



IDIA

Inter-University Institute
for Data Intensive Astronomy

Ilifu Online Training

Introduction to ilifu - 12 September 2023

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

- User documentation

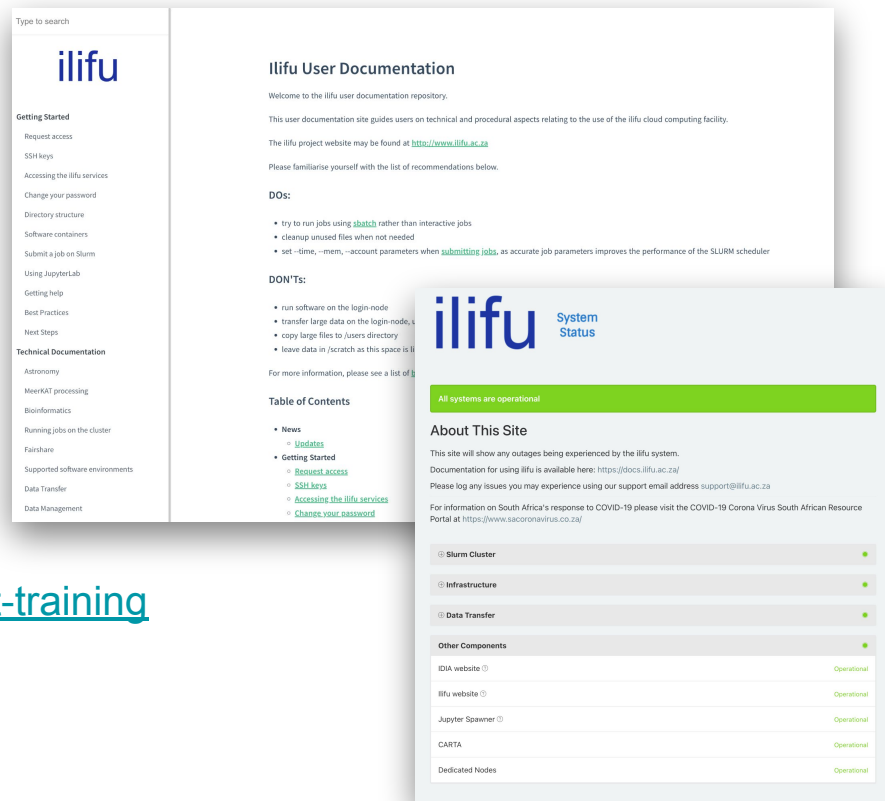
<http://docs.ilifu.ac.za>

- 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' homepage with a search bar and a navigation menu. The top-right screenshot is the 'ilifu User Documentation' page, which includes a welcome message, a list of recommendations (DOs), and a list of things to avoid (DON'Ts). The bottom-right screenshot is the 'ilifu System Status' page, which features a green banner stating 'All systems are operational' and a table listing various components and their status.

ilifu User Documentation

Welcome to the ilifu user documentation repository.

This user documentation site guides users on technical and procedural aspects relating to the use of the ilifu cloud computing facility.

The ilifu project website may be found at <http://www.ilifu.ac.za>

Please familiarise yourself with the list of recommendations below.

DOs:

- try to run jobs using `submit()` rather than interactive jobs
- cleanup unused files when not needed
- set `-time`, `-mem`, `-account` parameters when `submitting jobs`, as accurate job parameters improves the performance of the SLURM scheduler

DON'Ts:

- run software on the login node
- transfer large data on the login node,
- copy large files to `/users` directory
- leave data in `/scratch` as this space is limited

For more information, please see a list of links below.

Table of Contents

- News
 - [Updates](#)
- [Getting Started](#)
 - [Request access](#)
 - [SSH keys](#)
 - [Accessing the ilifu services](#)
 - [Change your password](#)

ilifu System Status

All systems are operational

About This Site

This site will show any outages being experienced by the ilifu system.

