# Topics

- Introduction to the ilifu research facility

- Directory structure

- Software environment

  - Singularity containers

  - Modules

- JupyterHub

- Introduction to Slurm

# Getting help

- Support contact

  support@ilifu.ac.za

- User documentation

  http://docs.ilifu.ac.za

- Ilifu System Status

  https://status.ilifu.ac.za

- Training videos

  https://www.ilifu.ac.za/latest-training

- Tutorials

  https://github.com/ilifu/ilifu_user_training
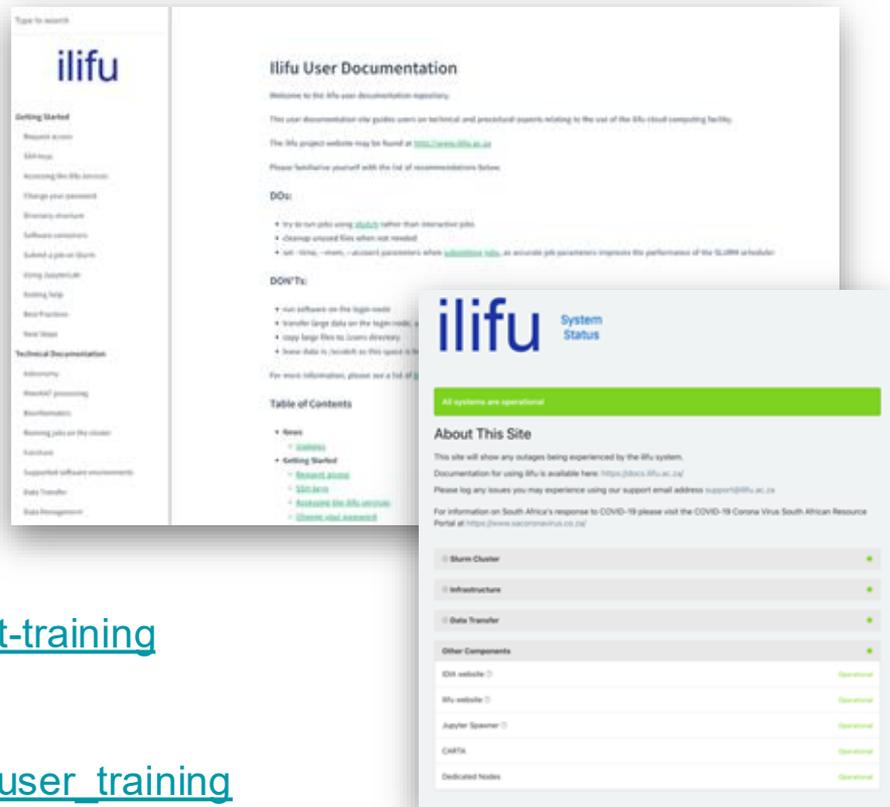
# High Performance Computing

Combining power of compute clusters

- collection of servers (computers)
- connected by fast local network
- to solve complex problems

Some terminology

- computer system/server also referred to as a node
- group of nodes is a cluster

# ilifu Research Facility

Cloud-based infrastructure for data-intensive research

Network of remote servers, accessed over the internet, to store, manage, and process data

- Open source cloud software to deploy infrastructure as a service (IaaS)
- Support variety of different scientific projects and requirements
- Flexible compute environment
  - Cluster environment with workload management, additional services
- Data management: storage, transfer

# ilifu Research Facility - interfaces

Command line interface

ssh - shell terminal



**ssh <username>@slurm.ilifu.ac.za**

**ssh <username>@transfer.ilifu.ac.za**

Web applications

IDIA Science Gateway - App Dashboard



https://gateway.idia.ac.za

# ilifu Research Facility - Web Applications

Application dashboard - accessed via https://gateway.idia.ac.za
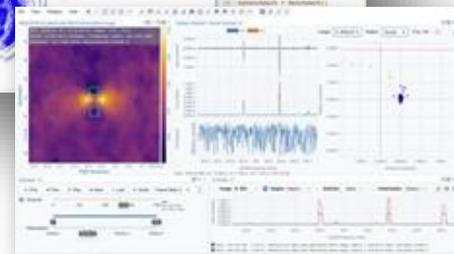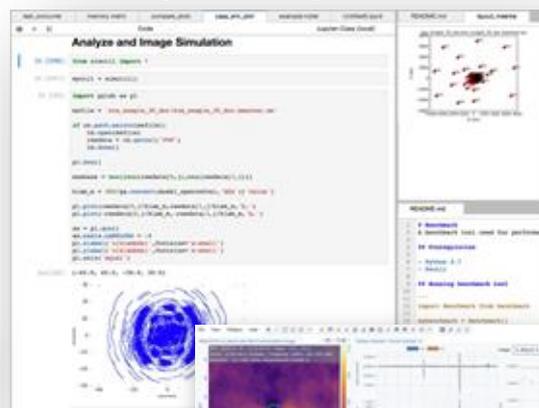


