RAMAIAH Institute of Technology

Department of Computer Science & Engineering

QUESTION BANK FOR VII Semester (Term: July – Nov 2023)

Big Data Laboratory (CSL76)

S.No	List of Programs: Big Data Laboratory:CSL76
	PART-A (Hadoop)
1.	Write a MapReduce program using Java, to analyze the given **Weather Report Data** and to generate a report with cities having maximum and minimum temperature for a particular year.
2.	Write a MapReduce program using Java, to analyze the given **Earthquake Data** and generate statistics with region and magnitude/ region and depth/ region and latitude/ region and longitude.
3.	Write a MapReduce program using Java, to analyze the given natural numbers and generate statistics for the number as **Odd or Even** and print their sum.
4.	Write a MapReduce program using Java, to analyze the given **Insurance Data** and generate a statistics report with the construction building name and the count of building/ county name and its frequency.
5.	Write a MapReduce program using Java, to analyze the given employee record data and generate a statistics report with the total number of **Female and Male Employees** and their average salary.
6.	Write a MapReduce program using Java, to analyze the given **Sales Records** over a period of time and generate data about the country's total sales, and the total number of the products. / Country's total sales and the frequency of the payment mode.
	PART-B (Spark and Pig)
1.	Write a spark program using Python , to analyze the given **Weather Report Data** and to generate a report with cities having maximum and minimum temperature for a particular year.
2.	Write a spark program using Python , to analyze the given **Earthquake Data** and generate statistics with region and magnitude/ region and depth/ region and latitude/ region and longitude
3.	Write a spark program using Python , to analyze the given **Insurance Data** and generate a statistics report with the construction building name and the count of building/ county name and its frequency
4.	Write a spark program using Python , to analyze the given **Sales Records** over a period of time and generate data about the country's total sales, and the total number of the products. / country's total sales and the frequency of the payment mode
5.	Write Pig Latin scripts to FILTER and GROUP **Student_details.txt** data. • FILTER student details BY city == 'Bangalore'; • GROUP student by age;
6.	Write Pig Latin scripts to JOIN and SORT **Customer_details.tx**t and **Order_details.txt** data. • JOIN customers BY id, • ORDER BY age ASC/ DESC;