

# Questions for the course "Introduction to running R, Python, and Julia in HPC", 17-19 October 2023

TUESDAY, 17 October 2023, Python

Please note the following:

- Add your questions below, numbering them continuously.
- Be careful in case someone else is writing at the same time.
- Please **DO NOT** delete your questions even when they have been answered, as we are planning to use them to improve our material.
- Do not share any sensitive information as this document is accessible to anyone with the correct link.
- You can start a new line within a question by pressing <SHIFT> + <ENTER>.

1. General information / summary:
2. Hostnames for login nodes (the main course project is at UPPMAX):
  - a. UPPMAX – rackham
    - i. **SSH:** rackham.uppmax.uu.se
    - ii. **ThinLinc:** rackham-gui.uppmax.uu.se
      1. 2FA may be needed, which can be handled by logging in with regular SSH, doing 2FA, logging out again, then there is a grace period of some minutes for you to login to ThinLinc. More here: [https://nbisweden.github.io/workshop-ngsintro/2203-canvas/lab\\_connect.html](https://nbisweden.github.io/workshop-ngsintro/2203-canvas/lab_connect.html)
    - iii. **From webbrowser:** <https://rackham-gui.uppmax.uu.se/>
  - b. HPC2N - Kebnekaise
    - i. **SSH:** kebnekaise.hpc2n.umu.se
    - ii. **ThinLinc:** kebnekaise-t1.hpc2n.umu.se
    - iii. **From webbrowser:** <https://kebnekaise-t1.hpc2n.umu.se:300/>
3. Course project at UPPMAX: NAISS2023-22-914
4. Storage area for the course project at UPPMAX: /proj/naiss2023-22-914
5. Course project at HPC2N: hpc2n2023-110
6. Storage area for the course project at HPC2N: /proj/nobackup/hpc2n2023-110
7. Create a directory in your project directory, at UPPMAX or HPC2N, and then you may get the course material from GitHub using: `git clone https://github.com/UPPMAX/R-python-julia-HPC.git`

## Q/A Tuesday Python

8. My UPPMAX account hasn't been enabled yet, is there anything I can do about it?
  - a. **A:** What is your name? You may be able to make it a private message in the chat, if you want. /Björn C!
9. In the project directory, what are the folders "nobackup" and "private" for?
  - a. **A:** Not needed for the course, but "private" has less reading permissions for others and /nobackup is not "backed up". The /proj/nobackup/ is the file server on kebnekaise that the project storage resides on. Yes, it is just named that to remind people it is not backed up.
10. I am trying to connect with ThinLinc to Rackham, but the ThinLinc client hangs and does not connect (or it is taking very long to connect). Any ideas?
  - a. **A:** Do you state "rackham-gui" instead of "rackham" ?
  - b. **A:** You can try the web version as a start: <https://rackham-gui.uppmax.uu.se>
  - c. There is a problem with 2-factor authentication using the ThinLinc desktop client. Connect first with SSH client and authenticate with 2FA, then connect with ThinLinc to use the grace period. See [https://nbisweden.github.io/workshop-ngsintro/2203-canvas/lab\\_connect.html](https://nbisweden.github.io/workshop-ngsintro/2203-canvas/lab_connect.html)
11. INFO: use uquota command to see paths to project folder(s)
12. I still haven't managed to login... wrong username/password. When a new password is set through naiss, is it propagated/linked immediately to the kebnekaise/rackham? I think that is what I "activated" at 9, and then it said it needs to be checked by someone before being accepted. When reloading the page after 5 min or so, it was activated, but I can still not login.
  - a. From terminal: `ssh <username>@rackham.uppmax.uu.se`  
Will take a check again.  
So I had not requested the account from UPPMAX, did that now. Can I use the hpc2n instead, that I should already have an account at?
    - i. **A:** Use Kebnekaise!!
13. Not activated accounts: Is there a button or box you can click to activate it in SUPR?
  - a. **A:** Did you follow all steps here: <https://www.uppmax.uu.se/support/getting-started/applying-for-a-user-account/>
14. Do you mind not saying our names on the open?
  - a. **A:** I can probably cut out that part from the recordings when I process them
15. When trying to import package the XQuartz application opens, but no window is shown. If quit application I get error message in terminal. Should I have loaded python/3.11.4 before writing command 'pip list' and then 'import <package>?'
  - a. **A:** Yes, you should load any python modules before doing pip list or import package. They need the python module. You also need to start python before doing import package. But pip list you do outside python but after loading
16. Sorry, but what is the difference between load python and start python? Am I not starting python by the command module load python/3.11.4 ?

