ilifu Research Cloud services

Jeremy Smith – ilifu System Administrator/Technical Support

University of Cape Town

04 October 2019

User Documentation

https://docs.ilifu.ac.za

- Getting Started
- Astronomy
- Bioinformatics
- Advanced topics
- Regularly updated

ilifu Research Cloud

Cloud-based infrastructure for data-intensive research

- Data management, storage, transfer
- Compute environment – flexible, cloud based computing
- Support variety of different scientific projects and requirements

Computing environment

ILIFU Cloud

Compute

Shared Storage

Ceph
BeeGFS

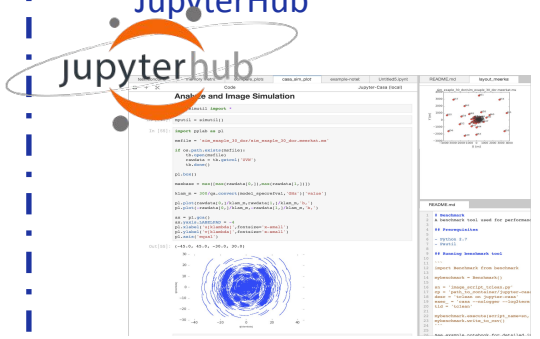
Virtual Networking

Cluster / Job Scheduler

Clusters

Partitions

JupyterHub



Singularity

…

Containers

Virtual Machines

VM VM VM

- services
- single use nodes
- etc.

SLURM workload manager

Structure:

- Head node (job submission & management)
- Partitions (organisation)
- Compute nodes (processing)

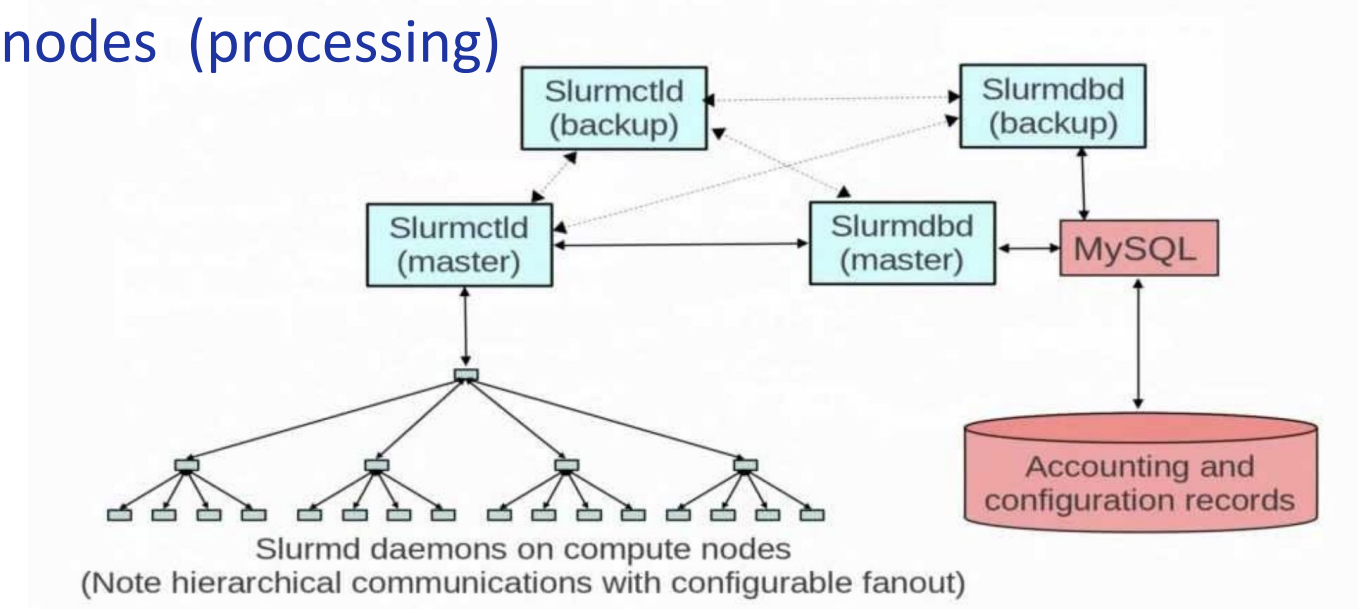
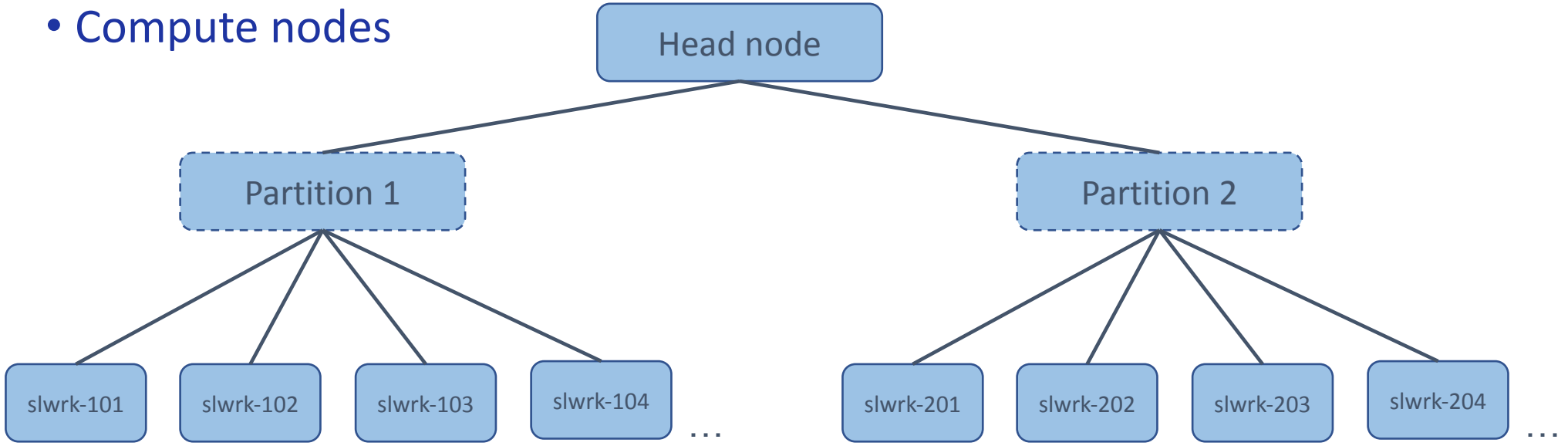


