

Server documentation RIMLSFNWI 2023

- [Introduction](#)
 - [Ubuntu](#)
 - [Contact](#)
 - [Access](#)
- [Server infrastructure](#)
 - [Staff servers](#)
 - [Home](#)
 - [Scratch](#)
 - [Ceph](#)
 - [Backup partition](#)
 - [mbconda](#)
 - [Student server](#)
 - [SLURM](#)
- [Server usage](#)
 - [Ceph layout](#)
 - [Conda](#)
 - [Working on our servers](#)
- [Network and connecting](#)
 - [Radboud network](#)
 - [SSH](#)
 - [SSH clients](#)
 - [Lilo](#)
 - [Network disks](#)

Introduction

This document is intended to provide you with essential information and guidelines for efficiently utilizing our server infrastructure. The primary purpose of this document is to:

- Explain the structure and resources available on our server infrastructure.
- Provide clear instructions on how to access, connect to, and utilize the servers.
- Describe data management practices and storage options.
- Offer guidance on network connections and remote access.
- Ensure that users can maximize the benefits of our servers while adhering to best practices and policies.

Prior to commencing work on our servers, it's essential to note that data retention is not indefinite. Our working directories undergo periodic cleaning, and access to your primary data storage will be available for a few months following the conclusion of your contract, afterwards this will be removed. Following our working [protocol](#) is crucial to prevent any unintended data loss.

Ubuntu

All our servers run on Ubuntu and use the command line to navigate. With Ubuntu comes Bash, a popular scripting language for command line interaction and automation. To get started, here are two tutorials on Ubuntu and Bash, as well as running scripts from the command line:

- <https://ubuntu.com/tutorials/command-line-for-beginners#1-overview>
- <https://www.freecodecamp.org/news/bash-scripting-tutorial-linux-shell-script-and-command-line-for-beginners/>

Contact

CNCZ

Our servers and the network they are connected to are managed by the Faculty of Science IT department [CNCZ](#). They are the primary contact for issues regarding science accounts, science e-mail, open VPN, hardware and network connection.

- Contact: postmaster@science.ru.nl
- Known issue page: <https://cncz.science.ru.nl/en/cpk/>

Radboud University ICT

The primary [ICT department of the Radboud University](#) handles everything regarding printing, office 365, eduroam, RU accounts and RU email.

- Contact: info@ru.nl
- Address: Houtlaan 4, 6525 XZ Nijmegen

Staff Bioinformatician

Our staff bioinformatician has **no** admin privileges. He cannot move, remove, or recover data he does not own.

Sybrein Rinzema, technician bioinformatics handles access for both our staff servers and student servers. He can also be contacted for any questions regarding bioinformatics and server usage.

- Contact: slrinzema@science.ru.nl

Access

Before continuing with the details, it's important to understand how you can gain access to our servers.

Network

All servers are accessed via the Huygens network, which is closed off from the outside by default. There are three main methods available to connect:

- Being physically present in the office or any Radboud building and connecting to eduroam.
- Connecting to the network via VPN ([eduVPN](#) or [open VPN](#)).
- Setting up an ssh proxyjump via [Lilo](#), a login server provided by CNCZ.

Science accounts

To gain access, please send an email to our [staff bioinformatician](#) with the following information:

- **Name** of your science account, *note: don't send your email address, if it contains an @ it is incorrect.*
- The team to which you belong.
- Your role within our department.
- Your supervisors full name, if you are an intern.

You use your science account to log in on our servers, these accounts are handled by CNCZ and a new account can be requested at the secretariat. More information on science accounts can be found at: <https://cncz.science.ru.nl/en/howto/login/>

In case you forgot your login details, you can use your Radboud account to login on dhz.science.ru.nl.

Server Infrastructure

Our server infrastructure is designed to meet the computational needs of our department effectively. In this section, we will explore the various components, including staff servers, storage locations, student servers, and the SLURM system.

