



# KATANA FOR R

Fonti Kar  
she/her

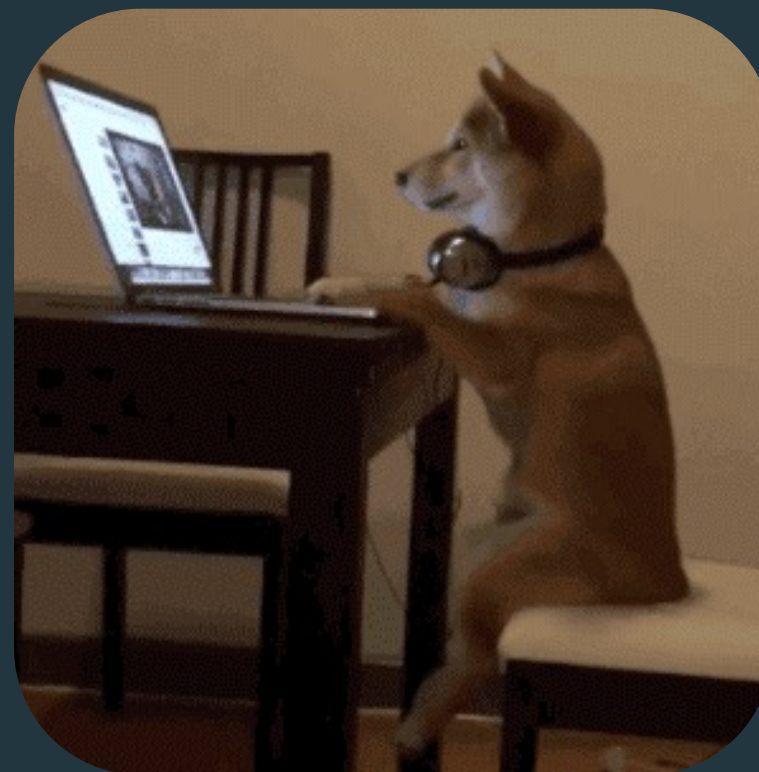
Workshop materials: <https://github.com/UNSW-codeRs/katana4R>

If you haven't already, please:

- Request for Katana access (takes a few days to process)
  - <https://research.unsw.edu.au/katana>
- Download and install FileZilla from <https://filezilla-project.org/>

# OUTLINE

- What is Katana?
- The Workflow
- File transfer with FileZilla
- Job scripts
- Shell terminal, Katana on Demand
- Using screen
- Submitting your job



# HI! THIS IS KATANA.



Log in node

# HI! THIS IS KATANA.



Log in node



Head node

Acquire resources  
Directed to a computing node

# HI! THIS IS KATANA.



Log in node



Head node

Acquire resources  
Directed to a computing node



Computing node(s) / cluster






UNSW wide / research center  
You get allocated to a node *within a cluster*\*

# WHEN DO I NEED KATANA?






- Time-consuming analyses
- Many iterative processes
- Computer is having a hard time



# WHAT DO I NEED FOR KATANA?

- Get access. Apply with UNSW IT Service 
- Download and install FileZilla 
- Project hygiene 
- Self-contained workflow 
- Shell Terminal / Katana On Demand (KOD) 

# WHAT DO I NEED FOR KATANA?

- Get access. Apply with UNSW IT Service 
- Download and install FileZilla 
- Project hygiene 
- Self-contained workflow 
- Shell Terminal / Katana On Demand (KOD) 



# THE WORKFLOW

## 1. Write a job script:



- How much computational resources you require
- R version you need
- Which .R script to run

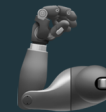
## 2. Upload data, .R script + job script:



- FileZilla
- Katana Data Mover

## 3. Submit job script via Shell Terminal

- Check the status of job
- Delete the job
- Run more jobs



## 4. Download outputs from your job:

- FileZilla
- Katana Data Mover



# FILE TRANSFER WITH FILEZILLA



- Best way to move files between Katana and your computer
- Files are moved to via the Katana Data Mover server `kdm.restech.unsw.edu.au`

Local site: /Users/fontikar/Dropbox/katana\_workshop/

Filename	Filesize	Filetype	Last modified
..			
.Rproj.user		Directory	08/04/2021 08:...
R		Directory	07/04/2021 15:4...
data		Directory	07/04/2021 15:4...
output		Directory	06/04/2021 14:2...
.RData	4,766,090	File	06/04/2021 02:0...
.Rhistry	6,116	File	08/04/2021 15:4...
Katana_101_Day1.mp4	87,769,669	mp4-file	08/04/2021 11:3...
Katana_101_slides.pdf	2,652,948	pdf-file	07/04/2021 13:5...

