

Mike Wallis - IS Research Services - 2022/23

# Introduction to RSS

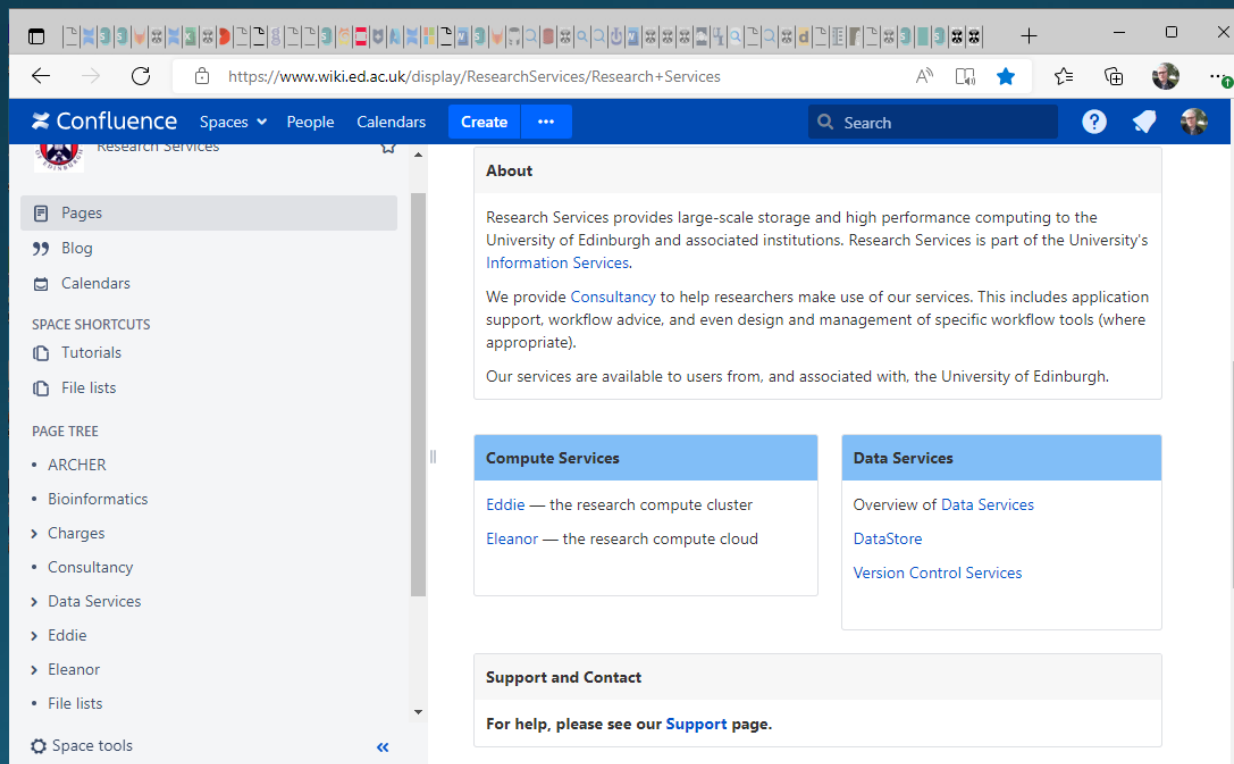
CERSE, 15<sup>th</sup> March 2023

# What is Research Computing?

- Doing anything that involves computers and research.
  - Stats
  - Visualisation
  - Image manipulation
  - Data entry and recording
  - Training programs and diet plans
  - Writing code to do things
  - Metanalysis
  - ...all kinds of things.
  - If users use a computer for research they can talk to us for help

# The RSS Wiki

- <https://www.wiki.ed.ac.uk/display/ResearchServices>



The screenshot shows a web browser displaying the Confluence Research Services Wiki page. The browser's address bar shows the URL <https://www.wiki.ed.ac.uk/display/ResearchServices/Research+Services>. The page features a blue header with the Confluence logo and navigation options like 'Spaces', 'People', 'Calendars', and 'Create'. A search bar is also present. The main content area is divided into sections: 'About', 'Compute Services', 'Data Services', and 'Support and Contact'. The 'About' section describes the services provided to the University of Edinburgh. The 'Compute Services' section lists 'Eddie' and 'Eleanor'. The 'Data Services' section lists 'Overview of Data Services', 'DataStore', and 'Version Control Services'. The 'Support and Contact' section provides a link to the 'Support' page.

