IDIA Inter-University Institute for Data Intensive Astronomy

Ilifu Online Training

Session 1: Introduction to ilifu

2 September 2025

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

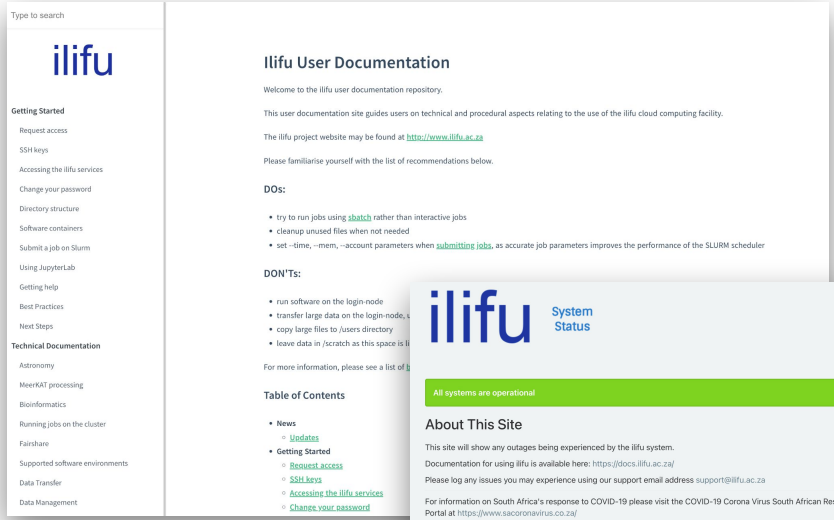
  support@ilifu.ac.za
- User documentation

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

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

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

  https://github.com/ilifu/ilifu_user_training



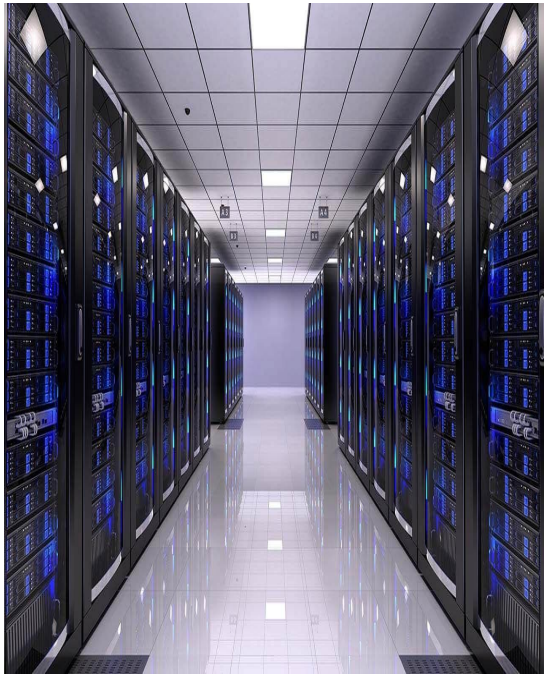
High Performance Computing

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

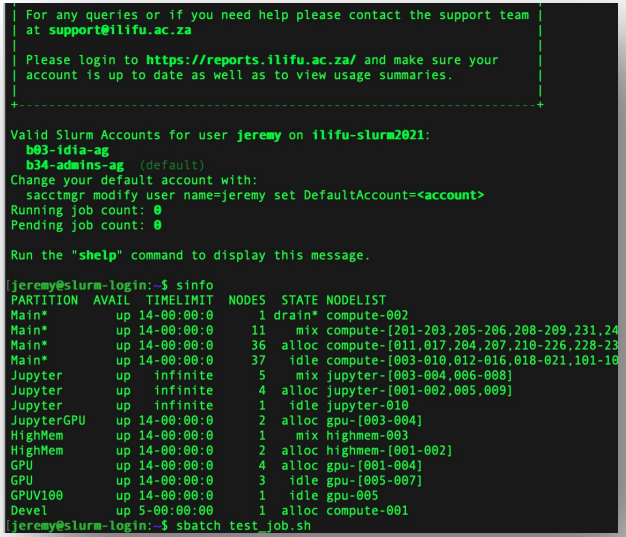
Network of remote servers, accessed over the internet, to store, manage, and process data