Image: https://slurm.schedmd.com

SLURM workload manager

Structure:

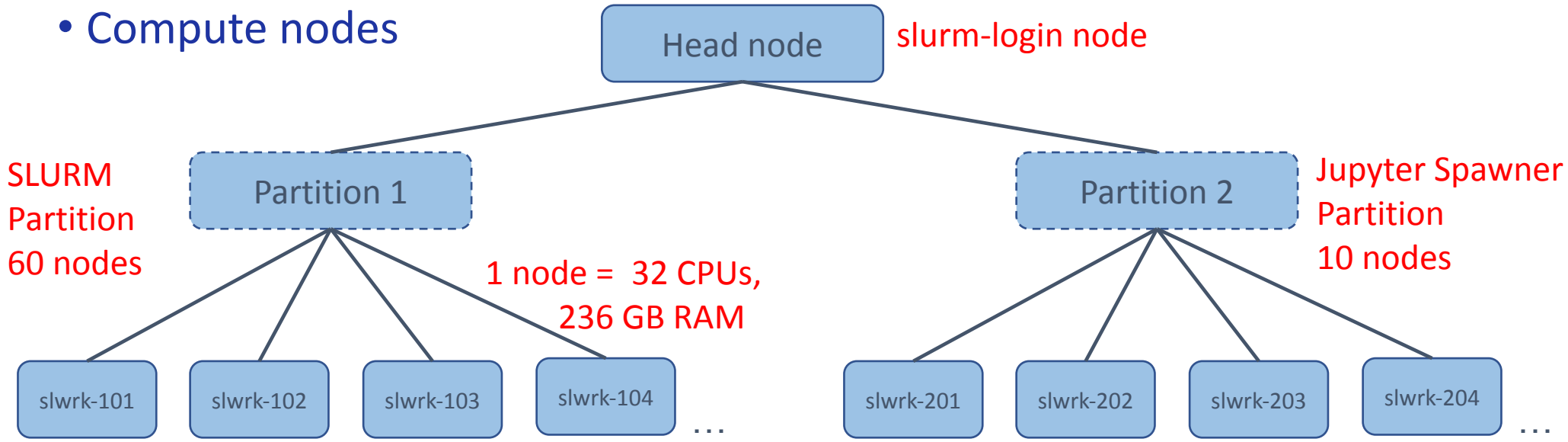
- Head node
- Partitions
- Compute nodes



SLURM workload manager

Structure:

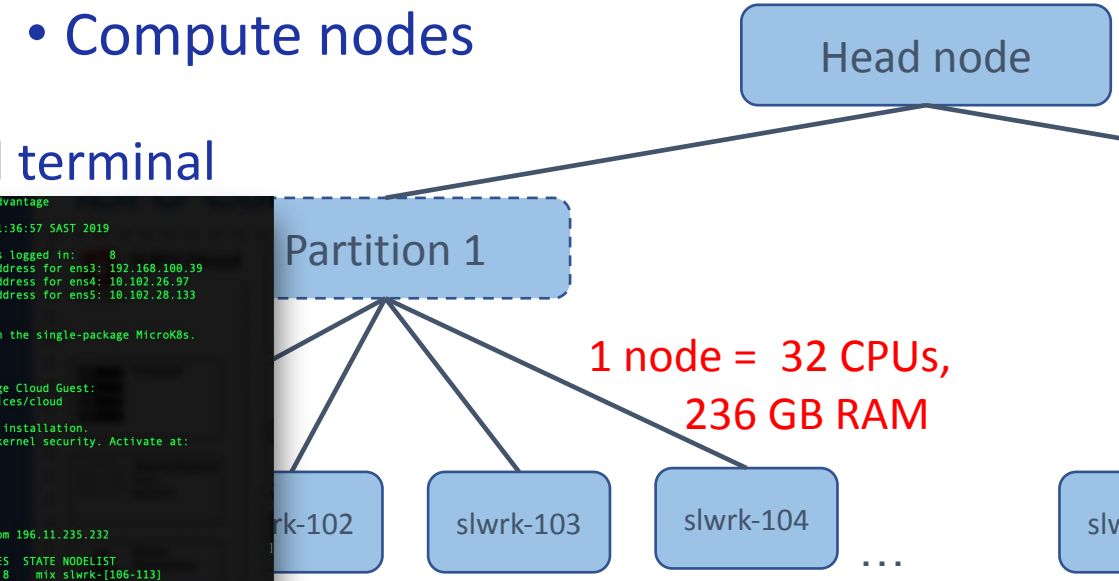
- Head node
- Partitions
- Compute nodes



