

# Package ‘BatchJobs’

March 18, 2015

**Title** Batch Computing with R

**Description** Provides Map, Reduce and Filter variants to generate jobs on batch computing systems like PBS/Torque, LSF, SLURM and Sun Grid Engine. Multicore and SSH systems are also supported. For further details see the project web page.

**Author** Bernd Bischl <bernd\_bischl@gmx.net>,  
Michel Lang <michellang@gmail.com>,  
Henrik Bengtsson <henrikb@braju.com>

**Maintainer** Bernd Bischl <bernd\_bischl@gmx.net>

**URL** <https://github.com/tudo-r/BatchJobs>

**BugReports** <https://github.com/tudo-r/BatchJobs/issues>

**MailingList** batchjobs@googlegroups.com

**License** BSD\_2\_clause + file LICENSE

**Depends** R (>= 2.15.0), BBmisc (>= 1.9), methods

**Imports** brew, checkmate (>= 1.5.1), DBI, digest, fail (>= 1.2),  
parallel, RSQLite (>= 1.0.0), sendmailR, stats, stringr, utils

**Suggests** MASS, testthat

**LazyData** yes

**ByteCompile** yes

**Version** 1.6

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2015-03-18 18:01:21

## R topics documented:

addRegistryPackages . . . . .	3
addRegistrySourceDirs . . . . .	4
addRegistrySourceFiles . . . . .	4
batchExpandGrid . . . . .	5

batchExport . . . . .	6
BatchJobs . . . . .	7
batchMap . . . . .	7
batchMapQuick . . . . .	8
batchMapResults . . . . .	9
batchReduce . . . . .	11
batchReduceResults . . . . .	12
batchUnexport . . . . .	13
callFunctionOnSSHWorkers . . . . .	14
cfBrewTemplate . . . . .	15
cfHandleUnknownSubmitError . . . . .	15
cfKillBatchJob . . . . .	16
cfReadBrewTemplate . . . . .	17
configuration . . . . .	17
debugMulticore . . . . .	18
debugSSH . . . . .	18
filterResults . . . . .	19
findDone . . . . .	20
findJobs . . . . .	22
getConfig . . . . .	22
getErrorMessage . . . . .	23
getJob . . . . .	24
getJobIds . . . . .	24
getJobInfo . . . . .	25
getJobNr . . . . .	26
getJobParamDf . . . . .	26
getJobResources . . . . .	27
getJobs . . . . .	27
getLogFiles . . . . .	28
getResources . . . . .	28
getSSHWorkersInfo . . . . .	29
grepLogs . . . . .	29
installPackagesOnSSHWorkers . . . . .	30
killJobs . . . . .	31
loadConfig . . . . .	32
loadExports . . . . .	33
loadRegistry . . . . .	34
loadResult . . . . .	34
loadResults . . . . .	35
makeClusterFunctions . . . . .	36
makeClusterFunctionsInteractive . . . . .	37
makeClusterFunctionsLocal . . . . .	38
makeClusterFunctionsLSF . . . . .	39
makeClusterFunctionsMulticore . . . . .	40
makeClusterFunctionsSGE . . . . .	41
makeClusterFunctionsSLURM . . . . .	42
makeClusterFunctionsSSH . . . . .	43
makeClusterFunctionsTorque . . . . .	44

makeJob . . . . . 45

makeRegistry . . . . . 45

makeSSHWorker . . . . . 47

makeSubmitJobResult . . . . . 48

reduceResults . . . . . 49

removeRegistry . . . . . 51

removeRegistryPackages . . . . . 52

removeRegistrySourceDirs . . . . . 52

removeRegistrySourceFiles . . . . . 53

resetJobs . . . . . 54

sanitizePath . . . . . 55

setConfig . . . . . 55

setJobFunction . . . . . 56

setJobNames . . . . . 57

showClusterStatus . . . . . 57

showLog . . . . . 58

showStatus . . . . . 58

sourceRegistryFiles . . . . . 59

submitJobs . . . . . 60

sweepRegistry . . . . . 61

testJob . . . . . 62

waitForJobs . . . . . 63

**Index** **65**

addRegistryPackages *Add packages to registry.*

**Description**

Mutator function for packages in [makeRegistry](#).

**Usage**

addRegistryPackages(reg, packages)

**Arguments**

reg	[Registry] Registry.
packages	[character] Packages to add to registry.

**Value**

[Registry](#) . Changed registry.

**See Also**

Other exports: [addRegistrySourceDirs](#); [addRegistrySourceFiles](#); [batchExport](#); [batchUnexport](#); [loadExports](#); [removeRegistryPackages](#); [removeRegistrySourceDirs](#); [removeRegistrySourceFiles](#)

---

addRegistrySourceDirs *Add source dirs to registry.*

---

**Description**

Mutator function for `src.dirs` in [makeRegistry](#).

**Usage**

```
addRegistrySourceDirs(reg, src.dirs, src.now = TRUE)
```

**Arguments**

<code>reg</code>	[Registry] Registry.
<code>src.dirs</code>	[character] Paths to add to registry. See <a href="#">makeRegistry</a> .
<code>src.now</code>	[logical(1)] Source files now on master? Default is TRUE.

**Value**

[Registry](#) . Changed registry.

**See Also**

Other exports: [addRegistryPackages](#); [addRegistrySourceFiles](#); [batchExport](#); [batchUnexport](#); [loadExports](#); [removeRegistryPackages](#); [removeRegistrySourceDirs](#); [removeRegistrySourceFiles](#)

---

addRegistrySourceFiles  
*Add source files to registry.*

---

**Description**

Mutator function for `src.files` in [makeRegistry](#).

**Usage**

```
addRegistrySourceFiles(reg, src.files, src.now = TRUE)
```

**Arguments**

reg	[Registry] Registry.
src.files	[character] Paths to add to registry. See <a href="#">makeRegistry</a> .
src.now	[logical(1)] Source files now on master? Default is TRUE.

**Value**

[Registry](#) . Changed registry.

**See Also**

Other exports: [addRegistryPackages](#); [addRegistrySourceDirs](#); [batchExport](#); [batchUnexport](#); [loadExports](#); [removeRegistryPackages](#); [removeRegistrySourceDirs](#); [removeRegistrySourceFiles](#)

---

batchExpandGrid	<i>Map function over all combinations.</i>
-----------------	--

---

**Description**

Maps an n-ary-function over a list of all combinations which are given by some vectors. Internally [expand.grid](#) is used to compute the combinations, then [batchMap](#) is called.

**Usage**

```
batchExpandGrid(reg, fun, ..., more.args = list())
```

**Arguments**

reg	[ <a href="#">Registry</a> ] Empty Registry that will store jobs for the mapping.
fun	[function] Function to map over the combinations.
...	[any] Vectors that are used to compute all combinations. If the arguments are named, these names are used to bind to arguments of fun.
more.args	[list] A list of other arguments passed to fun. Default is empty list.

**Value**

data.frame . Expanded grid of combinations produced by [expand.grid](#).

**Examples**

```

reg = makeRegistry(id = "BatchJobsExample", file.dir = tempfile(), seed = 123)
f = function(x, y, z) x * y + z
# lets store the param grid
grid = batchExpandGrid(reg, f, x = 1:2, y = 1:3, more.args = list(z = 10))
submitJobs(reg)
waitForJobs(reg)
y = reduceResultsVector(reg)
# later, we can always access the param grid like this
grid = getJobParamDf(reg)
cbind(grid, y = y)

```

---

batchExport

*Export R object to be available on the slaves.*


---

**Description**

Saves objects as RData files in the “exports” subdirectory of your file.dir to be later loaded on the slaves.

**Usage**

```
batchExport(reg, ..., li = list(), overwrite = FALSE)
```

**Arguments**

reg	[Registry] Registry.
...	[any] Objects to export. You must provide a valid name.
li	[list] More objects to export provided as a named list.
overwrite	[logical(1)] If set to FALSE (default), exported objects are protected from being overwritten by multiple calls of this function. Setting this to TRUE disables this check.

**Value**

character . Invisibly returns a character vector of exported objects.

**See Also**

Other exports: [addRegistryPackages](#); [addRegistrySourceDirs](#); [addRegistrySourceFiles](#); [batchUnexport](#); [loadExports](#); [removeRegistryPackages](#); [removeRegistrySourceDirs](#); [removeRegistrySourceFiles](#)

## Description

The BatchJobs package

## Additional information

**Homepage:** <https://github.com/tudo-r/BatchJobs>

**Wiki:** <https://github.com/tudo-r/BatchJobs/wiki>

**FAQ:** <https://github.com/tudo-r/BatchJobs/wiki/FAQ>

**Configuration:** <https://github.com/tudo-r/BatchJobs/wiki/Configuration>

The package currently support the following further R options, which you can set either in your R profile file or a script via `options`:

**BatchJobs.verbose** This boolean flag can be set to FALSE to reduce the console output of the package operations. Usually you want to see this output in interactive work, but when you use the package in e.g. knitr documents, it clutters the resulting document too much.

**BatchJobs.check.posix** If this boolean flag is enabled, the package checks your registry file dir (and related user-defined directories) quite strictly to be POSIX compliant. Usually this is a good idea, you do not want to have strange chars in your file paths, as this might results in problems when these paths get passed to the scheduler or other command-line tools that the package interoperates with. But on some OS this check might be too strict and cause problems. Setting the flag to FALSE allows to disable the check entirely. The default is FALSE on Windows systems and TRUE else.

## Description

You can then submit these jobs to the batch system.

## Usage

```
batchMap(reg, fun, ..., more.args = list(), use.names = FALSE)
```

**Arguments**

reg	[Registry] Empty Registry that will store jobs for the mapping.
fun	[function] Function to map over ....
...	[any] Arguments to vectorize over (list or vector).
more.args	[list] A list of other arguments passed to fun. Default is empty list.
use.names	[logical(1)] Store parameter names to enable named results in <code>loadResults</code> and some other functions. Default is FALSE.

**Value**

Vector of type integer with job ids.

**Examples**

```
reg = makeRegistry(id = "BatchJobsExample", file.dir = tempfile(), seed = 123)
f = function(x) x^2
batchMap(reg, f, 1:10)
print(reg)
```

---

batchMapQuick                      *Combination of makeRegistry, batchMap and submitJobs.*

---

**Description**

Combination of `makeRegistry`, `batchMap` and `submitJobs` for quick computations on the cluster. Should only be used by skilled users who know what they are doing. Creates the file.dir, maps function, potentially chunks jobs and submits them.

**Usage**

```
batchMapQuick(fun, ..., more.args = list(), file.dir = NULL,
  packages = character(0L), chunk.size, n.chunks,
  chunks.as.arrayjobs = FALSE, inds, resources = list())
```

**Arguments**

fun	[function] Function to map over ....
...	[any] Arguments to vectorize over (list or vector).



more.args	[list] A list of other arguments passed to fun. Default is empty list.
file.dir	[character] See <a href="#">makeRegistry</a> . Default is NULL, which means that it is created in the current directory under the name “bmq_[random alphanumerics]”.
packages	[character] See <a href="#">makeRegistry</a> .
chunk.size	[integer(1)] Preferred number of jobs in each chunk. Can not be used in combination with n.chunks. Note that the ids will get shuffled to balance out possible run time differences. Default is not to use chunking.
n.chunks	[integer(1)] Preferred number chunks. Can not be used in combination with chunk.size. Note that the ids will get shuffled to balance out possible run time differences. Default is not to use chunking.
chunks.as.arrayjobs	[logical(1)] Submit chunks as array jobs? Default is FALSE.
inds	[integer] Indices of ids / chunks to submit. Default is all. If ids get chunked, this subsets the list of shuffled ids.
resources	[list] Required resources for all batch jobs. Default is empty list.

