Amsterdam UMC

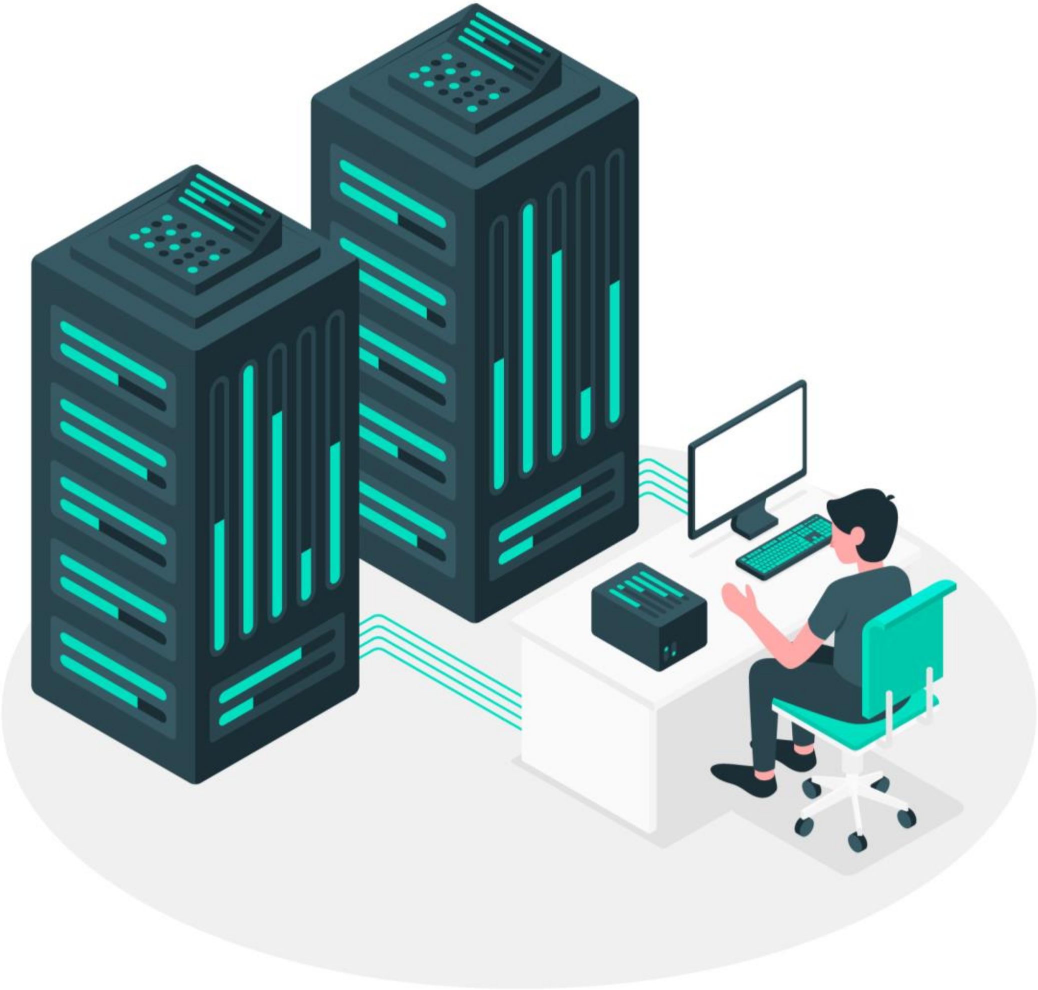
High Performance Computing

Introduction course

Outline of the course

- What is a compute cluster
- Overview of our cluster
- Connecting to the cluster
- How to store your data
- Module usage
- Slurm usage
- Open on demand
- Resources

What is a compute cluster



- A computing cluster is a group of connected computers (called nodes) that work together as a single system to perform tasks more efficiently
- 1994 first cluster of commodity PCs at NASA
- Most clusters run on Linux

