



# IDIA

Inter-University Institute  
for Data Intensive Astronomy

## Ilifu Online Training

### Introduction to ilifu - 19 March 2024

Jeremy Smith, ilifu Operations Manager  
University of Cape Town



# 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](mailto:support@ilifu.ac.za)

- User documentation

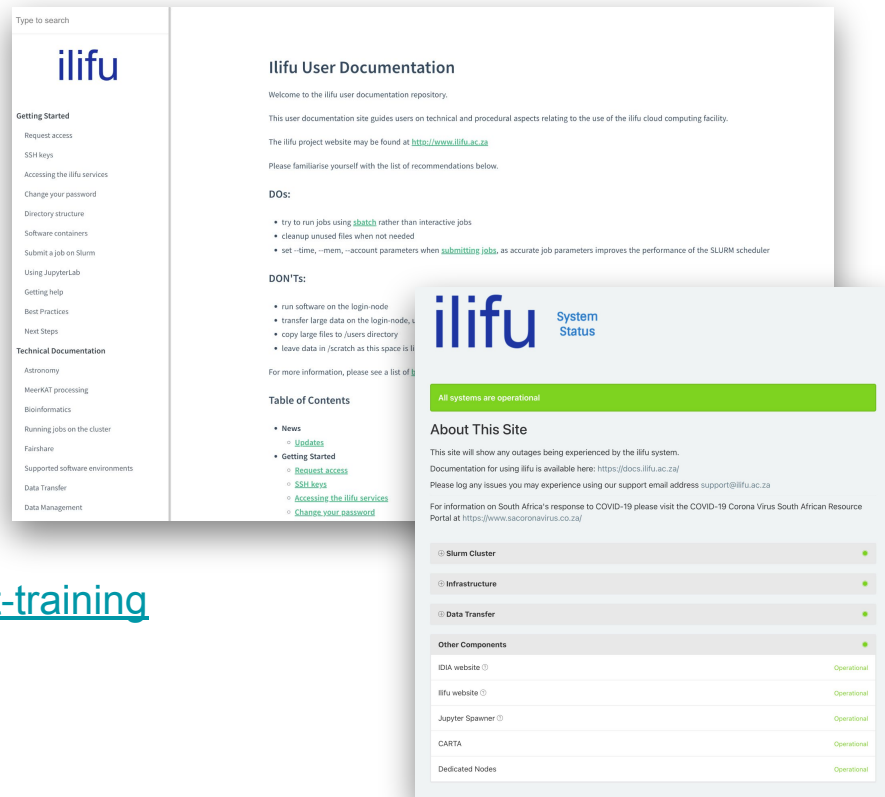
<http://docs.ilifu.ac.za>

- Ilifu System Status

<https://status.ilifu.ac.za>

- Training videos

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



The image shows three overlapping screenshots of the ilifu website. The top-left screenshot is the 'ilifu' user documentation repository page, featuring a search bar, a navigation menu with categories like 'Getting Started', 'Technical Documentation', and 'Getting help', and a main content area titled 'Ilifu User Documentation'. The top-right screenshot is the 'System Status' page, which has a green banner stating 'All systems are operational' and a table listing various components and their status. The bottom screenshot is a detailed view of the 'System Status' table.

Component	Status
Slurm Cluster	Operational
Infrastructure	Operational
Data Transfer	Operational
<b>Other Components</b>	
IDIA website	Operational
ilifu website	Operational
Jupyter Spawner	Operational
CARTA	Operational
Dedicated Nodes	Operational