SLURM workload manager

Structure:

- Head node
- Partitions
- Compute nodes

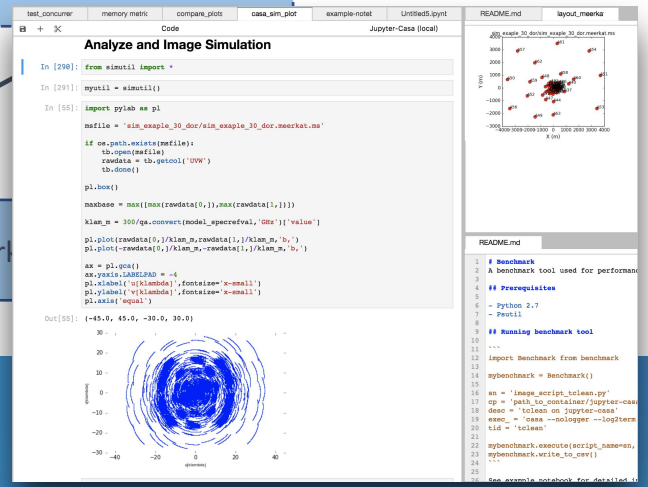


ssh – shell terminal

```
 * 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 NODELIST
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]
JupyerSpawnerONLY     up   infinite      4    mix slwrk-[201-202,205,209]
JupyerSpawnerONLY     up   infinite      4  alloc slwrk-[206-208,210]
JupyerSpawnerONLY     up   infinite      2   idle slwrk-[203-204]
jeremy@slurm-login:~$ sbatch compute_job.sh
```

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

Jupyterlab



Access

- Authentication
  - SSH key pairs – terminal (bash shell)
  - Password – Jupyterlab, CARTA
- If you lose your ssh key
  - New personal computer
  - Formating
  - https://docs.ilifu.ac.za/#/getting_started/ssh

