How to use the de.NBI cloud RStudio Server?

Some of the datasets are quite large, and might overwhelm some laptops with low memory. Hence, we have setup a RStudio Server on the de.NBI cloud, which can be used by those who find their own laptop limited.

Beware that the use of this Server requires quite some work, so this should be first restricted to those who have failed loading/processing the datasets on their laptop!
In case of trouble, get help from the tutors!

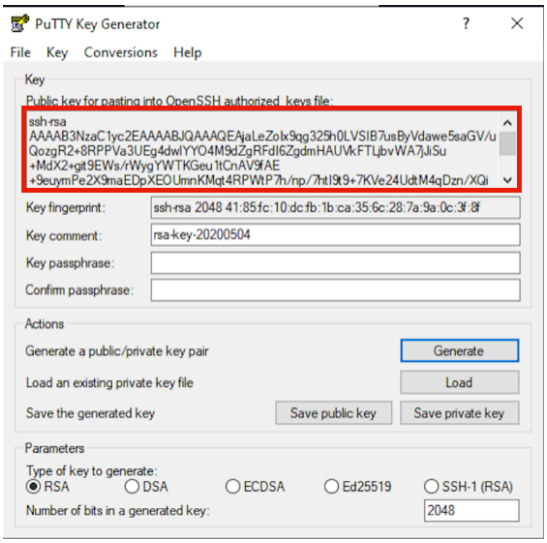
Here are the instructions, for Microsoft Windows (Windows) or MacOS (MacOS):

1. Create a pair of public/private ssh keys

In order to connect to the Virtual Machine (VM) hosting the RStudio server, you need to create a so-called **ssh key pair**:

- **Windows**: There are many different ways, here we focus on a robust solution based on PuTTY (https://www.putty.org/). PuTTY is an SSH Client enabling you to connect to your remote VM, please just install it on your local computer. The installation package also contains a tool called PuTTYGen. Click on 'Generate' and move your mouse cursor over the grey field to create some random input. Afterwards enter a passphrase and save your public and private keys into separate files e.g., bla_public.key (public key) and bla_private.ppk (private key).

Copy the public key (red frame) to the Google Sheet in the tab RStudio (beware, make sure to copy all lines!)
https://docs.google.com/spreadsheets/d/1jZ6fissYZsaxXeWwvzSioOhV_9f4er5HUzaLABEkJEk/edit?usp=sharing



- **MacOS**: open a terminal, and type the ssh-keygen (enter) command, and follow the instructions (press enter at each question). The command generates files in your home folder under .ssh/id_rsa.pub (public key) and .ssh/id_rsa (private key)
  Copy the string representing your public key (i.e. the content of the file .ssh/id_rsa.pub) to the corresponding column of the Google Sheet (tab Rstudio Server), with your name:
  https://docs.google.com/spreadsheets/d/1jZ6fissYZsaxXeWwvzSioOhV_9f4er5HUzaLABEkJEk/edit?usp=sharing

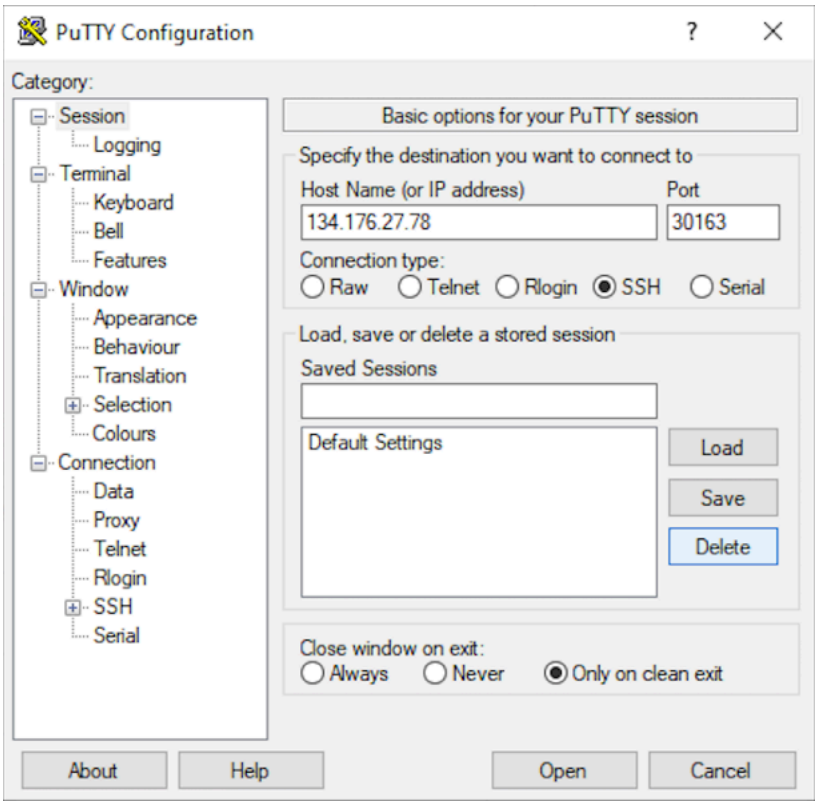
Once this is done, I will assign you a user name in the corresponding column.