**Confluence** Spaces People Calendars Create ... Search

Research Services

Pages  
Blog  
Calendars

SPACE SHORTCUTS  
Tutorials  
File lists

PAGE TREE  
• ARCHER  
• Bioinformatics  
> Charges  
• Consultancy  
> Data Services  
> Eddie  
> Eleanor  
• File lists  
Space tools

**About**

Research Services provides large-scale storage and high performance computing to the University of Edinburgh and associated institutions. Research Services is part of the University's [Information Services](#).

We provide [Consultancy](#) to help researchers make use of our services. This includes application support, workflow advice, and even design and management of specific workflow tools (where appropriate).

Our services are available to users from, and associated with, the University of Edinburgh.

**Compute Services**

[Eddie](#) — the research compute cluster  
[Eleanor](#) — the research compute cloud

**Data Services**

[Overview of Data Services](#)  
[DataStore](#)  
[Version Control Services](#)

**Support and Contact**

For help, please see our [Support](#) page.

# What I'm going to talk about

- Storage options (datastore)
- High-performance/throughput compute (eddie)
- Virtual machines (eleanor)

# What we offer for storage

- DataStore
- DataSync
- DataShare
- DataVault
- Version control systems
- Eddie parallel filesystem storage

# Why use Datastore?

- Accessible from anywhere on campus
- Resilient, backed up
- Free for majority of users
- Group space access managed by AD
- 500Gb quota for research staff

It is not encrypted at rest; for sensitive data we do suggest encrypting on disk.

# Why use DataSync?

- Available off-site
- Resilient
- Free at point of use
- Collaborative with external partners
- Can create shares with password/limited time expiry
- Secure transfer

# Eddie – HPC/HTC for UoE research





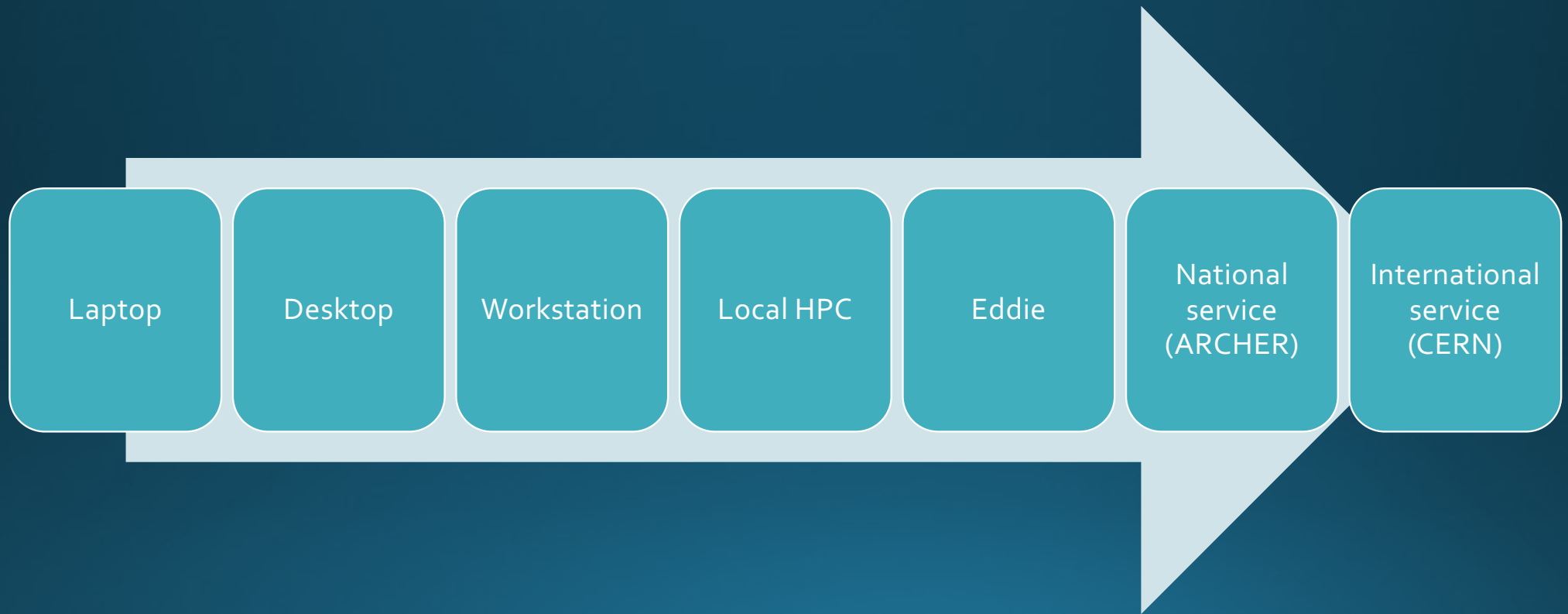
# What is Eddie?