White box cluster

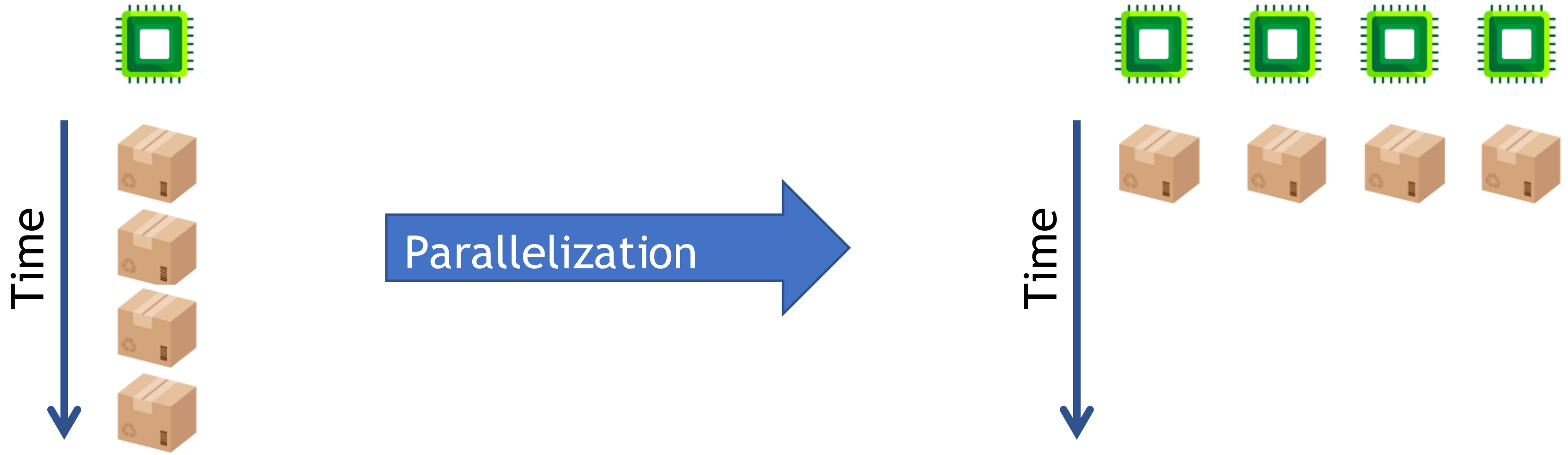
https://en.wikipedia.org/wiki/Beowulf_cluster#/media/File:Beowulf.jpg