Documentation for using ilifu is available here: <https://docs.ilifu.ac.za/>

Please log any issues you may experience using our support email address support@ilifu.ac.za

For information on South Africa's response to COVID-19 please visit the COVID-19 Corona Virus South African Resource Portal at <https://www.sacoronavirus.co.za/>

Component	Status
Slurm Cluster	Operational
Infrastructure	Operational
Data Transfer	Operational
Other Components	Operational
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

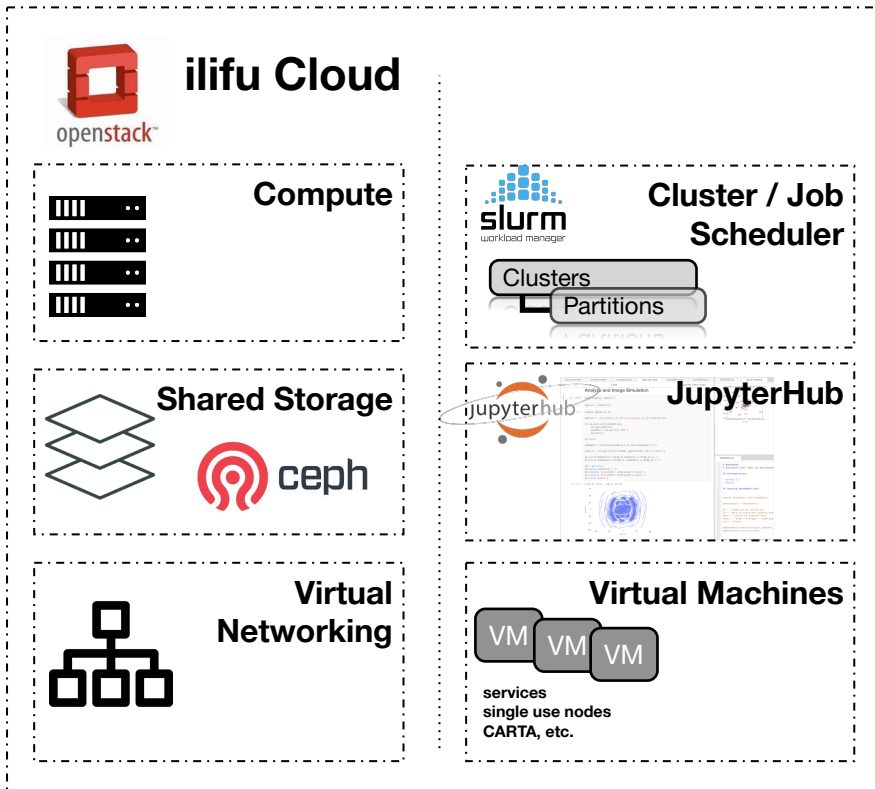


ilifu Research Facility

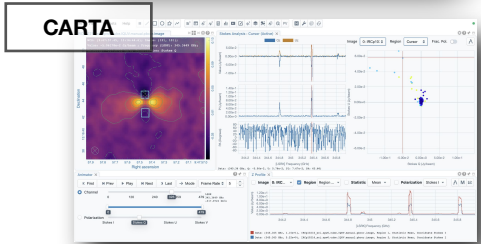
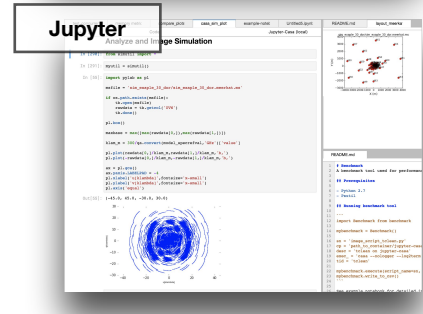
Cloud-based infrastructure for data-intensive research

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



```
ssh
[... text about interactive jobs ...]
[... text about login nodes ...]
[... text about support ...]
Last login: Fri Sep  2 08:37:30 2022 from 192.228.289.223
[... slurm partition table ...]
PARTITION    ANGL    TIME_LIMIT    NODES    STATE    MODEL_LIST
[... table content ...]
```





