NormalisedEfficiency.R

arcs

Tue Nov 28 15:25:16 2017

```
library(ggplot2)
library(scales)
library(data.table)
setwd("/home/arcs/Oct14/DataCSV")
getwd()
```

```
## [1] "/home/arcs/Oct14/DataCSV"
```

```
newdata <- fread("Oct2017Efficiency.csv")
```

```
##
Read 17.0% of 5876000 rows
Read 45.1% of 5876000 rows
Read 70.5% of 5876000 rows
Read 93.3% of 5876000 rows
Read 5876000 rows and 19 (of 19) columns from 0.404 GB file in 00:00:06
```

```
#########################################################################
############# Studying the structure of Data ############################
#########################################################################
names(newdata)
```

```
##  [1] "RequestCpus"              "MATCH_HEPSPEC"
##  [3] "MATCH_TotalCpus"          "RemoteWallClockTime"
##  [5] "ExitBySignal"             "ExitCode"
##  [7] "ExitSignal"               "ExitStatus"
##  [9] "RemoteSysCpu"             "RemoteUserCpu"
## [11] "CumulativeSuspensionTime" "RequestMemory"
## [13] "MemoryUsage"              "default_maxMemory"
## [15] "maxMemory"                "CumulativeRemoteSysCpu"
## [17] "CumulativeRemoteUserCpu"  "Remote_JobUniverse"
## [19] "JobUniverse"
```

```
str(newdata)
```

```
## Classes 'data.table' and 'data.frame':   5876000 obs. of  19 variables:
##  $ RequestCpus             : int  8 8 8 8 8 8 8 1 1 8 ...
##  $ MATCH_HEPSPEC           : chr  "None" "None" "None" "None" ...
##  $ MATCH_TotalCpus         : chr  "None" "None" "None" "None" ...
##  $ RemoteWallClockTime     : chr  "None" "None" "None" "None" ...
##  $ ExitBySignal            : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
##  $ ExitCode                : chr  "None" "None" "None" "None" ...
##  $ ExitSignal              : chr  "None" "None" "None" "None" ...
##  $ ExitStatus              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ RemoteSysCpu            : int  0 0 0 0 0 0 0 97 182 25311 ...
##  $ RemoteUserCpu           : int  0 0 0 0 0 0 0 49122 663 1323662 ...
##  $ CumulativeSuspensionTime: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ RequestMemory           : int  16000 16000 16000 16000 16000 16000 16000 2000 2000 16000 ...
##  $ MemoryUsage             : chr  "None" "None" "None" "None" ...
##  $ default_maxMemory       : int  2130 2130 2130 2130 2130 2130 2130 2130 2130 2130 ...
```

```
##  $ maxMemory               : chr  "16000" "16000" "16000" "16000" ...
##  $ CumulativeRemoteSysCpu  : chr  "0.0" "0.0" "0.0" "0.0" ...
##  $ CumulativeRemoteUserCpu : chr  "0.0" "0.0" "0.0" "0.0" ...
##  $ Remote_JobUniverse      : int  5 5 5 5 5 5 5 5 5 5 ...
##  $ JobUniverse             : int  5 5 5 5 5 5 5 5 5 5 ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

```
summary(newdata)
```

```
##   RequestCpus    MATCH_HEPSPEC      MATCH_TotalCpus    RemoteWallClockTime
##  Min.   :1.000   Length:5876000     Length:5876000     Length:5876000
##  1st Qu.:1.000   Class :character   Class :character   Class :character
##  Median :1.000   Mode  :character   Mode  :character   Mode  :character
##  Mean   :2.018
##  3rd Qu.:1.000
##  Max.   :8.000
##  ExitBySignal      ExitCode          ExitSignal          ExitStatus
##  Mode :logical   Length:5876000     Length:5876000     Min.   :0
##  FALSE:5755140   Class :character   Class :character   1st Qu.:0
##  TRUE :120860    Mode  :character   Mode  :character   Median :0
##                                                        Mean   :0
##                                                        3rd Qu.:0
##                                                        Max.   :0
##   RemoteSysCpu      RemoteUserCpu     CumulativeSuspensionTime
##  Min.   :     0.0   Min.   :      0   Min.   :0
##  1st Qu.:     0.0   1st Qu.:      2   1st Qu.:0
##  Median :     2.0   Median :      5   Median :0
##  Mean   :   294.6   Mean   :  15690   Mean   :0
##  3rd Qu.:   110.0   3rd Qu.:   9335   3rd Qu.:0
##  Max.   :298748.0   Max.   :1989119   Max.   :0
##  RequestMemory   MemoryUsage        default_maxMemory  maxMemory
##  Min.   :    0   Length:5876000     Min.   :2130      Length:5876000
##  1st Qu.: 1900   Class :character   1st Qu.:2130      Class :character
##  Median : 2130   Mode  :character   Median :2130      Mode  :character
##  Mean   : 4000                      Mean   :2130
##  3rd Qu.: 3700                      3rd Qu.:2130
##  Max.   :18000                      Max.   :2130
##  CumulativeRemoteSysCpu CumulativeRemoteUserCpu Remote_JobUniverse
##  Length:5876000         Length:5876000          Min.   :5
##  Class :character       Class :character        1st Qu.:5
##  Mode  :character       Mode  :character        Median :5
##                                                 Mean   :5
##                                                 3rd Qu.:5
##                                                 Max.   :5
##   JobUniverse
##  Min.   :5
##  1st Qu.:5
##  Median :5
##  Mean   :5
##  3rd Qu.:5
##  Max.   :5
```