Boelelaan MedFac

Meibergdreef Datacenter

General use cases

- Running long jobs
- Running many jobs at once
- Running parallel jobs



Overview of the cluster

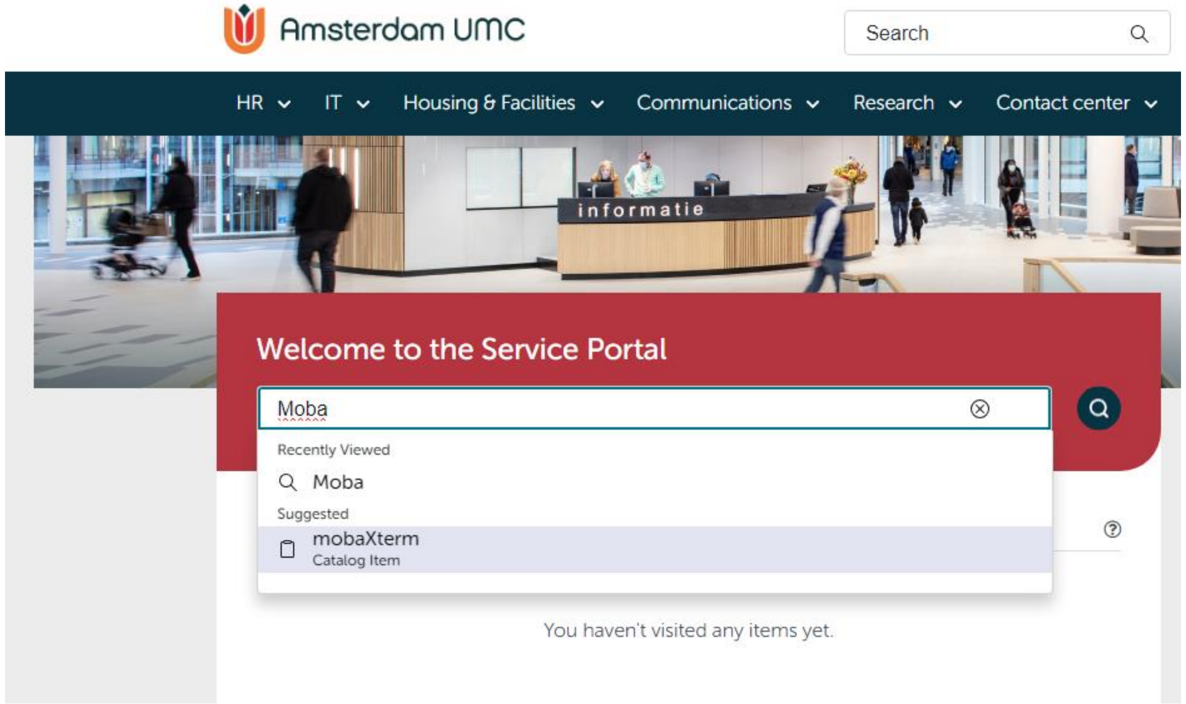
Overview

- Here is the list of available nodes
  - 9 x worker nodes
  - 3 x gpu nodes
  - 1 x 2.2 TB bigmem node
  - 2 x login nodes (one for zorg and one for research)
  - 1 x 100 TB fast storage node
  - 1 x 1 PB normal storage node

Connecting to the cluster

SSH Client

- How to get MobaXterm?
    - You can go to Service Portal
    - Search for 'mobaxterm' and order it.
    - You will get an e-mail once it is available for your account.



Welcome to the Service Portal

Moba

Recently Viewed

Moba

Suggested

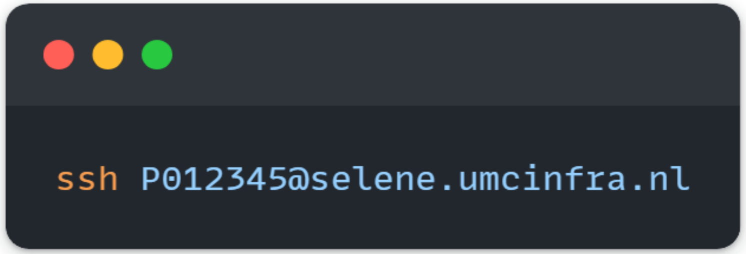
mobaXterm
Catalog Item

You haven't visited any items yet.

You can order easily from here: mobaXterm - Employee Center

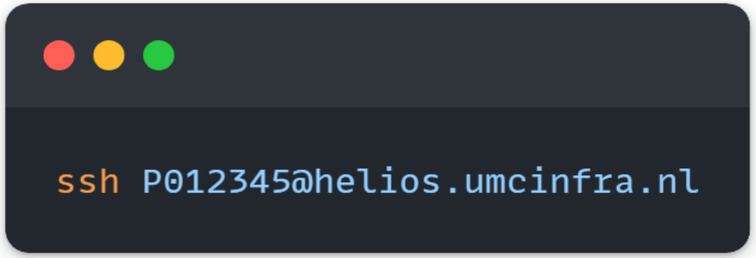
SSH Client

- How to connect to the cluster:
  - ○ Diagnostics login node:



```
ssh P012345@selene.umcinfra.nl
```

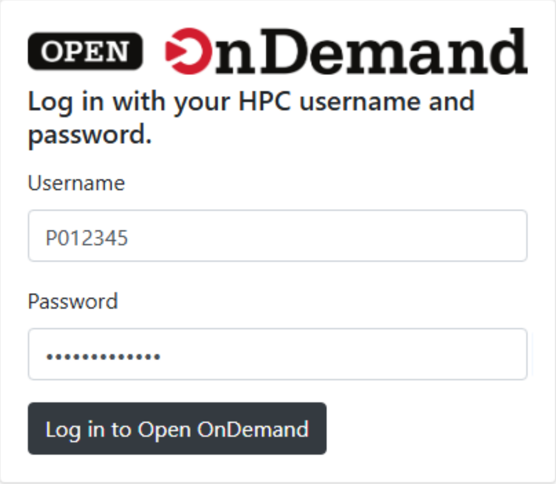
  - ○ Research login node:



```
ssh P012345@helios.umcinfra.nl
```

OpenOnDemand

Diagnostics: selene.umcinfra.nl **Research:** helios.umcinfra.nl



Log in with your HPC username and password.

