Problem

Create an inverted index for looking up the words in files

Workflow

- Mapper: fetch the file name of each record and split the record into words
  input: <offset, line>
  output: <word, fileName>

- Reducer: sum up the count for each word
  input: <word, (file1, file2, file1, ...)>
  output: <word, (file1=count1, file2=count2, ...)>

  Sample File(sample1.txt)

  I love Big Data

  Finally AI

  Sample File(sample2.txt)

  I love hello World

  Finally AI