sfCluster/snowfall: Managing parallel execution of R programs on a compute cluster

Jochen Knaus

Institute of Medical Biometry and Medical Informatics, University of Freiburg
DFG Forschergruppe FOR 534

jo@imbi.uni-freiburg.de

August 14, 2008

UNIVERSITÄTS FREIBURG KLINIKUM

Situation / Intention

➢ We wanted a solution for a heterogeneous infrastructure with many users with different knowledge levels running parallel R programs at the same time.

➢ Although there are many working cluster solutions for R, all of them need to have a running cluster available.

➢ Especially cluster setup and handling can be too difficult for users and therefore a barrier to get them into parallel computing.

Our solution: snowfall and sfCluster

sfCluster
Unix tool for automatic cluster management and monitoring.

snowfall
R package based on snow.
Can be used without sfCluster, but benefits of sfCluster environment.

snowfall R package

Design goals

- ➢ Connector to sfCluster.
- ➢ Easy access.
- ➢ Wrappers for essential snow functions.
- ➢ Fully supporting sequential execution without any code changes (all wrappers work in sequential mode, too) – also enable development/debugging on Windows laptops.
- ➢ Directly runnable everywhere (even without snow): programs are distributable inside packages.
- ➢ Extended error checks.
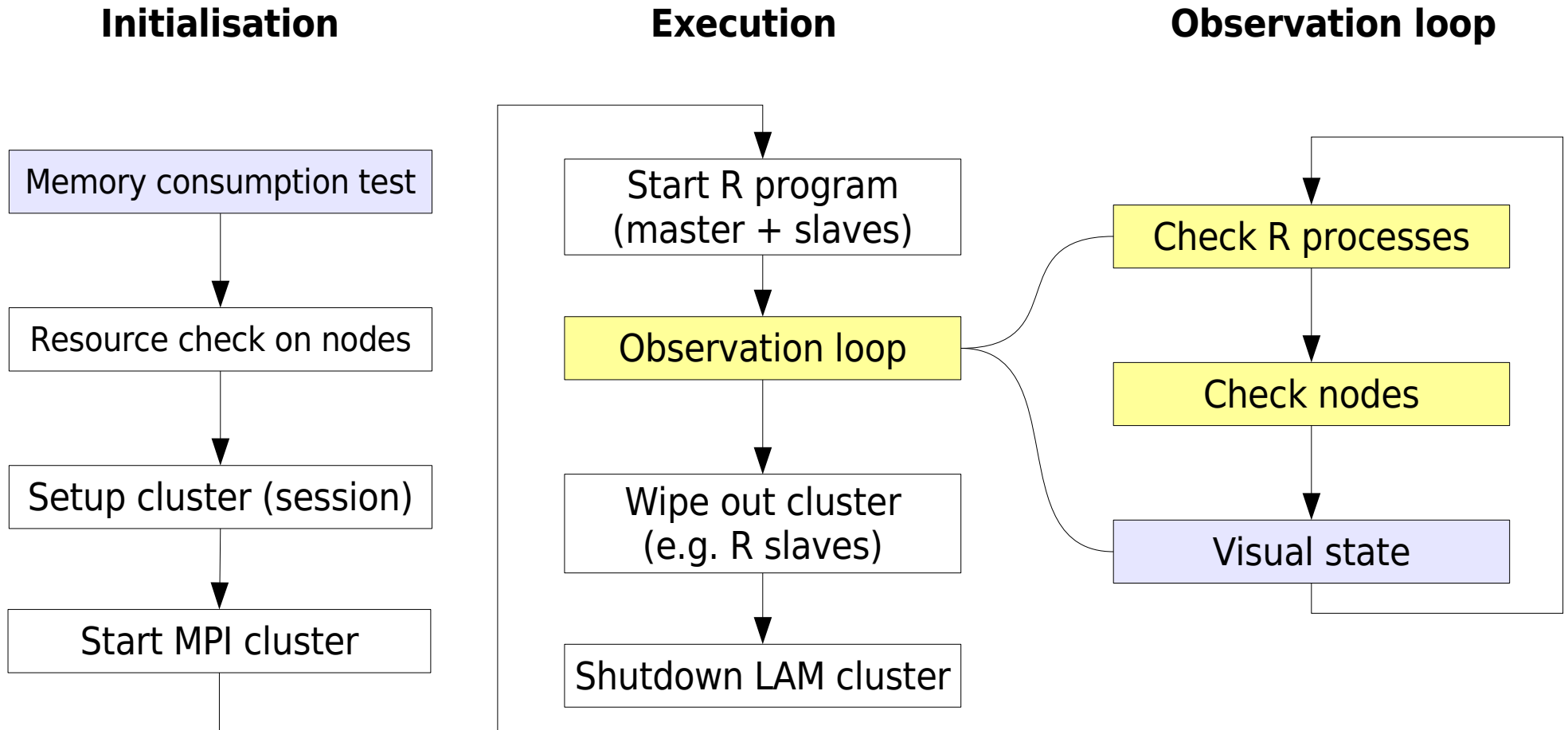- ➢ Function API equivalent to snow – porting is easy.

