



IDIA

Inter-University Institute
for Data Intensive Astronomy

Ilifu Online Training

Introduction to ilifu - 3 September 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

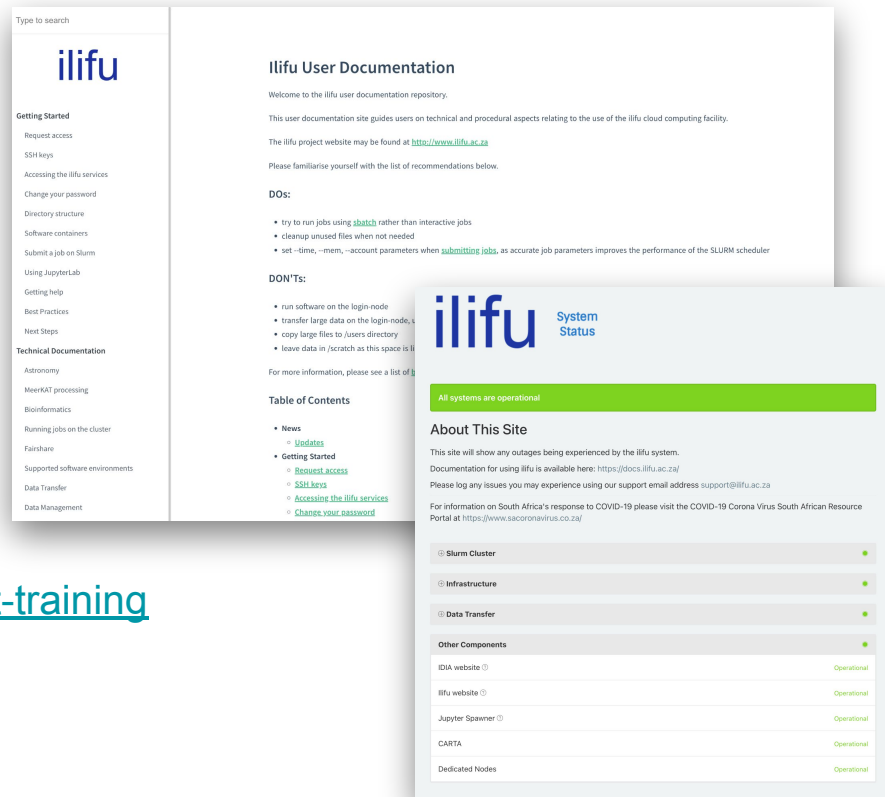
<http://docs.ilifu.ac.za>

- Ilifu System Status

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

- Training videos

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



The image shows three overlapping screenshots of the ilifu website. The top-left screenshot is the 'ilifu' homepage, featuring a search bar, a navigation menu with categories like 'Getting Started', 'Technical Documentation', and 'Fairshare', and a 'Table of Contents' section. 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 displays 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	
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

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 (default)
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

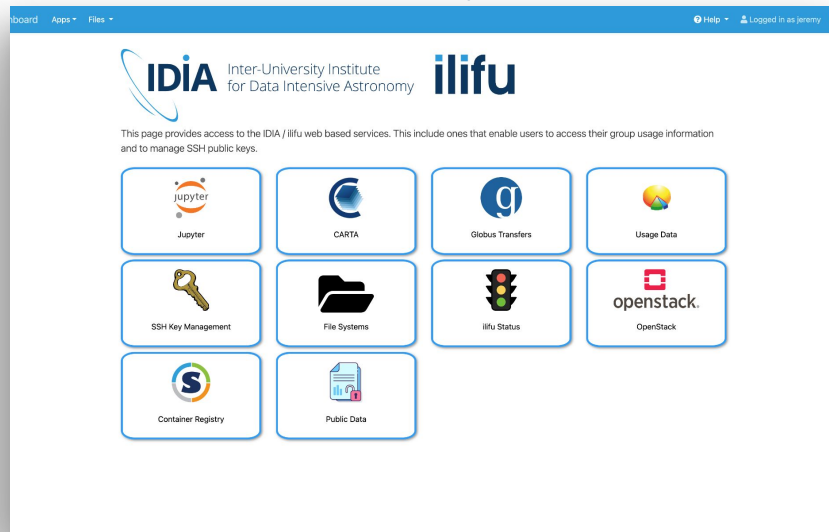
Run the "shelp" command to display this message.

