NHR FAU

FAU Friedrich-Alexander-Universität Erlangen-Nürnberg

HPC Café for Humanities and Social Sciences

Armine Garibyan

Big Data Linguistics, Liaison Scientist at NHR@FAU

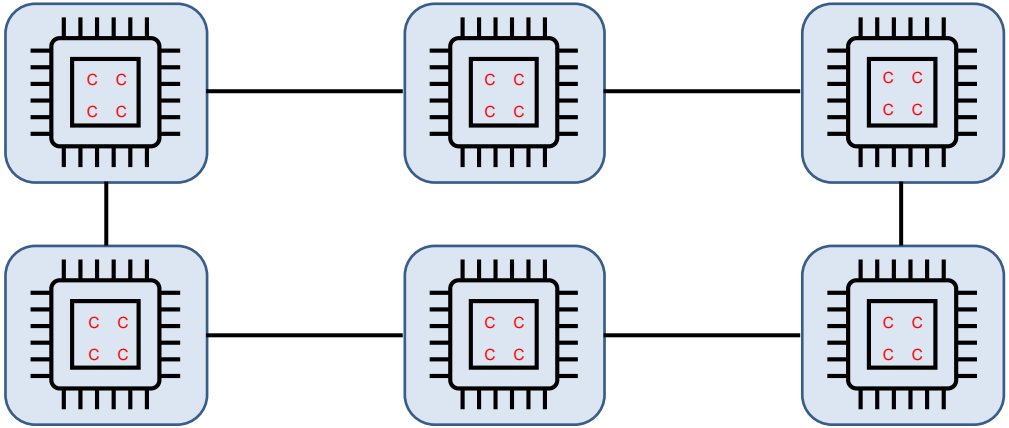
Poll

- Has your R (or other statistics software) ever crashed while running a heavy model or computation?
- Have you ever been afraid to close your laptop because some computation was still running?
- Have you ever waited forever for a model or another computation to finish and wished it were faster?

High-Performance Computing (HPC)

- Use of powerful computational systems to perform calculations and data processing at speeds and scales far beyond the capabilities of a typical desktop computer or a laptop
- Not one giant computer, but a cluster of many computers (= nodes)
- Computers (= nodes) > processors > cores

  → Each core can run a certain task at the same time as other cores.



Why HPC?

Example task: Automatic search for a certain gesture in a video corpus (36,296 hours) 30 Tb

- Search time for 1 hour of video: ~ 2 min
- Total search time for the whole corpus: ~ 72,592 min 50 days

- 1 core available: Video 1 → Video 2 → Video 3 → … → Video 36,296

- 4 cores available:
  - Core 1: Video 1 → Video 5 → …
  - Core 2: Video 2 → Video 6 → …
  - Core 3: Video 3 → Video 7 → …
  - Core 4: Video 4 → Video 8 → …

12 days

NHR FAU

FAU Friedrich-Alexander-Universität Erlangen-Nürnberg

Getting an HPC account

https://doc.nhr.fau.de/account/

How to get an account

- Depends on the status
  - account(s) for a lecturer and students for a seminar
  - FAU staff and students (publicly funded research, theses)
  - BayernKI and NHR users

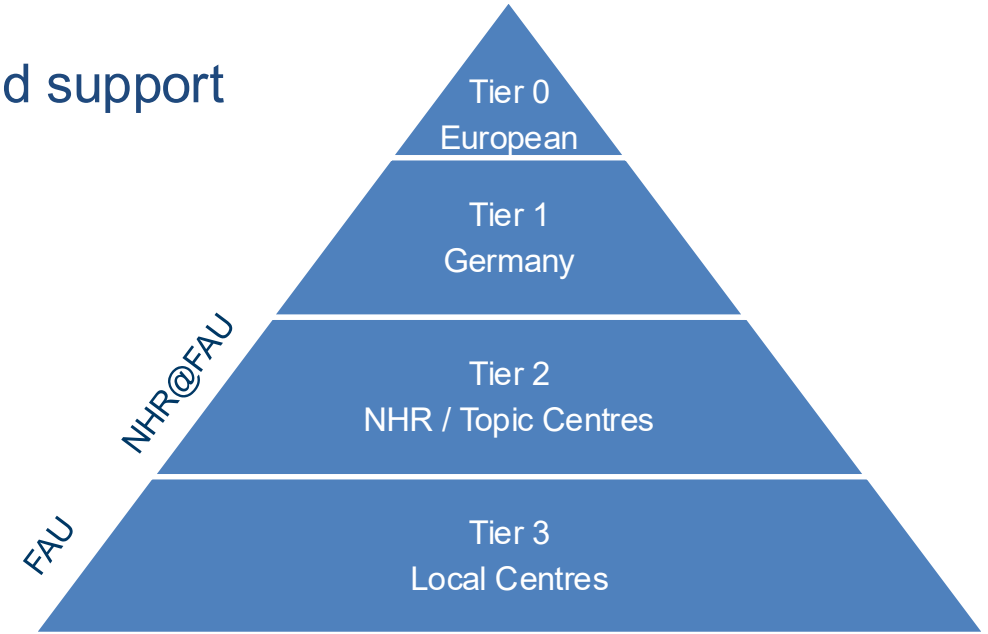
FAU service area “Tier3 Grundversorgung” (basic supply)

- invitation through the chair's contact person / PI in the HPC-Portal
- free of charge for publicly funded research and theses
- If your institution is represented in the HPC-Portal, the process is quick. Otherwise, the head of the institution needs to apply for a Tier3 basic project (a few days).

HPC Tier Pyramid

NHR (Nationales Hochleistungsrechnen) Alliance

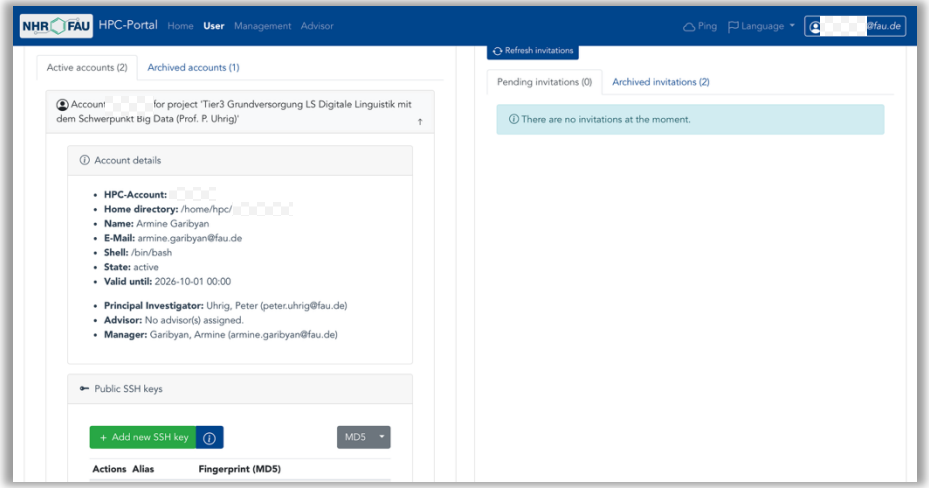
- Coordinated network of NHR centres funded by federal and state governments
- Provides resources, training and support



