**File naming convention of output files**

outputNYC\_[number of input files][number of nodes in cluster]\_[combiner status]

Ex - **outputNYC\_sixfiles8** => Six input files have been used in, and the Hadoop cluster consisted on 8 nodes. If no node number is provided in the output folder name, then the MapReduce was run on the default cluster of 2 nodes.

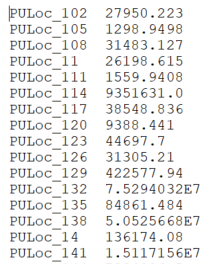
Ex - **outputNYC\_sixfiles\_combiner** => Six input files have been used in, and the Hadoop cluster consisted on 8 nodes. Standard combiner had been used.

**Sample Output File**

Each output file has three part:

* Part-r-0000
* Part-r-0001
* Part-r-0002

These output files parts can be open from WordPad. Ex – outputNYC\_sixfile – part-r-0000 opened as below:



The output shows the Reducer output pairs of Taxi Pickup Zone (i.e. PULoc\_xxx) and total revenue in $.

Output of all the parts can be concatenated and sorted (high to low) in Hadoop command prompt using below command:

hadoop fs -cat /user/s3400652/outputNYC\_sixfile/part-r-0000\* | sort -g -k2 -r | head -n10

Machine generated alternative text:
hadoope ip— 
-g 
—r 
192-168-24-120 
7.5294032E7 
5.052566BE7 
3.1618016E7 
2.8146592E7 
2.7911B9E7 
2 . •3402334E7 
2 . 656982BE7 
2.5625B96E7 
2 .335355BE7 
2. 308601E7 
had oop 
—cat /user/33400652/outputNYC I 
sort 
—k2 
head 
—n 10 
PULoc 
PULoc 
PULoc 
PULoc 
PULoc 
PULoc 
PULoc 
PULoc 
PULoc 
PCI Loc 
132 
13B 
161 
230 
237 
162 
186 
236 
170 