# 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

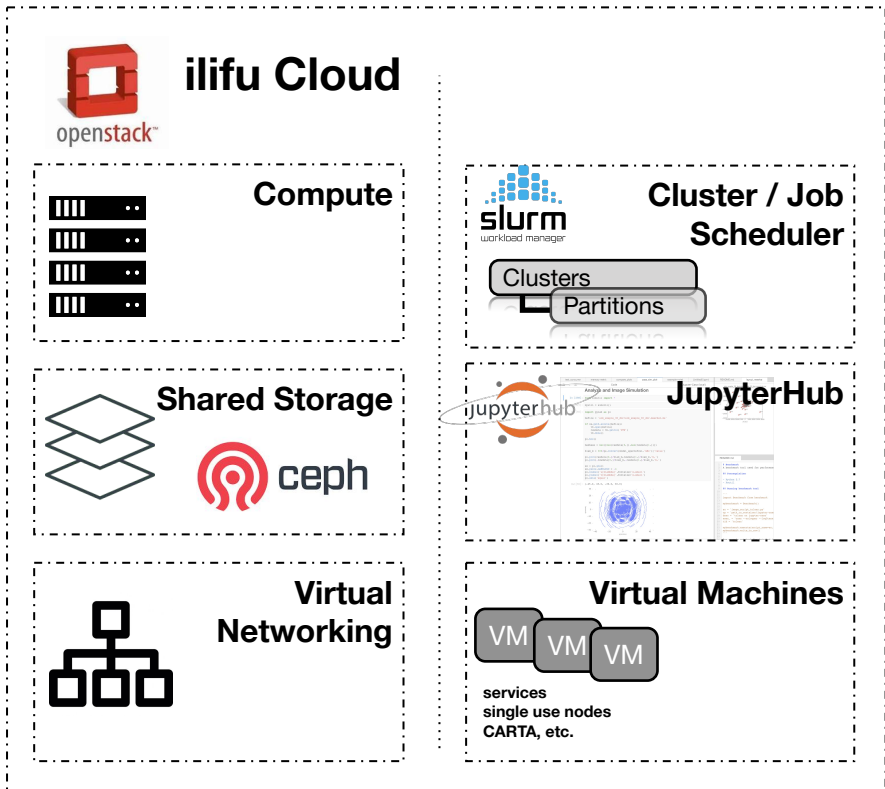
Cloud-based infrastructure for data-intensive research

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



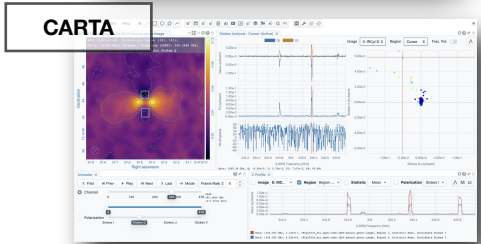
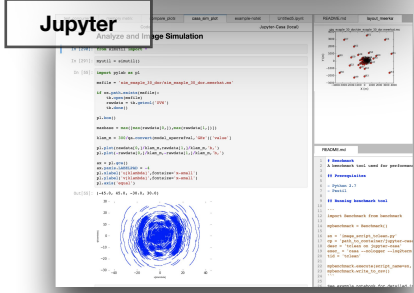
```
ssh
# If you need help please contact the support team
support@ilifu.ac.za

Please login to https://reports.ilifu.ac.za/ and make sure your
account is up to date as well as to view usage summaries.

Valid Slurm Accounts for user jeremy on ilifu-slurm021:
slurmadmin
Change your default account with:
sacctmgr modify user name=jeremy set DefaultAccount=account*
Running job count: #
Pending job count: #

Run the "help" command to display this message.

jeremy@slurm:~$ sinfo
PARTITION AVAIL TIMELIMIT NODES STATE MODELLIST
Main* up 14:00:00 1 drain compute-002
Main* up 14:00:00 11 mix compute-101-103,105-106,108-109,111,120
Main* up 14:00:00 26 alloc compute-1011,117,124,127,130,126,128,13
Main* up 14:00:00 27 idle compute-103-116,112,116,118-121,141-14
Jupyter up infinite 5 mix Jupyter-103-104,106-108
Jupyter up infinite 4 alloc Jupyter-101-102,105,109
Jupyter up infinite 1 idle Jupyter-010
Jupyter up infinite 2 alloc Jupyter-001
highmem up 14:00:00 1 mix T1ghmem-001
highmem up 14:00:00 2 alloc Jupyter-101-102
CPU up 14:00:00 4 alloc gpu-101-104
GPU up 14:00:00 2 idle gpu-105-107
GPU100 up 14:00:00 1 idle gpu-005
Revel up 3:00:00 1 alloc compute-001
jeremy@slurm:~$ sbatch test-job.sh
```