- a. **A:** Load python is the module load line. That activates important parts to python and its libraries. If not loading python you will get the system python!
  - b. **A:** Then you can start python (terminal) and do your (interactive) work.  
You can also run python scripts with `python <script>`.
17. I managed to login with ssh from terminal but from the ThinLinc Client I got the wrong user/password message. Server name : `kebnekaise-tl.hpc2n.umu.se`, User `adhs0001`
  - a. **A:** Did you put the password on the login box or wait for it to ask for password? It will not work unless you put the password in the login box:  
<https://www.hpc2n.umu.se/documentation/guides/thinlinc>  
- I followed the instruction, now I have a timed out message (more than three trials...)
18. NOTE: send an email to `support@uppmx.uu.se` if you have disabled accounts on UPPMAX
  - a. Citation from web page: "After applying it might take up to 2 working days before you receive 2 emails with information on how to login to UPPMAX."
  - b. But I have asked my colleagues to make a fast lane. Unfortunately they are all on a meeting right now.
  - c. We are sorry if you applied before Friday last week and still are waiting.
19. Sorry, why are we wasting so much time on this? There was too little time before and now we sit in breakout rooms forever to just import one package? When are we coming back to lecture?
  - a. Richel: I assume this is about the 'pip' part. The reason that this session took longer than needed, is that I could not predict how fast students are, where the previous time it did take 45 mins. I see many possible explanations for what caused this big difference.
20. LEARNINGS: The change of course structure means some tuning of timings! Sorry for that!
21. When do we want to only import packages and when do we want to install? Will you continue with this in next lecture?
  - a. Yes, in next session! If the package is not found you may need to install it.
22. sorry, was the task just to import pandas?
  - a. **A:** Yes, load python, see what packages exist (with `pip list`), and then start python and try import a library, for instance pandas
23. For future breakout rooms, will you bring us back to the 'general' room or will you give us a time when we should go back ourselves? This was a bit confusing this time.
  - a. Richel: I agree that this was unclear! And this by deliberate choice: if all went well, I would not have liked to disturb the students. If students would need help, I would call them back. I assumed the schedule and me going through the Breakout rooms would clarify things. Maybe next time I will share my plans :-)
24. Regarding using ThinLinc on rackham with: `rackham-gui.uppmx.uu.se`
  - a. 2FA may be needed, which can be handled by logging in with regular SSH (`rackham.uppmx.uu.se`), doing 2FA, logging out again, then there is a grace period of some minutes for you to login to ThinLinc. More here:  
[https://nbisweden.github.io/workshop-ngsintro/2203-canvas/lab\\_connect.html](https://nbisweden.github.io/workshop-ngsintro/2203-canvas/lab_connect.html)

25. Will you go through the material on the isolated env on the <https://uppmx.github.io/R-python-julia-HPC/python/isolated.html> later? Or will there be a recording of this?
- Richel: I am unsure if this question is about venv or conda. If 'venv': I assume you've found the recording at the bottom part of the page. If this is about conda: we'll discuss that later
26. When will the recordings be up?
- Later today/tomorrow
27. The error that you get in the youtube video for venv should be addressed because now I dont know how to install an older version of a package if the environment has a newer version already installed.
- A:** If you need a newer version, I know it works to use  
`pip install --upgrade <package>`
  - A:** If you need another/older different version you do:  
`pip install --upgrade package==version`
  - A:** If there are prerequisites, you need to do those first, or list all of them in order, e.g. if you install numba which depends on numpy and depends on scipy, you would do:  
`pip install --upgrade scipy numpy numba`
  - Richel: I agree that the video ideally did install something useful in the first demo. As 'Exercise 1' I felt it would be good enough. Also, when I scheduled to make the video, I expected the course materials to work. I was unexpectedly wrong. The material (and video) will likely be updated.
28. so we need to create a new environment with the old version and then upgrade the others?
- A:** You create a virtual environment with site-packages because there are many more than numpy in that. So use site-packages when you create the venv, and then upgrade the ones that are too old
29. After creating Example-gpu and installing numba, what are we supposed to do now? Is there more lecture coming?
- A:** There will be "informal talk" at 11:45, then break at 12:00, then we continue with lectures at 13:00
30. I feel like the instructions part goes really fast and is a bit unclear - what we are supposed to do on our own can be clarified. The working in break out room time feels quite long. I'd like to suggest spending more time on instructions and making sure the task is clear. At this time I do not know if I have done what I should have done, or if I have done more or less than instructed. Edit: All parts, general thought.
- Richel: I am unsure about which part this is about. As it is about a long breakout session, I assume this is in my part. In my part, I went through all rooms one by one, except the Silent room, as it is a silent room. I go through all rooms *exactly* because I check if all learners what to do. We hence agree that all learners should be checked upon :-)  
Question back: So, I assume you were in the Silent room then? If yes, would you recommend visit the Silent room too? I felt I should not. Would you accept my

reasoning being a valid way to look at it? Maybe I should state that there are rooms for 1 person too

:  
visiting the silent room. I think it is the time  
me off, as it is quite a long time. Then it  
feels like we are supposed to do more than what  
was understood by instructions the first time  
around. A bit more structured instructions before  
less confusing.

31. When updating numpy, scipy and numba – how do we know it is actually the newest versions that were updated/installed?
  - a. You can specify the version for instance with “pip install numpy==1.15.4”
  - b. But how do we know the version number?
  - c. pip list will show the versions that are installed
  - d. Sure that is correct, but if want to install a newer version – how now what version numbers one should specify? As the ones in pip list are the current ones right. Maybe I’m totally lost, I’ll make own research on this. Thanks.
32. If your run takes shorter amount of time than you allocated (e.g. 1 hour instead of 2) do you get charged with 1 hour?
  - a. **A:** at HPC2N, the project is charged with the time the simulation takes not the allocated time
  - b. **A:** if your job takes 2 hours and is scheduled for 1, it will stop after 1 hour and the scheduler will count 1 hour
  - c. **A:** also for UPPMAX the project is charged with the time the simulation takes not the allocated time
33. Were we supposed to do the first part of the exercises, where virtual environment called analysis was to be created?
  - a. Richel: I hope it got clear we did ‘Exercise 1’ from the slides, and then later create ‘GPU-example’, as I visited all Breakout Rooms twice, except for the Silent Room. If you were in the Silent Room, I hope you can at least understand why I let the Silent room be .. well .. quiet :-)
34. Can I also just type "conda activate <name>" in the batch script?
  - a. Richel: Yes! I am unsure if this is recommended though ...
35. Can I forward parameters from “sbatch file.sh 2 3” to the python code file?
  - a. Richel: Yes! Use the bash input variables, \$1, \$2, etc
36. How can I get cuda information (i.e. version) from Snowy? (need the pytorch package to match the cuda version).
  - a. **A:** Start an interactive job on Snowy and use module spider to find a module version (general).
37. Can anybody create a project? As a student for example. Or we can only be added as a user?