- Supercomputers can be very specialised hardware (eg Cray) or off-the shelf components
- Clever low latency networks
- Very fast shared (parallel) storage
- Lots of CPUs
- Lots of RAM
- Energy dense

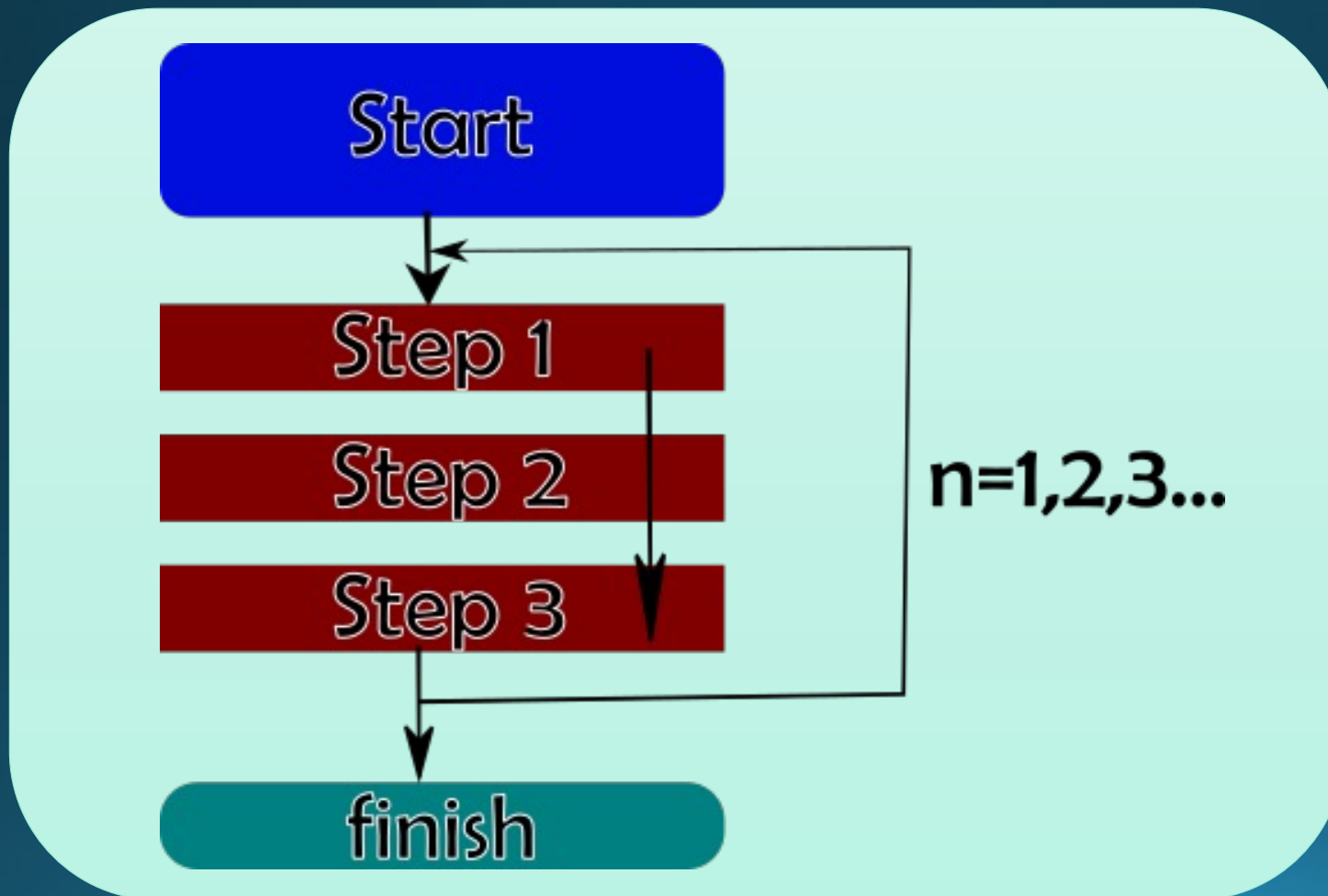
# Eddie

- 3<sup>rd</sup> generation: joint procurement between The University, IGMM, Roslin Institute and others
- Accessible by any researcher in The University
- Free at point of use
- ~400 nodes, ~11000 cores, ~110TB total RAM
- Workloads are scheduled to run when resources are available
- Eddie uses the Linux operating system (Scientific Linux 7.x)
- Quickstart: <https://edin.ac/38C3pxd>

