InfraLife

PReSTO – structural biology software in high perfomance compute environment

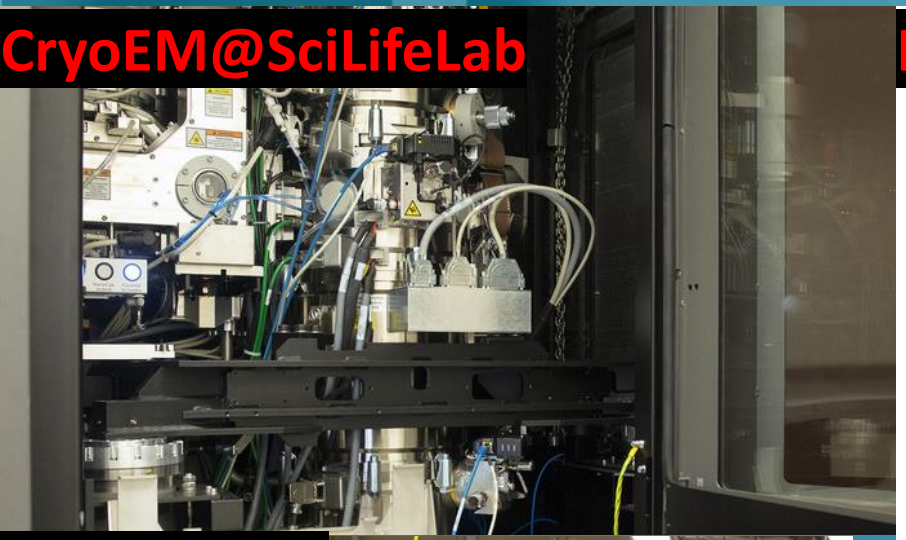
Integrative Structural Biology Course, 22 Aug 2022

Martin Moche

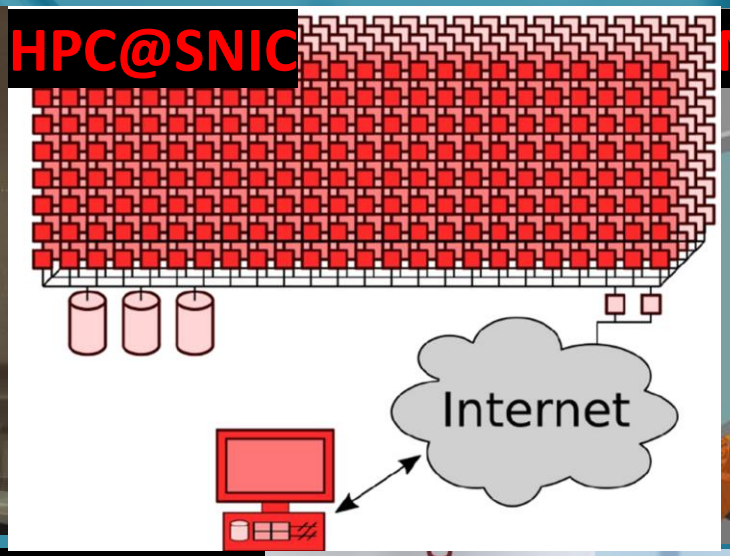
PReSTO ”Integrated Structural Biology computing”

The PReSTO project installs, adapt and test structural biology software for use in Swedish high-performance compute (HPC) environments

CryoEM@SciLifeLab



HPC@SNIC



NMX&ESS

MX@MAXIV

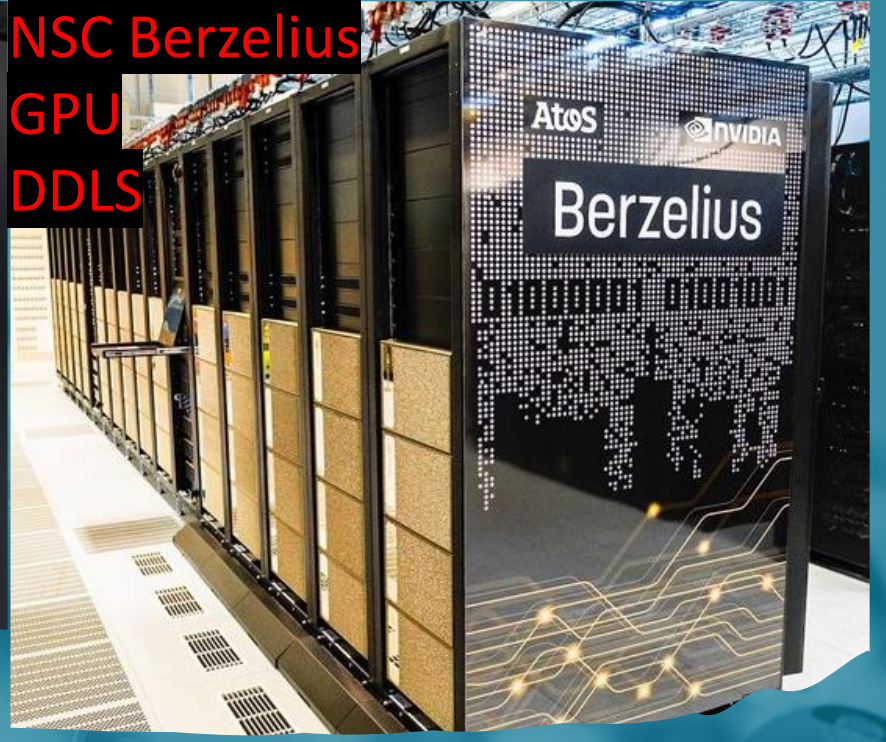
NMR@SNC

XFEL@(MAXIV)EU

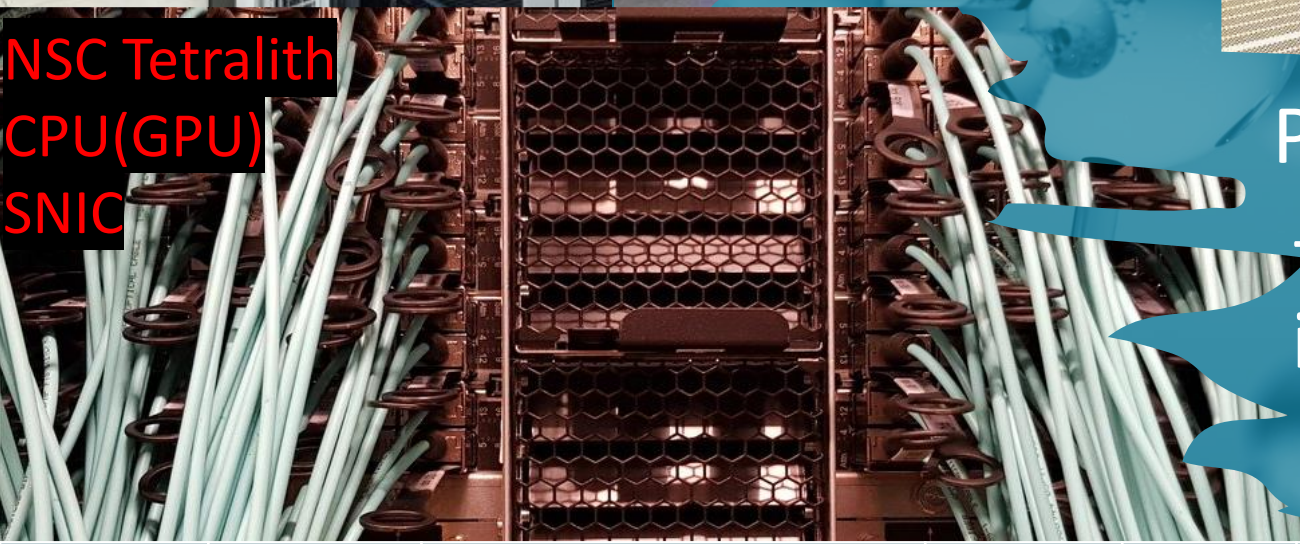
MAXIV
CPU
Local

LUNARC Aurora
CPU(GPU)
SNIC

NSC Berzelius
GPU
DDLS



NSC Tetralith
CPU(GPU)
SNIC



PReSTO HPC hardware in 2022
- GitLab and Easybuild for software installation transfer across hardware

NSC	Nodes	CPU type	CPUs	RAM	GPU	/scratch/local
Tetralith	170	2x Intel Xeon Gold 6130	32	96 GiB	1 NVIDIA® T4	2 TB NVMe SSD
Berzelius	60	2 AMD Epyc™ 7742	128	1 TB	8 NVIDIA® A100 Tensor Core	15 TB NVMe SSD

PReSTO history

2013

National Supercomputer Centre (NSC) visits Tällberg

2015

PReSTO gets MAXIV satellite status

2016

Software deployment tool Easybuild introduced

- Hardware transfer simplified

2018

Funding from Swedish Research Council (ends 2023)

- To involve Cryo-EM, NMR(X) and XFEL (serial crystallography)

