Lab Quiz 2

Instructions:

- Open the data generation notebook (attached) and generate text by entering your roll number (use capital alphabets E.g., MSDS21001)
- Create an input file in Hadoop File System and paste the generated text into the newly created file
- Write Java program implementing MapReduce for the following tasks:
  - Run a simple wordcount example (code is attached)
  - Modify the code of Mapper and Reducer functions and count:
    - Total number of words in the doc
    - Total number of lines of text in doc
- Submit
  - All program files as text files
  - Output files
- Time allocated for the task is 30 minutes
- Additional 10 minutes are for downloading/uploading input/output and programs files to/from the local system and submit to the classroom
- Submit only one zip file containing all program and data files
- Use of internet is not allowed
- Follow the following naming patterns:
  - For zip file: RollNo_LQ2.zip (MSDS21001_LQ2.zip)
  - For program files: task_program.txt (wordcount_program.txt)
  - For output files: task_output.txt (wordcount_output.txt)
  - Tasks include wordcount, totalcount, and linescount
  - This quiz will be checked by an automatic script so make sure that there is no mistake, especially in naming conventions

Troubleshooting Tips:

- If you are not able to copy all the text at once you can copy it partially (there is a limit of 512 characters in some versions of the virtual box to copy)
- If you are not able to copy anything from VM to the local system at all, please enable the feature (Machine > Setting > General > Advance > Shared Clipboard > Bidirectional)
- If there is an error while loading eclipse software, please restart the virtual machine

Note:

- You can use a single program for all three tasks (using multiple reducer functions and using the output of one reducer as input to the next, although it is not required). In this case, you need to submit three output files.