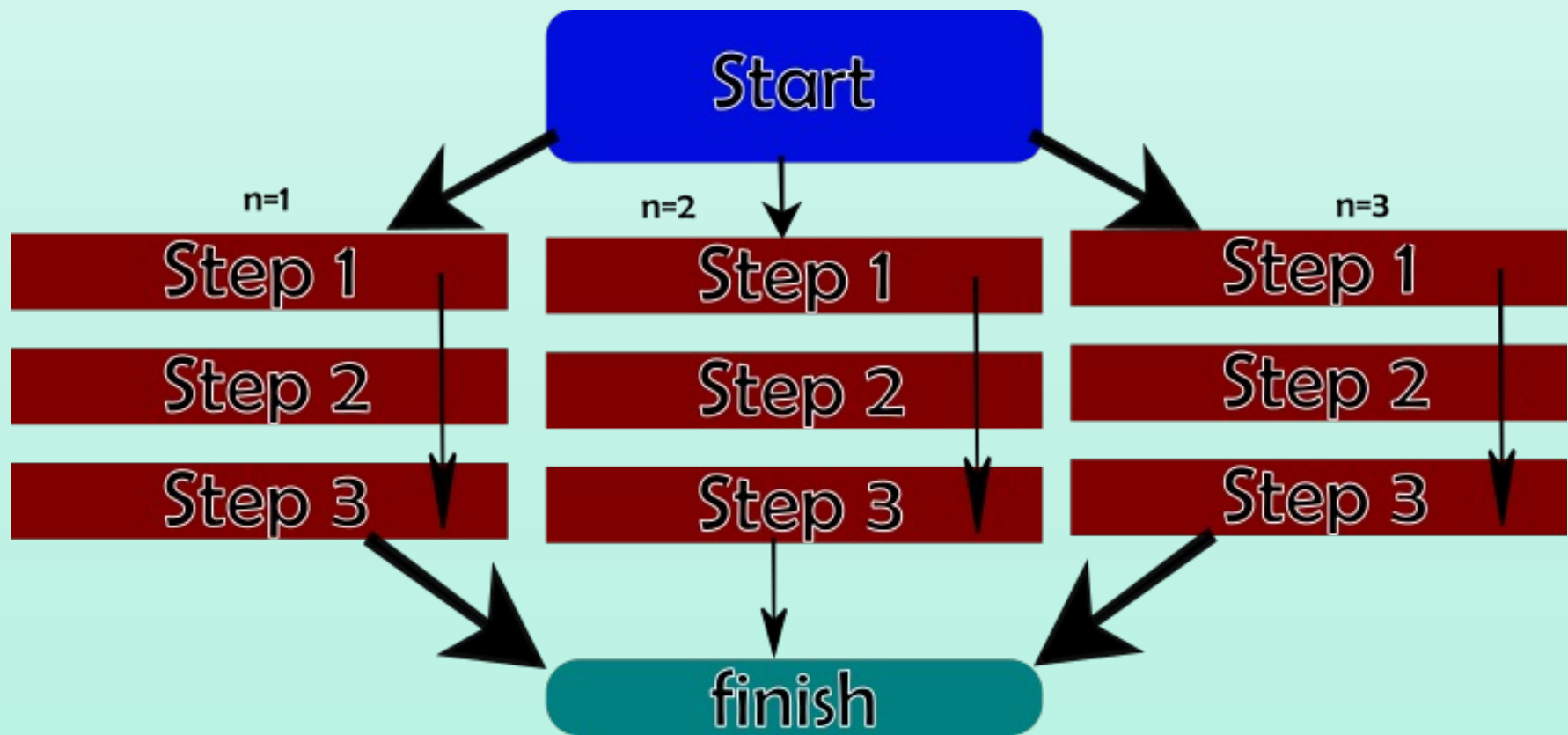
# Why use Eddie?



