[5919]-34

S.Y. M.Sc. (Computer Science)
CSDT - 234A : BIG DATA ANALYTICS
(2019 Pattern) (Semester - III)

Time : 2 Hours]   *[Max. Marks : 35*

Instructions to the candidates:

1)   Question 1 is compulsory.

2)   Solve any three questions from Q.2 to Q.5.

3)   Question 2 to 5 carry equal marks.

Q1)   Attempt any Five of the following :   **[5]**

   a)   Define Bigdata.

   b)   What is sharding?

   c)   Enlist various data integration patterns.

   d)   What is Hive?

   e)   What is IMDG?

   f)   Define data munging.

Q2)   Attempt the following :   **[10]**

   a)   i)   Explain big data analytics.   **[2]**

        ii) Explain 3V's of Bigdata.   **[4]**

   b)   Explain Hadoop framework in detail.   **[4]**

Q3)   Attempt the following :   **[10]**

   a)   i)   While using main memory, which weak points overcome by
             IMDG architecture?   **[2]**

        ii) Explain types of replication strategies.   **[4]**

   b)   Explain various features of HDFS.   **[4]**

Q4) Attempt the following : **[10]**

    a)   i)   Enlist the stages of SEMMA methodology.   **[2]**

        ii)   What are the benefits of Hadoop.   **[4]**

    b) Explain how to extract value from Bigdata.   **[4]**

Q5) Attempt any two of the following :   **[10]**

    a) Explain Map Reduce pattern with suitable example.   **[5]**

    b) Explain about components of YARN architecture.   **[5]**

    c) Explain CRISP methodology.   **[5]**

□□□