 6th Semester Examination, 2022

CMSADSE04T-COMPUTER SCIENCE (DSE3/4)

Time Allotted: 2 Hours　　　　　　　　　　　　　　　　　　　　　Full Marks: 40

The figures in the margin indicate full marks.
Candidates should answer in their own words and adhere to the word limit as practicable.

GROUP-A

1.　Answer any *four* questions from the following:　　　　　　　　2×4 = 8

　(a) What is sequence file in Hadoop?

　(b) Define the three key design principles of Pig Latin.

　(c) Define the various file formats supported in HIVE.

　(d) What is the difference between analysis and analytics?

　(e) What do you mean by semi-structured data?

　(f) How are views different from materialized view?

　(g) What is YARN?

GROUP-B

Answer any *four* questions from the following　　　　　　　8×4 = 32

2.　What is HDFS? What is Name Node and Data Node in HDFS? How　2+4+2 = 8
　NameNode tackle DataNode in HDFS?

3. (a) Define BIG Data.　　　　　　　　　　　　　　　　　　　　2+6 = 8

　(b) Explain the Evolution of Big Data and their characteristics.

4. (a) Illustrate the Hadoop Core Components with neat diagram.　　　4+4 = 8

　(b) Discuss the Hadoop system and Ecosystem components in four layers.

5. (a) Discuss the NOSQL data stores and their characteristic features.　4+4 = 8

　(b) Illustrate the Key Value pairs in data architectural patterns with an example.

6. (a) Discuss the functions of MongoDB Query Language and database Commands.

   (b) Describe the MapReduce execution steps with neat diagram.

7.   What would be the overall goals of big data in E-Commerce? Explain. How Big   3+5 = 8
       Data helps to identify risks and reduce fraud in E-Commerce?

   N.B. : *Students have to complete submission of their Answer Scripts through E-mail /
              Whatsapp to their own respective colleges on the same day / date of examination within
              1 hour after end of exam. University / College authorities will not be held responsible
              for wrong submission (at in proper address). Students are strongly advised not to submit
              multiple copies of the same answer script.*

——×——