# Serial processing



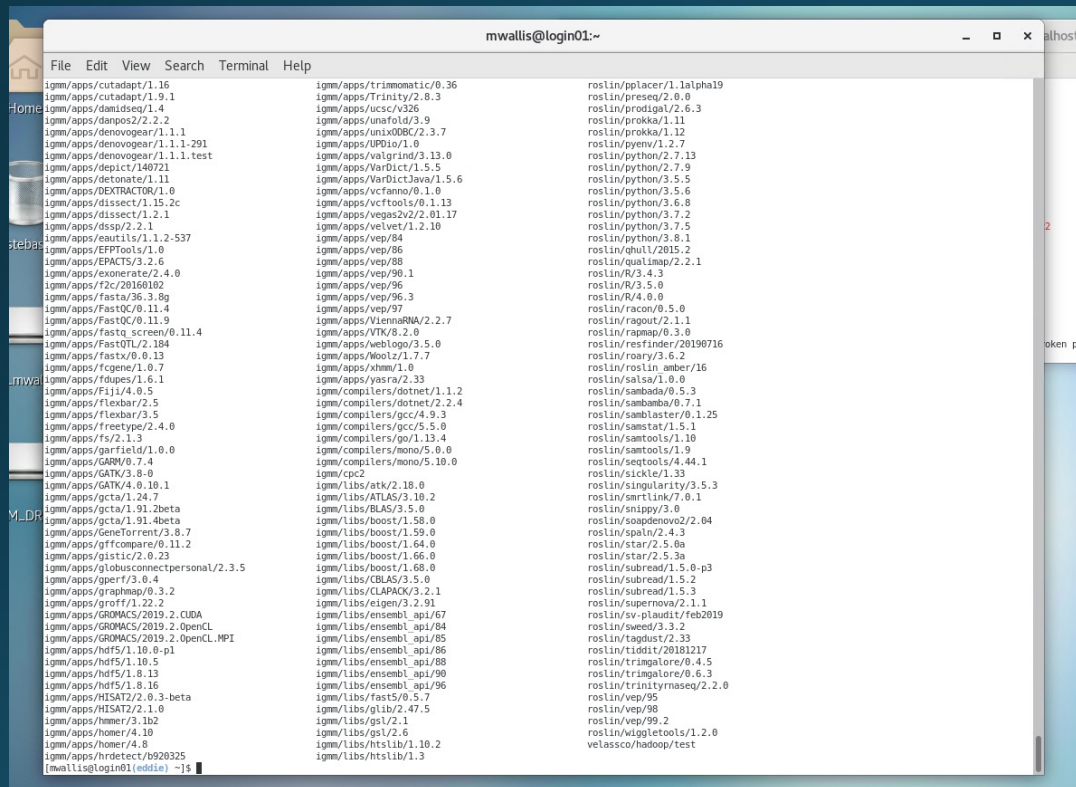
# Parallel processing



# Linux

- Understanding Linux is a *highly recommended* prerequisite for using Eddie
- Software Carpentry provides an online Linux course: <http://swcarpentry.github.io/shell-novice/>  
(at a minimum, be familiar with lessons 1,2,3 & 6)
- Research Services will do in-person Eddie training in the future, and are working on digital training on Learn