5 files and 4 directories. Total size: 95,195,028 bytes

Remote site: /home/z3516573/katana\_workshop

Filename	Filesize	Filetype	Last modified	Permissions	Owner/Group
..					
katana_workshop...	205	R Project	08/04/2021 1...	-rw-r--r--	z3516573 ...
Katana_101_slides...	2,652,948	pdf-file	08/04/2021 1...	-rw-r--r--	z3516573 ...
.Rhistry	6,085	File	08/04/2021 1...	-rw-r--r--	z3516573 ...
.RData	4,766,090	File	08/04/2021 1...	-rw-r--r--	z3516573 ...
.DS_Store	6,148	File	08/04/2021 1...	-rw-r--r--	z3516573 ...
output		Directory	08/04/2021 1...	drwxr-xr-x	z3516573 ...
data		Directory	08/04/2021 1...	drwxr-xr-x	z3516573 ...
R		Directory	08/04/2021 1...	drwxr-xr-x	z3516573 ...

5 files and 4 directories. Total size: 7,431,476 bytes

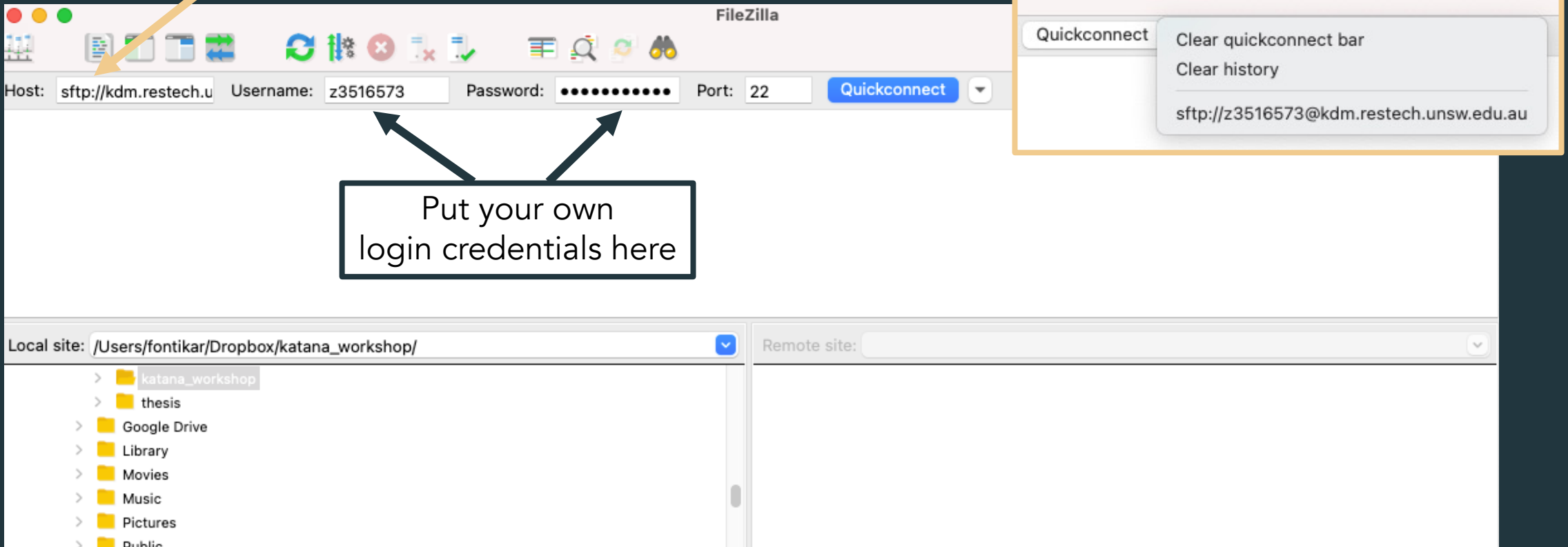
/home/z1234567

# SETTING UP FILEZILLA



via Quickconnect

sftp://kdm.restech.unsw.edu.au



- Set up once and find the KDM in drop down menu\*

# UPLOADING TO KOD



## Storage on Katana

Name of directory	File path	Backed up?	Storage limit	Best for
Home drive	/home/z1234567	Y	10GB	Code, small data (<1GB)
Global scratch	/srv/scratch/z1234567	N	16TB	Shared data, or big data

# EXERCISE



1. Using FileZilla, Create a **pbs/** folder on Katana in HOME directory
2. Upload **katana\_workshop** to GLOBAL SCRATCH

Name of directory	File path	Backed up?	Storage limit	Best for
Home drive	/home/z1234567	Y	10GB	Code, small data (<1GB)
Global scratch	/srv/scratch/z1234567	N	16TB	Shared data, or big data

# REQUESTING COMPUTATIONAL RESOURCES



**WALL TIME** - How long you need Katana for

- Biggest constraint
- <12 hrs can run anywhere, shorter wait
- >12 hrs can run on BEES only nodes



**RAM MEMORY** - How much working memory you need

- Important for spatial analyses e.g. raster data
- Lots of data? Need more memory for faster access



**CPU CORES** - How much processing power you need

- Less of an issue if working with a single model + simple tabulated data
- Can you optimise your code to run in parallel?

# REQUESTING COMPUTATIONAL RESOURCES

Yeah well... how much resources do I need? 🙄

Every project's needs are **different**