# ilifu Research Facility - interfaces

## Command line interface

ssh - shell terminal

```
For any queries or if you need help please contact the support team at support@ilifu.ac.za

Please login to https://reports.ilifu.ac.za/ and make sure your account is up to date as well as to view usage summaries.

Valid Slurm Accounts for user jeremy on ilifu-slurm2021:
b03-idia-ag
b34-admins-ag (default)
Change your default account with:
sacctmgr modify user name=jeremy set DefaultAccount=<account>
Running job count: 0
Pending job count: 0

Run the "shelp" command to display this message.

jeremy@slurm-login:~$ sinfo
PARTITION AVAIL TIMELIMIT NODES STATE NODELIST
Main* up 14-00:00:0 1 drain* compute-002
Main* up 14-00:00:0 11 mix compute-[201-203,205-206,208-209,231,24
Main* up 14-00:00:0 36 alloc compute-[011,017,204,207,210-226,228-23
Main* up 14-00:00:0 37 idle compute-[003-010,012-016,018-021,101-10
Jupyter up infinite 5 mix jupyter-[003-004,006-008]
Jupyter up infinite 4 alloc jupyter-[001-002,005,009]
Jupyter up infinite 1 idle jupyter-010
JupyterGPU up 14-00:00:0 2 alloc gpu-[003-004]
HighMem up 14-00:00:0 1 mix highmem-003
HighMem up 14-00:00:0 2 alloc highmem-[001-002]
GPU up 14-00:00:0 4 alloc gpu-[001-004]
GPU up 14-00:00:0 3 idle gpu-[005-007]
GPUV100 up 14-00:00:0 1 idle gpu-005
Devel up 5-00:00:00 1 alloc compute-001

jeremy@slurm-login:~$ sbatch test.job.sh
```

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

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

## Web applications

The screenshot displays a JupyterLab web interface. The main area shows a code editor with Python code for image simulation analysis. The code includes imports for NumPy and Matplotlib, and logic to load and plot raw data. Below the code, there are several interactive plots: a scatter plot of 'y' vs 'x' with red dots, a contour plot of a 2D function, a heatmap of a 2D function, and a line plot showing a signal over time. The interface also includes a sidebar with file explorer and a top navigation bar with various tool tabs.

