

R with future - Parallel backends across nodes

Fabian Freund, Brigitte Wellenkamp
(KIM Hohenheim, bwHPC)

July 26th, 2022
ZOOM



UNIVERSITY OF
HOHENHEIM



(Mediocre) Outlook: Multinode multicore plans

- ▶ Simple multinode plan should be possible via R PSOCK cluster. . .
- ▶ This causes problems when used on bwUniCluster
 - ▶ connection via ssh \Rightarrow Manual login with 2FA
 - ▶ R needs to be loaded on each node separately
- ▶ MPI cluster does (at least imo) also not work on bwUniCluster
 - ▶ future uses MPI cluster from parallel/snow, this has issues on bwUniCluster



Interactive multinode PSOCK cluster works

```
#In bash
```

```
salloc -t 30 -N 2 -p dev_multiple
```

```
#In R
```

```
library(parallelly);library(future.apply)
```

```
## Loading required package: future
```

```
library(parallel) #for stopCluster
```

```
#Get hostnames
```

```
nodes <- unique(availableWorkers())
```

```
nodes
```

```
cl1 <- makeClusterPSOCK(workers=nodes)
```

```
plan(strategy = cluster,workers=cl1)
```

```
future_replicate(2,Sys.info()["nodename"])
```

```
stopCluster(cl1)
```

[Dirty trick: I had to force module load math/R/4.1.2 via

.bashrc entry]



What does makeClusterPSOCK do?

```
makeClusterPSOCK(workers=c("node1", "node2"), dryrun=TRUE)
```

```
## -----  
## Manually, (i) login into external machine 'node1':  
##  
##   '/usr/bin/ssh' -R 11421:localhost:11421 node1  
##  
## and (ii) start worker #1 from there:  
##  
##   '/usr/lib/R/bin/Rscript' --default-packages=datasets,u  
##  
## Alternatively, start worker #1 from the local machine by  
##  
##   '/usr/bin/ssh' -R 11421:localhost:11421 node1 "'/usr/  
##  
## -----  
## Manually, (i) login into external machine 'node2':  
##  
##   '/usr/bin/ssh' -R 11422:localhost:11422 node2 "'/usr/  
##
```



Ideas for non-interactive construction of multinode PSOCK cluster

- ▶ In principle, you should be able to also span a cluster without ssh across nodes
 - ▶ Don't log into node, but use `srun -w <nodename>` to launch the worker
- ▶ Problem: `makeClusterPSOCK` does not directly allow this: It's always 1) log in via `ssh` or sth similar 2) Start R worker via `Rscript`
- ▶ Code is pure R, so one should be possible to change this
- ▶ May also work in `batchtools/future.batchtools` (the former allows also to set up a ssh-connected cluster)

In case this is relevant for you: let's try it out further (\Rightarrow contact us after the course)