Web application examples
https://jupyter.ilifu.ac.za
https://carta.idia.ac.za

# Compute environment - ssh

Your SSH key

- Used in the SSH (Secure Shell) protocol

- Authentication method for gaining access to encrypted connecting
  between systems

- Use connection to manage system remotely

- We need your SSH public key so our system knows to let you in

- Default:
  - ~/.ssh/id_rsa                              ~/.ssh/id_rsa.pub
  - ~/.ssh/id_ed25519        ~/.ssh/id_ed25519.pub

# Compute environment - ssh

Generating SSH key

- If you don't already have one

- New computer/formatted existing computer

GitHub docs on key generation:

https://docs.github.com/en/github/authenticating-to-github/

connecting-to-github-with-ssh/generating-a-new-ssh-key-and-

adding-it-to-the-ssh-agent

Manage your SSH keys associated with your ilifu account

https://usage.ilifu.ac.za/ssh_keys

# Compute environment - ssh

```
user-local:~$ ssh <username>@slurm.ilifu.ac.za

...

<username>@slurm-login:~$

<username>@slurm-login:~$ pwd

/users/<username>

<username>@slurm-login:~$ ls

README.md   workspace

<username>@slurm-login:~$ ls /idia/projects/

G4Jy           gamma-ray-binaries   meerlicht         shapley-uhf

M64-NGC151     goodsn               meerlirgs         share

adfs           grandspiral          meerrings         simba

antlia         hack4dev             merghers          supermightee

...
```

# Directory Structure

Common areas:

- /users
  - limited storage shared among all users, for scripts and small files – don't place data here, capping /users storage capacity can prevent access to the cluster for all users. **Limited to 200GB**

- /scratch3/users
  - directory space for processing data, temporary storage only, i.e. use this space during processing, and then clear all files immediately after processing. Remove unnecessary data and move data that you want to keep to project folder.
  - **90-day scratch storage auto-deletion policy**

Remaining storage separated by group: IDIA, CBio, ilifu

# Directory Structure

IDIA structure:

- /idia/users
  - user's private work directory, may store data products that are not ready to move to shared project space. **Limited to 10TB**

- /idia/projects
  - project specific directories. These directories are for sharing data and resources within project groups.  Raw data associated with a project will also be available from the project folder. Raw data should always be  read-only.

- /idia/software
  - software containers and the IDIA Pipelines software is stored here

# Directory Structure

Similar structure for /cbio and /ilifu groups:

- /cbio/users **(Limited to 10TB)**

- /cbio/projects

- /cbio/soft

- /ilifu/users **(Limited to 2TB)**

- /ilifu/software

- Exception for ilifu projects:
    - /ilifu/astro/projects
    - /ilifu/bio/projects

