

Tasks:

- HDFS
 1. Create input directory /data/input/mapreduce/data1 AND /data/input/mapreduce/data1 in HDFS
 2. Copy data files data1.log and data2.log in data 1 and data2 directories respectively
- Map Reduce Exercise 1: Run first MR code with single input data file
 1. Refer to the ide_guide document and develop your first mapreduce program
 2. Prepare JAR file
 3. Copy JAR file in VmWare machine
 4. Run your JAR file using command: *hadoop jar <JAR file name>*
 5. Check the output in output directory

Note: Use data1.log file as input
- Map Reduce Exercise 2: Use multiple input data files
 1. Modify your driver class to include multiple input files
 2. Rebuild the JAR
 3. Copy new JAR to VM
 4. Run code and verify output
- Map Reduce Exercise 3: Access HDFS programmatically
 1. Refer to the code snippet in hdfs folder and modify your program in such a way that the same command runs multiple times without any additional actions
- Map Reduce Exercise 4: Optimization
 1. Plugin Combiner class and rerun the code
- Scheduling end-to-end job
 1. Refer to the sample shell script given in hdfs folder