Code: 17F00407

MCA IV Semester Regular Examinations May 2019
BIG DATA ANALYTICS
(For 2017 admitted batches only)

Time: 3 hours   Max. Marks: 60

Answer all the questions
*****

1   What is cross-validation resampling technique? Explain in detail.

OR

2   Explain in detail about Boot-strapping.

3   (a)   Outline the stream processing issues.
    (b)   Write a short note on Bloom filtering.

OR

4   There are several ways that the bitstream 1001011011101 could be partitioned into buckets. Find and explain all of them.

5   What are the three different failure modes in map-reduce? Illustrate with example.

OR

6   Outline the Hadoop distributed file system design.

7   (a)   Explain 2-level network topology of hadoop cluster architecture.
    (b)   Discuss how bench marks are handled in hadoop.

OR

8   How hadoop can be run on Amazon EC2?

9   Differentiate between SQL and Hive QL.

OR

10   Explain in detail the Hive architecture.

*****