- Get to know your analyses
- Monitor computer activity (Task Manager, Activity Monitor)

**Trial-and-error**

- Test your analysis out with less data or fewer iterations



**If in doubt, start with defaults** and work from there

- Budgeting for slightly more but not too much

💡 **Tip:**  
Do these  
systematically!

# JOB SCRIPTS



- The instructions you leave with Katana so your job(s) can run autonomously
- Simple `.txt` file (Notepad, Text Editor, Sublime Text 3) with `.pbs` suffix
- Upload\* using FileZilla in a `pbs/` folder 
- Self-contained workflow 

`.pbs`

1. Resources you need (nodes, CPU, RAM, wall time)
2. R version you need
3. File path to the relevant working directory
4. Which `.R` script to run




# .pbs

1. Resources you need (nodes, CPU, RAM, wall time)
2. R version you need
3. File path to the relevant working directory
4. Which .R script to run

All scripts must begin with  
**`#!/bin/bash`**

```
1  #!/bin/bash *
```



```
2
3  #PBS -N job_name
4  #PBS -l select=1:ncpus=1:mem=1gb
5  #PBS -l walltime=1:00:00
6  #PBS -M your.name@student.unsw.edu.au
7  #PBS -m ae
8
9  #Clear Katana workspace
10 module purge
11
12 #Loading R version
13 module add R/4.0.2
14
15 #Navigate to working direction
16 cd /srv/scratch/z123456/katana_workshop
17
18 #Run R code
19 Rscript R/script_name.R
```

# .pbs

1. Resources you need (nodes, CPU, RAM, wall time)
2. R version you need
3. File path to the relevant working directory
4. Which .R script to run

Give your job a short, informative name  
e.g. `growth_glm_logit`

```
1  #!/bin/bash
2
3  #PBS -N job_name
4  #PBS -l select=1:ncpus=1:mem=1gb
5  #PBS -l walltime=1:00:00
6  #PBS -M your.name@student.unsw.edu.au
7  #PBS -m ae
8
9  #Clear Katana workspace
10 module purge
11
12 #Loading R version
13 module add R/4.0.2
14
15 #Navigate to working direction
16 cd /srv/scratch/z123456/katana_workshop
17
18 #Run R code
19 Rscript R/script_name.R
```

# .pbs

1. Resources you need (nodes, CPU, RAM, wall time)
2. R version you need
3. File path to the relevant working directory
4. Which .R script to run