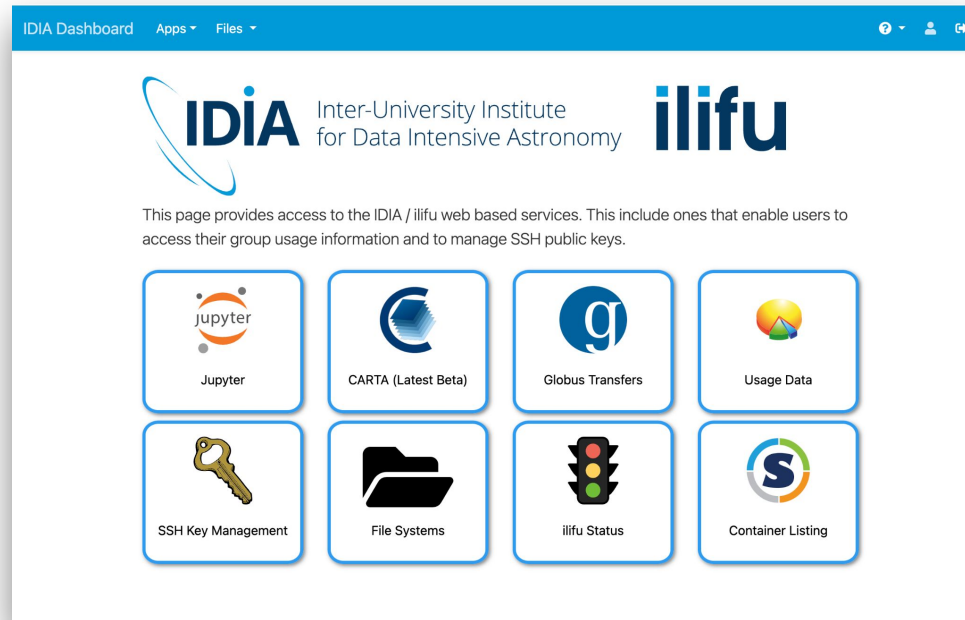
<https://jupyter.ilifu.ac.za>

<https://carta.idia.ac.za>



# ilifu Research Facility - Science Gateway

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



The screenshot shows the IDIA Dashboard application interface. At the top, there is a blue navigation bar with the text "IDIA Dashboard" and two dropdown menus labeled "Apps" and "Files". On the right side of the navigation bar, there are icons for a search, a user profile, and a refresh button. Below the navigation bar, the main content area features the IDIA logo (Inter-University Institute for Data Intensive Astronomy) and the ilifu logo. A paragraph of text states: "This page provides access to the IDIA / ilifu web based services. This include ones that enable users to access their group usage information and to manage SSH public keys." Below this text, there is a grid of eight service tiles, each with an icon and a label: Jupyter (orange circle with 'jupyter' text), CARTA (Latest Beta) (blue circle with a white 'C'), Globus Transfers (blue circle with a white 'g'), Usage Data (yellow and green triangle), SSH Key Management (yellow key icon), File Systems (black folder icon), ilifu Status (traffic light icon), and Container Listing (blue circle with a white 'S').



# Computing 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](https://usage.ilifu.ac.za/ssh_keys)

# 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.
- `/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.

Remaining storage separated by group: IDIA, CBio, ilifu

# Directory Structure

IDIA structure:

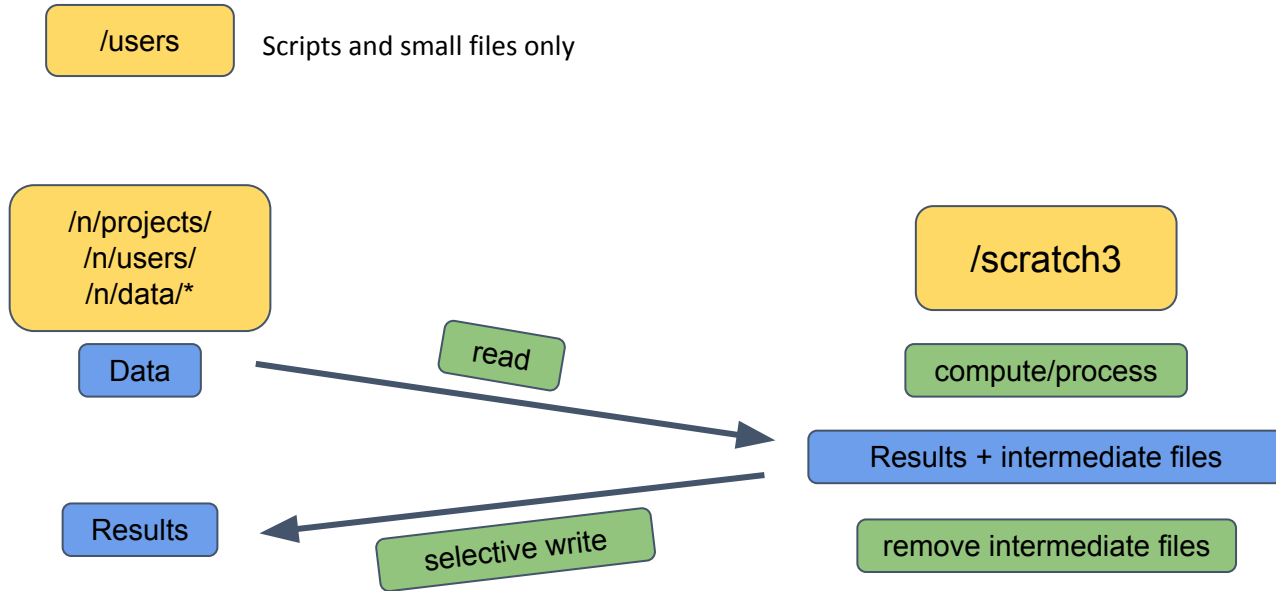
- `/idia/users`
  - user's private work directory, may store data products that are not ready to move to shared project space
- `/idia/projects`
  - project specific directories. These directories are for sharing data and resources within project groups. Raw data associated with a project will also be 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
- /cbio/projects
- /cbio/soft
- /ilifu/users
- /ilifu/software
- Exception for ilifu projects:
  - /ilifu/astro/projects
  - /ilifu/bio/projects

