How to use UCL HPC

Basic information on using High Performance Computing

Ph.D. student: Doy Kwon
doy.kwon.19@ucl.ac.uk
07946 599 645

Presenter

1. Doy Kwon (Chief Inspector)

2. Main arena of work
   - Crime prevention and security
   - Crime data science
   - Anti-crime policy making

3. History

03-10, in the field, patrol officer, detective, traffic, etc.
11-14, KNPA, Community Safety Bureau
14-’19, Lecturer in Police HRD Institute
19-, PhD student, Security and Crime Science in UCL
Prime Minister Awards in ‘The 33rd Public HRD Contest’ held at National Human Resources Development Institute in 2015

CONTENTS

CHAPTOR 01

Overview

Before, start a presentation

- **All information is in the websites below.**

(1) https://www.rc.ucl.ac.uk/docs/ UCL Research Computing Documentation

(2) http://rits.github-pages.ucl.ac.uk/hpc-intro/aio/index.html Introduction to HPC

(3) https://www.ucl.ac.uk/isd/services/research-it/research-computing-platforms-service
UCL Research Computing Platforms Service

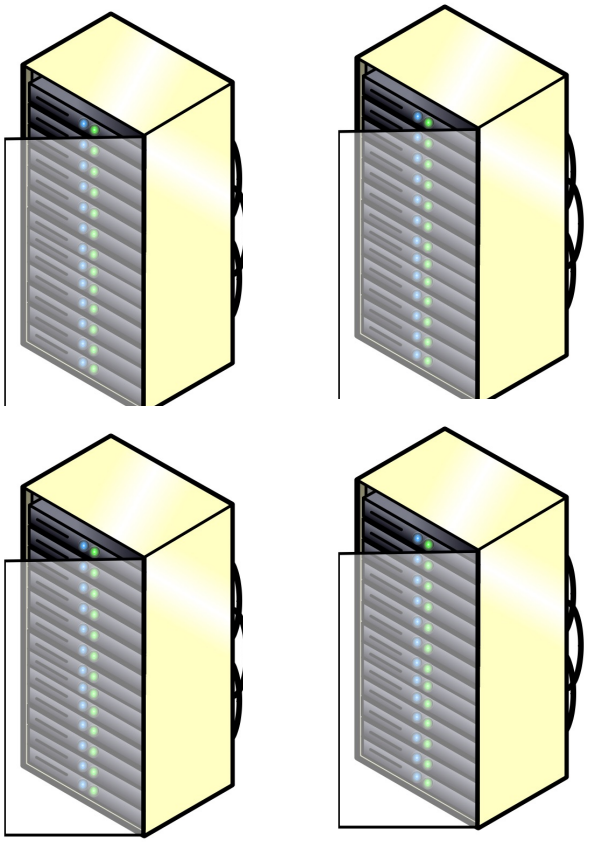
(4) https://www.rc.ucl.ac.uk/docs/Interactive_Jobs/ About using HPC interactively

(5) https://www.ucl.ac.uk/isd/services/get-connected/ucl-virtual-private-network-vpn/
UCL VPN

What is High Performance Computing?

- **People call it as “Server”, “Cluster”, and “Super-computer”.**
- **Anyway, HPC is bundles of computers out there to calculate large and complex data**

VS



Why do I use HPC?

- **Large memory : up to 1.5Tb RAM for single core**
- **Multicore : You can use over 128 cores at once (if you can code…)**
- **Save your resources : electricity, time, computer…**
- **You can run several codes simultaneously as if you have several computers at home**

(1) While you run codes in R on your laptop, you can not use R anymore for other job

(2) But! You can run several R codes simultaneously with HPC.

Differences of HPC compared to Personal com.

- **Basic command language is different : UNIX Shell (Linux)**

(1) Not based on GUI system such as Mac and Window. You should write codes in line like DOS program

(2) You can use R, Python or Matlab, but can't use Rstudio and Anaconda in HPC

- **You should submit a scheduling job script**

(1) According to the resources that you request, your job will be scheduled by the system organizer ("scheduler") automatically

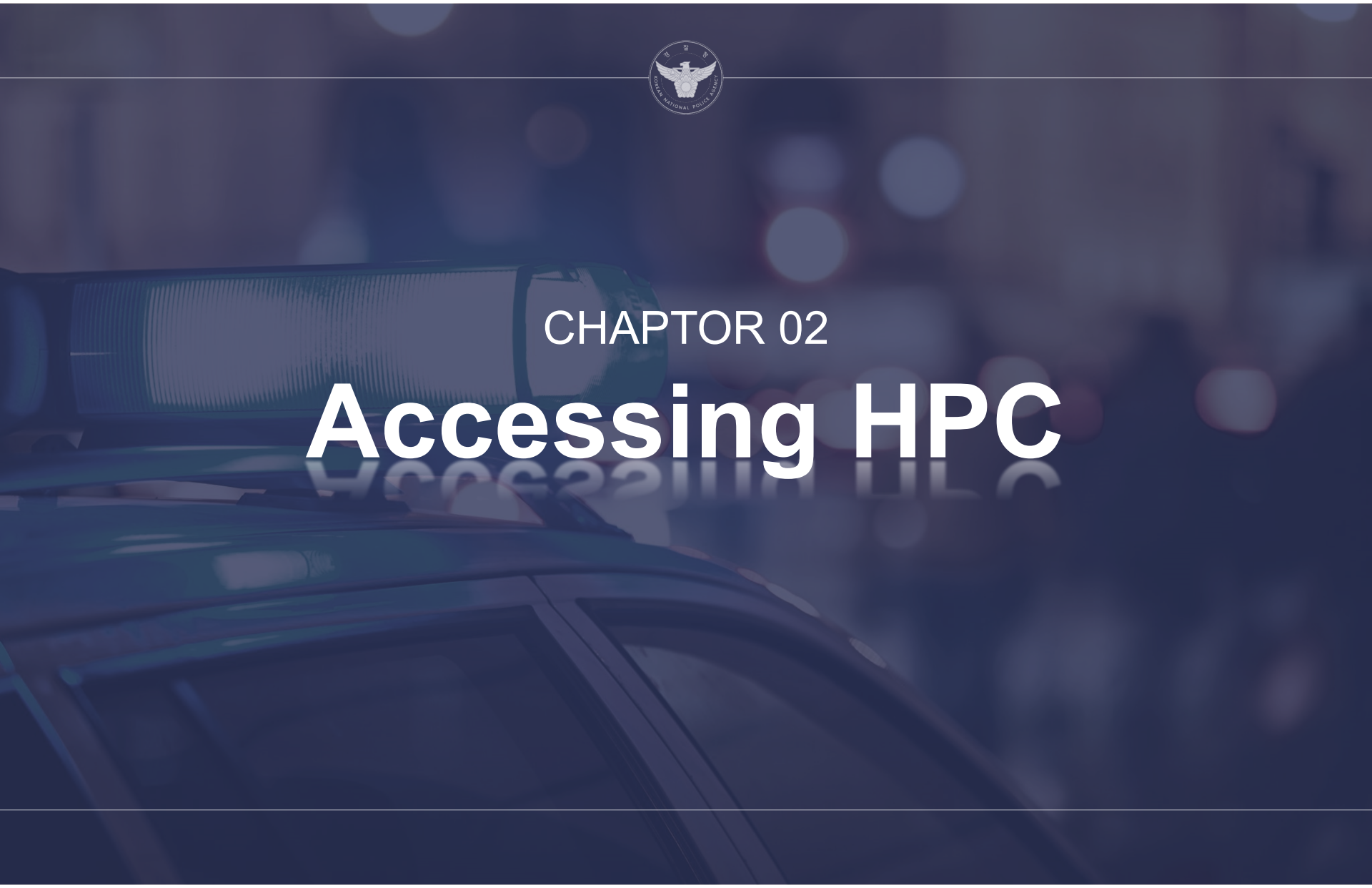
- **The versions of software is a little bit old for maintaining stability**

