Name: _____ Roll Number: _____ Section: _____

National University of Computer and Emerging Sciences, Lahore Campus

	Course:	**Fundamentals of Big Data (Lab)**	**Course Code:**
			Semester: **Spring 2023**
	Program:	**BS (Data Science)**	**Total Marks:** **30**
	Duration:	**100 Minutes**	**Weightage:** **25**
	Paper Date:	**29-March-2023**	**Page(s):** **2**
	Section:	**BDS-4B**	**Section:** _____
	Exam:	**Midterm**	**Roll No:** _____

Instruction/Notes:

- You can use either jupyter notebook via anaconda prompt or Google Collab to solve the exam.
- Properly comment your code.
- You **CANNOT** use any built in function **NOT** taught in class.
- If there is any ambiguity, take reasonable assumption. Questions during exam are not allowed.
- All other rules pertaining to examinations as per NUCES policy apply.
- Make a separate notebook to answer your questions and use the following syntax to name that notebook (21L-XXXX.ipynb)

Question 1 [Marks = 10]: Basic Python (Strings, Arrays, Loops)

1. Write a program that simulates a game of Rock-Paper-Scissors between two players. The program should prompt the users for their moves and then determine the winner.

Question - 02 [Marks = 10]: Map-Reduce

Example Input file

- L20-4305 Course: BigData Sem: Spring2020 Login:12-03-20-12:45 Logout:12-03-20-2:45 Accessed: stream, assignment
- L20-1111 Course:DataMining Sem:Spring2020 Login:12-03-20-11:00 Logout:12-03-20-12:00 Accessed: quiz, material, assignment
- L20-4305 Course:DataMining Sem:Spring2020 Login:12-03-20-12:00 Logout:12-03-20-2:00 Accessed: quiz, material, assignment
- L20-1111 Course:DataMining Sem:Spring2020 Login:12-03-20-2:00 Logout:12-03-20-3:00 Accessed: quiz, material, assignment
- L20-4305 Course:BigData Sem:Spring2020 Login:12-03-20-12:00 Logout:12-03-20-1:00 Accessed: quiz, material, assignment

Write a Map Reduce algorithm to find the number of times a student accessed his each class on google classroom during year 2020. You have to write python code for Mapper, Reducer .

Output for given input

- L20-4305 Course: BigData Sem: Spring2020 **2**
- L20-4305 Course: DataMining Sem:Spring2020 **1**
- L20-1111 Course:DataMining Sem:Spring2020 **1**

Question 3 [Marks =10]: For each course output the number of distinct students who have accessed the course classroom

Output:

- BigData **1**

Name: _____ Roll Number: _____ Section: _____

- DataMining **2**

Question - 04 [Marks=10]: Implement DBScan Clustering Algorithm on the given dataset. Online retail.

Online Retail Dataset is attached in the file.

- You need to load the dataset and preprocess it for missing values and outliers..
- You need to scale the data for better clusters.
- Truncate half of data set rows
- **BONUS[1 abs]:** Evaluate the model using classification matrix/ accuracy score.

Note:

- **To score full marks, properly execute your files and display the outputs for each cell separately. Moreover, submit .ipynb file.**
- **You have the choice to attempt any of the three questions.**

Good Luck!