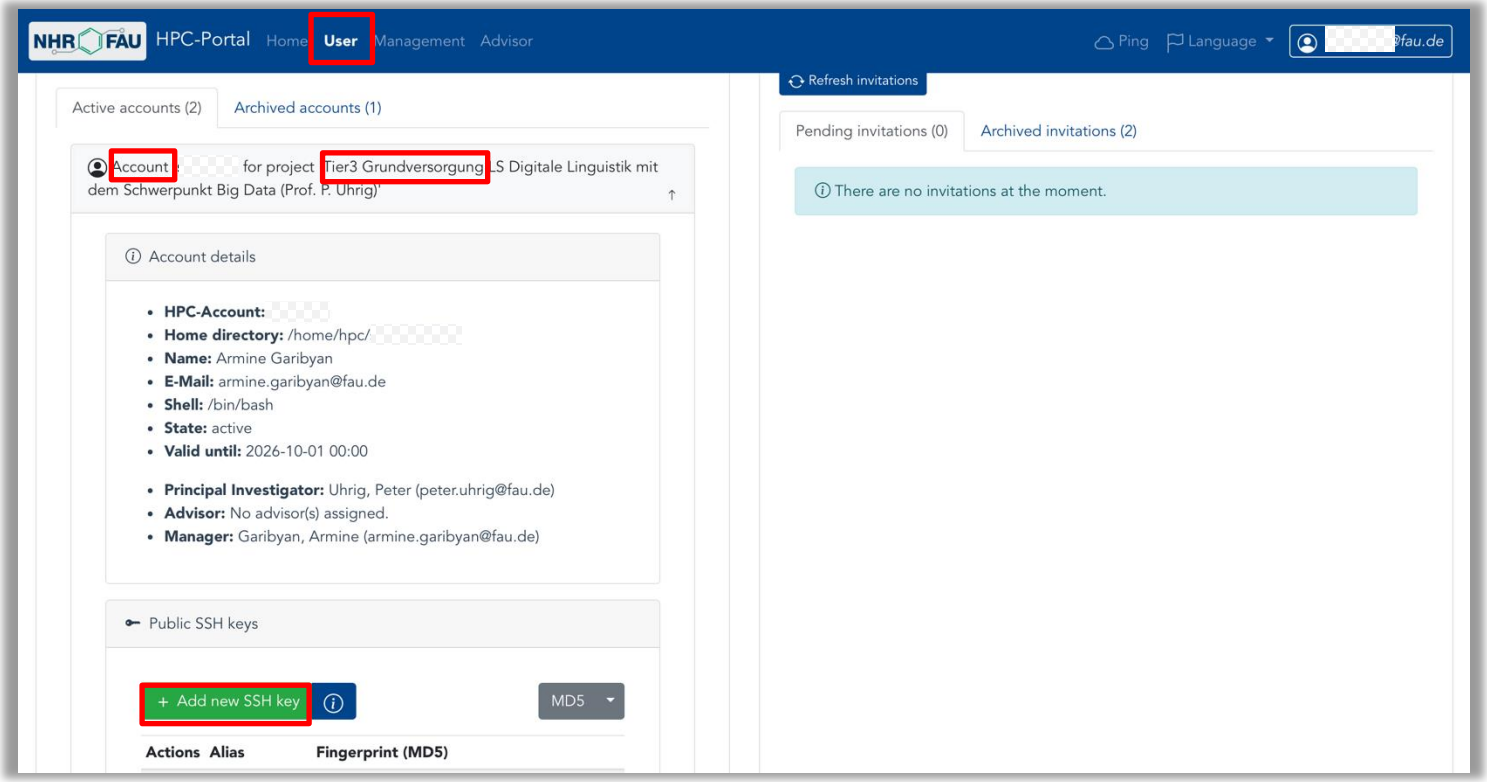
HPC-Portal

- Login at portal.hpc.fau.de with your FAU credentials via SSO

- Main functionality:
  - Track usage of resources
  - Manage your projects
  - Send out invitations to new users
  - Access ClusterCockpit (to monitor performance of your jobs)