(1) Recent version of R is 4.0.3, but the version of R in HPC is 3.6.0

- **Interactive mode can be tested with limited resources**

(1) We can use R interactively in HPC, but only for 2hours with 512mb RAM

- **Basically, you can use only 150Gb for a storage in the server**



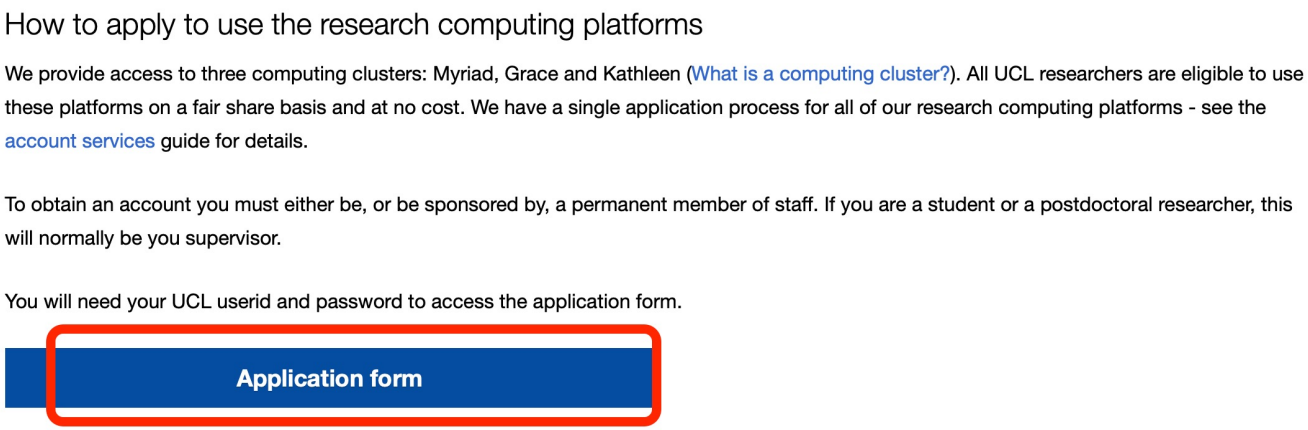
CHAPTOR 02

Accessing HPC

● You should get a permission to access HPC.

■ Write Application form

https://www.ucl.ac.uk/isd/services/research-it/research-computing-platforms-service



How to apply to use the research computing platforms

We provide access to three computing clusters: Myriad, Grace and Kathleen (What is a computing cluster?). All UCL researchers are eligible to use these platforms on a fair share basis and at no cost. We have a single application process for all of our research computing platforms - see the account services guide for details.

To obtain an account you must either be, or be sponsored by, a permanent member of staff. If you are a student or a postdoctoral researcher, this will normally be you supervisor.

You will need your UCL userid and password to access the application form.

Application form

■ The process of authorization

(1) You write these info in the application form : previous experience of using HPC, research project details, work descriptions, required resources, etc.

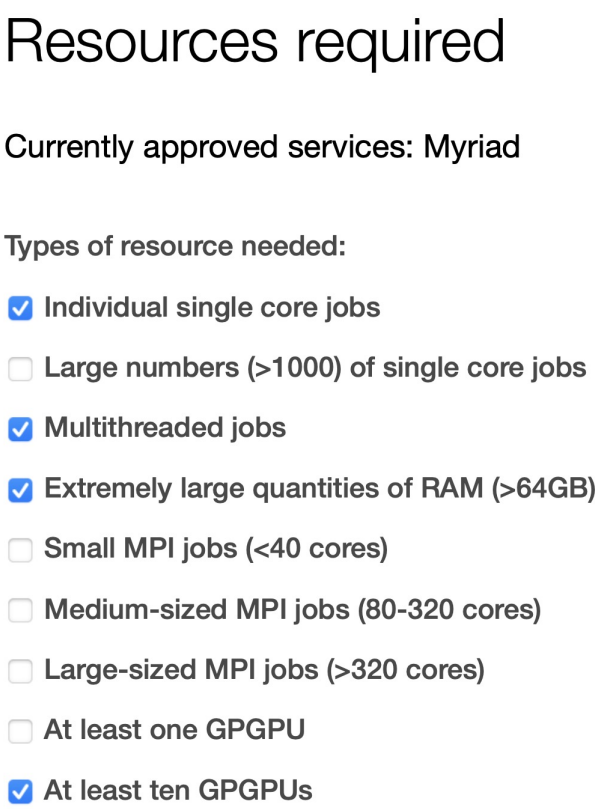
(2) Your supervisor will get email from RC-support team to approve your access using HPC.