2020

SNIC storage project

2021

NSC Berzelius

- CryoEM feasible

PReSTO@MAXIV

PReSTO for FragMAXapp

Online
Clu0-fe-1

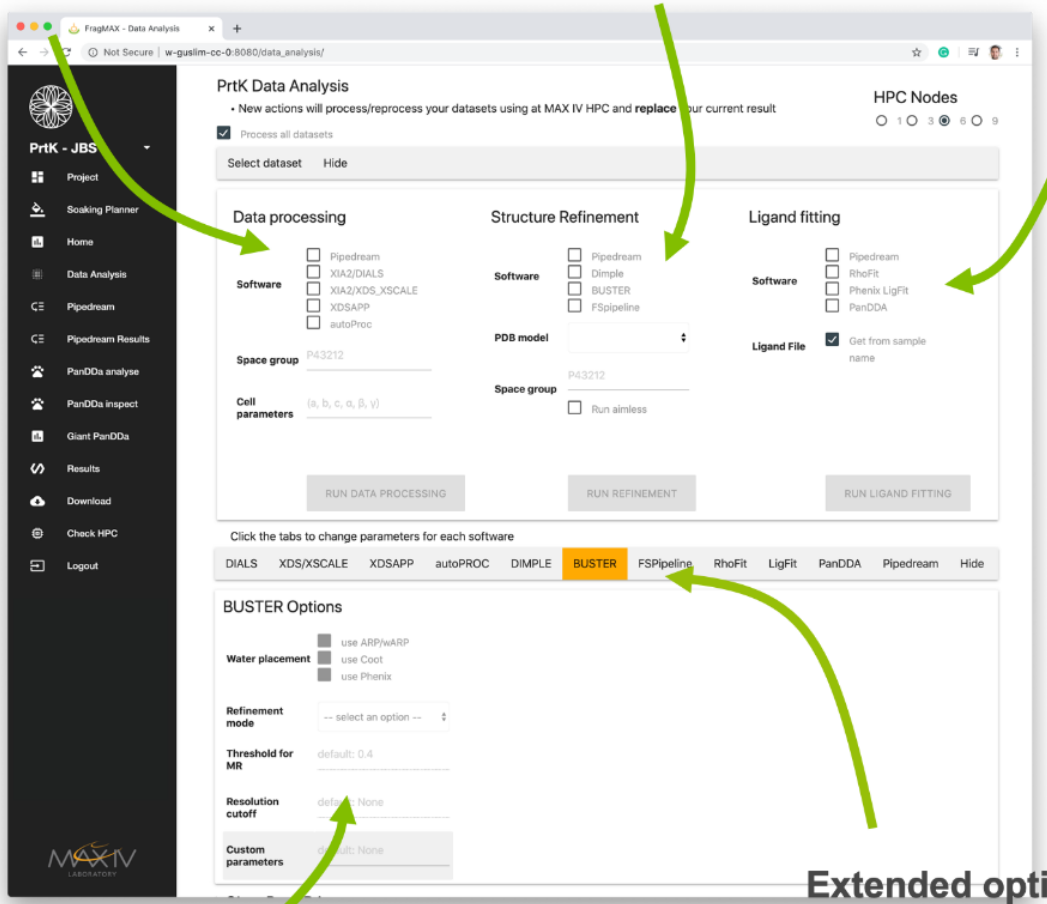
Offline
Offline-fe1

maxiv.lu.se/fragmax/fragmaxapp/

Data processing
software selection

Structure refinement using three automated pipelines based on different methods

Automatic ligand fitting
for each available refined structure



Extended options for refinement, water placement and blobs search

Custom definitions for each software applied to all datasets

Data processing takes place at
MAX IV Cluster

PReSTO@MAXIV

PReSTO for FragMAXapp

Online
Clu0-fe-1

Offline
Offline-fe1

Electron density
and structure viewer on the browser

Navigation tools for unexplained blobs and other structures

Information about the model and ligand fitting scores

Display control for ligand poses

Dual viewer for improved binding mode comparison

Ligand representation expected in the current dataset

Reciprocal lattice viewer

maxiv.lu.se/fragmax/fragmaxapp/

PReSTO in SB community 2022

	MX(NMX)	XFEL	CryoEM	NMR
infrastructure	MAXIV(ESS)	MAXIV	SciLifeLab	SNC
community	YES	YES	YES	yes
reason	available	available	data transfer	too few sw

User support

SNIC support - https://supr.snic.se/support/

login, using installed software/storage, SUPR, request software installation

PReSTO page: https://www.nsc.liu.se/support/presto/

Designed to support structural biology software use in HPC environment

No software manuals! Developers made these already

Phenix GUI slurm configuration

Known limitations (workarounds, issues, features)

Sub-page for MX, XFEL, NMR, Cryo-EM, NMX

List of software and releases in PReSTO

Acknowledgements and people involved

Access…

PReSTO training

- Upcoming InfraLife course practical´s
- Additional NMR, NMX, XFEL and CryoEM training once used by community
- For additional MX? contact PSF martin.moche@ki.se
  - Require your own dataset and account on PReSTO resource
  - Several 2-hour zoom sessions (1-5 participants)
  - MX curriculum:
    - https://www.nsc.liu.se/support/presto/MX-PReSTO%20training/

How to support PReSTO?

1. Use PReSTO! Report issues and suggestions to SNIC support
2. Suggest software to be added into PReSTO - PyMOL plugins, CryoEM software
3. Perform software updates (require PReSTO admin rights)

   Free up time for PReSTO staff to deal with new installs and issues
4. Add new software to PReSTO (require PReSTO admin rights)
5. Community representatives to be teached to perform 3 and 4

Trello for PReSTO management

Trello | Arbetsytor | Senaste | Stjärnmärkt | Mallar | Skapa

PReSTO
Gratis

- Tavlor
- Medlemmar
- Inställningar

Visningar av Arbetsyta
- *Tabell*
- *Kalender*

Dina tavlor
- PReSTO: Issues and work p...
- PReSTO: Cryo-EM software que...
- PReSTO: NMR
- PReSTO: Work packages on hold
- Project_meetings
- Start up
- WP3 - software platform at Ma...
- Visa mindre

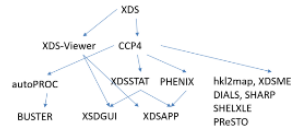
Tavla | **PReSTO: Issues and work packages** | PReSTO | Synlig Arbetsyta | +18 | Dela | Bevakar | Kalender Power-Up | Power-Ups

Information

Instruction on how to use this Trello board

MAX IV
MAX IV messages
46 6

MX software dependencies



MX sw dependencies
1 1

SHELXE mot multi-CPU
1 2

Version tracking
30 mar 4

SX data at MAX IV
7

SX data at Tetralith

Not started

MAX IV
Make generate_XDS.INP point to correct path for neggia/durin plugins

Suggestion
Install Arcimboldo
1 2

Bug
dials.image_viewer does not work in vglrun
4

MAX IV Bug in upstream software
Error loading pandda.inspect on gn/fe with missing numpy dep
7 1 0/3

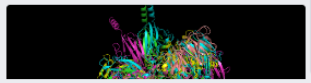
MAX IV
Add Cheetah GUI
9

MAX IV
Make forkxds issue a warning instead of shutting down when oversubscribing a node
1

Make xxdiff a dependency for XDSGUI

MAX IV Suggestion
Install PAIREF

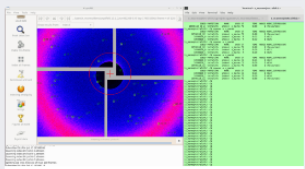
Contact XDSAPP3 developers



+ Lägg till ett kort

In progress

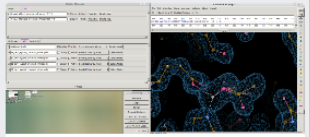
MAX IV
try MAX IV jupyterhub
3 1



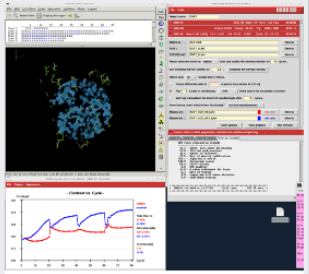
Adapt CrystFEL Slurm support to PReSTO?
16 8

xia2 -3dii fail with firmware upgraded Eiger@BioMAX
3 3

CrystFEL 0.10.1 GUI HDF5 read errors
3



COOT/refmac5 (CCP4) BUG report outside PReSTO
2 2



Coot does not launch from hkl2map

+ Lägg till ett kort

Done but not in repo (if applicable)

MAX IV Bug
crystfel - libslurm
5

compute node interactive software(XDSGUI, XDSAPP3 and hkl2map) does not start from presto menu at Berzelius
5

Reinstall GlobalPhasing sw to update .licence

Install DIALS 3.9.1 at NSC Tetralith
2

XDS help programs XSCALE_ISOCLUSTER and spot2pdb are missing at MAX IV
11 2

Install CCP4 8.0.001 at NSC Tetralith
1

Update autoPROC to 2022-06-08
1

Update generate_XDS.INP to 1.13
4

Update XDSSTAT to 20210827
3

update XDSAPP3 v 1.6 at Tetralith
2

+ Lägg till ett kort

Queued for 7.2

+ Lägg till ett kort

Released

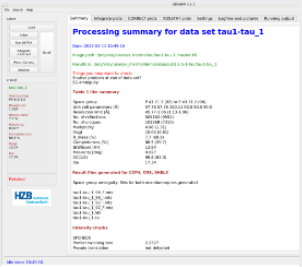
PReSTO menu does not work at NSC Tetralith for interactive jobs?
5

Stop importing 'os' in easyconfigs and start using hooks instead
3

COOT does not run as intended at Aurora
8

Add new nXDS to Tetralith & MAX IV
5

Suggestion Bug
BioMAX firmware update effect XDSAPP3?
11 3



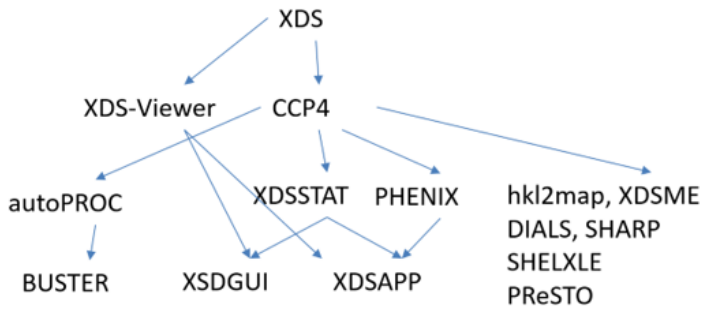
MAX IV Bug in upstream software
Send PR to EB upstream before a...
XDS default version expires 23 mars
22 2

