

Total No. of Questions : 12 ]

SEAT No. :

**P9522**

**[6182]-849A**

[Total No. of Pages :2

**M.E. (Computer Engineering) (Data Science)**  
**BIG DATA ANALYTICS**  
**(2017 Pattern) (Semester-I) (510303)**

*Time : 3 Hours]*

*[Max. Marks : 50*

*Instructions to the candidates:*

- 1) Attempt Q.1 or Q.2, Q.3 or Q.4, Q.5 or Q.6, Q.7 or Q.8, Q.9 or Q.10, Q.11 or Q.12.*
- 2) Neat diagram must be drawn whenever necessary.*
- 3) Figures to the right indicates full marks.*
- 4) Assume suitable data if necessary.*
- 5) Use of Scientific Calculator is permitted.*

**Q1)** Explain Data Analytics Cycle with suitable diagram. **[9]**

OR

**Q2)** Define Big data? What are the characteristics of Big Data? Explain any 2 Real-World Problems that Big Data Can Solve. **[9]**

**Q3)** What is Apache HBase? List out the name of the key components of HBase. **[8]**

OR

**Q4)** Explain the important features and Core Components of Hadoop. **[8]**

**Q5)** What is a Zookeeper? Describe the benefits of using a zookeeper? **[8]**

OR

**Q6)** Explain the Apache Flume in Hadoop? Explain the features of Apache Sqoop. **[8]**

**Q7)** What are the important components of the Spark ecosystem? What are the different cluster managers available in Apache Spark? **[8]**

OR

**P.T.O.**

**Q8)** Explain the functions of Spark SQL? How can you connect Hive to Spark SQL? [8]

**Q9)** What are the main characteristics of PySpark? What do you think PySpark is important in Data Science? [9]

OR

**Q10)** What is Text Mining in Data analytics? What is stemming and Lemmatization? [9]

**Q11)** Define Big Data Visualization? Explain the Benefits of good data visualization? [8]

OR

**Q12)** What are the different Data Visualization Techniques used in data science? Explain in brief uses of Tableau for data visualization? [8]