(3) After permission, you will get a specific server name that is proper to your job by email.

● Choose “Resources required” in the form

- **There are several HPC servers in UCL**

(1) According to the “resources required” selection, you will be allocated specific server that fits for your work.



- **My work uses lots of memories than CPUs, so I chose like this.**
- **So, server named “Myriad” were allocated to me.**

How to login the Myriad server

- **Now you want to connect to the remote server with your laptop**

(1) Secure Shell Protocol (SSH) https://en.wikipedia.org/wiki/Secure_Shell_Protocol

(2) Just remember "ssh" is a command that connects your laptop to the server

- **Turn on your "terminal (Mac)" or "remote desktop connection (Win)"**

(1) When you are in UCL, you can connect the server directly.

(2) However!!! If you are out of UCL firewall, at home, you should connect to "Socrates" VPN first for security matters

(3) After log into Socrates, you can connect to the HPC server

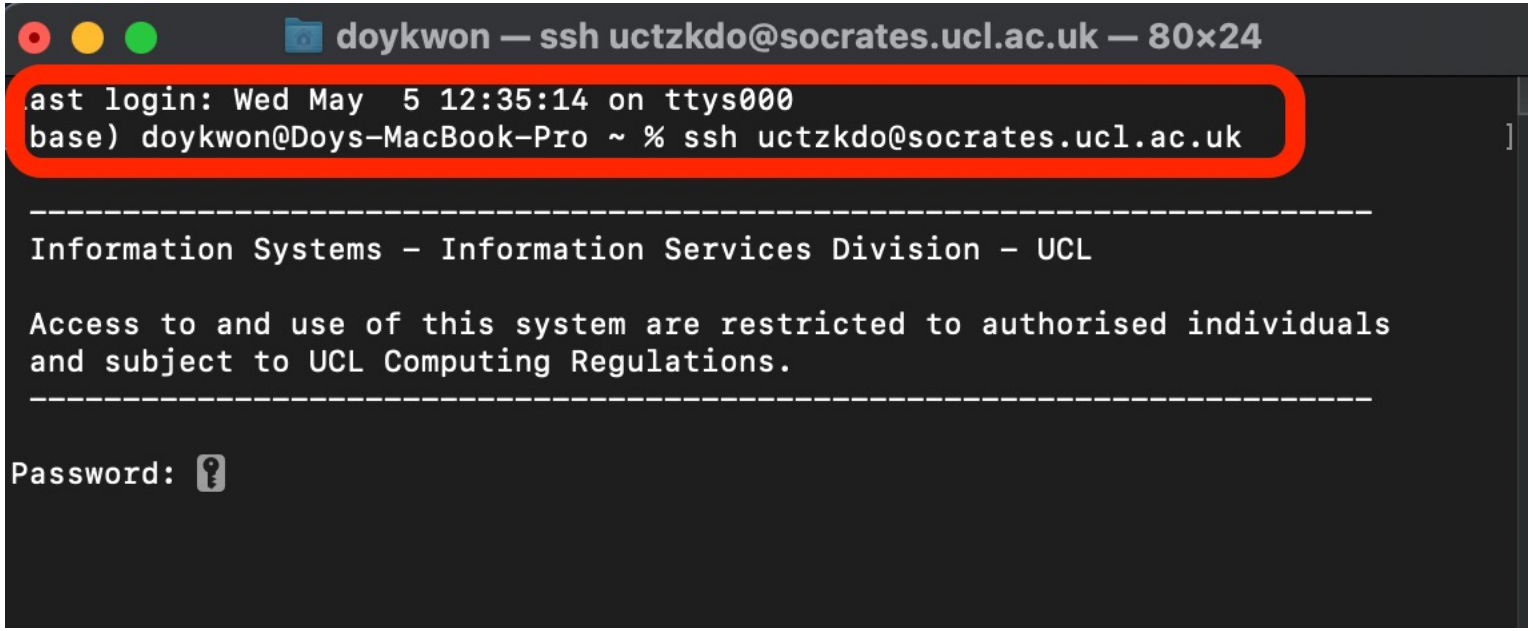
- **Connect to Myriad server with "ssh" command at home**

>ssh ucl_id@socrates.ucl.ac.uk
→ You should write your password
>ssh ucl_id@myriad.rc.ucl.ac.uk
→ You should write your password

Connect to Socrates VPN first 1

- **In your Mac, write codes below in terminal**

  > ssh userid@socrates.ucl.ac.uk



```
(base) doykwon@Doys-MacBook-Pro ~ % ssh uctzkdo@socrates.ucl.ac.uk

 ------------------------------------------------------------------------
 Information Systems - Information Services Division - UCL

 Access to and use of this system are restricted to authorised individuals
 and subject to UCL Computing Regulations.
 ------------------------------------------------------------------------

Password:
Last login: Wed May  5 12:41:52 2021 from ra-vpn179.ra-vpn.ucl.ac.uk

------------------------------------------------------------------------
---                This machine is managed by Puppet 6                ---
------------------------------------------------------------------------

Information Services Division - UCL

Access to and use of this system are restricted to authorised individuals
and subject to UCL Computing Regulations.

All IS systems run unattended overnight and at weekends.
If they fail services may not be restored until the next working day.

ISD Service Desk Hours

   Monday to Friday -  08.30 - 17.30  (Telephone Support)
   Monday to Friday -  09.30 - 17.00  (Personal Callers)

   Vacation times will vary

For ISD Service Desk contact information, see www.ucl.ac.uk/isd/servicedesk
For Service News, see www.ucl.ac.uk/isd/news

Please report any service issues to the ISD Service Desk.

PLEASE NOTE- Files in the /tmp file system are deleted at regular intervals.
Please ensure that any important files are stored in your home directory.
                    --------

5 %
```

