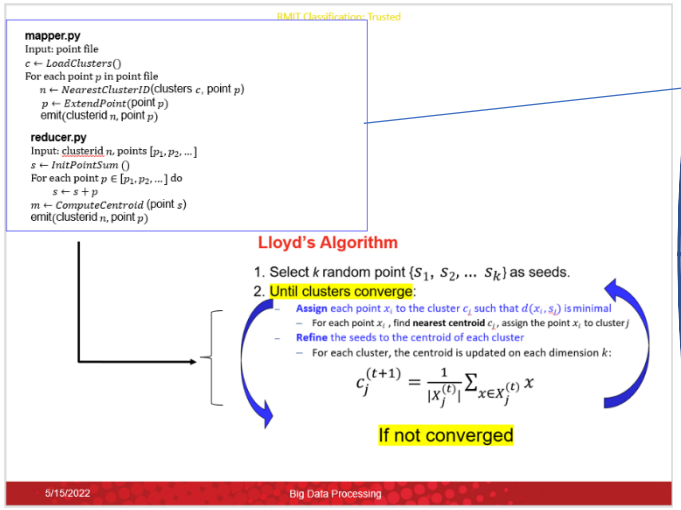


```
#!/bin/bash
i=1
while :
do
    hadoop jar ../../../hadoop-streaming-3.1.4.jar \
    -D mapred.reduce.tasks=1 \
    -D mapred.text.key.partitioner.options=-k1 \
    -file centroids.txt \
    -file ./mapper.py \
    -mapper ./mapper.py \
    -file ./reducer.py \
    -reducer ./reducer.py \
    -input /testMapReduce/dataset.txt \
    -output /testMapReduce/mapreduce-output$i \
    -partitioner org.apache.hadoop.mapred.lib.KeyFieldBasedPartitioner

    rm -f centroids1.txt
    hadoop fs -copyToLocal /testMapReduce/mapreduce-output$i/part-00000 centroids1.txt
    seeiftrue=`python reader.py`

    if [ $seeiftrue = 1 ]
    then
        rm centroids.txt
        hadoop fs -copyToLocal /testMapReduce/mapreduce-output$i/part-00000 centroids.txt
        break
    else
        rm centroids.txt
        hadoop fs -copyToLocal /testMapReduce/mapreduce-output$i/part-00000 centroids.txt
    fi
    i=$((i+1))
done
```

Is it okay if the reduce task number is not 1?

Current locations of k centroids are saved in this file.

What is the role of $i here?

Converged? - whether the new and old locations of each centroid are same or not

output final locations of k centroids

Update locations of the k centroids