`select`: Number of nodes  
`ncpu`: Number of CPU processors  
`mem`: Amount of RAM memory

`walltime`: Length of job (hours)

```
1  #!/bin/bash
2
3  #PBS -N job_name
4  #PBS -l select=1:ncpus=1:mem=1gb
5  #PBS -l walltime=1:00:00
6  #PBS -M your.name@student.unsw.edu.au
7  #PBS -m ae
8
9  #Clear Katana workspace
10 module purge
11
12 #Loading R version
13 module add R/4.0.2
14
15 #Navigate to working direction
16 cd /srv/scratch/z123456/katana_workshop
17
18 #Run R code
19 Rscript R/script_name.R
```

# .pbs

1. Resources you need (nodes, CPU, RAM, wall time)
2. R version you need
3. File path to the relevant working directory
4. Which .R script to run

Email for Katana to reach you




```
1  #!/bin/bash
2
3  #PBS -N job_name
4  #PBS -l select=1:ncpus=1:mem=1gb
5  #PBS -l walltime=1:00:00
6  #PBS -M your.name@student.unsw.edu.au
7  #PBS -m ae
8
9  #Clear Katana workspace
10 module purge
11
12 #Loading R version
13 module add R/4.0.2
14
15 #Navigate to working direction
16 cd /srv/scratch/z123456/katana_workshop
17
18 #Run R code
19 Rscript R/script_name.R
```

# .pbs

1. Resources you need (nodes, CPU, RAM, wall time)
2. R version you need
3. File path to the relevant working directory
4. Which .R script to run

When should Katana to email you?  
a = when job aborts with error  
e = when job ends naturally




```
1  #!/bin/bash
2
3  #PBS -N job_name
4  #PBS -l select=1:ncpus=1:mem=1gb
5  #PBS -l walltime=1:00:00
6  #PBS -M your.name@student.unsw.edu.au
7  #PBS -m ae
8
9  #Clear Katana workspace
10 module purge
11
12 #Loading R version
13 module add R/4.0.2
14
15 #Navigate to working direction
16 cd /srv/scratch/z123456/katana_workshop
17
18 #Run R code
19 Rscript R/script_name.R
```

# .pbs

1. Resources you need (nodes, CPU, RAM, wall time)
2. R version you need
3. File path to the relevant working directory
4. Which .R script to run

Clearing Katana's global environment. Ensures you are working on a clean slate



```
1  #!/bin/bash
2
3  #PBS -N job_name
4  #PBS -l select=1:ncpus=1:mem=1gb
5  #PBS -l walltime=1:00:00
6  #PBS -M your.name@student.unsw.edu.au
7  #PBS -m ae
8
9  #Clear Katana workspace
10 module purge
11
12 #Loading R version
13 module add R/4.0.2
14
15 #Navigate to working direction
16 cd /srv/scratch/z123456/katana_workshop
17
18 #Run R code
19 Rscript R/script_name.R
```

# .pbs

1. Resources you need (nodes, CPU, RAM, wall time)
2. R version you need
3. File path to the relevant working directory
4. Which .R script to run

Loads the R version you need\*

💡 **Tip:** In Katana, type `module avail R/` into terminal to see what R versions are available

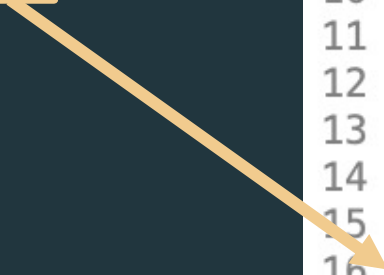
```
1  #!/bin/bash
2
3  #PBS -N job_name
4  #PBS -l select=1:ncpus=1:mem=1gb
5  #PBS -l walltime=1:00:00
6  #PBS -M your.name@student.unsw.edu.au
7  #PBS -m ae
8
9  #Clear Katana workspace
10 module purge
11
12 #Loading R version
13 module add R/4.0.2
14
15 #Navigate to working direction
16 cd /srv/scratch/z123456/katana_workshop
17
18 #Run R code
19 Rscript R/script_name.R
```



