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

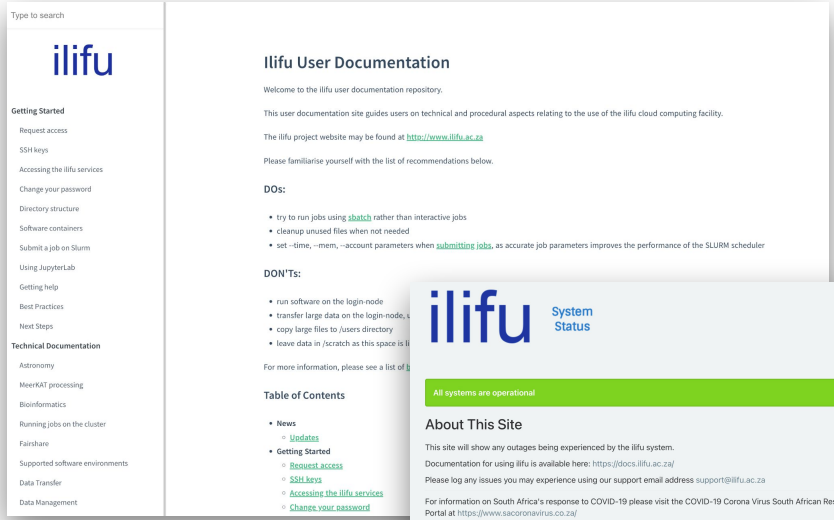
- Support contact

  support@ilifu.ac.za
- User documentation

  http://docs.ilifu.ac.za
- Ilifu System Status

  https://status.ilifu.ac.za
- Training videos

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



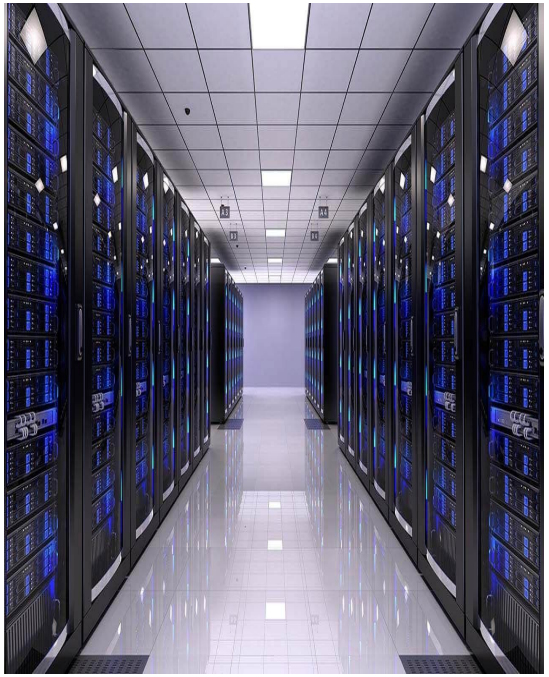
High Performance Computing

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

ilifu Cloud

openstack

Compute

Shared Storage

ceph

Virtual Networking

Cluster / Job Scheduler

slurm workload manager

Clusters

Partitions

JupyterHub

jupyterhub

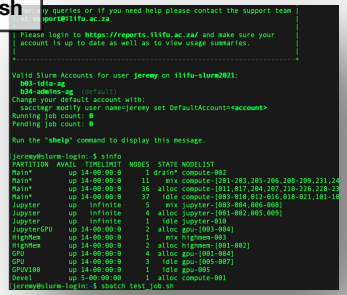
Virtual Machines

VM VM VM

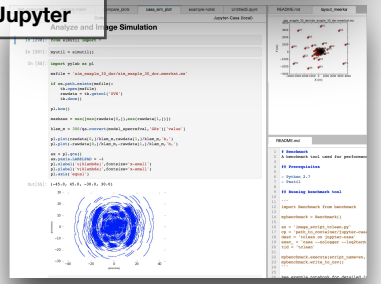
services
single use nodes
CARTA, etc.