Username: P012345

Password: •••••••••••••

Starting an open on demand session

- In Edge, go to https://helios.umcinfra.nl or https://selene.umcinfra.nl
- Enter you Pnumber
- Enter your amsterdamumc password
- Click ‘Log in to Open OnDemand’

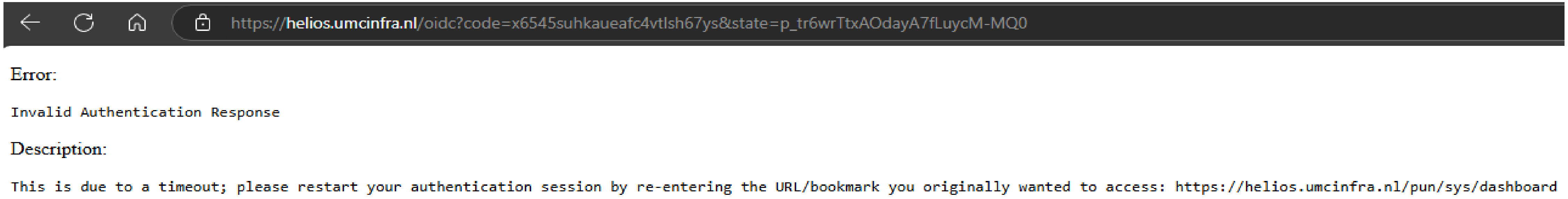
OpenOnDemand

Things to take into account

- When you enter OpenOnDemand you'll be automatically redirected to a temporary login key, looking something like this:

https://helios.umcinfra.nl/dex/auth/ldap/login?back=&state=eaap3mqivp7i2dp75mu2372ew

- If you don't log in within 5 minutes, this key will expire.



https://helios.umcinfra.nl/oidc?code=x6545suhkaueafc4vtlsh67ys&state=p_tr6wrTtxAOdayA7fLuycM-MQ0

Error:

```
Invalid Authentication Response
```

Description:

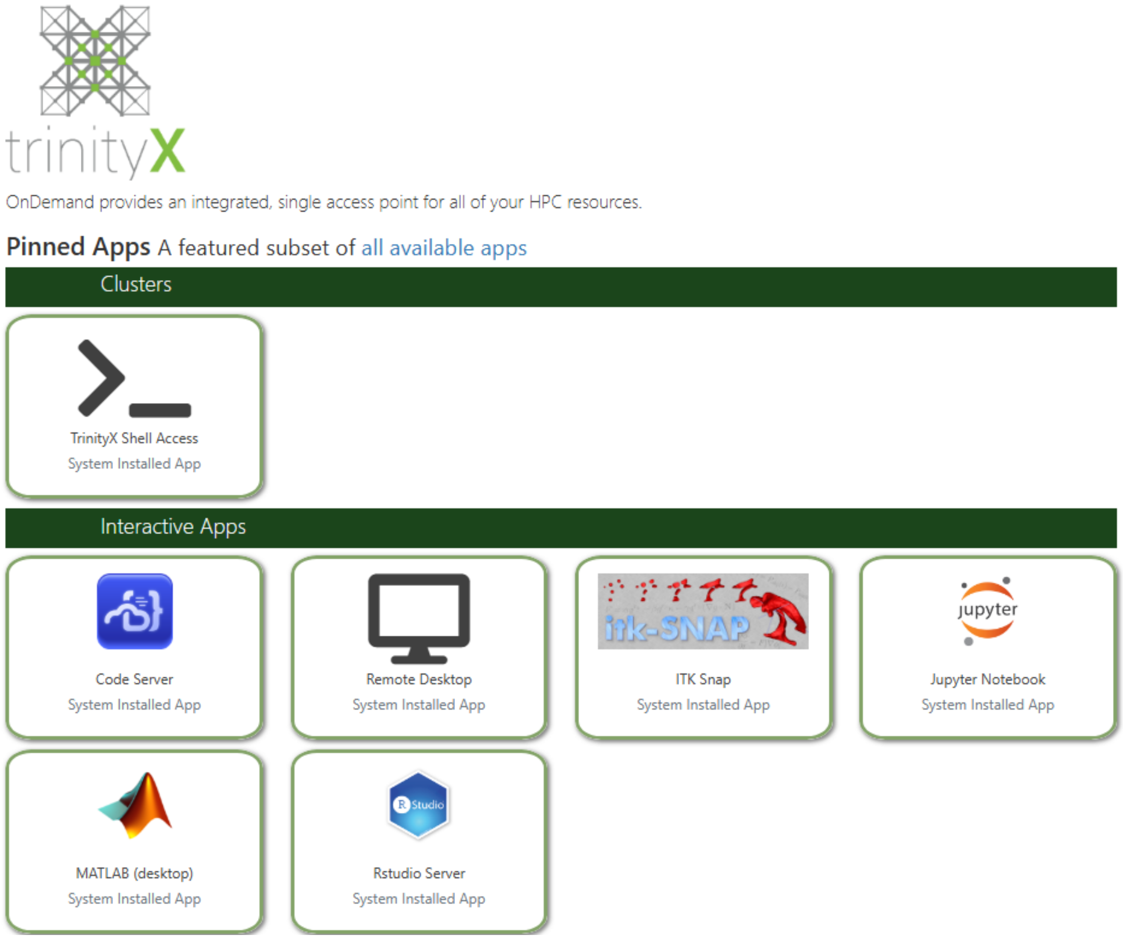
```
This is due to a timeout; please restart your authentication session by re-entering the URL/bookmark you originally wanted to access: https://helios.umcinfra.nl/pun/sys/dashboard
```

