cd /usr/local/mahout-0.7/bin

./mahout seqdirectory -c UTF-8 -i /reviews-extracted/ -o /reviews-sequential-svh110020

hadoop fs -ls /reviews-sequential-svh110020

hadoop fs -cat /reviews-sequential-svh110020/chunk-0

./mahout seqdumper -i /reviews-sequential-svh110020/chunk-0 | more

./mahout seq2sparse -i /reviews-sequential-svh110020/ -o /reviews-vectors-svh110020/ -ow -chunk 100 -x 80 -seq -ml 50 -n 2 -nv

./mahout seqdumper -i /reviews-vectors-svh110020/tfidf-vectors/part-r-00000

./mahout seqdumper -i /reviews-vectors-svh110020/tokenized-documents/part-m-00000

./mahout seqdumper -i /reviews-vectors-svh110020/dictionary.file-0 | more

./mahout kmeans -i /reviews-vectors-svh110020/tfidf-vectors/ -c /reviews-kmeans-centroids -cl -o /reviews-kmeans-clusters-svh110020 -k 60 -ow -x 10 -dm org.apache.mahout.common.distance.CosineDistanceMeasure

./mahout clusterdump -d /reviews-vectors-svh110020/dictionary.file-0 -dt sequencefile -i /reviews-kmeans-clusters-svh110020/clusters-\*-final/part-r-00000 -n 20 -b 100 -o /tmp/mahout-svh110020/Q1clusters.out -p /reviews-kmeans-clusters-svh110020/clusteredPoints/

more /tmp/mahout-svh110020/Q1clusters.out

./mahout seqdumper -i /reviews-kmeans-clusters-svh110020/clusteredPoints/part-m-00000 | more