

Introduction for MPI for Marine Microbiology

Using GWDG systems remotely: HPC, Bioinformatics

Christian Köhler
christian.koehler@gwdg.de

Hans-Georg Sommer
hans-georg.sommer@gwdg.de

Conrad Droste
cdroste@mpi-bremen.de

7. Juli 2021

hpc@gwdg.de

GWDG – Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen



Connecting remotely

HPC systems at GWDG

Data handling

Applications overview

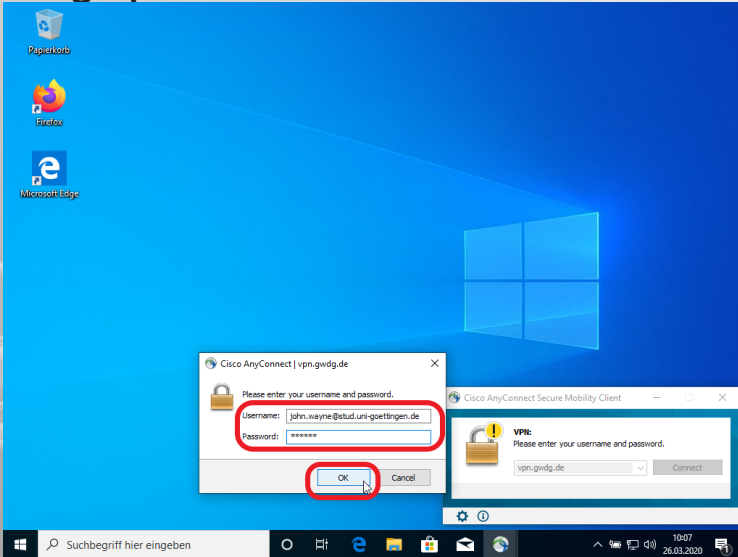
Further resources

Bioinformatics server

Connecting remotely

- visit <https://vpn.gwdg.de>
- click "AnyConnect", download the client
- straightforward installation, client runs in the background (systray icon)
- more info: https://docs.gwdg.de/doku.php?id=en:services:network_services:vpn:anyconnect

Setting up VPN



- on Linux/Mac clients: OpenSSH
- create keypair: `ssh-keygen -t rsa -b 2048 -f .ssh/id-rsa`
- upload public key on <https://www.gwdg.de>
- log in: `ssh -i <yourkey> <userid>@login-mdc.hpc.gwdg.de`
alternative: `login-fas`
- for Windows →

- MobaXterm
 - SSH (among other protocols) client for Windows
 - integrated SFTP file browser
 - bundled X-Server (for GUI applications)
 - <https://mobaxterm.mobatek.net/download.html>
Free "Home" and paid "Professional" Editions available
- X2Go: lightweight protocol client (cross-platform) for GUI applications

Session settings

SSH Telnet Rsh Xdmcp RDP VNC FTP SFTP Serial File Shell Browser Mosh Aws S3 WSL

Basic SSH settings

Remote host * gwdu19.gwdg.de ☒ Specify username ckoehle2  Port 22

Advanced SSH settings Terminal settings Network settings Bookmark settings

☒ X11-Forwarding ☒ Compression Remote environment: Interactive shell

Execute command: ☐ Do not exit after command ends

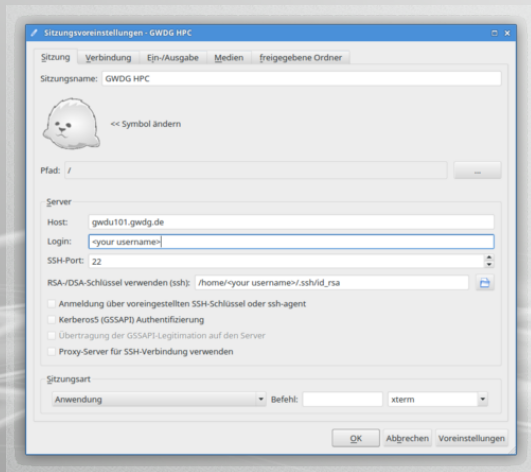
SSH-browser type: SFTP protocol ☐ Follow SSH path (experimental)

☒ Use private key C:\Users\Christian\.ssh\id_rsa.ppk ☐ Adapt locales on remote server

Execute macro at session start: <none>

OK Cancel

- Integrated X-Server
- forwarding of GUI applications
- SFTP file browser opened in parallel

