



**APJ ABDULKALAM TECHNOLOGICAL UNIVERSITY
08 PALAKKAD CLUSTER**

Q. P. Code : 2DB-17-1

(Pages: 2)

Name

Reg. No:

SECOND SEMESTER M.TECH. DEGREE EXAMINATION April/May 2017

Branch: Computer Science & Engg

Specialization: Computer Science & Engg

08CS6042(B) BIG DATA

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	Differentiate structured, semi-structured and unstructured data with examples.	3
	Answer b or c	
b	List the steps involved in data analytics life cycle. In which phase the team expect to invest most of the project time? Why? Where would the team expect to spend the least time?	6
c	The role of data scientist – 'a powerful and rare combination'. Why?	6
Q.no.	Module 2	Marks
2.a	Compare traditional RDBMS with Hadoop.	3
	Answer b or c	
b	List the steps for setting up a Hadoop cluster. Include commands and its explanation for each step.	6
c	Describe HDFS architecture in detail with proper diagrams.	6
Q.no.	Module 3	Marks
3.a	Briefly explain how MapReduce scale out to incorporate big data	3
	Answer b or c	
b	Differentiate the use of combiners and partitioners in a MapReduce program with proper examples.	6

- c Consider the following dataset regarding the electrical consumption of an organization. It contains the monthly electrical consumption and the annual average for various years. 6

Small Sample:

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1979	23	23	2	43	24	25	26	26	26	26	25	26
1980	26	27	28	28	28	30	31	31	31	30	30	30
1981	31	32	32	32	33	34	35	36	36	34	34	34
1984	39	38	39	39	39	41	42	43	40	39	38	38
1985	38	39	39	39	39	41	41	41	00	40	39	39

Write a MapReduce Application to find year wise average electrical consumption.

[Note: record output only if average > 30].

Q.no.	Module 4	Marks
4.a	Briefly discuss about CRUD operations in MongoDB .	3
	Answer b or c	
b	Write a short note on NoSQL databases and its types. Which NoSQL storage type is suitable for implementing in-memory cache? Justify your answer.	6
c	Write notes on HBase concepts. Compare row oriented and column oriented databases.	6

Q.no.	Module 5	Marks
5.a	Discuss HACE theorem of Big Data and how it.become a challenge in big data mining?	4
	Answer b or c	
b	Discuss about implementation of graph algorithms on MapReduce.	8
c	Describe methodology for implementing any one algorithm for clustering big data.	8

Q.no.	Module 6	Marks
6.a	Write about two real time big data analytics use cases.	4
	Answer b or c	
b	Discuss about various models used for big data analytics.	8
c	Write detailed notes about Apache Giraph and its features.	8