Staff servers

We currently have two servers available for our staff to use, below is a list of the names and addresses used to connect to them:

- **CN45 address:** cn45.science.ru.nl
- **CN106 address:** cn106.science.ru.nl

See [access](#) and [network and connecting](#) for more information on how to reach them.

Each staff server has access to multiple shared location to store data, as well as fast a local disk unique for each server called [scratch](#). All of these will be described in more detail later on in this chapter.

Home

- **Path:** `/home/[SCIENCE_ACCOUNT]`
- **Network address:** Check dhz.science.ru.nl
- **Description:** Your home directory is a network disk provided by CNCZ with limited capacity. Your session starts here when logging in on our staff servers, and here are all [Ubuntu](#) configuration files stored.
- **Usage:** Use this directory to store important (system) files, it can also be mounted as a [network disk](#) for ease of access. The address for this can be found via dhz.science.ru.nl.

? More information can be found at: <https://cncz.science.ru.nl/en/howto/storage/#home-directories>

! The soft capacity is 5GB, exceeding this will result in regular emails.

Scratch

- **Path:** `/scratch/`
- **Description:** The scratch directory is a storage space unique for each server. Directly connected to each server instead of via the network like most our data location, it is reliable and fast to work on. When running the `mbconda` setup script, your temporary directory will also be placed here.
- **Usage:** Copy the data necessary for your work here, making sure there's another version of your original data somewhere else, like on [ceph](#). This practice decreases the possibility of data loss due to user error, i.e., mistakes such as accidental deletions or overwrites.

⚠ This location is shared and has a limited capacity, be considerate of other users and keep in mind the possibility of scheduled wipes.

⚠ This location is not backed up, so please ensure data stored here is adequately backed up elsewhere. For example on Ceph.

⚠ This location is not accessible via network connection.

Ceph

- **Path:** `/ceph/rimlsfnwi/`
- **Network address:** `\\molbio-srv.science.ru.nl\[DEPARTMENT_ACRONYM]`
- **Description:** Our Ceph storage is the cornerstone of our data management. It is a large storage solution, subdivided into three primary subdirectories:
 - **data:** The data subdirectory is (among other things) divided into departments, which are divided into teams. These department directories can be mounted as a [network disk](#) with the address specified above.
 - **raw_data:** This location is specifically for data originating from our machines and fastq files generated from the BCL files.
 - **web_share:** This web-accessible location contains two subdirectories: *mbdata* (<https://mbdata.science.ru.nl>) for personal directories and *trackhub* (<https://trackhub.science.ru.nl/>) for hosting important trackhubs used in publications.
- **Usage:** You can store a large amount of data here in a directory of your own, making it easily accessible and when handled properly protecting your data against user error. For example by copying data to [scratch](#) for analysis or archiving important data on our [backup](#).

To make a directory of your own, navigate to your primary `department/team` subdirectory and make a directory of your own.

⚠ Your data will not be stored forever on ceph. After your contract ends you have a grace period of a few months (variable per person) to finish your work, before this period ends you must ensure all important data is either submitted or transferred to the relevant person.

⚠ Except for *raw_data*, Ceph is primarily safeguarded against disk failure, which provides adequate protection against all but the most extreme catastrophes (as CNCZ puts it: 'Not against meteor impact'). Since there is no version control, the possibility of user error still exists.

- ⚠ Please be aware that the *mbdata* web directory is not meant for permanent storage and may be periodically wiped.
- ⚠ Although the storage capacity is substantial and expandable, we have budget constraints. Please be mindful of your storage usage.
- ⚠ A well-known Ceph error is the 'segmentation fault,' which currently lacks a solution. Another reason to start working on [scratch](#).

Backup partition

- **Path:** TBD
- **Network address:** `\\[SCIENCE_ACCOUNT].z.science.ru.n\ [SCIENCE_ACCOUNT]`
- **Description:** A web accessible directory with a storage capacity of 1TB used for securely storing important data. This directory has version control.
- **Usage:** This location serves as a temporary storage space for truly essential data. It offers version control, allowing previous versions to be retrieved by [CNCZ](#) when needed.