# Directory Structure - Typical workflow



\* /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

```

> module avail
-----
bbartolo/1.6.1  casa/5.7.0  casa/5.8.0  casa/6.1.0-118-monolithic  casa/6.1.2.7-pipeline  /software/modules/astro  casa/6.2  casa/6.4  casa/6.4.4  pybids/1.9.2
bbartolo/1.6.7 (D)  casa/5.7.2-4  casa/6.1.0-118-monolithic  casa/6.1.2.7-modular  casa/6.3  casa/6.4.3  casa/6.5.0 (D)  pybids/1.10.1 (D)

cclink/ClinkV.1.0-0  exonerater/13.0.1  iotree/1.6.12  soopen/0.1  /software/modules/dio  svanno/1.0.1
GenomeBrowser/c11  fastqc/0.11.9  iotree/2.2.0 (D)  proteomizard/proteomizard_3.0.22167  treeP/nomewr
anomer/2020-08-07  fillup/0.2.1  listcol/2.3.4  proteomizard/3.0  treeP/nomewr
scbio/bcbio_container  fseqpipe/18.0  maffit/7.098  pypsa/0.12.0  trypcycler/trypcycler_bicontainers_0.5.4--pyhd778af7
scbio/1.1.3  gatk4/gatk_4.2.0.1  maspp/0.9  r2tools/0.1.1  trypcycler/trypcycler_stdphh_0.5.4
scbio/1.1.2.9  gatk4/4.2.5.0  nedsa/1.7.3  santools/1.10  vcfTools/0.1.16
scftools/1.20.2 (D)  gatk4/4.3.0  ngsnp/0.2.0  santools/1.12  vep/100.10
scfTools/1.17 (D)  gemini/gemini  nanoplot/1.14  santools/1.14  vep/100.11
scseq/10011.0.4  gemini/gemini  nanoplot/1.12  santools/1.12  vep/100.11
biolabhard/2.0.183  gromacs/2022.2  phenyl/Saring2822  savvysite/git
bwa/gt100  htlib/1.18.2  plicat/2.26.1  smptf/1.1  vep/100.11
cavus/1.40.0.1013  htlib/1.17  plink/2.0962.3  strtools/3.0.0
cd-hlt/4.0.2  nhyh/2.5.69  polypolish/0.5.0  stackz/2.01