# Directory Structure - Typical workflow

/users — Scripts and small files only

/n/projects/
/n/users/
/n/data/*

Data

Results

read

selective write

/scratch3

compute/process

Results + intermediate files

remove intermediate files
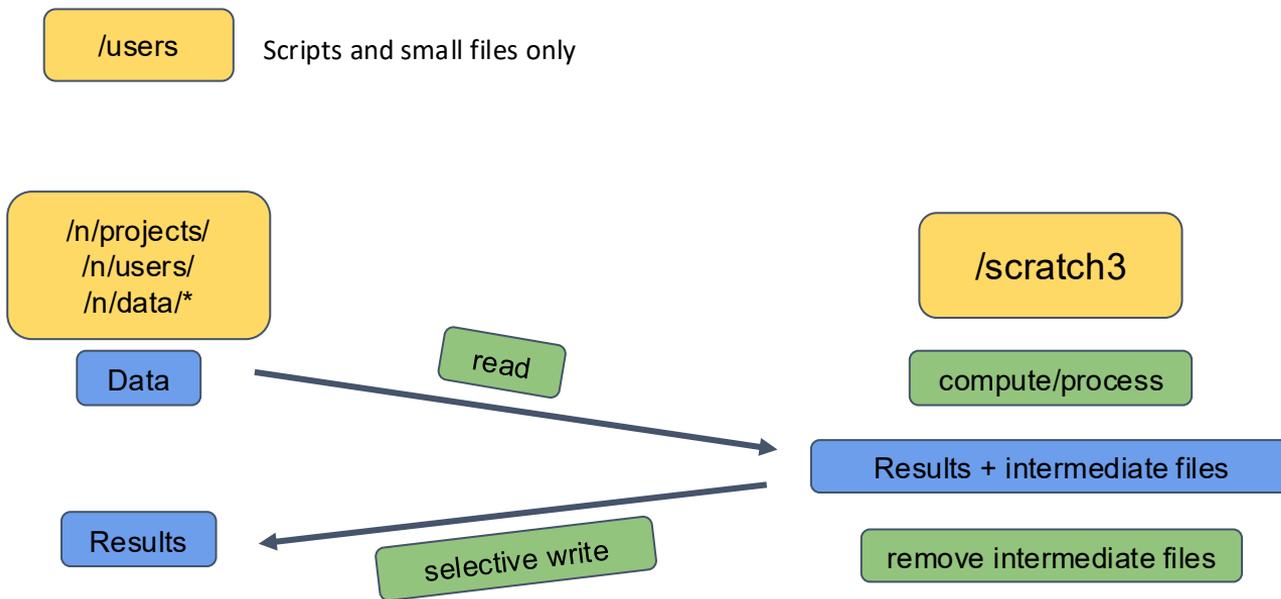
*/n/data generally read-only

n = idia,cbio,ilifu

# Ilifu Software Environment



- Software containers
  - encapsulated software environments
  - suite of applications and libraries
  - shareable, transferable to different hardware/environments
  - reproducible science

- Modules
  - common languages & utilities (CUDA, MPI)
  - ease of use
  - wrapper for containers

- Virtual environments
  - Python, conda
  - user created & managed
  - useful for prototyping & development

# Software environment - Singularity containers
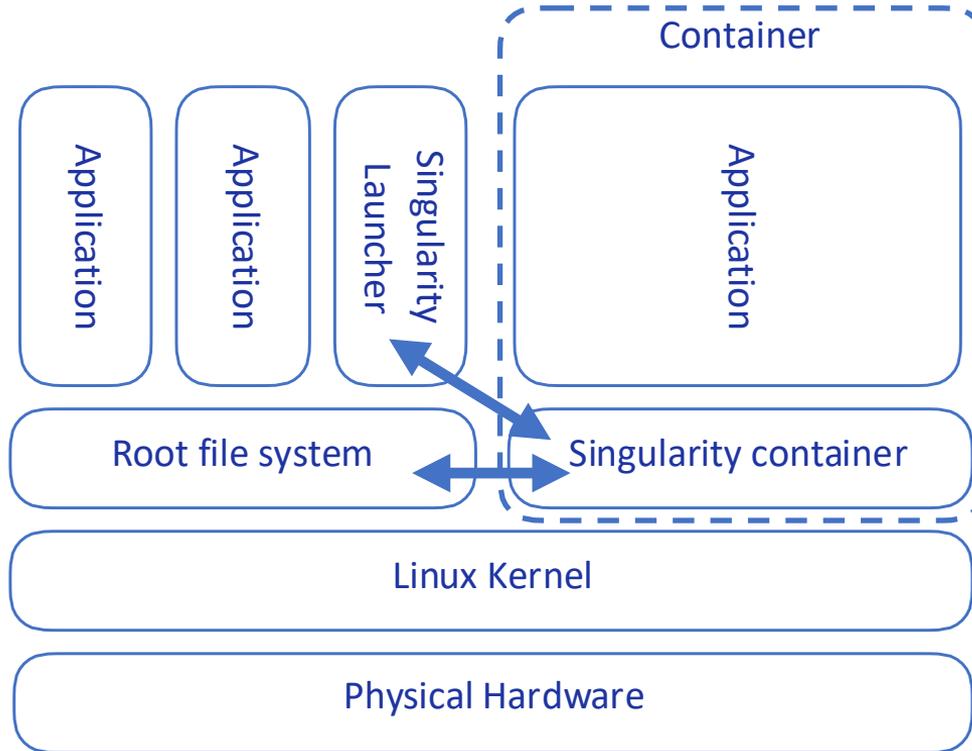
# Software environment - Singularity containers

Supported Containers:
- CASA 5, CASA 6
- Astronomy container (ASTRO-PY3, ASTRO-PY3.8, ASTRO-PY3.10)
- KERN suite
- GPU Python container
- Project containers:
  - MeerLICHT, LADUMA, HI Intensity mapping
- lots of others

Directories:
- /software
- /idia/software/containers
- /ilifu/software/containers

