Steps:

1. After having hadoop installed in our computer(ubuntu), we need to first open the terminal.
2. Then have permissions to access the user you have given to hadoop. We integrate hadoop to another user for security purpose.

Command: sudo su hduser

Here hduser is the user account we have given for hadoop.

1. Next change the directory for hadoop installation folder.

Command: cd /usr/local/hadoop

1. After going to hadoop directory type the command for starting hadoop distributed file system and environment.

Command: sbin/start-dfs.sh && sbin/start-yarn.sh

1. To check whether the environment had started correctly or not type the following command.

Command: jps

1. After hadoop starting correctly, create a input directory as “input” in hadoop directory.
2. Now move your input data to be processed to this directory by following command.

Command: sudo mv\* input-location target-location

Example: sudo mv\* /home/k/desktop /usr/local/hadoop/input

1. After moving data, type the following command to execute the program.

Command: hadoop jar filename input-directory output-directory

Example: hadoop jar final.jar /usr/local/hadoop/input/inp1 /usr/local/hadoop/outabc

1. The output folder should not be existed in the directory, if it already exists it throws an error.
2. Hadoop will create an output folder with the name given in the execution command “outabc” in the hadoop directory.
3. These outputs are created in dfs, So they are not visible in the file manager.
4. To access the output, Open the web browser and type the address “localhost:50070”.
5. This will open an html page. There on top you have utilities.
6. Upon clicking that you see “browse file system”. Click browse file system.
7. Here it will navigate you to hadoop root directory “/usr”.
8. Navigate to hadoop directory where you your output file.
9. Click the output file to download it and see your output.