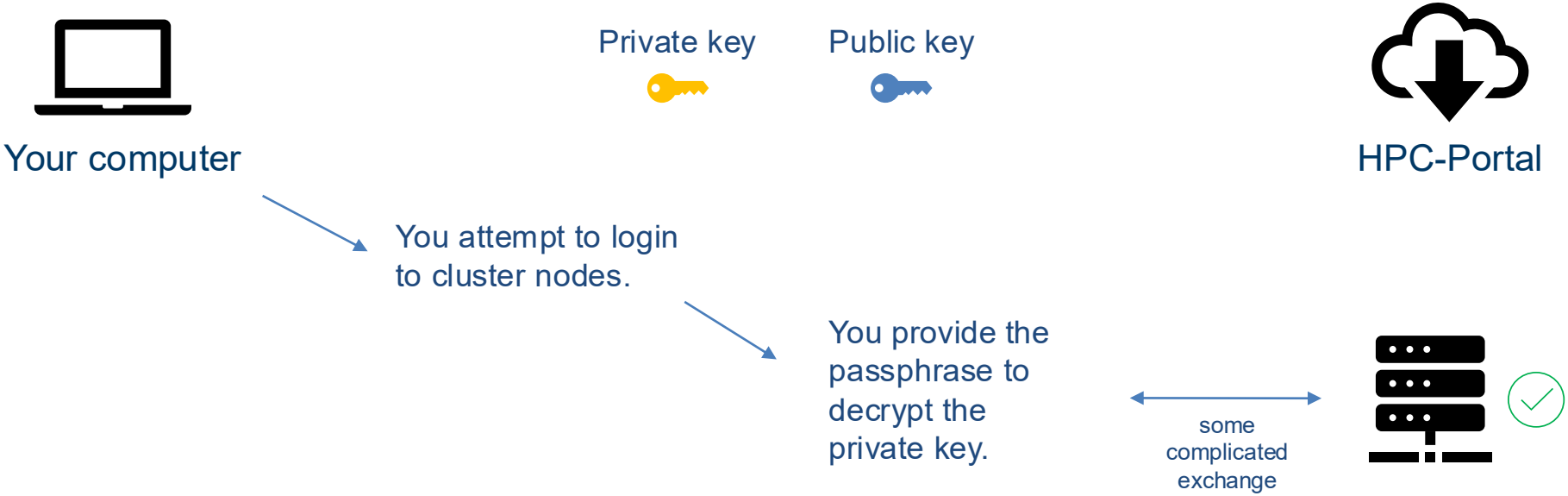
HPC-Portal


NHR FAU
HPC-Portal
Home
User
Management
Advisor
Ping
Language
Active accounts (2)
Archived accounts (1)
Account
for project Tier3 Grundversorgung LS Digitale Linguistik mit dem Schwerpunkt Big Data (Prof. P. Uhrig)
Account details
HPC-Account:
Home directory: /home/hpc/
Name: Armine Garibyan
E-Mail: armine.garibyan@fau.de
Shell: /bin/bash
State: active
Valid until: 2026-10-01 00:00
Principal Investigator: Uhrig, Peter (peter.uhrig@fau.de)
Advisor: No advisor(s) assigned.
Manager: Garibyan, Armine (armine.garibyan@fau.de)
Public SSH keys
+ Add new SSH key
MD5
Actions
Alias
Fingerprint (MD5)
Refresh invitations
Pending invitations (0)
Archived invitations (2)
There are no invitations at the moment.

SSH key pair

- Login to cluster nodes is not password-based, but SSH key pair based.
- The key pair consists of two parts: a public key and a private key

  → The key is basically a very long number (but encrypted).

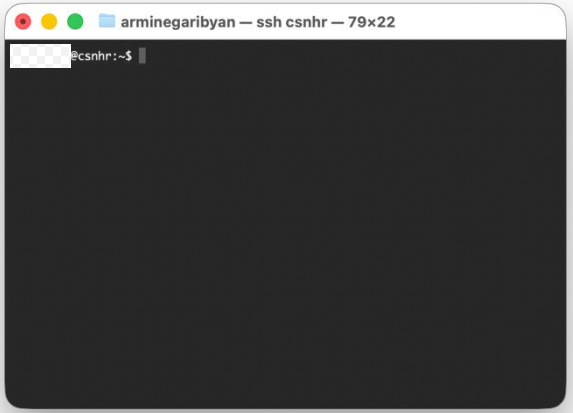


https://doc.nhr.fau.de/access/ssh-command-line/

Accessing the HPC systems

- No GUI is available on the HPC systems → text mode only.
- Basic knowledge of file handling, scripting, editing, etc. under Linux is therefore required.

```
touch   # create a file
mkdir   # create a directory
rm      # remove a file / directory
nano    # edit a file
cp      # copy a file / directory
mv      # move or rename smth
cat     # print / display some contents
ls      # list all files in a directory
```



Accessing the HPC systems

- From within the FAU network (or by using VPN) or via IPv6:

  woody.nhr.fau.de

  tinyx.nhr.fau.de

```
ssh abcd123@woody.nhr.fau.de
ssh woody.nhr.fau.de
ssh woody
```

- Outside the FAU network, use a jump host (ProxyJump):

  csnhr.nhr.fau.de

Change the configuration file!

https://doc.nhr.fau.de/access/ssh-command-line/

NHR FAU

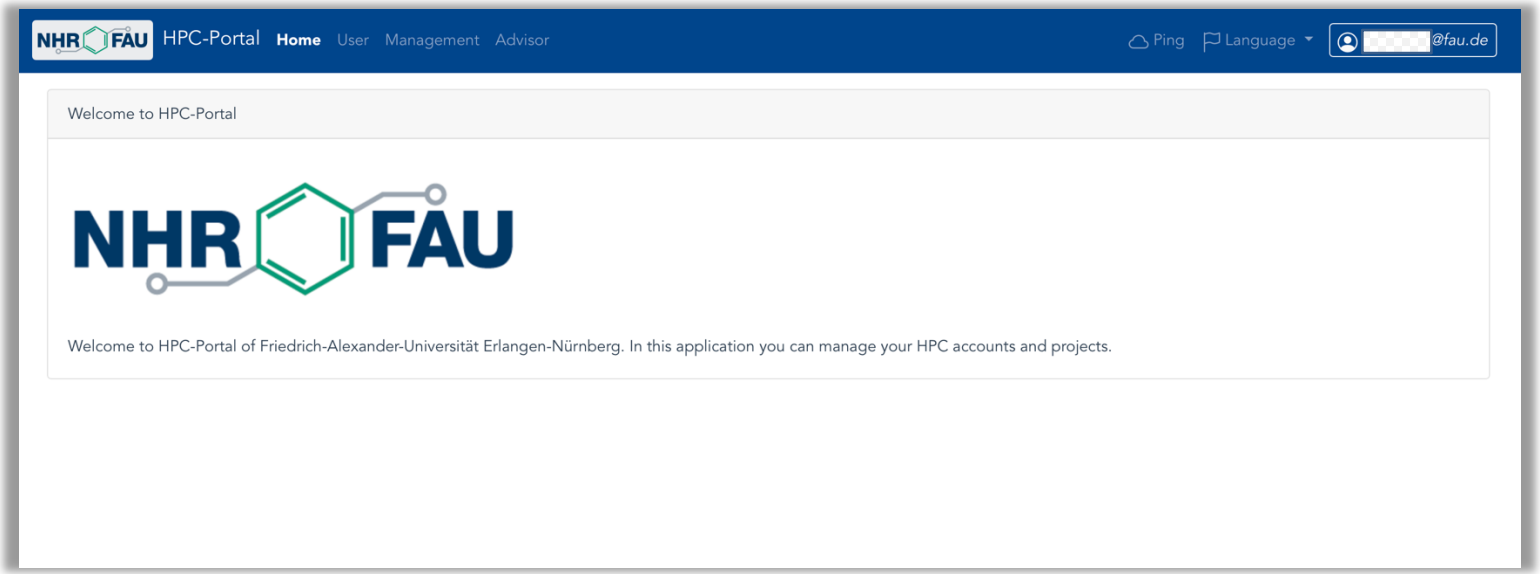
FAU Friedrich-Alexander-Universität Erlangen-Nürnberg