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-[201-203,205-206,208-209,211,24
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

```

Web applications

IDIA Science Gateway - App Dashboard



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

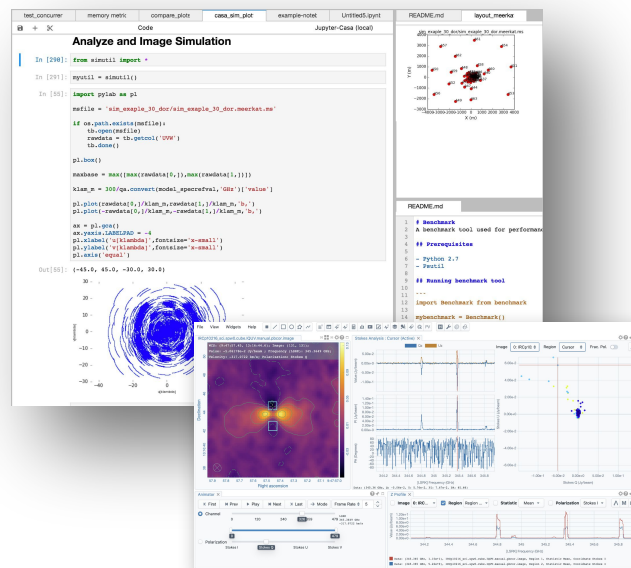
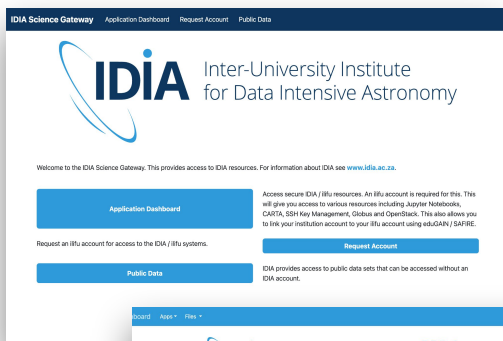
ssh <username>@transfer.ilifu.ac.za

<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.
- /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:

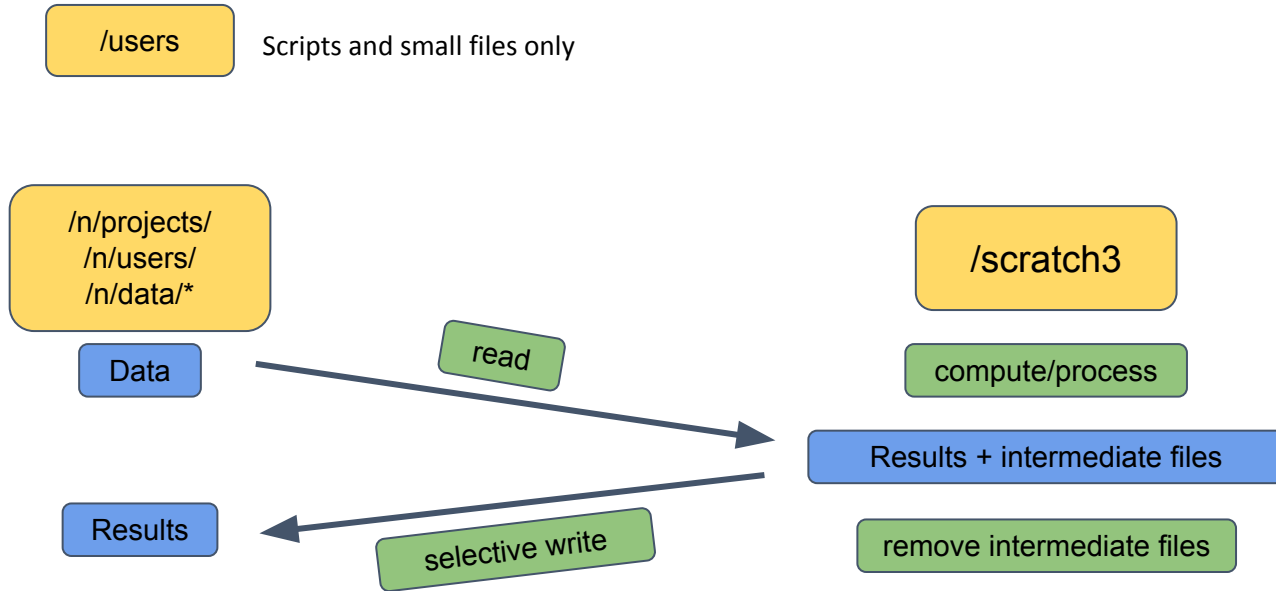
- `/idia/users`
 - user's private work directory, may store data products that are not ready to move to shared project space
- `/idia/projects`
 - project specific directories. These directories are for sharing data and resources within project groups. Raw data associated with a project will also be 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
-----
bbarelto/1.6.1      casa/5.7.0      casa/5.8.0      casa/6.1.2.7-pipeline  /software/modules/astro  casa/6.2  casa/6.4  casa/6.4.4  pybids/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) pybids/1.10.1 (D)

cclink/C110V/1.0.0  exonerater/13.0.1  itrtree/1.6.12  soopen/0.1  /software/modules/dio  svnmw/1.0.1
GenomeBrowser/c11  fastqc/0.11.9  itrtree/2.2.0 (D)  proteomizard/proteomizard_3.0.22167  treeP/nomewr
genome/2020-08-07  fillup/0.2.1  listcol/2.3.4  proteomizard/3.16  trypsin/trypsin
scbio/bcbio_container  fseqpipe/16.0  maffix/7.090  pypsa/0.12.0  trypsin/trypsin_biconainers_0.5.4--pyhd878af7
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  vcfutils/0.1.16
scfnalib/1.20.2 (D)  gatk4-3.0.0  ngsnp/0.2.0  santools/1.12  vep/100.1
scfnalib/1.17 (D)  gemini/gemini  nanoplot/1.14  santools/1.14  vep/100.1
scfnalib/1.18.4 (D)  gemini/gemini  nanoplot/22.0.0  santools/1.17  vep/100.1