# .pbs

1. Resources you need (nodes, CPU, RAM, wall time)
2. R version you need
3. File path to the relevant working directory
4. Which .R script to run

Change directory to global scratch to set it as a working directory



```
1  #!/bin/bash
2
3  #PBS -N job_name
4  #PBS -l select=1:ncpus=1:mem=1gb
5  #PBS -l walltime=1:00:00
6  #PBS -M your.name@student.unsw.edu.au
7  #PBS -m ae
8
9  #Clear Katana workspace
10 module purge
11
12 #Loading R version
13 module add R/4.0.2
14
15 #Navigate to working direction
16 cd /srv/scratch/z123456/katana_workshop
17
18 #Run R code
19 Rscript R/script_name.R
```



# .pbs

1. Resources you need (nodes, CPU, RAM, wall time)
2. R version you need
3. File path to the relevant working directory
4. Which .R script to run

Run specified .R script  
within R/ folder



```
1  #!/bin/bash
2
3  #PBS -N job_name
4  #PBS -l select=1:ncpus=1:mem=1gb
5  #PBS -l walltime=1:00:00
6  #PBS -M your.name@student.unsw.edu.au
7  #PBS -m ae
8
9  #Clear Katana workspace
10 module purge
11
12 #Loading R version
13 module add R/4.0.2
14
15 #Navigate to working direction
16 cd /srv/scratch/z123456/katana_workshop
17
18 #Run R code
19 Rscript R/script_name.R
```

# FILE PATHS IN .pbs + .R MUST BE RELATIVE

**.pbs**

```
1  #!/bin/bash
2
3  #PBS -N job_name
4  #PBS -l select=1:ncpus=1:mem=1gb
5  #PBS -l walltime=1:00:00
6  #PBS -M your.name@student.unsw.edu.au
7  #PBS -m ae
8
9  #Clear Katana workspace
10 module purge
11
12 #Loading R version
13 module add R/4.0.2
14
15 #Navigate to working direction
16 cd /srv/scratch/z123456/katana_workshop
17
18 #Run R code
19 Rscript R/script_name.R
```

**.R**

```
# Install/load any packages
install.packages("pacman")

# Installs and library() if package is not available
pacman::p_load(readr, dplyr, purrr)

# Read in data
plant <- read_csv("data/Plant_height.csv")

# Exclude Herb/Shrub
plant %>%
  filter(! growthform == "Herb/Shrub" &
         ! growthform == "Shrub/Tree" &
         ! growthform == "Fern") -> plant_processed

# Split data as a list
plant_ls <- split(plant_processed, plant_processed$growthform)

# Run model for each growth form and output the coefficients
plant_output <- plant_ls %>%
  map(~ lm(loght ~ temp, data = .)) %>%
  map(~ summary(.))

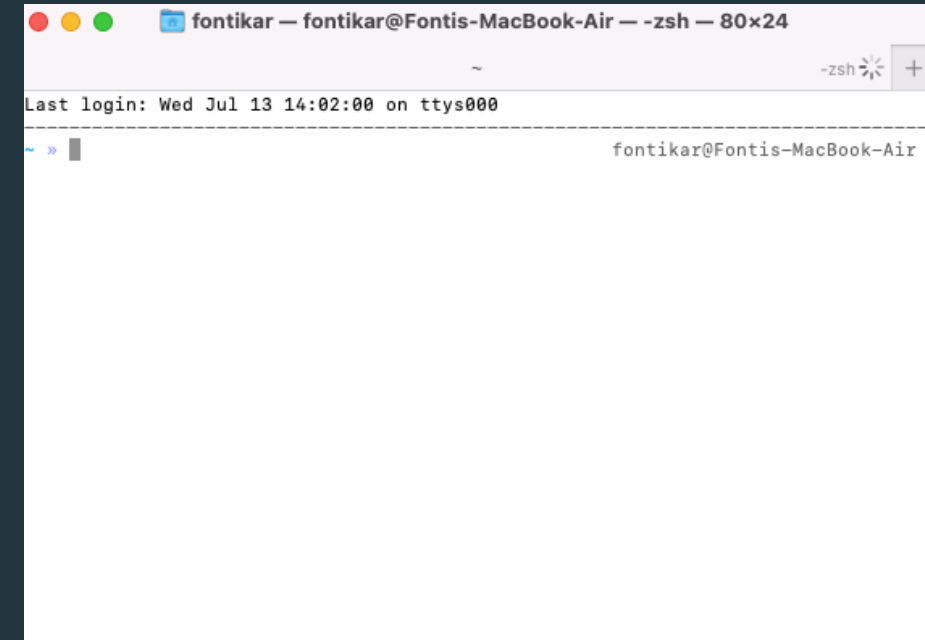
# Save the model output
saveRDS(plant_output, "output/growthform_lm_output")
```