⚠ Please carefully assess the data that warrants storage in this location. Ideally, when effectively utilizing all other available storage options, this location should be reserved for unreproducible or labor-intensive data.

mbconda

- **Path:** `/vol/mbconda`
- **Description:** This disk is shared among servers, allowing you to use your conda environments on both servers without the need to transfer them. It also makes environments created by others accessible to everyone.
- **Usage:** To begin using this directory, we offer an installation script at: `/vol/mbconda/install_conda.sh`. Running this script will set up conda for you and also correct temporary storage paths used by many scripts.

Conda environments enable the installation of tools without requiring admin rights in isolated environments, facilitating easy sharing of these environments while keeping it safe from other people. It's considered good practice to create small, dedicated environments for each analysis instead of a single, large environment.

? Read more on how to use conda in the [official documentation](#)

- ⚠ When experiencing long installation times, consider using [mamba](#) instead of conda. Mamba is pre-installed in the shared base environment.
- ⚠ This is not for indefinitely storing environments, ex-staff members their environments will be the first to go when room is needed.

Student server

For master and bachelor internships we have a separate server available called **mbscourse**, with comparable hardware to **cn45/cn106**. It is separated from our main data architecture, providing an environment for the students to work without the danger of deleting or changing vital data.

- **mbscourse address:** `mbscourse.science.ru.nl`

mbshome

- **Path:** `/mbshome/[SCIENCE_ACCOUNT]`
- **Description:** This home directory is a where your session starts when logging in, and here are all **Ubuntu** configuration files stored. Unlike our staff servers, this is not a network accessible disk. And unlike the staff servers this home directory has a larger (shared) capacity.
- **Usage:** Besides storing the standard ubuntu profile files, this directory can be used to store important scripts or files. The latter as long as they aren't too big. A private installation of conda can be installed into this directory.

⚠ This location is shared and has a limited capacity, be considerate of other users.

⚠ This location is also protected from other users, including reading rights. Make sure to copy everything to your supervisor before concluding your internship.

scratch

- **Path:** `/scratch`
- **Description:** The scratch directory is a large storage, it is reliable and fast to work on.
- **Usage:** Use this location to store your data and run your analysis. Make an appropriate directory in either bacint (*bachelor internship*), masint (*master internship*) or supervisors. When making this directory, keep in mind other people can't automatically write in it. Supervisors might need to copy data to their own directory on scratch or help their students with setting the correct permissions.

⚠ This location is shared and has a limited capacity, be considerate of other users.

⚠ Make sure to copy everything to your supervisor before concluding your internship.

SLURM

Our servers are set up as nodes by CNCZ to be part of the slurm cluster. This also explains our server names (i.e. cn45 is cluster node 45)! Our nodes are part of the `rimlsfnwi` partition, if you have access to our servers you have access to our SLURM partition.

If you are interested in using our SLURM partition, please read the CNCZ documentation:

<https://cncz.science.ru.nl/en/howto/slurm/>

Server usage

Ceph layout

Our partition on ceph is structured to maximize data safety and keep data clearly arranged. Below is a tree view displaying all important directories. For more information check the chapter [ceph](#).

```
/ceph/rimlsfnwi
├── data
│   ├── cellbio
│   ├── externals
│   ├── genomes
│   ├── humbio
│   ├── molbio
│   ├── moldevbio
│   └── share_all
├── raw_data
│   ├── bcl
│   ├── fastq
│   └── ms
└── web_share
    ├── mldata
    └── trackhub
```

Our ceph structure as of November 2023.

Some additional directories of interest might be:

- **genomes:** A shared directory for genomes, limiting the space the use and the time it takes to download these files.
- **share_all:** A shared directory for collaboration.

Conda

We use conda environments to ensure the stability and consistency of our scientific projects. Conda enables us to create isolated workspaces with specific software components, guaranteeing that our experiments and analyses produce reliable results across various computing systems and over time. This practice fosters collaboration, simplifies code sharing, and accelerates our research efforts while maintaining the integrity and transparency of our work. It also ensures that we do not get in each others way when installing different versions of the same software, and keeps your environments safe from outside influence.