- a. If you are a student, your teacher (as principal investigator) has to apply for the project
  - b. [https://uppmx.github.io/uppmx\\_intro/naiss.html#the-projects](https://uppmx.github.io/uppmx_intro/naiss.html#the-projects)
  - c. application for projects are done in `supr.naiss.se` There are national systems (Rackham, Dardel, Alvis, Tetralith) or local (HPC2N, Lunarc). If you are a PI that is a senior researcher you would be able to get medium and large projects  
If not you could apply for small projects
38. Also can you stream output from a batch job (e.g. a progress bar)? So you can see how it is doing?
- a. One way is to create a log file (a line in the batch file) that is saved somewhere. Then you can read it while it is updating. I suggest to use the command `tail` to just see the last rows of the file
  - b. More info about slurm at UPPMAX  
[https://www.uppmx.uu.se/digitalAssets/560/c\\_560271-l\\_1-k\\_uppmx-slurm-2023-02.pdf](https://www.uppmx.uu.se/digitalAssets/560/c_560271-l_1-k_uppmx-slurm-2023-02.pdf) and <https://www.uppmx.uu.se/support/user-guides/slurm-user-guide/>
  - c. email:
    - `--mail-type=BEGIN,END,FAIL,TIME_LIMIT_80`
    - o out/err redirection:
      - `--output=slurm-%j.out` and `--error=slurm-%j.err`  
.....by default, where `%j` will be replaced by the job ID
      - `--output=my.output.file`
      - `--error=my.error.file`
39. Just to confirm, so I add `#SBATCH --gres=gpu:x` in my script to activate GPU if needed?
- a. Yes
40. You (Birgitte) are in Python directory. Is this inside Exercises directory? Should we be in that as well I assume??  
Does it matter in which directory Exercises vs Python you are in while doing these batch jobs?
- a. The only thing that matters is that the exercises/programs that you want to run is in the directory you are in. Use the one you placed stuff in and change paths in virtual environments/batch scripts accordingly.
41. How to use conda with SLURM? Or SLURM with conda?
- a. Just load the conda module within the SLURM script and then do `conda activate`  
There will be something about conda later today
42. Does Thinlinc work with WSL2?
- a. At HPC2N, you can work with Thinlinc online: <https://kebnekaise-tl.hpc2n.umu.se:300/> . Thus, you dont need to install it.
  - b. I don't know if I would consider WSL2 in a thinlinc environment. You start thinlinc from windows not the linux terminal, I would say. There is no need to.
  - c. Yes, there are ThinLinc clients for Linux, Windows, macOS. Nothing to do with WSL2. You just run the client suitable for your OS. I don't know if ThinLinc can be

run from WSL2, maybe? I don't have access to a Windows computer to test right now. But no need to run it from WSL2 I guess.

43. Can I use dask for distributed computing on rackham?
- We haven't tried it fully yet but my plan is to test with slurm next week. I know a colleague has made it working a while ago. /Björn
  - I have made it work within one node (20 cores). You may test from the material presented in the **links** in the end of the parallel material, using you own installed dask package! /Björn
  - I can't find any dask code here:  
<https://uppmx.github.io/R-python-julia-HPC/common/parallel.html#dask>, or do you mean somewhere else?
    - <https://aaltoscicomp.github.io/python-for-scicomp/parallel/#dask-and-task-queues>
    - <https://enccs.github.io/hpda-python/dask/>
  - We will have another course for Python more dedicated to HPC in December. We plan to introduce DASK more thouroghly then. More info in a couple of weeks in the same info channels as you got the info about this course.
44. I tried out the instructions for jupyter (UPPMAX for rackham). How do I get out of the interactive session afterwards again? Or can I just stay in it for the other exercises too?
- \$ exit
  - you may want to test some more things! Then stay!
45. When I am using command "jupyter-notebook" I am getting following error. Bus::open: Can not get ibus-daemon's address.  
IBusInputContext::createInputContext: no connection to ibus-daemon  
I am using uppmx
46. my code is running but no jupyter notebook is showing off
- Which url did you copy? Should be the one with hostname. Are you inside ThinLinc (for HPC2N)?
  - I'm not in ThinLinc, I will retry
  - For HPC2N, you need to be within HPC2N domain, so easiest to do with ThinLinc
47. For running distributed code (parallel processes), do I need to be in interactive mode, or how do I combine that with SLURM?
- You can run with SLURM, you just need to ask for more cores and use srun for your programs. There are some examples on the Kebnekaise intro slides (near the end – download the PDF):  
<https://github.com/hpc2n/intro-course/blob/master/slides/3.usage.pdf>  
Also some examples here:  
<https://www.hpc2n.umu.se/documentation/batchsystem/mpi-examples>
48. [https://uppmx.github.io/bianca\\_workshop/slurm-intro/](https://uppmx.github.io/bianca_workshop/slurm-intro/)
49. For distributed computing (multi-node) I am aware of mpi4py (which uses MPI) and Dask
50. How to use this code on the cluster? (Multiprocessing in parallel session)
- This is the code of a Python script. That means that it can be `sbatch`ed later. Probably using the GPUs in the batch scripts