Don't worry, just go to OpenOnDemand again to repeat the login procedure

OpenOnDemand

You’ll find pinned apps that are used frequently

- Code Server
- Interactive Desktop
- Jupyter notebook
- Rstudio
- Shell command line interface



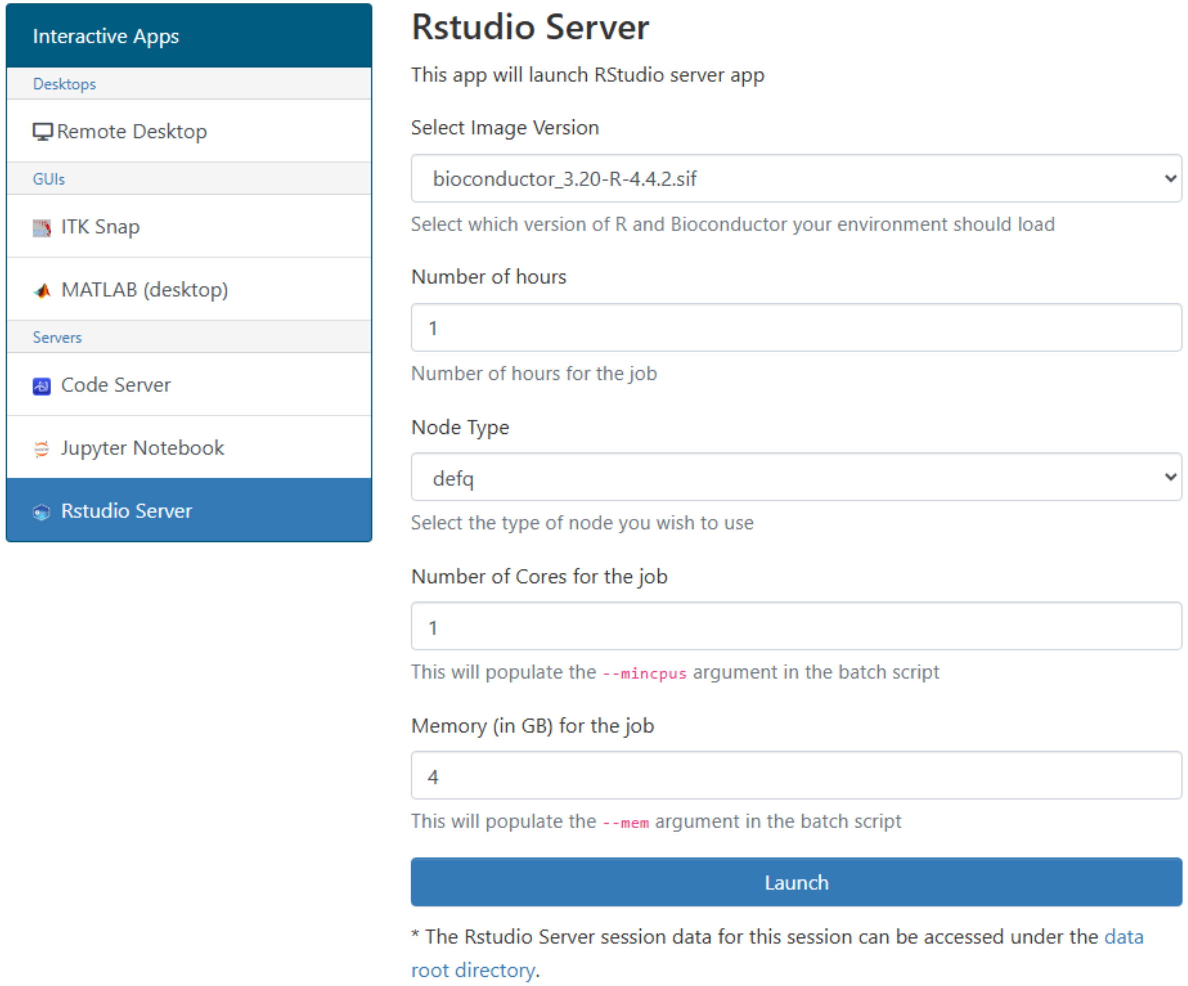
OnDemand provides an integrated, single access point for all of your HPC resources.