New ADXV version 1.15 to NSC Tetralith
2

Install ShelXle 1370 at NSC Tetralith

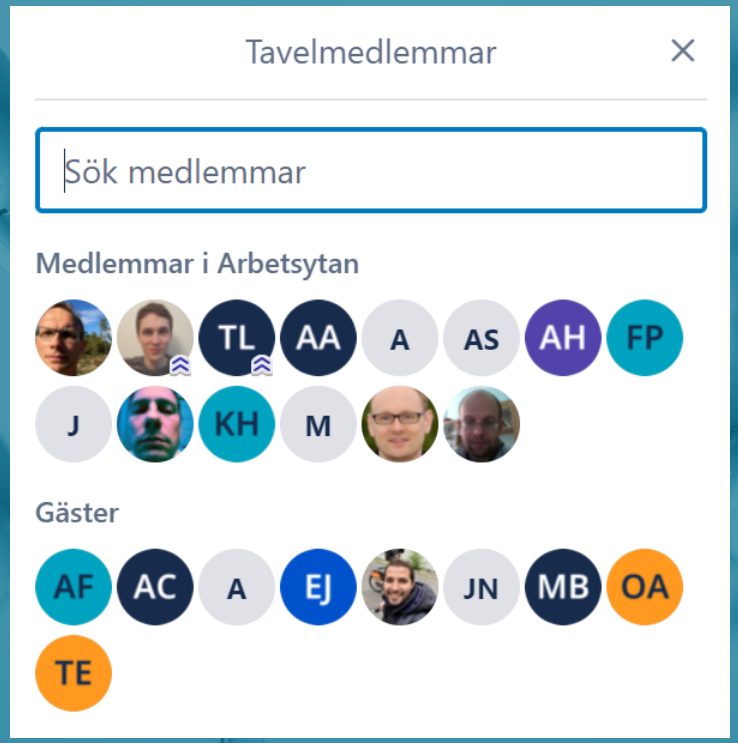
+ Lägg till ett kort

MX software dependencies



PReSTO team acknowledgement

Sebastian Thorarensen
Karl Hörnell
Torben Rasmussen
Torbjörn Lönnemark
Filip Polbratt
Anton Haglund
Johan Raber



Zdenek Matej
Jie Nan
Oskar Aurelius
Aaron Finke
Elmir Jagudin
Thomas Eriksson
Ana Gonzales
Aleksander Cehovin
Johan Unge
Gustavo Lima
Anastasia Shilova