Interfaces

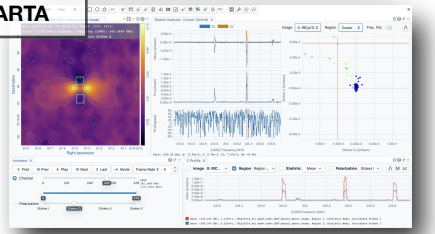
ssh



Jupyter



CARTA



ilifu Research Facility - interfaces

Command line interface

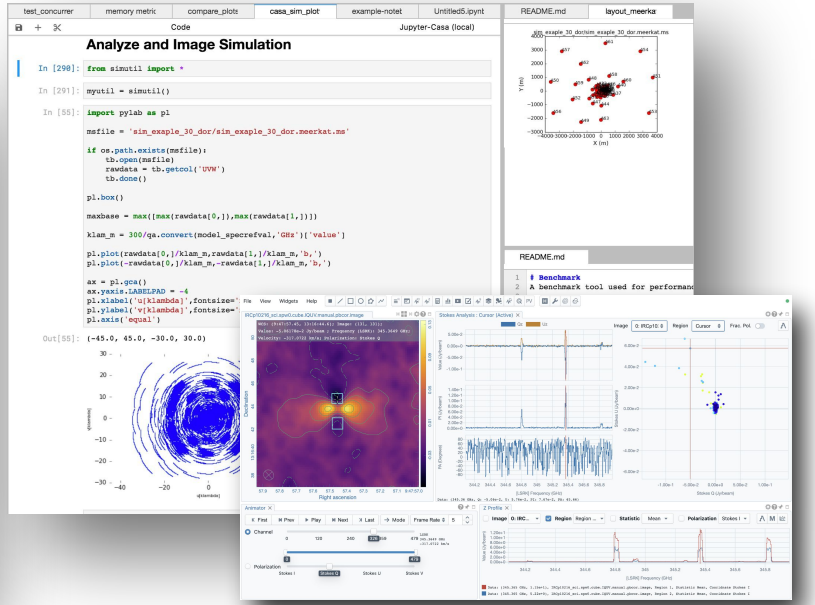
ssh - shell terminal

```
+-------------------------------------------------------------------------+
| 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

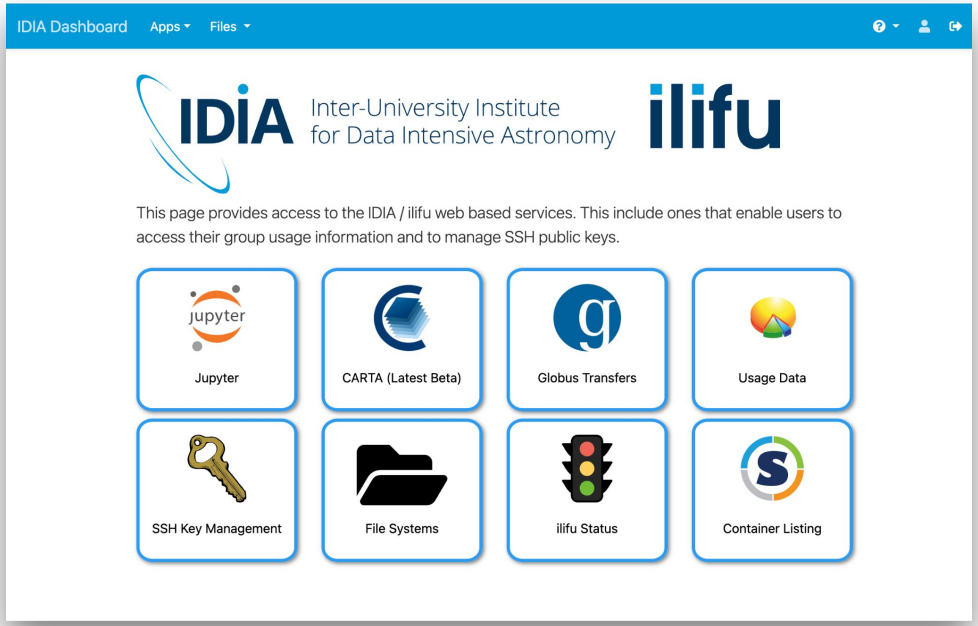


https://jupyter.ilifu.ac.za

https://carta.idia.ac.za

ilifu Research Facility - Science Gateway

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



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