Pinned Apps A featured subset of all available apps

OpenOnDemand - RStudio

Specify your slurm settings for you job:

- Bioconductor version
- Number of hours for your job
- Node type (worker or big memory)
- Number of cores
- Amount of memory (in gigabytes)



Rstudio Server

This app will launch RStudio server app

Select Image Version: bioconductor_3.20-R-4.4.2.sif

Select which version of R and Bioconductor your environment should load

Number of hours: 1

Number of hours for the job

Node Type: defq

Select the type of node you wish to use

Number of Cores for the job: 1

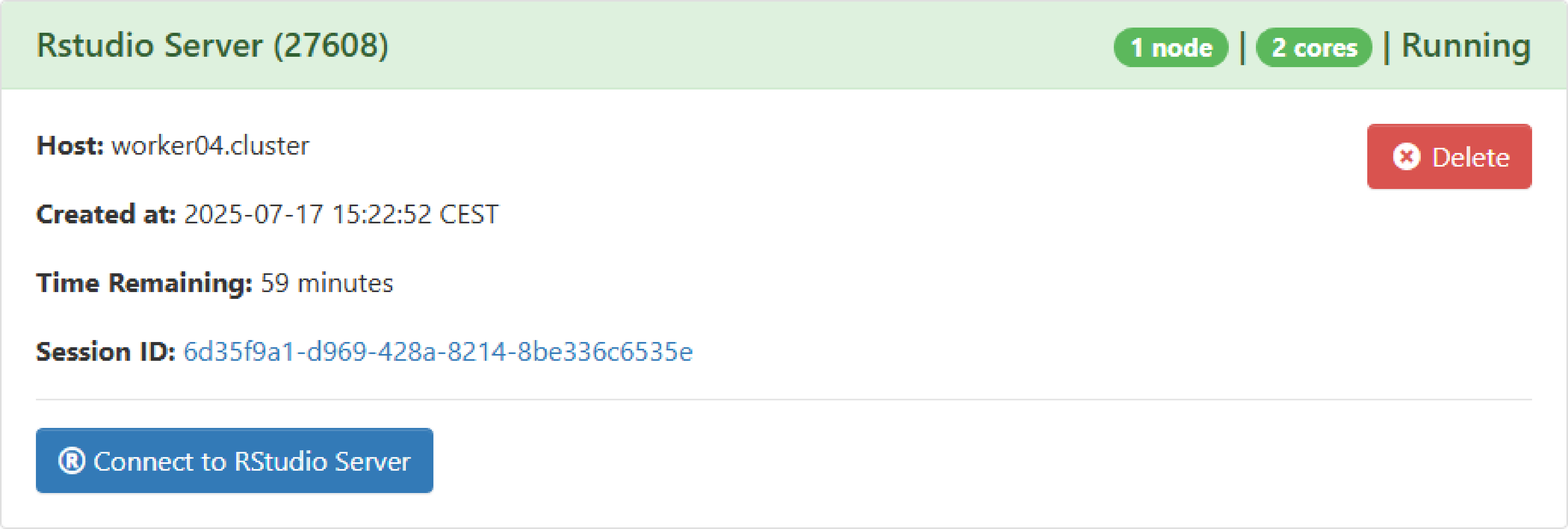
This will populate the `--mincpus` argument in the batch script

Memory (in GB) for the job: 4

This will populate the `--mem` argument in the batch script

* The Rstudio Server session data for this session can be accessed under the data root directory.

OpenOnDemand - RStudio



Rstudio Server (27608)