Anders Sjöström
Anders Follin

Michael Hall
Stefan Fleichmann
Jose Miguel de la Rosa Trevin
Björn Forsberg

Maria Sunnerhagen
Alexandra Ahlner

THANK YOU SPONSORS

EUROPEAN SPALLATION SOURCE

SciLifeLab

MAX IV

The research-based pharmaceutical industry

Swedish Research Council

CryoNET

xenocs
Exploring the very small

NBIS

excillum

BRUKER

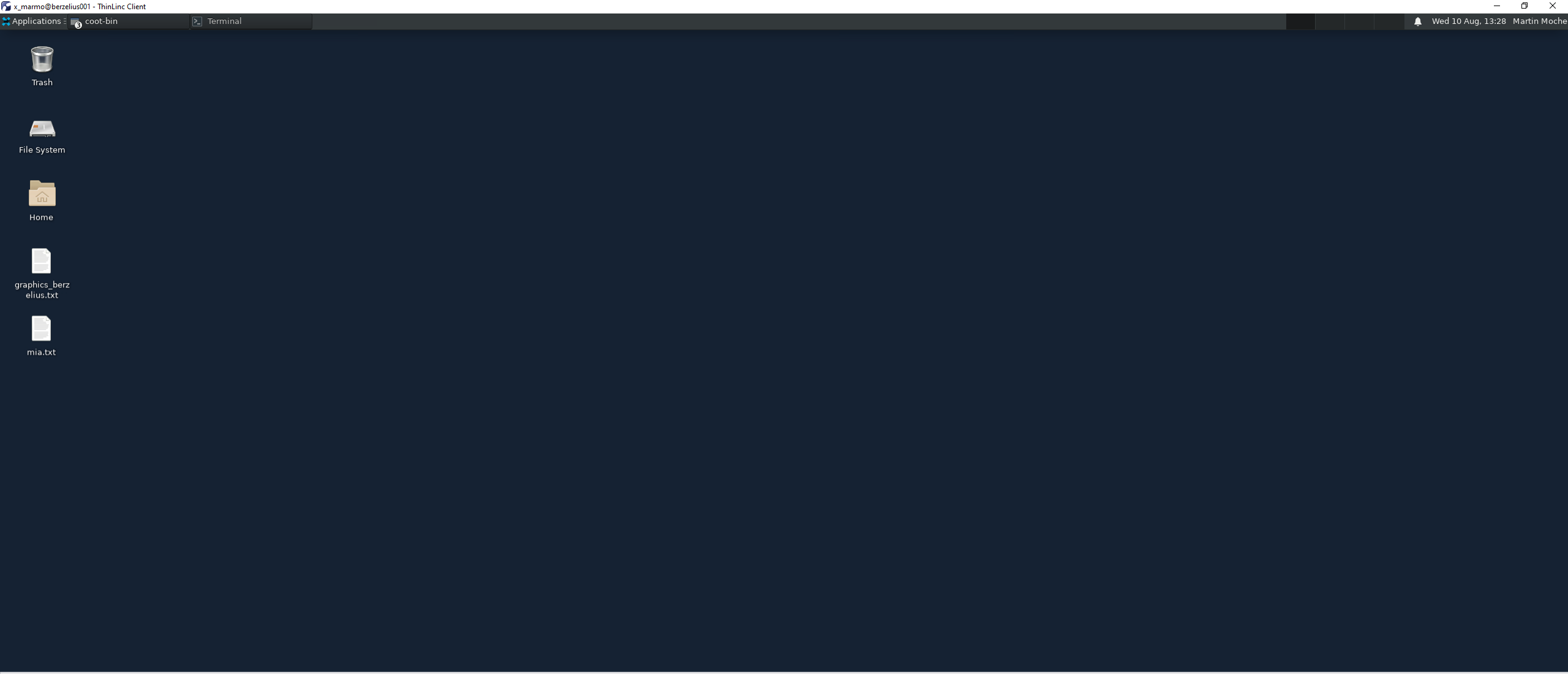
SwedNMR

LINXS

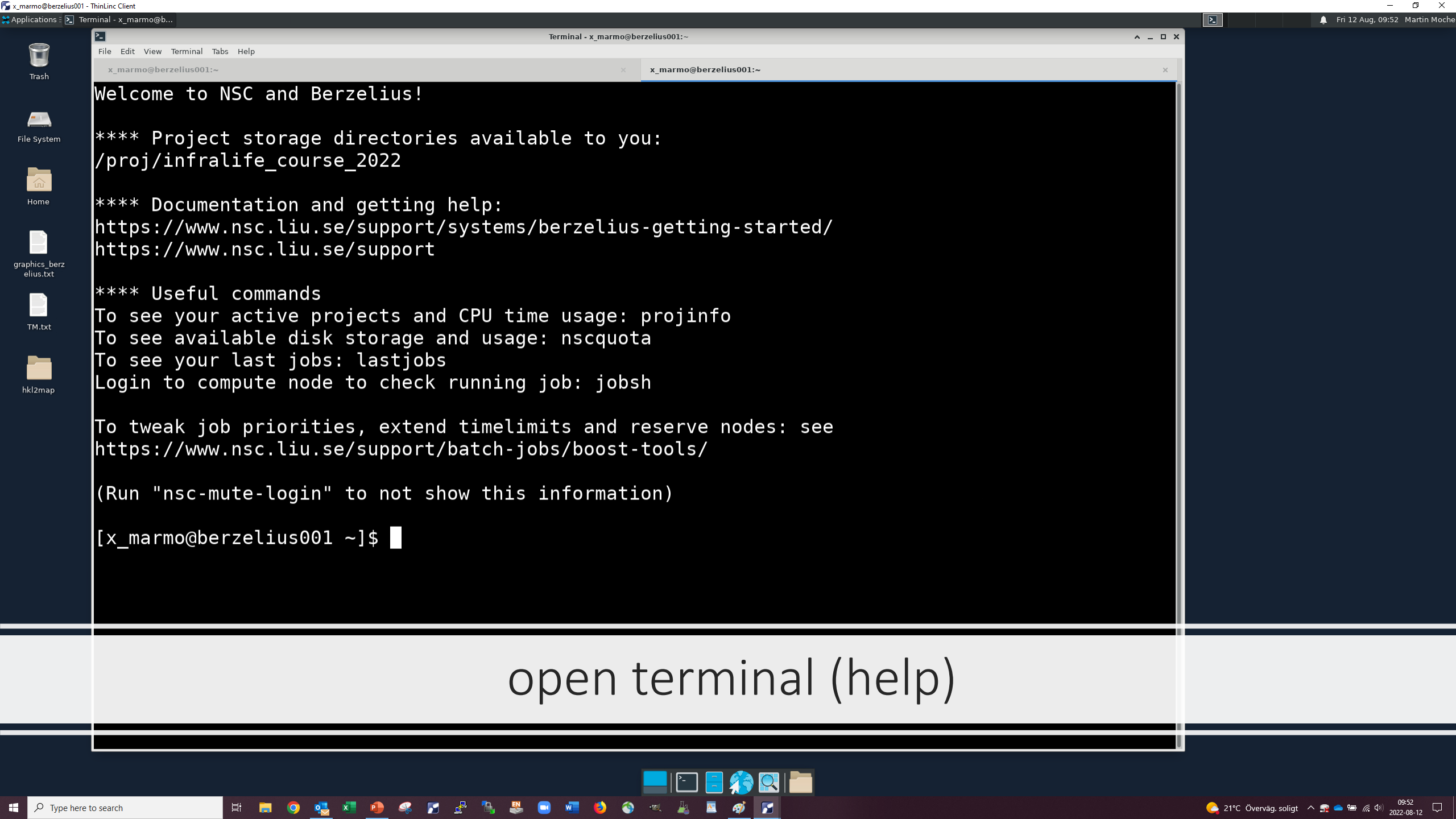
ThermoFisher SCIENTIFIC

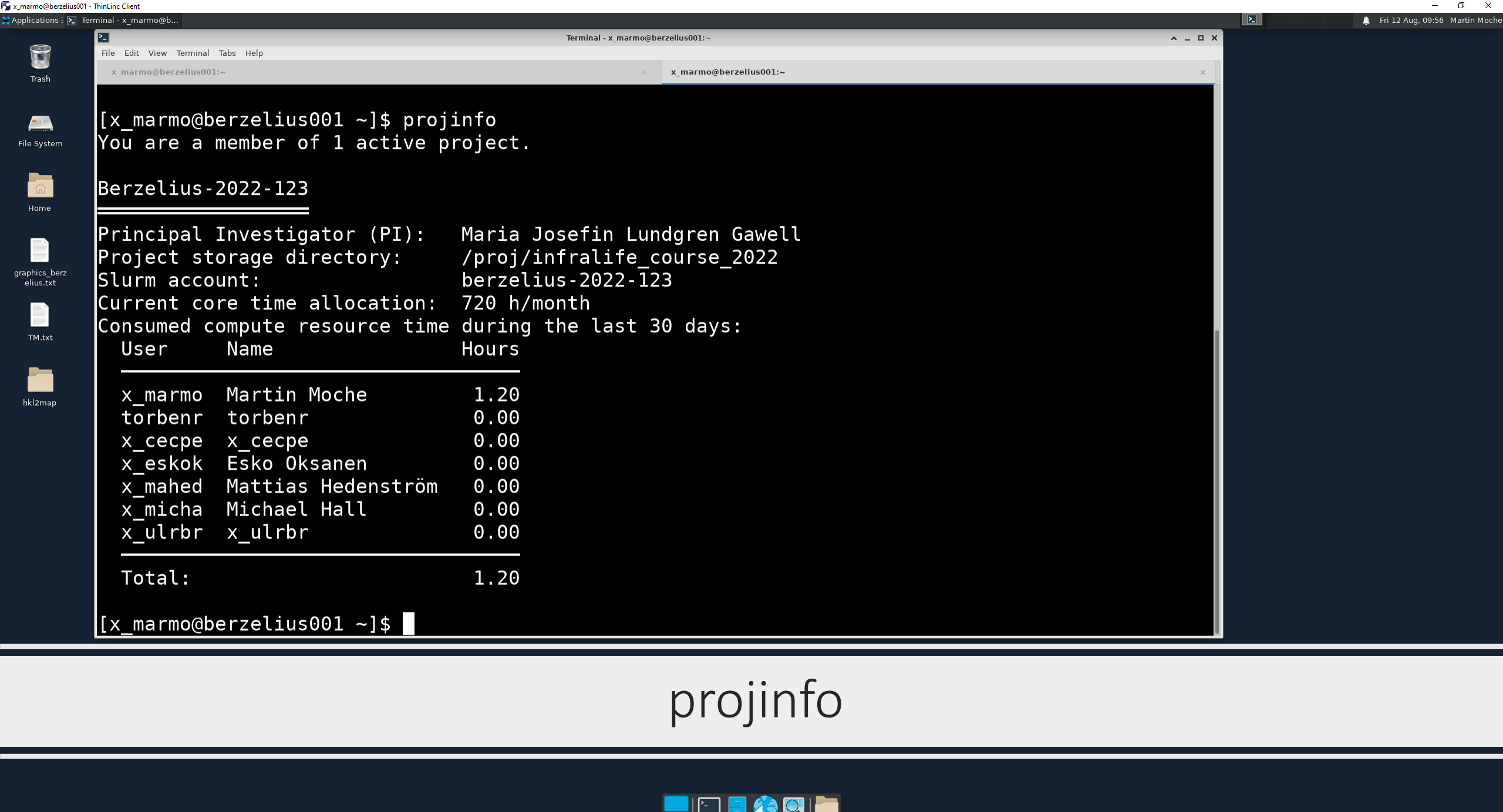
InfraLife

Live DEMO start

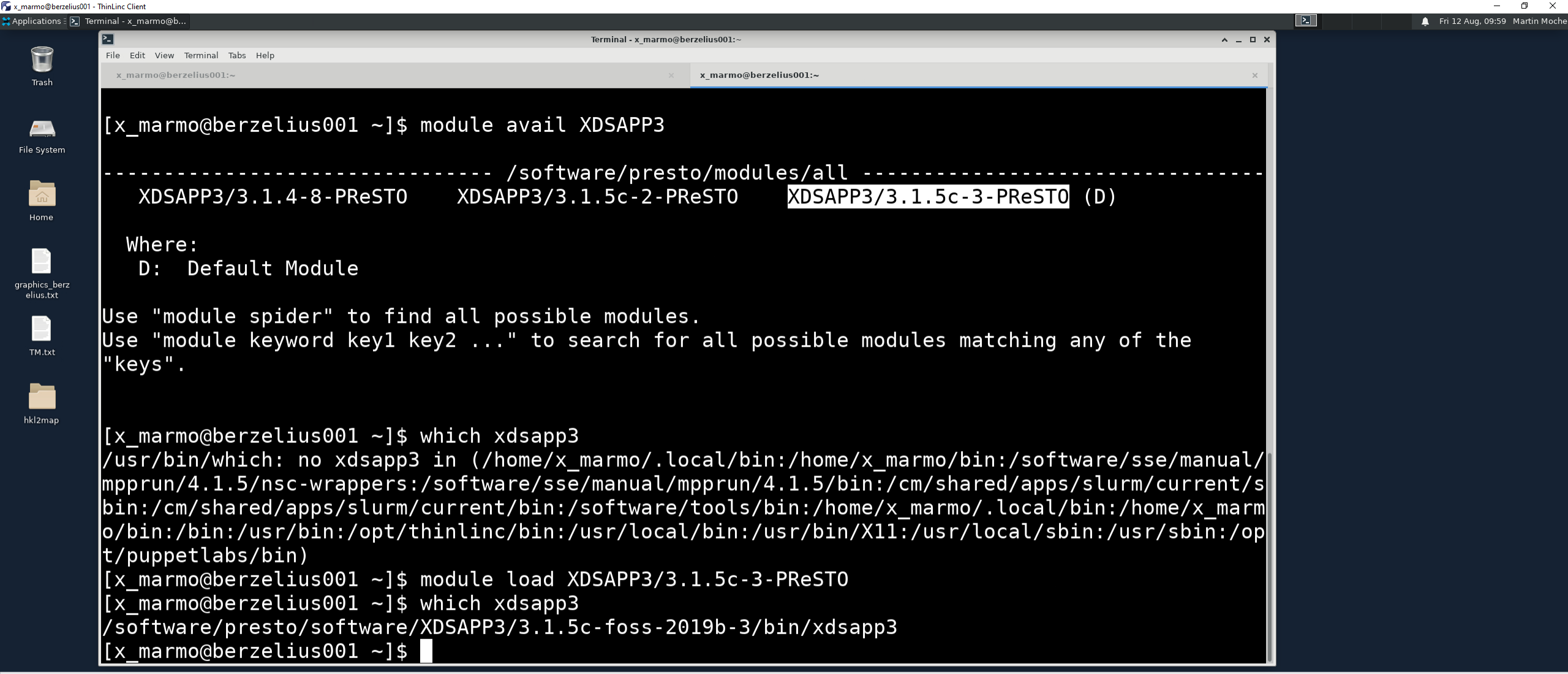


Remote desktop at Berzelius login node

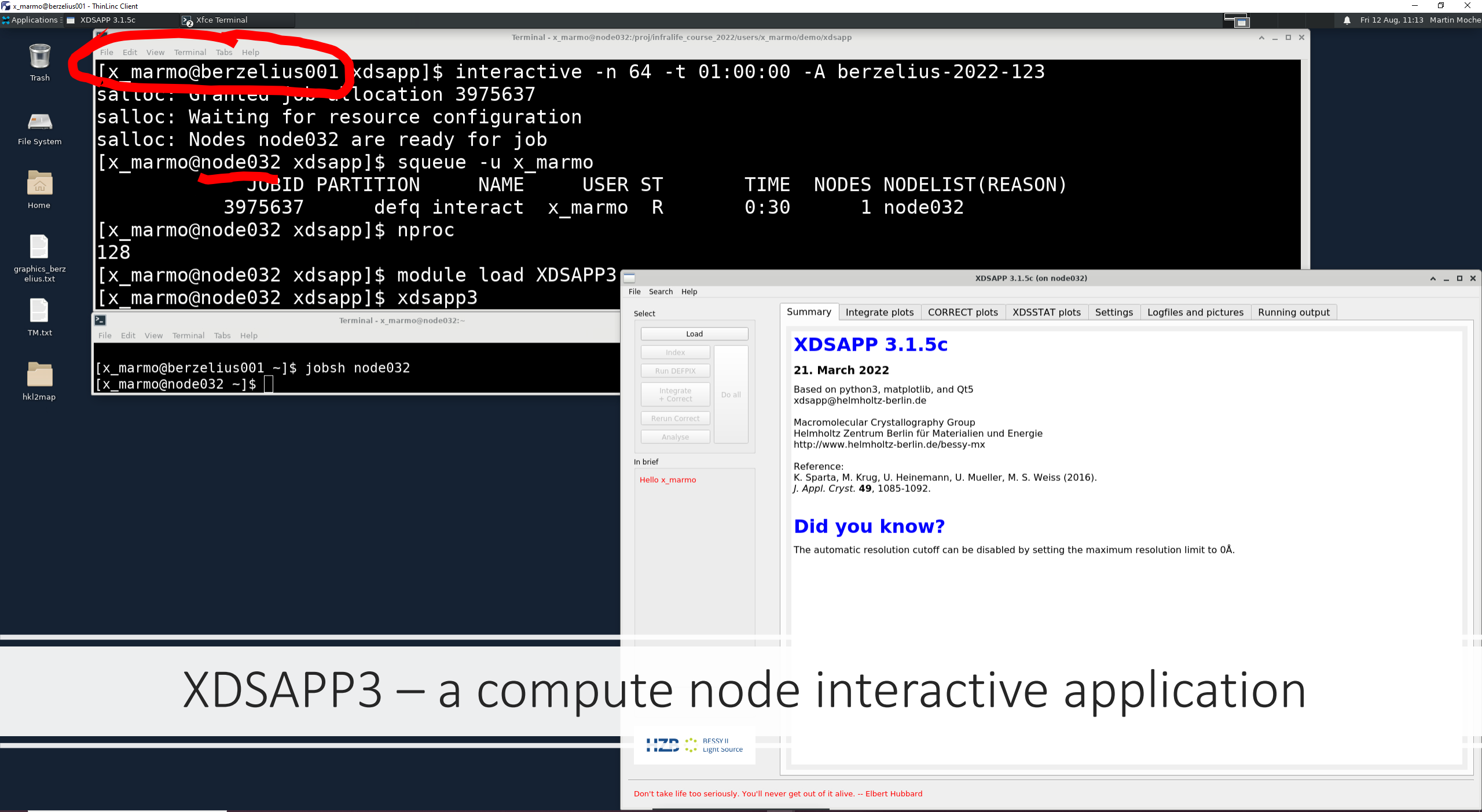

x_marmo@berzelius001 - ThinLinc Client
Applications
Terminal - x_marmo@b...
Fri 12 Aug, 09:52 Martin Moche
Trash
File System
Home
graphics_berzelius.txt
TM.txt
hkl2map
Terminal - x_marmo@berzelius001:~
File Edit View Terminal Tabs Help
x_marmo@berzelius001:~
x_marmo@berzelius001:~
Welcome to NSC and Berzelius!
**** Project storage directories available to you:
/proj/infralife_course_2022
**** Documentation and getting help:
https://www.nsc.liu.se/support/systems/berzelius-getting-started/
https://www.nsc.liu.se/support
**** Useful commands
To see your active projects and CPU time usage: projinfo
