

SWE 507 PARALLEL PROGRAMMING PROJECT STUDY 5 (2025)

Due date: 23 May 2025, class time.

In this project, you are expected to do Hadoop MapReduce parallel computing. The following tasks are expected to be done in terms of project:

- Install Hadoop as single node cluster.
- Show that HDFS is working properly by copying some files from host to HDFS and reverse direction.
- Obtain or design the word count program that uses Hadoop-MapReduce method.
- Build your application and obtain the jar file.
- Move that jar file and the input data to the configured directories.
- Show that your program runs under Hadoop and expected output is generated.
- You should use a text data as an input which has a size not less than 256 Mbyte. You can create a large text file by copying and pasting segments repeatedly.

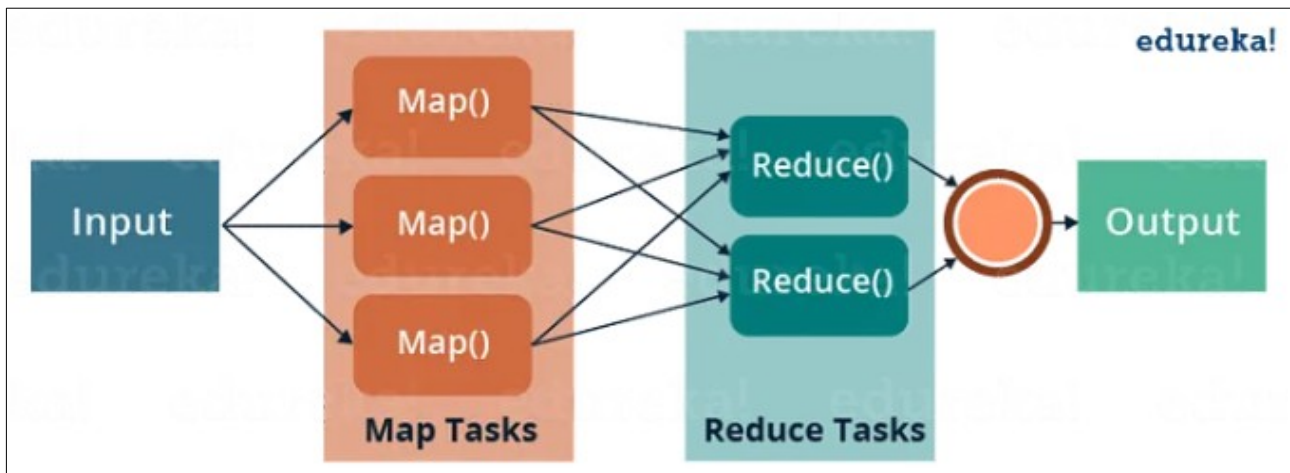


Figure 1. Hadoop Map-Reduce method (<https://medium.com/edureka/mapreduce-tutorial-3d9535ddbe7c>)

No	Task	Grade
1	Hadoop single node installation and HDFS working properly.	40
2	MapReduce distributed parallel computing is working as expected.	40
3	Presentation and discussions are done.	20

PS: In class presentation is required.