- 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

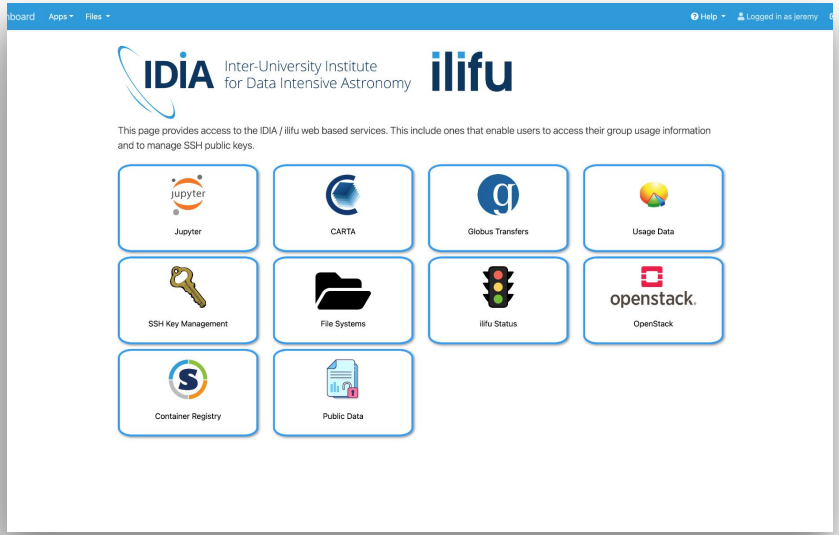


`ssh <username>@slurm.ilifu.ac.za`

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

Web applications

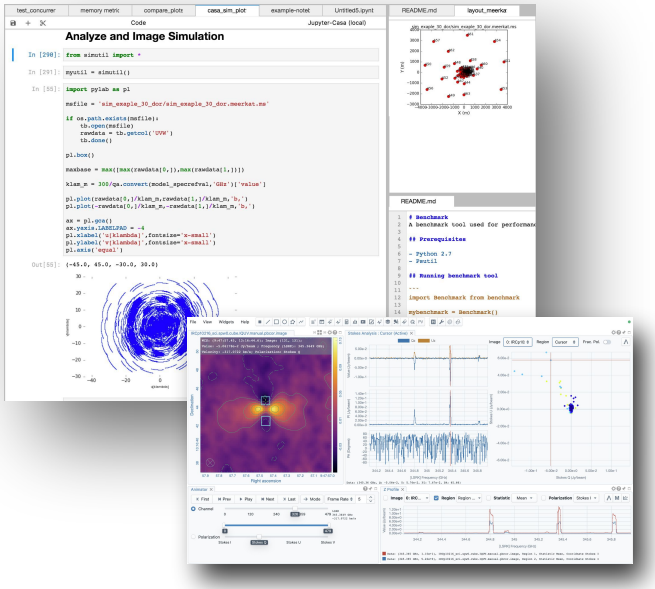
IDIA Science Gateway - App Dashboard



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:

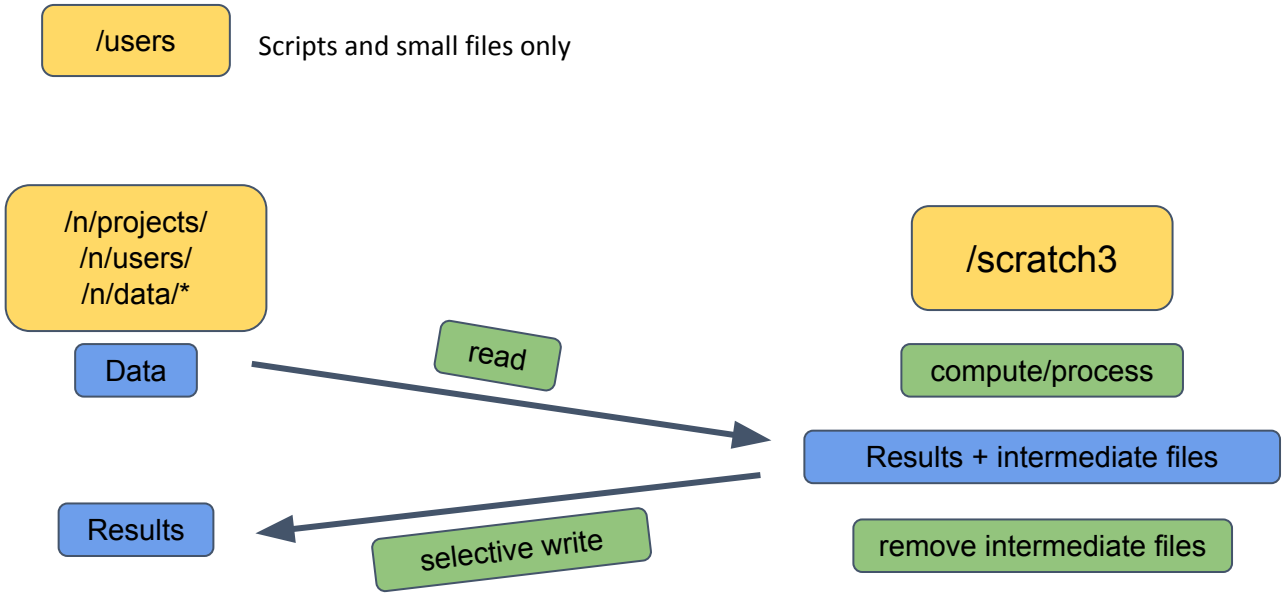
- /idia/users
  - user's private work directory, may store data products that are not ready to move to shared project space. **Limited to 10TB**
- /idia/projects
  - project specific directories. These directories are for sharing data and resources within project groups. Raw data associated with a project 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