To see available disk storage and usage: nscquota
To see your last jobs: lastjobs
Login to compute node to check running job: jobsh
To tweak job priorities, extend timelimits and reserve nodes: see
https://www.nsc.liu.se/support/batch-jobs/boost-tools/
(Run "nsc-mute-login" to not show this information)
[x_marmo@berzelius001 ~]$
open terminal (help)
Type here to search
21°C Överväg. soligt
09:52
2022-08-12


x_marmo@berzelius001 - ThinLinc Client
Applications
Terminal - x_marmo@b...
Fri 12 Aug, 09:56 Martin Moche
Trash
File System
Home
graphics_berzelius.txt
TM.txt
hkl2map
Terminal - x_marmo@berzelius001:~
File Edit View Terminal Tabs Help
x_marmo@berzelius001:~
x_marmo@berzelius001:~
[x_marmo@berzelius001 ~]$ projinfo
You are a member of 1 active project.
Berzelius-2022-123
Principal Investigator (PI): Maria Josefin Lundgren Gawell
Project storage directory: /proj/infralife_course_2022
Slurm account: berzelius-2022-123
Current core time allocation: 720 h/month
Consumed compute resource time during the last 30 days:
User Name Hours
x_marmo Martin Moche 1.20
torbenr torbenr 0.00
x_cecpe x_cecpe 0.00
x_eskok Esko Oksanen 0.00
x_mahed Mattias Hedenström 0.00
x_micha Michael Hall 0.00
x_ulrbr x_ulrbr 0.00
Total: 1.20
[x_marmo@berzelius001 ~]$
Type here to search
21°C Överväg. soligt
09:56
2022-08-12