**katana\_workshop** is the project root

# SHELL TERMINAL



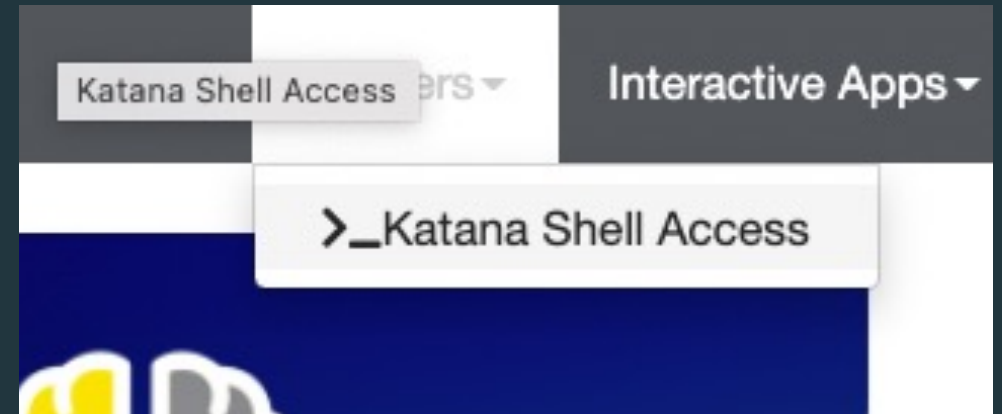
- Windows: Windows Terminal App
- Mac: Terminal zsh
- Katana on Demand:
  - <https://kod.restech.unsw.edu.au>
  - Caveats: Access via UNSW network or VPN



# KATANA ON DEMAND (KOD)

<https://kod.restech.unsw.edu.au>

- Click on Clusters
- Click on \_Katana Shell Access
- Enter your credentials
- You're at the head node of Katana!



```
z3516573@katana1.gen's password:
Last login: Wed Jul 13 14:02:07 2022 from l-fvff936zq6lt.ad.unsw.edu.au

*****
*       You have reached the KATANA High Performance Computing Cluster       *
*****

This service is for authorised clients only.  Please LOG OFF now if you are
not authorised to use this service.
```