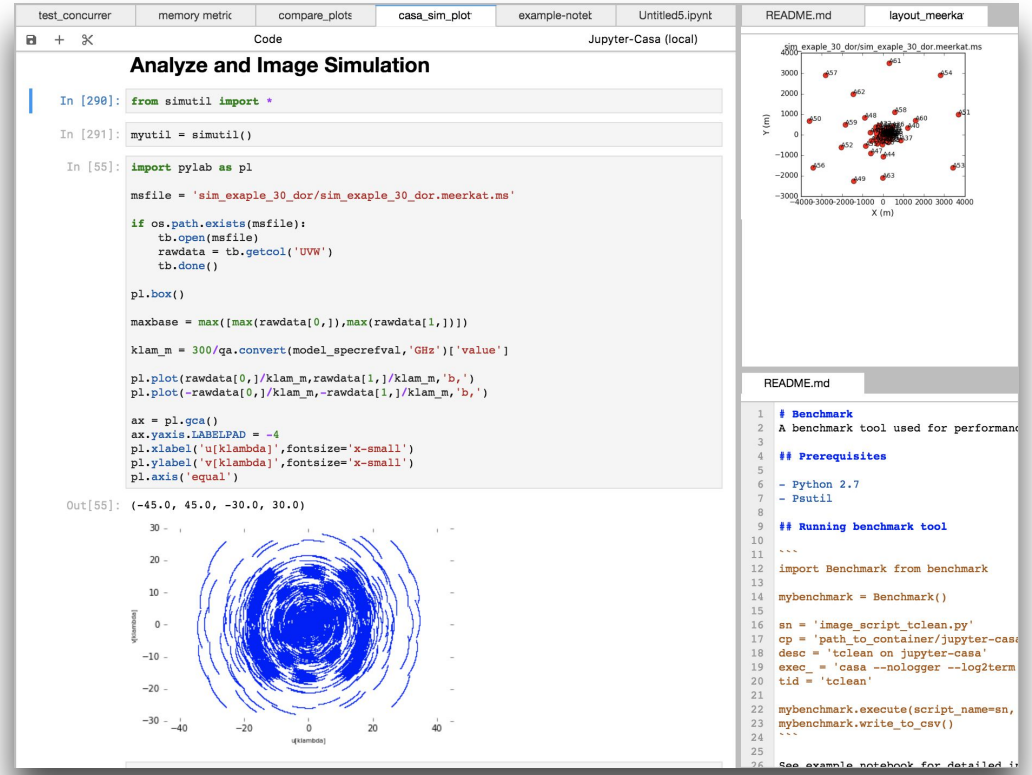
Interfaces

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 NODELIST
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]
JupyerSpawnerONLY    up   infinite      4    mix slwrk-[201-202,205,209]
JupyerSpawnerONLY    up   infinite      4  alloc slwrk-[206-208,210]
JupyerSpawnerONLY    up   infinite      2   idle slwrk-[203-204]
jeremy@slurm-login:~$ sbatch compute_job.sh
```

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

Jupyterlab



https://jupyter.ilifu.ac.za

Directory Structure

- Separated by group: IDIA, CBIO, SANBI, etc
- Common areas:
  - /users - only 20 TB shared among all users, for scripts and small files – **don't place data here**
  - /scratch/users - storage space for processing data, only temporary storage , ie during processing
    - currently 100 TB, will expand to 300 TB soon
- IDIA specific:
  - /idia the base directory for all IDIA related projects

Directory Structure

- IDIA structure:
- /idia/users - user's directory for storing long-term data and data products that are not project specific
- /idia/projects - project specific directories. These directories are for sharing data and resources within project groups. Project archive data, data products, intermediate data and project specific resources, such as script or software containers, are stored here. Raw data associated with a project will also be available from the project folder. Raw data folders should always be read-only.
- /idia/software - software containers and the IDIA Pipelines software is stored here
- /idia/data/public - public data and data products available to all users are stored here.

Storage infrastructure

BeeGFS

/users

/scratch

- Small (100 TB currently)
- Fast
- Ideal for processing
- ± 15% greater speeds

CephFS

/idia

/cbio

/sanbi

- Large (1.5 PB +)
- High read speeds
- Ideal for storage
- Still being optimized

Singularity containers

- https://docs.ilifu.ac.za/#/getting_started/container_environments
- Encapsulated software environments
- A software stack that contains everything required to run an application/workflow, including files, environmental variables, libraries and dependencies
- Containers accessible across platforms and services, allowing sharing of application environments

Singularity containers

Singularity containers

Supported Containers:

- CASA
- KERN suite
- Source Finding
- Python 2.7, Python 3.6, R
- Project containers:
  - MeerLICHT,
  - LADUMA
  - HI intensity map

Singularity containers

Open container as an interactive shell:

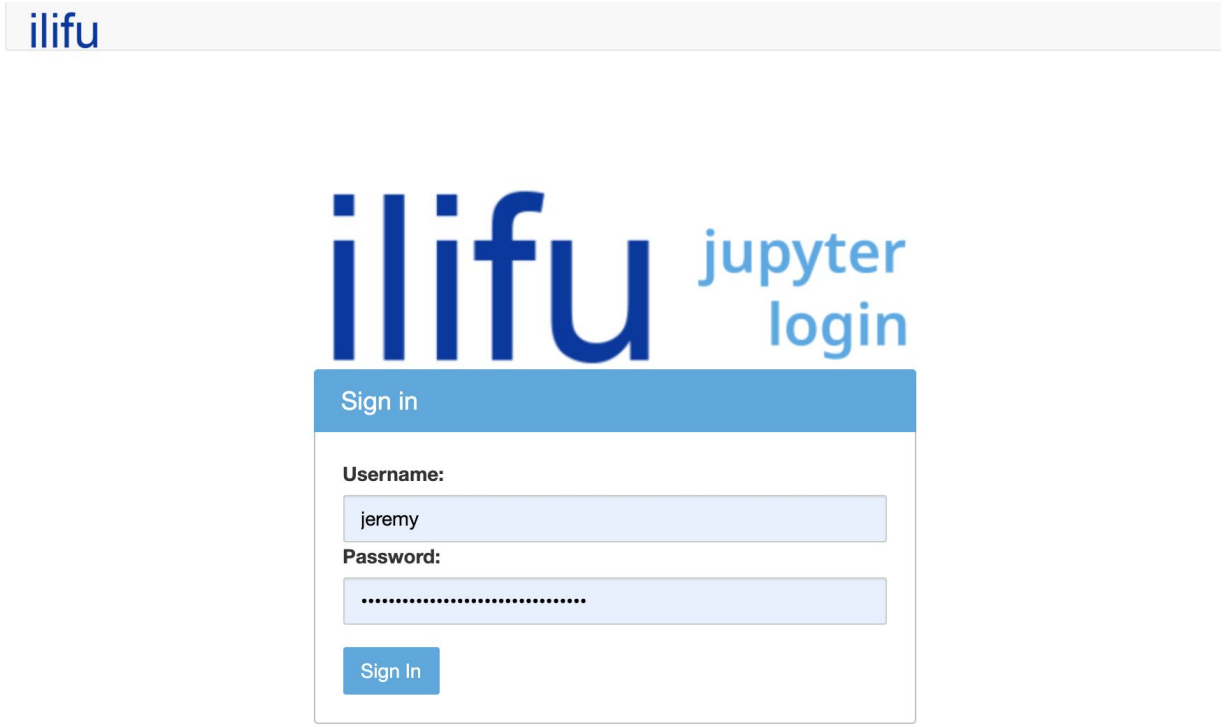
```
$ singularity shell /idia/software/containers/SF-PY3-bionic.simg
```

Run a script/workflow using a container environment:

```
$ singularity exec /idia/software/containers/casa-stable.img casa
-c myscript.py
```

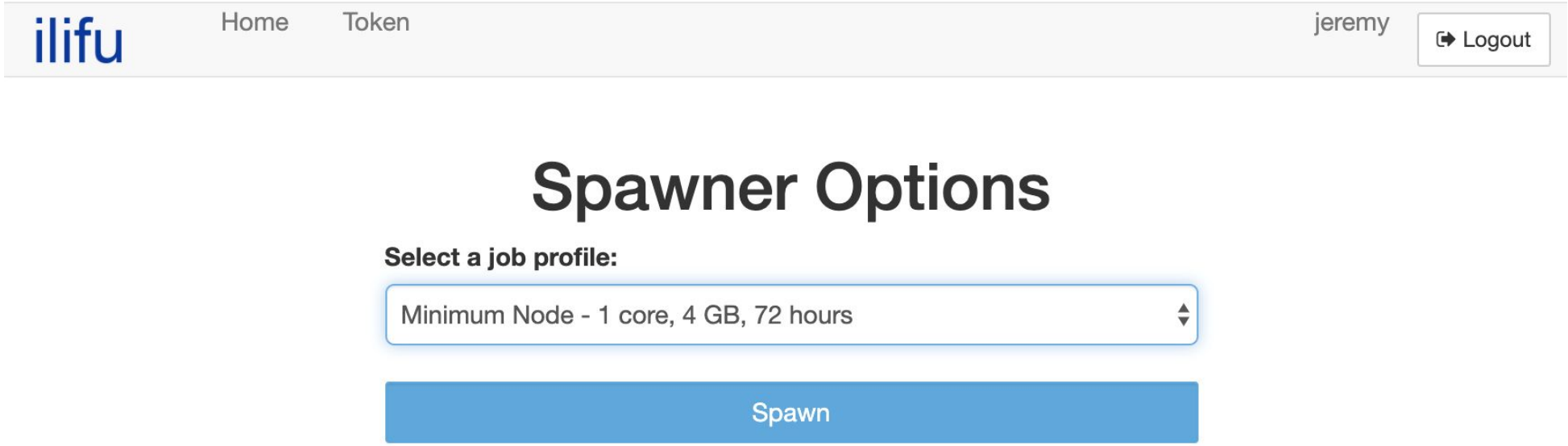
Note: singularity not available from the SLURM login node