Use cases

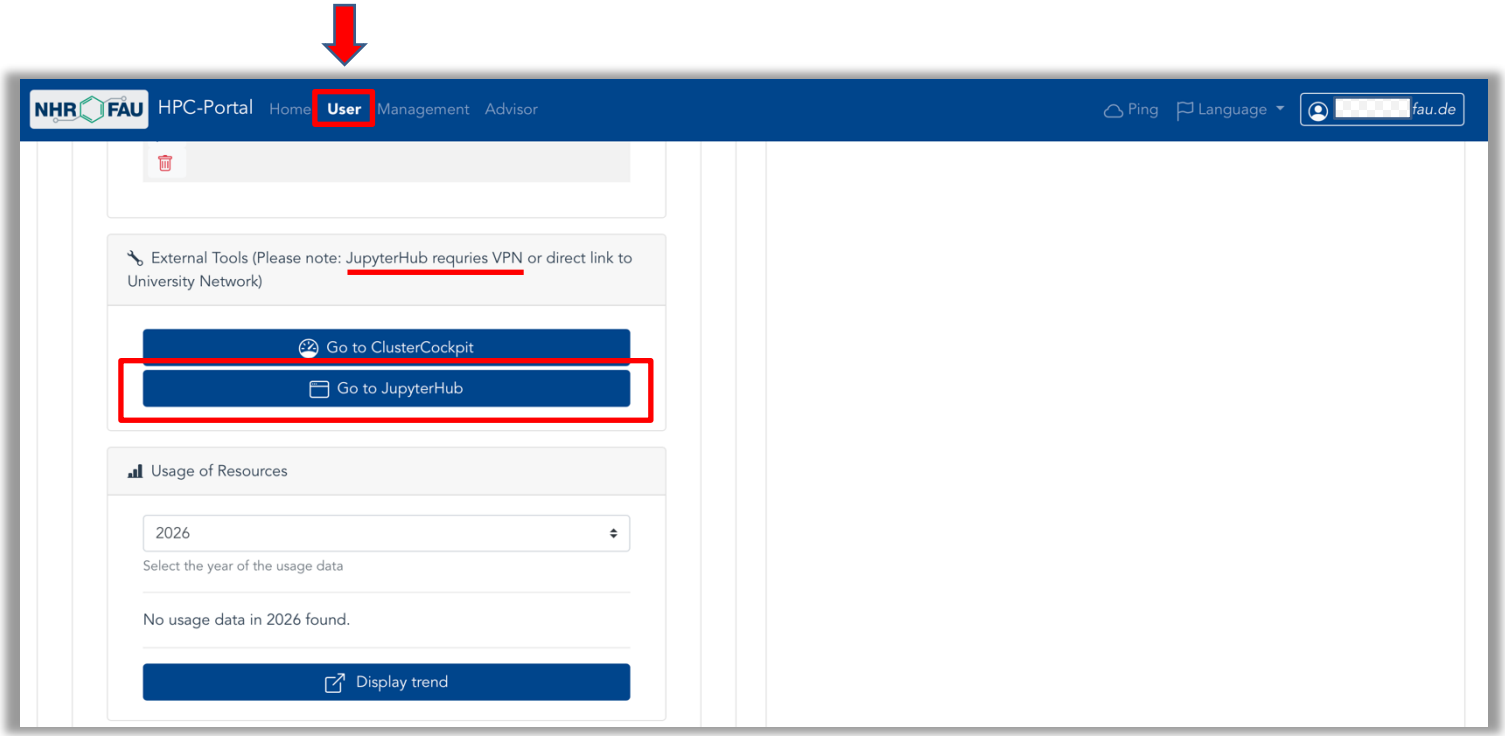
1. GUI-based workflow
2. Terminal-based workflow

Use case 1: Using RStudio through JupyterHub

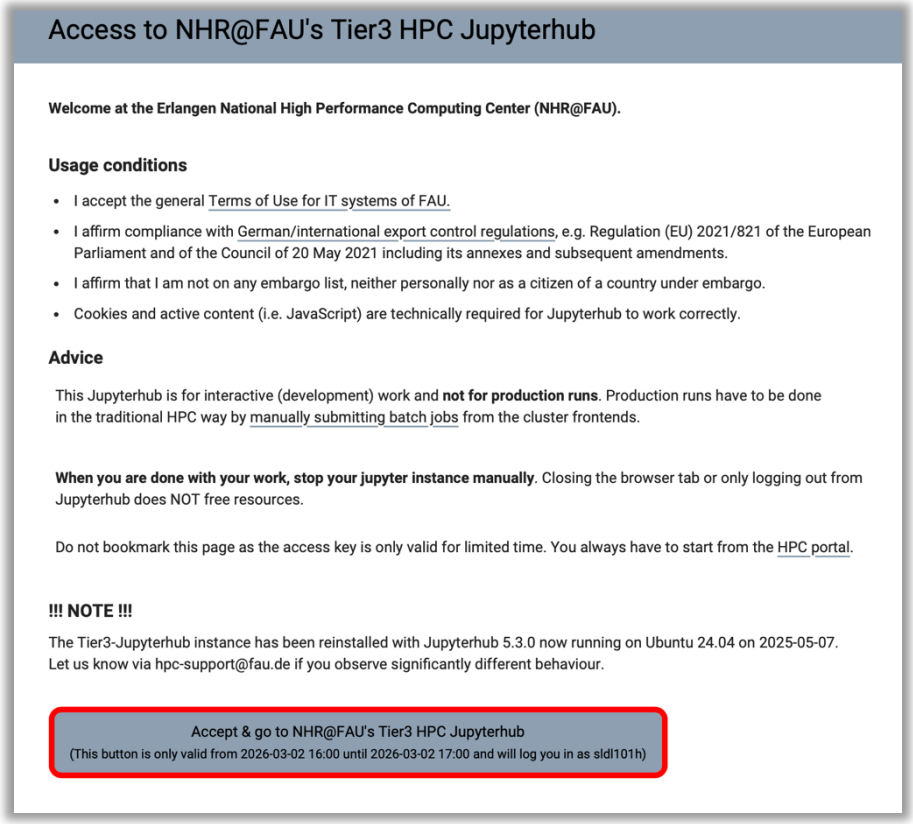
- JupyterHub – multi-user environment running on a remote server
- Access via web browser through HPC-Portal



Use case 1: Using RStudio through JupyterHub



Use case 1: Using RStudio through JupyterHub



Access to NHR@FAU's Tier3 HPC Jupyterhub

Welcome at the Erlangen National High Performance Computing Center (NHR@FAU).

Usage conditions

- I accept the general Terms of Use for IT systems of FAU.
- I affirm compliance with German/international export control regulations, e.g. Regulation (EU) 2021/821 of the European Parliament and of the Council of 20 May 2021 including its annexes and subsequent amendments.
- I affirm that I am not on any embargo list, neither personally nor as a citizen of a country under embargo.
- Cookies and active content (i.e. JavaScript) are technically required for Jupyterhub to work correctly.

Advice

This Jupyterhub is for interactive (development) work and **not for production runs**. Production runs have to be done in the traditional HPC way by manually submitting batch jobs from the cluster frontends.

When you are done with your work, stop your jupyter instance manually. Closing the browser tab or only logging out from Jupyterhub does NOT free resources.

