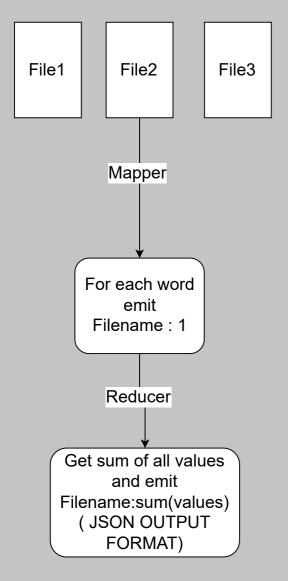
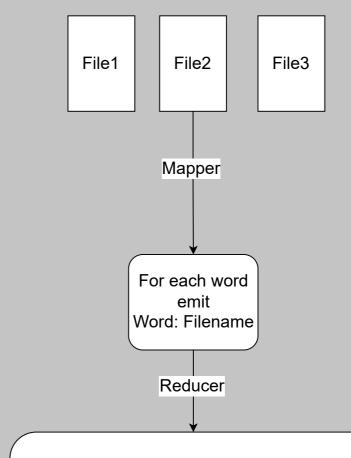


- 3- replace seed points with NewInput file
- 4- Apply Map reduce ( Convert txt to json ) on URLS.txt
- 5 goto step 1

## Calc file length with map reduce

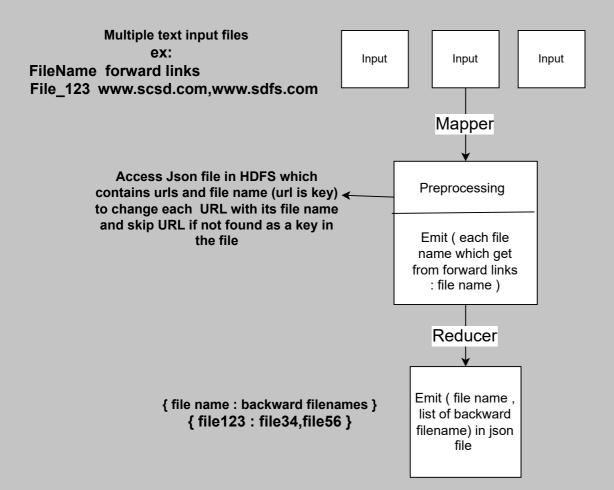


## Calc Inverted index and tf\_idf with map reduce



- 1 Calc frequency of the word in each file
- 2 access json file in HDFS have length of each file
- 3- calc number of unique files contain this word 4-Calc tf-idf for each file 5-emit (word,Filenames sorted by tf\_idf)

## **Backward link generate Map reduce**



## Map reduce to convert txt files to Json

