

Big Data Analytics

Assignment-01 (Spring 2024)

Submission to be done on portal

Due Date: Sunday, April 21, 2024, before 11:59 PM

- Use Assignment pages and scan your assignment. (Hand written Only)
- Upload .pdf file on portal
- [Note: Title page must have Student Name, Registration Number, Section, and Date of Submission.
- Don't use register pages
- Submission will be on portal only

Assignment 01

- Q #01: *What is Map Reduce? Explain working of various phases of Map Reduce with appropriate example and diagram* 5 Marks
- Q #02: *Explain Job Scheduling in Map Reduce. How it is done in case of (i) The Fair Scheduler (ii) The Capacity Scheduler* 5 Marks
- Q #03: *Given the input file containing multiple lines of integer values, could you demonstrate how to apply the MapReduce algorithm to count the occurrences of each number? Please detail each step of the process, including mapping, shuffling, sorting, and reducing, to clearly illustrate how MapReduce operates on this dataset.* 10 Marks
- Given Data Set:*
- | | | | | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 23 | 23 | 27 | 43 | 24 | 25 | 26 | 26 | 26 | 26 | 25 | 26 | 25 |
| 26 | 27 | 28 | 28 | 28 | 30 | 31 | 31 | 31 | 30 | 30 | 30 | 29 |
| 31 | 32 | 32 | 32 | 33 | 34 | 35 | 36 | 36 | 34 | 34 | 34 | 34 |
| 39 | 38 | 39 | 39 | 39 | 41 | 42 | 43 | 40 | 39 | 38 | 38 | 40 |
| 38 | 39 | 39 | 39 | 39 | 41 | 41 | 41 | 28 | 40 | 39 | 39 | 45 |
| 23 | 23 | 27 | 43 | 24 | 25 | 26 | 26 | 26 | 26 | 25 | 26 | 25 |
| 26 | 27 | 28 | 28 | 28 | 30 | 31 | 31 | 31 | 30 | 30 | 30 | 29 |