

Name :

B

Reg No :

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

07 THRISSUR CLUSTER

THIRD SEMESTER M.TECH. DEGREE EXAMINATION DEC 2017

COMPUTER SCIENCE DEPARTMENT

COMPUTER SCIENCE AND ENGINEERING

07CS7115 BIGDATA ANALYTICS

Time: 3 hours

Max. marks: 60

Answer all six questions. Part 'a' of each question is compulsory.

Answer either part 'b' or part 'c' of each question

Q.no.	Module 1	Marks
1a	State three attributes that define the characteristics of Big Data?	4
	Answer b or c	
b	Describe the four main types of data structures to define the data growth from non structured data type.	5
c	Discuss the challenges of the current analytical architecture for data scientists.	5
Q.no.	Module 2	Marks
2a	How is ETLT used to perform extract, transform and load processes to extract data?	4
	Answer b or c	
b	In which phase would the team expect to invest most of the project time? Why? Where would the team expect to spend the least time?	5
c	Elaborate the key roles for a successful analytics project.	5
Q.no.	Module 3	Marks
3a	Discuss R Graphical user interfaces and explain how R software uses CLI?	4
	Answer b or c	
b	Give the comparative analysis of data exploration and presentation in R.	5
c	Identify suitable hypothesis test for evaluating and comparing the means of 2 samples.	5
Q.no.	Module 4	Marks
4a	List and explain the steps involved in the Box-Jenkins methodology for time series analysis	4

Answer b or c

- | | | |
|----------|---|----------|
| b | Enumerate the difference between AR and MA model. | 5 |
| c | Construct a time series with a trend and time series for differencing using neat diagram. | 5 |

Q.no.	Module 5	Marks
5a	State the goals of POS tagging in text analysis.	5

Answer b or c

- | | | |
|----------|---|----------|
| b | Describe sentiment analysis used in text analysis. How and why is it important? | 7 |
| c | Elaborate why can't we use TF alone to measure the usefulness of the words? | 7 |

Q.no.	Module 6	Marks
6a	Which structured query language would be applied to unstructured data and what are its major categories?	5

Answer b or c

- | | | |
|----------|--|----------|
| b | Relate Map reduce and Hadoop and explain how Hadoop stores data in a distributed system. | 7 |
| c | Use the tools Pig and Hive in Big Data ecosystem. How do you choose among them for batch applications? | 7 |