LAPACK/3.9.0  R/4.2.0  cudf/11.4.2.478.57.02  /software/modules/common  nodejs/18.13.0  peti/1.33.0  python/3.9.4
R/RStudio1.2.5042-04.0.0  R/4.2.0  cudf/11.0.0.510.39.01  java/7re-1.0.0.241  nodejs/19.4.0  python/3.9.7
R/RStudio1.2.5042-04.0.4  anaconda3/login.oid  cudf/11.7.0.511.43.04 (D)  java/openjdk-14.0.2  openssl/3.0.2  python/2.7.18  python/3.9.12
R/RStudio1.2.5042-04.2.0  anaconda3/login  dotnet/6.0.381  java/openjdk-17.0.2  openssl/1.1.1  python/3.6.15  python/3.9.16
R/RStudio1.2.5042-04.2.1  anaconda3/2020.07  draw.io/1.13  java/openjdk-18.0.2 (D)  openssl/1.1.1  python/2.7.7  python/3.10.0
R/RStudio2022.12.0-303-R4.2.2  anaconda3/2021.05  gitHubcli/2.0.0  julia/1.5.3  openssl/1.1.1  python/3.7.16  python/3.10.1
R/4.0.0  anaconda3/2021.11  go/1.16.3  libqjs/1.7.1.3  openssl/1.4.0  python/3.8.2  python/3.10.4
R/4.0.0  cuda/10.0.130_A10-A8  go/1.17.3  maven/3.6.3  openssl/1.4.0.5  python/3.8.3  python/3.10.9
R/4.0.2  cuda/10.1.243_A10-07.00  go/1.18.4  mongo/6.0.0.123  openssl/1.1.0  python/3.8.0  python/3.11.0
R/4.0.3  cuda/10.2.89_A10-20.05  gromacs/22.49.1  mongo/6.0.2  openssl/1.1.1 (D)  python/3.8.10  python/3.11.1
R/4.1.1  cuda/11.0.2.450-11.05  horebrev/2.4.13  neo4j/0.8.2  openssl/1.1.1  python/3.9.0  python/3.11.2

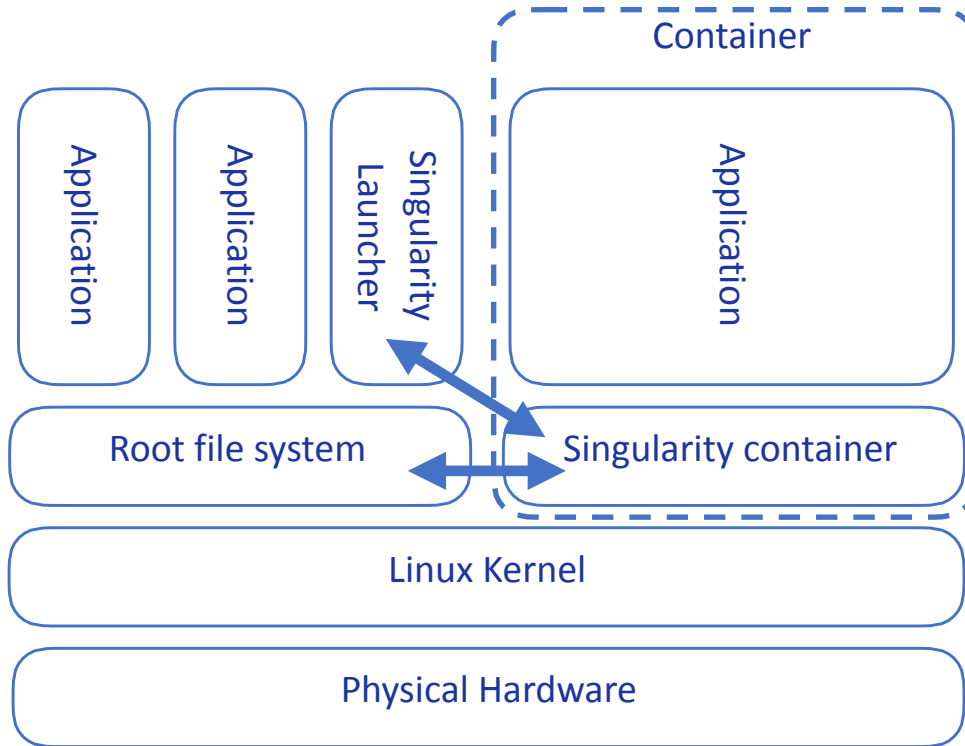
Core/1mod/6.0  Core/setting/6.0  /usr/share/1mod/1mod/modulefiles

Where:
L: Module is loaded
D: Default Module

Use "module spider" to find all possible modules.
Use "module keyword key1 key2 ..." to search for all possible modules matching any of the "keys".

