

Unix HPC Environment – Linux and Solaris

Samuel Sarholz
sarholz@rz.rwth-aachen.de

Center for Computing and Communication
RWTH Aachen University

Agenda

- o HPC @ RZ
- o Cluster Overview
- o Filesystems
- o Software
- o Batch System
- o IDEs

2

The RZ Compute Cluster

Vendor	Machine Type	Machine Count	Processor	Sockets	Cores / Threads
Sun	X4450	10	Xeon	4	24
IBM	LS42 Blade	3	Opteron	4	16
Fujitsu Siemens	Rx200	50	Xeon	2	8
Fujitsu Siemens	Rx600	2	Xeon	4	16
Sun	T5120	20	US T2	1	8 / 64
Sun	X4600	2	Opteron	8	12
Sun	V40z	4	Opteron	4	8
Sun	V40z	64	Opteron	4	4

3 o About 12 TFlop/s peak performance

Integrative Hosting

- o Institutes buy the hardware
- o The RZ manages the OS and software in the same manners as the RZ cluster
 - Linux, Solaris, (Windows)
 - File systems are shared
- o Resources preferably will be shared
 - Institutes can use more nodes than they own in peak
 - On average they get their share
 - E.g. the Institute for Scientific Computing cluster
- o Login and batch jobs may be restricted to certain user groups

Some Hosted Machines

Vendor	Machine Type	Machine Count	Processor	Sockets	Cores
IBM	LS42 Blade	72	Opteron	4	16
Fujitsu Siemens	Rx200	232	Xeon	2	8
Sun	X2200	36	Opteron	2	4
Sun	X4100	18	Opteron	2	4
Dell	1950	8	Xeon	2	4
Sun	V20z	10	Opteron	2	2

- Hosted compute power about 4x more than the RZ cluster



Some institutes we host for

- o Lehr- und Forschungsgebiet Theoretische Chemie ([IPC](#)), FB 1
Prof. Arne Lüchow
- o Lehrstuhl für Mathematik ([CCES](#)), FB 1
Prof. Dieter Bothe
- o Lehrstuhl für Theoretische Physik C und Institut für Theoretische Physik ([ThPhC](#)), FB 1
Prof. Ulrich Schollwöck
- o Lehrstuhl für Werkstoffchemie ([MCh](#)), FB 1
Prof. Jochen Schneider
- o Lehrstuhl für Informatik 12 (Hochleistungsrechnen) ([SC](#)), FB 1
Prof. Christian Bischof
- o Ultra High Speed Mobile Information and Communication ([UMIC](#))
Exzellenzcluster
- o Aachen Institute for advanced study in Computational Engineering Science ([AICES](#))
Graduiertenschule im Rahmen der Exzellenzinitiative
- o Tailor-Made Fuels from Biomass ([TMFB](#))
Exzellenzcluster

6

Center for

Computing and

Communication

HPC@ RZ

Cluster

Filesystems

Software

Batch

IDEs

Agenda

- o HPC @ RZ
- o Cluster Overview
- o Filesystems
- o Software
- o Batch System
- o IDEs

7

Login (1/2)

o Frontend nodes for login

- Unix Frontends (ssh protocol) (or NX* on Linux)

cluster.rz.RWTH-Aachen.DE (Sparc-Solaris)

cluster-solaris.rz.. (Sparc-Solaris)

cluster-solaris-sparc.rz... (cl-sol-s)

cluster-solaris-opteron.rz... (cl-sol-o)

cluster-linux.rz... (Xeon-Linux)

cluster-linux-xeon.rz... (cl-lin-x)

cluster-linux-operon.rz... (cl-lin-o)

8

* You can get an NX client from www.nomachine.com

Login (1/2)

- o NX (graphical)
 - www.nomachine.com
- o ssh (command line)
 - Comes by default with Linux distributions
 - `ssh [-Y] cluster.rz.rwth-aachen.de`
 - Get Windows clients from
 - Putty: <http://www.chiark.greenend.org.uk/~sgtatham/putty/>
 - ssh for Windows: <ftp://ftp.cert.dfn.de/pub/tools/net/ssh>