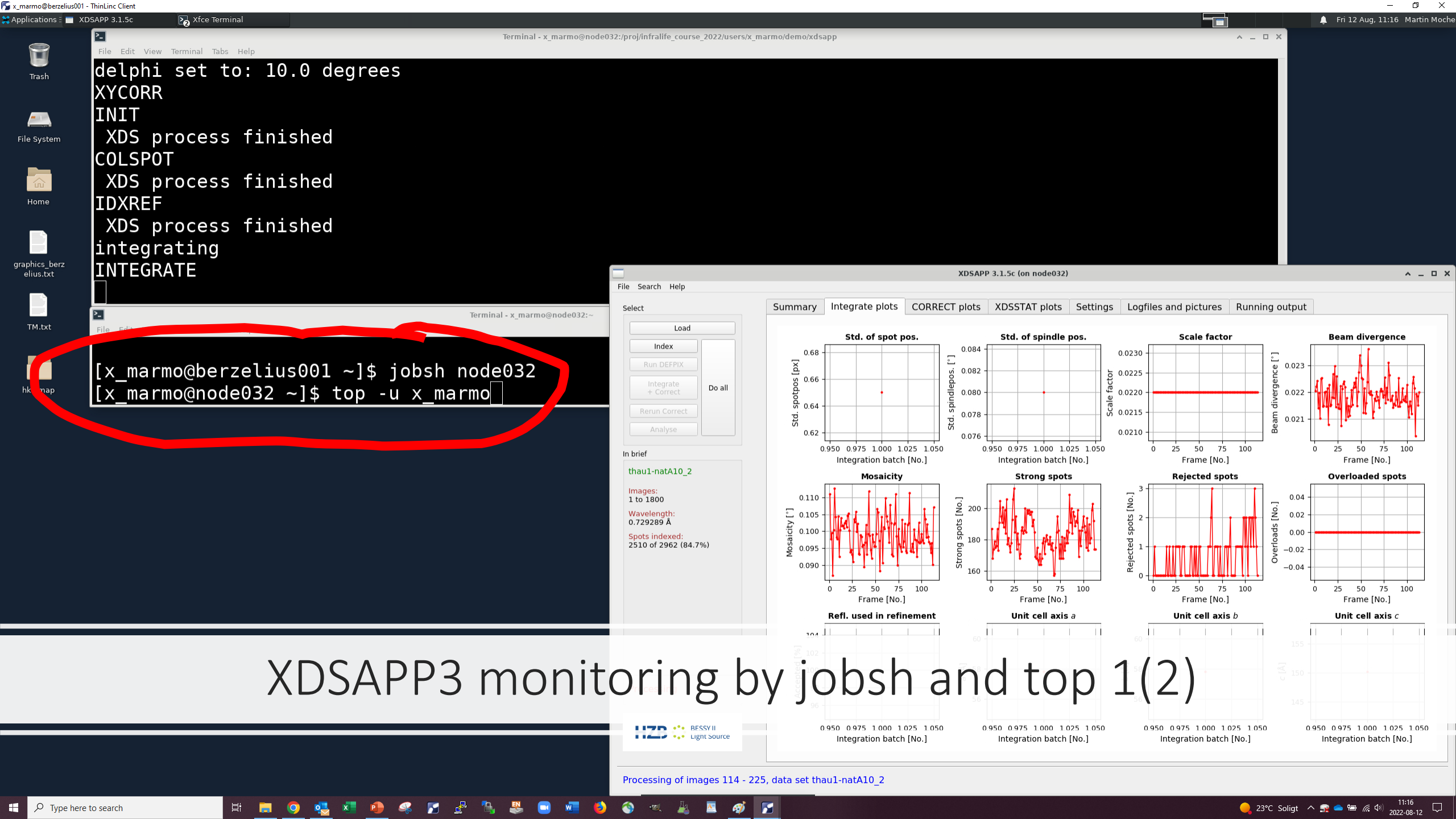
projinfo



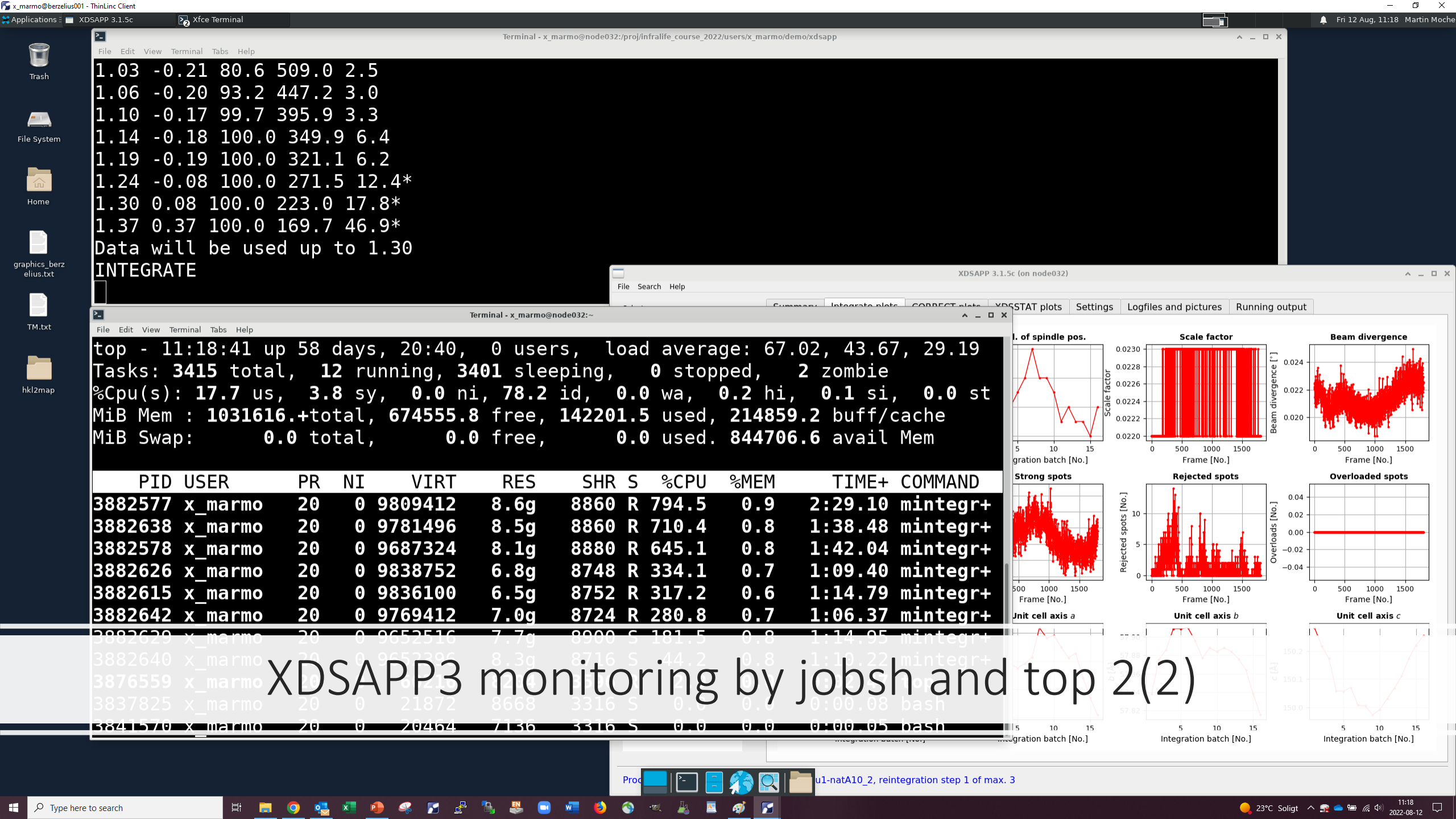
module avail and module load

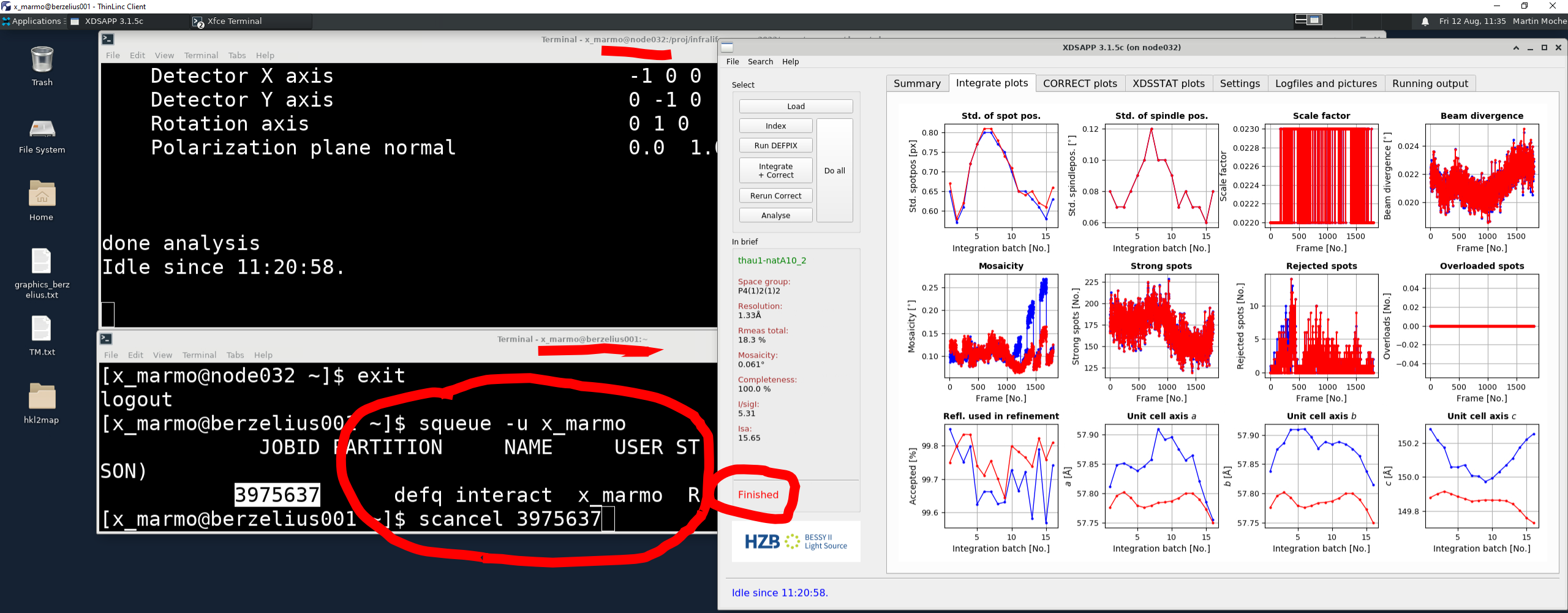


XDSAPP3 – a compute node interactive application



XDSAPP3 monitoring by jobsh and top 1(2)


x_marmo@berzelius001 - ThinLinc Client
Applications
XDSAPP 3.1.5c
Xfce Terminal
Fri 12 Aug, 11:18 Martin Moche
Terminal - x_marmo@node032:/proj/infralife_course_2022/users/x_marmo/demo/xdsapp
File Edit View Terminal Tabs Help
1.03 -0.21 80.6 509.0 2.5
1.06 -0.20 93.2 447.2 3.0
1.10 -0.17 99.7 395.9 3.3
1.14 -0.18 100.0 349.9 6.4
1.19 -0.19 100.0 321.1 6.2
1.24 -0.08 100.0 271.5 12.4*
1.30 0.08 100.0 223.0 17.8*
1.37 0.37 100.0 169.7 46.9*
Data will be used up to 1.30
INTEGRATE
Trash
File System
Home
graphics_berzelius.txt
TM.txt
hkl2map
Terminal - x_marmo@node032:~
File Edit View Terminal Tabs Help
top - 11:18:41 up 58 days, 20:40, 0 users, load average: 67.02, 43.67, 29.19
Tasks: 3415 total, 12 running, 3401 sleeping, 0 stopped, 2 zombie
%Cpu(s): 17.7 us, 3.8 sy, 0.0 ni, 78.2 id, 0.0 wa, 0.2 hi, 0.1 si, 0.0 st
MiB Mem : 1031616.+total, 674555.8 free, 142201.5 used, 214859.2 buff/cache
MiB Swap: 0.0 total, 0.0 free, 0.0 used. 844706.6 avail Mem
PID USER PR NI VIRT RES SHR S %CPU %MEM TIME+ COMMAND
3882577 x_marmo 20 0 9809412 8.6g 8860 R 794.5 0.9 2:29.10 mintegr+
3882638 x_marmo 20 0 9781496 8.5g 8860 R 710.4 0.8 1:38.48 mintegr+
3882578 x_marmo 20 0 9687324 8.1g 8880 R 645.1 0.8 1:42.04 mintegr+
3882626 x_marmo 20 0 9838752 6.8g 8748 R 334.1 0.7 1:09.40 mintegr+
3882615 x_marmo 20 0 9836100 6.5g 8752 R 317.2 0.6 1:14.79 mintegr+
3882642 x_marmo 20 0 9769412 7.0g 8724 R 280.8 0.7 1:06.37 mintegr+
XDSAPP 3.1.5c (on node032)
File Search Help
Summary
Integrate plots
CORRECT plots
XDSTAT plots
Settings
Logfiles and pictures
Running output
Scale factor
Beam divergence
Strong spots
Rejected spots
Overloaded spots
Unit cell axis a
Unit cell axis b
Unit cell axis c
Frame [No.]
Integration batch [No.]
Scale factor
Beam divergence [°]
Rejected spots [No.]
Overloads [No.]
XDSAPP3 monitoring by jobsh and top 2(2)
Type here to search
23°C Soligt
11:18
2022-08-12