- b. Do I need to sbatch it exactly with the number of processes as in the python script? with n=6?
  - c. I predict: yes. I predict if you do wrong, it will work, yet less efficient
  - d. you can allocate more cores I guess, but they will not be used. you will be accounted for the allocated cores though so it is more "expensive"
  - e. note that your CPU usage counted will be the sum of all cores. Hence if you use 10x more CPUs for a 2x speed increase, you won't be in the front of the queue long anymore
  - f. you may use  $\leq$  number of allocated cores inside of your python script but not more. Otherwise, your script will create an overhead on the nodes
  - g. Ok<y, and this will Work easily with n >> 20?
  - h. if there are >20 cores in the node yes. Dardel for instance has 128 cores/node
  - i. you can allocate more cores I guess, but they will not be used
  - j. Note that your CPU usage counted will be the sum of all cores. Hence if you use 10x more CPUs for a 2x speed increase, you won't be in the front of the queue long anymore :-)
51. Question: I am using Bianca, for a script that generate figures, do I just add #SBATCH --gres=gpu:1 within my script? Do I need #SBATCH -M bianca as well?
52. When working in interactive mode, how does one know how many tasks/nodes to command?
53. Hoa about cross-node distributed computing?
- a. For distributed computing (multi-node) I am aware of mpi4py (which uses MPI) and Dask
54. (from chat) How do you load cuda, is that simply set up with certain python versions?
- a. There will be a session about it later, but in general it is included with some versions (though at HPC2N at least you mean need to load separately)
55. HPC2N YouTube channel will have the recordings later today or tomorrow:  
<https://www.youtube.com/user/HPC2N/videos>
56. Direct link to the recordings playlist on HPC2N's YouTube: <https://www.youtube.com/watch?v=aCSMEGgITG8&list=PL6jMHLEmPVLxdCIIWrzseYkGoCCXjrukM>
57. Evaluation survey for today: <https://forms.office.com/e/NxtwCk7w5H>



## WEDNESDAY, 18 October 2023, Julia

1. Please note the following:
2. Add your questions below, numbering them continuously.
3. Be careful in case someone else is writing at the same time.
4. Please **DO NOT** delete your questions even when they have been answered, as we are planning to use them to improve our material.
5. Do not share any sensitive information as this document is accessible to anyone with the correct link.
6. You can start a **new line within a question by pressing <SHIFT> + <ENTER>**.
7. General information / summary:
8. Hostnames for login nodes (the main course project is at UPPMAX):
  - a. UPPMAX – rackham
    - i. **SSH:** rackham.uppmax.uu.se
    - ii. **ThinLinc:** rackham-gui.uppmax.uu.se
    - iii. **From webbrowser:** <https://rackham-gui.uppmax.uu.se/>
  - b. HPC2N - Kebnekaise
    - i. **SSH:** kebnekaise.hpc2n.umu.se
    - ii. **ThinLinc:** kebnekaise-t1.hpc2n.umu.se
    - iii. **From webbrowser:** <https://kebnekaise-t1.hpc2n.umu.se:300/>
9. Course project at UPPMAX: NAISS2023-22-914
10. Storage area for the course project at UPPMAX: /proj/naiss2023-22-914
11. Course project at HPC2N: hpc2n2023-110
12. Storage area for the course project at HPC2N:  
/proj/nobackup/hpc2n2023-110
13. Zoom ONLY for the parallel sessions: [https://umu.zoom.us/j/64351885207?  
pwd=RVpwYnFRdINLQkM5WWR0VWsrSzUwQT09](https://umu.zoom.us/j/64351885207?pwd=RVpwYnFRdINLQkM5WWR0VWsrSzUwQT09)

## Q/A Julia

14. println("Hello World!"), Why do you use println?
  - a. That gives a new line after the print
  - b. Look in screen!
15. Editors:
  - a. Nano easiest
  - b. vim if you are used to itexit
  - c. emacs if you are used to it
  - d. mobaxterm has a built in editor
  - e. Gedit graphical
16. If I run julia serial-sum.jl 4 8 in the julia env it gives errjuliaor (syntax: extra token "4" after end of expression)? The other way round works
  - a. You should run from **batch env (\$)**, exit julia REPL first!
  - b. Interesting that 8 4 did work, though

17. In julia, when writing 'ls' i got this message: **ERROR:** UndefVarError: Is not defined
- 'ls' will work in the shell mode, use ';'
 

But it seemed like Pedro was not in shell mode when giving command 'ls'?
  - Go back to julian mode with <backspace>
  - He is using shell mode now.
 