Usage Concept

- o Interactive frontend systems are meant for short test runs, GUIs and debugging.
 - Compute jobs should be submitted to the batch system
 - On Linux we offer a few nodes for MPI tests and debugging to avoid overloaded frontends
 - mpiexec determines the hosts with the lowest load to run on
- o The batch system manages the distribution of compute jobs
 - Makes sure systems are not overloaded to guarantee good performance
 - On Unix the Sun Grid Engine (SGE) is employed

10

Agenda

- o HPC @ RZ
- o Cluster Overview
- o Filesystems
- o Software
- o Batch System
- o IDEs

11

File Storage Strategies

- o Snapshots in home are in .snapshot/.snapshot
 - The naming and the location has changed recently
 - Naming scheme: hourly.X , nightly.X, weekly.X
 - Snapshot date is saved in the access time
 - Use `ls -ltr` to list them in chronological order

Agenda

- o HPC @ RZ
- o Cluster Overview
- o Filesystems
- o Software
- o Batch System
- o IDEs

13

Programming Tools (1/2)

- o Compiler^{*}
 - Intel
 - GNU
 - Sun
- o MPI^{*}
 - Intel
 - Sun
- o Debugger
 - TotalView

^{*} We have others installed as well but there are the recommended

Programming Tools (2/2)

- o Correctness tools
 - Sun Thread Analyzer
 - Intel Thread Checker
 - Marmot
- o Performance analysis
 - Intel VTune
 - Sun Analyzer
 - Vampir
 - Scalasca
 - Acumem

15

Software Packages

FEM

- o Abaqus
- o Ansys
- o Hyperworks
- o Ls-Dyna
- o Marc/Mentat
- o Nastran/Patran

Chemie

- o Gaussian
- o Turbomole
- o Vasp
- o abinit
- o meep

Site license

Parallel version available

No site license

CFD

- o CFX/TASCFlow
- o Fluent
- o ICEM-CFD
- o StarCD
- o gerris

Mathematik

- o Maple
- o Mathematica

Misc

- o Matlab/Simulink
- o Tecplot
- o Siesta

16

Center for

Computing and

Communication

HPC@ RZ

Cluster

Filesystems

Software

Batch

IDEs

Agenda

- o HPC @ RZ
- o Cluster Overview
- o Filesystems
- o Software
- o Batch System
- o IDEs

17

Batch System (1/4)

- o We employ the Sun Grid Engine (SGE)
- o It manages the scheduling and distribution of compute jobs
 - Makes sure systems are not overloaded
- o Jobs have to request resources
 - or a small default value will be chosen
- o A batch job is a shell script
 - Resources can requested in the batch script or while submitting
- o find more detailed infos and examples on our web page
 - www.rz.rwth-aachen.de/ca/k/otd/ (only in German)
 - our user's guide (Primer)

18

Batch System (2/4)

o Job management commands

- qsub : submits a batch job
 - qsub batch.job
 - qsub -l ostype=linux batch.job
 - -P for RZ projects
- qdel : delete jobs
 - qdel <job-ID>
- qstat : show status of you jobs
- qmon : GUI for managing jobs

Batchsystem (3/4)

o Resource parameters and options to qsub

- Output file -o \$HOME/batchout.txt
- Runtime -l h_rt=01:30:27
- Memory requirement -l h_vmem=1.5GB #per process!
- Operating System -l ostype=linux | solaris
- Platform -l platform=x86 | sparc
- HW Architecture -l march=intel64|amd64|ultraT2
- Number of Threads -l num_procs=4
- Number of MPI processes -pe mpi* 16
 - mpi*4tasks 16 for bundeling MPI processes to nodes
- Licenses -l ISVsoftware=1
- Email notifications -M beas, -m e@mail.de

20

Batch System (4/4)

Example batch script:

```
#$ -N Hybrid-Test-Job
#$ -l h_rt=00:15:00
#$ -l h_vmem=500M
#$ -l num_proc=4
#$ -l ostype=sunos
#$ -pe mpi* 5
#$ -l platform=x86
$PSRC/pex/462 # build the example program a.out
cd 462
$MPIEXEC $FLAGS_MPI_BATCH a.out
```