# EXERCISE



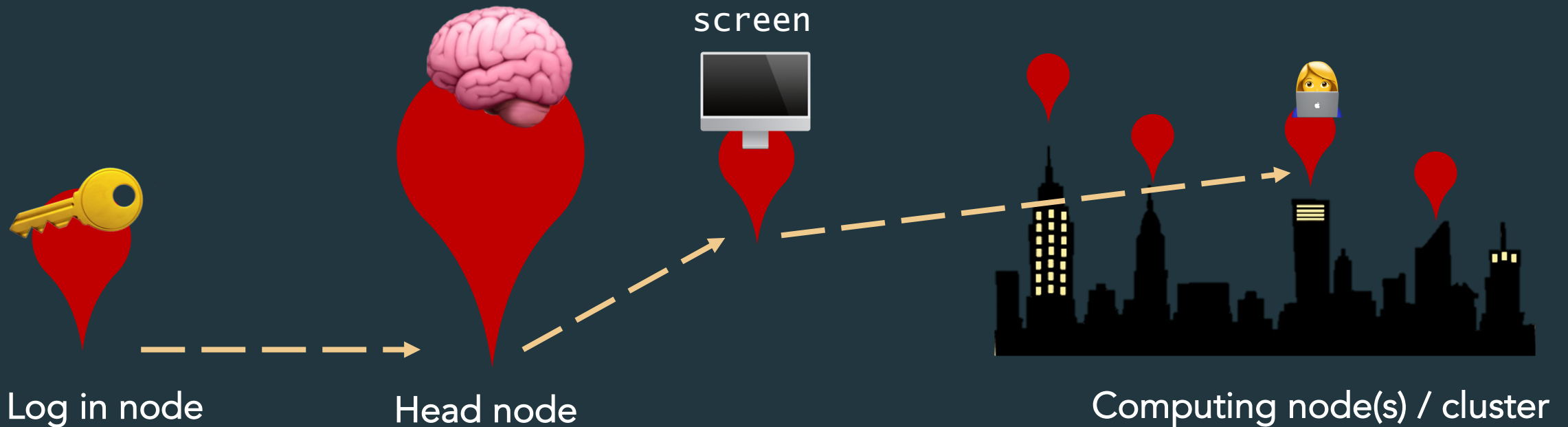
1. Make a copy of the `pbs_jobscript_template`
2. Change the email and zID and save it for future use!
3. Edit your own **.pbs** script
  - Request 30 mins, 1GB RAM, 1 CPU
  - Leave **module add/** as template
  - Navigate to **/srv/scratch/katana\_workshop**
  - Rscript **R/growthform\_lm\_analysis.R**
4. Upload your **.pbs** script to **HOME/pbs/**

Give it a go!

# SUBMITTING YOUR JOB



- `screen` is a program that allows you to submit jobs to Katana
- Think of it like a waiting room



# SUBMITTING YOUR JOB



- **screen** is a program that allows you to submit jobs to Katana
- Think of it like a waiting room



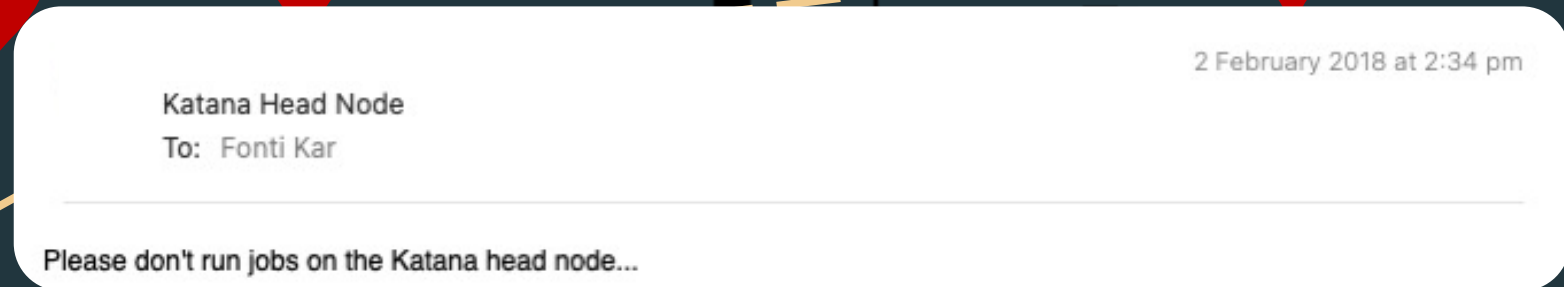
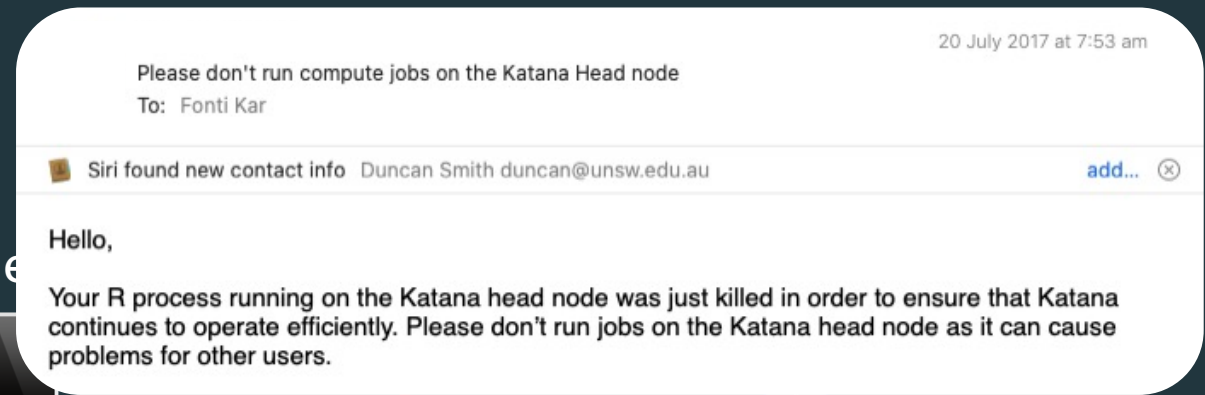
Log in node



