

Name:	 UPES UNIVERSITY WITH A PURPOSE
Enrolment No:	

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, May 2019

Course:	Big Data Analytics	Semester: II
Program:	M.Tech CS	Time 03 hrs.
Course Code:	CSDA7003	Max. Marks: 100

Instructions:

SECTION A

S. No.		Marks	CO
Q 1	Name the file systems used in Big data(any 2) and non bigdata systems (any 2).	4	CO1
Q 2	Explain Apache Zookeeper.	4	CO3
Q 3	Define Pig Grunt.	4	CO2
Q 4	Differences between HBase and HDFS.	4	CO4
Q 5	Explain a regular expression to capture mailed.	4	CO2

SECTION B

Q 6	Justify the advantages of using Apache Spark over Hadoop. Explain in brief two major libraries of Apache Spark.	10	CO4
Q 7	Elaborate Pig Latin and explain Pig Latin's data flow.	10	CO2
Q 8	Justify why we need big data frame works. Briefly discuss the applications of big data.	10	CO1
Q 9	Show commands with appropriate arguments to perform data transfer between local file system and HDFS. Explain Storage mechanisms in H Base or Explain data serialization. With proper examples, discuss and differentiate between structured, unstructured and semi-structured data.	10	CO4

SECTION-C

Q 10	Demonstrate vector product with map-reduce of a $n \times n$ matrix M with a vector v of length n .	20	CO2
Q 11	With suitable block diagram explain architecture of HDFS. Discuss role of Data node and Name node in HDFS. Or Define Map Reduce. Explain working of various phases of Map Reduce with appropriate examples and diagrams.	20	CO3