*/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

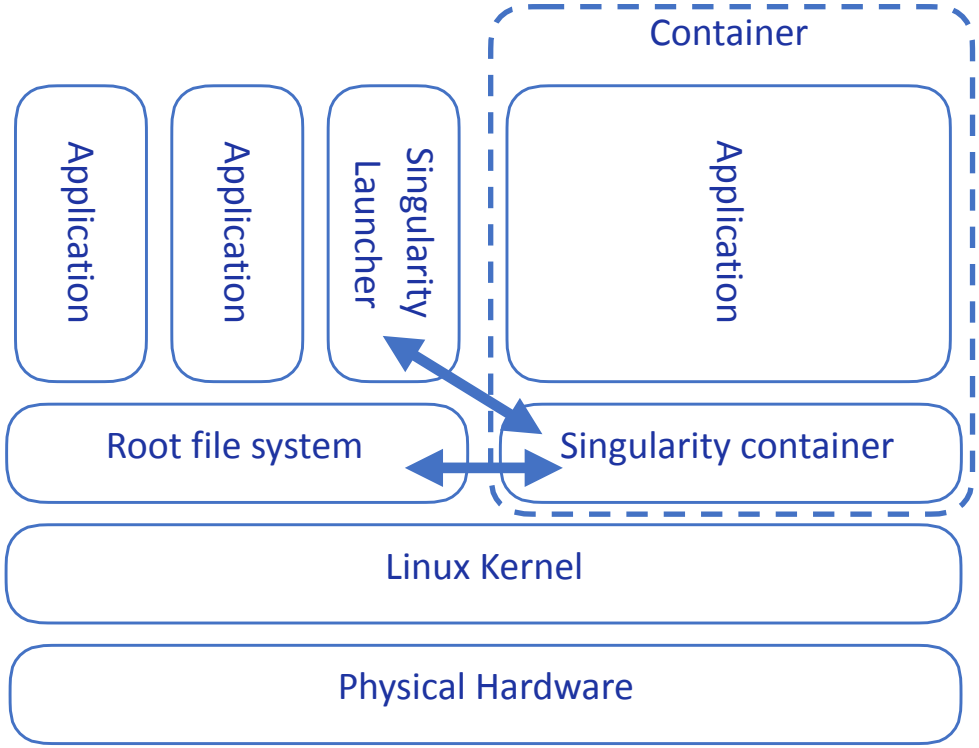
```
> module avail

---------------------------------------- /software/modules/astro ----------------------------------------
   bbarolo/1.6.1       casa/5.7.0      casa/5.8.0                    casa/6.1.2.7-pipeline    casa/6.2    casa/6.4      casa/6.4.4        pybdsf/1.9.2
   bbarolo/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)    pybdsf/1.10.1 (D)

---------------------------------------- /software/modules/bio ----------------------------------------
   ClinSV/ClinSV_1.0.0          exomiser/13.0.1          iqtree/1.6.12          popgen/0.1                                   svanna/1.0.1
   GenomeBrowser/cli            fastqc/0.11.9            iqtree/2.2.0     (D)   proteowizard/proteowizard_3.0.22167          treePL/homebrew
   annovar/2020-06-07           filtlong/0.2.1           lirical/1.3.4          prsice-2/2.3.1d                              trycycler/trycycler_biocontainers.0.5.4--pyhdfd78af_0
   bcbio/bcbio_container        fragpipe/18.0            mafft/7.490            pysam/0.21.0                                 trycycler/trycycler_staphb.0.5.4                      (D)
   bcbio/1.2.3                  gatk/gatk_4.2.6.1        mash/2.3               rclone/1.60.1                                varfish/0.32
   bcbio/1.2.9           (D)    gatk/4.2.5.0             medaka/1.7.3           samtools/1.10                                vcftools/0.1.16
   bcftools/1.10.2              gatk/4.3.0.0       (D)   minimap2/2.24          samtools/1.13                                vep/singularity
   bcftools/1.17         (D)    gemini/gemini            nanoplot/1.41.0        samtools/1.14                                vep/101.0
   beast/beast1.10.4            genomestrip/2.00.1958    nextflow/22.10.7       samtools/1.17                         (D)    vep/106.1                                             (D)
   biobambam2/2.0.183           gromacs/2022.2           phyml/Spring2022       savvysuite/git
   bwa/github                   htslib/1.10.2            picard/2.26.11         snpEff/5.1
   canvas/1.40.0.1613           htslib/1.17        (D)   plink/2.00a2.3         sratoolkit/3.0.0