Host: worker04.cluster

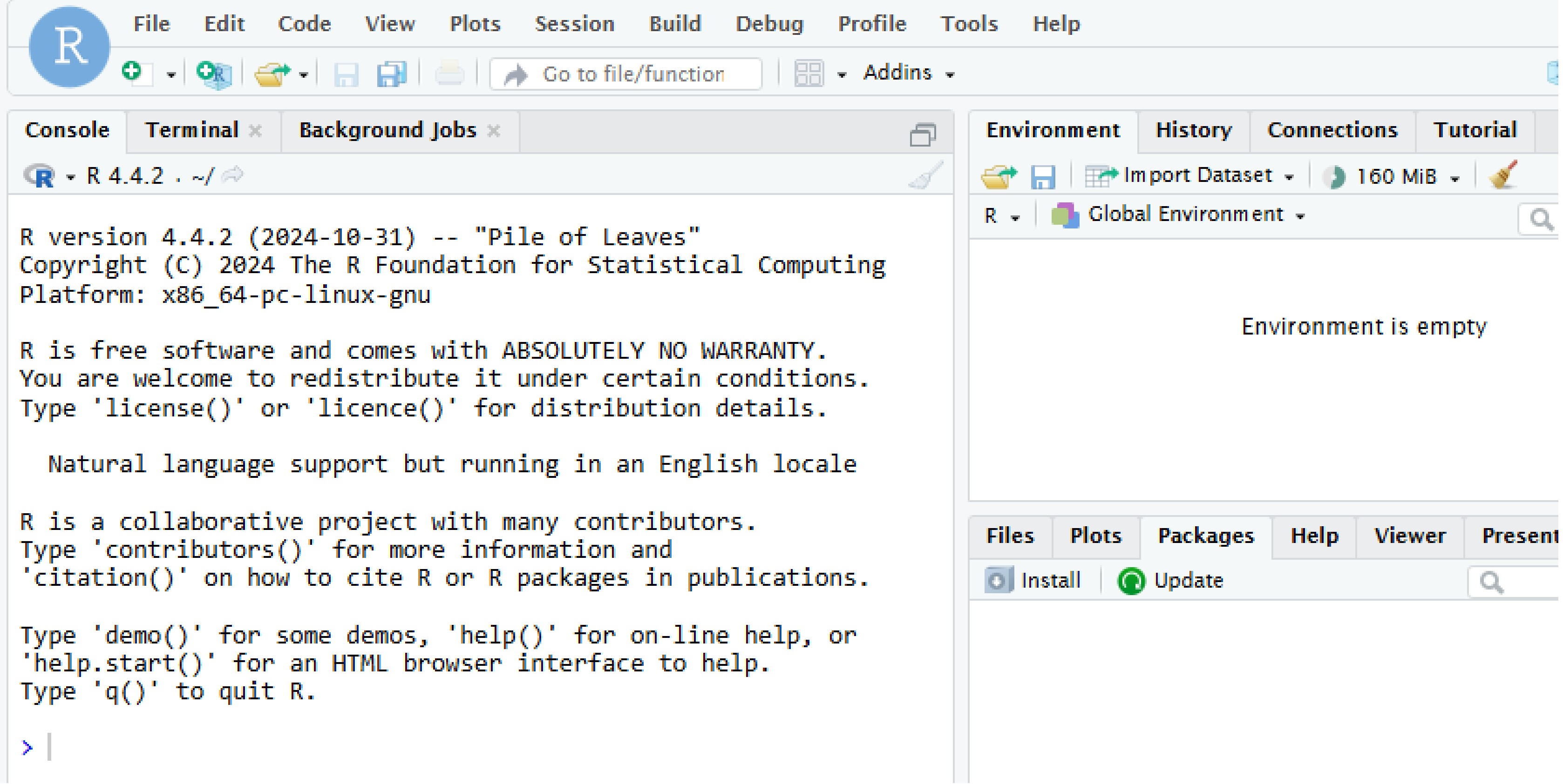
Created at: 2025-07-17 15:22:52 CEST

Time Remaining: 59 minutes

Session ID: 6d35f9a1-d969-428a-8214-8be336c6535e

When your job is ready, you can click on “Connect to RStudio Server”

OpenOnDemand - RStudio



```
R version 4.4.2 (2024-10-31) -- "Pile of Leaves"
Copyright (C) 2024 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> 
```

Now you can use RStudio

How to store your data

Folder structure

- How it is structured:
  - *`/net/[type]/[groups/projects]`*

  - Example:
    - /net/beegfs/groups/cfg
    - /net/beegfs/projects/RODAM-pros

Your directories

- Home folder [/home]:
  - Once logged in, you will be here `/home/<P012345>`.
  - Landingpage
- Job Scratch folder [/scratch]:
  - Fast storage
  - Where you can store your active data during Slurm executions.
- Personal folder [/net/beegfs/user/<P012345>]:
  - Where you can store personal data.
- Group folder [/net/beegfs/groups/<group_name>]
- Project folder [/net/beegfs/projects/<project_name>]

