Total No. of Questions : 12]

SEAT No. :

P3501

[Total No. of Pages : 2

[6005]-663

F.Y. M.E. (Computer Engineering) (Master of Data Science)

BIG DATA ANALYTICS

(2017 Pattern) (Semester -I) (510303)

Time :3 Hours] *[Max. Marks : 50*

Instructions to the candidates:

1) ***Answer question Q1or Q2, Q3 or Q4, Q5or Q6, Q7 orQ8, Q9 or Q10, Q11 or Q12***
2) ***Neat diagrams must be drawn wherever necessary.***
3) ***Figures to the right indicate full marks.***
4) ***Assume suitable data if necessary.***
5) ***Use of scientific calculator is permitted.***

Q1) What is Data analysis process? what are different types of Data Analytics? Explain the Importance of Data Analytics in Research? **[9]**

OR

Q2) What are different Tools used for Big data Analytics? What are the Business Drivers for Big Data Analytics? **[9]**

Q3) What are different Technologies used for Handling Big data? Differentiate between Distributed and Parallel Computing. **[8]**

OR

Q4) Explain in brief the hadoop components HDFC and YARN. **[8]**

Q5) What is Hadoop ecosystem ? Explain the role of each component in Hadoop ecosystem. **[8]**

OR

Q6) Explain Features of Apache Pig for Big Data Analysis? Differentiate between Pig and Map Reduce. **[8]**

Q7) What is Apache Spark?Explain any 2 Apache Spark Machine Learning Algorithms detail. **[8]**

OR

Q8) What is Apache Spark RDD?What are the limitations of RDD in Apache Spark and how to overcome those limitations? **[8]**

Q9) What is PySpark? What are industrial benefits of PySpark? **[9]**

OR

P.T.O.

Q10) What is Scala?What are important features of Scala?Explain in brief benefits of using Scala. **[9]**

Q11) What are important features of Data Visualization? Explain the Tools used in data visualization with appropriate examples. **[8]**

OR

Q12) Explain the benefits of good data visualization? What are the different Data Visualization Techniques used in data science? **[8]**

♦ ♦ ♦