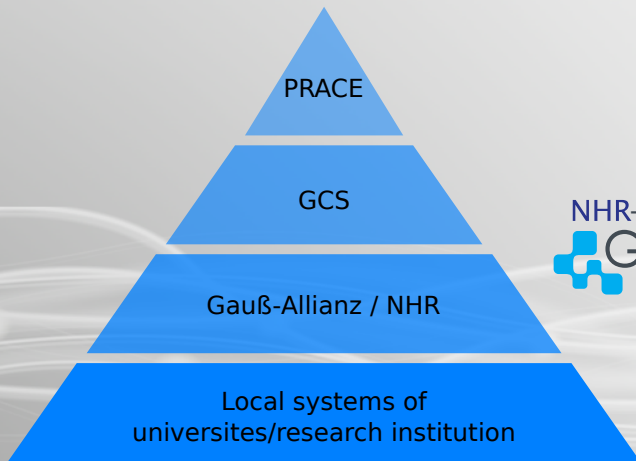


- cross-platform client
- NoMachine NX protocol between client and login node
- useful if network BW/latency is insufficient for X-Forwarding
- Session type: Application xterm

HPC systems at GWDDG

HPC landscape in Germany

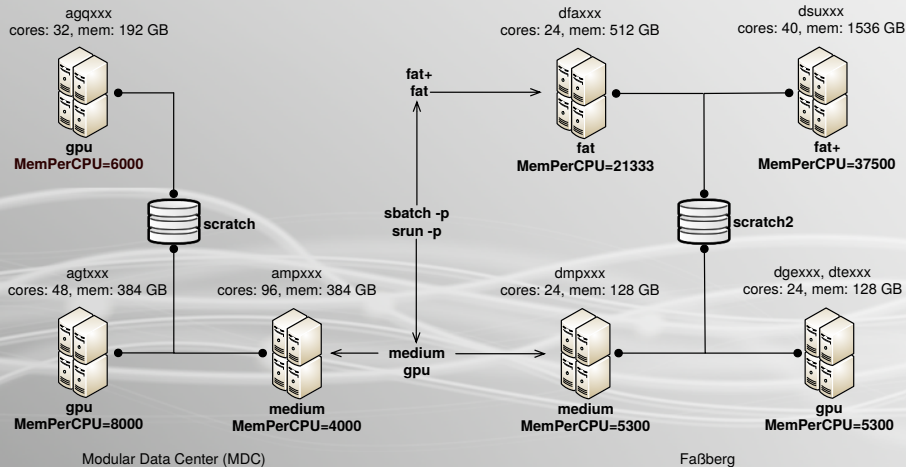
T0: Europe, T1/2: Germany, T3: local



Scientific
Compute
Cluster

- <https://hpc.gwdg.de>
- documentation for the services/applications shown here:
https://docs.gwdg.de/doku.php?id=en:services:application_services:high_performance_computing:start
- compute partition: 255 systems with 5416 cores
- fat partition (high mem): 36 systems with 256 GB - 2TB mem
- GPU partition: 20 Tesla K40, 32 GeForce GTX 980, 14 GeForce GTX 1080
- Storage: StorNext connection (HOME), 2.15 PB HPC FS (BeeGFS)
- 3x login nodes (reachable from GöNET)
- integration of systems from various institutes

Scientific Compute Cluster (SCC)



- 96 standard nodes (2 Xeon Platinum 9242, 2x48 cores, 384 GB memory)
- 2 GPU nodes with 8 Tesla V100/32
 - 2 Xeon Gold 6252, 2x24 cores, 384 GB memory
- 12 GPU nodes with 4 Quadro RTX5000
 - 2 Xeon Gold 6242, 2x16 cores, 192 GB memory
- 100 GBit/s Intel Omni-Path interconnect
- 1.7 PB /scratch storage (of which 100TB are flash based)

New Data Center



hosting DLR system (currently being installed), NHR-Nord expansion
(procurement in preparation)

Data handling

Servermanager

Eintrag auswählen:

- Eigene Server
 - GöHPC Web
 - GWDG**
 - scratch3

Neuer Server Neues Verzeichnis
Neues Lesezeichen Umbenennen
Löschen Kopieren

Allgemein Erweitert Übertragungs-Einstellungen Zeichensatz

Server: Port:

Protokoll:

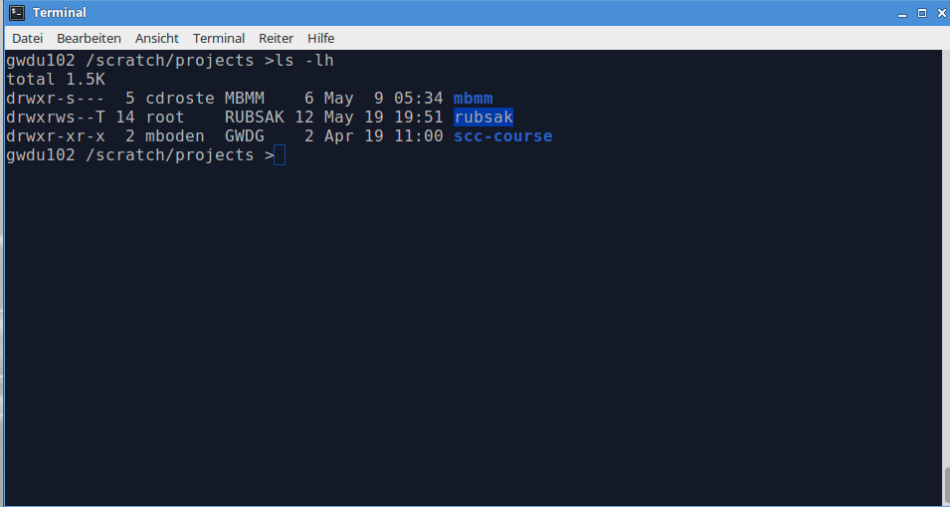
Verbindungsart:

Benutzer:

Schlüsseldatei:

Kommentare:

- very efficient way to synchronize directories
- transfers only delta (new/removed/changed files)
- remote usage possible (uses SSH in the background)
- `rsync -avvH SRC-DIR USERID@transfer-scc.gwdg.de:DST-DIR`

A screenshot of a terminal window titled "Terminal". The window has a blue title bar and a menu bar with options: "Datei", "Bearbeiten", "Ansicht", "Terminal", "Reiter", and "Hilfe". The terminal content shows a user at the prompt "gwdul02 /scratch/projects" running the command "ls -lh". The output lists three directories: "mbmm", "rubsak", and "scc-course", each with its permissions, size, owner, group, creation date, and time. The "rubsak" directory is highlighted in blue in the original image.

```
gwdul02 /scratch/projects >ls -lh
total 1.5K
drwxr-s---  5 cdroste MBMM      6 May  9 05:34 mbmm
drwxrws--T 14 root    RUBSAK 12 May 19 19:51 rubsak
drwxr-xr-x  2 mboden  GWDG   2 Apr 19 11:00 scc-course
gwdul02 /scratch/projects >
```

- file system overview
 - your HOME is available on all nodes
 - /scratch high performance BeeGFS for parallel IO (site-local!)
 - on gwdu101,gwdu102,amp,agq,agt nodes: scratch
 - on gwdu103,dmp,dfa,dsu,dge,dte nodes: scratch2
- archive directory
 - \$AHOME points to your tape archive directory
 - data is migrated to tapes → only store few large files
 - recipe: `tar -czvf $AHOME/data.tgz data`

- recommended for home/office use: ownCloud sync client
- easily transfer to/from HPC by running cadaver

```
$ cadaver https://owncloud.gwdg.de/remote.php/nonshib-webdav
Authentication required for ownCloud on server 'owncloud.gwdg.de'
Username: your-user-name
Password:
dav:/remote.php/nonshib-webdav/>
```

- example commands: **lls**, **lpwd**, **put**, **get**
- for recursive directory access: **module load rclone**