```
######################################################################
############# Conversion to numeric values ###########################
######################################################################
```

```
newdata[,"RemoteWallClockTime"] <- as.numeric(unlist(newdata[,"RemoteWallClockTime"])) #RemoteWallClock
```

```
## Warning: NAs introduced by coercion
```

```
newdata[, "ExitCode"] <- as.numeric(unlist(newdata[, "ExitCode"]))
```

```
## Warning: NAs introduced by coercion
```

```
newdata[, "MemoryUsage"] <- as.numeric(unlist(newdata[, "MemoryUsage"]))
```

```
## Warning: NAs introduced by coercion
```

```
newdata[, "MATCH_HEPSPEC"] <- as.numeric(unlist(newdata[, "MATCH_HEPSPEC"]))
```

```
## Warning: NAs introduced by coercion
```

```
newdata[, "MATCH_TotalCpus"] <- as.numeric(unlist(newdata[, "MATCH_TotalCpus"]))
```

```
## Warning: NAs introduced by coercion
```

```
########################################################################
###################### Data Cleansing ##################################
########################################################################

newdata1 <- subset(newdata, newdata$ExitCode == 0) # Removing the failed Jobs
newdata2 <- subset(newdata1, select = c(RemoteSysCpu, RemoteUserCpu, RemoteWallClockTime, MATCH_HEPSPEC
########################################################################
############### Computation of efficiency ##############################
########################################################################

newdata2$CPUTime <- newdata2$RemoteSysCpu + newdata2$RemoteUserCpu
newdata2$WallTime <- newdata2$RemoteWallClockTime #- newdata2$CumulativeSuspensionTime
newdata2$HEPSPEC_TotalCpus <- newdata2$MATCH_HEPSPEC/ newdata2$MATCH_TotalCpus
newdata2$NWallTime <- newdata2$WallTime * newdata2$RequestCpus * newdata2$HEPSPEC_TotalCpus
newdata2$NCPUTime <- newdata2$CPUTime * newdata2$HEPSPEC_TotalCpus
newdata2$NEfficiency <- newdata2$NCPUTime/ newdata2$NWallTime



#Cleanseing data by removing NA rows

newdata2 <- na.omit(newdata2)
newdata2 <- subset(newdata2, newdata2$NWallTime != 0)
summary(newdata2)
```

```
##   RemoteSysCpu      RemoteUserCpu     RemoteWallClockTime MATCH_HEPSPEC
##  Min.   :     0.0   Min.   :      0   Min.   :     1      Min.   : 35.00
##  1st Qu.:     0.0   1st Qu.:      3   1st Qu.:    45      1st Qu.: 82.27
##  Median :     2.0   Median :      5   Median :   142      Median : 83.84
##  Mean   :   311.5   Mean   :  14987   Mean   :  8326      Mean   : 82.82
##  3rd Qu.:    12.0   3rd Qu.:    171   3rd Qu.:   771      3rd Qu.:104.00
##  Max.   :298748.0   Max.   :1983871   Max.   :812433      Max.   :310.00
##  MATCH_TotalCpus   RequestCpus       CPUTime           WallTime
##  Min.   : 4.000   Min.   :1.000   Min.   :      0   Min.   :     1
##  1st Qu.: 8.000   1st Qu.:1.000   1st Qu.:      4   1st Qu.:    45
##  Median : 8.000   Median :1.000   Median :      7   Median :   142
##  Mean   : 8.564   Mean   :1.337   Mean   :  15299   Mean   :  8326
##  3rd Qu.:10.000   3rd Qu.:1.000   3rd Qu.:    196   3rd Qu.:   771
```

```
##  Max.   :72.000   Max.   :8.000   Max.   :2000324   Max.   :812433
##  HEPSPEC_TotalCpus   NWallTime           NCPUTime
##  Min.   : 1.111    Min.   :       7   Min.   :       0
##  1st Qu.: 8.750    1st Qu.:     531   1st Qu.:      39
##  Median :10.400    Median :    1466   Median :      61
##  Mean   : 9.631    Mean   :  214608   Mean   :  146817
##  3rd Qu.:10.480    3rd Qu.:    7994   3rd Qu.:    1824
##  Max.   :20.000    Max.   :38862969   Max.   :20897762
##   NEfficiency
##  Min.   :   0.0000
##  1st Qu.:   0.0258
##  Median :   0.1163
##  Mean   :   0.3448
##  3rd Qu.:   0.6710
##  Max.   :2650.0000
```

```
TotalCPUTime <- sum(as.numeric(newdata2$NCPUTime))
TotalWallTime <- sum(newdata2$NWallTime)
TotalCPUTime
```

```
## [1] 430180240464
```

```
TotalWallTime
```

```
## [1] 628809466701
```

```
CumulativeEfficiency <- TotalCPUTime/TotalWallTime
CumulativeEfficiency
```

```
## [1] 0.6841186
```