Head node



screen



Computing node(s) / cluster

# screen CHEATSHEET

## screen

Function	Keystrokes [    ] = enter together	Example
Create new screen	<code>screen -S name_of_screen</code>	<code>[z3516573@katana1 ~]\$ screen -S test</code>
Detach from screen	<code>[Ctrl a] d</code>	<code>[detached from 11723.test]</code>
Reattach to screen	<code>screen -r name_of_screen</code> <code>screen -r #prefix</code> <code>screen -r -d name_of_screen</code> *for attached screens	<code>[z3516573@katana1 ~]\$ screen -r test</code> <code>[z3516573@katana1 ~]\$ screen -r 117</code>
List all screens	<code>screen -ls</code>	<code>[z3516573@katana1 ~]\$ screen -ls</code> There are screens on: 6968.test (Detached) 11723.test (Detached)
Delete a screen	<code>screen -X -S name_of_screen quit</code>	<code>[z3516573@katana1 ~]\$ screen -X -S 69 quit</code>



# KATANA CHEATSHEET

## Katana

### Function

### Keystrokes

### Example

Submit your job `qsub pbs/job_script_name.pbs`

```
[z3516573@katana1 ~]$ qsub pbs/my_job_script.pbs
```

Check on the  
status of job(s)

```
qstat -u $USER
qstat -u z123456
qstat -su $USER
```

```
[z3516573@katana1 ~]$ qstat -u $USER
```

for more detailed info

Delete a job

```
qdel JOBID
```

```
Job ID      Username Queue
-----
1134119.kman.re z3516573 bees12
[z3516573@katana1 ~]$ qdel 1134119
```

Reduce  
resources  
\*only for queued jobs

```
qalter -l select=1:ncpus=4:mem=512mb
JOBID
```

```
[z3516573@katana1 ~]$ qalter -l select=1:ncpus=4:mem=4 1134122
```

# EXERCISE



Let's give it a go!

1. Log into Katana OnDemand <https://kod.restech.unsw.edu.au>
2. Go to Clusters > \_Katana Shell Access, enter credentials
3. Create a new screen

```
@katana1 ~]$ screen -S test
```
4. Queue your .pbs job and detach

```
@katana1 ~]$ qsub pbs/my_job_script.pbs|
[Ctrl a] d
```
5. Check status

```
@katana1 ~]$ qstat -u $USER
```
6. Alter your RAM memory to 2GB

```
@katana1 ~]$ qalter -l
```

# SO, YOU WANT TO BYPASS KOD?

- Shell is platform and OS specific 🤪
- Katana access via secure shell (ssh)
  - e.g `ssh z123456@katana1.restech.unsw.edu.au`
  - Two servers `katana1` and `katana2`
  - If you just type `katana@unsw.edu.au` = random 🙈

💡 **Idea:** Find people using the same OS and set it up together if you are super keen!

# WE MADE IT FOLKS.



## Want more?

- Coding and R workshops in T2 <http://environmentalcomputing.net/>
- Drop-in help sessions, tutorials + nerdy events <https://unsw-coders.netlify.app/>

## Need help?

- <https://research.unsw.edu.au/katana>
- [https://unsw-restech.github.io/help\\_and\\_support.html](https://unsw-restech.github.io/help_and_support.html)
- Hacky Hour <https://research.unsw.edu.au/hacky-hour>