scfnalib/1.18.4 (D)  gemini/gemini  nanoplot/22.0.0  santools/1.17  vep/100.1
biolabhard/2.0.183  gromacs/2022.2  psvl/Svrlng2822  savvysite/git
bioconda/1.40.0.1013  htlib/1.18.2  psvl/Svrlng2822  savvysite/git
cavus/1.40.0.1013  htlib/1.17 (D)  plink/2.0902.3  strtools/3.0.0
cd-hlt/4.0.2  hyphy/2.5.49  polypolish/0.5.0  strtools/3.0.0

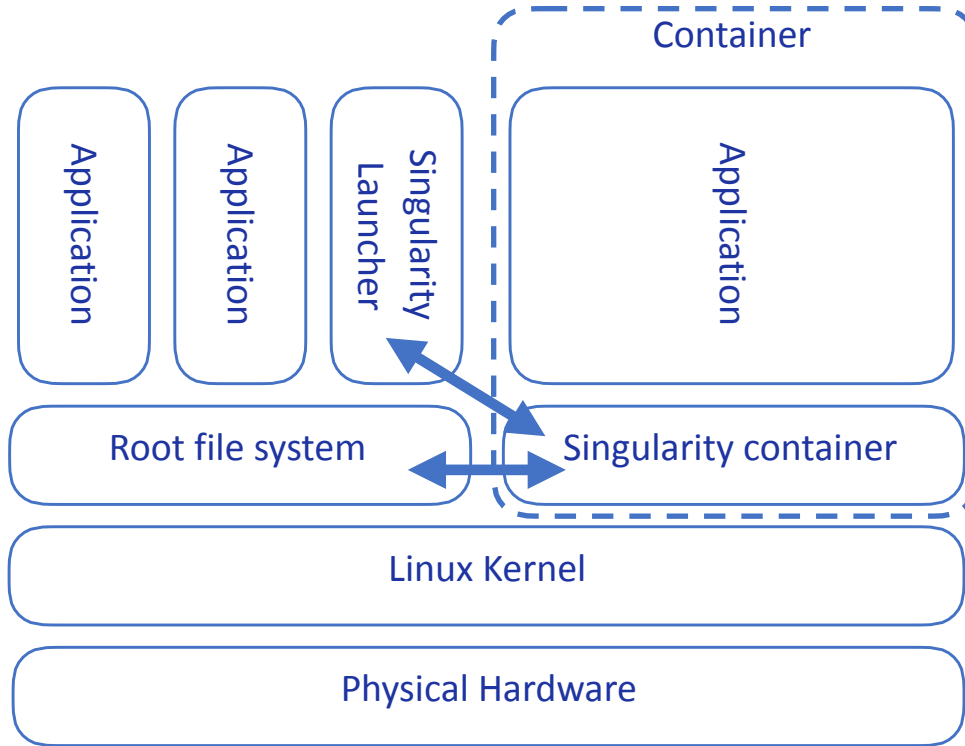
LAPACK/3.9.0  R/4.2.0  cudf/11.4.2.470.57.02  /software/modules/common  nodejs/18.13.0  peti/1.33.0  python/3.9.4
R/RStudio/2.0.842-R4.0.0  R/4.2.0  cudf/11.0.0.510.39.01  java/7re-1.0.0.241  nodejs/19.4.0  python/3.9.7
R/RStudio/2.0.842-R4.0.4  anaconda3/login  cudf/11.7.0.511.43.04 (D)  java/openjdk-14.0.2  oemmpi/3.0.3  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  oemmpi/2.1.1  python/3.6.15  python/3.9.16
R/RStudio/2.0.842-R4.2.1  anaconda3/2020.07  draw/1.1.1  java/openjdk-18.0.2 (D)  oemmpi/2.1.4  python/2.7.7  python/3.10.0
R/RStudio2022.12.0-303-R4.2.2  anaconda3/2021.05  gitHubutils/2.0.0  julia/1.5.3  oemmpi/3.1.4  python/3.7.16  python/3.10.1
R/4.0.0  anaconda3/2021.11  go/1.16.3  libq3/1.7.1.3  oemmpi/4.0.2  python/3.8.2  python/3.10.4
R/4.0.0  cuda/10.0.130_118_48  go/1.17.3  maven/3.6.3  oemmpi/4.0.5  python/3.8.3  python/3.10.9
R/4.0.2  cuda/10.1.243_118_07_00  go/1.18.4  mongo/6.0.0.123  oemmpi/4.1.0  python/3.8.0  python/3.11.0
R/4.0.3  cuda/10.2.89_118_20_05  gromacs/22.49.1  mongo/6.0.2  oemmpi/4.1.4 (D)  python/3.10.16  python/3.11.3
R/4.1.1  cuda/11.0.2.450-11.05  horebwr/2.4.13  neo4j/0.8.2  pandas/2.0.2  python/3.9.0  python/3.11.2 (D)

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

Where:
  I: 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, 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
```