Simpler functions for common tasks

➢ Loading libraries and sources in the cluster.

➢ Variable handling over the cluster (with exporting and removal).

➢ Additional: parallel call with intermediate result save and restore (results are not lost on single node shutdowns/crashes) – this can also be used for "dynamical" cluster resizing.

sfCluster management tool

➢ Hide cluster handling, setup and shutdown from user.

➢ Implementation as Unix command line tool (written in Perl).

➢ Using only open source tools.

➢ Build upon MPI (currently LAM, OpenMPI in the future).

➢ Automatic resource allocation, depending on current usage of universe. Partly usage of machines is possible.

➢ One LAM cluster per program (means: multiple clusters per user): clusters are independent.

➢ Monitoring the execution of parallel R programs with detection of problems.

sfCluster workflow

Initialisation

```
Memory consumption test
        ↓
Resource check on nodes
        ↓
Setup cluster (session)
        ↓
Start MPI cluster
```

Execution

```
Start R program
(master + slaves)
        ↓
Observation loop
        ↓
Wipe out cluster
(e.g. R slaves)
        ↓
Shutdown LAM cluster
```

Observation loop

```
Check R processes
        ↓
Check nodes
        ↓
Visual state
```

(optional) stop on error   optional step

sfCluster execution modes

Execution modes for running sfCluster

- ➤ batch (-b)   like "R CMD BATCH". *Default*.
- ➤ interactive (-i)   interactive R shell
- ➤ monitor (-m)   batch + debugging informations.

- ➤ sequential (-s):   sequential execution without cluster.

Optionally, these modes can be installed as R addition like "R CMD par", "R CMD parmon" etc.

