U.S.N

P.E.S. College of Engineering, Mandya - 571 401
(An Autonomous Institution affiliated to VTU, Belagavi)
Eighth Semester, B.E. - Information Science and Engineering
Semester End Examination; May/ June - 2019
Big Data

Time: 3 hrs   *Max. Marks: 100*

Note: Answer **FIVE** full questions, selecting **ONE** full question from each unit.

UNIT - I

1 a.	What is big data? Discuss different types and sources of data.	5
b.	As an HR manager of a company providing big data solutions to clients, what characteristics would you look for while recruiting a potential candidate for the position of a data analyst?	5
c.	List the steps that SNA follows to detect fraud.	5
d.	Write a note on social customer relationship management.	5
2 a.	How does the use of RFID tags help in inventory control? Explain.	5
b.	Distinguish between different types of data analytics.	6
c.	Discuss the advantages of big data analytic with the help of manufacturing unit.	9

UNIT - II

3 a.	What is Hadoop? Why do we need Hadoop? Distinguish between traditional RDBMS and Hadoop.	8
b.	Compare relational database with NOSQL database.	12
4 a.	Why do organizations store large volumes of data? How does Namenode tackle Datanode failures?	7
b.	Write an algorithm for selection, projection and union operations in Map-reduce.	13

UNIT - III

5 a.	Discuss the different interfaces and classes used in org**.**apache**.**hadoop**.**io package.	15
b.	Differentiate between streaming information access and low latency information access.	5
6 a.	Discuss the different techniques used to optimize Map-reduce jobs.	6
b.	List and explain the main features of Map-reduce.	10
c.	Describe the working of the Map-reduce algorithm.	4

UNIT - IV

7 a.	How to organize output data with output formats? Explain.	8
b.	Write a code for creation of multiple input splits by using multifile split class.	7
c.	Explain the types of Map-reduce applications.	5
8 a.	How will you perform local application testing on Map-reduce applications with Eclipse?	5
b.	Write a note on defensive programming in Map-reduce.	5
c.	What is logging? Explain the role it performs in testing Hadoop applications.	10

UNIT - V

9 a. Discuss the advantages of YARN over Map-reduce. 5

  b. Explain the YARN architecture with neat diagram. 7

  c. What are the various statements used in flow of data processing in pig latin? 5

  d. Write a Hive command for creating tables. 3

10 a. What are joins? How many types of joins are there in pig latin? Explain. 7

  b. Explain pig architecture. 5

  c. Discuss the two modes used for running the pig scripts. 4

  d. What are the main reasons for devoloping pig latin? 4

✳ ✳ ✳ ✳