**Value**[Registry](#)


---

batchMapResults	<i>Maps a function over the results of a registry by using batchMap.</i>
-----------------	--

---

**Description**

Maps a function over the results of a registry by using batchMap.

**Usage**

```
batchMapResults(reg, reg2, fun, ..., ids, part = NA_character_,
  more.args = list())
```

**Arguments**

reg	[Registry] Registry whose results should be mapped by fun.
reg2	[Registry] Empty registry that should store the job for the mapping.
fun	[function(job, res, ...)] Function to map over results of reg. Further arguments come from ... of batchMapResults and more.args.
...	[any] Further arguments to vectorize over (list or vector). Must all be the same length as number of results in reg.
ids	[integer] Ids of jobs whose results should be mapped with fun. Default is all jobs.
part	[character] Only useful for multiple result files, then defines which result file part(s) should be loaded. NA means all parts are loaded, which is the default.
more.args	[list] A list of other arguments passed to fun. Default is empty list.

**Value**

Vector of type integer with job ids.

**Examples**

```
reg1 = makeRegistry(id = "BatchJobsExample1", file.dir = tempfile(), seed = 123)
# square some numbers
f = function(x) x^2
batchMap(reg1, f, 1:10)

# submit jobs and wait for the jobs to finish
submitJobs(reg1)
waitForJobs(reg1)

# look at results
reduceResults(reg1, fun = function(aggr,job,res) c(aggr, res))

reg2 = makeRegistry(id = "BatchJobsExample2", file.dir = tempfile(), seed = 123)

# define function to tranform results, we simply do the inverse of the squaring
g = function(job, res) sqrt(res)
batchMapResults(reg1, reg2, fun = g)

# submit jobs and wait for the jobs to finish
submitJobs(reg2)
waitForJobs(reg2)

# check results
reduceResults(reg2, fun = function(aggr,job,res) c(aggr, res))
```

---

batchReduce	<i>Reduces via a binary function over a list adding jobs to a registry.</i>
-------------	---

---

**Description**

Each jobs reduces a certain number of elements on one slave. You can then submit these jobs to the batch system.

**Usage**

```
batchReduce(reg, fun, xs, init, block.size, more.args = list())
```

**Arguments**

reg	[Registry] Empty Registry.
fun	[function(aggr, x, ...)] Function to reduce xs with.
xs	[vector] Vector to reduce.
init	[any] Initial object for reducing.
block.size	[integer(1)] Number of elements of xs reduced in one job.
more.args	[list] A list of other arguments passed to fun. Default is empty list.

**Value**

Vector of type integer with job ids.

**Examples**

```
# define function to reduce on slave, we want to sum a vector
f = function(aggr, x) aggr + x
reg = makeRegistry(id = "BatchJobsExample", file.dir = tempfile(), seed = 123)

# sum 20 numbers on each slave process, i.e. 5 jobs
batchReduce(reg, fun = f, 1:100, init = 0, block.size = 5)
submitJobs(reg)
waitForJobs(reg)

# now reduce one final time on master
reduceResults(reg, fun = function(aggr,job,res) f(aggr, res))
```

---

batchReduceResults	<i>Reduces results via a binary function and adds jobs for this to a registry.</i>
--------------------	--

---

### Description

Each jobs reduces a certain number of results on one slave. You can then submit these jobs to the batch system. Later, you can do a final reduction with `reduceResults` on the master.

### Usage

```
batchReduceResults(reg, reg2, fun, ids, part = NA_character_, init,
  block.size, more.args = list())
```

### Arguments

reg	[Registry] Registry whose results should be reduced by fun.
reg2	[Registry] Empty registry that should store the job for the mapping.
fun	[function(aggr, job, res, ...)] Function to reduce results with.
ids	[integer] Ids of jobs whose results should be reduced with fun. Default is all jobs.
part	[character] Only useful for multiple result files, then defines which result file part(s) should be loaded. NA means all parts are loaded, which is the default.
init	[any] Initial object for reducing.
block.size	[integer(1)] Number of results reduced in one job.
more.args	[list] A list of other arguments passed to fun. Default is empty list.

### Value

Vector of type integer with job ids.

### Examples

```
# generating example results:
reg1 = makeRegistry(id = "BatchJobsExample1", file.dir = tempfile(), seed = 123)
f = function(x) x^2
batchMap(reg1, f, 1:20)
submitJobs(reg1)
waitForJobs(reg1)
```

```
# define function to reduce on slave, we want to sum the squares
myreduce = function(aggr, job, res) aggr + res

# sum 5 results on each slave process, i.e. 4 jobs
reg2 = makeRegistry(id = "BatchJobsExample2", file.dir = tempfile(), seed = 123)
batchReduceResults(reg1, reg2, fun = myreduce, init = 0, block.size = 5)
submitJobs(reg2)
waitForJobs(reg2)

# now reduce one final time on master
reduceResults(reg2, fun = myreduce)
```

---

batchUnexport	<i>Unload exported R objects.</i>
---------------	-----------------------------------

---

## Description

Removes RData files from the “exports” subdirectory of your `file.dir` and thereby prevents loading on the slave.

## Usage

```
batchUnexport(reg, what)
```

## Arguments

reg	[Registry] Registry.
what	[character] Names of objects to remove.

## Value

character . Invisibly returns a character vector of unexported objects.

## See Also

Other exports: [addRegistryPackages](#); [addRegistrySourceDirs](#); [addRegistrySourceFiles](#); [batchExport](#); [loadExports](#); [removeRegistryPackages](#); [removeRegistrySourceDirs](#); [removeRegistrySourceFiles](#)

---

 callFunctionOnSSHWorkers

*Call an arbitrary function on specified SSH workers.*


---

### Description

Calls can be made in parallel or consecutively, the function waits until all calls have finished and returns call results. In consecutive mode the output on the workers can also be shown on the master during computation.

Please read and understand the comments for argument `dir`.

Note that this function should only be used for short administrative tasks or information gathering on the workers, the true work horse for real computation is [submitJobs](#).

In [makeSSHWorker](#) various options for load management are possible. Note that these will be ignored for the current call to execute it immediatly.

### Usage

```
callFunctionOnSSHWorkers(nodenames, fun, ..., consecutive = FALSE,
  show.output = consecutive, simplify = TRUE, use.names = TRUE,
  dir = getwd())
```

### Arguments

<code>nodenames</code>	[character] Nodenames of workers to call function on. Only workers which were specified in your <a href="#">makeClusterFunctionsSSH</a> configuration can be used.
<code>fun</code>	[function] Function to call on workers.
<code>...</code>	[any] Arguments for <code>fun</code> .
<code>consecutive</code>	[logical(1)] Do calls consecutively and always wait until each worker is done. Default is FALSE.
<code>show.output</code>	[logical(1)] Show output of workers on master during call. Can be useful to see what is happening. Can only be used in consecutive mode. Default is consecutive.
<code>simplify</code>	[logical(1)] Should the result be simplified? See <a href="#">sapply</a> . Default is TRUE.
<code>use.names</code>	[logical(1)] Name results by <code>nodenames</code> . Default is TRUE.
<code>dir</code>	[character(1)] Directory under which a temporary registry will be created in a subdirectory for communication. This has to be somewhere on the shared filesystem. The created subdirectory will be cleaned up on exit. Default is current working directory.

**Value**

Results of function calls, either a list or simplified.

---

cfBrewTemplate	<i>Cluster functions helper: Brew your template into a job description file.</i>
----------------	--

---

**Description**

This function is only intended for use in your own cluster functions implementation.

Calls brew silently on your template, any error will lead to an exception. If debug mode is turned on in the configuration, the file is stored at the same place as the corresponding R script in the “jobs”-subdir of your files directory, otherwise in the temp dir via [tempfile](#).

**Usage**

```
cfBrewTemplate(conf, template, rscript, extension)
```

**Arguments**

conf	[environment] BatchJobs configuration.
template	[character(1)] Job description template as a char vector, possibly read in via <a href="#">cfReadBrewTemplate</a> .
rscript	[character(1)] File path to your corresponding R script for the job.
extension	[character(1)] Extension for the job description file, e.g. “pbs”.

**Value**

character(1) . File path of result.

---

cfHandleUnknownSubmitError	<i>Cluster functions helper: Handle an unknown error during job submission.</i>
----------------------------	---

---

**Description**

This function is only intended for use in your own cluster functions implementation.

Simply constructs a [SubmitJobResult](#) object with status code 101, NA as batch job id and an informative error message containing the output of the OS command in output.

**Usage**

```
cfHandleUnknownSubmitError(cmd, exit.code, output)
```

**Arguments**

cmd	[character(1)] OS command used to submit the job, e.g. qsub.
exit.code	[integer(1)] Exit code of the OS command, should not be 0.
output	[character] Output of the OS command, hopefully an informative error message. If these are multiple lines in a vector, they are automatically pasted together.

**Value**

[SubmitJobResult](#) .

---

cfKillBatchJob

*Cluster functions helper: Kill a batch job via OS command*

---

**Description**

This function is only intended for use in your own cluster functions implementation.

Calls the OS command to kill a job via system like this: “cmd batch.job.id”. If the command returns an exit code > 0, the command is repeated after a 1 second sleep `max.tries-1` times. If the command failed in all tries, an exception is generated.

**Usage**

```
cfKillBatchJob(cmd, batch.job.id, max.tries = 3L)
```

**Arguments**

cmd	[character(1)] OS command, e.g. “qdel”.
batch.job.id	[character(1)] Id of the batch job on the batch system.
max.tries	[integer(1)] Number of total times to try execute the OS command in cases of failures. Default is 3.

**Value**

Nothing.



---

cfReadBrewTemplate	<i>Cluster functions helper: Read in your brew template file.</i>
--------------------	---

---

### Description

This function is only intended for use in your own cluster functions implementation.

Simply reads your template and returns it as a character vector. If you do this in the constructor of your cluster functions once, you can avoid this repeated file access later on.

### Usage

```
cfReadBrewTemplate(template.file)
```

### Arguments

template.file	[character(1)] File path.
---------------	------------------------------

### Value

character .

---

configuration	<i>BatchJobs configuration.</i>
---------------	---------------------------------

---

### Description

In order to understand how the package should be configured please read <https://github.com/tudo-r/BatchJobs/wiki/Configuration>.

### See Also

Other conf: [getConfig](#); [loadConfig](#); [setConfig](#)

---

debugMulticore	<i>Helper function to debug multicore mode.</i>
----------------	---

---

**Description**

Useful in case of severe errors. Tries different operations of increasing difficulty and provides debug output on the console

**Usage**

```
debugMulticore(r.options)
```

**Arguments**

r.options	[list] Options for R and Rscript, one option per element of the vector, a la “vanilla”. Default is c(“--no-save”, “--no-restore”, “--no-init-file”, “--no-site-file”).
-----------	--

**Value**

Nothing.

**See Also**

Other debug: [debugSSH](#); [getErrorMessages](#); [getJobInfo](#); [getLogFiles](#); [grepLogs](#); [killJobs](#); [resetJobs](#); [setJobFunction](#); [showLog](#); [testJob](#)

---

debugSSH	<i>Helper function to debug SSH mode.</i>
----------	---

---

**Description**

Useful in case of configuration problems. Tries different operations of increasing difficulty and provides debug output on the console.

Note that this function does not access nor use information specified for your cluster functions in your configuration.

