

<b>Name:</b>  <b>Enrolment No:</b>	
--	--

**UNIVERSITY OF PETROLEUM AND ENERGY STUDIES**  
**End Semester Examination, May 2022**

**Course: Big Data Analytics**  
**Program: M.Tech(CSE)**  
**Course Code: CSDA7003**

**Semester: II**  
**Time : 03 hrs.**  
**Max. Marks: 100**

**Instructions:**

Section A: Short answer type questions. Attempt all the questions, each Question carries 4 Marks

Section B: Medium answer type questions. Attempt all the questions, each Question carries 10 Marks

Section C: Long answer type questions. Attempt all the questions, each Question carries 20 Marks

**SECTION A**  
**(5Qx4M=20Marks)**

S. No.		Marks	CO
Q 1	Write a note on challenges with storage and processing of Big data.	4	CO1
Q 2	Write a shot note on Apache Spark.	4	CO4
Q 3	Define Data Wrangling.	4	CO1
Q 4	What is the role of SSH?	4	CO3
Q 5	What is the role of HBase in Big Data Processing?	4	CO4

**SECTION B**  
**(4Qx10M= 40 Marks)**

Q 6	The algorithms developed and maintained by the Apache Hadoop projects are implemented in the form of Hadoop MapReduce, write your views on it with reference to the use of MapReduce framework. What are the different features of MapReduce and draw a diagram to represent the working of MapReduce.	10	CO4
Q 7	Why we use YARN? What are the advantages of using YARN? Draw a diagram to represent the working of YARN architecture and explain.	10	CO3
Q 8	Why we use Big data? What are the different characteristics of Big Data? Explain the role of Big Data with examples in the following domains: a) Web b) Financial c) Healthcare d) Internet of Things e) Environment f) Logistics and transportation	10	CO1

	g) Industry h) Retail What are the different benefits of big data systems in Healthcare applications?		
Q 9	Define Hadoop ecosystem. Draw a neat diagram to demonstrate all the Hadoop Ecosystem elements at various stages of Data Processing.	10	CO4
<b>SECTION-C</b> <b>(2Qx20M=40 Marks)</b>			
Q 10	Illustrate the significance of HDFS in Hadoop ecosystem. What are the different features of HDFS? Draw a neat diagram of HDFS architecture and explain the concept of blocks in HDFS architecture. What are the roles of NameNode, DataNode, Secondary NameNode, Job tracker and task tracker? Explain in detail.	20	CO3
Q 11	Analytics is the process of contextualize and process the raw into useful information, explain how? What are the different goals of Analytics task? According to National Research Council, what are the characterization of computational tasks? Explain the following types of analytics with its statistical/ML models: <ul style="list-style-type: none"> <li>• Descriptive Analytics</li> <li>• Diagnostic Analytics</li> <li>• Predictive Analytics</li> <li>• Prescriptive Analytics</li> </ul>	20	CO1