   cd-hit/4.8.2                 hyphy/2.5.49             polypolish/0.5.0       stacks/2.61

---------------------------------------- /software/modules/common ----------------------------------------
   LAPACK/3.9.0                     R/4.2.0                       cuda/11.4.2_470.57.02         hwloc/1.11.13                nodejs/18.13.0        perl/5.33.0          python/3.9.4
   R/RStudio1.2.5042-R4.0.0         R/4.2.2                 (D)   cuda/11.6.0_510.39.01         java/jre-1.8.0_261           nodejs/19.4.0   (D)   perlbrew/perlbrew    python/3.9.7
   R/RStudio1.2.5042-R4.0.4         anaconda3/login.old           cuda/11.7.0_515.43.04 (D)     java/openjdk-14.0.1          openBLAS/0.3.9        python/2.7.18        python/3.9.12
   R/RStudio1.2.5042-R4.2.0         anaconda3/login               dotnet/5.0.301                java/openjdk-17.0.2          openmpi/2.1.1         python/3.6.15        python/3.9.16
   R/RStudio1.2.5042-R4.2.1         anaconda3/2020.07             drmaa/1.1.1                   java/openjdk-18.0.2 (D)      openmpi/2.1.6         python/3.7.7         python/3.10.0
   R/RStudio2022.12.0-353-R4.2.2    anaconda3/2021.05             githubcli/2.0.0               julia/1.5.3                  openmpi/3.1.6         python/3.7.16        python/3.10.1
   R/3.6.3                          anaconda3/2021.11       (D)   go/1.16.3                     libgsl/2.7.1                 openmpi/4.0.3         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                  openmpi/4.0.5         python/3.8.3         python/3.10.9
,D)
   R/4.0.2                          cuda/10.1.243_418.87.00       go/1.18.4             (D)     mono/6.8.0.123               openmpi/4.1.0         python/3.8.6         python/3.11.0
   R/4.0.3                          cuda/10.2.89_440.33.01        graphviz/2.49.1               mpich/3.3a2                  openmpi/4.1.4   (D)   python/3.8.16        python/3.11.1
   R/4.1.1                          cuda/11.0.2_450.51.05         homebrew/2.4.13               neovim/0.8.2                 pandoc/2.18           python/3.9.0         python/3.11.2 (D)

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

  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".
```

