# Hadoop Streaming – Wordcount Using Map reducer in Hadoop

**Steps:**

1. Open command prompt and run as administrator Go to hadoop sbin directory



Note:

* 1. Check hadoop/data/datanode and hadoop/data/namenode and if both folders are empty, type “hdfs namenode -format”.
  2. Check python version with “python --version”.
  3. Check “C:\Python39\” is added in Environment variables > System variables > Path, if not add your python path.
  4. Check Environment variables > System variables > HADOOP\_HOME is set as “C:\Hadoop”.



1. Start Hadoop Services start-dfs.cmd start-yarn.cmd



1. Open the browser and go to the URL localhost:9870



1. Create a Directory in HDFS hdfs dfs -mkdir -p /user/hadoop/input



1. Copy the Input File to HDFS hdfs dfs -put C:/Users/Admin/input.txt /user/hadoop/input



Note: mapper.py:



reducer.py:



1. Run the Hadoop Streaming Job hadoop jar

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

-files

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

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

-output /user/output ^

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

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





1. View the Output

hadoop dfs -cat /user/output/part-00000

1. Once the map reduce operations are performed successfully, the output will be present in the specified directory.

“/user/output/part-00000”



1. Stop Hadoop Services stop-dfs.cmd stop-yarn.cmd



1. Stop Hadoop Services stop-dfs.cmd stop-yarn.cmd



# RESULT:

Thus the implementation of the python mapper and reducer programs using MapReduce to count the words in a text file using Hadoop is executed successfully.