Yes now, but not before. Never mind, I'll sort it out:)
18. **Note:** ] is for package mode
19. **NOTE:** As it happened to me (Björn) you may see after some time with installing packages that your disk quota is full. This happens when you install in the \$HOME area. It may be better to put your files to a project area.
20. In the beginning when using the command 'Pkg.add("DFTK")' (after 'using Pkg') the only package failed to download was DFTK. Cannot continue from there.
- Kebnekaise or Rackham? Rackham.
  - What was the error?

```

Precompiling project...
 × MPI
 × DFTK
 46 dependencies successfully precompiled in 148 seconds. 59 already precompiled.
 2 dependencies errored. To see a full report either run `import Pkg; Pkg.precompile()` or load the packages

```

When using command 'import Pkg; Pkg.precompile()' as instructed i get:

```

[julia> import Pkg; Pkg.precompile()
Precompiling project...
 × DFTK
 0 dependencies successfully precompiled in 9 seconds. 106 already precompiled.

ERROR: The following 1 direct dependency failed to precompile:
DFTK [acf6eb54-70d9-11e9-0013-234b7a5f5337]

```

... and more error text.

Seems like library 'libmpi' could not be loaded and it errored from there. Says there is no such file or directory.

I loaded and started julia from /home/<user>/R-python-julia-HPC/Exercises : is this incorrect?

- Can you try "using Pkg" "Pkg.add("DFTK")" as I did?

Yes, I get:

**Resolving** package versions...

**No Changes** to `domus/h1/<user>/julia/environments/v1.8/Project.toml`

**No Changes** to `domus/h1/<user>/julia/environments/v1.8/Manifest.toml`

- What is the output of LOAD\_PATH ? (Björn)

"4-element Vector{String}:

"@"

"@v#.#"

"@stdlib"

"/sw/comp/julia/1.8.5/rackham/lib/glob\_pkg/environments/v1.8"

i. LOOKS OK

- I think the issue can come from the "/sw/comp/..." environment, can you try the following: exit() the julia session, and start a new one

f. Then,

- empty!(LOAD\_PATH)

- h. `push!(LOAD_PATH, "@")`
  - i. `push!(LOAD_PATH, "@v#.#")`
  - j. `push!(LOAD_PATH, "@stdlib")`
  - k. After, this try the “using Pkg” and add the package DFTK  
Thanks! Did as stated, and after `Pkg.add(“DFTK”)` i got the “No changes” responses again. Should I try activating my-first-env again and add DFTK in there?
  - l. Yes, if the initial installation of DFTK worked, let me know  
Had to install MPIPreferences by ‘add MPIPreferences’ in Pkg mode. Then ‘using MPIPreferences’, then `MPIPreferences.use_system_binary` – and then redo adding DFTK.  
Now it works.
21. For DFTK it needed MPI to precompile.
- a. On UPPMAX follow: [https://www.uppmx.uu.se/support/user-guides/julia-user-guide/#tocjump\\_22751075665337883\\_13](https://www.uppmx.uu.se/support/user-guides/julia-user-guide/#tocjump_22751075665337883_13)
  - b. In other words do this in terminal and restart julia  
For julia/1.8.5:
  - c. `$ module load gcc/11.3.0 openmpi/4.1.3`
22. This; `#SBATCH --time=00:10:00` # Asking for 10 minutes. If you specific the time and later it take longer time. Will it be terminated automatically?
- a. Terminated, yes
  - b. So better overestimate the times a little bit
23. PLACEHOLDER
24. After command ‘vim name.jl’, and after creating whatever function in the new window – how save and exit from there? I tried ending with ‘:wq’ and then pushing enter, did not work.
- a. `<esc> :wq`
  - b. This is why vim is not recommended for non-vim users in a course. We should have been more clear!  
Okay. Yes could have been clearer, thanks, but no worries ofc. What use otherwise?  
- Thanks, see the editors below now.
  - c. Follow-up question: do you write e.g. Nano in shell mode in julia, like ‘Nano name.jl’ instead of vim then?
  - d. Yes.
    - i. I did not know you could start from the REPL shell! Otherwise you do it from the bash shell  
sorry, which one of these is the shell mode in julia?
25. Editors:
- a. Nano easiest — ^ means <ctrl>
  - b. vim if you are used to it
  - c. emacs if you are used to it
  - d. mobaxterm has a built in graphical editor
  - e. Gedit graphical (installed on both centres!)
26. When will XQuartz, that the instructions stated was good to download for graphics in terminal, be useful?

- a. For Jupyter and RStudio, and if you used the terminal. But you must have logged in with -X or -Y (ordinary dash/hyphen)
    - i. Like `ssh -X <username>@centre-address`
  - b. Otherwise ThinLinc is good
27. I did this Submitted batch job 8495841 on cluster snowy and got this: `scontrol show job 8495841 ---> slurm_load_jobs error: Invalid job id specified`
  - a. **On Snowy: add - -M snowy on the scontrol line**
28. In running julia in batch mode. I'm unsure how to access this 'batch mode' – do I do this after accessing rackham/uppmax and then my directory?  
Do I write these lines:  
`#!/bin/bash -l`  
`#SBATCH -A naiss2023-22-914`  
`#SBATCH --time=00:10:00`  
`#etc....`  
straight into terminal or do I need to load something before, to activate batch mode?
  - a. You need to have it in a script file
  - b. Than `sbatch <filename>`
29. When having stared jupiter etc. How when finished exit the code running in julia and get back to terminal?
  - a. `<ctrl> - C` works, that is how I do it but there may be some other way. Does that work for you? Thank you! That worked.
30. PLACEHOLDER
31. PLACEHOLDER
32. PLACEHOLDER
33. PLACEHOLDER
34. Recordings for today are now uploaded to the same playlist as yesterday, on HPC2N's YouTube channel: <https://www.youtube.com/watch?v=aCSMEGgITG8&list=PL6jMHLEmPVLxdCIWrzseYkGoCCXjrukM>
35. Evaluation survey for today: <https://forms.office.com/e/bDAFJwXRK9>