**Usage**

```
debugSSH(nodename, rhome = "", r.options = c(“--no-save”, “--no-restore”,
“--no-init-file”, “--no-site-file”), dir = getwd())
```

**Arguments**

nodename	[character(1)] Node on which worker should be constructed for the test.
rhome	[character(1)] Path to R installation on the worker. "" means R installation on the PATH is used, of course this implies that it must be on the PATH (also for non-interactive shells)! Default is "".
r.options	[list] Options for R and Rscript, one option per element of the vector, a la "--vanilla". Default is c("--no-save", "--no-restore", "--no-init-file", "--no-site-file").
dir	[character(1)] Path where internally used test registries can be created. Note that this must be shared for the worker. Default is current working directory.

**Value**

Nothing.

**See Also**

Other debug: [debugMulticore](#); [getErrorMessages](#); [getJobInfo](#); [getLogFiles](#); [grepLogs](#); [killJobs](#); [resetJobs](#); [setJobFunction](#); [showLog](#); [testJob](#)

---

filterResults	<i>Find all results where a specific condition is true.</i>
---------------	---

---

**Description**

Find all results where a specific condition is true.

**Usage**

```
filterResults(reg, ids, fun, ...)
```

**Arguments**

reg	[ <a href="#">Registry</a> ] Registry.
ids	[integer] Ids of jobs whose results you want to test for the condition. Default is all jobs for which results are available.
fun	[fun(job, res)] Predicate function that returns TRUE or FALSE.
...	[any] Additional arguments to fun.

**Value**

integer . Ids of jobs where fun(job, result) returns TRUE.

**Examples**

```
reg = makeRegistry(id = "BatchJobsExample", file.dir = tempfile(), seed = 123)
f = function(x) x^2
batchMap(reg, f, 1:10)
submitJobs(reg)
waitForJobs(reg)

# which square numbers are even:
filterResults(reg, fun = function(job, res) res %% 2 == 0)
```

---

findDone

*Find jobs depending on computational state.*

---

**Description**

findDone: Find jobs which successfully terminated.

findNotDone: Find jobs for which results are still missing.

findMissingResults: Deprecated. Alias for findNotDone.

findErrors: Find jobs where errors occurred.

findNotErrors: Find jobs where no errors occurred.

findTerminated: Find jobs which have terminated (done / error).

findNotTerminated: Find jobs which have not terminated (not done / no error).

findSubmitted: Find jobs which have been submitted.

findNotSubmitted: Find jobs which have not been submitted.

findOnSystem: Find jobs which are present on the batch system at the moment.

findNotOnSystem: Find jobs which are not present on the batch system at the moment.

findRunning: Find jobs which are running.

findNotRunning: Find jobs which are not running.

findStarted: Find jobs which have been started on the batch system.

findNotStarted: Find jobs which have not been started on the batch system. findNotRunning: Find jobs which are not running.

findExpired: Find jobs where walltime was probably hit. Right now the heuristic is as follows: Find all jobs that have started, did not abort with an error, did not complete with a result and are not submitted or running anymore. Note that this heuristic does not include jobs the scheduler loses before starting.

findDisappeared: Find jobs which disappeared from the system. Right now the heuristic is as follows: Find all jobs that are submitted but not started nor on the system anymore.

**Usage**

```
findDone(reg, ids, limit = NULL)
findNotDone(reg, ids, limit = NULL)
findMissingResults(reg, ids, limit = NULL)
findErrors(reg, ids, limit = NULL)
findNotErrors(reg, ids, limit = NULL)
findTerminated(reg, ids, limit = NULL)
findNotTerminated(reg, ids, limit = NULL)
findSubmitted(reg, ids, limit = NULL)
findNotSubmitted(reg, ids, limit = NULL)
findOnSystem(reg, ids, limit = NULL)
findNotOnSystem(reg, ids, limit = NULL)
findRunning(reg, ids, limit = NULL)
findNotRunning(reg, ids, limit = NULL)
findStarted(reg, ids, limit = NULL)
findNotStarted(reg, ids, limit = NULL)
findExpired(reg, ids, limit = NULL)
findDisappeared(reg, ids, limit = NULL)
```

**Arguments**

reg	[Registry] Registry.
ids	[integer] Subset of job ids to restrict the result to. Default is all jobs.
limit	[integer(1)] Limit the number of returned ids. Default is all ids.

**Value**

integer . Ids of jobs.

---

findJobs	<i>Finds ids of jobs that match a query.</i>
----------	--

---

**Description**

Finds ids of jobs that match a query.

**Usage**

```
findJobs(reg, ids, pars, jobnames)
```

**Arguments**

reg	[Registry] Registry.
ids	[integer] Subset of job ids to restrict the result to. Default is all jobs.
pars	[R expression] All jobs whose parameters match the given expression are selected. This implies that you have named the parameters when you passed the vectors. If you forgot to do this you can use .arg1, .arg2, etc., to refer to the the unnamed ones.
jobnames	[character] Restrict to jobs with stored names. Exact matching is used.

**Value**

integer . Ids for jobs which match the query.

**Examples**

```
reg = makeRegistry(id = "BatchJobsExample", file.dir = tempfile(), seed = 123)
f = function(x, y) x * y
batchExpandGrid(reg, f, x = 1:2, y = 1:3)
findJobs(reg, pars = (y > 2))
```

---

getConfig	<i>Returns a list of BatchJobs configuration settings</i>
-----------	---

---

**Description**

Returns a list of BatchJobs configuration settings

**Usage**

```
getConfig()
```

**Value**

list of current configuration variables with class "Config".

**See Also**

Other conf: [.BatchJobs.R](#), [configuration](#); [loadConfig](#); [setConfig](#)

---

getErrorMessage	<i>Get error messages of jobs.</i>
-----------------	------------------------------------

---

**Description**

Get error messages of jobs.

**Usage**

```
getErrorMessage(reg, ids)
```

**Arguments**

reg	[ <a href="#">Registry</a> ] Registry.
ids	[integer] Ids of jobs. Default is all jobs with errors.

**Value**

character . Error messages for jobs as character vector  
NA if job has terminated successfully.

**See Also**

Other debug: [debugMulticore](#); [debugSSH](#); [getJobInfo](#); [getLogFiles](#); [grepLogs](#); [killJobs](#); [resetJobs](#); [setJobFunction](#); [showLog](#); [testJob](#)

---

getJob	<i>Get job from registry by id.</i>
--------	-------------------------------------

---

**Description**

Get job from registry by id.

**Usage**

```
getJob(reg, id, check.id = TRUE)
```

**Arguments**

reg	[Registry] Registry.
id	[integer(1)] Id of job.
check.id	[logical(1)] Check the job id? Default is TRUE.

**Value**

[Job](#) .

---

getJobIds	<i>Get ids of jobs in registry.</i>
-----------	-------------------------------------

---

**Description**

Get ids of jobs in registry.

**Usage**

```
getJobIds(reg)
```

**Arguments**

reg	[Registry] Registry.
-----	-------------------------

**Value**

character .



---

getJobInfo	<i>Get computational information of jobs.</i>
------------	---

---

### Description

error messages (shortened, see [showLog](#) for detailed error messages), Returns time stamps (submitted, started, done), time running, approximate memory usage (in Mb, see note) time in queue, hostname of the host the job was executed, assigned batch ID, the R PID and the seed of the job.

### Usage

```
getJobInfo(reg, ids, pars = FALSE, prefix.pars = FALSE, select,
           unit = "seconds")
```

### Arguments

reg	[Registry] Registry.
ids	[integer] Ids of jobs. Default is all jobs.
pars	[logical(1)] Include job parameters in the output? Default is FALSE.
prefix.pars	[logical(1)] Should a prefix be added to job parameter names (column names) to avoid name clashes? Default is FALSE.
select	[character] Select only a subset of columns. Usually this is not required and you can subset yourself, but in some rare cases it may be advantageous to not query all information. Note that the column "id" (job id) is always selected. If not provided, all columns are queried and returned.
unit	[character(1)] Unit to convert execution and queing times to. Possible values: "seconds", "minutes", "hours", "days" and "weeks". Default is "seconds".

### Value

data.frame .

### Note

To estimate memory usage the sum of the last column of [gc](#) is used.

### See Also

Other debug: [debugMulticore](#); [debugSSH](#); [getErrorMessages](#); [getLogFiles](#); [grepLogs](#); [killJobs](#); [resetJobs](#); [setJobFunction](#); [showLog](#); [testJob](#)

---

getJobNr	<i>Get number of jobs in registry.</i>
----------	--

---

**Description**

Get number of jobs in registry.

**Usage**

```
getJobNr(reg)
```

**Arguments**

reg	[Registry] Registry.
-----	-------------------------

**Value**

integer(1) .

---

getJobParamDf	<i>Returns parameters for all jobs as the rows of a data.frame.</i>
---------------	---

---

**Description**

Returns parameters for all jobs as the rows of a data.frame.

**Usage**

```
getJobParamDf(reg, ids)
```

**Arguments**

reg	[Registry] Registry.
ids	[integer] Ids of jobs. Default is all jobs.

**Value**

data.frame . Rows are named with job ids.

**Examples**

```
# see batchExpandGrid
```

---

getJobResources      *Function to get the resources that were submitted for some jobs.*

---

**Description**

Throws an error if call it for unsubmitted jobs.

**Usage**

```
getJobResources(reg, ids, as.list = TRUE)
```

**Arguments**

reg	[ <a href="#">Registry</a> ] Registry.
ids	[integer] Ids of jobs. Default is all submitted jobs.
as.list	[integer(1)] If FALSE a data.frame will be returned. Default is TRUE.

**Value**

list | data.frame . List (or data.frame) of resource lists as passed to [submitJobs](#).

---

getJobs      *Get jobs from registry by id.*

---

**Description**

Get jobs from registry by id.

**Usage**

```
getJobs(reg, ids, check.ids = TRUE)
```

**Arguments**

reg	[ <a href="#">Registry</a> ] Registry.
ids	[integer] Ids of jobs. Default is all jobs.
check.ids	[logical(1)] Check the job ids? Default is TRUE.

**Value**

list of [Job](#) .

---

getLogFiles	<i>Get log file paths for jobs.</i>
-------------	-------------------------------------

---

**Description**

Get log file paths for jobs.

**Usage**

```
getLogFiles(reg, ids)
```

**Arguments**

reg	[Registry] Registry.
ids	[integer] Ids of jobs. Default is all jobs.

**Value**

character . Vector of file paths to log files.

**See Also**

Other debug: [debugMulticore](#); [debugSSH](#); [getErrorMessages](#); [getJobInfo](#); [grepLogs](#); [killJobs](#); [resetJobs](#); [setJobFunction](#); [showLog](#); [testJob](#)

---

getResources	<i>Function to get job resources in job function.</i>
--------------	---

---

**Description**

Return the list passed to [submitJobs](#), e.g. nodes, walltime, etc.

**Usage**

```
getResources()
```

**Details**

Can only be called in job function during job execution on slave.

**Value**

list .

---

getSSHWorkersInfo      *Print and return R installation and other information for SSH workers.*

---

### Description

Workers are queried in parallel via [callFunctionOnSSHWorkers](#).

The function will display a warning if the first lib path on the worker is not writable as this indicates potential problems in the configuration and [installPackagesOnSSHWorkers](#) will not work.

### Usage

```
getSSHWorkersInfo(nodenames)
```

### Arguments

nodenames      [character]  
Nodenames of workers.

### Value

list . Displayed information as a list named by nodenames.

### See Also

[callFunctionOnSSHWorkers](#)

---

grepLogs      *Grep log files for a pattern.*

---

### Description

Searches for occurrence of pattern in log files.