```

# Software environment - Singularity containers





# Software environment - Singularity containers

## Supported Containers:

- CASA 5, CASA 6
- Astronomy container (ASTRO-PY3, ASTRO-PY3.8)
- 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-PY3.8.simg
```



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
```



# Software environment - modules

## module avail

\$ module avail

```

----- /software/modules/common -----
LAPACK/3.9.0          anaconda3/2020.07      githubcli/2.0.0        mono/6.8.0.123        perlbrew/perlbrew     python/3.10.0
R/RStudio1.2.5042-R4.0.0  anaconda3/2021.05      go/1.16.3              mpich/3.3a2           python/2.7.18         python/3.10.1 (D)
R/RStudio1.2.5042-R4.0.4  anaconda3/2021.11      go/1.17.3              openBLAS/0.3.9       python/3.6.15        ruby/2.6.6
R/3.6.3              cuda/10.0.130_410.48   graphviz/2.49.1        openmpi/2.1.1        python/3.7.7         singularity/2.6.1
R/4.0.0              cuda/10.1.243_418.87.00  homebrew/2.4.13       openmpi/2.1.6        python/3.8.2         singularity/3.7.3
R/4.0.2              cuda/10.2.89_440.33.01  hwloc/1.11.13         openmpi/3.1.6        python/3.8.3         singularity/3.8.3
R/4.0.3              cuda/11.0.2_450.51.05   java/jre-1.8.0_261     openmpi/4.0.3        python/3.8.6         singularity/3.9.0
R/4.1.1              cuda/11.4.2_470.57.02   java/openjdk-14.0.1 (D)  openmpi/4.0.5        python/3.9.0         singularity/3.9.1 (L,D)
anaconda3/login.old    dotnet/5.0.301         julia/1.5.3            openmpi/4.1.0 (D)    python/3.9.4         user_tools
anaconda3/login        drmaa/1.1.1            maven/3.6.3            perl/5.33.0           python/3.9.7

----- /software/modules/astro -----
casa/5.7.0          casa/5.8.0          casa/6.1.2.7-pipeline  casa/6.2          casa/6.4 (D)
casa/5.7.2-4       casa/6.1.0-118-monolithic  casa/6.1.2.7-modular  casa/6.3          pybdsf/1.9.2

----- /software/modules/bio -----
bcbio/bcbio_container  biobambam2/2.0.183  genomestrip/2.00.1958  plink/2.00a2.3  samtools/1.13  vep/singularity
bcbio/1.2.3            canvas/1.40.0.1613  htstlib/1.10.2          popgen/0.1      samtools/1.14 (D)  vep/101.0 (D)
bcbio/1.2.9            cd-hit/4.8.2        mafft/7.490             prsice-2/2.3.1d  treePL/homebrew
bcftools/1.10.2        gemini/gemini        mash/2.3                 samtools/1.10   vcftools/0.1.16

----- /usr/share/lmod/lmod/modulefiles -----
Core/lmod/6.6      Core/settarg/6.6

```

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://jupyter.ilifu.ac.za>

ilifu

Sign in to your ilifu account

Username

Password

Sign In

Or sign in with / link account to

eduGAIN / SAFIRE

New user? [Register](#)



