

UCS 548 – Foundation of Data Science

Assignment 11

Write the Map/Reduce Programs in R for the following.

1. To read a line of text and tokenize it.
2. To read a line of text and tokenize it, and find the occurrence of each token.
3. To read a text file ORDER.TXT containing the order placed by several customers and find the unique product ordered from the file.

Solution 1: (Sample Program)

mapper.R - Wordcount program in R to tokenize

```
trimWhiteSpace <- function(line) gsub("(^ +)|(+ $)", "", line)
splitIntoWords <- function(line) unlist(strsplit(line, "[[:space:]]+"))

line <- "This is a demo program for R. Map Reduce is demo in R."
line <- trimWhiteSpace(line)
words <- splitIntoWords(line)
## **** can be done as cat(paste(words, "\t1\n", sep=""), sep="")
for (w in words)
  cat(w, "\t1\n", sep="")
```

reducer.R - Wordcount program in R to count the occurrence

```
l=length(words)
env <- new.env(hash = TRUE)
count <- 1
for( w in words)
{
  word <- w
  if (exists(word, envir = env, inherits = FALSE)) {
    oldcount <- get(word, envir = env)
    assign(word, oldcount + count, envir = env)
  }
  else assign(word, count, envir = env)
  l=l+1
}
```

```
for (w in ls(env, all = TRUE))
  cat(w, "\t", get(w, envir = env), "\n", sep = "")
```