

**APJ ABDULKALAM TECHNOLOGICAL UNIVERSITY
08 PALAKKAD CLUSTER**

Q. P. Code : 2DB182

(Pages: 2)

Supple

Name

Reg. No:

SECOND SEMESTER M.TECH. DEGREE EXAMINATION, APRIL 2018

Branch: CSE

Specialization : Computer Science and Engineering

08 CS 6042 (B) Big Data Essentials

Time: 3 hours

Max. marks: 60

Answer all six questions.

Modules 1 to 6: Part 'a' of each question is compulsory and answer either part 'b' or part 'c' of each question.

Q.no.	Module 1	Marks
1.a	What is the role of a data scientist?	3
	Answer b or c	
b	Explain the different phases in data analytics life cycle	6
c	Describe the different characteristics of big data. Explain how data analytics can be used in real world applications.	6
Q.no.	Module 2	Marks
2.a	What are the functions of data node and name node?	3
	Answer b or c	
b	Explain the hadoop architecture in detail.	6
c	Describe the different steps involved in setting up a hadoop cluster	6
Q.no.	Module 3	Marks
3.a	Explain the concept of partitioner and combiner.	3
	Answer b or c	
b	What are the steps involved in developing a map reduce application	6
c	Describe the dataflow in map reduce	6

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

4.a	Write short notes on MongoDB?	3
-----	-------------------------------	---

Answer b or c

b	Graph databases embrace relationships. Justify your answer.	6
c	Discuss about the data storage in big data environment	6

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

5.a	State the limitations of map reduce framework	4
-----	---	---

Answer b or c

b	How is big data mining different from normal data mining?	8
c	What are different clustering techniques that can be applied on big data?	8

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

6.a	Write short notes on Apache Giraph	4
-----	------------------------------------	---

Answer b or c

b	Discuss about bulk synchronous parallel model.	8
c	Explain the different phases in real time data analytics.	8