GETTING STARTED WITH HIGH PERFORMANCE COMPUTING IN RESEARCH

Richard Polzin <rpolzin@ukaachen.de>

LAST UPDATED: 24. JULY 2024

TABLE OF CONTENT

TABLE OF CONTENT

1. The What, Why, and How of HPC

TABLE OF CONTENT

1. The What, Why, and How of HPC
2. Linux Crashcourse

TABLE OF CONTENT

1. The What, Why, and How of HPC
2. Linux Crashcourse
3. RWTH Compute Cluster

TABLE OF CONTENT

1. The What, Why, and How of HPC
2. Linux Crashcourse
3. RWTH Compute Cluster
4. Tips and Tricks

TABLE OF CONTENT

1. The What, Why, and How of HPC
2. Linux Crashcourse
3. RWTH Compute Cluster
4. Tips and Tricks
5. SLURM Job Manager

TABLE OF CONTENT

1. The What, Why, and How of HPC
2. Linux Crashcourse
3. RWTH Compute Cluster
4. Tips and Tricks
5. SLURM Job Manager
6. Compute Time Application

THE WHAT, ~~WHY, AND HOW~~ OF HPC

- **What** is HPC?
  - **H**igh **P**erformance **C**omputing: Systems thousands of times faster than your laptop
  - Large bandwidth, store hundreds of GB in RAM, Massive parallelization, ...
  - Usually running some form of Linux

THE ~~WHAT~~, WHY, ~~AND HOW~~ OF HPC

- **Why** would you want HPC?
  - Basically: If you could profit from your programs running faster or with more data
  - Highly Parallelizable, runs "in the background"
  - 96 CPU cores per node, 1.5TB SSDs, up to 1TB RAM, 4 GPUs ...

THE ~~WHAT, WHY, AND~~ HOW OF HPC

- **How** do you get access to HPC?
  - All RWTH-affiliates are granted access to some extend
  - NHR-Verbund offers free compute resources for scientists working at universities in Germany

HPC.NRW

INTRODUCTION TO LINUX

(in an HPC context)

Version 20.09 | HPC.NRW Competence Network

CC BY SA

THE COMPETENCE NETWORK FOR HIGH-PERFORMANCE COMPUTING IN NRW.

RWTH COMPUTE CLUSTER

ACCOUNT CREATION

Use Selfservice to create your account (https://idm.rwth-aachen.de/selfservice/)

1. Accounts and Passwords
2. Account Overview
3. Create Account

RWTH COMPUTE CLUSTER

LOGIN

```
ssh ab1234@login23-1.hpc-itc.rwth-aachen.de
```

- Login with secure shell (ssh): **ssh tim@host**

Different nodes available for different use cases

Login Node	Recommended Usage	Remarks	Hardware Specifications
`login23-1.hpc.itc.rwth-aachen.de`	General use	Same architecture and software stack as computing nodes	2x Intel Xeon 8468 540 GB memory
`login23-2.hpc.itc.rwth-aachen.de`	General use Development		
`login23-3.hpc.itc.rwth-aachen.de`			
`login23-4.hpc.itc.rwth-aachen.de`			
`login23-x-1.hpc.itc.rwth-aachen.de`	General use Remote desktop	Runs FastX webserver Architecture and Softwarestack is **different** to computing nodes	2 x Intel Xeon 8360Y 540 GB memory
`login23-x-2.hpc.itc.rwth-aachen.de`			
`login23-g-1.hpc.itc.rwth-aachen.de`	Development Testing	Same architecture and software stack as computing nodes Has GPUs	2x Intel Xeon 8468 540 GB memory 4x NVIDIA H100
`copy23-1.hpc.itc.rwth-aachen.de`	File transfer	High network speed No x-server No job management No module system	2 x Intel Xeon 8360Y 134 GB memory 2 x 100 GBit/s network
`copy23-2.hpc.itc.rwth-aachen.de`			