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:

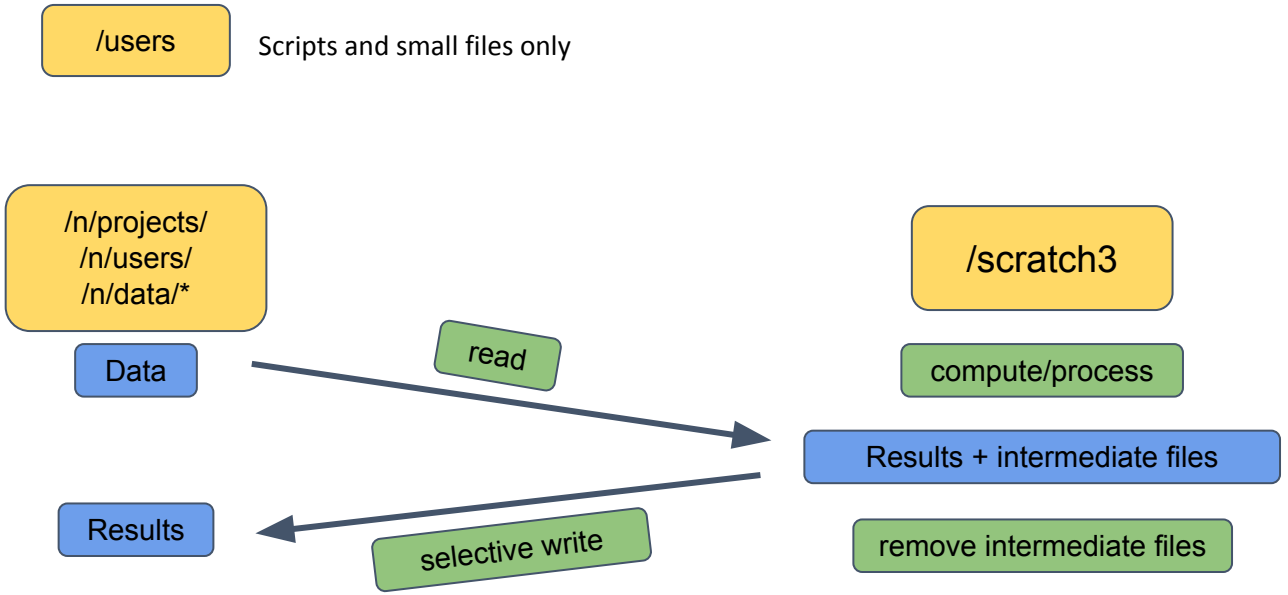
- /idia/users
  - user's private work directory, may store data products that are not ready to move to shared project space
- /idia/projects
  - project specific directories. These directories are for sharing data and resources within project groups. Raw data associated with a project will also be available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



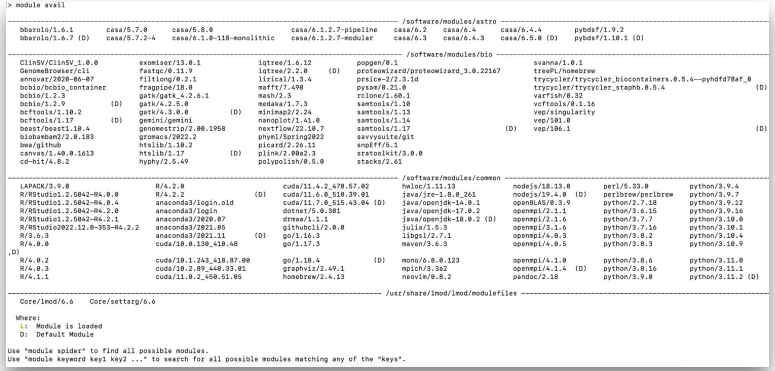
*/n/data generally read-only

n = idia,cbio,ilifu

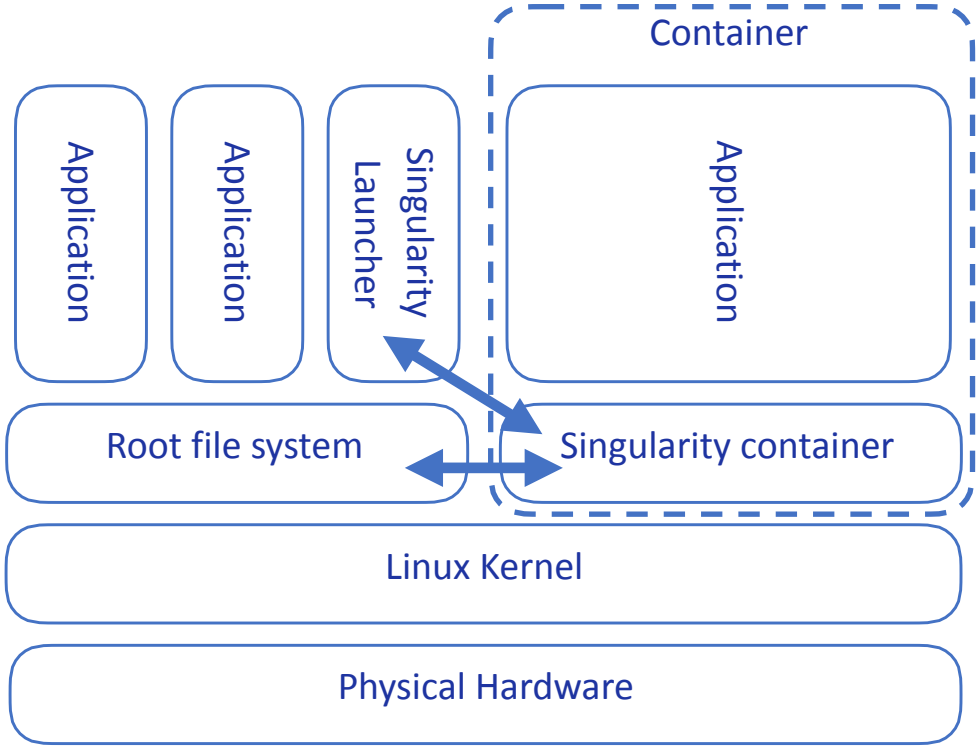