When using our servers, please make sure you use the *conda installation script* found at: [/vol/mbconda/install_conda.sh](#). This automatically provides you with a base environment and the ability to create your own specific environments.


Working on our servers

To re-emphasize the warning mentioned in this document's introduction: your data is not kept indefinitely on our servers. Working directories like [scratch](#) and the web share [mbdata](#) are wiped clean periodically. And our archival directories like [ceph](#) and our [backup partition](#).

In our ceph directory you can make a personal directory, in the team directory you belong to. It is customary to name your personal directory after your science account name, but not mandatory.

Below is a protocol outlining the safest and most efficient methods for working on our servers.

1. Create a conda environment for your analyses, or use an existing one.
2. Copy or link all crucial data to your directory on [/scratch](#), ensuring there is a safe copy somewhere else.
3. Execute your analysis.
4. Safeguard significant intermediate files by backing them up on [/ceph](#).
5. **Upon completion of your analyses or a specific stage within a larger project, move all (important) data you generated to [/ceph](#). Deleting your data on [/scratch](#) ensures there is space for your colleagues.**

 Due to the large amount of data we produce, it is strongly discouraged to copy fastq files. Instead use symbolic links, as detailed here: https://linuxhint.com/create_symbolic_link_ubuntu/

Network and connecting

Radboud network

All servers are only accessible via the radboud network (eduroam). Make sure you are either connected to the correct wifi or are using a [vpn](#) to connect to the Radboud Network.

SSH

SSH is a secure network protocol for remote computer access and management. It uses encryption to protect data and commands, ensuring privacy and security when working with remote systems over unsecured networks. Here are two example commands on how to use connect to `cn45` from the command line, make sure to replace the address with the server you are trying to connect to.

Standard SSH Command:

```
ssh [SCIENCE_ACCOUNT]@cn45.science.ru.nl
```

SSH Command with X Forwarding:

```
ssh -X [SCIENCE_ACCOUNT]@cn45.science.ru.nl
```