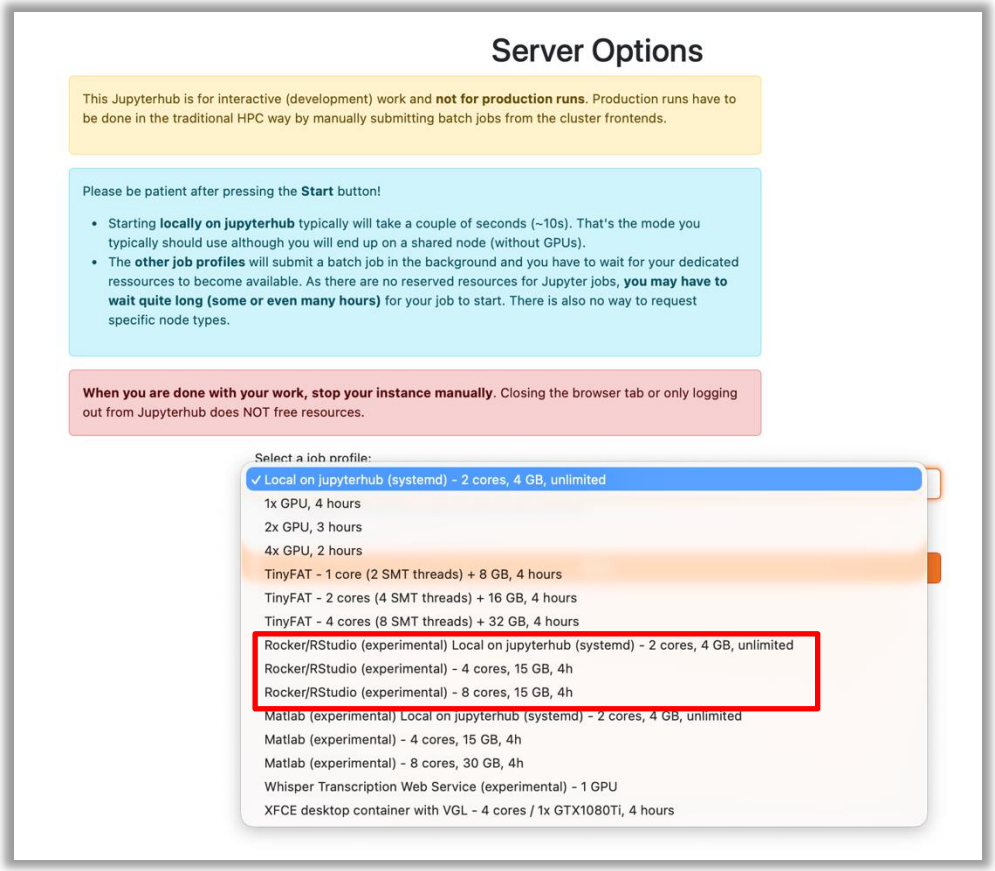
Do not bookmark this page as the access key is only valid for limited time. You always have to start from the HPC portal.

!!! NOTE !!!

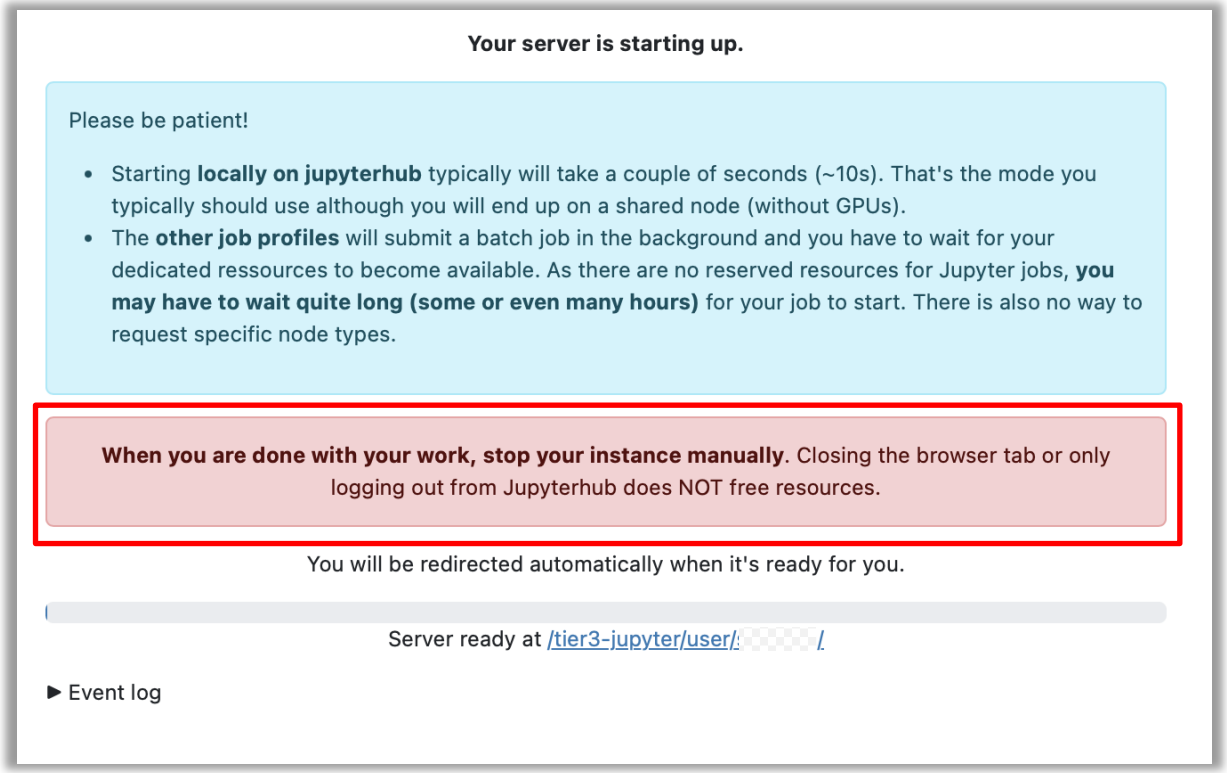
The Tier3-Jupyterhub instance has been reinstalled with Jupyterhub 5.3.0 now running on Ubuntu 24.04 on 2025-05-07.
Let us know via hpc-support@fau.de if you observe significantly different behaviour.

