Linux für bwGRiD

Sabine Richling, Heinz Kredel

Universitätsrechenzentrum Heidelberg
Rechenzentrum Universität Mannheim

27. June 2011

Introduction

What is the bwGRiD cluster MA/HD?

- Compute cluster with 2 $\times$ 140 compute nodes
- Each compute node has 8 CPUs
- Operating system: Scientific Linux
- Access via access node `frbw.grid.uni-mannheim.de`
- Allocation of compute nodes via batch system

What do I need to know?

- Basic Linux concepts and commands
- Working with a batch system

Where do I find information on the bwGRiD cluster?
`http://www.uni-mannheim.de/rum/zs/hpc/bwgrid_cluster`

Content

- Introduction to Linux
- Access to Linux servers
- Working with files and directories
- Text editors on the bwGRiD Cluster
- Environment and configurations files
- Introduction to the batch system

Introduction to Linux

http://krum.rz.uni-mannheim.de/linux-kurs-2011/

Access to Linux servers

From Linux/Max

- `ssh` for login
- `scp` for file transfer

From Windows

- `Putty/Xming` for login
- `winscp` for file transfer

For download links see:
http://www.uni-mannheim.de/rum/zs/hpc/bwgrid_cluster/loginserver

Working with files and directories

- Moving within a directory tree: `cd ls`
- Creating/removing directories: `mkdir rmdir`
- Viewing content of text files: `cat more less`
- Changing access rights: `chmod`

Text editors on bwGRiD

- `nano`: simple text editor controlled with control keys
- `emacs`: extensible and customizable text editor controlled with commands or via menu
- `vi`: standard Linux text editor controlled via commands

bash Environment

- Environment variables: `env`
- Setting environment variables: `export echo`
- Setting aliases: `alias`
- Configuration files: `.bashrc`

Introduction to PBS

What is PBS?

- PBS = Portable Batch System
- System to send jobs to distributed computing resources
- Scheduling of jobs (Ablaufsplanung)
- Monitoring of jobs and queues
- bwGRiD: PBS Torque and scheduler Moab

How does PBS work?

- User determines resource requirements for a job and writes a batch script.
- User submits job to PBS with the `qsub` command.
- PBS places the job into a queue based on its resource requests and runs the job when those resources become available.
- The job runs until it either completes or exceeds one of its resource request limits.
- PBS copies the job's output into the directory from which the job was submitted.

How to use PBS on bwGRID?

- Two ways to use compute nodes via PBS:
  - Interactive Batch → for tests and development
  - Submitting Batch Jobs → for productive work
- Compute nodes are exclusively allocated!
- Compute nodes can be used for
  - serial jobs (one core, memory up to 16 GB)
  - parallel jobs with shared memory (up to 8 cores, OpenMP)
  - parallel jobs with distributed memory (up to 64 nodes, MPI)

Interactive Batch

Interactive Batch

Start an interactive job with X11 forwarding for 2 hours
(default 1 hour, maximum 8 hours):

```
qsub -I -X -l walltime=2:00:00
```

If resources are available, you enter a node with a new prompt:

```
[userid@n0xxxxx ~] $
```

Now start working interactively.

Remember:

Always use the interactive batch to develop and test your programs.
Do not start programs on the login servers frbw1, frbw2 or frbw3.

Useful PBS options

- `-I`
  run as an interactive job
- `-I -X`
  run as an interactive job with X11 forwarding
- `-l nodes=N:ppn=M`
  request N nodes with M processors per node (default: nodes=1)
- `-l walltime=N`
  maximum elapsed time in `hh:mm:ss` form (default: 1 hour)
- `-o outfile`
  redirect standard output to `outfile`
- `-e errfile`
  redirect standard error to `errfile`
- `-j oe`
  combine standard output and standard error
- `-N jobname`
  rename the job to `jobname`

Using Software Packages

Setting the environment with the `module` command:

```
module subcommand [module-name]
```

- `module ava`
  shows available software modules
- `module help module-name`
  shows information for the software module
- `module load module-name`
  load a software module
- `module unload module-name`
  unloads a software module
- `module list`
  shows loaded software modules

Statistics/Mathematics Modules

- R (www.r-project.org)
  `math/R/2.13.0`
- Stata (MA only)
  `math/stata/11`
- Matlab (MA only) & Clones
  `math/matlab/R2011a`
  `math/octave/3.0.3`
  `math/scilab/4.1.2`
- Mathematica (MA only)
  `math/mathematica/8.0`
- Maple
  `math/maple/14`

Read `module help` for more information!

Batch Jobs

PBS Job Scripts

- PBS job scripts are regular shell scripts with options for PBS (Lines starting with `#PBS`).
- PBS job scripts always start in your home directory `$HOME`. Use `cd` to change in another directory.

Example scripts in: `/opt/system/moab/6.0.0/examples`

Submitting Batch Jobs

- To submit a job to PBS, use the `qsub` command.
- `qsub` can take all PBS options on the command line as well; specifying these on the command line overrides settings in the job script.
- Notice that `qsub` prints the job number. This number is important for finding output files generated by this run as well as for deleting the job if necessary.

Monitoring

Job Status:

- `qstat`
  shows the status of your jobs
  (more details with `-a` or `-n` or `-f`)
- `checkjob myjobid`
  shows detailed information of job

Node Status:

- `freenodes`
  shows the number of free nodes
- `pbsnodes -a [nodeid]`
  detailed information on the status of all or particular nodes

Batch Queues

`qstat -q`
shows available batch queues:

- `single` – 1 nodes, 120 hours run-time limit
- `normal` – up to 64 nodes, 48 hours run-time limit
- `interactive` – up to 16 nodes, 8 hours run-time limit
- `batch` – for new Jobs before before assignment to single or normal

Do not specify a batch queue in a job script!
The assignment to a queue happens automatically.