# Pre-installed applications



```
mwallis@login01:~$ module avail
igmm/apps/cutadapt/1.16
igmm/apps/cutadapt/1.9.1
igmm/apps/damseq/1.4
igmm/apps/damseq2/2.2.2
igmm/apps/denovoassembler/1.11.1
igmm/apps/denovoassembler/1.1.1-291
igmm/apps/denovoassembler/1.1.1.1.test
igmm/apps/depict/140721
igmm/apps/dextrator/1.0
igmm/apps/dissect/1.15.2c
igmm/apps/dissect/1.2.1
igmm/apps/dssp/2.2.1
igmm/apps/eautils/1.1.1.2-537
igmm/apps/EFPTools/1.0
igmm/apps/EFACTS/2.2.6
igmm/apps/exonerate/2.4.0
igmm/apps/f2c/20160162
igmm/apps/fastq/36.3.8g
igmm/apps/FastQC/0.11.4
igmm/apps/FastQC/0.11.9
igmm/apps/fastq_screen/0.11.4
igmm/apps/FastOTL/2.184
igmm/apps/fastx/0.13
igmm/apps/fcgene/1.0.7
igmm/apps/fdupes/1.6.1
igmm/apps/Fiji/4.0.5
igmm/apps/Flexbar/2.5
igmm/apps/flexbar/3.5
igmm/apps/freetype/2.4.0
igmm/apps/fs/2.1.3
igmm/apps/garf/1.0.0
igmm/apps/GARV/0.7.4
igmm/apps/GATK/3.8-0
igmm/apps/GATK/4.0.10.1
igmm/apps/gcta/1.24.7
igmm/apps/gcta/1.91.2beta
igmm/apps/gcta/1.91.4beta
igmm/apps/GeneTorrent/3.8.7
igmm/apps/gfcompare/0.11.2
igmm/apps/gistic/2.0.23
igmm/apps/globusconnectpersonal/2.3.5
igmm/apps/gperff/3.0.4
igmm/apps/grainmap/0.3.2
igmm/apps/groff/1.22.2
igmm/apps/GROMACS/2019.2.CUDA
igmm/apps/GROMACS/2019.2.OpenCL
igmm/apps/hdifs/1.10.0-p1
igmm/apps/hdifs/1.10.5
igmm/apps/hdifs/1.8.13
igmm/apps/hdifs/1.8.16
igmm/apps/HISAT2/2.0.3-beta
igmm/apps/HISAT2/2.1.0
igmm/apps/homer/3.162
igmm/apps/homer/4.10
igmm/apps/homer/4.8
igmm/apps/hrdetect/b920325
igmm/apps/trimomatic/0.36
igmm/apps/Trinity/2.8.3
igmm/apps/ucsc/v326
igmm/apps/unaFold/3.9
igmm/apps/uniXORBC/2.3.7
igmm/apps/UPDio/1.0
igmm/apps/valgrind/3.13.0
igmm/apps/VarDict/1.5.5
igmm/apps/VarDictLava/1.5.6
igmm/apps/vcfanno/0.1.0
igmm/apps/vcftools/0.1.13
igmm/apps/vegas2v2/2.01.17
igmm/apps/velvet/1.2.10
igmm/apps/vep/84
igmm/apps/vep/86
igmm/apps/vep/88
igmm/apps/vep/90.1
igmm/apps/vep/96
igmm/apps/vep/96.3
igmm/apps/vep/97
igmm/apps/VicunaRNA/2.2.7
igmm/apps/VTK/8.2.0
igmm/apps/weblogo/3.5.0
igmm/apps/Woolz/1.7.7
igmm/apps/xhm/1.0
igmm/apps/yasra/2.33
igmm/compiler/dotnet/1.1.2
igmm/compiler/dotnet/2.2.4
igmm/compiler/gcc/4.9.3
igmm/compiler/gcc/5.5.0
igmm/compiler/gp/1.13.4
igmm/compiler/mono/5.0.0
igmm/compiler/mono/5.10.0
igmm/cpc2
igmm/Libs/atk/2.18.0
igmm/Libs/ATLAS/3.10.2
igmm/Libs/BLAS/3.5.0
igmm/Libs/boost/1.58.0
igmm/Libs/boost/1.59.0
igmm/Libs/boost/1.64.0
igmm/Libs/boost/1.66.0
igmm/Libs/boost/1.68.0
igmm/Libs/CBLAS/3.5.0
igmm/Libs/CLAPACK/3.2.1
igmm/Libs/eigen/3.2.91
igmm/Libs/ensembl_api/67
igmm/Libs/ensembl_api/84
igmm/Libs/ensembl_api/85
igmm/Libs/ensembl_api/86
igmm/Libs/ensembl_api/88
igmm/Libs/ensembl_api/90
igmm/Libs/ensembl_api/96
igmm/Libs/fast5/0.5.7
igmm/Libs/glib/2.47.5
igmm/Libs/gst/2.1
igmm/Libs/gst/2.6
igmm/Libs/htslib/1.10.2
igmm/Libs/htslib/1.3
roslin/pplacer/1.1alpha19
roslin/preseq/2.0.0
roslin/prodigal/2.6.3
roslin/prokka/1.11
roslin/prokka/1.12
roslin/pyenv/1.2.7
roslin/python/2.7.13
roslin/python/2.7.9
roslin/python/3.5.5
roslin/python/3.5.6
roslin/python/3.6.8
roslin/python/3.7.2
roslin/python/3.7.5
roslin/python/3.8.1
roslin/qualimap/2.2.1
roslin/R/3.4.3
roslin/R/3.5.0
roslin/R/4.0.0
roslin/racon/0.5.0
roslin/ragout/2.1.1
roslin/rapmap/0.3.0
roslin/resfinder/20190716
roslin/roary/3.6.2
roslin/roslin_amber/16
roslin/salsa/1.0.0
roslin/sambada/0.5.3
roslin/sambamba/0.7.1
roslin/sambaster/0.1.25
roslin/samstat/1.5.1
roslin/santools/1.10
roslin/santools/1.9
roslin/seqtools/4.44.1
roslin/sickle/1.33
roslin/singularity/3.5.3
roslin/smtlink/7.0.1
roslin/snappy/3.0
roslin/soappdenovo2/2.04
roslin/spaln/2.4.3
roslin/star/2.5.0a
roslin/star/2.5.3a
roslin/subread/1.5.0-p3
roslin/subread/1.5.2
roslin/subread/1.5.3
roslin/supernova/2.1.1
roslin/sv-plaudit/feb2019
roslin/sveed/3.3.2
roslin/tappus/2.33
roslin/tiddit/20181217
roslin/trimgalore/0.4.5
roslin/trimgalore/0.6.3
roslin/trinityrnaseq/2.2.0
roslin/vep/95
roslin/vep/98
roslin/vep/99.2
roslin/wgdltools/1.2.0
velascco/hadoop/test
```

- module avail will show you everything
- Some applications are installed by RSS
- Some are installed by local communities
- Very good chance that what you want is already there
- You can also install your own applications into user space
- First point of call if you think your application should be a module is your local IT service