# 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 of 2024-03-18 15:42):

Job Profile	Available Jobs
Minimum (1 core)	58
Small (2 cores)	29
Medium (4 cores)	14
Large (8 cores)	6
Half-Max (16 cores)	3
Max (32 cores)	0

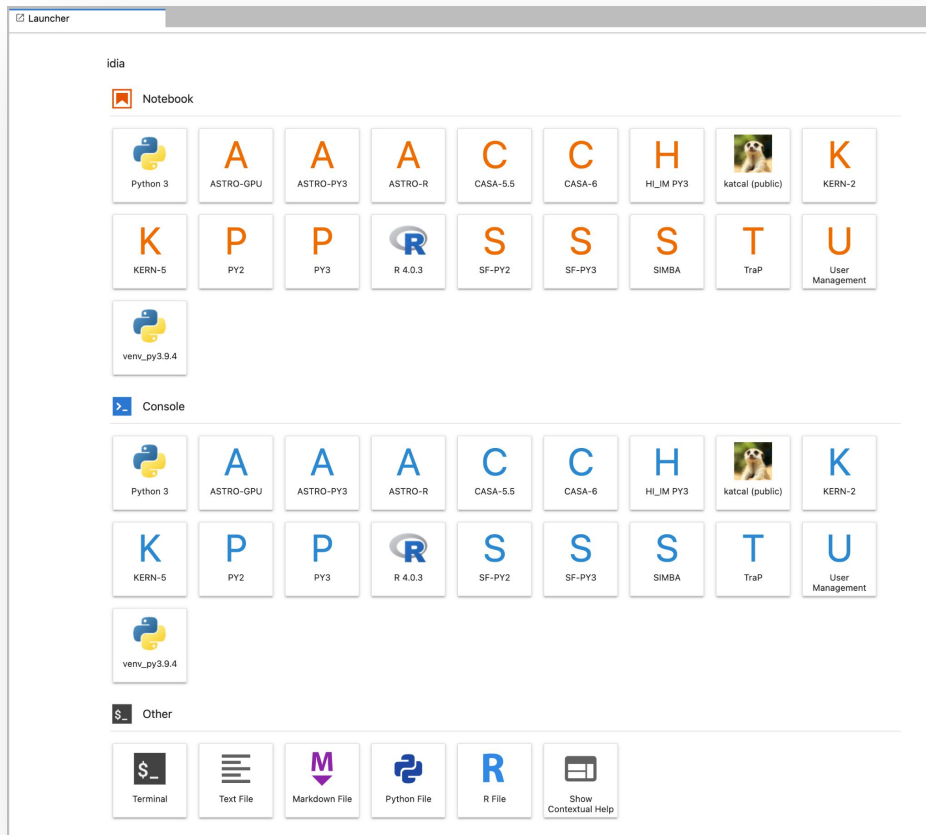
Select a job profile:

Minimum Node - 1 core, 7 GB, 18 hours idle timeout, max 5 days lifespan



# JupyterHub

Choose kernel  
in launcher



The screenshot shows the JupyterHub Launcher interface for a user named 'idia'. It is divided into three main sections: Notebook, Console, and Other.

- Notebook:** Contains a grid of kernel options. The first row includes Python 3, ASTRO-GPU, ASTRO-PY3, ASTRO-R, CASA-5.5, CASA-6, HI\_M PY3, katcal (public), and KERN-2. The second row includes KERN-5, PY2, PY3, R 4.0.3, SF-PY2, SF-PY3, SIMBA, TraP, and User Management. Below the grid is a 'venv\_py3.9.4' option.
- Console:** Contains an identical grid of kernel options as the Notebook section.
- Other:** Contains icons for Terminal, Text File, Markdown File, Python File, R File, and Show Contextual Help.



## Demo

Demo resources

[https://github.com/ilifu/ilifu\\_user\\_training](https://github.com/ilifu/ilifu_user_training)

ilifu