### Usage

```
grepLogs(reg, ids, pattern = "warn", ignore.case = TRUE, verbose = FALSE,  
range = 2L)
```

**Arguments**

reg	[Registry] Registry.
ids	[integer] Ids of jobs to grep. Default is all terminated jobs (done + errors).
pattern	[character(1)] Pattern to search for. See <a href="#">grep</a> . Default is "warn".
ignore.case	[logical(1)] Ignore case. See <a href="#">grep</a> . Default is TRUE.
verbose	[logical(1)] Print matches. Default is FALSE.
range	[integer(1)] If verbose is set to TRUE, print range leading and trailing lines for contextual information about the warning. Default is 2.

**Value**

integer . Ids of jobs where pattern was found in the log file.

**See Also**

Other debug: [debugMulticore](#); [debugSSH](#); [getErrorMessages](#); [getJobInfo](#); [getLogFiles](#); [killJobs](#); [resetJobs](#); [setJobFunction](#); [showLog](#); [testJob](#)

---

installPackagesOnSSHWorkers

*Install packages on SSH workers.*

---

**Description**

Installation is done via [callFunctionOnSSHWorkers](#) and [install.packages](#).

Note that as usual the function tries to install the packages into the first path of `.libPaths()` of each each worker.

**Usage**

```
installPackagesOnSSHWorkers(nodenames, pkgs, repos = getOption("repos"),
  consecutive = TRUE, show.output = consecutive, ...)
```

**Arguments**

nodenames	[character] Nodenames of workers.
pkgs	[character] See <a href="#">install.packages</a> .
repos	[character] See <a href="#">install.packages</a> . If the user must be queried this is of course done on the master.
consecutive	[logical(1)] See <a href="#">callFunctionOnSSHWorkers</a> . Default is TRUE.
show.output	[logical(1)] See <a href="#">callFunctionOnSSHWorkers</a> . Default is consecutive.
...	[any] Passed to <a href="#">install.packages</a> .

**Value**

Nothing.

**See Also**

[callFunctionOnSSHWorkers](#)

---

killJobs	<i>Kill some jobs on the batch system.</i>
----------	--

---

**Description**

Kill jobs which have already been submitted to the batch system. If a job is killed its internal state is reset as if it had not been submitted at all.

The function informs if (a) the job you want to kill has not been submitted, (b) the job has already terminated, (c) for some reason no batch job id is available. In all 3 cases above, nothing is changed for the state of this job and no call to the internal kill cluster function is generated.

In case of an error when killing, the function tries - after a short sleep - to kill the remaining batch jobs again. If this fails again for some jobs, the function gives up. Only jobs that could be killed are reset in the DB.

**Usage**

```
killJobs(reg, ids, progressbar = TRUE)
```

**Arguments**

reg	[Registry] Registry.
ids	[integer] Ids of jobs to kill. Default is none.
progressbar	[logical(1)] Set to FALSE to disable the progress bar. To disable all progress bars, see <a href="#">makeProgressBar</a> .

**Value**

integer . Ids of killed jobs.

**See Also**

Other debug: [debugMulticore](#); [debugSSH](#); [getErrorMessages](#); [getJobInfo](#); [getLogFiles](#); [grepLogs](#); [resetJobs](#); [setJobFunction](#); [showLog](#); [testJob](#)

**Examples**

```
## Not run:
reg = makeRegistry(id = "BatchJobsExample", file.dir = tempfile(), seed = 123)
f = function(x) Sys.sleep(x)
batchMap(reg, f, 1:10 + 5)
submitJobs(reg)
waitForJobs(reg)

# kill all jobs currently _running_
killJobs(reg, findRunning(reg))
# kill all jobs queued or running
killJobs(reg, findNotTerminated(reg))

## End(Not run)
```

---

loadConfig	<i>Load a specific configuration file.</i>
------------	--

---

**Description**

Load a specific configuration file.

**Usage**

```
loadConfig(conffile = ".BatchJobs.R")
```



**Arguments**

conffile [character(1)]  
Location of the configuration file to load. Default is “.BatchJobs.conf” in the current working directory.

**Value**

Invisibly returns a list of configuration settings.

**See Also**

Other conf: [.BatchJobs.R](#), [configuration](#); [getConfig](#); [setConfig](#)

---

loadExports	<i>Load exported R data objects.</i>
-------------	--------------------------------------

---

**Description**

Loads exported RData object files in the “exports” subdirectory of your file.dir and assigns the objects to the global environment.

**Usage**

```
loadExports(reg, what = NULL)
```

**Arguments**

reg [Registry]  
Registry.

what [character]  
Names of objects and corresponding “RData” files which should be loaded. Default NULL loads all files.

**Value**

character . Invisibly returns a character vector of loaded objects.

**See Also**

Other exports: [addRegistryPackages](#); [addRegistrySourceDirs](#); [addRegistrySourceFiles](#); [batchExport](#); [batchUnexport](#); [removeRegistryPackages](#); [removeRegistrySourceDirs](#); [removeRegistrySourceFiles](#)

---

loadRegistry	<i>Load a previously saved registry.</i>
--------------	--

---

**Description**

Load a previously saved registry.

**Usage**

```
loadRegistry(file.dir, work.dir)
```

**Arguments**

file.dir	[character(1)] Location of the file.dir to load the registry from.
work.dir	[character(1)] Location of the work. Unchanged if missing.

**Details**

Loads a previously created registry from the file system. The `file.dir` is automatically updated upon load, so be careful if you use the registry on multiple machines simultaneously, e.g. via sshfs or a samba share.

**Value**

Registry .

---

loadResult	<i>Loads a specific result file.</i>
------------	--------------------------------------

---

**Description**

Loads a specific result file.

**Usage**

```
loadResult(reg, id, part = NA_character_, missing.ok = FALSE)
```

**Arguments**

reg	[Registry] Registry.
id	[integer(1)] Id of job.
part	[character] Only useful for multiple result files, then defines which result file part(s) should be loaded. NA means all parts are loaded, which is the default.
missing.ok	[logical(1)] If FALSE an error is thrown if no result file is found. Otherwise NULL is returned. Default is FALSE.

**Value**

any . Result of job.

**See Also**

[reduceResults](#)

---

loadResults	<i>Loads result files for id vector.</i>
-------------	--

---

**Description**

Loads result files for id vector.

**Usage**

```
loadResults(reg, ids, part = NA_character_, simplify = FALSE,
            use.names = "ids", missing.ok = FALSE)
```

**Arguments**

reg	[Registry] Registry.
ids	[integer] Ids of jobs. Default is all done jobs.
part	[character] Only useful for multiple result files, then defines which result file part(s) should be loaded. NA means all parts are loaded, which is the default.
simplify	[logical(1)] Should the result be simplified to a vector, matrix or higher dimensional array if possible? Default is TRUE.
use.names	[character(1)] Name the results with job ids ("ids"), stored job names ("names") or return an unnamed result ("none"). Default is ids.

missing.ok [logical(1)]  
 If FALSE an error is thrown if the results are not found. Otherwise missing results are imputed to NULL. Default is FALSE.

### Value

list . Results of jobs as list, possibly named by ids.

### See Also

[reduceResults](#)

---

makeClusterFunctions *Create a ClusterFuntions object.*

---

### Description

Use this funtion when you implement a backend for a batch system. You must define the functions specified in the arguments.

### Usage

```
makeClusterFunctions(name, submitJob, killJob, listJobs, getArrayEnvirName,
  class = NULL, ...)
```

### Arguments

name	[character(1)] Name of cluster functions.
submitJob	[function(conf, reg, job.name, rscript, log.file, job.dir, resources, ...)] Function to submit a new job. Must return a <a href="#">SubmitJobResult</a> object. The arguments are: conf [environment]: The user configuration. reg [ <a href="#">Registry</a> ]: The registry. job.name [character(1)]: Name of job, used if the job is displayed on the batch system. This is just for display and not an id! rscript [character(1)]: File path to R script that is used to execute the job. log.file [character(1)]: File path where log file (.Rout) has to be placed. job.dir [character(1)]: Directory where all files relating to this job are placed. resources [list]: Freely definable list of required resources for this job, e.g. walltime or memory.
killJob	[function(conf, reg, batch.job.id)] Function to kill a job on the batch system. Make sure that you definately kill the job! Return value is currently ignored. The arguments are: conf [environment]: The user configuration. reg [ <a href="#">Registry</a> ]: The registry.

	batch.job.id [character(1)]: Batch job id, as produced by submitJob. Set killJob to NULL if killing jobs cannot be supported.
listJobs	[function(conf, reg)] List all jobs on the batch system for the current user / registry. This includes queued, running, held, idle, etc. jobs. Must return an integer vector of batch job ids, same format as they are produced by submitJob. It does not matter if you return a few job ids too many (e.g. all for the current user instead of all for the current registry), but you have to include all relevant ones. The arguments are: conf [environment]: The user configuration. reg [Registry]: The registry. Set listJobs to NULL if listing jobs cannot be supported.
getArrayEnvirName	[function()] Returns the name of the environment variable specifying the array ID. Should return NA if not supported.
class	[character(1)] Optional class name for cluster functions object. Useful to provide a nice print method which might show additional information about the workers. Default is NULL.
...	[any] Currently ignored.

**See Also**

Other clusterFunctions: [makeClusterFunctionsInteractive](#); [makeClusterFunctionsLSF](#); [makeClusterFunctionsLocal](#); [makeClusterFunctionsMulticore](#); [makeClusterFunctionsSGE](#); [makeClusterFunctionsSLURM](#); [makeClusterFunctionsSSH](#); [makeClusterFunctionsTorque](#)

---

makeClusterFunctionsInteractive

*Create cluster functions for sequential execution in same session.*

---

**Description**

All jobs executed under these cluster functions are executed sequentially, in the same interactive R process that you currently are. That is, submitJob does not return until the job has finished. The main use of this ClusterFunctions implementation is to test and debug programs on a local computer.

Listing jobs returns an empty vector (as no jobs can be running when you call this) and killJob returns at once (for the same reason).

**Usage**

```
makeClusterFunctionsInteractive(write.logs = TRUE)
```

**Arguments**

`write.logs` [logical(1)]  
Sink the output to log files. Turning logging off can increase the speed of calculations but makes it next to impossible to debug. Default is TRUE.

**Value**

[ClusterFunctions](#) .

**See Also**

Other clusterFunctions: [ClusterFunctions](#), [makeClusterFunctions](#); [makeClusterFunctionsLSF](#); [makeClusterFunctionsLocal](#); [makeClusterFunctionsMulticore](#); [makeClusterFunctionsSGE](#); [makeClusterFunctionsSLURM](#); [makeClusterFunctionsSSH](#); [makeClusterFunctionsTorque](#)

---

`makeClusterFunctionsLocal`

*Create cluster functions for sequential execution on local host.*

---

**Description**

All jobs executed under these cluster functions are executed sequentially, but in an independent, new R session. That is, `submitJob` does not return until the job has finished. The main use of this `ClusterFunctions` implementation is to test and debug programs on a local computer.

`Listing jobs` returns an empty vector (as no jobs can be running when you call this) and `killJob` returns at once (for the same reason).

**Usage**

```
makeClusterFunctionsLocal()
```

**Value**

[ClusterFunctions](#) .