RWTH COMPUTE CLUSTER

FILE SYSTEMS

File System	Type	Path	Persistence	Snapshots	Backup	Quota (space)	Quota (#files)	Use Cases
$HOME	NFS/CIFS	/home/<user-name>	permanent	$HOME_SNAPSHOT	yes	150 GB	-	Source code, configuration files, important results
$WORK	NFS/CIFS	/work/<user-name>	permanent	$WORK_SNAPSHOT	no	250 GB	-	Many small working files
$HPCWORK	Lustre	/hpcwork/<user-name>	permanent	-	no	1000 GB	1 million	I/O intensive computing jobs, large files

- $HOME for code/configuration, $WORK for output and working data, $HPCWORK for large data

RWTH COMPUTE CLUSTER

PROJECT MANAGEMENT

- Groups are created for every project
- Every group consists of owners (PC/PI), Managers and Members
- Granted computation time and storage space is shared through groups
- In Aachen project storage is deleted **8 months** after a project's conclusion. Make sure you migrate the data by then!

RWTH COMPUTE CLUSTER

PROJECT MANAGEMENT

Users can be added to and removed from groups/ projects using their TIM-ID.

```
member add --name <project-id> <user-id>
member delete --name <project-id> <user-id>
member finger # view group affiliations
```

- append **--manager** to any command above to assign or revoke the manager role

TIPS AND TRICKS

This is a collection of tips and tricks that make working with the cluster easier and more convenient.

TIPS AND TRICKS

1. You can mount the remote cluster file system to your local machine

TIPS AND TRICKS

1. You can mount the remote cluster file system to your local machine

```
sshfs ab1234@copy23node:/home/ab1234 /mnt/clusterhome
```

TIPS AND TRICKS

1. You can mount the remote cluster file system to your local machine

```
sshfs ab1234@copy23node:/home/ab1234 /mnt/clusterhome
```

unmount with:

```
sudo umount -l /mnt/clusterhome
```

TIPS AND TRICKS

2. If you are outside of eduroam, you can connect through VPN and access the cluster *everywhere*

Go to http://help.itc.rwth-aachen.de
and search for "VPN"

SLURM JOB MANAGER

Simple **L**inux **U**tility for **R**esource **M**anagement

SLURM JOB MANAGER

Simple **L**inux **U**tility for **R**esource **M**anagement

- Job scheduler often used in supercomputers and compute clusters

SLURM JOB MANAGER

Simple **L**inux **U**tility for **R**esource **M**anagement

- Job scheduler often used in supercomputers and compute clusters
- Provides many advantages for utilizing HPC hardware with many users, such as..
  - ... Accounting, Containerization, Priorities, Chain- and Array-Jobs, ...

SLURM JOB MANAGER

SLURM JOB MANAGER

- Users can interact with SLURM from login nodes

SLURM JOB MANAGER

- Users can interact with SLURM from login nodes
- Users may request cores, memory and time, then send their programs to be queued

SLURM JOB MANAGER

- Users can interact with SLURM from login nodes
- Users may request cores, memory and time, then send their programs to be queued
- SLURM reserves these resources and waits till they are available

SLURM JOB MANAGER

- Users can interact with SLURM from login nodes
- Users may request cores, memory and time, then send their programs to be queued
- SLURM reserves these resources and waits till they are available
- Once available, the code will then be run

SLURM JOB MANAGER

SLURM JOB MANAGER

SLURM is fed **jobscripts**, which contain all information the scheduler needs to run a program

These jobscripts consist of three parts:

SLURM JOB MANAGER

SLURM is fed **jobscripts**, which contain all information the scheduler needs to run a program

These jobscripts consist of three parts:

1. Shebang

SLURM JOB MANAGER

SLURM is fed **jobscripts**, which contain all information the scheduler needs to run a program

These jobscripts consist of three parts:

1. Shebang
2. Job Parameters

SLURM JOB MANAGER

SLURM is fed **jobscripts**, which contain all information the scheduler needs to run a program

These jobscripts consist of three parts:

1. Shebang
2. Job Parameters
3. Actual Job Code

SLURM JOB MANAGER

EXAMPLE

SLURM JOB MANAGER

EXAMPLE

```
#!/usr/bin/zsh
```

SLURM JOB MANAGER

EXAMPLE

```
#!/usr/bin/zsh

### Job Parameters
#SBATCH --cpus-per-task=8

#SBATCH --time=00:15:00
#SBATCH --job-name=example_job
#SBATCH --output=stdout.txt
#SBATCH --account=<project-id>
```

