

```
#!/usr/bin/python3

import os
import sys
import time

file=sys.argv[1]
k=sys.argv[2]

try:
    init=sys.argv[3]
except:
    init=[]

try:
    maxIter=int(sys.argv[4])
except:
    maxIter=100

try:
    metricOrder=sys.argv[5]
except:
    metricOrder=2
os.system("rm input*")
os.system("rm temp")
start=time.time()
os.system('python3 preprocess.py '+str(file)+' '+str(k)+' '+str(init)+' '+str(metricOrder))
os.system('hdfs dfs -put data.txt input')
iter=0
while(iter<maxIter):
    os.system('cp centroids.npy centroids_old.npy')
    os.system('hadoop jar $HADOOP_HOME/share/hadoop/tools/lib/hadoop-streaming-2.8.1.jar -
input input -output output -mapper mapper.py -reducer reducer.py')
    os.system('python3 postprocess.py')
    os.system('rm -r output')
    iter+=1
    if(os.system('cmp centroids.npy centroids_old.npy')==0):
        break
end=time.time()
print("\nIterations: "+str(iter))
print("Total runtime: ")
print("%.2f" % (end-start)+str(" seconds"))
os.system('python3 results.py '+str(file)+' '+str(k)+' '+str(metricOrder))
```