How to connect to Hadoop cluster via ssh

Igor Yakushin
`ivy2@uchicago.edu`

ssh

- For the next few lessons we would be connecting to Hadoop cluster via ssh
- You might want to review section 3 from my "Introduction to Linux and RCC" tutorial

  https://canvas.uchicago.edu/courses/23040

  where this is described in more details. If you do not have an access to the tutorial, ask at the office. It is a good idea to refresh your knowledge of Linux and review the whole tutorial since the lack of Linux skills is a continuous source of problems.
- You would need to have ssh client on your laptop.
  - Mac and Linux users already have it by default
  - Windows users might have to install, for example, MobaXterm

    https://mobaxterm.mobatek.net/
- Make sure to enroll in 2-factor authentication:

  https://2fa.rcc.uchicago.edu/

ssh

- You need to be either on campus network or use VPN to be able to connect to Hadoop cluster. VPN is managed by the university IT and can be downloaded here:
  https://uchicago.service-now.com/it?id=kb_article&sys_id=8d7e1ed0db91c0d0432f7f8cbf96195f
- Notice: you cannot use ThinLinc to connect to Hadoop cluster, it only works with midway
- To ssh to Hadoop cluster:
  - Connect to VPN
  - Open a terminal and type there:

    ssh YourUserName@hadoop.rcc.uchicago.edu
  - You'll be prompted for password and 2-factor authentication
  - Notice: if you mistype your password 3 times, you would be locked out of your account for 4 hours