CLOUD COMPUTING APPLICATIONS

MapReduce Example: Word Count

Roy Campbell & Reza Farivar

Word Count

- Have a large file of words, many words in each line
- Count the number of times each distinct word appears in the file

Word Count Program

MAP(KEY = LINE, **VALUE** = **CONTENTS**):
        FOR EACH WORD W IN VALUE:
            EMIT INTERMEDIATE(w,1)

REDUCE(KEY, VALUES):
 //key: a word; values: an iterator
//over counts
    RESULT = 0
    FOR EACH V IN **INTERMEDIATE VALUES**:
            RESULT += V
    EMIT(KEY, RESULT)

Word Count Illustrated

map(**key**=line, **values**=contents):
  for each **word** w in contents:
   emit(w,"1")

reduce(**key**=line, **values**=uniq_counts):
  sum all 1's in **values** list
   emit result (word, sum)

File
see bob run
see spot
throw

MAP
see   1
bob   1
run   1
see   1
spot   1
throw   1

REDUCE
bob   1
run   1
see   2
spot   1
throw   1