

Practical Assignment

Objective: Apply MapReduce concepts in a real-world data processing scenario using Python.

1. **Dataset Selection:** Choose a large dataset (e.g., a collection of text documents, social media data, etc.) that requires significant computational resources to process.
2. **Problem Statement:** Define a clear problem statement that can be addressed using the MapReduce model. For example, word count, sentiment analysis, etc.
3. **Implement MapReduce:**
 - Use the provided **parallel_map_reduce** function as a starting point.
 - Develop a custom map function specific to your problem statement.
 - Develop a corresponding reduce function.
4. **Performance Analysis:**
 - Run the MapReduce job on your dataset.
 - Measure the execution time and compare it with a non-parallelized approach.
 - Analyze the scalability of your solution by varying the **pool_size** parameter.
5. **Report Writing (ONLY ONE. ipynb FILE):**
 - Document your approach, code, results, and observations.
 - Discuss the scalability and efficiency of the MapReduce model based on your findings.
 - Reflect on the challenges faced and potential improvements.