Ilifu Software Environment

- 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

SINGULARITYCE



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       (D)    go/1.17.3            (D)    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                  (D)    cuda/11.4.2_470.57.02   (D)    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    htslib/1.10.2            popgen/0.1         samtools/1.14   (D)    vep/101.0       (D)
   bcbio/1.2.9          (D)    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

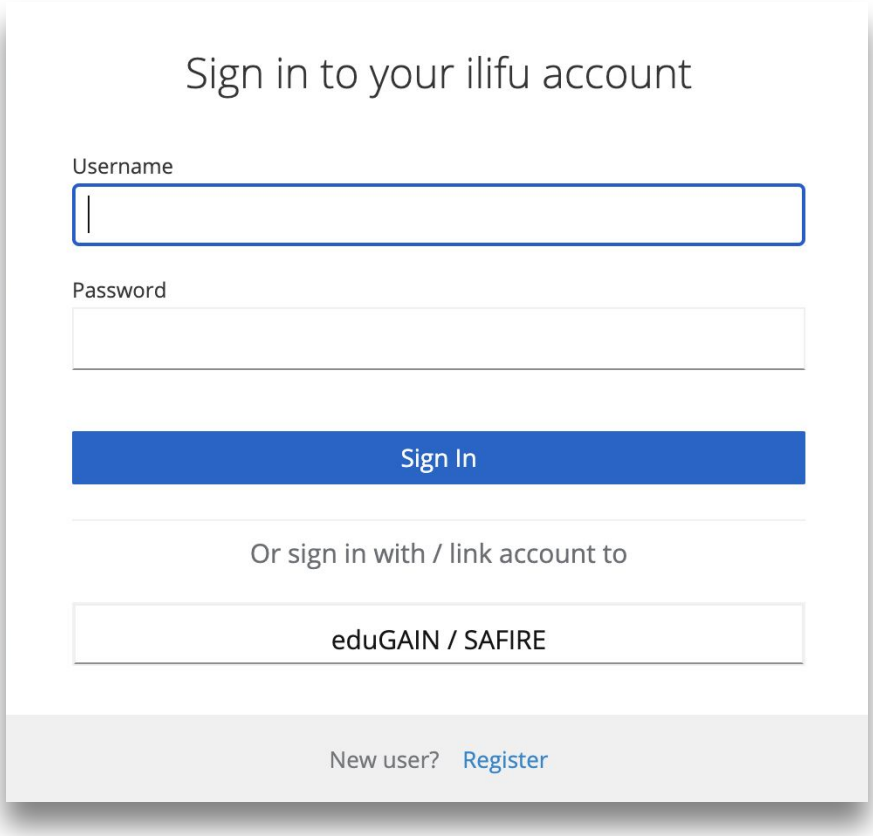
- module avail
- module help <module>