Only special queues need the PBS option `-q queuename`

- `itp,testitp` – only for members of the ITP

Deleting Batch Jobs

- `qdel jobid`
  deletes queued or running jobs

Statistics Software in Batch Mode

R Batch Jobs

Read Module Help

```
module help math/R/2.13.0
```

Starting a single R program

- example script: bwgrid-r.pbs
- replace example R-program with your R-program
- only reasonable if your program needs a lot of memory

Starting multiple R programs on one node

- example script: bwgrid-r-multi.pbs + task.sh
- determine number of cores in bwgrid-r-multi.pbs
- edit task.sh to fit your needs

Parallel R packages

Package	Version	Websites	Technology
Computer Cluster			
Rmpi	0.5-6	http://cran.r-project.org/web/packages/Rmpi http://www.stats.uwo.ca/faculty/yu/Rmpi	MPI
rpvm	1.0.2	http://cran.r-project.org/web/packages/rpvm http://www.biostat.umn.edu/~nali/SoftwareListing.html	PVM
nws	1.7.0.0	http://cran.r-project.org/web/packages/nws http://nws-r.sourceforge.net	NWS and socket
snow	0.3-3	http://cran.r-project.org/web/packages/snow http://www.cs.uiowa.edu/~luke/R/cluster	Rmpi, rpvm, nws, socket
snowFT	0.0-2	http://cran.r-project.org/web/packages/snowFT	rpvm, snow
snowfall	1.60	http://cran.r-project.org/web/packages/snowfall http://www.imbi.uni-freiburg.de/parallel	snow
papply	0.1	http://cran.r-project.org/web/packages/papply http://math.acadiau.ca/ACMMaC/software/papply.html	Rmpi
biopara	1.5	http://cran.r-project.org/web/packages/biopara http://hedwig.mgh.harvard.edu/biostatistics/node/20	socket
taskPR	0.31	http://cran.r-project.org/web/packages/taskPR http://users.ece.gatech.edu/~gte810u/Parallel-R	MPI (only LAM/MPI)
Grid Computing			
GridR	0.8.4	http://cran.r-project.org/web/packages/GridR http://www.stefan-rueping.de	web service, ssh, condor, globus
multiR	-	http://e-science.lancs.ac.uk/multiR	3 tier client/server architecture
Biocep-R	NA	http://biocep-distrib.r-forge.r-project.org	java 5
Multi-core System			
pnmath(0)	0.2	http://www.cs.uiowa.edu/~luke/R/experimental	openMP, Pthreads
fork	1.2.1	http://cran.r-project.org/web/packages/fork	Unix: fork
R/Parallel	0.6-20	http://www.rparallel.org	C++, file
romp	0.1a	http://code.google.com/p/romp	openMP

Table 1: Overview about parallel R packages for computer clusters, grid computing and multi-core machines, including the latest version numbers, the corresponding hyperlinks and technologies.

M. Schmidberger et al. 2009 (Tech. Report 47, Department of Statistics, LMU)

Rmpi Package

- Package information: http://www.stats.uwo.ca/faculty/yu/Rmpi/
- Provides interface to MPI functions from R
- Requires installed MPI implementation
- Package includes scripts to launch R instances
- Programmer determines workload of the processes and the communication between the processes
- Rmpi tutorial: http://math.acadiau.ca/ACMMaC/Rmpi/
- Rmpi is integrated in `math/R/2.13.0` (see example)

Stata Batch Jobs

Read Module Help

```
module help math/stata/11
```

Starting a stata program

- example script: bwgrid-stata.pbs
- replace example do-file with your do-file
- submit PBS script

Hints

- `stata-mp` uses shared-memory parallelization
- allocate only one node per job

Matlab Batch Jobs

Read Module Help

```
module help math/matlab/2011a
```

First try the Matlab Compiler

- Compile your program BEFORE job submission
- Start the compiled binary in your job script

Only if your program does not compile or when you use the PCT

- allocate only one node per job
- use the PBS software option
- start m-file directly in job script
- you can start multiple Matlab programs on one node
- do not submit multiple Matlab jobs at a time

Parallelization with Matlab

- Starting up to 8 programs on a node (with `pbsdsh`)
- Automatic Multithreading in Matlab
- Shared memory computing with Parallel Computing Toolbox
- Distributed computing with Distributed Computing Server
  Additional licenses necessary!

Multithreading in Matlab

Automatic Multithreading:

- Matlab runs computations on multiple threads
- No changes to Matlab code required
- Users can change behavior via preference
- Maximum gain in element-wise operations and BLAS routines

On bwGRiD:

- Default: Automatic. Use as many threads as cores: 8
- Change number of used threads with: `numThreads=2`

Matlab Parallel Computing Toolbox

Basic commands

- `matlabpool open 2;` – start Matlab workers (max. 8)
- `matlabpool close;` – stop all Matlab workers

Parallel Tools

- use parallel Matlab functions
- `parfor` – parallel loops
- `spmd` – for working with distributed arrays
- `pmode` – interactive multithreading

Mathematica Batch Jobs

Read Module Help

```
module help math/mathematica/8.0
```

Starting a single Mathematica program

- example script: bwgrid-math.pbs
- replace example Mathematica-program with your program

Hints

- allocate only one node per job
- you can use up to 4 subprocesses in your program
- do not start multiple Mathematica programs in one job
- do not submit multiple Mathematica jobs at a time