**See Also**

Other clusterFunctions: [ClusterFunctions](#), [makeClusterFunctions](#); [makeClusterFunctionsInteractive](#); [makeClusterFunctionsLSF](#); [makeClusterFunctionsMulticore](#); [makeClusterFunctionsSGE](#); [makeClusterFunctionsSLURM](#); [makeClusterFunctionsSSH](#); [makeClusterFunctionsTorque](#)

---

`makeClusterFunctionsLSF`*Create cluster functions for LSF systems.*

---

## Description

Job files are created based on the brew template `template.file`. This file is processed with `brew` and then submitted to the queue using the `bsub` command. Jobs are killed using the `bkill` command and the list of running jobs is retrieved using `bjobs -u $USER -w`. The user must have the appropriate privileges to submit, delete and list jobs on the cluster (this is usually the case).

The template file can access all arguments passed to the `submitJob` function, see here [ClusterFunctions](#). It is the template file's job to choose a queue for the job and handle the desired resource allocations. Examples can be found on <https://github.com/tudo-r/BatchJobs/tree/master/examples/cfLSF>.

## Usage

```
makeClusterFunctionsLSF(template.file, list.jobs.cmd = c("bjobs", "-u $USER", "-w"))
```

## Arguments

<code>template.file</code>	[character(1)] Path to a brew template file that is used for the PBS job file.
<code>list.jobs.cmd</code>	[character] Change default system command / options to list jobs. The first entry is the command, the following the options. See <a href="#">system3</a> .

## Value

[ClusterFunctions](#) .

## See Also

Other clusterFunctions: [ClusterFunctions](#), [makeClusterFunctions](#); [makeClusterFunctionsInteractive](#); [makeClusterFunctionsLocal](#); [makeClusterFunctionsMulticore](#); [makeClusterFunctionsSGE](#); [makeClusterFunctionsSLURM](#); [makeClusterFunctionsSSH](#); [makeClusterFunctionsTorque](#)

---

makeClusterFunctionsMulticore

*Use multiple cores on local Linux machine to spawn parallel jobs.*

---

### Description

Jobs are spawned by starting multiple R sessions on the commandline (similar like on true batch systems). Packages `parallel` or `multicore` are not used in any way.

### Usage

```
makeClusterFunctionsMulticore(ncpus = max(getOption("mc.cores", detectCores())
  - 1, 1), max.jobs, max.load, nice, r.options = c("--no-save",
  "--no-restore", "--no-init-file", "--no-site-file"), script)
```

### Arguments

<code>ncpus</code>	[integer(1)] Number of VPUs of worker. Default is to use all cores but one, where total number of cores "available" is given by option <code>mc.cores</code> and if that is not set it is inferred by <code>detectCores</code> .
<code>max.jobs</code>	[integer(1)] Maximal number of jobs that can run concurrently for the current registry. Default is <code>ncpus</code> .
<code>max.load</code>	[numeric(1)] Load average (of the last 5 min) at which the worker is considered occupied, so that no job can be submitted. Default is <code>ncpus-1</code> .
<code>nice</code>	[integer(1)] Process priority to run R with set via <code>nice</code> . Integers between -20 and 19 are allowed. If missing, processes are not nice'd and the system default applies (usually 0).
<code>r.options</code>	[character] Options for R and Rscript, one option per element of the vector, a la "-vanilla". Default is <code>c("--no-save", "--no-restore", "--no-init-file", "--no-site-file")</code> .
<code>script</code>	[character(1)] Path to helper bash script which interacts with the worker. You really should not have to touch this, as this would imply that we have screwed up and published an incompatible version for your system. This option is only provided as a last resort for very experienced hackers. Note that the path has to be absolute. This is what is done in the package: <a href="https://github.com/tudo-r/BatchJobs/blob/master/inst/bin/linux-helper">https://github.com/tudo-r/BatchJobs/blob/master/inst/bin/linux-helper</a> Default means to take it from package directory.

### Value

`ClusterFunctions` .



**See Also**

Other clusterFunctions: [ClusterFunctions](#), [makeClusterFunctions](#); [makeClusterFunctionsInteractive](#); [makeClusterFunctionsLSF](#); [makeClusterFunctionsLocal](#); [makeClusterFunctionsSGE](#); [makeClusterFunctionsSLURM](#); [makeClusterFunctionsSSH](#); [makeClusterFunctionsTorque](#)

---

makeClusterFunctionsSGE

*Create cluster functions for Sun Grid Engine systems.*

---

**Description**

Job files are created based on the brew template `template.file`. This file is processed with `brew` and then submitted to the queue using the `qsub` command. Jobs are killed using the `qdel` command and the list of running jobs is retrieved using `qselect`. The user must have the appropriate privileges to submit, delete and list jobs on the cluster (this is usually the case).

The template file can access all arguments passed to the `submitJob` function, see here [ClusterFunctions](#). It is the template file's job to choose a queue for the job and handle the desired resource allocations. Examples can be found on <https://github.com/tudo-r/BatchJobs/tree/master/examples/cfSGE>.

**Usage**

```
makeClusterFunctionsSGE(template.file, list.jobs.cmd = c("qstat", "-u $USER"))
```

**Arguments**

<code>template.file</code>	[character(1)] Path to a brew template file that is used for the PBS job file.
<code>list.jobs.cmd</code>	[character] Change default system command / options to list jobs. The first entry is the command, the following the options. See <a href="#">system3</a> .

**Value**

[ClusterFunctions](#) .

**See Also**

Other clusterFunctions: [ClusterFunctions](#), [makeClusterFunctions](#); [makeClusterFunctionsInteractive](#); [makeClusterFunctionsLSF](#); [makeClusterFunctionsLocal](#); [makeClusterFunctionsMulticore](#); [makeClusterFunctionsSLURM](#); [makeClusterFunctionsSSH](#); [makeClusterFunctionsTorque](#)

---

`makeClusterFunctionsSLURM`*Create cluster functions for SLURM-based systems.*

---

## Description

Job files are created based on the brew template `template.file`. This file is processed with `brew` and then submitted to the queue using the `sbatch` command. Jobs are killed using the `scancel` command and the list of running jobs is retrieved using `squeue`. The user must have the appropriate privileges to submit, delete and list jobs on the cluster (this is usually the case).

The template file can access all arguments passed to the `submitJob` function, see here [ClusterFunctions](#). It is the template file's job to choose a queue for the job and handle the desired resource allocations. Examples can be found on <https://github.com/tudo-r/BatchJobs/tree/master/examples/cfSLURM>.

## Usage

```
makeClusterFunctionsSLURM(template.file, list.jobs.cmd = c("squeue", "-h",  
  "-o %i", "-u $USER"))
```

## Arguments

<code>template.file</code>	[character(1)] Path to a brew template file that is used for the PBS job file.
<code>list.jobs.cmd</code>	[character] Change default system command / options to list jobs. The first entry is the command, the following the options. See <a href="#">system3</a> .

## Value

[ClusterFunctions](#) .

## See Also

Other clusterFunctions: [ClusterFunctions](#), [makeClusterFunctions](#); [makeClusterFunctionsInteractive](#); [makeClusterFunctionsLSF](#); [makeClusterFunctionsLocal](#); [makeClusterFunctionsMulticore](#); [makeClusterFunctionsSGE](#); [makeClusterFunctionsSSH](#); [makeClusterFunctionsTorque](#)

---

 makeClusterFunctionsSSH

*Create an SSH cluster to execute jobs.*


---

### Description

Worker nodes must share the same file system and be accessible by ssh without manually entering passwords (e.g. by ssh-agent or passwordless pubkey). Note that you can also use this function to parallelize on multiple cores on your local machine. But you still have to run an ssh server and provide passwordless access to localhost.

### Usage

```
makeClusterFunctionsSSH(..., workers)
```

### Arguments

...	[SSHWorker] Worker objects, all created with <a href="#">makeSSHWorker</a> .
workers	[list of SSHWorker] Alternative way to pass workers.

### Value

ClusterFunctions .

### See Also

[makeSSHWorker](#)

Other clusterFunctions: [ClusterFunctions](#), [makeClusterFunctions](#); [makeClusterFunctionsInteractive](#); [makeClusterFunctionsLSF](#); [makeClusterFunctionsLocal](#); [makeClusterFunctionsMulticore](#); [makeClusterFunctionsSGE](#); [makeClusterFunctionsSLURM](#); [makeClusterFunctionsTorque](#)

### Examples

```
## Not run:

# Assume you have three nodes larry, curley and moe. All have 6
# cpu cores. On curley and moe R is installed under
# "/opt/R/R-current" and on larry R is installed under
# "/usr/local/R/". larry should not be used extensively because
# somebody else wants to compute there as well.
# Then a call to 'makeClusterFunctionsSSH'
# might look like this:

cluster.functions = makeClusterFunctionsSSH(
  makeSSHWorker(nodename = "larry", rhome = "/usr/local/R", max.jobs = 2),
  makeSSHWorker(nodename = "curley", rhome = "/opt/R/R-current"),
```

```

makeSSHWorker(nodename = "moe", rhome = "/opt/R/R-current"))
## End(Not run)

```

---

```
makeClusterFunctionsTorque
```

*Create cluster functions for torque-based systems.*

---

### Description

Job files are created based on the brew template `template.file`. This file is processed with `brew` and then submitted to the queue using the `qsub` command. Jobs are killed using the `qdel` command and the list of running jobs is retrieved using `qselect`. The user must have the appropriate privileges to submit, delete and list jobs on the cluster (this is usually the case).

The template file can access all arguments passed to the `submitJob` function, see here [ClusterFunctions](#). It is the template file's job to choose a queue for the job and handle the desired resource allocations. Examples can be found on <https://github.com/tudo-r/BatchJobs/tree/master/examples/cfTorque>.

### Usage

```
makeClusterFunctionsTorque(template.file, list.jobs.cmd = c("qselect",
  "-u $USER", "-s EHQRTW"))
```

### Arguments

<code>template.file</code>	[character(1)] Path to a brew template file that is used for the PBS job file.
<code>list.jobs.cmd</code>	[character] Change default system command / options to list jobs. The first entry is the command, the following the options. See <a href="#">system3</a> .

### Value

[ClusterFunctions](#) .

### See Also

Other clusterFunctions: [ClusterFunctions](#), [makeClusterFunctions](#); [makeClusterFunctionsInteractive](#); [makeClusterFunctionsLSF](#); [makeClusterFunctionsLocal](#); [makeClusterFunctionsMulticore](#); [makeClusterFunctionsSGE](#); [makeClusterFunctionsSLURM](#); [makeClusterFunctionsSSH](#)

---

makeJob *Creates a job description.*

---

### Description

Usually you will not do this manually. Every object is a list that contains the passed arguments of the constructor.

### Usage

```
makeJob(id = NA_integer_, fun, fun.id = digest(fun), pars, name, seed)
```

### Arguments

id	[integer(1)] Job id, determined by DB autoincrement. Default is NA.
fun	[function] Job function to apply on parameters.
fun.id	[character(1)] Id used to store function on disk. Default is digest(fun).
pars	[list] Parameter list for job function.
name	[character(1)] Alias name for this job.
seed	[integer(1)] Random seed for job.

---

makeRegistry *Construct a registry object.*

---

### Description

Note that if you don't want links in your paths (file.dir, work.dir) to get resolved and have complete control over the way the path is used internally, pass an absolute path which begins with "/".