Example interactive mode

```
jo@biom9:~$ sfCluster -i --cpus=16 --mem=200
Session-ID  : bjrrj9v2_R
biom8.imbi.uni-freiburg.de: 1 CPUs assigned (1 possible).
biom9.imbi.uni-freiburg.de: 1 CPUs assigned (1 possible).
biom10.imbi.uni-freiburg.de: 1 CPUs assigned (1 possible).
knecht5.fdm.uni-freiburg.de: 8 CPUs assigned (8 possible).
knecht4.fdm.uni-freiburg.de: 5 CPUs assigned (8 possible).
ASSIGNED 16 cpus on 5 machines (16 requested).

-- sfCluster: START R-interactive session --

> library(snowfall)
> sfInit()
        16 slaves are spawned successfully. 0 failed.
Startup Lockfile removed: /h/jo/.sfCluster/SFINIT_jo_bjrrj9v2_R_1113_080820
JOB STARTED AT Wed Aug 20 11:14:08 2008 ON biom9 (OSLinux) 2.6.18-6-686-bigmem

R Version:  R version 2.5.1 (2007-06-27)
snowfall 1.43 initialized (parallel=TRUE, CPUs=16)

> q()
Save workspace image? [y/n/c]: n

-- sfCluster: INTERACTIVE session finished. --
LAM/MPI cluster successfully halted
```
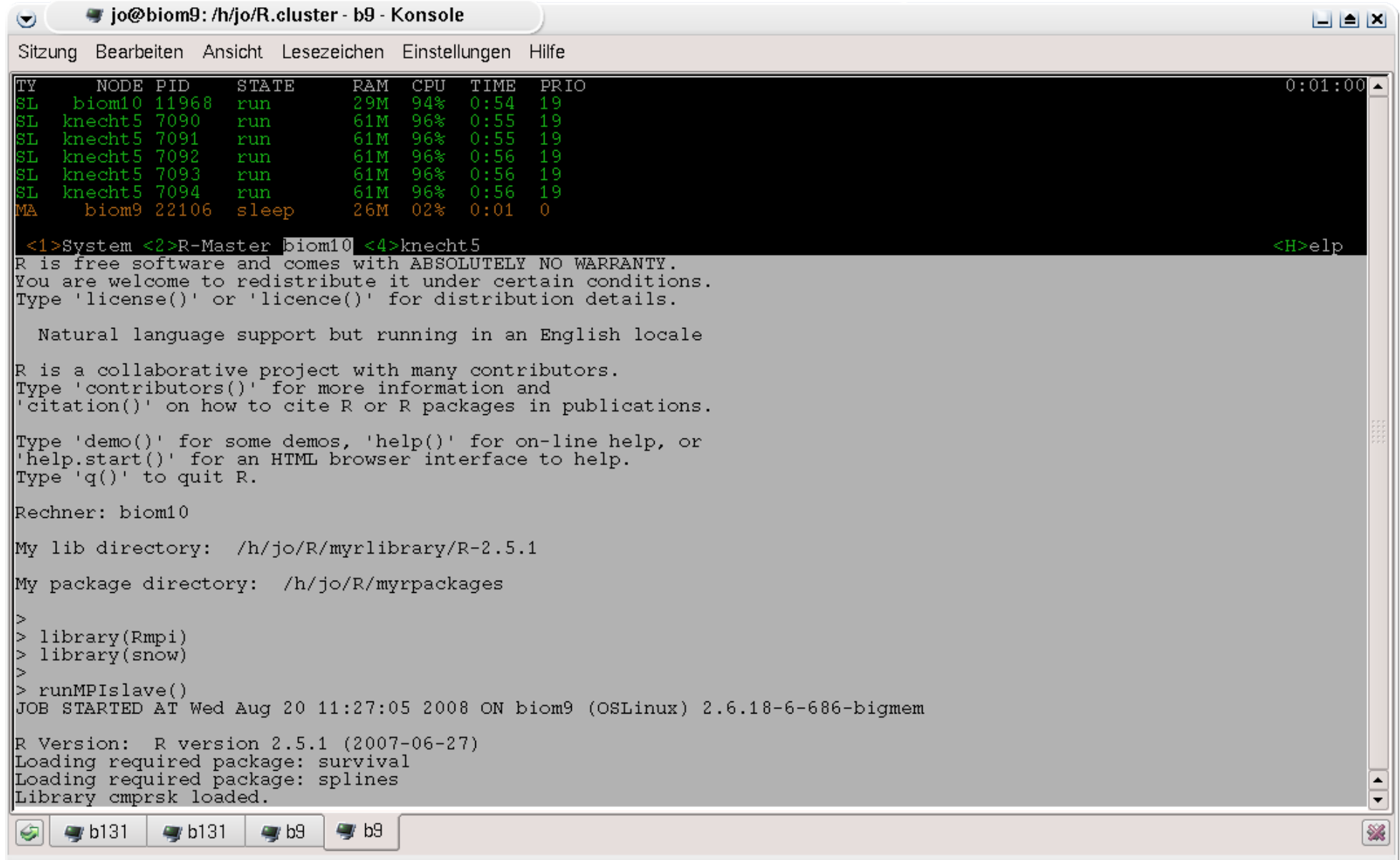
Example screenshot monitoring mode

sfCluster options

- ➢ Request specific number of CPUs.
- ➢ Request specific R version for execution.
- ➢ Send mail at success or failure.
- ➢ Set nice level of all slaves ...