ilifu Research Facility - interfaces

Command line interface

ssh - shell terminal

```
* Support: https://ubuntu.com/advantage

System information as of Fri Aug 23 11:36:57 SAST 2019

System load: 0.49      Users logged in: 8
Usage of /: 35.9% of 21.15GB  IP address for ens3: 192.168.100.39
Memory usage: 5%      IP address for ens4: 10.102.26.97
Swap usage: 0%       IP address for ens5: 10.102.28.133
Processes: 396

* Keen to learn Istio? It's included in the single-package MicroK8s.
https://snapcraft.io/microk8s

Get cloud support with Ubuntu Advantage Cloud Guest:
http://www.ubuntu.com/business/services/cloud

* Canonical Livepatch is available for installation.
- Reduce system reboots and improve kernel security. Activate at:
https://ubuntu.com/livepatch

170 packages can be updated.
75 updates are security updates.

Last login: Fri Aug 23 09:08:21 2019 from 196.11.235.232
jeremy@slurm-login:~$ sinfo
PARTITION AVAIL TIMELIMIT NODES STATE MODELIST
Main* up 14-00:00:0 8 mix slwrk-[106-113]
Main* up 14-00:00:0 14 alloc slwrk-[101,104-105,114-124]
Main* up 14-00:00:0 38 idle slwrk-[102-103,125-160]
JupyterSpawnerONLY up infinite 4 mix slwrk-[201-202,205,209]
JupyterSpawnerONLY up infinite 4 alloc slwrk-[206-208,210]
JupyterSpawnerONLY up infinite 2 idle slwrk-[203-204]
jeremy@slurm-login:~$ sbatch compute_job.sh
```

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

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

Web applications

The screenshot displays a JupyterLab web application interface. The main area shows a code editor with Python code for image simulation analysis. The code includes imports for 'simutil', 'pylab', and 'os', followed by file operations and plotting functions. The output area shows a plot of 'k' vs 'x' with a circular pattern of data points. To the right, there are several smaller plots and a 'README.md' file. The interface includes a top navigation bar with tabs for 'test_corpusmer', 'memory metrics', 'compare_plots', 'casa_sim_plot', 'example-note', 'Untitled5.py.txt', 'README.md', and 'layout_meeke'. The bottom status bar shows 'Jupyter-Casa (local)' and 'JupyterLab 0.35.0'.

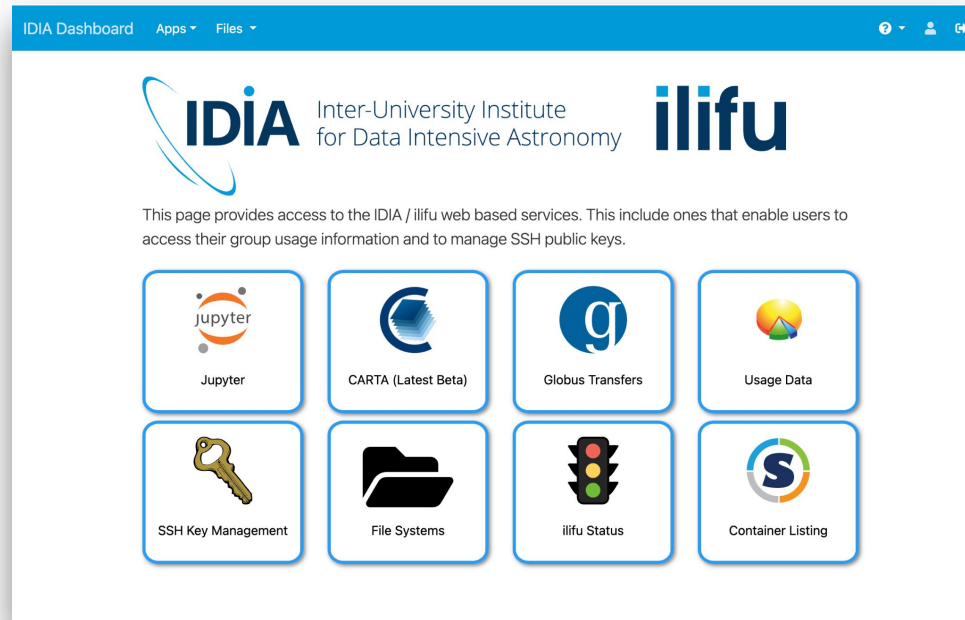
<https://jupyter.ilifu.ac.za>

<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 dropdown menus for "Apps" and "Files". On the right side of the bar 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 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 and green 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

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

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:

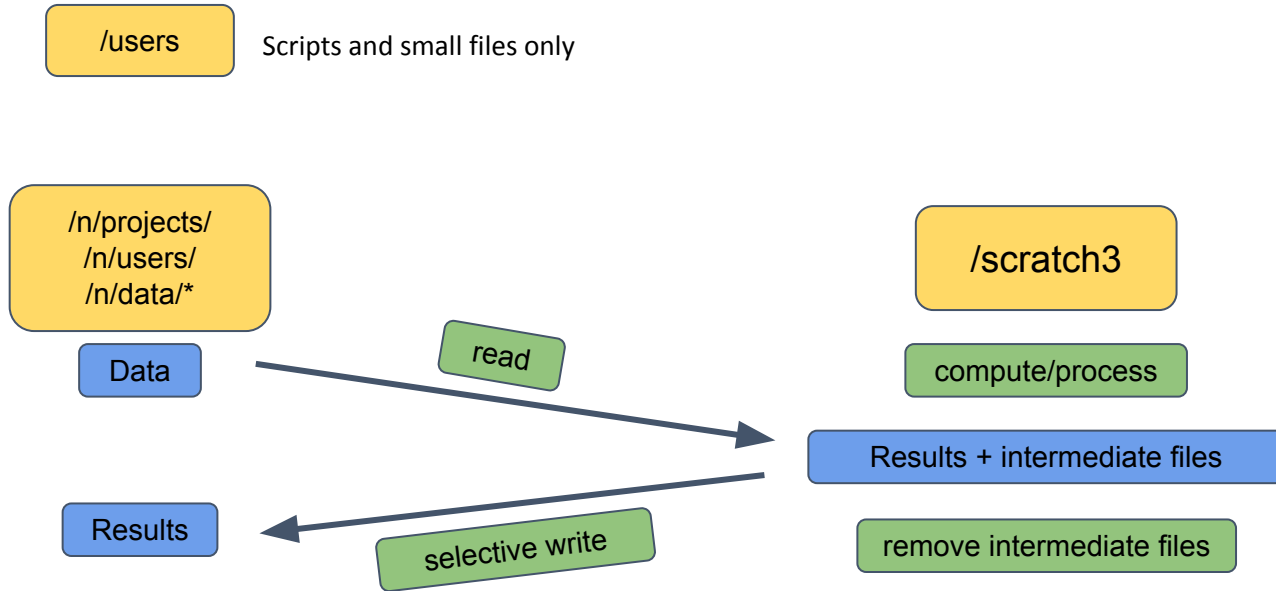
- `/idia/users`
 - user's private work directory, may store data products that are not ready to move to shared project space
- `/idia/projects`
 - project specific directories. These directories are for sharing data and resources within project groups. Raw data associated with a project will also be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

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
-----
bbarelto/1.6.1      casa/5.7.0      casa/5.8.0      casa/6.1.2.7-pipeline      /software/modules/astro      casa/6.4      pybdsf/1.9.2
bbarelto/1.6.7 (D) casa/5.7.2-4      casa/6.1.8-118-monolithic casa/6.1.2.7-modular      casa/6.3      casa/6.4.3      casa/6.5.0 (D) pybdsf/1.10.1 (D)

cinsv/C100V/1.0-0  exonerater/13.0.1  itrtree/1.6.12  soopen/0.1.1  /software/modules/dio  svanno/1.0.1
GenomeBrowser/c11 fastqc/0.11.9  itrtree/2.2.0 (D)  proteomizard/proteomizard_3.0.22167  treeP/homewr
anomer/2020-08-07  fillup/0.2.1  listcol/2.3.4  proteomizard/3.16  trypcycler/trypcycler.biocntainers.0.5.4--pyhd78af78a
scbio/bcbio_container  fseqpipe/16.0  maffix/7.090  pypsa/0.12.0  trypcycler/trypcycler.staph.0.5.4
scbio/1.1.3  gatk4-gt/4.2.0.1  msnv/0.9  r2tools/0.1.1  vartools/0.5
scbio/1.1.2.9 (D)  gatk4-2.5.0  nedsa/1.7.3  santools/1.10  vcfTools/0.1.16
scftools/1.20.2 (D)  gatk4-3.0.0  ngsutils/0.12  santools/1.12  vep/100.1
scftools/1.17 (D)  gemini/gemini  nanoplot/1.41.0  santools/1.14  vep/100.1
scsoft/blast/1.0.4  geminirep/2.00.1958  nantflow/22.0.0  santools/1.17  vep/100.1
biobambard/2.0.183  gromacs/2022.2  phenyl/Saring2822  savvysite/git
bwa/gt/0.8  anaconda/2021.11  htlib/1.18.2  pipx/2.0.1  vep/100.1
cavus/1.40.0.1013  htlib/1.17 (D)  plink/2.0902.3  strtools/3.0.0
cd-hlt/4.0.2  nhyh/2.5.49  polyolith/0.5.0  stackz/2.01