# Software environment - Singularity containers

Open container as an interactive shell:

```
singularity shell /path/to/container
Example:
$ singularity shell /idia/software/containers/ASTRO-PY3.10.si
```

Run a script/workflow using a container environment:

```
singularity exec /path/to/container <software> <script/input_parameters>
$ singularity exec /idia/software/containers/casa-6.simg python
myscript.py
```

Look at what is inside a container by viewing its build script:

```
singularity inspect -d /path/to/container
```

# Software environment - modules

## module avail

```
$ module avail

------------------------------------------------------------------- /software/modules/astro -------------------------------------------------------------------
   atnfcat/2.0.1          casa/5.7.2-4              casa/6.2       casa/6.5.0      cfitsio/4.3.1  (D)   pybdsf/1.9.2          splinter/a968918
   bbarolo/1.6.1          casa/5.8.0               casa/6.3       casa/6.5.5      pgplot/5.2           pybdsf/1.10.1  (D)   tempo/0b487e2
   bbarolo/1.6.7 (D)      casa/6.1.0-118-monolithic  casa/6.4     casa/6.6.0      psrcat/1.50          qd/2.3.12            tempo2/2024.02.1
   caloeph/2.3.2          casa/6.1.2.7-pipeline    casa/6.4.3     casa/6.6.4  (D) psrdada/008afa7      sigproc/28ba4f4
   casa/5.7.0             casa/6.1.2.7-modular     casa/6.4.4     cfitsio/3.450   psrxml/1.01          sofa/20180130

-------------------------------------------------------------------- /software/modules/bio --------------------------------------------------------------------
   ClinSV/ClinSV_1.0.0    filtlong/0.2.1           merqury/1.3                       samtools/1.18
   DIANN/1.8.1            fragpipe/18.0            meryl/1.3                         samtools/1.19
   GenomeBrowser/cli      gatk/gatk_4.2.6.1        meryl/1.4.1             (D)       samtools/1.19.2
   ambertools/23          gatk/4.2.5.0             minimap2/2.24                     samtools/1.20                                           (D)
   annovar/2020-06-07     gatk/4.3.0.0             miniwdl/1.12.1                    savvysuite/git
   bamtools/2.5.2         gatk/4.4.0.0             mtoolbox/1.2.1.1                  seqkit/2.6.0
   bcbio/bcbio_container  gatk/4.5.0.0       (D)   multiqc/1.17                      seqtk/1.4
   bcbio/1.2.3            gemini/gemini            multiqc/1.22.3          (D)       seqwish/0.7.9-2
   bcbio/1.2.9       (D)  gemma/0.98.5             mummer/4.0.0rc1                   shapeit5/5.1.1
   bcftools/1.10.2        genomestrip/2.00.1958    nanocomp/1.23.1                   snpEff/5.1
   bcftools/1.17          gfastats/1.3.6           nanofilt/2.8.0                    snpdists/0.8.2
   bcftools/1.19     (D)  glimpse/2.0.0            nanoplot/1.41.0                   snpsites/2.5.1
 ...
------------------------------------------------------------------ /software/modules/common ------------------------------------------------------------------
   LAPACK/3.9.0                 cuda/11.0.2_450.51.05       gnuplot/default        neovim/0.8.2              python/3.9.7
   LAPACK/3.10.1                cuda/11.4.2_470.57.02       go/1.16.3              neovim/0.9.4       (D)    python/3.9.12
   LAPACK/3.12.0          (D)   cuda/11.6.0_510.39.01       go/1.17.3              nodejs/18.13.0           python/3.9.16
   R/RStudio1.2.5042-R4.0.0     cuda/11.7.0_515.43.04       go/1.18.4              nodejs/19.4.0      (D)   python/3.9.19
   R/RStudio1.2.5042-R4.0.4     cuda/11.8.0_520.61.05       go/1.20.4              openBLAS/0.3.9           python/3.10.0
   R/RStudio1.2.5042-R4.2.0     cuda/12.1.0_530.30.02       go/1.21.6              openBLAS/0.3.20          python/3.10.1
   R/RStudio1.2.5042-R4.2.1     cuda/12.2.0_535.54.03       go/1.22.4          (D) openBLAS/0.3.25    (D)   python/3.10.4
   R/RStudio2022.12.0-353-R4.2.2  cuda/12.4.0_550.54.14 (D) graphviz/2.49.1        openmpi/2.1.1            python/3.10.9
 ...
------------------------------------------------------------------ /opt/lmod/lmod/modulefiles/Core ------------------------------------------------------------------
   lmod    settarg

  Where:
   L:  Module is loaded
   D:  Default Module
```