### Usage

```
makeRegistry(id, file.dir, sharding = TRUE, work.dir,  
  multiple.result.files = FALSE, seed, packages = character(0L),  
  src.dirs = character(0L), src.files = character(0L), skip = TRUE)
```

**Arguments**

<code>id</code>	[character(1)] Name of registry. Displayed e.g. in mails or in cluster queue.
<code>file.dir</code>	[character(1)] Path where files regarding the registry / jobs should be saved. Default is “<id>-files” in current working directory if <code>id</code> is set.
<code>sharding</code>	[logical(1)] Enable sharding to distribute result files into different subdirectories? Important if you have many experiments. Default is TRUE.
<code>work.dir</code>	[character(1)] Working directory for R process when experiment is executed. Default is the current working directory when registry is created.
<code>multiple.result.files</code>	[logical(1)] Should a result file be generated for every list element of the returned list of the job function? Note that the function provided to <code>batchMap</code> or <code>batchReduce</code> must return a named list if this is set to TRUE. The result file will be named “<id>-result-<element name>.RData” instead of “<id>-result.RData”. Default is FALSE.
<code>seed</code>	[integer(1)] Start seed for experiments. The first experiment in the registry will use this seed, for the subsequent ones the seed is incremented by 1. Default is a random number from 1 to <code>.Machine\$integer.max/2</code> .
<code>packages</code>	[character] Packages that will always be loaded on each node. Default is <code>character(0)</code> .
<code>src.dirs</code>	[character] Directories containing R scripts to be sourced on registry load (both on slave and master). Files not matching the pattern “\.[Rr]\$" are ignored. Useful if you have many helper functions that are needed during the execution of your jobs. These files should only contain function definitions and no executable code. Default is <code>character(0)</code> .
<code>src.files</code>	[character] R scripts files to be sourced on registry load (both on slave and master). Useful if you have many helper functions that are needed during the execution of your jobs. These files should only contain function and constant definitions and no long running, executable code. These paths are considered to be relative to your <code>work.dir</code> . As a last remedy in problematic cases you can use absolute paths, by passing paths that start with “/”, see the comment about <code>file.dir</code> and <code>work.dir</code> above, where we allow the same thing. Note that this is a less portable approach and therefore usually a less good idea. Default is <code>character(0)</code> .
<code>skip</code>	[logical(1)] Skip creation of a new registry if a registry is found in <code>file.dir</code> . Defaults to TRUE.

**Details**

Every object is a list that contains the passed arguments of the constructor.

**Value****Registry****Examples**

```
reg = makeRegistry(id = "BatchJobsExample", file.dir = tempfile(), seed = 123)
print(reg)
```

---

makeSSHWorker	<i>Create SSH worker for SSH cluster functions.</i>
---------------	---

---

**Description**

Create SSH worker for SSH cluster functions.

**Usage**

```
makeSSHWorker(nodename, rhome = "", ncpus, max.jobs, max.load, nice,
  r.options = c("--no-save", "--no-restore", "--no-init-file",
    "--no-site-file"), script)
```

**Arguments**

nodename	[character(1)] Host name of node.
rhome	[character(1)] Path to R installation on worker. "" means R installation on the PATH is used, of course this implies that it must be on the PATH (also for non-interactive shells)! Default is "".
ncpus	[integers(1)] Number of VPUs of worker. Default means to query the worker via "/proc/cpuinfo".
max.jobs	[integer(1)] Maximal number of jobs that can run concurrently for the current registry. Default is ncpus.
max.load	[numeric(1)] Load average (of the last 5 min) at which the worker is considered occupied, so that no job can be submitted. Default is ncpus-1.
nice	[integer(1)] Process priority to run R with set via nice. Integers between -20 and 19 are allowed. If missing, processes are not nice'd and the system default applies (usually 0). Default is no niceing.

r.options	[character] Options for R and Rscript, one option per element of the vector, a la “-vanilla”. Default is c(“--no-save”, “--no-restore”, “--no-init-file”, “--no-site-file”).
script	[character(1)] Path to helper bash script which interacts with the worker. You really should not have to touch this, as this would imply that we have screwed up and published an incompatible version for your system. This option is only provided as a last resort for very experienced hackers. Note that the path has to be absolute. This is what is done in the package: <a href="https://github.com/tudo-r/BatchJobs/blob/master/inst/bin/linux-helper">https://github.com/tudo-r/BatchJobs/blob/master/inst/bin/linux-helper</a> Default means to take it from package directory.

**Value**

[SSHWorker](#) .

---

makeSubmitJobResult    *Create a SubmitJobResult object.*

---

**Description**

Use this function in your implementation of [makeClusterFunctions](#) to create a return value for the submitJob function.

**Usage**

```
makeSubmitJobResult(status, batch.job.id, msg, ...)
```

**Arguments**

status	[integer(1)] Launch status of job. 0 means success, codes between 1 and 100 are temporary errors and any error greater than 100 is a permanent failure.
batch.job.id	[character(1)] Unique id of this job on batch system. Note that this is not the usual job id used in BatchJobs! Must be globally unique so that the job can be terminated using just this information.
msg	[character(1)] Optional error message in case status is not equal to 0. Default is “OK”, “TEMPERR”, “ERROR”, depending on status.
...	[any] Currently unused.

**Value**

[SubmitJobResult](#) . A list, containing status, batch.job.id and msg.



---

reduceResults	<i>Reduce results from result directory.</i>
---------------	--

---

### Description

The following functions provide ways to reduce result files into either specific R objects (like vectors, lists, matrices or data.frames) or to arbitrarily aggregate them, which is a more general operation.

### Usage

```
reduceResults(reg, ids, part = NA_character_, fun, init, impute.val,
  progressbar = TRUE, ...)
```

```
reduceResultsList(reg, ids, part = NA_character_, fun, ...,
  use.names = "ids", impute.val, progressbar = TRUE)
```

```
reduceResultsVector(reg, ids, part = NA_character_, fun, ...,
  use.names = "ids", impute.val)
```

```
reduceResultsMatrix(reg, ids, part = NA_character_, fun, ..., rows = TRUE,
  use.names = "ids", impute.val)
```

```
reduceResultsDataFrame(reg, ids, part = NA_character_, fun, ...,
  use.names = "ids", impute.val,
  strings.as.factors = default.stringsAsFactors())
```

### Arguments

reg	[ <a href="#">Registry</a> ] Registry.
ids	[integer] Ids of selected jobs. Default is all jobs for which results are available.
part	[character] Only useful for multiple result files, then defines which result file part(s) should be loaded. NA means all parts are loaded, which is the default.
fun	[function] For reduceResults, a function function(aggr, job, res, ...) to reduce things, for all others, a function function(job, res, ...) to select stuff. Here, job is the current job descriptor (see <a href="#">Job</a> ), result is the current result object and aggr are the so far aggregated results. When using reduceResults, your function should add the stuff you want to have from job and result to aggr and return that. When using the other reductions, you should select the stuff you want to have from job and result and return something that can be coerced to an element of the selected return data structure (reasonable conversion is tried internally). Default behavior for this argument is to return res, except for reduceResults where no default is available.

<code>init</code>	[ANY] Initial element, as used in <a href="#">Reduce</a> . Default is first result.
<code>impute.val</code>	[any] For <code>reduceResults</code> : If not missing, the value of <code>impute.val</code> is passed to function <code>fun</code> as argument <code>res</code> for jobs with missing results. For the specialized reduction functions <code>reduceResults[Type]</code> : If not missing, <code>impute.val</code> is used as a replacement for the return value of <code>fun</code> on missing results.
<code>progressbar</code>	[logical(1)] Set to FALSE to disable the progress bar. To disable all progress bars, see <a href="#">makeProgressBar</a> .
<code>...</code>	[any] Additional arguments to <code>fun</code> .
<code>use.names</code>	[character(1)] Name the results with job ids (“ids”), stored job names (“names”) or return a unnamed result (“none”). Default is <code>ids</code> .
<code>rows</code>	[logical(1)] Should the selected vectors be used as rows (or columns) in the result matrix? Default is TRUE.
<code>strings.as.factors</code>	[logical(1)] Should all character columns in result be converted to factors? Default is <code>default.stringsAsFactors()</code> .

**Value**

Aggregated results, return type depends on function. If `ids` is empty: `reduceResults` returns `init` (if available) or NULL, `reduceResultsVector` returns `c()`, `reduceResultsList` returns `list()`, `reduceResultsMatrix` returns `matrix(0, 0, 0)`, `reduceResultsDataFrame` returns `data.frame()`.

**Examples**

```
# generate results:
reg = makeRegistry(id = "BatchJobsExample", file.dir = tempfile(), seed = 123)
f = function(x) x^2
batchMap(reg, f, 1:5)
submitJobs(reg)
waitForJobs(reg)

# reduce results to a vector
reduceResultsVector(reg)
# reduce results to sum
reduceResults(reg, fun = function(aggr, job, res) aggr+res)

reg = makeRegistry(id = "BatchJobsExample", file.dir = tempfile(), seed = 123)
f = function(x) list(a = x, b = as.character(2*x), c = x^2)
batchMap(reg, f, 1:5)
submitJobs(reg)
waitForJobs(reg)
```

```

# reduce results to a vector
reduceResultsVector(reg, fun = function(job, res) res$a)
reduceResultsVector(reg, fun = function(job, res) res$b)
# reduce results to a list
reduceResultsList(reg)
# reduce results to a matrix
reduceResultsMatrix(reg, fun = function(job, res) res[c(1,3)])
reduceResultsMatrix(reg, fun = function(job, res) c(foo = res$a, bar = res$c), rows = TRUE)
reduceResultsMatrix(reg, fun = function(job, res) c(foo = res$a, bar = res$c), rows = FALSE)
# reduce results to a data.frame
print(str(reduceResultsDataFrame(reg)))
# reduce results to a sum
reduceResults(reg, fun = function(aggr, job, res) aggr+res$a, init = 0)

```

---

removeRegistry	<i>Remove a registry object.</i>
----------------	----------------------------------

---

### Description

If there are no live/running jobs, the registry will be closed and all of its files will be removed from the file system. If there are live/running jobs, an informative error is generated. The default is to prompt the user for confirmation.

### Usage

```
removeRegistry(reg, ask = c("yes", "no"))
```

### Arguments

reg	[Registry] Registry.
ask	[character(1)] If "yes" the user is prompted to confirm the action. If trying to prompt the user this way in a non-interactive session, then an informative error is generated. If "no", the registry will be removed without further confirmation.

### Value

logical[1]

---

removeRegistryPackages

*Remove packages from registry.*

---

### Description

Mutator function for packages in [makeRegistry](#).

### Usage

```
removeRegistryPackages(reg, packages)
```

### Arguments

reg	[Registry] Registry.
packages	[character] Packages to remove from registry.

### Value

[Registry](#) . Changed registry.

### See Also

Other exports: [addRegistryPackages](#); [addRegistrySourceDirs](#); [addRegistrySourceFiles](#); [batchExport](#); [batchUnexport](#); [loadExports](#); [removeRegistrySourceDirs](#); [removeRegistrySourceFiles](#)

---

removeRegistrySourceDirs

*Remove packages from registry.*

---

### Description

Mutator function for src.dirs in [makeRegistry](#).

### Usage

```
removeRegistrySourceDirs(reg, src.dirs)
```

### Arguments

reg	[Registry] Registry.
src.dirs	[character] Paths to remove from registry.

**Value**

[Registry](#) . Changed registry.

**See Also**

Other exports: [addRegistryPackages](#); [addRegistrySourceDirs](#); [addRegistrySourceFiles](#); [batchExport](#); [batchUnexport](#); [loadExports](#); [removeRegistryPackages](#); [removeRegistrySourceFiles](#)

---

`removeRegistrySourceFiles`

*Remove source files from registry.*

---

**Description**

Mutator function for `src.files` in [makeRegistry](#).

**Usage**

```
removeRegistrySourceFiles(reg, src.files)
```

**Arguments**

<code>reg</code>	[Registry] Registry.
<code>src.files</code>	[character] Paths to remove from registry.