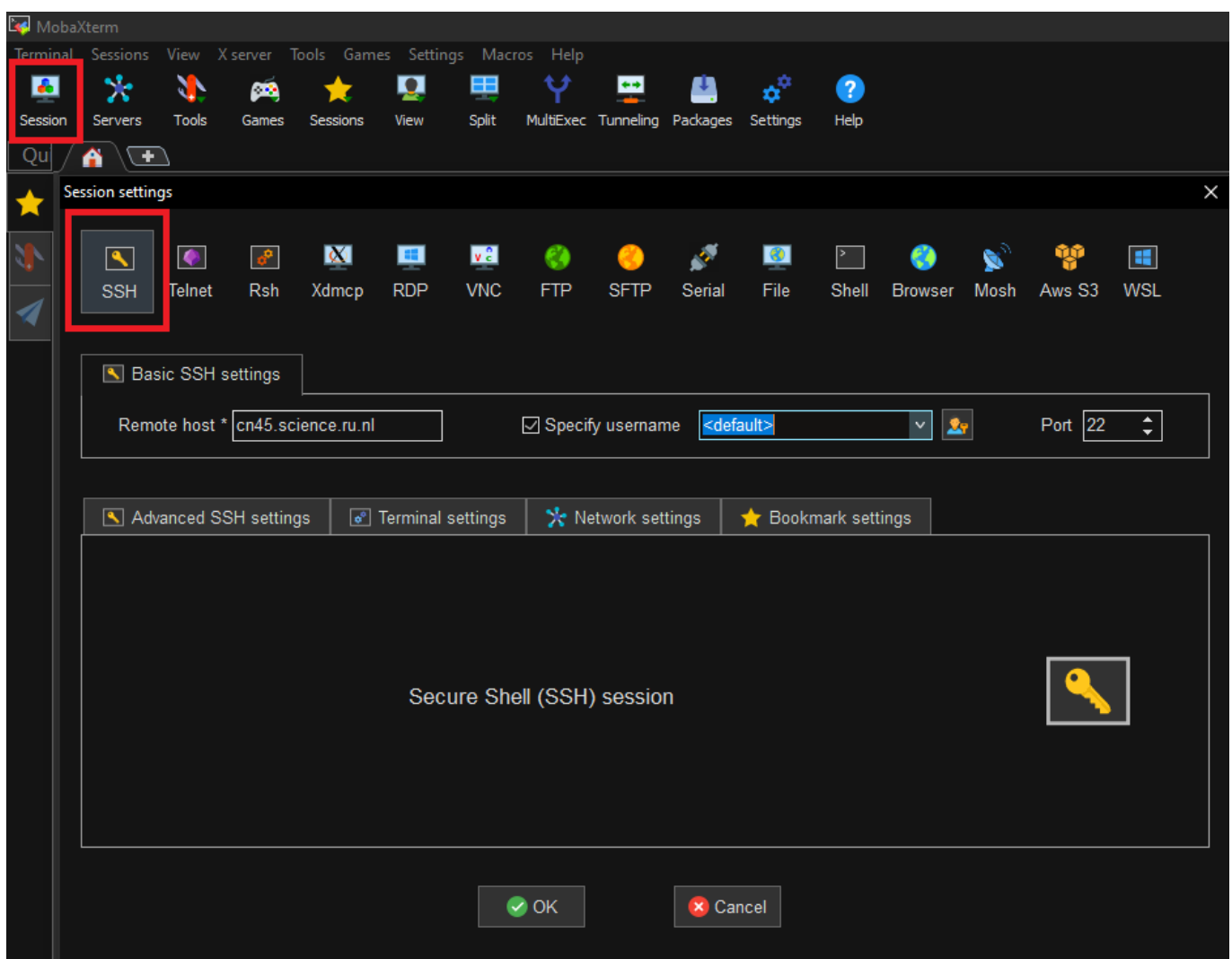
This command enables X forwarding, allowing you to run graphical applications on the remote server, and their graphical output is displayed on your local machine.

SSH Clients

SSH clients are essential tools that facilitate secure remote access and offer a range of features to simplify the management of remote systems. These tools utilize 'sessions' to store server addresses and, if permitted, passwords, thereby eliminating the need to repeatedly log in. We recommend the following:

Mobaxterm - [site](#)

MobaXterm is a versatile SSH client and remote server management tool, compatible with the Windows operating system. It provides a wide range of features, including an integrated X server for graphical applications, tabbed sessions, and a local terminal. It offers SCP (Secure Copy Protocol) as an alternative for secure file transfers between local and remote systems.



Termius - [site](#)

Termius is a secure and user-friendly SSH client available on multiple platforms, including Windows, macOS, and Linux. It offers features like X-forwarding for graphical applications, tabbed sessions, session sharing, strong security through SSH key management, and SFTP support for secure file transfers.