-----
LAPACK/3.9.0      R/4.2.0      cuda/11.4.2_478.57.02  /software/modules/common  nodejs/18.13.0  peti/1.33.0  python/3.9.4
R/RStudio/2.0.842-R4.0.0  anaconda3  cuda/11.0.0_510.39.01  java/jre-1.8.0_261  nodejs/19.4.0  python/3.9.7
R/RStudio/2.0.842-R4.0.4  anaconda3/login.oid  cuda/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/RStudio/2.0.842-R4.2.0  anaconda3/login  dotnet/6.0.301  java/openjdk-17.0.2  openssl/1.1.1  python/3.6.15  python/3.7.56  python/3.10.4
R/RStudio/2.0.842-R4.2.1  anaconda3/2020.07  draw.io/1.13  java/openjdk-18.0.2 (D)  openssl/1.1.0  python/2.7.7  python/3.10.0  python/3.10.8
R/RStudio2022.12.0-303-R4.2.2  anaconda3/2021.05  gitHubcli/2.0.0  julia/1.5.3  openssl/1.1.0  python/3.7.16  python/3.10.1  python/3.10.4
R/4.0.0  anaconda3/2021.11  go/1.16.3  libqjs/2.7.1.3  openssl/1.0.2  python/3.8.2  python/3.10.4
R/4.0.0  cuda/10.0.130_410.48  go/1.17.3  maven/3.6.3  openssl/1.0.5  python/3.8.3  python/3.10.9
R/4.0.2  cuda/10.1.243_410.07.00  go/1.18.4  mongo/6.0.0_323  openssl/1.1.0  python/3.8.0  python/3.11.0
R/4.0.3  cuda/10.2.09_440.30.05  go/1.18.4  mongo/6.0.2  openssl/1.1.1 (D)  python/3.8.10  python/3.11.3
R/4.1.1  cuda/11.0.2_450.51.05  go/1.18.4  neo4j/0.8.2  openssl/1.1.1  python/3.9.0  python/3.11.2

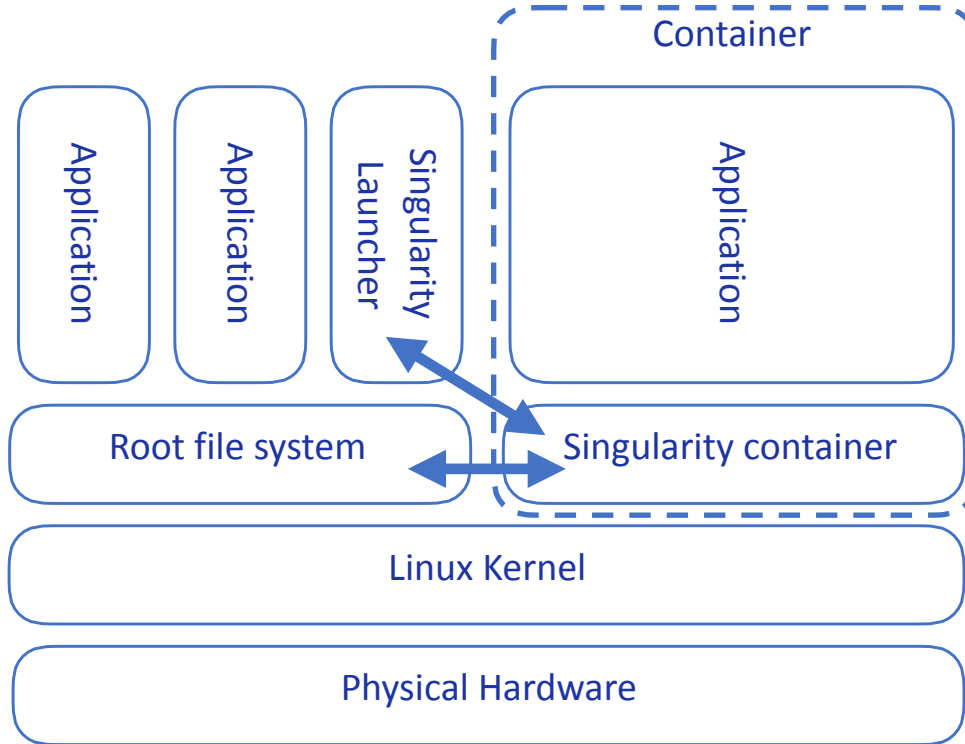
-----
Core/1mod/6.0  Core/settings/6.0  /usr/share/1mod/modulefiles