**Value**

[Registry](#) . Changed registry.

**See Also**

Other exports: [addRegistryPackages](#); [addRegistrySourceDirs](#); [addRegistrySourceFiles](#); [batchExport](#); [batchUnexport](#); [loadExports](#); [removeRegistryPackages](#); [removeRegistrySourceDirs](#)

---

resetJobs	<i>Reset computational state of jobs.</i>
-----------	---

---

### Description

Reset state of jobs in the database. Useful under two circumstances: Either to re-submit them because of changes in e.g. external data or to resolve rare issues when jobs are killed in an unfortunate state and therefore blocking your registry.

The function internally lists all jobs on the batch system and if those include some of the jobs you want to reset, it informs you to kill them first by raising an exception. If you really know what you are doing, you may set `force` to `TRUE` to omit this sanity check. Note that this is a dangerous operation to perform which may harm the database integrity. In this case you **HAVE** to make externally sure that none of the jobs you want to reset are still running.

### Usage

```
resetJobs(reg, ids, force = FALSE)
```

### Arguments

reg	[Registry] Registry.
ids	[integer] Ids of jobs to kill. Default is none.
force	[logical(1)] Reset jobs without checking whether they are currently running. <b>READ THE DETAILS SECTION!</b> Default is FALSE.

### Value

Vector of reseted job ids.

### See Also

Other debug: [debugMulticore](#); [debugSSH](#); [getErrorMessages](#); [getJobInfo](#); [getLogFiles](#); [grepLogs](#); [killJobs](#); [setJobFunction](#); [showLog](#); [testJob](#)

---

sanitizePath	<i>Sanitize a path</i>
--------------	------------------------

---

**Description**

Replaces backward slashes with forward slashes and optionally normalizes the path.

**Usage**

```
sanitizePath(path, make.absolute = TRUE, normalize.absolute = FALSE)
```

**Arguments**

path	[character] Vector of paths to sanitize.
make.absolute	[logical] If TRUE convert to an absolute path.
normalize.absolute	[logical] Also call <a href="#">normalizePath</a> on absolute paths? This will immediately resolve symlinks.

**Value**

character with sanitized paths.

---

setConfig	<i>Set and overwrite configuration settings</i>
-----------	---

---

**Description**

Set and overwrite configuration settings

**Usage**

```
setConfig(conf = list(), ...)
```

**Arguments**

conf	[Config or list] List of configuration parameters as returned by <a href="#">loadConfig</a> or <a href="#">getConfig</a> .
...	[ANY] Named configuration parameters. Overwrites parameters in conf, if provided.

**Value**

Invisibly returns a list of configuration settings.

**See Also**

Other conf: [.BatchJobs.R, configuration](#); [getConfig](#); [loadConfig](#)

---

setJobFunction	<i>Sets the job function for already existing jobs.</i>
----------------	---

---

**Description**

Use this function only as last measure when there is a bug in a part of your job function and you have already computed a large number of (unaffected) results. This function allows you to fix the error and to associate the jobs with the corrected function.

Note that by default the computational state of the affected jobs is also reset.

**Usage**

```
setJobFunction(reg, ids, fun, more.args = list(), reset = TRUE,
              force = FALSE)
```

**Arguments**

reg	[ <a href="#">Registry</a> ] Registry.
ids	[integer] Ids of jobs. Default is all jobs.
fun	[function] Replacement function.
more.args	[list] A list of other arguments passed to fun. Default is empty list.
reset	[logical(1)] Reset job status via <a href="#">resetJobs</a> . Default is TRUE.
force	[logical(1)] See <a href="#">resetJobs</a> . Default is FALSE.

**Value**

Nothing.

**See Also**

Other debug: [debugMulticore](#); [debugSSH](#); [getErrorMessages](#); [getJobInfo](#); [getLogFiles](#); [grepLogs](#); [killJobs](#); [resetJobs](#); [showLog](#); [testJob](#)



---

setJobNames	<i>Set job names.</i>
-------------	-----------------------

---

**Description**

Set job names.

**Usage**

```
setJobNames(reg, ids, jobnames)
```

**Arguments**

reg	[Registry] Registry.
ids	[integer] Ids of jobs. Default is all jobs.
jobnames	[character] Character vector with length equal to length(ids). NA removes the names stored in the registry. A single NA is replicated to match the length of ids provided.

**Value**

Named vector of job ids.

---

showClusterStatus	<i>Show information about available computational resources on cluster.</i>
-------------------	---

---

**Description**

Currently only supported for multicore and SSH mode. Displays: Name of node, current load, number of running R processes, number of R processes with more than 50 The latter counts either jobs belonging to reg or all BatchJobs jobs if reg was not passed.

**Usage**

```
showClusterStatus(reg)
```

**Arguments**

reg	[Registry] Registry. Must not be passed and this is the default.
-----	---

**Value**

data.frame . Returns the displayed table invisibly.

---

showLog	<i>Display the contents of a log file.</i>
---------	--

---

### Description

Display the contents of a log file, useful in case of errors.

Note this rare special case: When you use chunking, submit some jobs, some jobs fail, then you resubmit these jobs again in different chunks, the log files will contain the log of the old, failed job as well. showLog tries to jump to the correct part of the new log file with a supported pager.

### Usage

```
showLog(reg, id, pager = getOption("pager"))
```

### Arguments

reg	[Registry] Registry.
id	[integer(1)] Id of selected job. Default is first id in registry.
pager	[any] Pager to use to display the log. Defaults to <code>getOption("pager")</code> . This option is passed to <code>file.show</code> and is highly OS dependant and GUI dependant. If either R's pager or the environment variable "PAGER" is set to "less" or "vim", the correct part of the log file will be shown. Otherwise you find information about the correct part in the beginning of the displayed file.

### Value

character(1) . Invisibly returns path to log file.

### See Also

Other debug: [debugMulticore](#); [debugSSH](#); [getErrorMessages](#); [getJobInfo](#); [getLogFiles](#); [grepLogs](#); [killJobs](#); [resetJobs](#); [setJobFunction](#); [testJob](#)

---

showStatus	<i>Retrieve or show status information about jobs.</i>
------------	--

---

### Description

E.g.: How many there are, how many are done, any errors, etc. showStatus displays on the console, getStatus returns an informative result without console output.

**Usage**

```
showStatus(reg, ids, run.and.exp = TRUE, errors = 10L)
```

```
getStatus(reg, ids, run.and.exp = TRUE)
```

**Arguments**

reg	[Registry] Registry.
ids	[integer] Ids of selected jobs. Default is all jobs.
run.and.exp	[logical(1)] Show running and expired jobs? Requires to list the job on the batch system. If not possible, because that cluster function is not available, this option is ignored anyway. Default is TRUE.
errors	[integer(1)] How many of the error messages should be displayed if any errors occurred in the jobs? Default is 10.

**Value**

list . List of absolute job numbers. showStatus returns them invisibly.

**Examples**

```
reg = makeRegistry(id = "BatchJobsExample", file.dir = tempfile(), seed = 123)
f = function(x) x^2
batchMap(reg, f, 1:10)
submitJobs(reg)
waitForJobs(reg)

# should show 10 submitted jobs, which are all done.
showStatus(reg)
```

---

sourceRegistryFiles    *Source registry files*

---

**Description**

Sources all files found in src.dirs and specified via src.files.

**Usage**

```
sourceRegistryFiles(reg, envir = .GlobalEnv)
```

**Arguments**

reg	[Registry] Registry.
envir	[environment] Environment to source the files into. Default is the global environment.

**Value**

Nothing.

---

submitJobs	<i>Submit jobs or chunks of jobs to batch system via cluster function.</i>
------------	--

---

**Description**

If the internal submit cluster function completes successfully, the `retries` counter is set back to 0 and the next job or chunk is submitted. If the internal submit cluster function returns a fatal error, the submit process is completely stopped and an exception is thrown. If the internal submit cluster function returns a temporary error, the submit process waits for a certain time, which is determined by calling the user-defined `wait`-function with the current `retries` counter, the counter is increased by 1 and the same job is submitted again. If `max.retries` is reached the function simply terminates.

Potential temporary submit warnings and errors are logged inside your file directory in the file “`submit.log`”. To keep track you can use `tail -f [file.dir]/submit.log` in another terminal.

**Usage**

```
submitJobs(reg, ids, resources = list(), wait, max.retries = 10L,
           chunks.as.arrayjobs = FALSE, job.delay = FALSE, progressbar = TRUE)
```

**Arguments**

reg	[Registry] Registry.
ids	[integer] Vector for job id or list of vectors of chunked job ids. Only corresponding jobs are submitted. Chunked jobs will get executed sequentially as a single job for the scheduler. Default is all jobs which were not yet submitted to the batch system.
resources	[list] Required resources for all batch jobs. The elements of this list (e.g. something like “ <code>walltime</code> ” or “ <code>nodes</code> ” are defined by your template job file. Defaults can be specified in your config file. Default is empty list.
wait	[function(retries)] Function that defines how many seconds should be waited in case of a temporary error. Default is exponential back-off with $10 \cdot 2^{\text{retries}}$ .

<code>max.retries</code>	[integer(1)] Number of times to submit one job again in case of a temporary error (like filled queues). Each time wait is called to wait a certain number of seconds. Default is 10 times.
<code>chunks.as.arrayjobs</code>	[logical(1)] If ids are passed as a list of chunked job ids, execute jobs in a chunk as array jobs. Note that your scheduler and your template must be adjusted to use this option. Default is FALSE.
<code>job.delay</code>	[function(n, i) or logical(1)] Function that defines how many seconds a job should be delayed before it starts. This is an expert option and only necessary to change when you want submit extremely many jobs. We then delay the jobs a bit to write the submit messages as early as possible to avoid writer starvation. n is the number of jobs and i the number of the ith job. The default function used with <code>job.delay</code> set to TRUE is no delay for 100 jobs or less and otherwise <code>runif(1, 0.1*n, 0.2*n)</code> . If set to FALSE (the default) delaying jobs is disabled.
<code>progressbar</code>	[logical(1)] Set to FALSE to disable the progress bar. To disable all progress bars, see <a href="#">makeProgressBar</a> .

**Value**

`integer` . Vector of submitted job ids.

**Examples**

```
reg = makeRegistry(id = "BatchJobsExample", file.dir = tempfile(), seed = 123)
f = function(x) x^2
batchMap(reg, f, 1:10)
submitJobs(reg)
waitForJobs(reg)

# Submit the 10 jobs again, now randomized into 2 chunks:
chunked = chunk(getJobIds(reg), n.chunks = 2, shuffle = TRUE)
submitJobs(reg, chunked)
```

---

sweepRegistry

*Sweep obsolete files from the file system.*

---

**Description**

Removes R scripts, log files, resource informations and temporarily stored configuration files from the registry's file directory. Assuming all your jobs completed successfully, none of these are needed for further work. This operation potentially releases quite a lot of disk space, depending on the number of your jobs. BUT A HUGE WORD OF WARNING: IF you later notice something strange and need to determine the reason for it, you are at a huge disadvantage. Only do this at your own risk and when you are sure that you have successfully completed a project and only want to archive your produced experiments and results.

**Usage**

```
sweepRegistry(reg, sweep = c("scripts", "conf"))
```

**Arguments**

reg	[Registry] Registry.
sweep	[character] Possible choices: Temporary R scripts of jobs, really not needed for anything else then execution (“scripts”), log file of jobs, think about whether you later want to inspect them (“logs”), BatchJobs configuration files which are temporarily stored on submit, really not needed for anything else then execution (“conf”), resource lists of <a href="#">submitJobs</a> which are temporarily stored on submit, think about whether you later want to inspect them (“resources”), Default is <code>c("scripts", "conf")</code> .