Accept & go to NHR@FAU's Tier3 HPC Jupyterhub
(This button is only valid from 2026-03-02 16:00 until 2026-03-02 17:00 and will log you in as sldl101h)

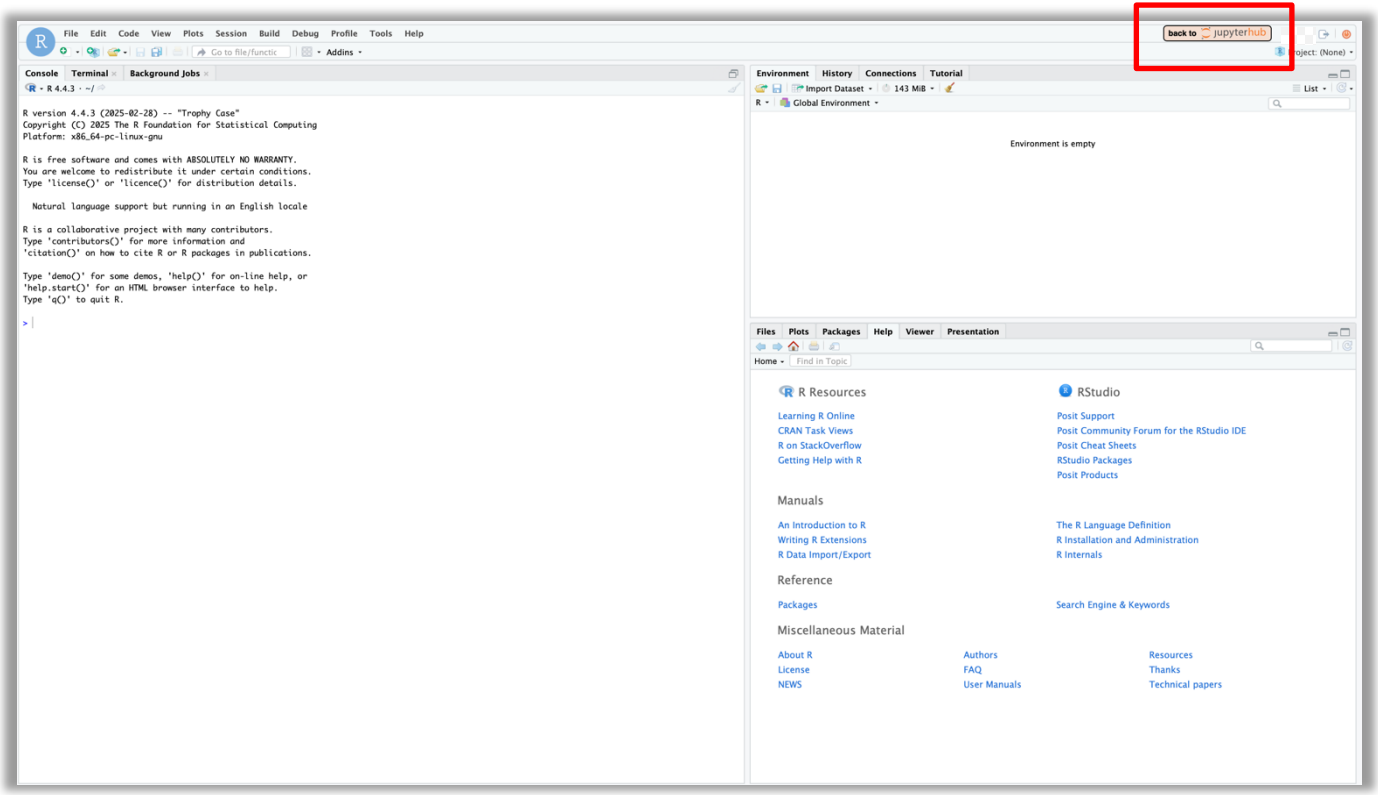
Use case 1: Using RStudio through JupyterHub



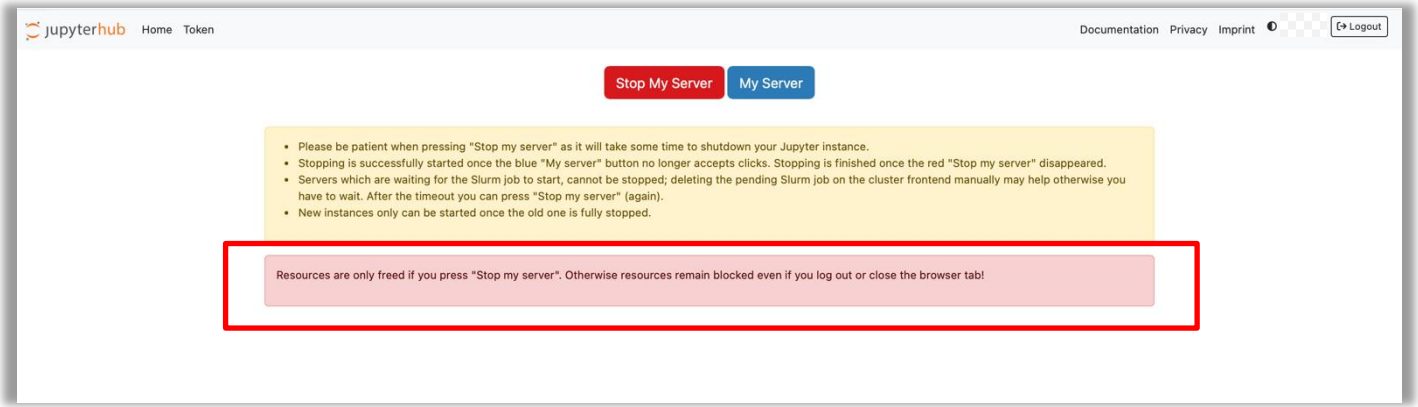
Use case 1: Using RStudio through JupyterHub



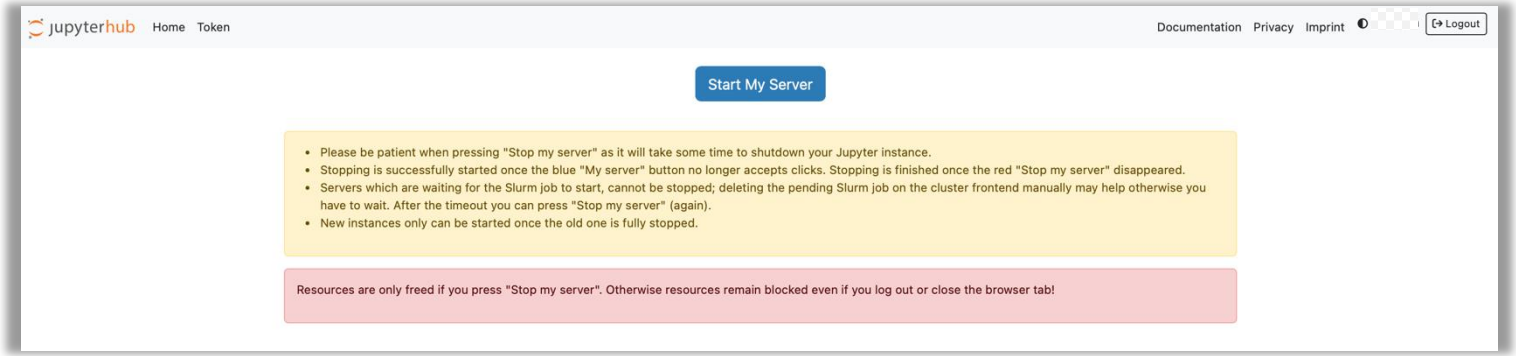
Use case 1: Using RStudio through JupyterHub