```
$ 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



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:

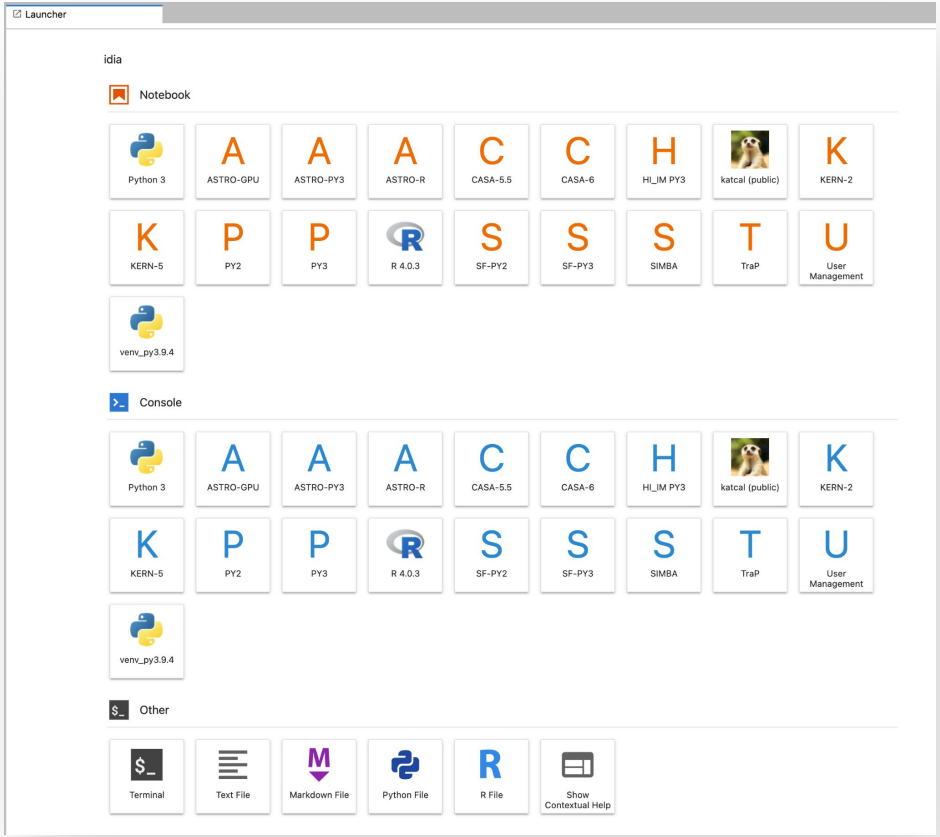
Minimum Node - 1 core, 7 GB, 18 hours idle timeout, max 5 days lifespan

Start

JupyterHub

Choose kernel

in launcher



Demo

Demo resources

https://github.com/ilifu/ilifu_user_training