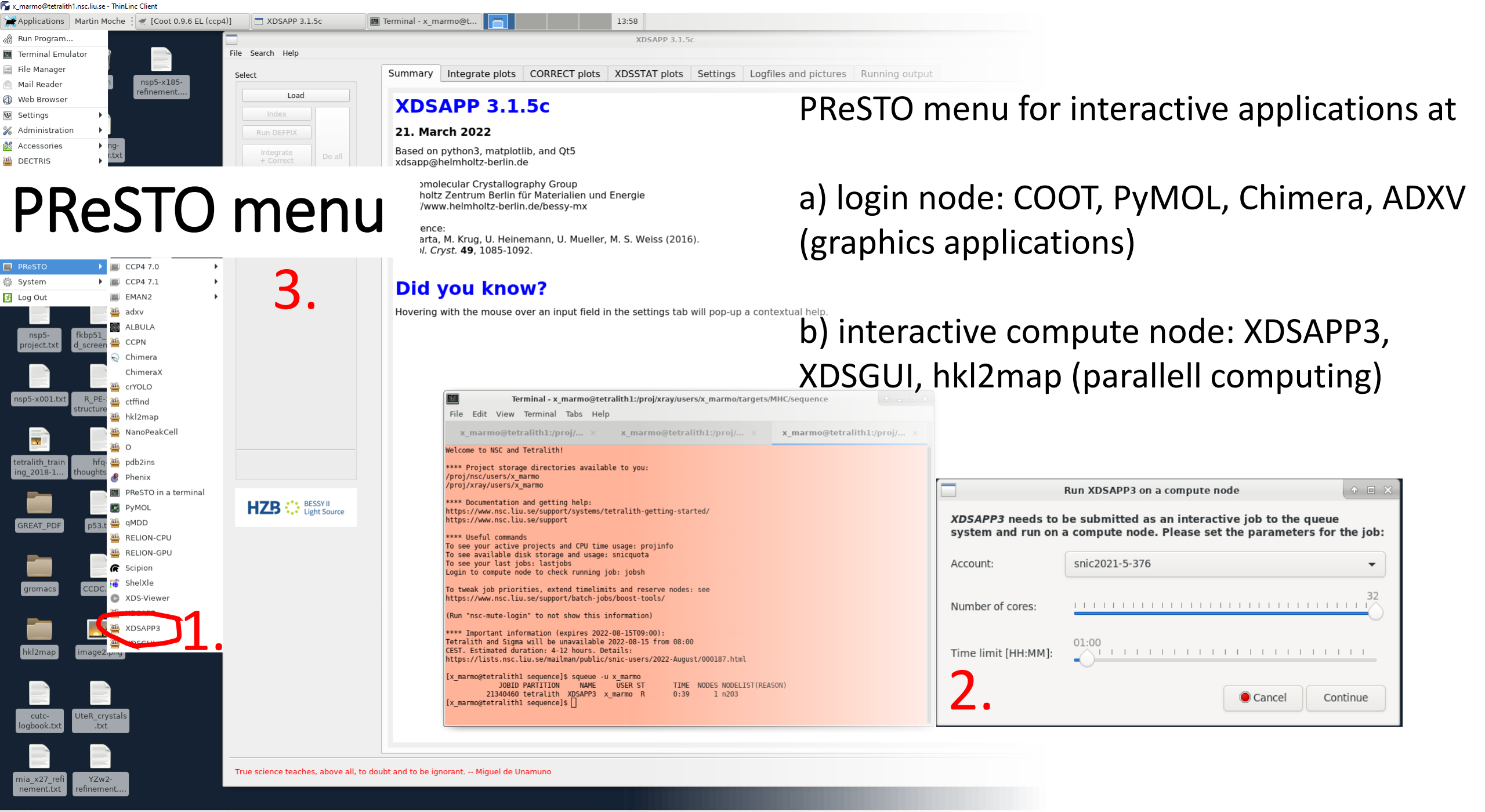
scancel or exit to save compute time

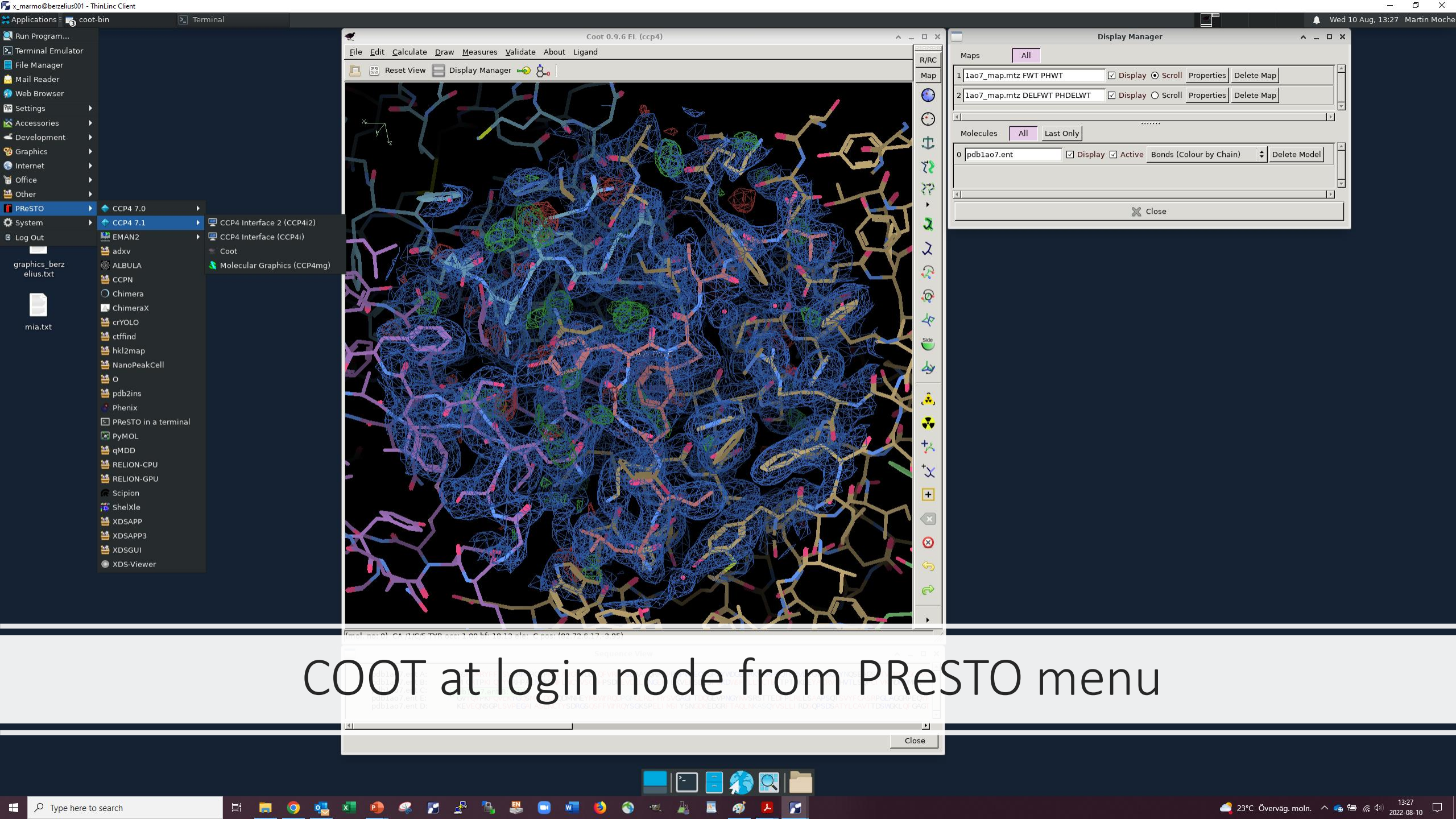
PReSTO menu

PReSTO menu for interactive applications at

a) login node: COOT, PyMOL, Chimera, ADXV (graphics applications)

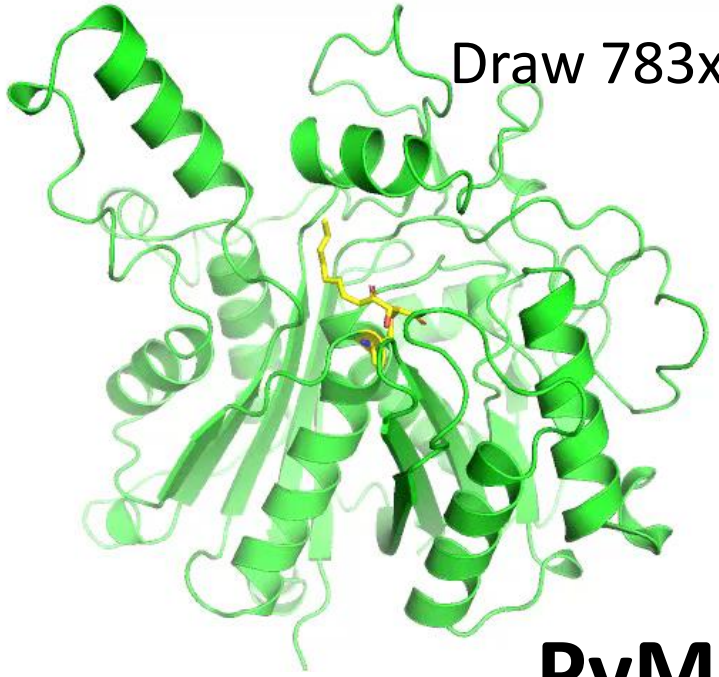
b) interactive compute node: XDSAPP3, XDSGUI, hkl2map (parallell computing)



