U.S.N.										

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

Autonomous Institute Affiliated to VTU

October 2024 Supplementary Examinations

Programme: B.E. **Semester: VI**
Branch: Information Science and Engineering **Duration: 3 hrs.**
Course Code: 22IS6PEBDA **Max Marks: 100**
Course: Big Data Analytics

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

		UNIT - I	CO	PO	Marks
1	a)	Differentiate between Traditional Business Intelligence (Bi) Versus Big Data	CO1		**05**
	b)	Mention any five **top challenges** facing Big Data.	CO1		**05**
	c)	Consider the data: <body> Bda exam Public IP address of user is : </body> What are the challenges faced with the data given also provide possible solution to extract useful information from such data.	CO1		**10**
		UNIT - II			
2	a)	Explain the working principles of map reduce and combiner with the diagram.	CO1		**05**
	b)	Explain with a neat diagram the architecture of HDFS and its components.	CO1		**10**
	c)	Develop a mapper class in Java to search a word using Hadoop framework.	CO3	PO2	**05**

			UNIT - III			
	3	a)	Design a Table 'person' with fields, SL No, Name, Department and fees. **Fig-1** Perform the following: 1. Create the keyspace with the name Student 2. Insert into the table with the values as shown in fig.1. 3. Alter the table by adding a column Email_ID 4. Update the table by replacing the Sita in the place of Raj 5. Delete the first row from the table 6. Apply and demonstrate the concept of index using a query. 7. Use suitable collections concept for inserting a group of email_id for each student.	*CO4*	*PO3*	**08**
		b)	Explain the Hive architecture with a diagram.	*CO1*		**6**
		c)	Explain 4 data types of Hive with an example each.	*CO1*		**6**
			OR			
	4	a)	Explain the features of Cassandra.	*CO1*		**6**
		b)	Identify appropriate key terms of Casandra architecture i) Basic functional unit of Casandra ii) A collection of related Casandra nodes iii) A unit that contains all the multiple servers all stacked on top of another iv) A temporary memory location where we write data during updates and deletion.	*CO1*		**4**
		c)	Create a table employee (e_id,name, Designation, Experience, salary) and Faculty (F_id,Subject,Sem,E_id) and perform the below queries in Hive framework. 1. Query to find the names and their experience of all the employees who handle "OS" as a subject. 2. Query to find the names of faculty,experience,subject handling who has minimum salary. 3. Query to find the subjects handled by faculties whose experience is more than 5 years, Print Sem and designation along with it. 4. Drop a column "sem" from the Faculty table. 5. Drop both the tables.	*CO4*	*PO3*	**10**

Sl No	Name	Department	Fees
1	Rahul	ISE	5,00,500
2	Ram	CSE	6,00,000
3	Raj	Mech	4,00,000

			UNIT - IV			
	5	a)	Define Transformation in Spark. Explain filter and union transformation with an example.	*CO1*		**06**
		b)	Create a table CAR in mysql which has (car_id, Car_name, Model). Design suitable queries/ commands to do the following in the Sqoop framework. 1. Import the CAR table from MySql database server to HDFS. 2. Verify the imported data in HDFS 3. Import CAR table data into '/queryresult' directory. 4. The subset query to retrieve the CAR id and name, whose make is Toyota. 5. Imagine a manager adds another row with data " 1010, Hycross, toyota". Import this data in the CAR table. 6. Export the CAR data from the HDFS to the CAR table in MySql database server.	*CO4*	*PO3*	**10**
		c)	Explain the following transformations with an example each: ● coalesce() ● union(dataset) ● reduceByKey(func, [numTasks]) ● groupByKey()	*CO2*	*PO1*	**04**
			OR			
	6	a)	Illustrate the two types of Operations supported by Apache Spark and specify the importance of portioning of data set using parallelized collections.	*CO1*		**08**
		b)	Identify and explain the ways in which RDDs can be created in spark.	*CO1*		**06**
		c)	A Student registration table has 30 entries. Design a query to import the table from MySql into HDFS. Later, a new entry was made "1006, Ramu, ISE". Write the appropriate import commands in order to import the new data into HDFS. Export the Student registration data from the HDFS to the student table in MySql database server	*CO3*	*PO2*	**06**
			UNIT - V			
	7	a)	Explain how Zookeeper in Hadoop works with a neat diagram.	*CO1*		**07**
		b)	Explain the apache flume architecture with a neat diagram.	*CO1*		**08**
		c)	Explain the limitations of Apache Flume.	*CO1*		**05**

****