In MacOS,

1. nano ~/.ssh/config
   Host 192.168.64.3
   HostkeyAlgorithms +ssh-rsa
   PubkeyAcceptedKeyTypes +ssh-rsa

2. scp /path/to/your/logfile.csv cloudera@<ip_address_of_cloudera>:/home/cloudera/
   put 'cloudera' as password when prompted

In Cloudera now,

3. Open Eclipse
   File → New → Java Project -> Name it (e.g., `LogFile`)

4. Right-click the project → Build Path → Configure Build Path → Libraries → Add External JARs.
   Add all Hadoop jar files under a folder
   All .jar files under
   - /usr/lib/hadoop
   - /usr/lib/hadoop-mapreduce
   - /usr/lib/hadoop/client

5. Right-click src → New → Package -> logfile
   Right-click src on package → New → Class -> LogMapper + LogReducer + LogTime(paste appropriate code in appropriate class)

6. Build/Export Your JAR File
   This JAR file is what you'll run with Hadoop.

   How to Export JAR from Eclipse:
   - Right-click your project (in the Project Explorer).
   - Choose Export…
   - Under the Java folder, select JAR file and click Next.
   - Select your project and source files (leave everything checked).
   - Export Destination: Enter /home/cloudera/Desktop/logFile.jar (or any path you prefer on the VM).
   - Click Finish.
   - This creates logFile.jar in your home directory (or where you specified).

7. Store logTime.csv in Desktop/data folder
   ls -l ~/home/cloudera/Desktop/data/logfile.csv

8. hdfs dfs -ls
   lists input, output folder of hadoop

9. hdfs dfs -mkdir /inputF1
   inputF1 is a hdfs folder

When you run your MapReduce job, this folder (/inputF1) is often given as the input path.
Hadoop will read all files inside /inputF1 as input for the job.

10. hdfs dfs -put /home/cloudera/Desktop/data/logTime.csv /inputF1

11. hdfs dfs - cat /inputF1/logTime.csv
    To see the .csv file

12. hadoop jar <your-jar-file>.jar <package.main-class / main-class> <input-path> <output-path>
    hadoop jar /home/cloudera/Desktop/logFile.jar logfile.LogTime /inputF1/logTime.csv /logFileOutput
    To run your MapReduce job using hadoop jar

13. View Results
    on terminal - hdfs dfs -cat /user/cloudera/output/part-r-00000
    in text file - hdfs dfs -get /user/cloudera/output/part-r-00000 ~/max_login_result.txt