> ssh userid@myriad.rc.ucl.ac.uk

```
PLEASE NOTE- Files in the /tmp file system are deleted at regular intervals.
Please ensure that any important files are stored in your home directory.
----------------------------

5 % ssh uctzkdo@myriad.rc.ucl.ac.uk
Password:
```

Then, connect to the specific HPC server

```
5 % ssh uctzkdo@myriad.rc.ucl.ac.uk
Password:
Last login: Tue Apr 27 12:38:20 2021 from socrates-a-d05.isdapp.ucl.ac.uk
                                UCL MYRIAD
                        Research IT Services

          ~/
             * is backed up
             * has a 150GB quota limit

          ~/Scratch
             * is NOT backed up
             * should be considered "at risk"
             * has a 1TB quota limit (by default)

          To request quota increases, and other support, contact:
             rc-support@ucl.ac.uk

 Access to and use of this system are restricted to authorised individuals and
 subject to UCL Computing Regulations.

             https://www.ucl.ac.uk/information-security/sites/information-security/files/Regulations_0.pdf

 In particular, please note that this service should not be used to store data
 for which you have obligations under the Data Protection Act 2018:

             https://www.ucl.ac.uk/legal-services/data-protection-overview

 For datasets obtained under contract, please ensure you are complying with
 contractual requirements.

 Myriad announcements mailing list:
             myriad-users@ucl.ac.uk

 Help and Support Pages:
             https://wiki.rc.ucl.ac.uk/wiki/Category:User_Guide

[uctzkdo@login12 ~]$
```

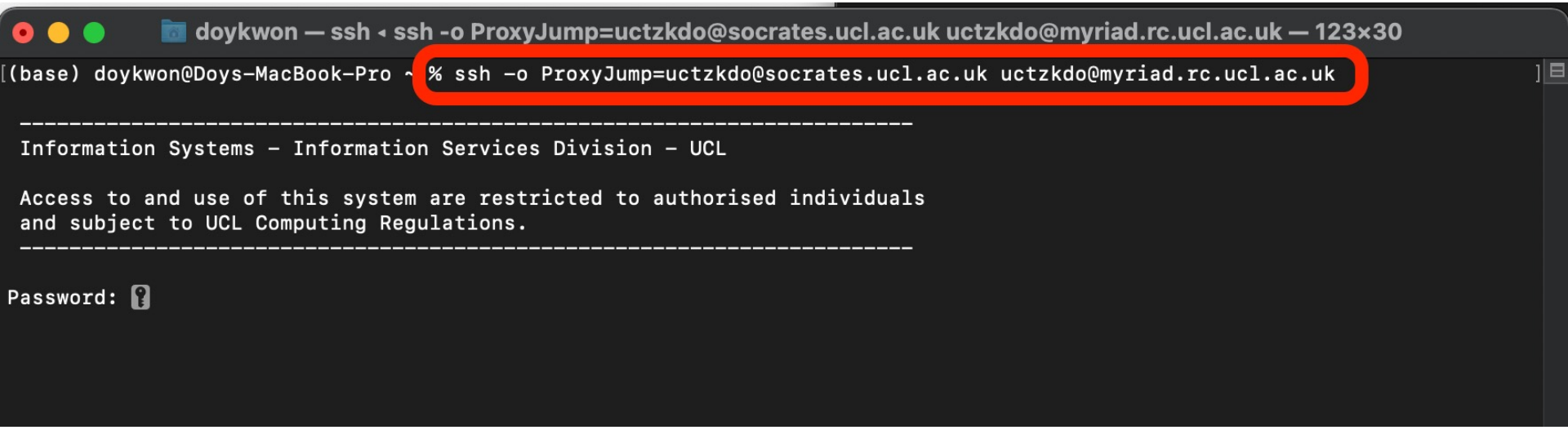
More easy way of logging to HPC by tunnelling

- **Instead, you can log in with one line of code by tunnelling**

(1) On your laptop "terminal", write code below

> ssh -o ProxyJump=uctzid@socrates.ucl.ac.uk uctzid@myriad.rc.ucl.ac.uk

you should write your ucl-id in "uctzid" above



- **After input passwords two times, you can log into HPC.**

CHAPTOR 03

Basic commands

UNIX Shell (Linux) basic commands

- **After logging on the server, you can input commands interactively.**
- **Here are basic commands that you will use on someday.**

No.	Command	Meaning	Usage
1	pwd	Path of this working directory	>pwd
2	ls	List the files and directories	> ls -l
3	cd	Change directory	> cd ..
4	cat	Concatenate. Make new txt file / show file contents	>cat file.txt
5	cp	Copy file or directory	>cp A B
6	mv	Move file	>mv A B
7	mkdir / rmdir	Make directory Remove directory	>mkdir name
8	rm	Remove file or directory	>rm file.name
9	*	all	>rm *.zip
10	nano	Open and edit file	>nano file.name

UNIX Shell (Linux) basic commands

```
[[uctzkdo@login12 ~]$ pwd
/home/uctzkdo
[[uctzkdo@login12 ~]$ ls -l
total 12
drwxr-xr-x 3 uctzkdo uctzp0 4096 Apr 13 17:11 R
lrwxrwxrwx 1 uctzkdo uctzp0   24 Feb  2 22:56 Scratch -> /scratch/scratch/uctzkdo
drwxr-xr-x 3 uctzkdo uctzp0 4096 Apr 15 09:30 data
drwxr-xr-x 2 uctzkdo uctzp0 4096 Mar  2 15:55 temp
[uctzkdo@login12 ~]$
```