# Batch processing

- Asynchronous workload
- Ability to submit thousands of jobs at the same time
- Great for montecarlo simulations
- Simple parallelisation
- Lots of RAM



# Interactive sessions

- Prototyping
- Jupyter
- Rstudio
- Watching time-step processes in real time
- `qlogin -l h_rt=04:00:00,h_vmem=8G`

RStudio@node2f24.ecdf.ed.ac.uk

File Edit Code View Plots Session Build Debug Profile Tools Help

```
3 library(ggplot2)
4 library(ggribbles)
5
6 setwd = ("/exports/eddie3_homes_local/mwallis/")
7
8 pulsar <- read.csv(file = "rscripts/pulsar.csv", header=FALSE) %>%
9   mutate(row = row number()) %>%
10  gather(col, height, -row) %>%
11  mutate(
12    col = sub("^V", "", col) %>% as.integer()
13  )
14
15 ggplot(pulsar, aes(x = col, y = row, height = height, group = row)) +
16   geom_ridgeline(min_height = min(pulsar$height),
17                scale= 0.2,
18                size = 1,
19                fill = "black",
20                colour = "white") +
21   scale_y_reverse() +
22   theme_void() +
23   theme(
24     panel.background = element_rect(fill = "black"),
25     plot.background = element_rect(fill = "black", color = "black"),
26   )
27
28
```

Environment History Connections Tutorial

R Global Environment

Data

pulsar	24000 obs. of 3 variables
--------	---------------------------

Values

setwd	"/exports/eddie3_homes_local/mwallis/"
-------	----------------------------------------

Files Plots Pack

Zoom Export Publish

Console Terminal Jobs

```
R 3.6.3 · /exports/eddie3_homes_local/mwallis/
+ pulsar <- read.csv(file = "rscripts/pulsar.csv", header=FALSE) %>%
+   mutate(row = row number()) %>%
+   gather(col, height, -row) %>%
+   mutate(
+     col = sub("^V", "", col) %>% as.integer()
+   )
+ > ggplot(pulsar, aes(x = col, y = row, height = height, group = row)) +
+   geom_ridgeline(min_height = min(pulsar$height),
+                 scale= 0.2,
+                 size = 1,
+                 fill = "black",
+                 colour = "white") +
+   scale_y_reverse() +
+   theme_void() +
+   theme(
+     panel.background = element_rect(fill = "black"),
+     plot.background = element_rect(fill = "black", color = "black"),
+   )
+ >
```

R RStudio@node2f24.ecdf.ed.ac.uk

File Edit Code View Plots Session B

# Eleanor

- Sandbox virtual environment
- Low cost
- Simple interface
- Great for experimenting
- Quickstart: <https://edin.ac/39Hylfq>

horizon.ecdf.ed.ac.uk/dashboard/project/instances/84df1a4a-f77b-4c70-9b4b-1a2279a9fd92/

openstack. ed • ft\_mwallis

### Instance Console

If console is not responding to keyboard input: click the grey status bar below. [Click here to show only console](#)  
To exit the fullscreen mode, click the browser's back button.

Connected (encrypted) to: QEMU (instance-00010ed9) Send CtrlAltDel

```
[GCC 4.8.5 20150623 (Red Hat 4.8.5-39)] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> import numpy as np
>>> from astroplan import download_IERS_A
Downloading https://hpiers.obspm.fr/iers/bul/bulc/Leap_Second.dat
|-----| 1.3k/1.3k (100.00%) 8s
>>> from astroplan import Observer
>>> from astropy.time import Time
>>> import datetime
File "<stdin>", line 1
import datetime
^
SyntaxError: invalid syntax
>>> import datetime
>>> today=datetime.date.today()
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
NameError: name 'today' is not defined
>>> today=datetime.date.today()
>>> loc = Observer(lat=55.95*u.deg, long=3.19*u.deg)
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
NameError: name 'u' is not defined
>>> import astropy.units as u
>>> loc = Observer(lat=55.95*u.deg, long=3.19*u.deg)
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
File "/usr/local/lib64/python3.6/site-packages/astropy/units/decorators.py", line 168, in wrapper
    bound_args = wrapped_signature.bind(*func_args, **func_kwargs)
File "/usr/lib64/python3.6/inspect.py", line 2997, in bind
    return args[0]._bind(args[1:], kwargs)
File "/usr/lib64/python3.6/inspect.py", line 2988, in _bind
    arg=next(iter(kwargs)))
TypeError: got an unexpected keyword argument 'lat'
>>> loc = Observer(latitude=55.95*u.deg, longitude=3.19*u.deg)
>>> loc
<Observer: location (lon, lat, el)=(3.19 deg, 55.95 deg, -1.702236185308295e-09 m),
    timezone=<UTC>>
>>> sun_rse = (log.sun_rise_time(t, which="nearest")
... )
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
NameError: name 'log' is not defined
>>> sun_rse = (loc.sun_rise_time(t, which="nearest"))
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
NameError: name 't' is not defined
>>> t = Time('{} 12:00:00'.format
```