Applications overview

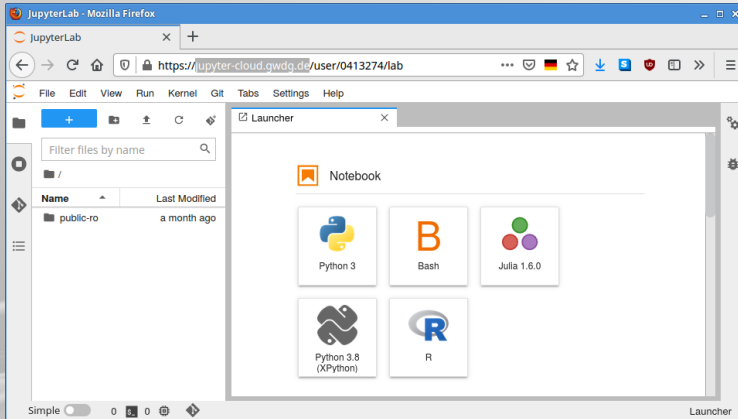
- Application domains & software
 - Quantum chemistry, Molecular dynamics
 - Bioinformatics, Genomics, Evolutionary Biology
 - Astrophysics, Cosmology, Numerical Fluid Dynamics
 - Computer Algebra
 - Machine Learning
 - Medical Imaging (e.g. MRI)
 - Data Analytics: Python, R, Spark, ...
- Module system Lmod → software catalogue module avail

```
Terminal
Datei Bearbeiten Ansicht Terminal Reiter Hilfe

----- BIOINFORMATICS -----
abyss/1.5.2
abyss/1.9.0
abyss/2.1.0 (D)
allpaths-lg/52488
amos/3.1.0
augustus/3.2.2
augustus/3.3.2
augustus/3.3.3 (D)
bambam/1.4
bamtools/2.4.0
bamtools/2.4.1
bamtools/2.5.1 (D)
bbmap/37.66
bbmap/38.68 (D)
bcftools/1.8
bcftools/1.9 (D)
bcl2fastq/2.20
beagle/4.1
beast/1.8.2
beast/2.4.7
beast/2.6.1 (D)
bedtools/2.18.0
filtlong/0.2.0
fithichip/6.0
flappie/1.1.0
flash/1.2.11
flye/2.5
freebayes/1.2.0
funannotate/1.7.4
gapfiller/1.10
genomethreader/1.7.1
genometools/1.5.6
gmap/20190610
grace/5.1.24
grass/7.4.2
great/1.5
guppy/CPU/3.1.5
guppy/GPU/3.1.5 (D)
hail/0.2_novenv
hail/0.2 (D)
hic-pro/2.11.1
hic-pro/3.0.0 (D)
hichipper/0.7.5
hisat2/2.1.0
picard/2.10.5
picard/2.20.2 (D)
pilon/1.22
platanus/1.2.1
plink/1.0.7
plink/1.90 (D)
pomoxis/0.2.2
popbam/0.3
porechop/0.2.3
poretools/0.6.0
prank/170427
profphd/1.0.42
pyrad/3.0.66
qgis/2.14
quast/5.0.1
racon/0.5.0
racon/1.4.3 (D)
raxml-ng/0.9.0
raxml/8.2.4
raxml/8.2.12 (D)
recon/1.08
repeatmasker/4.0.6

lines 56-78
```

Data Analytics Jupyter



- JupyterHub on the SCC: https://docs.gwdg.de/doku.php?id=en:services:application_services:jupyter:hpc
- JupyterCloud instance: <https://jupyter-cloud.gwdg.de>

Further resources

Where to get more information?



- **Info wiki** <https://info.gwdg.de> shows you information about
 - Basic access
 - Hardware overview
 - How to submit jobs
 - Advice on most common applications
- **Support** mail address: hpc@gwdg.de
- Rocket.Chat channel [#hpc-users](#)

- We offer **courses** on SCC usage, programming with MPI, OpenMP, CUDA, Python, Research Data Management and more
→ **GWDG Academy** <https://www.gwdg.de/academy>
- Next Events **“Using the Scientific Compute Cluster”**
 - 19.07.2021 09:30 - 16:00
 - 01.11.2021 09:30 - 16:00
 - will be held online via BigBlueButton

Bioinformatics server

- Bioinformatics server
 - dedicated high mem (1T) system `gwd100.gwdg.de`
 - also hosts RStudio server (web-based IDE)
 - more information: <https://bioinfo.gwdg.de/>
- Dedicated Windows Terminal server

- Bioinformatics server
 - dedicated high mem (1T) system gwdu100.gwdg.de
 - also hosts RStudio server (web-based IDE)
 - more information: <https://bioinfo.gwdg.de/>
- Dedicated Windows Terminal server
- **Thanks for your attention! Questions?**

- Bioinformatics server
 - dedicated high mem (1T) system gwdu100.gwdg.de
 - also hosts RStudio server (web-based IDE)
 - more information: <https://bioinfo.gwdg.de/>
- Dedicated Windows Terminal server
- **Thanks for your attention! Questions?**

→ Hans-Georg Sommer