Advanced HPC Cluster Usage with R: Introduction to batchtools

If you ever need to configure your personal LRZ account to work well with batchtools, you can have a look at the **Guide to the LRZ**.

Exercise 1

Consider the following code:

```
library(batchtools)
unlink("test_reg", recursive = TRUE)
reg = makeExperimentRegistry("test_reg")
addProblem(name = "p1", data = 1, seed = 1,
  fun = function(data, job) runif(data))
addAlgorithm(name = "a1",
  fun = function(job, data, instance) 2 * instance)
addAlgorithm(name = "a2",
  fun = function(job, data, instance) data + instance)
addExperiments(repls = 2)
submitJobs()
res = reduceResultsDataTable()
getJobPars()[res]
```

a. The function `runif` outputs random numbers between 0 and 1. Delete the argument `seed = 1` in function `addProblem` and execute the whole code several times. What happens with the results? Do they change?
b. With `repls = 2`, the experiment above is repeated twice. Run the code using the argument `seed = 1` in function `addProblem` and find out the value of the random number that was generated by `runif` in each of the two repetitions.
c. Include a third algorithm `a3` (see code below) and run the whole code above multiple times (using `seed = 1` in `addProblem`). What do you observe and why?

```
addAlgorithm(name = "a3",
  fun = function(job, data, instance) data + instance + runif(1))
```

Exercise 2

Read the introductory text of *Example 2: Machine Learning* from the online **batchtools vignette**:

a. Open a new `.R` file and copy-paste the executable code of each single R chunk from section *Example 2: Machine Learning* (until section *Submitting and Collecting Results*).
b. Run the code step by step and get it work either on your local machine or on **www.rstudio.lrz.de**. You can skip the last part *Submitting and Collecting Results* for now.
c. Load the data `data(PimaIndiansDiabetes, package = "mlbench")` and change the code from the previous exercise so that it works with the `PimaIndiansDiabetes` data.
Hint: For the `iris` data, the target column is `Species`, for the `PimaIndiansDiabetes` data, the target column is `diabetes`.