



**MCA II Year I Semester I-Midterm Examinations**

**Course: Data Science**

**(Only for MCA)**

**Date: 19/10/2023**

**Session: FN**

**Time: 90 Minutes**

**Max Marks: 30**

**Section – A**

**Answer all questions**

**(12 x 1 = 12 Marks)**

1. What are the characteristics of big data?(CO1,L1)
2. Define Data science.(CO1,L1)
3. Compare Datamart and Datalakes.(CO1,L1)
4. Define Chinese walls. (CO1,L1)
5. Define Big Data. (CO1,L1)
6. Mention any two problems encountered when working with more data than can fit in memory. (CO2, L1)
7. Mention any two tips for general programming best practices when working with large data. (CO2, L1)
8. List few python tools that can help you deal with large data. (CO2, L1)
9. Define Hash table.(CO2, L1)
10. Compare Online algorithm Vs. Streaming algorithm.(CO2, L1)
11. Write a pandas program to create series with dictionary. (CO3, L1)
12. Write the command to check the version of pandas.(CO3, L1)

## Section—B

Answer any three questions

(3 x 6 =18 Marks)

1. Explain the classified components of big data technologies.(CO1,L2)
2. Describe the steps involved in the data retrieval. (CO1,L2)
3. What are the general techniques for handling large volumes of data? Explain how to apply training principle to perceptron with online learning. (CO2,L3)
4. Explain the following terms with example. (CO2,L2)
  - a. Tree Structures
  - b. Sparse Data
5. Write a Pandas program to create and display a Data Frame from a Table:1 which has the index labels.

	area	population
California	423967	38332521
Florida	170312	19552860
Illinois	149995	12882135
New York	141297	19651127
Texas	695662	26448193

Table:1