**Value**

logical . Invisibly returns TRUE on success and FALSE if some files could not be removed.

---

testJob	<i>Tests a job by running it with Rscript in a new process.</i>
---------	---

---

**Description**

Useful for debugging. Note that neither the registry, database or file directory are changed.

**Usage**

```
testJob(reg, id, resources = list(), external = TRUE)
```

**Arguments**

reg	[Registry] Registry.
id	[integer(1)] Id of job to test. Default is first job id of registry.
resources	[list] Usually not needed, unless you call the function <a href="#">getResources</a> in your job. See <a href="#">submitJobs</a> . Default is empty list.
external	[logical(1)] Run test in an independent external R session instead of current. The former allows to uncover missing variable definitions (which may accidentally be defined in the current global environment) and the latter is useful to get traceable exceptions. Default is TRUE.

**Value**

any . Result of job. If the job did not complete because of an error, NULL is returned.

**See Also**

Other debug: [debugMulticore](#); [debugSSH](#); [getErrorMessages](#); [getJobInfo](#); [getLogFiles](#); [grepLogs](#); [killJobs](#); [resetJobs](#); [setJobFunction](#); [showLog](#)

**Examples**

```
reg = makeRegistry(id = "BatchJobsExample", file.dir = tempfile(), seed = 123)
f = function(x) if (x==1) stop("oops") else x
batchMap(reg, f, 1:2)
testJob(reg, 2)
```

---

waitForJobs	<i>Wait for termination of jobs on the batch system.</i>
-------------	--

---

**Description**

Waits for termination of jobs while displaying a progress bar containing summarizing informations of the jobs. The following abbreviations are used in the progress bar: “S” for number of jobs on system, “D” for number of jobs successfully terminated, “E” for number of jobs terminated with an R exception and “R” for number of jobs currently running on the system.

**Usage**

```
waitForJobs(reg, ids, sleep = 10, timeout = 604800, stop.on.error = FALSE,
  progressbar = TRUE)
```

**Arguments**

reg	[ <a href="#">Registry</a> ] Registry.
ids	[integer] Vector of job ids. Default is all submitted jobs not yet terminated.
sleep	[numeric(1)] Seconds to sleep between status updates. Default is 10.
timeout	[numeric(1)] After waiting timeout seconds, show a message and return FALSE. This argument may be required on some systems where, e.g., expired jobs or jobs on hold are problematic to detect. If you don't want a timeout, set this to Inf. Default is 604800 (one week).
stop.on.error	[logical(1)] Immediately return if a job terminates with an error? Default is FALSE.
progressbar	[logical(1)] Set to FALSE to disable the progress bar. To disable all progress bars, see <a href="#">makeProgressBar</a> .

**Value**

logical(1) . Returns TRUE if all jobs terminated successfully and FALSE if either an error occurred or the timeout is reached.



# Index

- [.BatchJobs.R](#), [23](#), [33](#), [56](#)
- [.BatchJobs.R \(configuration\)](#), [17](#)
- [addRegistryPackages](#), [3](#), [4–6](#), [13](#), [33](#), [52](#), [53](#)
- [addRegistrySourceDirs](#), [4](#), [4](#), [5](#), [6](#), [13](#), [33](#), [52](#), [53](#)
- [addRegistrySourceFiles](#), [4](#), [4](#), [6](#), [13](#), [33](#), [52](#), [53](#)
- [batchExpandGrid](#), [5](#)
- [batchExport](#), [4](#), [5](#), [6](#), [13](#), [33](#), [52](#), [53](#)
- [BatchJobs](#), [7](#)
- [BatchJobs-package \(BatchJobs\)](#), [7](#)
- [batchMap](#), [5](#), [7](#), [8](#), [46](#)
- [batchMapQuick](#), [8](#)
- [batchMapResults](#), [9](#)
- [batchReduce](#), [11](#), [46](#)
- [batchReduceResults](#), [12](#)
- [batchUnexport](#), [4–6](#), [13](#), [33](#), [52](#), [53](#)
- [callFunctionOnSSHWorkers](#), [14](#), [29–31](#)
- [cfBrewTemplate](#), [15](#)
- [cfHandleUnknownSubmitError](#), [15](#)
- [cfKillBatchJob](#), [16](#)
- [cfReadBrewTemplate](#), [15](#), [17](#)
- [ClusterFunctions](#), [38–44](#)
- [ClusterFunctions \(makeClusterFunctions\)](#), [36](#)
- [configuration](#), [17](#), [23](#), [33](#), [56](#)
- [debugMulticore](#), [18](#), [19](#), [23](#), [25](#), [28](#), [30](#), [32](#), [54](#), [56](#), [58](#), [63](#)
- [debugSSH](#), [18](#), [18](#), [23](#), [25](#), [28](#), [30](#), [32](#), [54](#), [56](#), [58](#), [63](#)
- [detectCores](#), [40](#)
- [expand.grid](#), [5](#)
- [filterResults](#), [19](#)
- [findDisappeared \(findDone\)](#), [20](#)
- [findDone](#), [20](#)
- [findErrors \(findDone\)](#), [20](#)
- [findExpired \(findDone\)](#), [20](#)
- [findJobs](#), [22](#)
- [findMissingResults \(findDone\)](#), [20](#)
- [findNotDone \(findDone\)](#), [20](#)
- [findNotErrors \(findDone\)](#), [20](#)
- [findNotOnSystem \(findDone\)](#), [20](#)
- [findNotRunning \(findDone\)](#), [20](#)
- [findNotStarted \(findDone\)](#), [20](#)
- [findNotSubmitted \(findDone\)](#), [20](#)
- [findNotTerminated \(findDone\)](#), [20](#)
- [findOnSystem \(findDone\)](#), [20](#)
- [findRunning \(findDone\)](#), [20](#)
- [findStarted \(findDone\)](#), [20](#)
- [findSubmitted \(findDone\)](#), [20](#)
- [findTerminated \(findDone\)](#), [20](#)
- [gc](#), [25](#)
- [getConfig](#), [17](#), [22](#), [33](#), [55](#), [56](#)
- [getErrorMessages](#), [18](#), [19](#), [23](#), [25](#), [28](#), [30](#), [32](#), [54](#), [56](#), [58](#), [63](#)
- [getJob](#), [24](#)
- [getJobIds](#), [24](#)
- [getJobInfo](#), [18](#), [19](#), [23](#), [25](#), [28](#), [30](#), [32](#), [54](#), [56](#), [58](#), [63](#)
- [getJobNr](#), [26](#)
- [getJobParamDf](#), [26](#)
- [getJobResources](#), [27](#)
- [getJobs](#), [27](#)
- [getLogFiles](#), [18](#), [19](#), [23](#), [25](#), [28](#), [30](#), [32](#), [54](#), [56](#), [58](#), [63](#)
- [getResources](#), [28](#), [62](#)
- [getSSHWorkersInfo](#), [29](#)
- [getStatus \(showStatus\)](#), [58](#)
- [grep](#), [30](#)
- [grepLogs](#), [18](#), [19](#), [23](#), [25](#), [28](#), [29](#), [32](#), [54](#), [56](#), [58](#), [63](#)
- [install.packages](#), [30](#), [31](#)
- [installPackagesOnSSHWorkers](#), [29](#), [30](#)

- Job, [24](#), [27](#), [49](#)
- Job (makeJob), [45](#)
- killJobs, [18](#), [19](#), [23](#), [25](#), [28](#), [30](#), [31](#), [54](#), [56](#), [58](#), [63](#)
- loadConfig, [17](#), [23](#), [32](#), [55](#), [56](#)
- loadExports, [4–6](#), [13](#), [33](#), [52](#), [53](#)
- loadRegistry, [34](#)
- loadResult, [34](#)
- loadResults, [8](#), [35](#)
- makeClusterFunctions, [36](#), [38](#), [39](#), [41–44](#), [48](#)
- makeClusterFunctionsInteractive, [37](#), [37](#), [38](#), [39](#), [41–44](#)
- makeClusterFunctionsLocal, [37](#), [38](#), [38](#), [39](#), [41–44](#)
- makeClusterFunctionsLSF, [37](#), [38](#), [39](#), [41–44](#)
- makeClusterFunctionsMulticore, [37–39](#), [40](#), [41–44](#)
- makeClusterFunctionsSGE, [37–39](#), [41](#), [41](#), [42–44](#)
- makeClusterFunctionsSLURM, [37–39](#), [41](#), [42](#), [43](#), [44](#)
- makeClusterFunctionsSSH, [14](#), [37–39](#), [41](#), [42](#), [43](#), [44](#)
- makeClusterFunctionsTorque, [37–39](#), [41–43](#), [44](#)
- makeJob, [45](#)
- makeProgressBar, [32](#), [50](#), [61](#), [63](#)
- makeRegistry, [3–5](#), [8](#), [9](#), [45](#), [52](#), [53](#)
- makeSSHWorker, [14](#), [43](#), [47](#)
- makeSubmitJobResult, [48](#)
- mc.cores, [40](#)
- normalizePath, [55](#)
- options, [7](#)
- Reduce, [50](#)
- reduceResults, [12](#), [35](#), [36](#), [49](#)
- reduceResultsDataFrame (reduceResults), [49](#)
- reduceResultsList (reduceResults), [49](#)
- reduceResultsMatrix (reduceResults), [49](#)
- reduceResultsVector (reduceResults), [49](#)
- Registry, [3–5](#), [8–12](#), [19](#), [21–24](#), [26–28](#), [30](#), [32](#), [34–37](#), [47](#), [49](#), [51–54](#), [56–60](#), [62](#), [63](#)
- Registry (makeRegistry), [45](#)
- removeRegistry, [51](#)
- removeRegistryPackages, [4–6](#), [13](#), [33](#), [52](#), [53](#)
- removeRegistrySourceDirs, [4–6](#), [13](#), [33](#), [52](#), [52](#), [53](#)
- removeRegistrySourceFiles, [4–6](#), [13](#), [33](#), [52](#), [53](#), [53](#)
- resetJobs, [18](#), [19](#), [23](#), [25](#), [28](#), [30](#), [32](#), [54](#), [56](#), [58](#), [63](#)
- sanitizePath, [55](#)
- sapply, [14](#)
- setConfig, [17](#), [23](#), [33](#), [55](#)
- setJobFunction, [18](#), [19](#), [23](#), [25](#), [28](#), [30](#), [32](#), [54](#), [56](#), [58](#), [63](#)
- setJobNames, [57](#)
- showClusterStatus, [57](#)
- showLog, [18](#), [19](#), [23](#), [25](#), [28](#), [30](#), [32](#), [54](#), [56](#), [58](#), [63](#)
- showStatus, [58](#)
- sourceRegistryFiles, [59](#)
- SSHWorker, [43](#), [48](#)
- SSHWorker (makeSSHWorker), [47](#)
- SubmitJobResult, [15](#), [16](#), [36](#), [48](#)
- SubmitJobResult (makeSubmitJobResult), [48](#)
- submitJobs, [8](#), [14](#), [27](#), [28](#), [60](#), [62](#)
- sweepRegistry, [61](#)
- system3, [39](#), [41](#), [42](#), [44](#)
- tempfile, [15](#)
- testJob, [18](#), [19](#), [23](#), [25](#), [28](#), [30](#), [32](#), [54](#), [56](#), [58](#), [62](#)
- waitForJobs, [63](#)