SLURM JOB MANAGER

EXAMPLE

```
#!/usr/bin/zsh

### Job Parameters
#SBATCH --cpus-per-task=8

#SBATCH --time=00:15:00
#SBATCH --job-name=example_job
#SBATCH --output=stdout.txt
#SBATCH --account=<project-id>

### Program Code
echo "Hello SLURM"
```

SLURM JOB MANAGER

EXAMPLE

- Safe the file

SLURM JOB MANAGER

EXAMPLE

- Safe the file
- Submit the job

SLURM JOB MANAGER

EXAMPLE

- Safe the file
- Submit the job

```
> sbatch testjob.sh
```

SLURM JOB MANAGER

EXAMPLE

- Safe the file
- Submit the job

```
> sbatch testjob.sh
```

- Check its state

SLURM JOB MANAGER

EXAMPLE

- Safe the file
- Submit the job

```
> sbatch testjob.sh
```

- Check its state

```
> squeue --me
JOBID     PARTITION  NAME        USER      ST    TIME  NODES
12345678  c23ms      example_job AB123456  R     0:02      1
```

SLURM JOB MANAGER

COMMON PARAMETERS

- Number of cores: **-c /--cpus-per-task <numpcus>**
- Memory: **-m /--mem=<amount>G**
- Human readable Job name: **-J /--job-name**
- Reporting File: **-o /--output=<filename>**
- Runtime: **-t /--time=d-hh:mm:ss**
- Account: **-A /--account=<project>**
- GPUs: **--gres=gpu:<type>:<amount>**

SLURM JOB MANAGER

COMMANDS IN A NUTSHELL

Submit jobs

```
> sbatch <BATCH_SCRIPT> [ADDITIONAL ARGUMENTS]
```

SLURM JOB MANAGER

COMMANDS IN A NUTSHELL

Display Job Queue

```
> squeue [OPTIONS]
```

--me shows only your jobs
--start shows the estimated starttime
--format can be used to filter e.g. for GPU jobs

SLURM JOB MANAGER

COMMANDS IN A NUTSHELL

Cancel Jobs

```
> scancel [OPTIONS] <JOB-ID>
```

--me cancel all of your jobs

--v provide details of the cancellation process

SLURM JOB MANAGER

COMMANDS IN A NUTSHELL

Request an interactive job

```
> salloc [OPTIONS]
```

Job parameters just like in the jobscript. Redirects the shell to the head node. Job ends when the shell terminates.

SLURM JOB MANAGER

COMMANDS IN A NUTSHELL

Request an interactive job

```
> salloc [OPTIONS]
```

Job parameters just like in the jobscript. Redirects the shell to the head node. Job ends when the shell terminates.

```
salloc --gres=gpu:1 -n 24 -t 1:00:00 # 24c+GPU node for 1 hour
```

SLURM JOB MANAGER

COMMANDS IN A NUTSHELL

Display Accounting Information

```
> sacct [OPTIONS]
```

Print details abound pending, running, or past jobs.

```
> sacct -S $(date -I --date="yesterday")
```

Would show all jobs submitted since yesterday.

SLURM JOB MANAGER

COMMANDS IN A NUTSHELL

```
> r_wlm_usage
```

Command for only the RWTH HPC to display available / used resources against the account / group quota

SLURM JOB MANAGER

Monitoring tools are improving constantly, with *perfmon* currently being established across compute clusters

SLURM JOB MANAGER

General / Job Overview SingleHost

2024-03-27 10:23:00 to 2024-03-27 18:23:00

JobID 44628605 (jupyterhub) | Host i23m0005 | Reset time | Joblist | Multi Host

Metadata

Project	Partition	Cores allocated	State
supp0006	c23test	16	TIMEOUT

Start	End	Runtime	Timelimit
2024-03-27 10:23:29	2024-03-27 18:23:54	08:00:25	08:00:00

CPU Average Power	CPU Total Energy	GPU Average Power	GPU Total Energy
405 W	3.24 kWh	No data	No data

General

CPU

Memory

Hardware

FLOP/s (DP)

FLOP/s (SP)

