

PE-850
(634) M.Sc. in Information Technology (Fourth Semester)
Examination June 2022
Compulsory/Optional
BIG DATA ANALYTICS
Paper-

Time: Three Hours

Maximum Marks: 080

Minimum Passing Marks:

Note: Answer from both section as directed. The figures in the right-hand margin indicate marks.

SECTION-A

1. Answer the following questions in short:-

[1 X 10 = 10]

- (i) Define block in HDFS.
- (ii) What is data mining?
- (iii) Write the definition of Hadoop.
- (iv) Who invented the term big data?
- (v) What do you mean by heartbeat in HDFS?
- (vi) What is "map" in Hadoop?
- (vii) What is shuffling in map reduce?
- (viii) Can we change the block size of HDFS files?
- (ix) What will be the problem in having lots of small files in HDFS?
- (x) Define Fstmage.

2. Answer the following questions:-

[2 X 5 = 10]

- (i) Define partitioner.
- (ii) What is checkpointing in Hadoop?
- (iii) Define data locality in HADOOP.
- (iv) Define Hive.
- (v) Explain any two challenges of big data.

SECTION-B

[12 X 5 = 60]

3. Answer any five of the following questions:-

- (i) What is Data Warehouse? Explain Data Warehouse architecture in detail.
- (ii) Explain the difference between Data Warehouse and OLAP.
- (iii) What are the challenges in handling big data? Explain scope of big data in the near future.
- (iv) What are the various applications of big data? Explain with example.
- (v) How MapReduce organizes work? What are the advantages of using MapReduce with Hadoop?
- (vi) What is YARN? Explain YARN framework. Compare classic MapReduce with YARN.
- (vii) What is NOSQL? Explain types of NOSQL databases.
- (viii) What is Mongo DB? Explain the data types and features of Mongo DB in detail.
- (ix) What are the features of Apache Pig? Explain in detail the applications of Apache Pig.
- (x) Explain the architecture of Hive in detail. Also explain the difference between RDBMS AND Hive.