Google “linux server basic commands” for more understanding

https://www.hostinger.com/tutorials/linux-commands

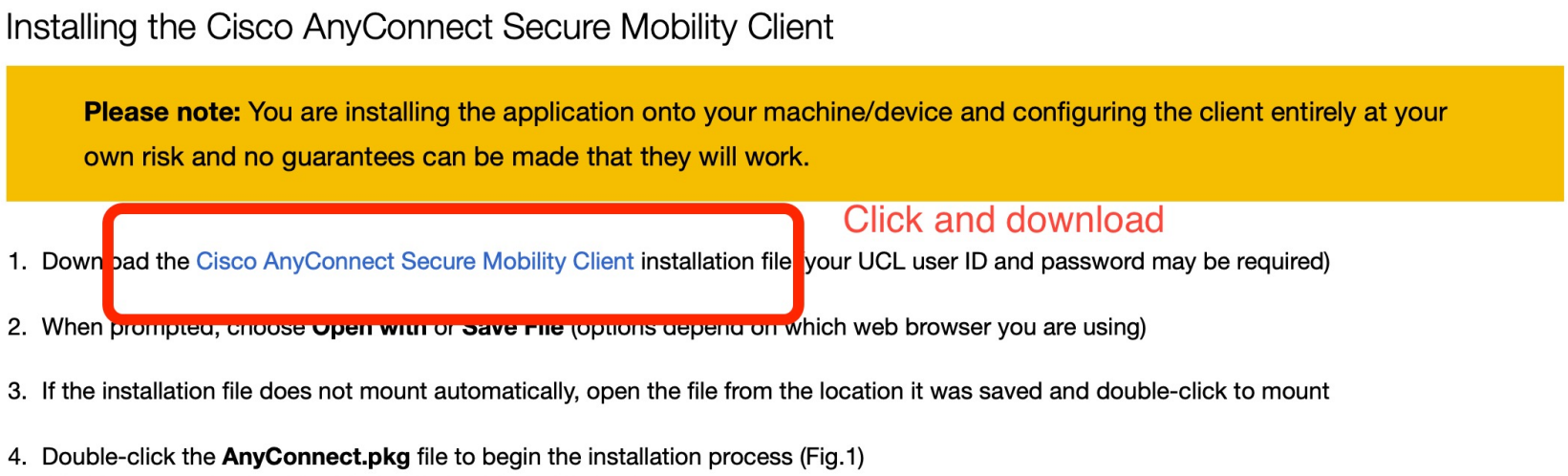
How can I transfer files in and out of HPC

- **If you are at home, outside of UCL firewall, you should use UCL VPN**
- **Without using UCL VPN, you cannot transfer files in and out of HPC**

https://www.ucl.ac.uk/isd/services/get-connected/ucl-virtual-private-network-vpn/

(1) Super easy, just go to the website, and choose proper file for your O/S

(2) You should install "Cisco AnyConnect Secure Mobility Client"



Codes for transferring file from laptop to HPC

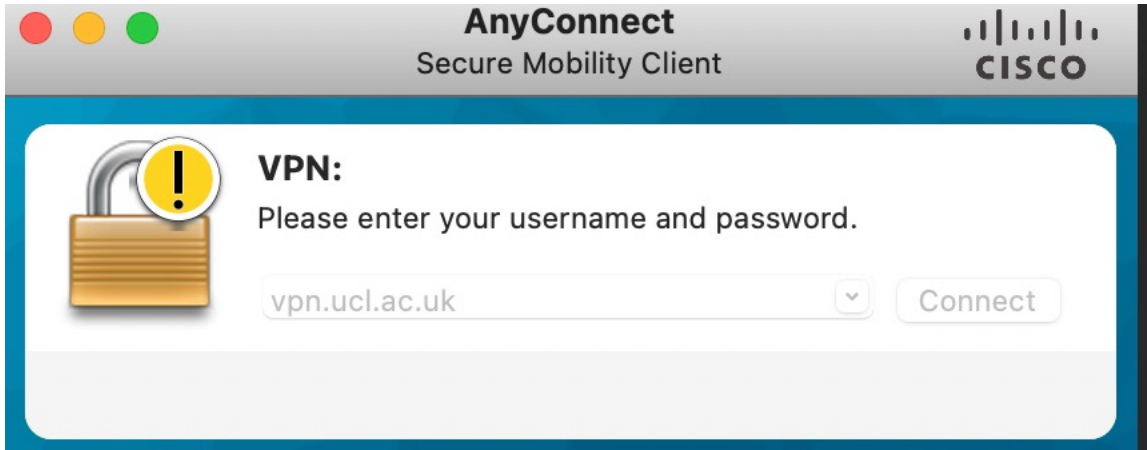
- **After log into UCL VPN (Cisco app), you can transfer file remotely in and out of HPC**

> rsync -avzP local/path/folder/file.csv uctzid@myriad.rc.ucl.ac.uk:target/directory

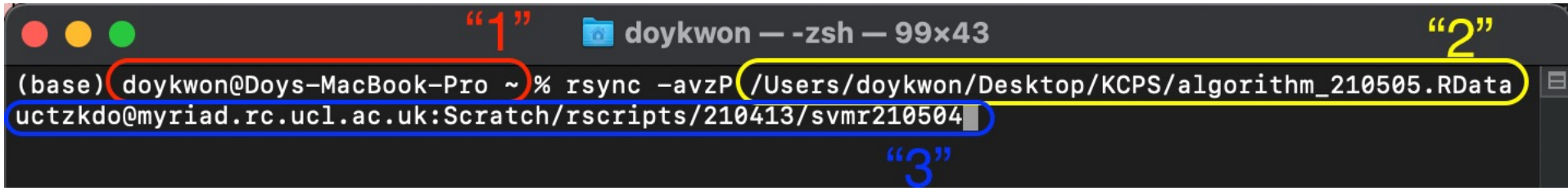
(1) Now, you want to move a file in your laptop to HPC.

(2) So, you should input your codes in laptop terminal not HPC terminal (you can turn on several terminals at the same time)

(3) Parameters : a (archive), v (verbose), z (compression), P (partial/progress)



● Codes for transferring a file from laptop to HPC



(1) "1" : you should write codes on your laptop terminal.