- o qsub options can be specified in a batch script
 - the line has to begin with “#\$”
 - Caution: commenting \$VAR out may cause an error;
add a space: # \$MPIEXEC

21

Agenda

- o HPC @ RZ
- o Cluster Overview
- o Filesystems
- o Software
- o Batch System
- o IDEs

22

Program development Tools

- o IDEs: eclipse, sunstudio, kdevelop
 - All have some (dis)advantages
 - So we don't recommend one
- o non-integrated tools for program development
 - Text editor (kate, gedit, nedit, gvim)
 - Compiler
 - Use modules to choose one
 - There is a default one available
 - make

Modules

- o Setup the environment for a software package
 - e.g. set PATH and other needed environment variables
- o `module help`
 - `list` : show loaded modules
 - `load`, `unload` : add / remove software to the environment
 - `avail` : show available modules (long list)
 - structured in categories, `unload` to shorten the list
 - `switch` : exchange a module
 - all others will be reloaded so that dependencies work correctly
 - `reload` : load all modules again, may fix the environment
- o modules can depend on other modules
 - e.g. MPI needs a compiler

24

Module environment

o Compiler and MPI environment variables

- \$CC \$CXX \$FC
- \$MPICC \$MPICXX \$MPIFC \$MPIEXEC
- \$FLAGS_FAST \$FLAGS_FAST_NO_FPOPT
- \$FLAGS_LPATH \$FLAGS_RPATH
- \$FLAGS_OPENMP \$FLAGS_AUTOPAR
- \$FLAGS_DEBUG
- \$FLAGS_MATH_LINKER \$FLAGS_MATH_INCLUDE
- \$PATH \$LD_LIBRARY_PATH

Compiling and Linking

- o Compiling source
 - `$CC hello.c`
- o Building objects
 - `$CC hello.c -c hello.o`
- o Linking objects to executables
 - `$CC hello.o -o hello.exe [-l<library>]`

GNU make

- o Make is a tool for software-project management
- o Determines which pieces needs to be rebuild
- o Dependencies are defined in a makefile
- o Default filenames: Makefile, GNUmakefile, makefile
- o Allows parallel builds: `make -j 16`

Makefile (1/2)

o Rules:

```
target  : prerequisite list  
    < tab > command  
    < tab > next command
```

- o Target usually refers to an output file
- o Prerequisites : other rules or files (e.g. source code)
- o Command: action to be performed
- o Phony Targets don't refer to a file

```
.PHONY: clean  
clean:  
    < tab > rm -rf *.o  
    core*
```

Makefile (2/2)

o Variables:

```
OBJ = main.o first.o second.o  
all: $(OBJ)  
    < tab > cc -o myprog.exe  $(OBJ)
```

o Automatic Variables:

```
$< - first prerequisite  
$@ - target of the rule
```

o Defining implicit rules: Pattern Rules

```
%.o : %.c  
    < tab > $(CC) -c $(CFLAGS) $< -o $@
```

o Note: on Solaris: make is not gmake

Starting Parallel Programs

- o OpenMP (multi threading)
 - export OMP_NUM_THREADS=*n*
 - start the program
- o MPI (message passing)
 - mpiexec -np *X program*
 - RZ wrapper
 - -m 3
 - --show
- o We offer several example programs in \$PSRC
 - gmake run build and execute them

30

The End

Thank you for your attention!

Questions?

Links and Support

- o **Information related to HPC at the RWTH Aachen**
<http://www.rz.rwth-aachen.de/hpc/>
- o **RWTH Compute Cluster User's Guide**
<http://www.rz.rwth-aachen.de/hpc/primer>
- o **Usage of the Windows HPC systems**
<http://www.rz.rwth-aachen.de/hpc/win>
- o **Information on Hosting**
<http://www.rz.rwth-aachen.de/ca/c/pgo>
- o **Questions or Problems?**
Just send an e-mail to our trouble ticket system:
hpc@rz.rwth-aachen.de

32