Jupyterlab



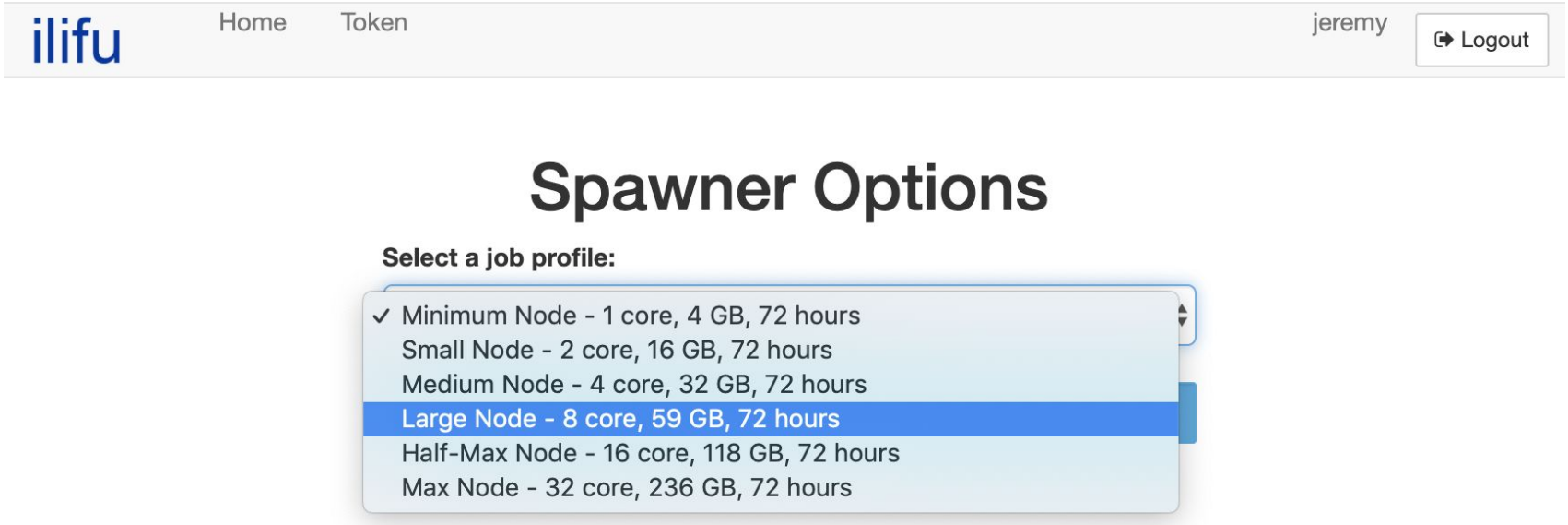
- https://jupyter.ilifu.ac.za

Jupyterlab



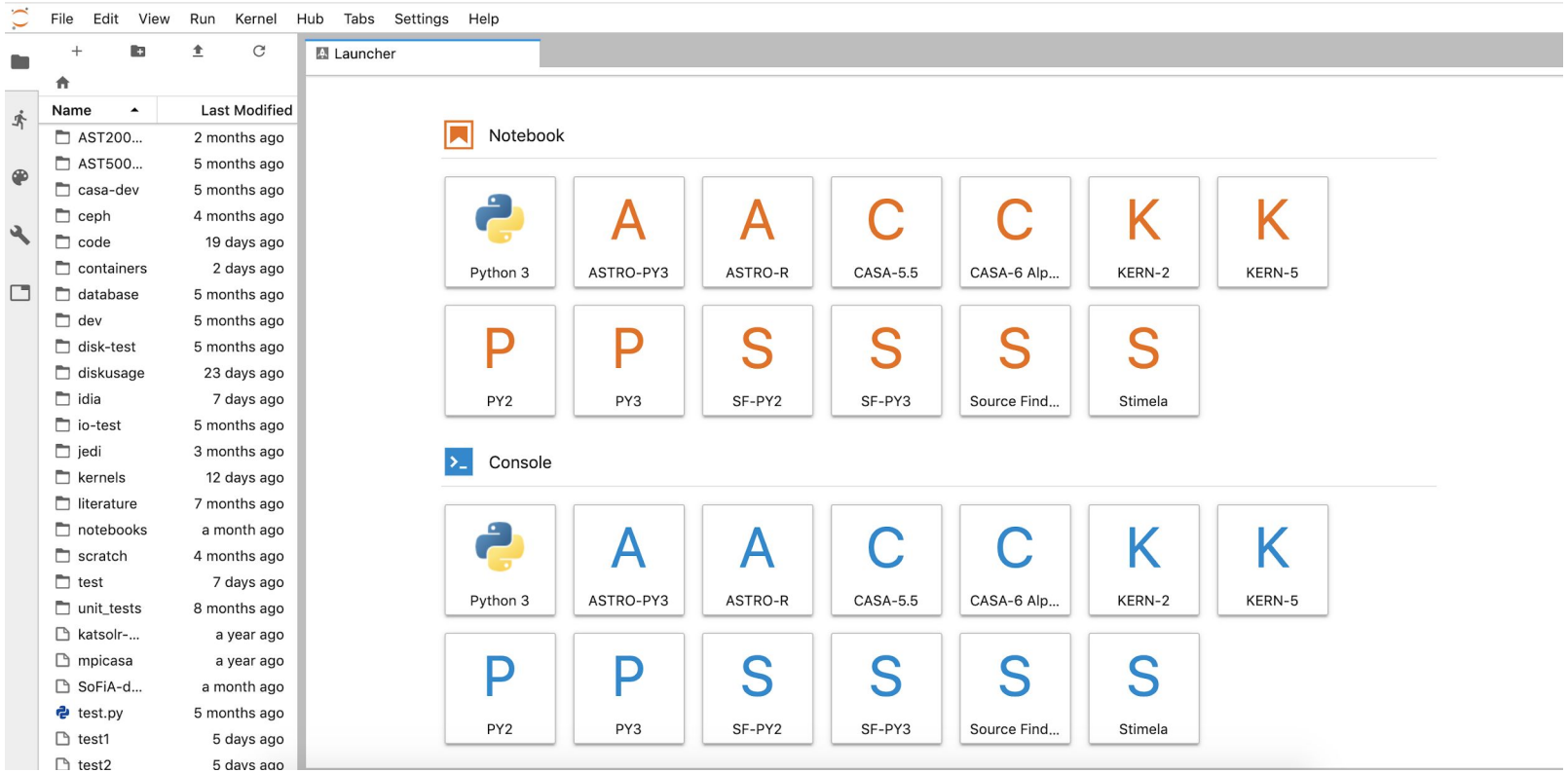
- https://jupyter.ilifu.ac.za

Jupyterlab