THURSDAY, 19 OCTOBER 2023, R

1. Please note the following:
2. Add your questions below, numbering them continuously.
3. Be careful in case someone else is writing at the same time.
4. Please **DO NOT** delete your questions even when they have been answered, as we are planning to use them to improve our material.
5. Do not share any sensitive information as this document is accessible to anyone with the correct link.
6. You can start a **new line within a question by pressing <SHIFT> + <ENTER>**.
7. General information / summary:
8. Hostnames for login nodes (the main course project is at UPPMAX):
  - a. UPPMAX – rackham
    - i. **SSH:** rackham.uppmax.uu.se
    - ii. **ThinLinc:** rackham-gui.uppmax.uu.se
    - iii. **From webbrowser:** <https://rackham-gui.uppmax.uu.se/>
  - b. HPC2N - Kebnekaise
    - i. **SSH:** kebnekaise.hpc2n.umu.se
    - ii. **ThinLinc:** kebnekaise-tl.hpc2n.umu.se
    - iii. **From webbrowser:** <https://kebnekaise-tl.hpc2n.umu.se:300/>
9. Course project at UPPMAX: NAISS2023-22-914
10. Storage area for the course project at UPPMAX: /proj/naiss2023-22-914
11. Course project at HPC2N: hpc2n2023-110
12. Storage area for the course project at HPC2N:  
/proj/nobackup/hpc2n2023-110
13. Zoom ONLY for the parallel sessions: [https://umu.zoom.us/j/64351885207?  
pwd=RVpwYnFRdINLQkM5WWR0VWsrSzUwQT09](https://umu.zoom.us/j/64351885207?pwd=RVpwYnFRdINLQkM5WWR0VWsrSzUwQT09)
14. PLACEHOLDER
15. PLACEHOLDER
16. I runned this: notebook(dir="/proj/naiss2023-22-914/nazib/Julia") but got this Info: running  
`/sw/comp/python/3.10.8/rackham/bin/jupyter notebook`
17. On HPC2N, you can copy the exercises in a tarball from  
/proj/nobackup/hpc2n2023-110/Exercises.tar.gz
18. On UPPMAX you can copy the exercises in a tarball from /proj/naiss2023-22-914/Exercises.tar.gz
19. To copy the file, go to your personal subdirectory under /proj/naiss2023-22-914/ (UPPMAX) or under /proj/nobackup/hpc2n2023-110 (HPC2N) and do: cp <path listed above> .
20. Example for me (bbrydsoe), on Rackham:
  - a. cd /proj/naiss2023-22-914/bbrydsoe
  - b. cp /proj/naiss2023-22-914/Exercises.tar.gz .

- c. Then you need to extract the gzipped tar file: `tar xzvf Exercises.tar.gz`  
xzvf right?
  - d. Yes
  - e. `tar xzvf Exercises.tar.gz`
  - f. `cd Exercises/`
  - g. `cd R`
  - h.
21. I created a .Renviron file in my home directory and update it using this code: `echo R_LIBS_USER="$HOME/R-packages-%V" > ~/.Renviron` I checked it in my texteditor and it looks fine. However, when I try to install a new package in my R session it cannot find this path still it looks at the default library path which is `"/sw/apps/R/4.1.1/rackham/lib64/R/library"` which I dont have writing access to.
- a. The system may not read your .Renviron unless you start a new shell or source .Renviron
  - b. What is the output of `.libPaths()` in your R terminal?
22. Now it is these two:
- ```
[1] "/domus/h1/sarho/R-packages-4.0.4"
[2] "/sw/apps/R/x86_64/4.0.4/rackham/lib64/R/library"
```
- It was only the second one yesterday. So does it mean that I had to log in again and it got fixed? Because I basically didnt do anything since yesterday.
- a. **A:** Yes, logout and login is the easiest way to get the system to detect the new settings.
23. There is no /proj directory in my home folder
- a. What system are you on? Rackham or Kebnekaise: rackham. Solved. I understand I need to move to /proj/naiss first.
24. How long should it take for a package to get installed from an R session? I started to install stringr and it has been several minutes now.
- a. Some take longer! But do you see progress? Cluster?
25. Could you please share the code for login to directory in reckham
- a. Login is one thing and go to a directory another, that you can do when you are logged in
  - b. Log in from terminal: `ssh -Y <username>@rackham.uppmax.uu.se`
    - i. Password
  - c. Orient in the file trees on Rackham
    - i. `cd /proj/naiss2023-22-914` (note the starting `"/`)
    - ii. `ls`
      - 1. Did you create a user directory already, otherwise:
    - iii. `mkdir <yourname>`
    - iv. `cd <yourname>`
    - v. `pwd`
      - 1. If the path looks correct you are now in your own folder/directory
  - d. Then follow step 20 above to get the exercises
  - e.