Software environment - modules

module avail

```
$ module avail
```

```
-----/software/modules/astro-----
atnfcats/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      psrpat/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   parxml/1.01      sofa/20180130

-----/software/modules/bio-----
ClinSV/ClinSV_1.0.0  filtlong/0.2.1      mercury/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      samtools/1.19.2 (D)
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       gfstats/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>

Sign in to your ilifu account

Username

Password

Sign In

Or sign in with / link account to

eduGAIN / SAFIRE

New user? [Register](#)

ilifu

board Apps Files Help Logged in as Jeremy

IDIA Inter-University Institute for Data Intensive Astronomy **ilifu**

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.

 Jupyter	 CARTA	 Globus Transfers	 Usage Data
 SSH Key Management	 File Systems	 ilifu Status	 OpenStack
 Container Registry	 Public Data		



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

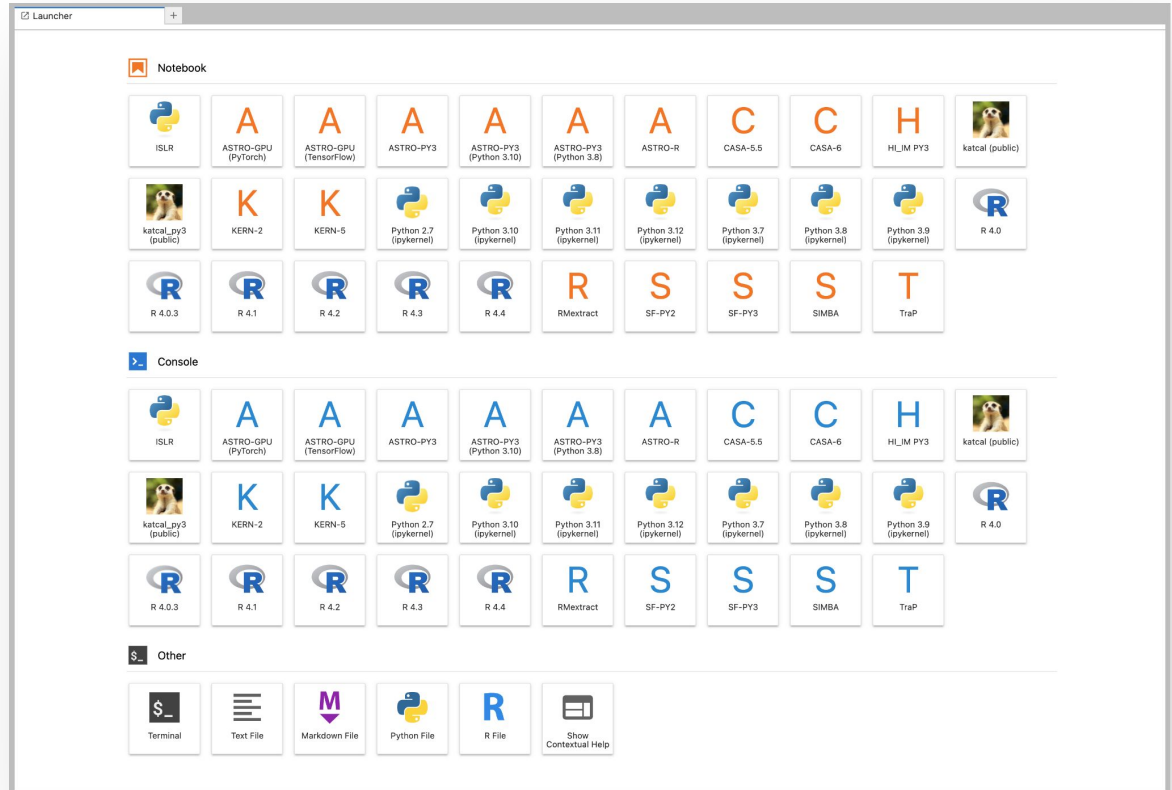
Select a job profile:

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

ilifu