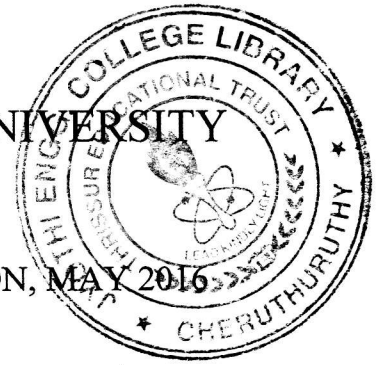# APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

08 PALAKKAD CLUSTER

SECOND SEMESTER M.TECH DEGREE EXAMINATION, MAY 2016

Computer Science and Engineering

08 CS 6042 (B) Big Data Essentials

Max. Marks : 60   Duration : 3 Hours

Answer ALL six questions. Part (a) of each question is compulsory.

Answer EITHER part (b) *or* part (c) of each question.

Q.No.   Marks

Module I

1a. What is the role of a data scientist?   3

Answer b *or* c

b. Describe the different characteritics of big data. Explain how data anayltics can be used in real

world applications.   6

c. Describe the data analytics life cycle.   6

Module II

2a. What are the functions of datanode and namenode.   3

Answer b *or* c

b. Explain the hadoop architecture in detail.   6

c. Explain the map reduce frame work   6

Module III

3a. Explain the concept of partitioner and combiner.   3

Answer b *or* c

b. Describe the different design patterns used in map reduce algorithm design.   6

c. What are the steps involved in developing a map reduce application   6

Module IV

4a. Write short notes on MongoDB?   3

Answer b *or* c

b. Discuss about the data storage in big data environment. 6

c. Explain the concept of graph databases 6

Module V

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

Answer b *or* c

b. Explain the issues and challenges involved in big data mining. 8

c. What are different clustering techniques that can be applied on big data? 8

Module VI

6a. Write short notes on Apache Giraph 4

Answer b *or* c

b. Explain the different phases in real time data analytics. 8

c. Explain bulk synchronous parallel model. 8