

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 BIG DATA 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.	What are the different data structures supported by Big Data? Give examples?	4
	Answer b or c	
b.	Evolution of big data brings with it new security challenges. Give your view.	5
c.	Is EDW a good choice for data repository from an analyst perspective? Give reasons to support your answer.	5
Q.no.	Module 2	Marks
2a	'Data conditioning is a data pre-processing step'. Comment.	4
	Answer b or c	
b	The analytics team needs to envision the risks involved in doing the analysis project. Suggest the phase of data analytics life cycle which is suitable for this operation.	5
c	Data analytic team found out that neural network is a good technique for meeting the business requirement. Give a detailed analysis of the data analytic phase they are undergoing, highlighting the useful tools.	5
Q.no.	Module 3	Marks
3a	What are the NOIR Attributes in R?	4
	Answer b or c	
b	Consider an analysis context to test how weight of a person can be related to health problems. There are 4 categories of people, namely underweight, normal weight, overweight and obese. Which statistical method is suitable for the situation? Explain with support of R programming language.	5

- | | | |
|--|---|---|
| | c Give a comparative study on descriptive statistical analysis and exploratory data analysis, with example. | 5 |
|--|---|---|

Q.no.	Module 4	Marks
--------------	-----------------	--------------

- | | | |
|--|---|---|
| | 4a What are the components of a time series? Explain. | 4 |
|--|---|---|

Answer b or c

- | | | |
|--|---|---|
| | b Is it necessary to use ACF in certain time series for further processing? Give your view. | 5 |
| | c For an MA(q) model, autocorrelations have a value of zero beyond lag 'q'. Give reasons. | 5 |

Q.no.	Module 5	Marks
--------------	-----------------	--------------

- | | | |
|--|---|---|
| | 5a What are the different approaches to represent text during analysis? | 5 |
|--|---|---|

Answer b or c

- | | | |
|--|---|---|
| | b Consider a scenario where a company stopped manufacturing a product based on the reviews it received from social networking sites. What kind of analysis would they have applied to make such a decision? | 7 |
| | c Large collection of unstructured text data in Tera byte size need to be analysed in short time. Suggest a suitable method for performing it in a time bound fashion. | 7 |

Q.no.	Module 6	Marks
--------------	-----------------	--------------

- | | | |
|--|---|---|
| | 6a What is a NoSQL data store? Explain its different types. | 5 |
|--|---|---|

Answer b or c

- | | | |
|--|--|---|
| | b Suppose a website wants to store its real-time data feeds. Which Hadoop technology can be used for this scenario? Give reasons. | 7 |
| | c Suppose that finding Word count from a terra byte size of data is an urgent requirement. Suggest a technology that can handle this requirement | 7 |