Use case 1: Using RStudio through JupyterHub



Use case 1: Using RStudio through JupyterHub



Use case 2: Using R through the terminal

RStudio through JupyterHub	R through the terminal
Automatic setup	Manual setup
Small, medium-sized datasets	Large datasets
Pre-configured time, memory, cores	More control over time, memory, cores
Interactive work	Batch job

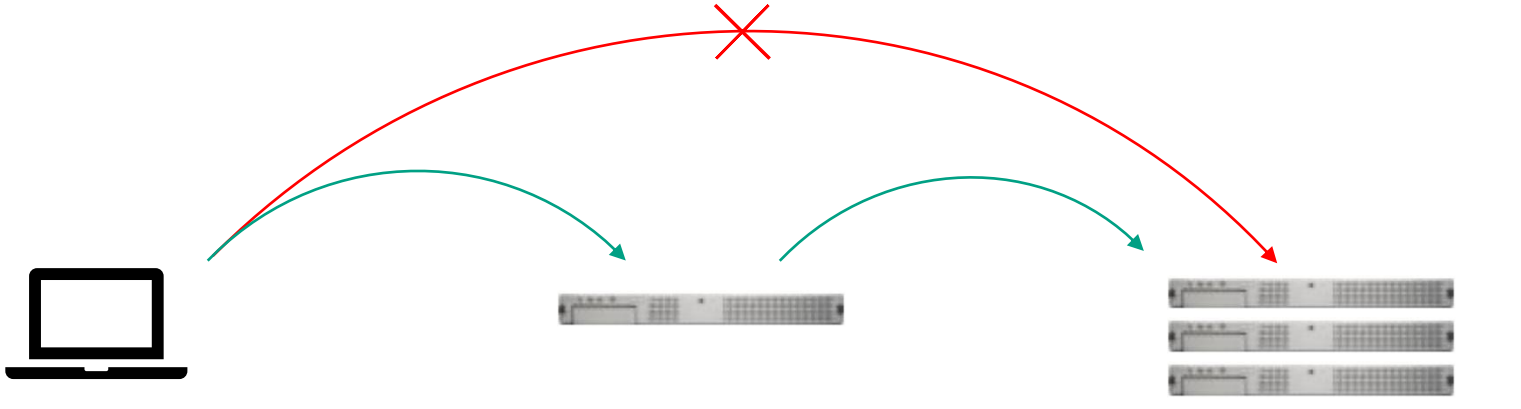
Use case 2: Using R through the terminal

```
$ ssh woody.nhr.fau.de          # log in to the cluster front-end
$ module avail r                # check available R versions
$ module load r/4.3.3-conda     # load the version you need
$ R                             # start R in the terminal

> install.packages("arrows")    # install packages into user library
> q()                           # quit R
```

- Front-end nodes are not suitable for computationally intensive work, but for interactive work, like editing, preparing input, etc.
  - They are shared among all users, so be considerate.
  - All processes will be automatically killed after 1 hour!
- Heavy work must be done on compute nodes (dedicated resources).

Login vs. compute nodes



Batch system

- HPC systems are provided on demand.
- Resources are provided when they become available.
- Batch system will handle queuing of jobs, resource distribution and allocation.

- “Batch jobs” contain the following:
  - Resource requirements (number of nodes, cores, …)
  - Job runtime (usually max. 24 hours)
  - Setup of runtime environment (e.g., environment with installed packages)
  - Commands for application/script run

Example: Batch script for Woody

```
#!/bin/bash -l
#SBATCH --nodes=1                # 1 node/machine/computer
#SBATCH --ntasks=4               # how many "tasks" the job needs to run
#SBATCH --time=07:00:00          # runtime for the job
#SBATCH --job-name=job123        # name of the job
#SBATCH --mail-user=<email>      # get notified on status changes
#SBATCH --export=NONE            # ensure clean environment

unset SLURM_EXPORT_ENV           # prevent export of env. to job

module load r                    # load module(s)

srun Rscript my_script.R         # execute parallel application
```

More templates for all clusters as well as descriptions of all options at:

https://hpc.fau.de/systems-services/documentation-instructions/batch-processing/job-script-examples-slurm/

SLURM batch job submission

```
abcd123@woody4:~$ sbatch my_job_script.sh
Submitted batch job 675329

abcd123@woody4:~$ squeue -l
JOBID     PARTITION NAME      USER      STATE     TIME TIME_LIMI NODES NODELIST(REASON)
675329    work      job123    abcd123   RUNNING   0:06 07:00:00  1     w2504

abcd123@woody4:~$ scancel 675329
```

For all jobs, you get an output (.out) and an error (.err) file

- Options `--output=<outfile>` and `--error=<errfile>` can change their names and directories

General SLURM documentation:

https://doc.nhr.fau.de/batch-processing/batch_system_slurm/

NHR FAU

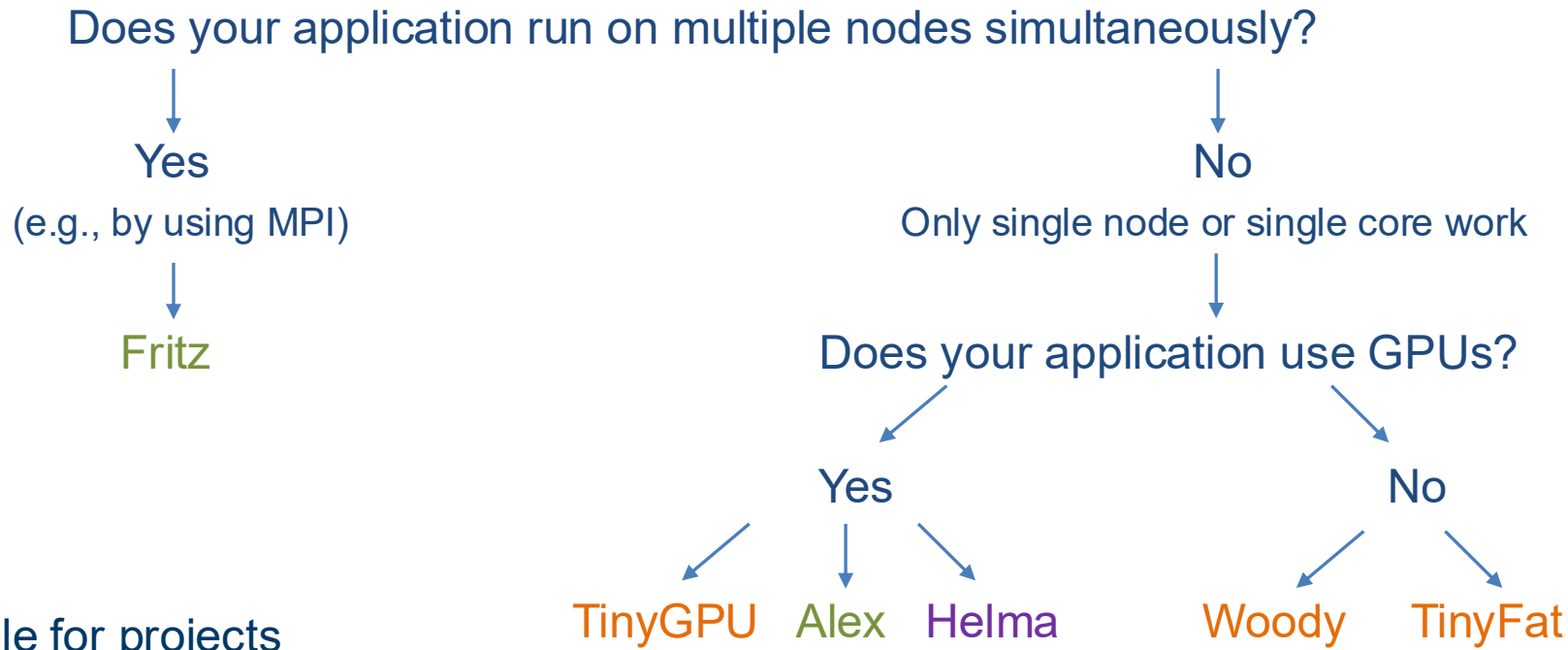
FAU Friedrich-Alexander-Universität Erlangen-Nürnberg

HPC clusters

HPC Clusters

- The HPC clusters at NHR@FAU are tailored to different use cases:
  - multi-node MPI-parallel jobs: Fritz
  - GPU jobs: TinyGPU, Alex
  - single-core or single node jobs: Woody
  - single-core or single node with larger main memory requirement: TinyFat

Which clusters should I use?



Accessible for projects

- “Tier3-Grundversorgung”
- NHR (Tier2 after application)
- Only upon invitation

https://hpc.fau.de/files/2025/10/2025-10-15_HPC_in_a_Nutshell.pdf

Cluster information (Woody)

Hostnames	# nodes	CPUs and # cores per node	Main memory per node	node-local SSD	Slurm partition	Slurm constraint
`w12xx`, `w13xx`	64	1 x Intel Xeon E3-1240 v5 ("Skylake"), 4 cores @3.5 GHz	32 GB	1 TB	`work`	`sl`
`w14xx`, `w15xx`	112	1 x Intel Xeon E3-1240 v6 ("Kaby Lake"), 4 cores @3.7 GHz	32 GB	900 GB	`work`	`kl`
`w22xx` - `w25xx` **(1)**	110	2 x Intel Xeon Gold 6326 ("Ice Lake"), 2 x 16 cores @2.9 GHz	256 GB	1.8 TB	`work`	`icx`

Information about all clusters at: **https://doc.nhr.fau.de/clusters/overview/**

NHR FAU

FAU Friedrich-Alexander-Universität Erlangen-Nürnberg

Storage

File systems

Available file systems differ in size, backup and intended use.

Mount point	Access	Purpose	Technology	Backup	Snap-shots	Data lifetime	Quota
/home/hpc	$HOME	Source, input, important results	NFS	YES	YES	Account lifetime	50 GB
/home/vault	$HPCVAULT	Mid-/long-term storage	NFS	YES	YES	Account lifetime	500 GB
/home/{woody, saturn, titan, janus, atuin}	$WORK	General-purpose, log files	NFS	NO	NO	Account lifetime	500 GB NHR project
/lxfs /lustre	$FASTTMP (fritz)	High performance parallel I/O	Lustre via InfiniBand	NO	NO	High watermark	Only inodes
/???	$TMPDIR	Node-local job-specific dir	SSD/ ramdisk	NO	NO	Job runtime	NO

More detailed information about all file systems and what they are for:
https://doc.nhr.fau.de/data/filesystems/

File systems

- Important results → put in $HOME or $HPCVAULT (backed up, persistent)
- Working files / intermediate outputs → $WORK (okay to delete later)
- High-volume temporary I/O → $FASTTMP or $TMPDIR
- Never rely on $TMPDIR for anything permanent — it's erased when the job ends.

```
abcd123@woody4:~$ shownicerquota.pl              # only for RRZE systems

Path              Used  SoftQ   HardQ   Gracetime  Filec  FileQ  FiHaQ   FileGrace
/home/hpc         5.7G  52.5G   104.9G  N/A        72K    500K   1,000K  N/A
/home/woody       112G  333.0G  499.5G  N/A        188K   N/A
```

Data transfer

- **`scp/ rsync`** is used to transfer files from and to the outside

  https://doc.nhr.fau.de/data/copying/

- Both commands below are executed from your local machine.

```
SOURCE / DESTINATION

# local file to remote
scp my_data.csv abcd123@csnhr.nhr.fau.de:/home/woody/abcd/abcd01/

# remote file to local
scp abcd123@csnhr.nhr.fau.de:~/project/output.txt .
```

Good practices

- Be considerate of the valuable HPC resources.
- When unsure about the amount of resources necessary, do scaling experiments → test interactively.
- Use the appropriate file system; delete obsolete files.
- Check your jobs regularly:
  - Log files (.out and .err)
  - Job monitoring via ClusterCockpit (usage of resources, performance)
- Talk to your colleagues having experience working with HPC
- Have a look at the past HPC Café talks

Good practices

When contacting **hpc-support@fau.de**:

- Provide as much detail as possible
  - “My jobs always crash” will not do
  - Cluster, JobID, file system, time of event, …
  - Batch script, output files, …

Bad practices

- Using login nodes for computational work
- Requesting more cores or memory than you actually need
- Re-using other people's job scripts you don't understand
- Storing large datasets in $HOME

Useful resources

- HPC-Support at **hpc-support@fau.de**

- NHR@FAU documentation: **https://doc.nhr.fau.de**
- ClusterCockpit for Job Monitoring
  - **https://doc.nhr.fau.de/job-monitoring-with-clustercockpit/**
- Past HPC Café talks (slides and recordings)
  - **https://hpc.fau.de/systems-services/support/hpc-cafe/**

- Courses, workshop, consultation hours:
  - **https://www.nhr-verein.de/en/courses-and-workshops**

NHR FAU

FAU Friedrich-Alexander-Universität Erlangen-Nürnberg

Thank you

NHR@FAU

https://doc.nhr.fau.de

hpc-support@fau.de