Storage Policies

- Home folder [/home]:
  - Capped 10GB
  - Retention Time: Unlimited
- Job Scratch folder [/scratch]:
  - No limit
  - Retention Time: Job time - If not removed via job, then 2 weeks
- Personal user folder [/net/beegfs/users/*<P012345>*]:
  - No limit but please don't exceed 10TB
- Group folder [/net/beegfs/groups/*<group_name>*]:
  - No limit but please don't exceed 50TB
- Project folder [/net/beegfs/projects/*<project_name>*]:
  - No limit but please don't exceed 50TB

FAIR storage usage

- Please use it in a FAIR way:
    - Don’t store too much data for too long
- We are working on FAIR data policies
    - Currently not limited / capped
    - Currently not billed
- We will check regularly for disk usage for FAIR use

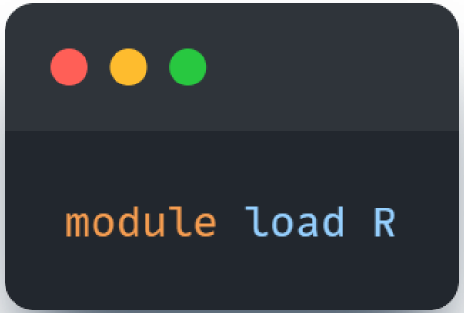
Module usage

Module commands

List available

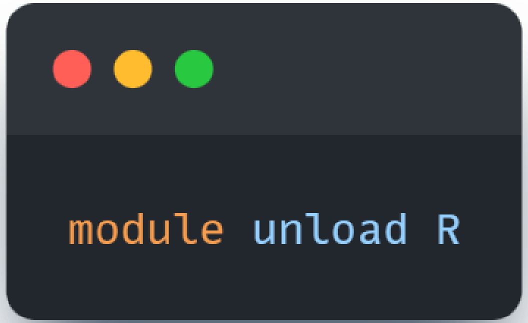
```
module avail
```

Load module



```
module load R
```

Unload module



```
module unload R
```

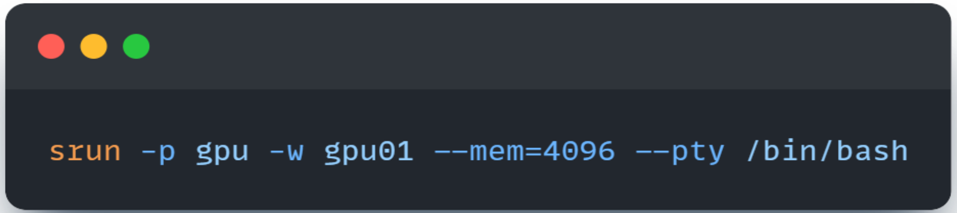
Slurm usage

How to use Slurm

- *`srun`*
  - Use it to execute your script without detaching from the job.
- *`sbatch`*
  - Use it to submit your job to cluster and detach from it.
- *`scancel`*
  - Cancel your job with the job id you get from squeue.
- *`squeue`*
  - List jobs that are currently in the queue.

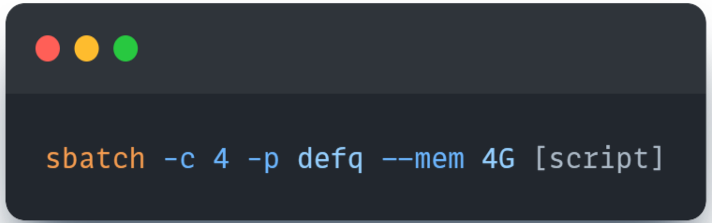
How to use Slurm

- srun usage:



```
srun -p gpu -w gpu01 --mem=4096 --pty /bin/bash
```

- sbatch usage:

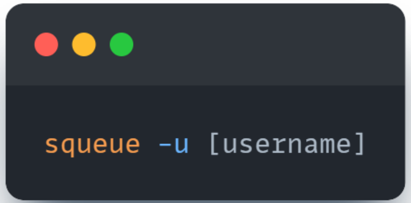


```
sbatch -c 4 -p defq --mem 4G [script]
```

- scancel usage:

```
scancel [job-id]
```

- squeue usage



```
squeue -u [username]
```

Billing

- Every group is billed via the declaration number (kostenplaats nummer) used on slurm jobs.
- The cost is calculated based on cluster use.

Resources

1. HPC Wiki
2. Email: hpc-support@amsterdamumc.nl
3. Servicenow ticket

   For HPC support questions, please refer to the IT servicedesk and mention "HPC support".