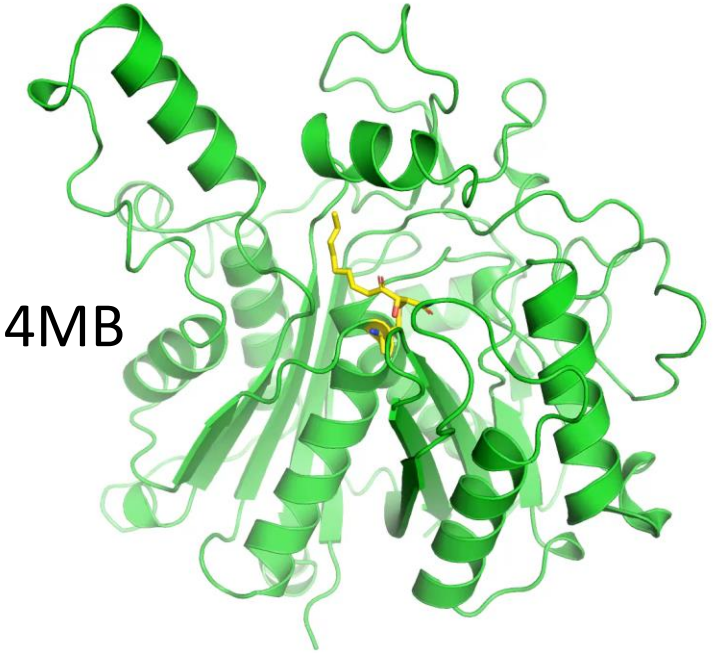


COOT at login node from PReSTO menu

Draw 783x529 14MB

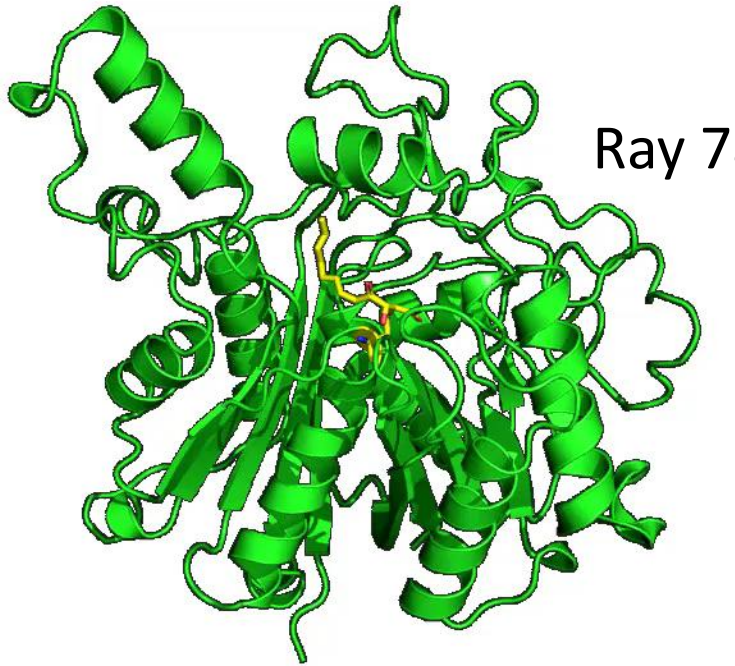


Draw 1566x1058 34MB

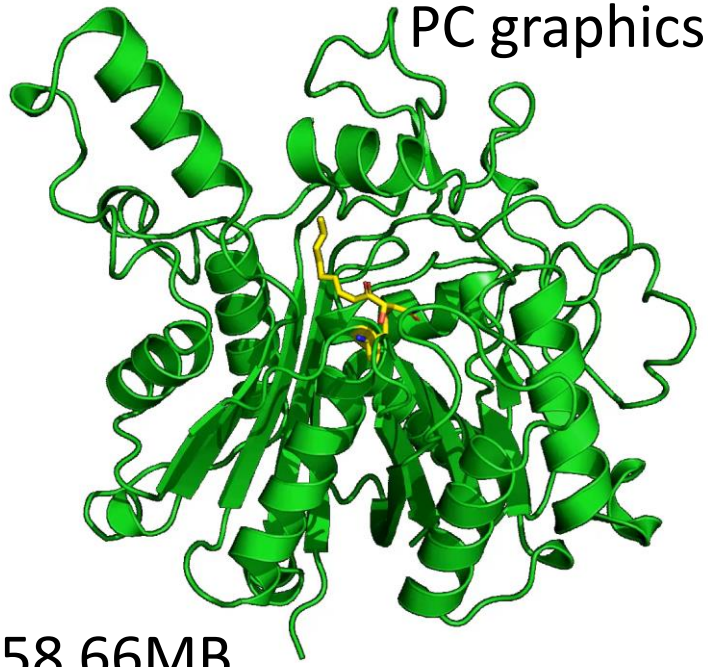


PyMOL movies for PowerPoint

Ray 783x529 29MB



PC graphics card...



Ray 1566x1058 66MB

Live DEMO end

Berzelius test-login

- Download and install thinlinc from https://www.cendio.com/thinlinc/download
- Install ”Google authenticator” on your phones
- Can you login to Berzelius?
- Can you see the PReSTO menu?

THE END
Enjoy the course!

Login and compute nodes for MX software

Use login node for:

1. Interactive graphics software: PyMOL, COOT, Chimera, ADXV, ALBULA
2. GUIs doing non-parallel computing: ccp4i, ccp4i2
   PHASER is parallel => ccp4i, ccp4i2 at interactive compute node
3. GUIs with SLURM: Phenix, CrystFEL

Use interactive compute node for

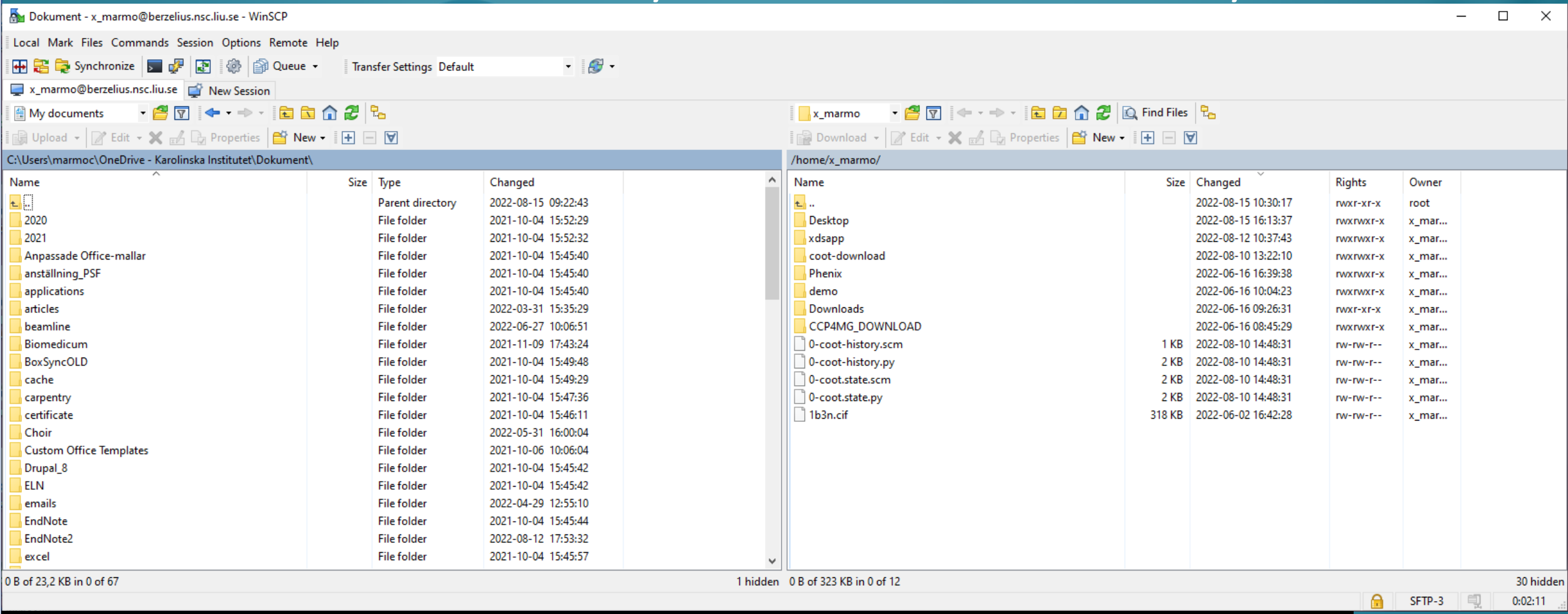
1. XDSAPP3, hkl2map, XDSGUI

i.e. software running in parallel mode once started from GUI

Use sbatch script at compute node for

1. BUSTER, XDS, DIALS, autoPROC, SHARP, archimboldo

File transfer by WinSCP, filezilla, rsync




```
rsync -rvplt x_marmo@berzelius.nsc.liu.se:/proj/infralife_course_2022/users/x_marmo/ .
```

Ways to transfer large amount of data and many files?

Things not discussed in this presentation

- Easybuild structure
  - easyconfigs, easyblocks
- sbatch script examples
- How to write compute and storage applications to SNIC

PReSTO acknowledgements

Lima, G.M.A. et al. FragMAXapp: crystallographic fragment-screening data-analysis and project-management system. *Acta Crystallogr D Struct Biol* **77**, 799-808 (2021).

Anandapadamanaban, M. et al. E3 ubiquitin-protein ligase TRIM21-mediated lysine capture by UBE2E1 reveals substrate-targeting mode of a ubiquitin-conjugating E2. *J Biol Chem* **294**, 11404-11419 (2019).

Anandapadamanaban, M. et al. Mutation-Induced Population Shift in the MexR Conformational Ensemble Disengages DNA Binding: A Novel Mechanism for MarR Family Derepression. *Structure* **24**, 1311-1321 (2016).

Brock, J.S. et al. A dynamic Asp-Arg interaction is essential for catalysis in microsomal prostaglandin E2 synthase. *Proc Natl Acad Sci U S A* **113**, 972-7 (2016).

Janfalk Carlsson, A. et al. Laboratory-Evolved Enzymes Provide Snapshots of the Development of Enantioconvergence in Enzyme-Catalyzed Epoxide Hydrolysis. *Chembiochem* (2016).