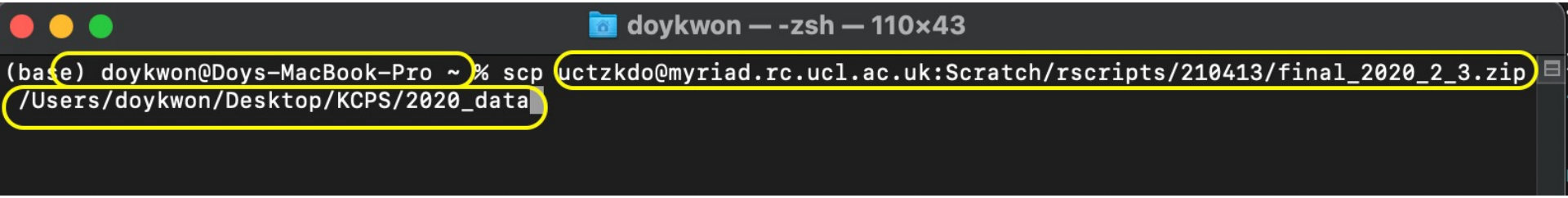
(2) "2" : Local path of file

(3) "3" : Target directory in HPC where the file will transfer

(4) You can transfer many file by making archive file (zip, tar... commands)

> rsync -avzP local/path/folder/file.csv uctzid@myriad.rc.ucl.ac.uk:target/directory

● Codes for transferring a file from HPC to laptop



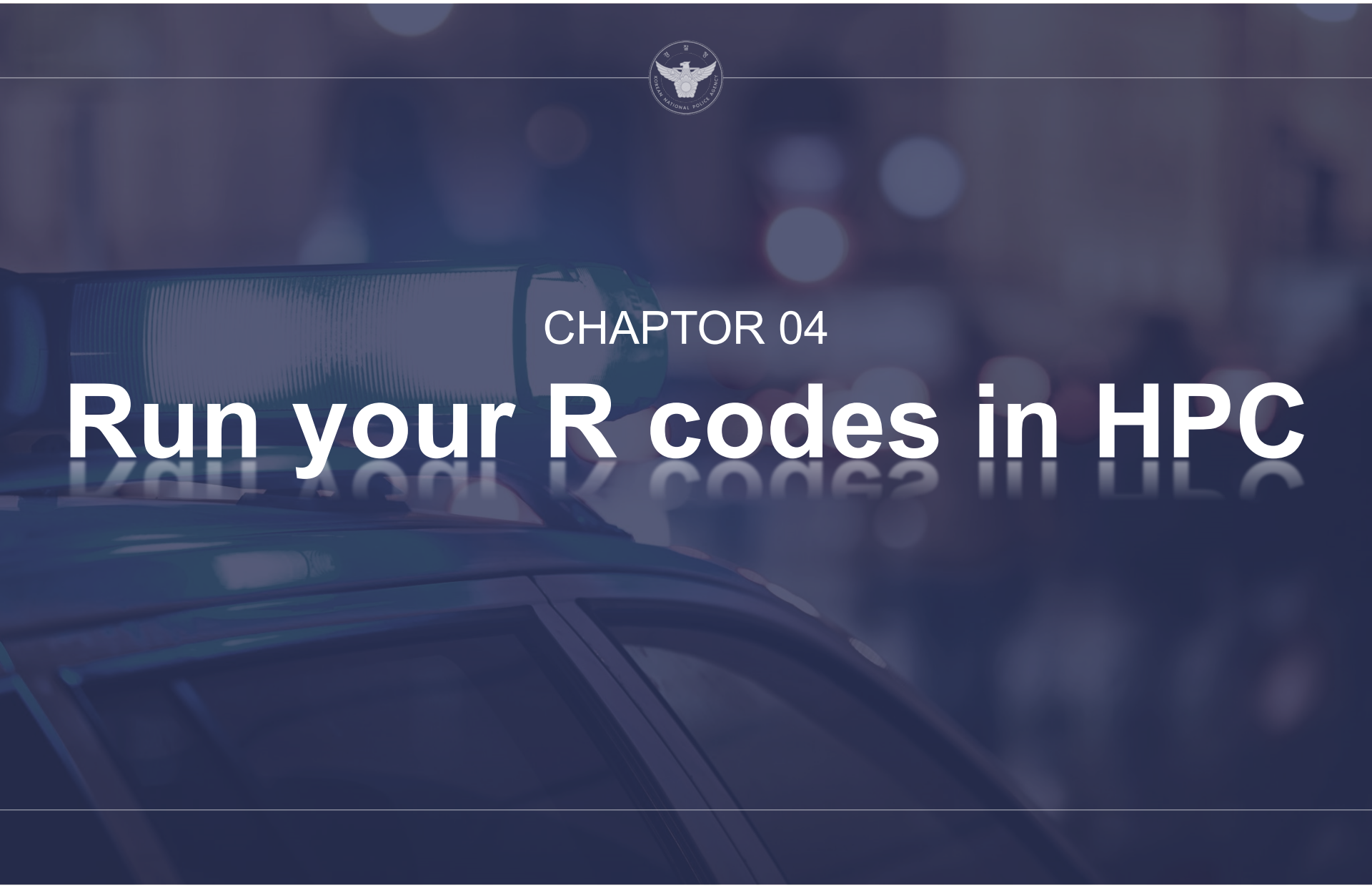
(1) You want to get a file from HPC to your laptop

(2) You should write codes on your laptop terminal.

(3) Write directory and file name in HPC first and then write local path

(4) "scp" commands also conduct file transferring, but "rsync" is much faster

- rsync -avzP uctzid@myriad.rc.ucl.ac.uk:directory/file.csv local/path/folder
- scp uctzid@myriad.rc.ucl.ac.uk:directory/file.csv local/path/folder



CHAPTOR 04

Run your R codes in HPC

Job script file? Batch job?

- **In HPC server, you should make and submit a job script file for each task.**
- **HPC scheduler program will read your job script file and consider where to put the job on the nodes according to the size of requested resources.**
- **This process is called batch job submission.**

Job script file is ending with "file.sh"

- Requesting resources : memory, CPU cores, processing time
- Running batch file
- https://www.rc.ucl.ac.uk/docs/Software_Guides/R/ examples of R job script file

So, you need to have two files for a task in HPC.

One is "Job script file" and "Code file" that you want to run.

