

coggle

made for free at coggle.it DATA (FILE) READ

INPUT DATA (text file)

TEXTINPUTFORMAT()

record format: kvp=(#, line)

MAPPER()

tokenization()

caseNormalization()

stemming()

intermediate data: kvp=(token_string, 1)

PARTITIONER()

no change in intermediate data

SHUFFLE_AND_SORT()

sort key

REDUCER()

final data output