26. Q: On rackham, my current library path is not writable. R 4.0.4
- a. Do you get some more info that it will try to install at another place?
    - i. Yes, "would you like to install in a personal library instead"
    - ii. Good! That is the way to do it.
27. **Info:** devtools is installed on UPPMAX
28. When you use `renv::init()`, do you need to be in a specific R version? I mean does it affect the package versions that are later installed?
- a. The virtual environment is connected to one specific R version! Same for python
29. How can you confirm you have created `.Renviron`?
- I do the 'touch...' but when check by 'ls' there is no `.Renviron` - maybe you shouldn't see this there anyway?
- a. `ls -a` will list hidden (<.something>) files
  - b. It is in you home folder (/home/<username>)
  - c. Check with `ls -a $HOME`
30. When you load a specific R version (e.g., 4.0.4), and then you download packages from CRAN. Will the packages will be the latest package version on CRAN or the one that it is compatible with the R version you loaded?
- .If the version of an R package available at CRAN and installable via `install.packages()`, which is always the latest version, is compatible with the version of R that you are running, then you will be able to install that version. If it is not compatible with the latest version of R, then it will not be available at CRAN and installable via `install.packages()`.
- .Note that CRAN is also not sophisticated in another way. If you try to install a CRAN package that is no longer available for the version of R you are running, you will get the message "this package is not available for your version of R" or something like that. If you try to install a BioConductor package with `install.packages()`, instead of `BiocManager::install()`, you'll get the same message. If you misspell the package name, or give it a nonsensical name, you'll get the same message.
31. What would happen to my virtual environments when my one-year UPPMAX account expires?
- a. You should move your data and files to a local backup! And possible move them to a new cluster
  - b. Thanks for your response! Focusing my question a bit, i wonder whether I can, as part of my back up, also save the R packages versions I installed or were ready available when I created my virtual environment? Let's assume that my virtual environment is compatible with R 4.0.4
  - c.
32. During installation the selection for crane mirrors will be SWEDEN UMEA?

- a. I think that is default, but you may get a pop up screen in ThinLinc or if you use x-forwarding in terminal
33. Why do we have some of function with – and other with – below- Example time and n
- ```
#!/bin/bash
#SBATCH -A naiss2023-22-914 # Course project id. Change to your
own      project ID after the course
#SBATCH --time=00:10:00 # Asking for 10 minutes
#SBATCH -n 1 # Asking for 1 core
```
- a. Not sure I understand your question. This script is needed to reach the compute nodes
- why sometimes '-' as before time and why sometimes '-' as before n i think person means
- b. For most commands, '-' (hyphen) is used for single letters options/flags, whereas two '-' is used for whole words
- i. Example:
1. `ls -lrt` is shorthand for `ls -l -r -t`
  2. What about the nodes you allocated in the batch script header then? Is it overwritten by the `makeCluster()` function of the `doParallel` package?
  3. `ls --help` gives help output for the `ls` command
34. Rackham, sh script job section, is there difference between `R < code.R` and `Rscript code.R`?
- a. `Rscript` passes some arguments into `R`
35. I'm wondering about the `makeCluster` and `stopCluster` functions. Can you elaborate on that? Don't we already define the number of allocated nodes in the batch script header?
- a. Allocated cores are an upper limit for what you can define from `R`
- b. In the batch script/allocation commands you ask for some resource, a number of cores etc. Then when you run `R` or whatever, you say how many of the cores etc. you will use. It could vary what you need in the same session.
36. What about the nodes you allocated in the batch script header then? Is it overwritten by the `makeCluster()` function of the `doParallel` package?
- A:** No, you have the number of nodes and cores allocated that you asked for in the batch script, but maybe your script is not asking to use all of them. It can use fewer, just not more.
- a. They are still active and the CPU-hrs used will be based on that.
- b. You may book more cores to get more memory even though you are not running parallel work, **typically 6-8 GB per core**
- c. The number of cores in `makeCluster` should not exceed the number in your batch script (`-n` flag). Otherwise, the script will overload the cores if computations are expensive
- d. something that could be useful is to allocate some number of cores (`-n`) and then use less in `makeCluster` to get more RAM memory per core. Even you can allocate the entire node and use only few cores in `makeCluster` if your application is memory consuming
- e.
37. If you run the command `squeue -u <username>` and get this: `JOBID PARTITION NAME USER ST TIME NODES NODELIST(REASON)`. It means you have no job running?



- a. Yes and no in the queue either
  - b. Did you already start a job with sbatch? Perhaps it is already finished?
    - i. Do `ls -lrt` and you can see the latest file and it might be `slurm-XXX`
  - c. Yes i did. Thanks. I can see it with `-out` extension.
38. If you want to load an R- package for the `hello.R` command. Do you do it inside the `R` command or in the `slurm`?
- a. The library lines has to be in the code
  - b. Depending on cluster you may have to in the **batch script** load `R_packages` on UPPMAX or other modules at Kebnekaise, like `R-bundle-Bioconductor`
39. In the Rscript (Example taken from <https://github.com/lgreski/datasciencectacontent/blob/master/markdown/pml-randomForestPerformance.md> `library(mlbench)` `data(Sonar)`). The question is, if you have your own data. You have to set path to the location of the data in your directory inside UPPMAX?
- a. Yes you have to set full paths to data. If you always run from the same place (present working directory) it can be sufficient with relative paths. If you run it from somewhere else the full path is a must
40. Do you choose to either run `rscript` with `'R --no-save --quiet < hello.R'` or send script to batch? Cause if copy and use whole example serial code it runs `rscript` – so then why send the script to batch?
- a. Sbatch lets you use the compute nodes. Otherwise it will run from the login node (if you are there). That is the difference. Did I understand your question?
  - b. So one should use either or, but `sbatch` command if want to run on batch node and not in login node? I'm not sure if you understand, or if it's rather I who are not understanding. If look at slides, the 'send script to batch' command is stated/explained after the example serial code – I assumed this step was to come after running `Rscript`, but after running it its done right. So that is why I asked if you choose to do either or.
41. When trying the example serial code, but with `add2.R` - do I just add e.g. `2 4` after `'R --no-save --quiet < add2.R 2 4'` like so, or where add arguments?
- a. `R CMD BATCH --no-save --no-restore '--args a=1 b=c(2,5,6)' test.R test.out &`
  - b. From <https://stackoverflow.com/questions/14167178/passing-command-line-arguments-to-r-cmd-batch>
  - c. Thanks! So the code in the "UPPMAX solution" under the exercise is not correct? Because this is what I did, and did not work. By doing so, the arguments are ignored and answer is NA.
    - i. Are you in the folder where `add2` is? Yes, used `'ls'` and `add2.R` is there.
  - d. A can reproduce this error
  - e. Try: `Rscript add2.R 2 3`