- ➢ *... and many more*

sfCluster administration options

➢ Show current usage of resources in cluster universe (with determination of free resources).

➢ Show current running sessions (per user or all users).

➢ Convenient session shutdown (kill). Can be used by (administration user) *root*.

➢ sfCluster allows the definition of "subuniverses" in the whole cluster universe, which are accessible to specific user groups.

➢ Installation via Tarball or Debian package.

Examples administration

```
jo@biom9:~$ sfCluster -o --all
SESSION            | STATE | USR     | M  | MASTER        #N    RUNTIME R-FILE / R-OUT
-------------------+-------+---------+----+------------------------------------------------
MWhCBAj6_R         | run   | jo      | MO | biom9.imbi    6   0:00:09 boot.R / boot.Rout
4DTqQJWF_R-2.7.1   | run   | arthur  | BA | biom9.imbi   20   1:24:54 simul_pcsh.R / [...]

jo@biom9:~$ sfCluster --universe --mem=0.5G
Assumed memuse: 512M (use '--mem' to change).

Node                           | Max-Load | CPUs | RAM     | Free-Load | Free-RAM | FREE-TOTAL
-------------------------------+----------+------+---------+-----------+----------+-----------
biom8.imbi.uni-freiburg.de     |        5 |    8 | 15.9G   |     1     |   13.6G  |     1
biom9.imbi.uni-freiburg.de     |        7 |    8 | 15.9G   |     1     |   12.4G  |     1
biom10.imbi.uni-freiburg.de    |        8 |    8 | 15.9G   |     1     |   12.4G  |     1
biom11.imbi.uni-freiburg.de    |        2 |    4 |  7.9G   |     0     |    4.6G  |     0
knecht5.fdm.uni-freiburg.de    |        8 |    8 | 15.7G   |     8     |    0.7G  |     1
knecht4.fdm.uni-freiburg.de    |        8 |    8 | 15.7G   |     8     |    3.0G  |     6
knecht3.fdm.uni-freiburg.de    |        8 |    8 | 15.7G   |     7     |    4.3G  |     7
knecht1.fdm.uni-freiburg.de    |        4 |    4 |  7.8G   |     4     |    7.5G  |     4
biom6.imbi.uni-freiburg.de     | no-sched |    4 |  7.9G   |     -     |      -   |     -

Potential usable CPUs: 21

jo@biom9:~$ sfCluster --kill MWhCBAj6_R
Try to "smart" shutdown remote sfCluster (biom9.imbi.uni-freiburg.de, pid 15491)
Waiting for sfCluster to halt: ..... succeeded. Force wipeout remains.
[...]
```

Summary

➢ We have very good experiences running sfCluster/snowfall in our institute for several months now.

➢ Many users run parallel programs without even knowing how to setup clusters.

For more informations visit and download:

http://www.imbi.uni-freiburg.de/parallel

References

R packages: **snow**, **Rmpi**.

Ananth Grama, Anshul Gupta, Vipin Kumar, and George Karypis. *Introduction to Parallel Computing*. Pearson Education, second edition, 2003.

G. Burns, R. Daoud, and J. Vaigl. LAM: *An Open Cluster Environment for MPI*. Technical report, 1994.
http://www.lam-mpi.org/download/files/lam-papers.tar.gz

A. Rossini, L. Tierney, and N. Li. *Simple parallel statistical computing in R*. Journal of Computational and Graphical Statistics, 16(2): 399-420, 2007.