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

Server Options

Nodes Free

as at Tue Mar 8 14:44:01 SAST 2022

83 Minimum

40 Small

18 Medium

7 Large

1 Half-Max

0 Max

0 GPU

Select a job profile:

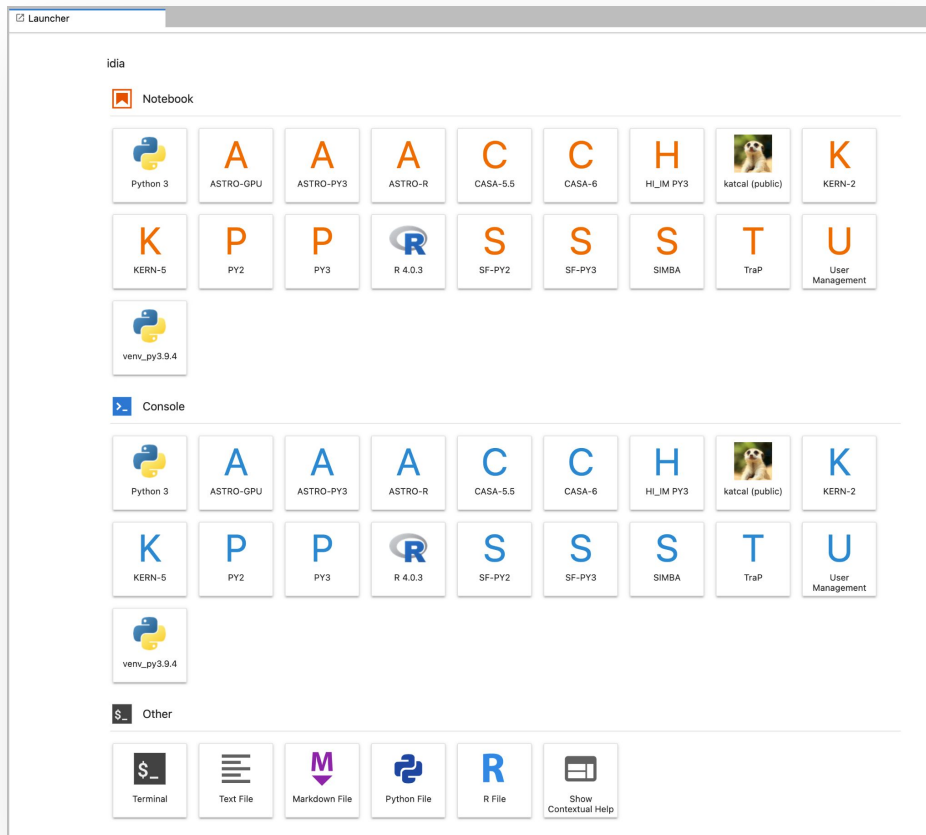
Minimum Node - 1 core, 7 GB, 18 hours idle timeout, max 5 days lifespan



Start

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:** A grid of 28 kernel options. The first row contains: Python 3, ASTRO-GPU, ASTRO-PY3, ASTRO-R, CASA-5.5, CASA-6, HI_M PY3, katcal (public), and KERN-2. The second row contains: 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:** A grid of 28 kernel options, identical to the Notebook section.
- Other:** A row of six utility icons: Terminal, Text File, Markdown File, Python File, R File, and Show Contextual Help.



Demo

Demo resources https://github.com/ilifu/ilifu_user_training