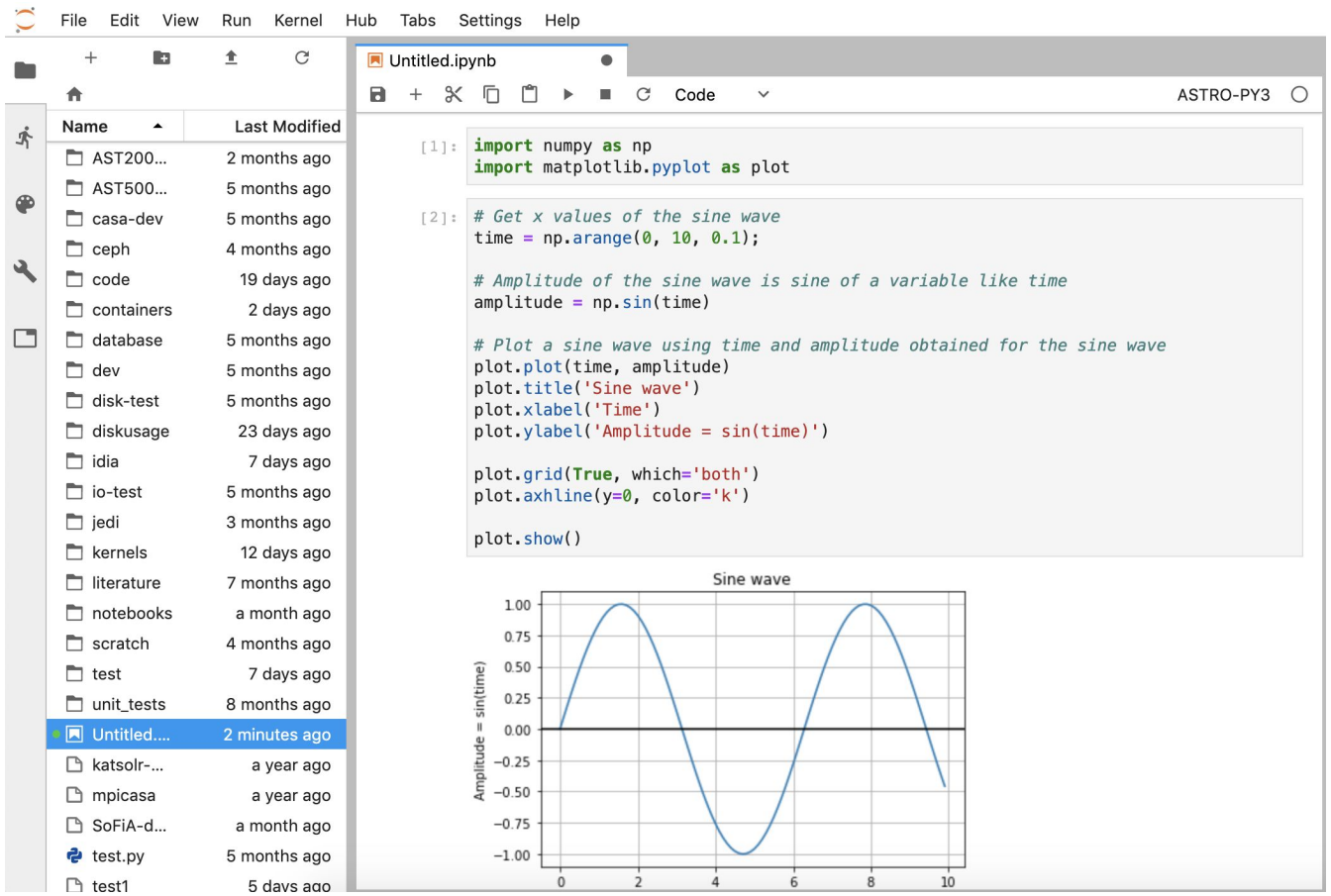
- https://jupyter.ilifu.ac.za

Jupyterlab



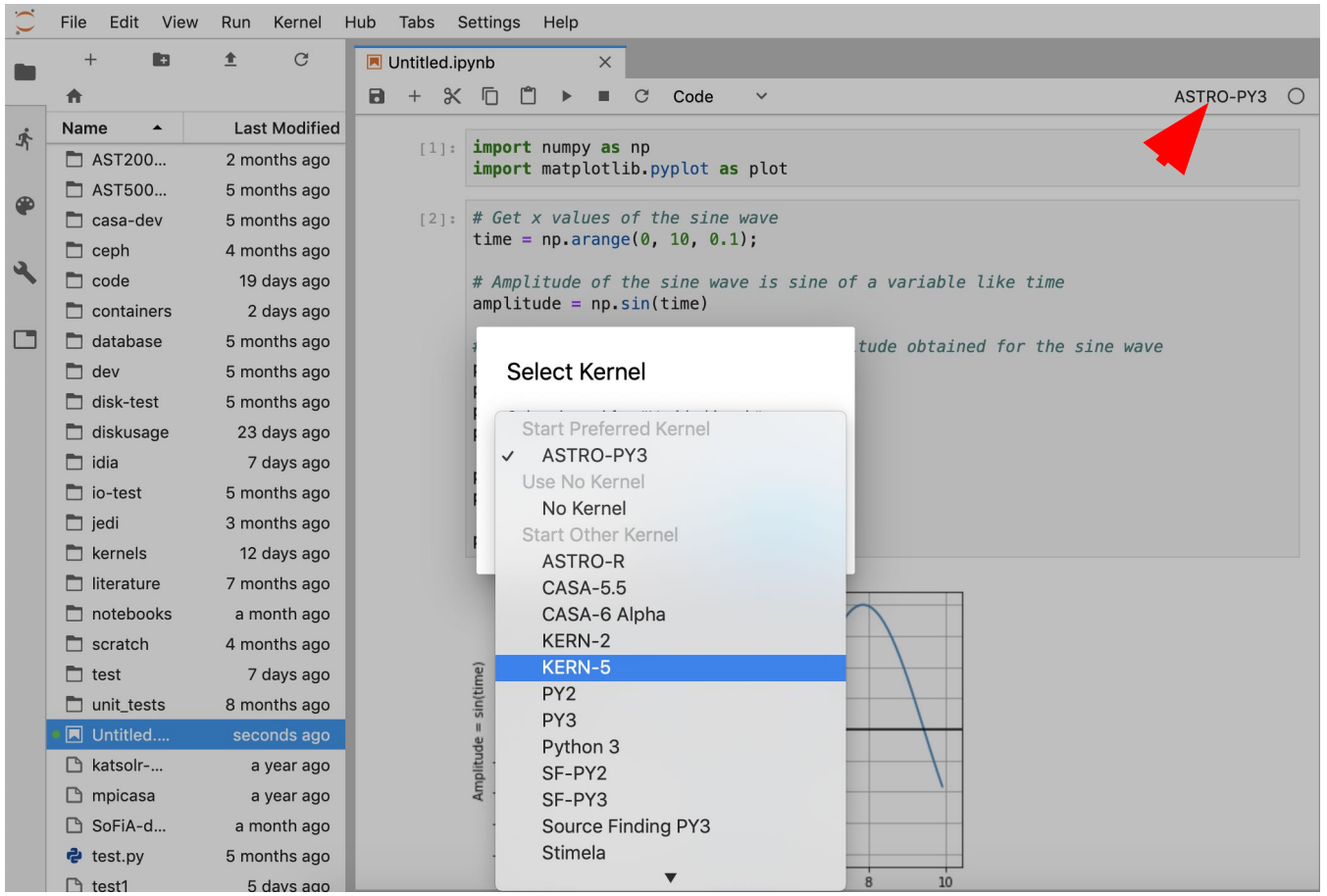
- https://jupyter.ilifu.ac.za

Jupyterlab



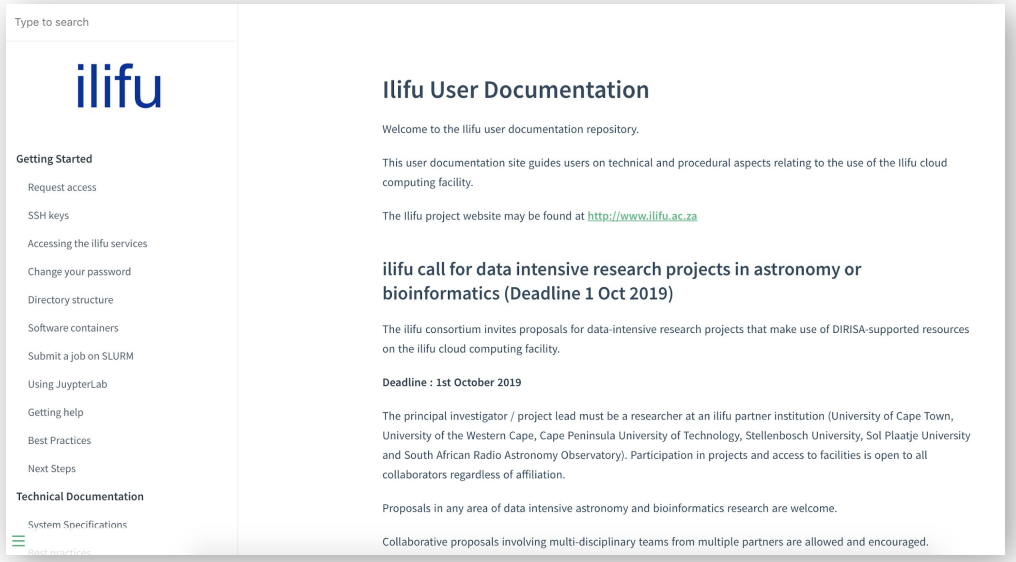