# Software environment - modules

- module avail

- module help <module>

```
$ module help python


----------------------------- Module Specific Help for "python/3.10.1" --------------------------------
This module configures Python 3.10.1 for use
```

- module load <module>

- module list

- module purge

- module --help

# JupyterHub

https://gateway.idia.ac.za

OR

https://jupyter.ilifu.ac.za

# JupyterHub

## Session size



Launch Jupyter Lab

Hi jeremy. Remember to try and choose the smallest profile that fits your task. This helps us to make sure that everyone has access to the resources they need. Please visit the user documentation to learn more about Jupyter on ilifu. If you have any more questions, please send an email to ilifu support.

The following table shows the job profiles available on the ilifu cluster (as at 2024-08-30 10:41):

| Job Profile | Available Jobs |
|---|---|
| GPU Session (16 cores, 1 GPU) | 3 |
| Minimum Session (1 core, dedicated) | 17 |
| Small Session (2 cores, dedicated) | 8 |
| Medium Session (4 cores, dedicated) | 4 |
| Large Session (8 cores, dedicated) | 2 |
| Half-Max Session (16 cores, dedicated) | 0 |
| Max Session (32 cores, dedicated) | 0 |

Select a job profile:

Development Session - 2 core, 3 GB RAM, shared, 18 hrs idle timeout, max 14 days lifespan

Start

# JupyterHub

Choose kernel

in launcher
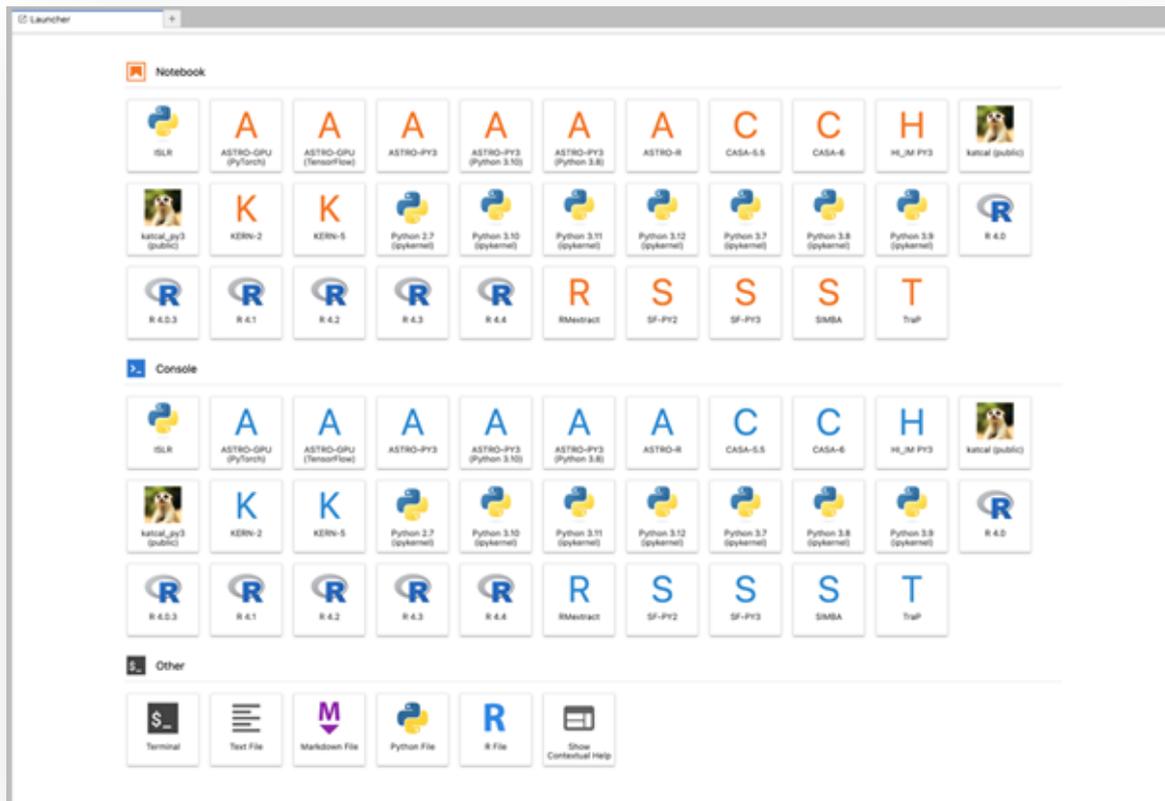
# Demo

Demo resources

https://github.com/ilifu/ilifu_user_training