- f. Does still not work for me. When using Rscript instead, says 'No such file or directory' - but still in directory where add2.R is so it should find it.
- g. Can "cd" to the right directory in the batch script, like after the "module loads". Like so:
- ```
# Load any modules you need, here R/4.0.4
module load R/4.0.4
cd /proj/naiss2023-22-914/bjornc/Exercises/
# Run your R script (here 'add2.R')
Rscript add2.R 2 3
```
- h. Nope, still error:
- ```
[<user>@rackham4 R]$ Rscript --no-save --quiet < add2.R 2 3
Fatal error: cannot open file '2': No such file or directory
```
- i. You are not in the batch script but on command line now
- i. And remove `-no-save --quiet`
- j. So if I put in these commands, see below, I am still not in batch script?
- ```
#!/bin/bash
#SBATCH -A naiss2023-22-914 # Change to your own after the course
#SBATCH --time=00:10:00 # Asking for 10 minutes
#SBATCH -n 1 # Asking for 1 core

# Load any modules you need, here for R/4.0.4
module load R/4.0.4
```

42. This is really basic, but I mainly use Bianca and RStudio. Should I be saving my code and then running it in through the terminal with a SLURM request rather than using an interactive session with 2 cores or so and running the code as I go along? I guess it also depends on the size of the project probably.
- We'll talk about rstudio after break
  - Was this answered?
43. HPC2N parallel session: <https://umu.zoom.us/j/64351885207?pwd=RVpwYnFRdINLQkM5WWR0VWsrSzUwQT09>
44. How do you change the number of core if you are already in the interactive mode?
- You cannot change the max number of cores, but you could use something like "srun -n 2 <your commands>" to only run on 2 of them, for instance
45. (solved) Rackham; dependency between batch jobs, what's the best practice? Dependency between shells (each contain one R script), dependency between R scripts and all scripts in one shell, or Rscript <script1.R >script2.R ? E.g., a large simulation with  $10^4$  iterations, each require lots of memory. I divide the job to say 1000 (batch) \* 10 iters (per batch) to get the total  $10^4$  iters. The second job requires all iters from the first job.
- For dependency on the job level:  
[https://www.uppmax.uu.se/digitalAssets/560/c\\_560271-l\\_1-k\\_uppmax-slurm-2022.pdf](https://www.uppmax.uu.se/digitalAssets/560/c_560271-l_1-k_uppmax-slurm-2022.pdf)
  - For R: please help!
  - ( solved) Also, how to submit the batch, non-interactive job to the test node (may called otherwise)?

- i. Same as for interactive: `-p devcore` is the same as `p -core` whereas `-p devel` is the same as `-p node`
- ii. In the batch script: `#SBATCH -p devcore` for instance

46. How did you copy into thinlinc terminal again? shift + insert doesn't seem to work for me

- a. within ThinLinc
- b. copy: `<ctrl>-<insert>`
- c. paste: `<shift>-<insert>`
- d. On Bianca it is harder to copy between your computer and thinklinc
- e. but there is a "menu" to the left of the ThinLinc screen where there is a clipboard

47.

48. Suppose I have a long R script and I want to select some lines of the code. How can I do that?

- a. I am not sure what is meant. Could you clarify?

49. PLACEHOLDER

50. How to even enter batch mode? Feel like a missing part in instruction. Vim? Nano? Why not in instructions?

- a. Do you mean how to write a batch script? Use any editor you like; vim, nano, emacs...  
- Okay thank you for clarification. This is not stated anywhere in slides (or I might be 'blind' for this instruction ) so did not understand at all. Thanks!
- b. Then you submit the batch script with `sbatch`
- c. Yes, the batch script is an instruction that you send to the slurm scheduler. It contains both the commands that should be run (the job itself) and what allocations on the compute part of the cluster are needed

51. PLACEHOLDER

52. PLACEHOLDER

53. PLACEHOLDER

54. PLACEHOLDER

55. PLACEHOLDER

56. PLACEHOLDER

57. All recordings for the three days of the course will be uploaded to HPC2N's YouTube channel, to a playlist for the course: <https://www.youtube.com/watch?v=aCSMEGgITG8&list=PL6jMHLEmPVLxdCIIWrzseYkGoCCXjrukM>

58. Evaluation survey for today: <https://forms.office.com/e/05mVUtbn9>

59. Evaluation survey for the whole course: <https://forms.office.com/e/dnTnevNTeA>