IDIA

SINGULARITYCE

ilifu

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.sif
```

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
$ singularity inspect -d /idia/software/containers/casa-6.simg
```

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
   calceph/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
```

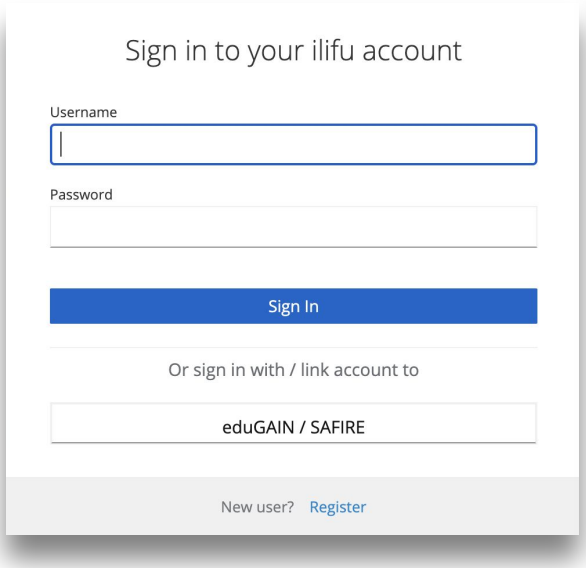
- module load <module>
- module list
- module purge
- module --help

JupyterHub

https://gateway.idia.ac.za

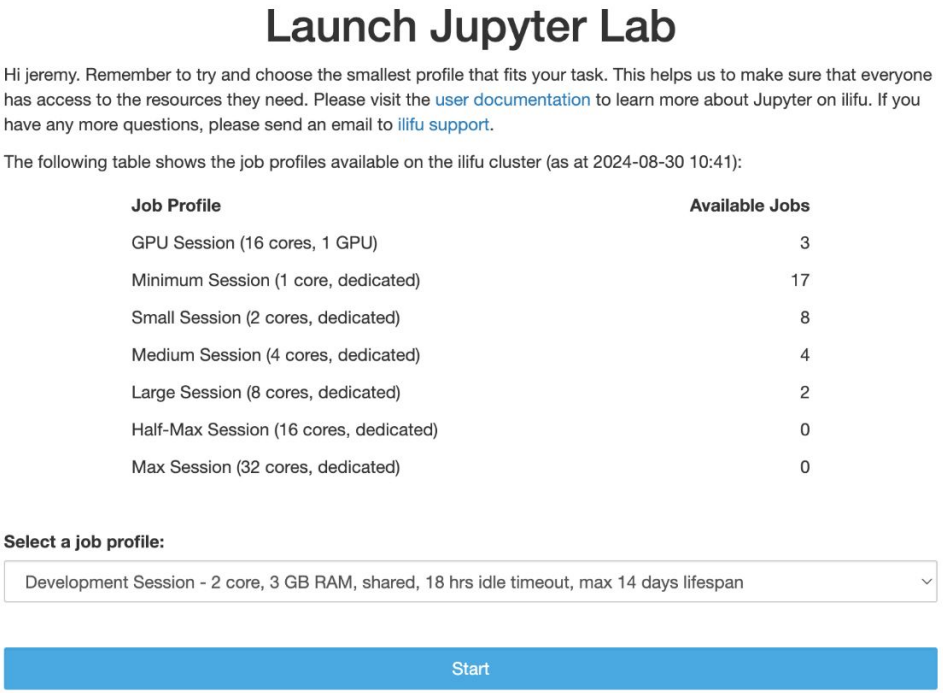
OR

https://jupyter.ilifu.ac.za



JupyterHub

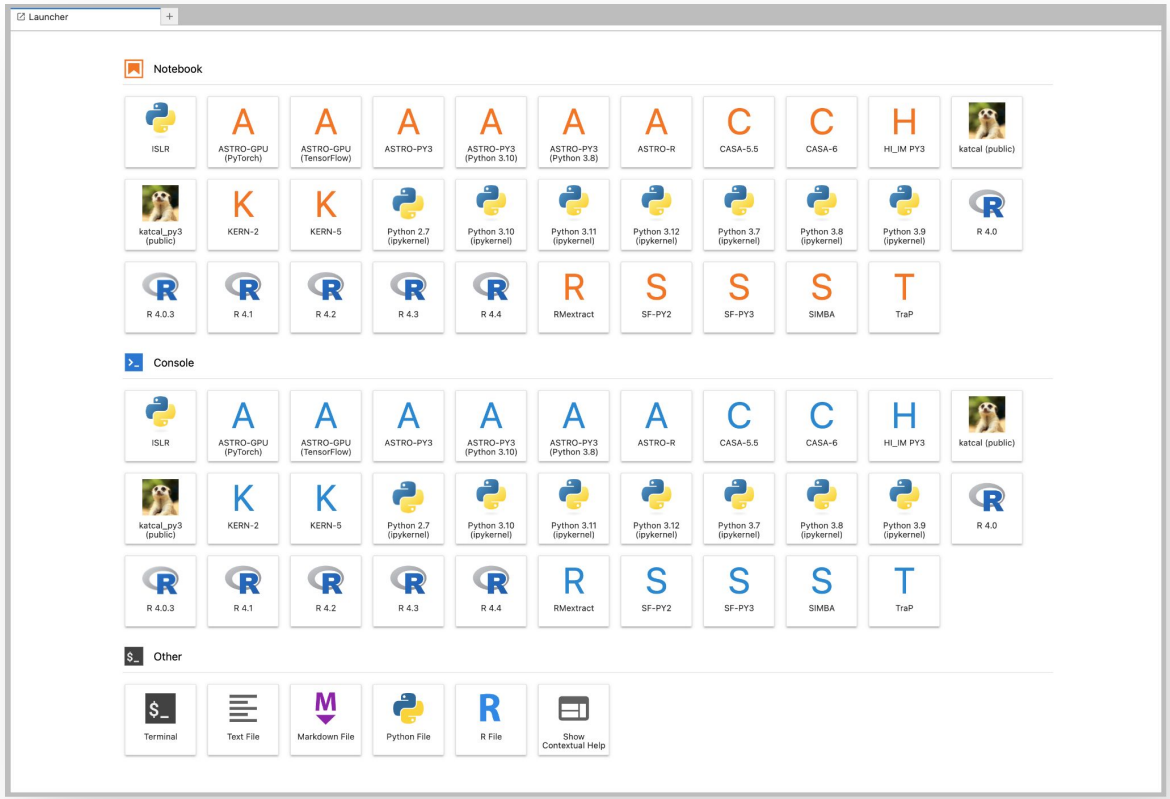
Session size



JupyterHub

Choose kernel

in launcher



Demo

Demo resources

https://github.com/ilifu/ilifu_user_training