

Model Question Paper

Seventh Semester BE Degree Examination

Big Data Analytics

Time: 3 Hours (180 Minutes)

Max. Marks: 100

Note: 1. Answer any FIVE full questions, choosing ONE full question from each module.

2. M: Marks, L: RBT (Revised Bloom's Taxonomy) level, C: Course outcomes.

Module -1			M	L	C
Q1	a.	Define Big Data. Explain its 3V's characteristics	06	L2	CO1
	b.	Explain the applications of Big Data Analytics in following domain 1. Web Domain. 2. Healthcare	08	L2	CO1
	c.	List & explain the various classification of Big Data Types	06	L2	CO1
OR					
Q2	a.	Discuss about the classification of data with neat example.	06	L2	CO1
	b.	State the functions of NameNode and DataNode in Hadoop DFS with neat sketch.	08	L2	CO1
	c.	Write elaborately about Block replication in Hadoop with neat sketch.	06	L2	CO1
Module- 2					
Q3	a.	With suitable business application, apply Map and Reduce model on Big Data in HDFS with suitable diagrams.	08	L3	CO3
	b.	Explain the working of combiner to the map task output in MapReduce	06	L2	CO2
	c.	Discuss fault tolerance and speculative execution in MapReduce	06	L2	CO2
OR					
Q4	a.	Illustrate the working of YARN Hadoop V2 Architecture with suitable example. List out its pros and cons over Hadoop V1.	08	L3	CO3
	b.	List & explain MapReduce Advantages	06	L2	CO2
	c.	With a neat diagram, explain Sqoop Import operation	06	L2	CO2
Module - 3					
Q5	a.	What is NoSQL data store? explain CAP Theorem in Big Data Solutions.	06	L2	CO2
	b.	Discuss the characteristics of Schema-less architecture and its advantages.	06	L2	CO2
	c.	Considering a student database, justify how the NoSQL databases provide increased flexibility compared to traditional relational databases. Illustrate it.	08	L3	CO4
OR					
Q6	a.	State the advantages of Key-Value Pair in NoSQL Data Store.	06	L2	CO2
	b.	Discuss the MongoDB architecture and its key features.	06	L2	CO2
	c.	With an example illustrate HBase operations: create, insert, scan, get.	08	L3	CO4
Module – 4					
Q7	a.	Discuss the contribution of HIVE to HDFS in details.	06	L2	CO2
	b.	Compare and contrast HIVE with traditional RDBMS.	06	L3	CO3
	c.	Write a HiveQL statement to Create a table named as student with four columns	08	L3	CO4

		as USN, Name, Dept, Phone and load the data by considering the following text file named as “ss.txt” stored in directory /home/user1 4MT15CS001, SS1, CSE, 8978987894 4MT15CS002, SS2, CSE, 8978987895 4MT15CS003, SS3, CSE, 8978987896 Write a query to alter the table name to “CSEStudent” and third column name to “Department”.			
OR					
Q8	a.	With a neat diagram, explain Apache Pig Architecture.	06	L2	CO2
	b.	Compare Apache Pig versus MapReduce in details.	06	L3	CO3
	c.	Considering the following student.csv file USN NAME DEPT Phone 4MT15CS001, SS1, CSE, 8978987894 4MT15CS002, SS2, CSE, 8978987895 4MT15CS003, SS3, CSE, 8978987896 a) Write & illustrate the working of the Pig Latin Script Statement to load a student.csv file into relation “A” b) Write & illustrate the working of the Pig Latin “foreach” statement to dump values of first two columns of student.csv file	08	L3	CO4
	Module - 5				
Q9	a.	Describe Apache Spark architecture and its key features.	07	L2	CO3
	b.	What is Spark Streaming? With a neat diagram, explain in details.	06	L2	CO3
	c.	Write a Spark program for word count using PySpark.	07	L3	CO4
OR					
Q10	a.	Explain Web Content Mining with an example.	07	L2	CO3
	b.	Discuss the concept of Text Mining with suitable examples	06	L2	CO3
	c.	Illustrate the working of the PageRank algorithm using a simple Web Graph give below <pre> graph TD a((a)) --> b((b)) a((a)) --> c((c)) b((b)) --> c((c)) c((c)) --> d((d)) a((a)) <--> d((d)) </pre>	07	L3	CO4

