

Simple Test

Aaron Palumbo

9/13/2015

```
library(rmr2)
```

```
## Warning: S3 methods 'gorder.default', 'gorder.factor', 'gorder.data.frame',  
## 'gorder.matrix', 'gorder.raw' were declared in NAMESPACE but not found
```

```
## Please review your hadoop settings. See help(hadoop.settings)
```

```
library(rhdfs)
```

```
## Loading required package: rJava  
##  
## HADOOP_CMD=/home/apalumbo/workspace/cuny_msda_is622/hadoop-2.7.1/bin/hadoop  
##  
## Be sure to run hdfs.init()
```

```
hdfs.init()
```

```
#####  
# Simple Test #  
#####
```

```
rmr.options(backend="hadoop")
```

```
## NULL
```

```
mapper <- function(null, val){  
  df.m <- val[val$m == 'm', ]  
  df.n <- val[val$m == 'n', ]  
  list.m <- list()  
  for (j in 1:2){  
    df.m['i'] <- df.m$r  
    df.m['j'] <- j  
    list.m[[j]] <- df.m  
  }  
  df.m <- do.call(rbind, list.m)  
  list.n <- list()  
  for (i in 1:2){  
    df.n['i'] <- i  
    df.n['j'] <- df.n$c  
    list.n[[i]] <- df.n  
  }  
  df.n <- do.call(rbind, list.n)  
  
  df <- rbind(df.m, df.n)
```

```

    return(keyval(df[, c('i', 'j')], df$s))
}

reducer <- function(key, val.list){
  return(keyval(key, val.list))
}

#####
# small.df creates two small 2x2 matrices      #
# m and n.                                     #
# the goal of this example is to map the       #
# elements of m and n to the correct position #
# in p according to m * n = p.                 #
# i.e.:                                         #
#                                               #
# | a, b |   | e, f |   | ae + bg, af + bh | #
# |     | * |     | = |                     | #
# | c, d |   | g, h |   | ce + dg, cf + dh | #
#                                               #
# if we call the rows of p "i" and the cols  #
# of p "j" we should see:                     #
#                                               #
# i j   val                                     #
# 1 1    a                                     #
# 1 1    b                                     #
# 1 1    e                                     #
# 1 1    g                                     #
# 2 1    c                                     #
# 2 1    d                                     #
# 2 1    e                                     #
# 2 1    g                                     #
# 1 2    a                                     #
# 1 2    b                                     #
# 1 2    f                                     #
# 1 2    h                                     #
# 2 2    c                                     #
# 2 2    d                                     #
# 2 2    f                                     #
# 2 2    h                                     #
#                                               #
#####

small.df <- data.frame(m=c(rep('m', 4), rep('n', 4)),
                      r=c(1, 1, 2, 2, 1, 1, 2, 2),
                      c=c(1, 2, 1, 2, 1, 2, 1, 2),
                      s=letters[1:8],
                      v=1:8)

mat <- to.dfs(small.df)

if (exists('mr')){ rm(mr) }
mr <- mapreduce(input=mat, map=mapper)

```

```
result <- from.dfs(mr)

print(cbind(result$key, result$val))
```

```
##      i j result$val
## 1    1 1          a
## 1.1 1 1          b
## 1.2 1 1          e
## 1.3 1 1          g
## 1.4 2 1          c
## 1.5 2 1          d
## 1.6 2 1          e
## 1.7 2 1          g
## 2    1 2          a
## 2.1 1 2          b
## 2.2 1 2          f
## 2.3 1 2          h
## 2.4 2 2          c
## 2.5 2 2          d
## 2.6 2 2          f
## 2.7 2 2          h
```