

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>

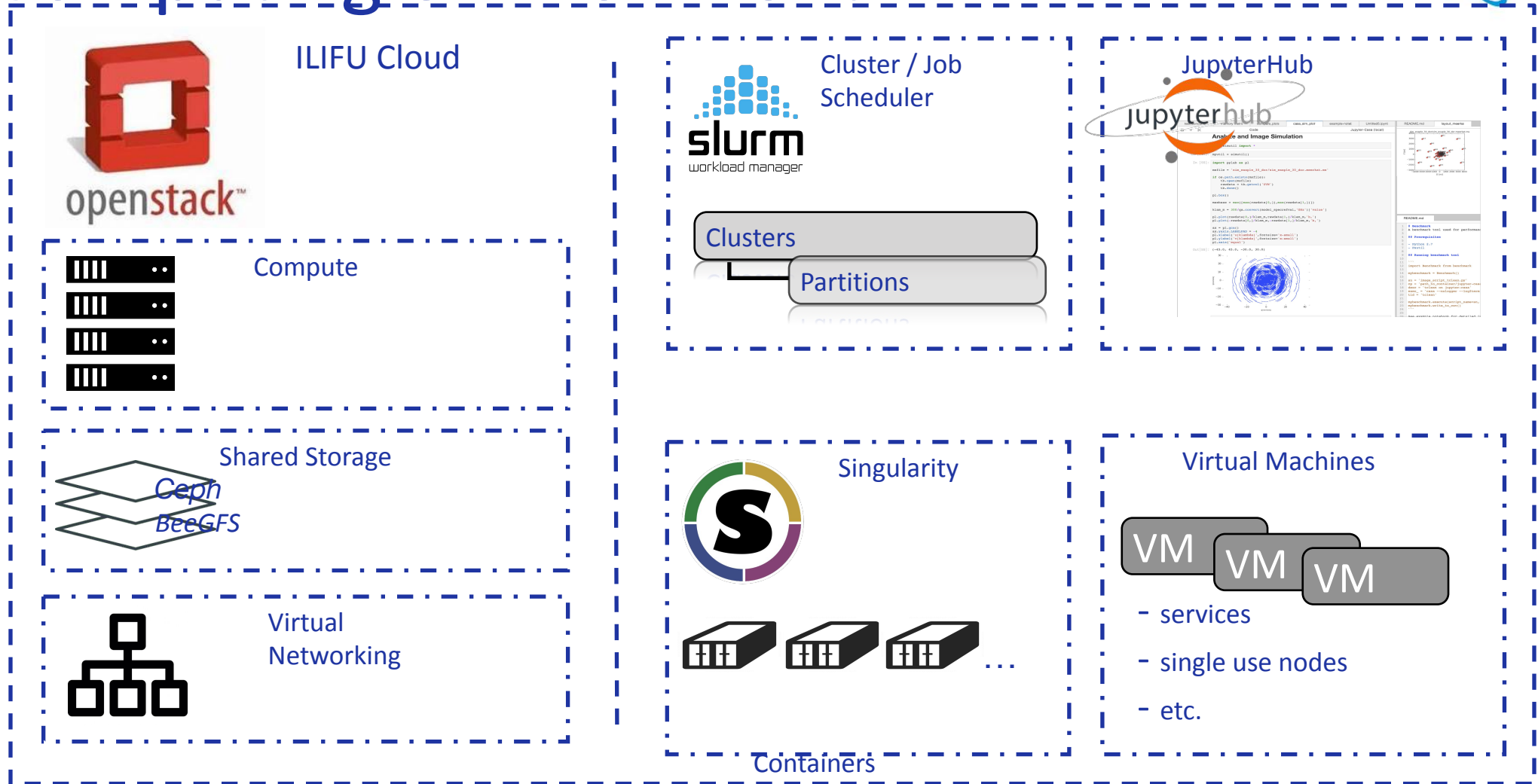
- Getting Started
- Astronomy
- Bioinformatics
- Advanced topics
- Regularly updated