2. Connect to the virtual machine

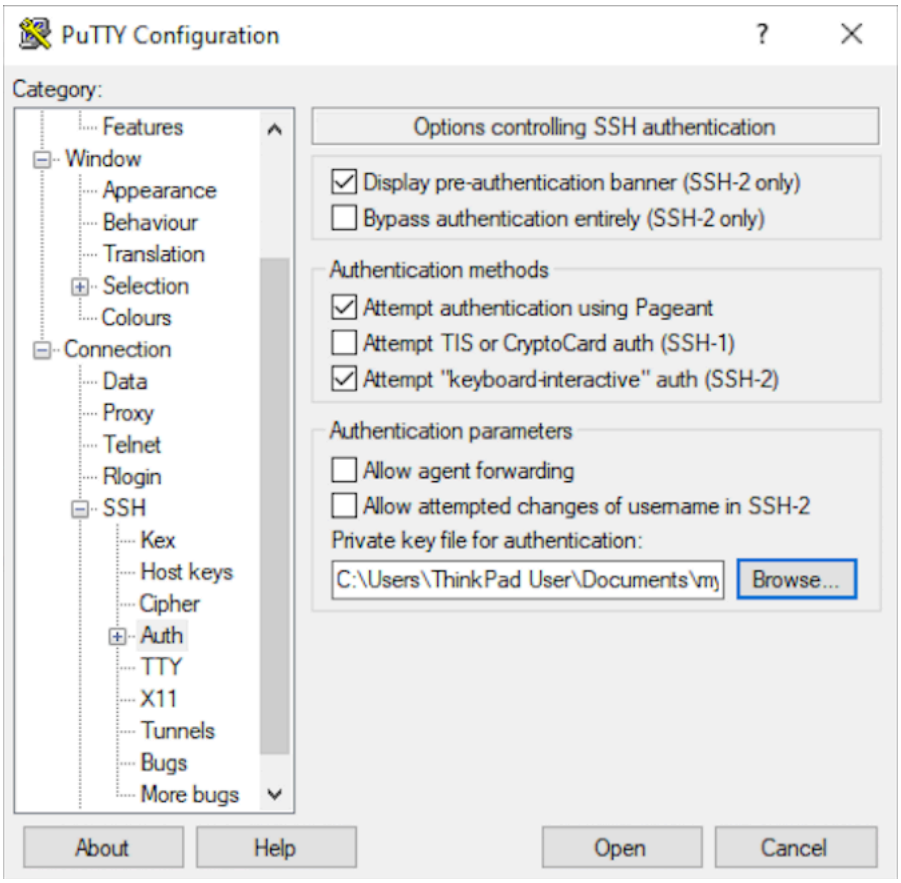
Once you have a user name indicated in the Google sheet, you can connect via ssh to the VM, using the following method:

Windows: use PUTTY

- In the session window, give the IP address 134.176.27.78 and the port 30163



In the Connection > SSH > auth window, indicate the path to your **private ssh key**:



- In the Connection > SSH > Tunnels window, indicate

    - **Source port: 8787**
    - **Destination: localhost:8787**
    - **Click add**

- Now go back to the Session window, and click **Save**
- Now click **Open**
- At the “login as” prompt, given your username (from the Google Sheet)

MacOS: use the command line

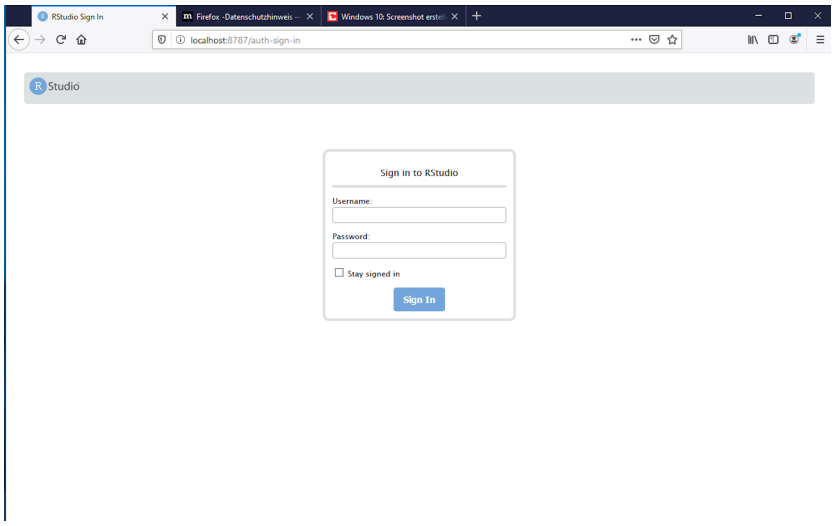
Open a terminal and type:

```
ssh -i ~/.ssh/id_rsa XXXXX@134.176.27.78 -p 30163 -L
8787:localhost:8787
```

Once you are connected…

- When you are connected (either through Windows or MacOS), open a web browser, and in the address bar, type
  `localhost:8787`

- You should now see a login screen for RStudio



- Now log-in using the same username, and the username as Password

3. Accessing datasets

- Once you are logged in RStudio, you can access the datasets for Project-01 in the folder
  `/home/ubuntu/data/project-01`

- Create a variable:

```
> data.dir = '/home/ubuntu/data/project-01'
> my.dir = '/home/XXXXX'
> setwd(my.dir)
```

  where XXXXX is your user name (the one you used to log into rstudio). The last command defines /home/XXXXX as the default working directory.

- Load the datasets with

```
> hallmarks = readRDS(file.path(data.dir,'hallmarks_genesets.rds'))
```

And likewise for the other datasets.