horizon.ecdf.ed.ac.uk/dashboard/project/images

openstack. ed • mwallis

### Images

Click here for filters or full text search Create Image Delete Images

Key Pairs	Name	Type	Status	Visibility	Protected	Disk Format	Size
Server Groups	Centos 6 - 1612	Image	Active	Public	Yes	QCOW2	734.50 MB
Rating	Centos 7 - 1612	Image	Active	Public	Yes	QCOW2	1.25 GB
Volumes	Centos 7 1004	Image	Active	Public	Yes	QCOW2	891.75 MB
Network	elms	Image	Active	Public	No	QCOW2	12.65 MB
Orchestration	EPCC Container Factory	Snapshot	Active	Public	No	RAW	20.00 GB
Data Processing	Scientific Linux 6.8	Image	Active	Public	Yes	RAW	8.00 GB
Object Store	Scientific Linux 7.2	Image	Active	Public	Yes	QCOW2	704.34 MB
Identity	test-ovos	Image	Active	Public	No	QCOW2	12.65 MB
	Ubuntu 14.04	Image	Active	Public	No	QCOW2	248.19 MB
	Ubuntu 14.04 Spark 1.3.1	Image	Active	Public	No	QCOW2	913.03 MB
	Ubuntu 14.04 Vanilla 2.7.1	Image	Active	Public	No	QCOW2	1.10 GB
	Ubuntu 16.04	Image	Active	Public	Yes	QCOW2	296.25 MB
	Ubuntu 18.04	Image	Active	Public	Yes	QCOW2	320.56 MB
	Windows 10 1709	Image	Active	Public	Yes	QCOW2	14.54 GB

www.ecdf.ed.ac.uk/display/Research/Services/Cloud+Quickstart

Confluence

### Steps

- Login to the Web Dashboard
- Create an SSH Key Pair
- Launch an Instance
- Associate a Floating IP address
- Connect to your instance with SSH

### Login to the Web Dashboard

Go to <https://horizon.ecdf.ed.ac.uk> and login with your Active Directory username and password. (This is usually the same as your EASE username and password. See Authentication and Authorisation for more details.)  
The Domain field should be set to 'ed'.  
If you do not have an account, please see Getting Access for information on getting one.

### Create an SSH Key Pair

Click here to expand...

### Launch an Instance

In the left-hand navigation panel, open the Project section, then open the Compute section and click on Instances.  
Choose the Launch instance button towards the top right.  
In the Launch instance dialog box that appears, specify the values in each of the tabs, navigating between them with the Next and Back buttons:

### Details tab

Click here to expand...

This tab allows you to specify the name and number of instances you want to launch.

#### Instance Name

Assign a name to the instance.

The name you assign here becomes the initial host name of the instance. If the name is longer than 63 characters, it will be truncated.  
After the instance is built, if you change the instance name in the API or change the host name directly the name is not updated in the dashboard.  
Instance names are not guaranteed to be unique when created, so you could have two instances with the same host name.

#### Availability Zone

# Why use Eleanor?



## Sandbox

- Just testing a theory before scaling up



## Application server

- Don't want to run a webserver on your laptop?



## It's your machine

- And it's a VM! Break something? Delete it & start over!

And so much more!

# Available resources

- GPUs\* on both Eddie and Eleanor
- Eddie: 000's of cores, TB of RAM, Pb of storage
  - Scalable, inexpensive (or free) for compute
- Eleanor: 16 cores, 96Gb of RAM, Tb of storage
  - Sandbox, expandable, (slightly) chargeable
- DataStore, DataSync, DataVault, Secure Storage
  - Shared resilient storage; external collab; archive; special circs
  - Accessible anywhere on campus
- Version control service
  - SVN and GitLab

# Plan your research compute needs

- Have ideas!
- Talk to us about your ideas!
- We can offer advice, support and training
- ...we cannot run your services for you (so think about this!)
- Use the free tier to test as much as you like
- Paid for when you need serious resources
- Remember to have a data management plan
- Think about your application as a publishable resource

# Summary

- You create data (so have a plan!)
- Your workload will get bigger than your laptop
- The University has services you can use
- ...and people to help you use them



# Help and support

- <https://www.wiki.ed.ac.uk/display/ResearchServices>
- [is.helpline@ed.ac.uk](mailto:is.helpline@ed.ac.uk)