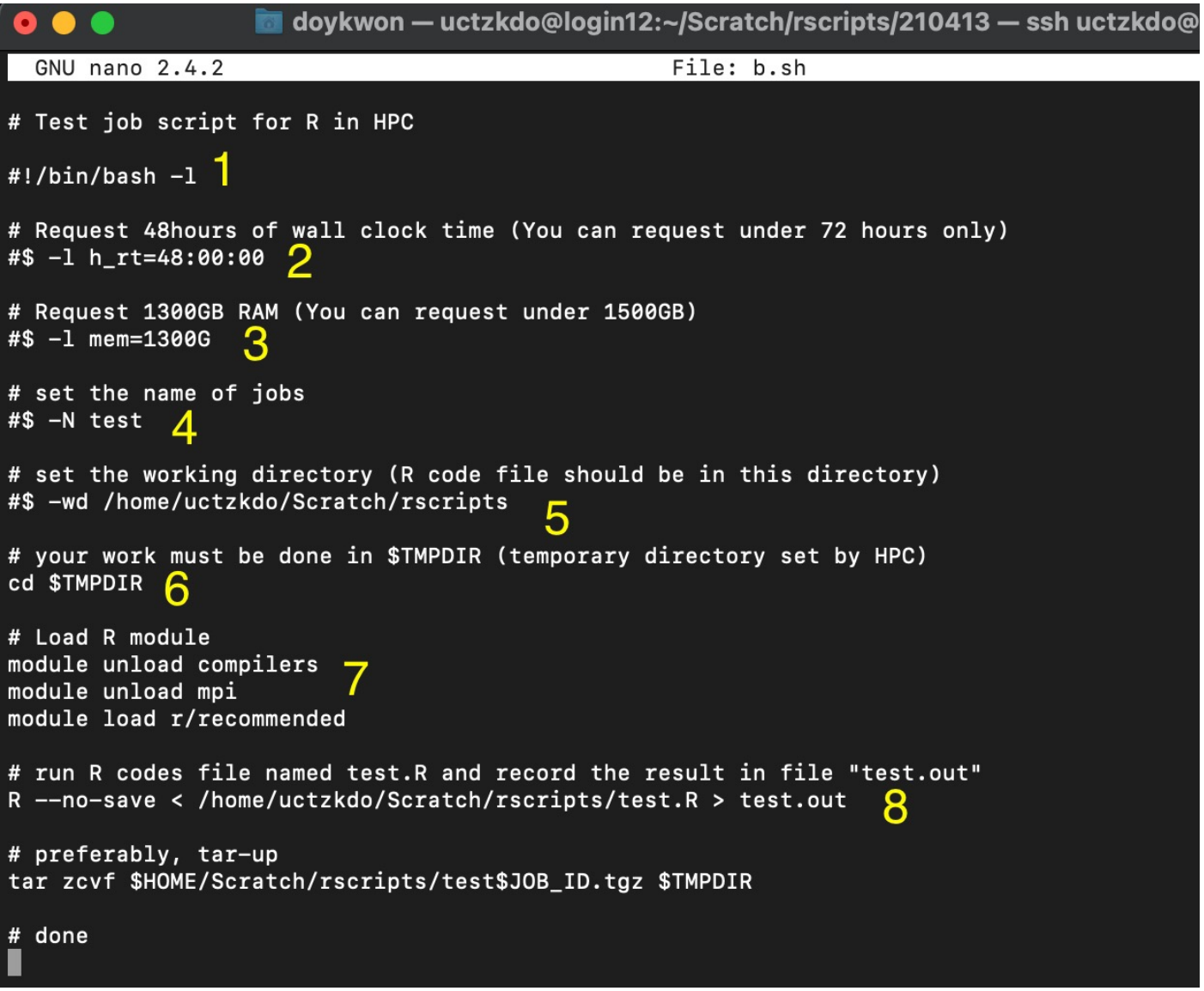
● Nano editor

- **When you make code files including batch, R, python, etc, you will use nano editor commands in HPC.**

No.	Command	Meaning	Usage
1	nano	Open the file	>nano batch.R
2	Ctrl + o	Save the file	
3	Ctrl + x	Exit nano editor	

- **# : when you start the code with # (Hash), it means comments, so the program ignore the line after #.**

My real Job script code for R



>cat batch.sh

>nano batch.sh

1: Hash Bang (compulsory)

2: Request time

3: Request RAM

4: set the name of job

5: Set WD

6: Just do it (compulsory)

7: load software R

8: Set the location of R code file (VIP)

My real R code for HPC

```
  GNU nano 2.4.2                    File: r_algo_svmr_210505.R

# HPC scripts for building Kernel Support Vector Machine classifier for cybercrime with KCPS dataset
options(java.parameters="-Xmx8g")

library(rJava)
library(NLP)
library(NLP4kec)  # This package is for morphology anlaysis of Korean, based on rJava
library(data.table)
library(dplyr)
library(lubridate)
library(tidyverse)
library(stringi)
library(stringr)
library(ggplot2)
library(tm)
library(tidymodels)
library(chron)
library(magrittr)
library(RTextTools)
library(gmodels)
library(e1071)
library(caret)
library(readxl)

setwd("/lustre/home/uctzkdo/Scratch/rscripts/210413/svmr210504")

load("algorithm_210505.RData")

cf_svmr <- train(form = labels ~., data = training_set_svm2, method = 'svmRadial')

saveRDS(cf_svmr, "cf_SVMR_0505.rds")

predict_svmr <- predict(cf_svmr, test_set_svm)

save.image("algorithm_210507.RData")

# END
```

>cat file.R

>nano file.R

- Just R script
- Load libraries
- Set working directory
- Build codes for my task

