U.S.N.										

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

July 2023 Semester End Main Examinations

Programme: B.E. **Semester: VI**
Branch: Computer Science and Engineering **Duration: 3 hrs.**
Course Code: 20CS6PEBDA **Max Marks: 100**
Course: Big Data Analytics **Date: 17.07.2023**

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

Important Note: Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.

			UNIT - I	CO	PO	Marks
1	a)		Define Big Data. List out the sources from which big data gets generated.	*CO1*	*PO1*	**5**
	b)		Discuss the different types of NoSQL databases with example for each.	*CO1*	*PO1*	**5**
	c)		List the various characteristics of data and what are the challenges with big data, explain.	*CO1*	*PO1*	**10**
			UNIT - II			
2	a)		Using CQL write queries for the following: i) Create a Keyspace Hospital and Create the Column Family Doctor (ID, Name, Reg_no, Salary, Department, Designation, Specializations, VisitingHospitals) assuming appropriate data type. ii) Insert required row to the Column Family. iii) Display Name and Department whose designation is "Senior Surgeon" and salary is greater than 1,00,000 in decreasing order. iv) Create a table to add patient_name and disease. Insert values which will be valid for 30 days. Import an existing csv file into the current column family.	*CO2*	*PO3*	**10**
	b)		Demonstrate various features of Cassandra in detail.	*CO2*	*PO1*	**10**
			UNIT - III			
3	a)		Write map and reduce functions to find the sum of even and odd numbers.	*CO3*	*PO3*	**10**
	b)		Write HDFS Commands for the following i. To create a directory by name BMS ii. To copy a file from HDFS to local file system iii.To show the Access Control Lists (ACLs) of files and directories	*CO3*	*PO3*	**5**

			iv. To retrieve all files that match to the source path entered by the user in HDFS, and creates a copy of them to one single, merged file in the local file system identified by local destination v. To copy a file from one directory to another directory within the HDFS			
		c)	Differentiate between Combiner and Reducer with an example	*CO2*	*PO1*	**5**
			OR			
	4	a)	Write a map and reduce functions to find out number of products sold in each country for the following data. **Transaction_ date** | **Product Name** | **Price** | **Payment Mode** | **Name** | **Country** 01-01-1900 00:00 | 1 | 1200 | Mastercard | Carolina | UK 02-01-1900 00:00 | 2 | 1200 | Visa | Betina | US 03-01-1900 00:00 | 3 | 1200 | Mastercard | Gourya | Australia	*CO3*	*PO3*	**7**
		b)	Describe Hadoop high-level architecture with a neat diagram.	*CO2*	*PO1*	**8**
		c)	Distributed data bases relax ACID properties. Analyze the properties supported by distributed systems with a scenario. Reflect databases that follow one of the three possible combinations with a neat diagram.	*CO2*	*PO1*	**5**
			UNIT - IV			
	5	a)	Discuss in detail about maps and tuples in Scala. Also Write Spark SQL queries for the following by making necessary imports and creating appropriate contexts. i. Create a SchemaRDD by loading a JSON file which contains details of mobile phones. ii. Select mobile name, mobile_price in the increasing order of mobile_price for 10 mobiles. iii. Access the first column from the SchemaRDD created above.	*CO3*	*PO3*	**10**
		b)	Demonstrate the three main capabilities provided by Spark SQL with a neat diagram.	*CO1*	*PO1*	**5**
		c)	Write the output of the following scala operations i. val s = "Hello", s(4) ii. 1.to(10) iii. val xmax, ymax = 100 iv. val s=if (n > 0) { r = r * n; n -= 1 }(Given r=1 n=4) v. def decorate(str: String, left: String = "[", right: String = "]") =left + str + right, decorate("Hello")	*CO3*	*PO3*	**5**
			OR			
	6	a)	Write a program in Apache spark to count the word occurrences in a given text file and display only those words appearing more than five times. Briefly discuss all the transformations and actions used in the program.	*CO3*	*PO3*	**10**

		b)	Write a Scala program to create a user defined function to return largest number among two numbers.	*CO3*	*PO3*	**5**
		c)	Demonstrate the process of creating SchemaRDD using the following methods i. registerTempTable(Use Scala or Python) ii. inferSchema() or case class	*CO1*	*PO1*	**5**
			UNIT - V			
	7	a)	Explain how collaborative filtering can be employed in developing a product recommender system in online shopping portal.	*CO3*	*PO3*	**10**
		b)	Compare content based filtering and collaborative filtering. Describe the steps involved in building a content-based recommender system.	*CO1*	*PO1*	**10**

****