- https://jupyter.ilifu.ac.za

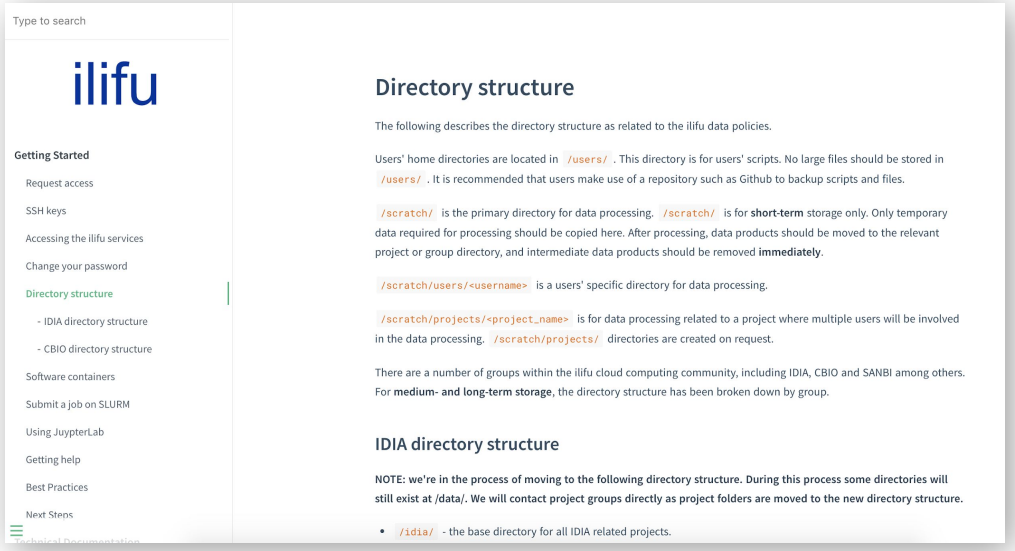
Jupyterlab



- https://jupyter.ilifu.ac.za

User Documentation – docs.ilifu.ac.za





ilifu support team: support@ilifu.ac.za