# HadoopStreaming–WordcountUsingMapreducerinHadoop

**Steps:**

1. OpencommandpromptandrunasadministratorGoto hadoopsbin directory



Note:

* 1. Check hadoop/data/datanode and hadoop/data/namenode and if both foldersareempty, type “hdfsnamenode-format”.
  2. Checkpythonversionwith“python--version”.
  3. Check “C:\Python39\” is added in Environment variables > System variables >Path,if not addyourpython path.
  4. CheckEnvironmentvariables>Systemvariables >HADOOP\_HOMEissetas“C:\Hadoop”.



1. StartHadoopServicesstart-dfs.cmdstart-yarn.cmd



1. OpenthebrowserandgototheURLlocalhost:9870



1. CreateaDirectoryinHDFShdfsdfs-mkdir-p/user/hadoop/input



1. CopytheInputFiletoHDFShdfsdfs-putC:/Users/Admin/input.txt/user/hadoop/input



Note:mapper.py:



reducer.py:



1. RuntheHadoop StreamingJobhadoopjar

hadoopjarC:\hadoop\share\hadoop\tools\lib\hadoop-streaming-3.3.1.jar^

-files

/Users/monid/OneDrive/Documents/DataAnalytics/mapper.py,/Users/monid/OneDrive/Documents/DataAnalytics/reducer.py ^

-input/user/hadoop/input/data.txt ^

-output/user/output^

-mapper"pythonC:/Users/monid/OneDrive/Documents/DataAnalytics/mapper.py"^

-reducer"pythonC:/Users/monid/OneDrive/Documents/DataAnalytics/reducer.py"





1. ViewtheOutput

hadoopdfs-cat/user/output/part-00000

1. Oncethemapreduceoperationsareperformedsuccessfully,theoutputwillbepresentinthespecifieddirectory.

“/user/output/part-00000”



1. StopHadoopServicesstop-dfs.cmdstop-yarn.cmd



1. StopHadoopServicesstop-dfs.cmdstop-yarn.cmd



# RESULT:

ThustheimplementationofthepythonmapperandreducerprogramsusingMapReducetocountthewords in atextfileusing Hadoop is executedsuccessfully.