FLOP/s (HP)

Memory Bandwidth

IO

Fabric | Lustre | NFS

COMPUTE TIME APPLICATION

COMPUTE TIME APPLICATION

- Resources are measured in core hours (core-h)

COMPUTE TIME APPLICATION

- Resources are measured in core hours (core-h)
  - A Macbook Pro has roughly 70k core-h/year

COMPUTE TIME APPLICATION

- Resources are measured in core hours (core-h)
  - A Macbook Pro has roughly 70k core-h/year
- The smallest project already guarantees 360k core-h

COMPUTE TIME APPLICATION

- Resources are measured in core hours (core-h)
  - A Macbook Pro has roughly 70k core-h/year
- The smallest project already guarantees 360k core-h
- Estimate your needs based on previous work

COMPUTE TIME APPLICATION

- Resources are measured in core hours (core-h)
  - A Macbook Pro has roughly 70k core-h/year
- The smallest project already guarantees 360k core-h
- Estimate your needs based on previous work
- Write and submit your application!

COMPUTE TIME APPLICATION

HPC resources in Germany are arranged hierarchically in the HPC Performance Pyramid.

- **Tier-0**: PRACE and EuroHPC
- **Tier-1**: Gauss Center for Supercomputing (GCS, JSC, HLRS, LRZ)
- **Tier-2**: HPC Centres with supra-regional tasks and thematically dedicated HPC/centres
- **Tier-3**: Regional HPC centres

COMPUTE TIME APPLICATION

Tier 0/1
Tier 2
Tier 3
Institute Clusters,
etc.

COMPUTE TIME APPLICATION

You want to bring your research to another level?
Use our computing time!

And how?

Easy access: NHR starter projects

- For applicants without a deeper background knowledge of High Performance Computing
- Your application will be assigned to one of the NHR centers, from where you will receive advice, and limited HOC resources for use of one year.

Try it out: PREP-projects

- Maximum duration of six months
- For porting applications or scalability analysis
- Continuous application possible

Continuous: Normal projects

- Up to 12 Mio. core-h per year
- GPUs in Aachen: max. 2 Mio. core-h
- GPUs in Darmstadt: max. 1 Mio. core-h
- Continuous application possible

Think big: Large projects

- From 12 Mio. core-h per year up to 50 Mio. core-h per year
- GPUs in Aachen: max. 4 Mio. core-h
- GPUs in Darmstadt: max. 2 Mio. core-h
- Application possible four times a year.

Large project application deadlines:

project start January 1: deadline October 1 (of the year before)	project start April 1: deadline January 1 (same year)	project start July 1: deadline April 1 (same year)	project start October 1: deadline July 1 (same year)
OCT 1st → JAN 1st	JAN 1st → APR 1st	APR 1st → JUL 1st	JUL 1st → OCT 1st
Deadline → Start	Deadline → Start	Deadline → Start	Deadline → Start

NHR Normal

New application
(< 8 Mio. Core-h)
Technical Review
(< 2 weeks)
Formal reject:
Resubmission of
improved application
recommended
Application technical approved
Significant technical issues
Provisioning of start contingent
(0.24 Mio Core-h p. a. / 0.02 Mio Core-h p. m.)
User will be informed about start contingent
(w.r.t planned the project start)
Scientific Review (< 5 weeks)
2 scientific reviews >0
Only 1 sci review after 7 weeks >0
At least one reject
OR
No answer from reviewer
Conditional Allocation
(Minimum of reviews)
Conditional Allocation
(Minimum of reviews)
Provisioning of conditional contingent
User will be informed about conditional
contingent
RAB decision
Final resource allocation
User will be informed about final contingent

ACKNOWLEDGEMENT (RWTH/JARA)

Computations were performed with computing resources granted by RWTH Aachen University under project .

ACKNOWLEDGEMENT (RWTH/JARA)

The authors gratefully acknowledge the computing time provided to them at the NHR Center NHR4CES at RWTH Aachen University (project number <your-project-id: p0020XXX>). This is funded by the Federal Ministry of Education and Research, and the state governments participating on the basis of the resolutions of the GWK for national high performance computing at universities (www.nhr-verein.de/unsere-partner).