Lilo

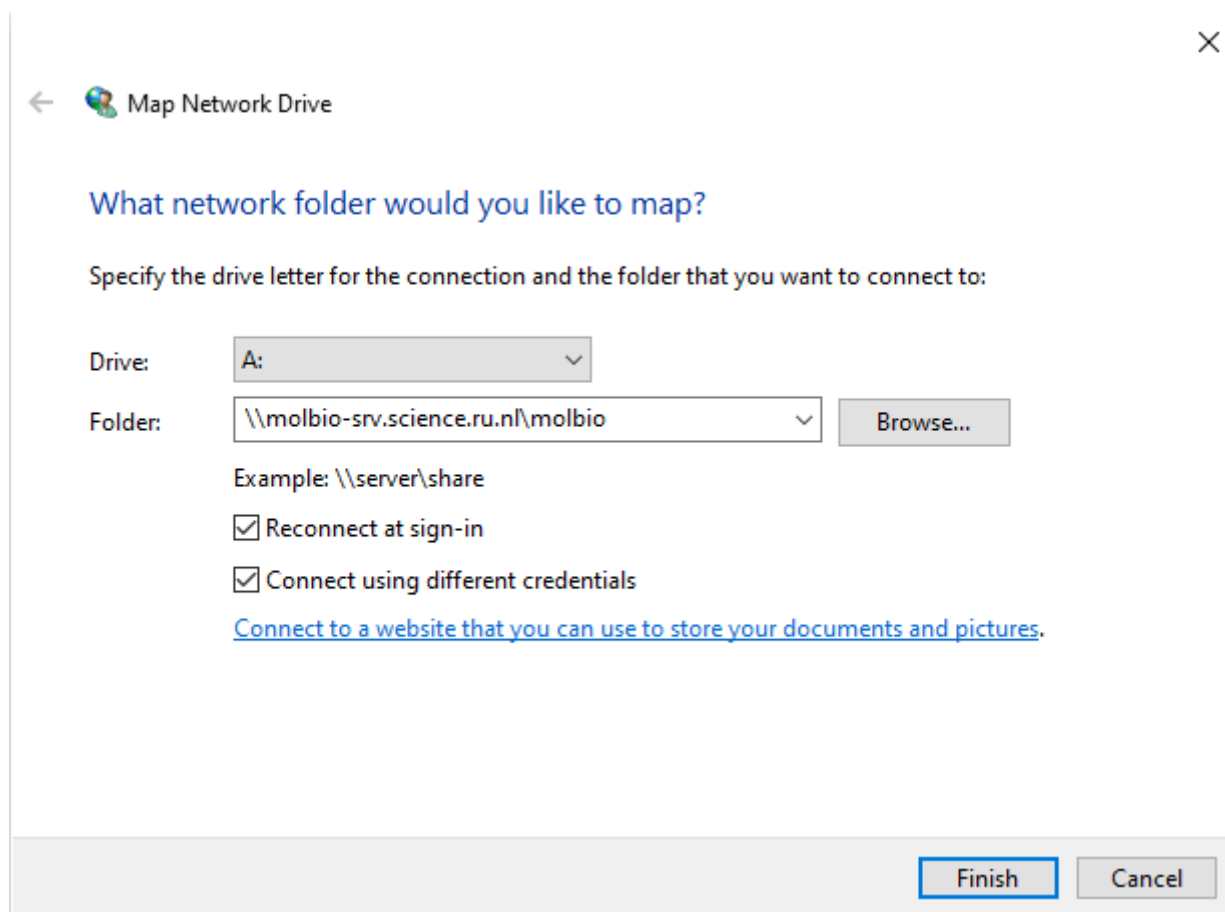
CNCZ offers a couple of login servers, known under the collective name as Lilo. These servers are accessible from outside of the Radboud network and can also be used to access our servers via proxy jumping. If you can not log in on Lilo, you might need to contact [cncz](#). More information on Lilo can be found [here](#).

Network disks

Throughout this documentation you've seen network addresses for some data locations, these addresses lead to network disks. Network disks, also known as Samba shares, are shared folders on a network that allow users to store and access files from different devices. They enable easy file sharing and collaboration among users connected to the same network, regardless of their operating system.

Windows:

1. Open File Explorer.
2. Click on "This PC" in the left sidebar.
3. Click on the "Computer" tab and select "Map network drive."
4. Choose a drive letter for the share.
5. In the "Folder" field, enter the share path (e.g., \server.science.ru.nl\share).
6. Check the box that says "Reconnect at sign-in" to make the mount permanent.
7. Select "Connect using different credentials"
8. Click "Finish" and enter your **science credentials** if prompted.



Mac:

1. Open Finder.
2. Click "Go" in the menu and select "Connect to Server."

3. Enter the share path (smb://server.science.ru.nl/share, please correct the supplied network locations.) and click "Connect."
4. Authenticate with **science credentials** if required.
5. The share will appear on the desktop and in the "Finder" sidebar. To make it permanent, add it to your Login Items in System Preferences > Users & Groups.