Submit my job and check the situation

- **Submission my job**

  > qsub batchfile.sh

- **Check the situation of my job once**

  > qstat

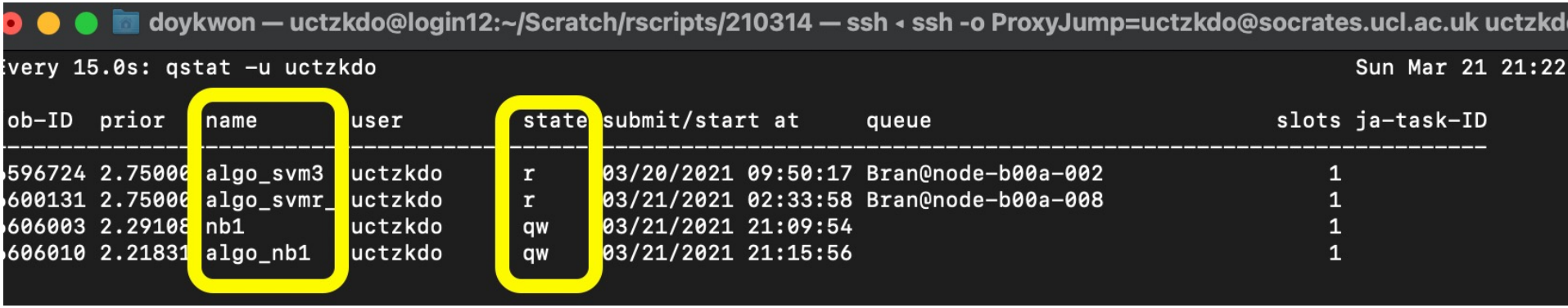
- **Check the situation of my job constantly**

  > watch -n 30 –u uctzid

- **Killing (withdrawal) my job**

  > qdel job-number

Submit my job and check the situation



- **Watch the job situation constantly every 15 sec.**

Name : the job name you wrote in batch file (job submission file)

State : the state of each job

- **r = running**
- **qw = queuing**
- **Eqw = Error** → you should fix codes in batch file
- dr = this job is being deleted
- Rr = this job was rescheduled but is now running on a new node (rare)

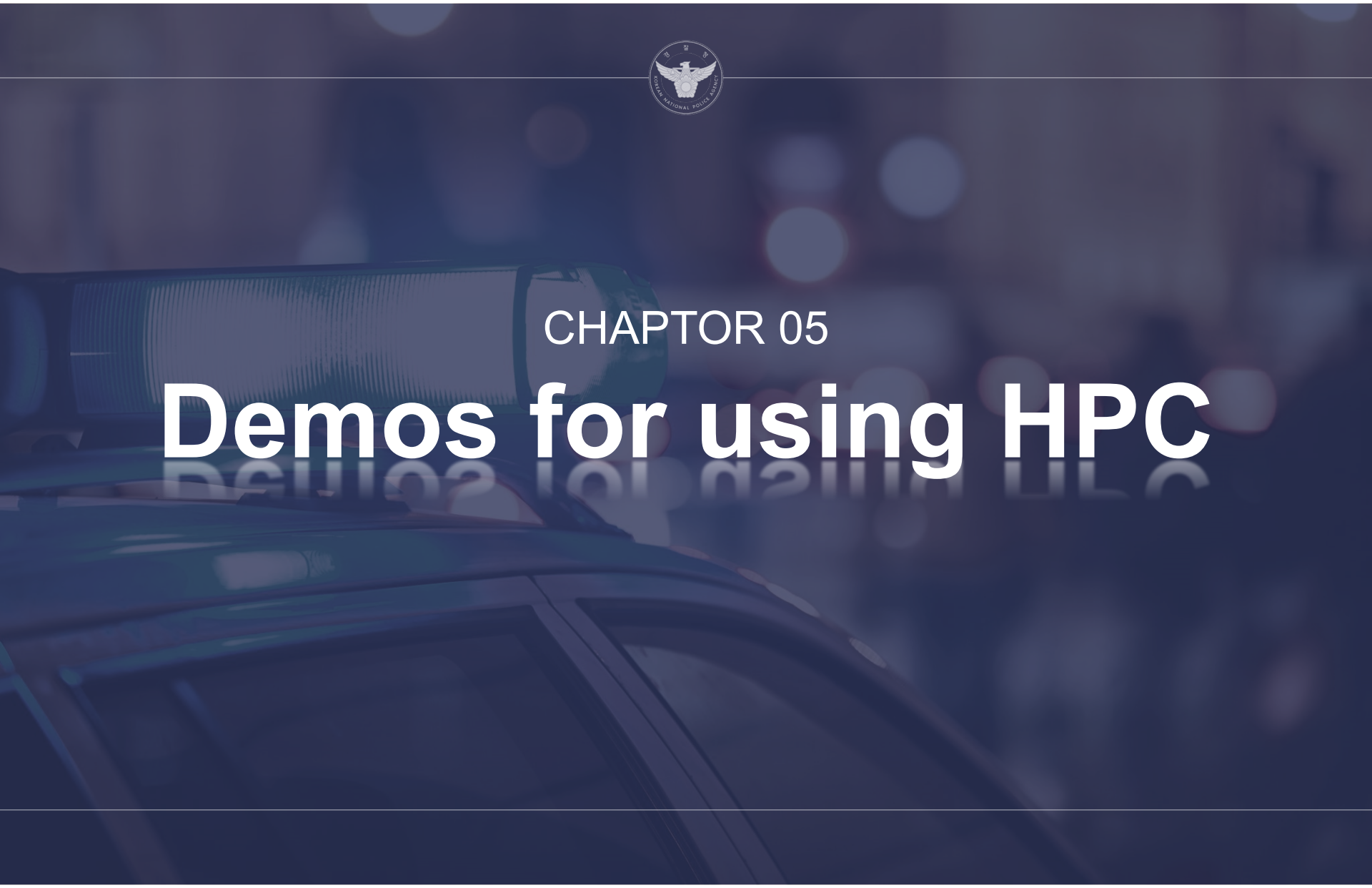
Using R interactively in HPC

- **When you install R packages or test your code, you can use R interactively in HPC**
- **But, the resources are limited (512mb RAM, 2hours).**

Just type these 4 lines in your HPC terminal

```
> module unload compilers
> module unload mpi
> module load r/recommended
> R --interactive
```

You can use R interactively. (not like Rstudio, just "R")



CHAPTOR 05

Demos for using HPC

Question & Discussion