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

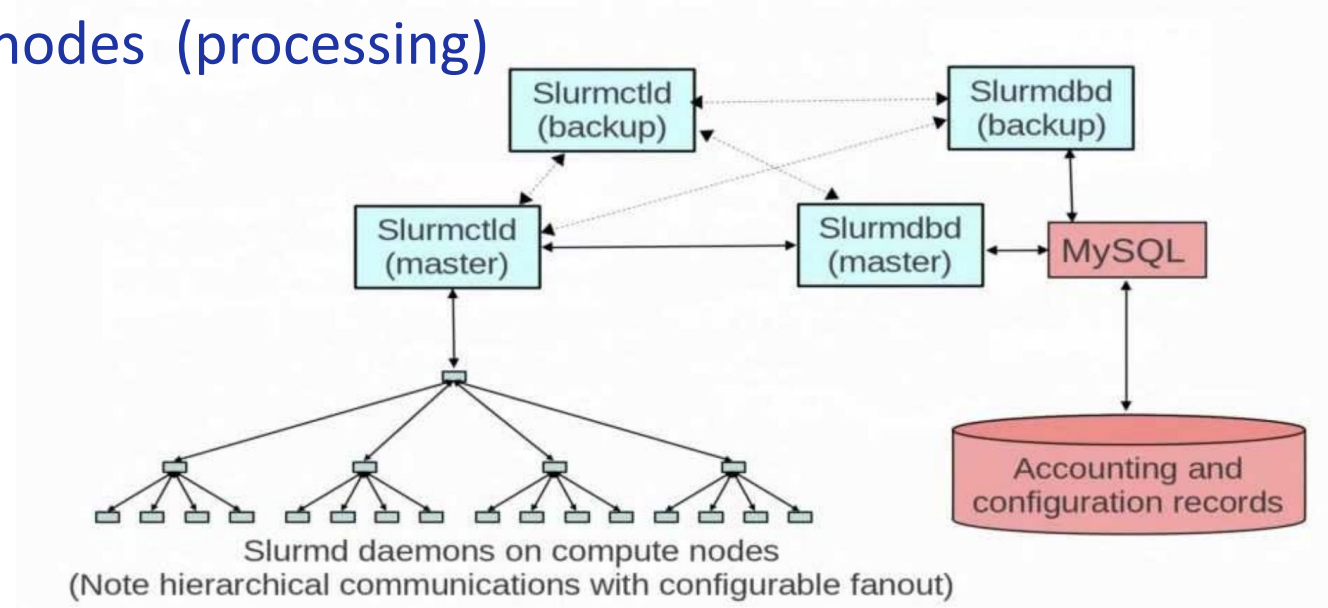
Computing environment



SLURM workload manager

Structure:

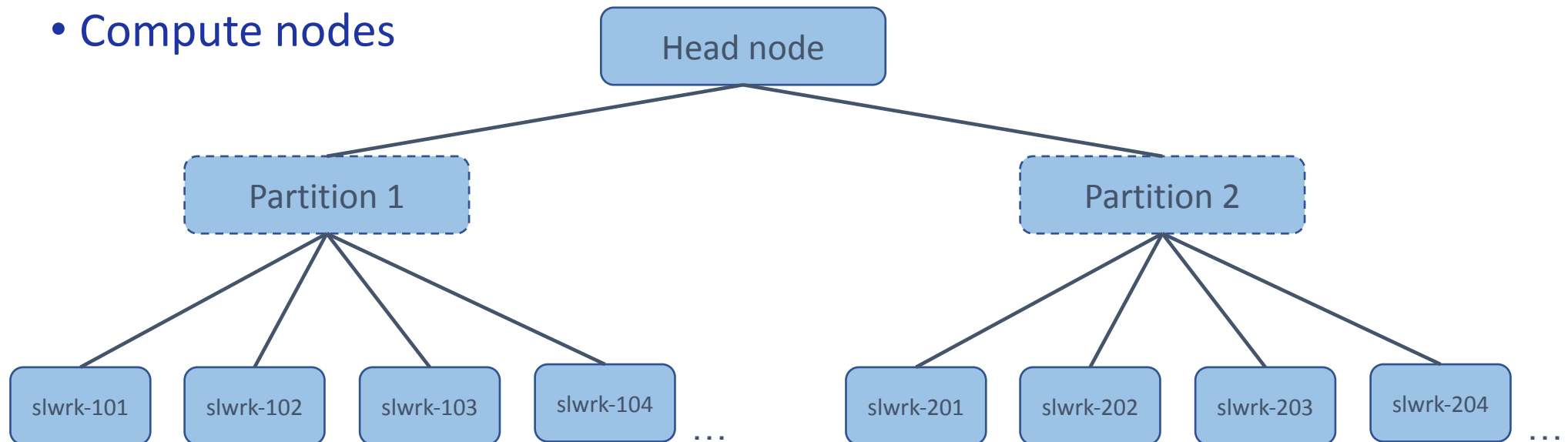
- Head node (job submission & management)
- Partitions (organisation)
- Compute nodes (processing)



SLURM workload manager

Structure:

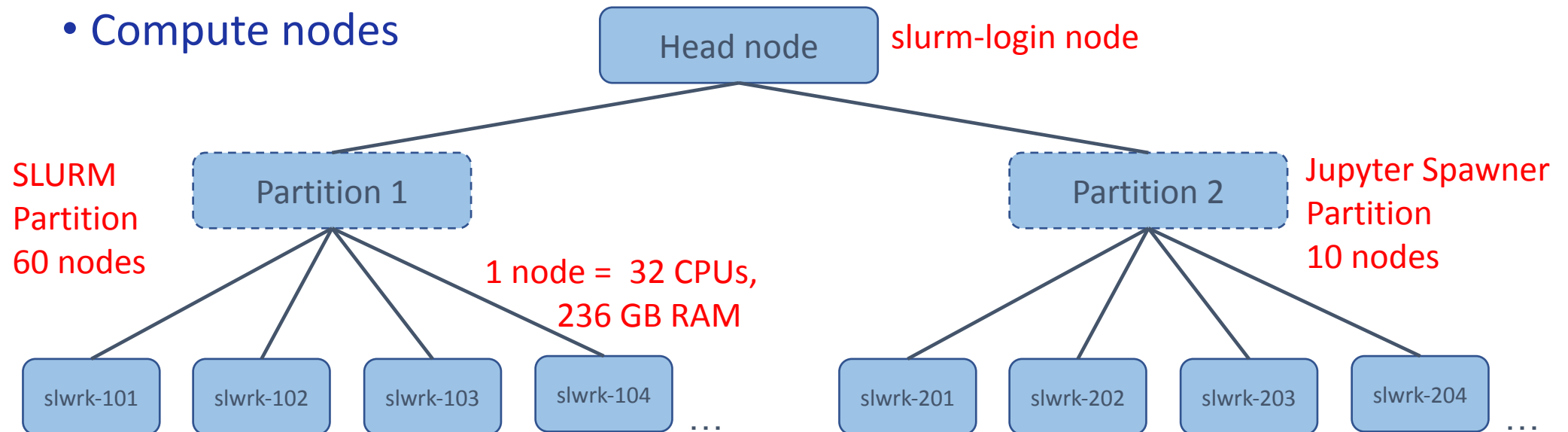
- Head node
- Partitions
- Compute nodes



SLURM workload manager

Structure:

- Head node
- Partitions
- Compute nodes



Structure:

- Head node
- Partitions
- Compute nodes

```
* 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 21.5GB  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

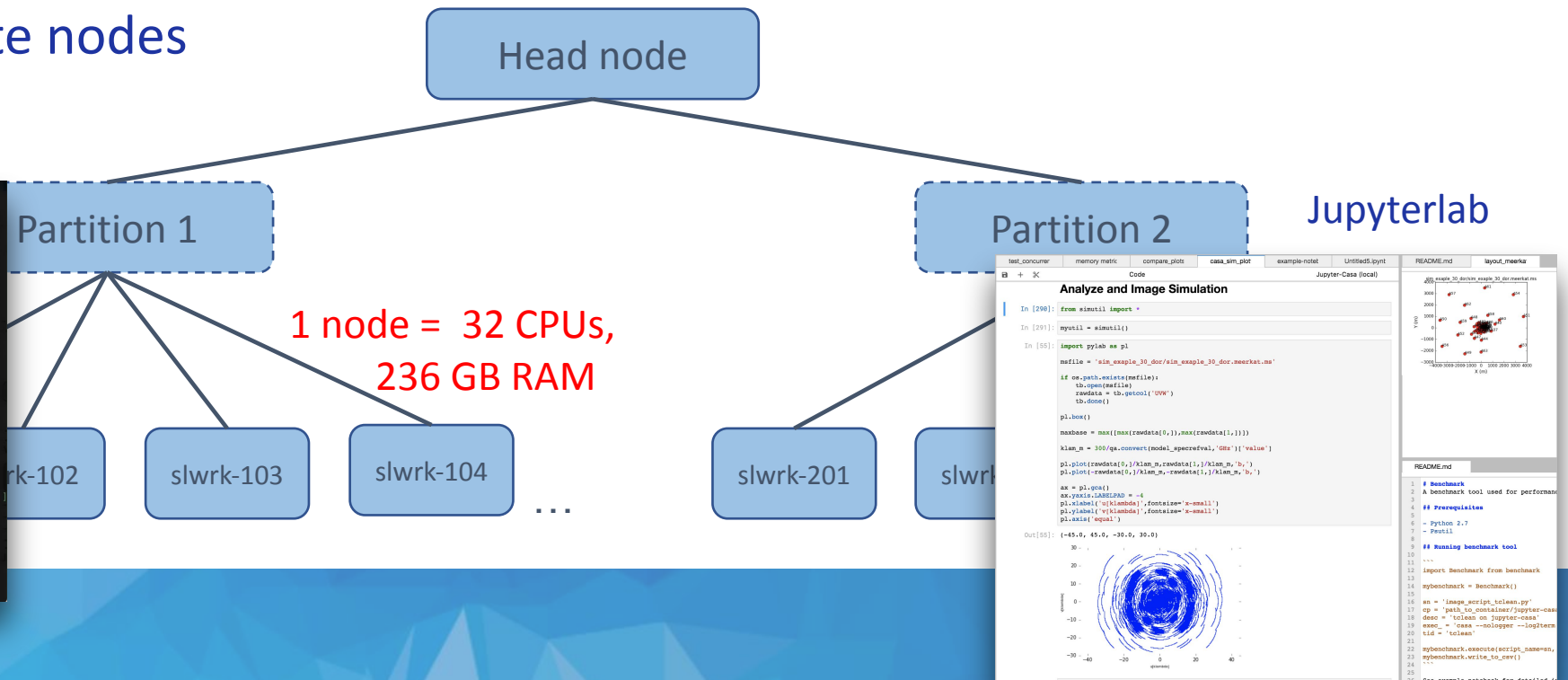
* Canonical Livepatch is available for installation.
- Reduce system reboots and improve kernel security. Activate at:
https://ubuntu.com/livepatch

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:~$ sudo

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]
JupyterSpannerONLY up infinite  4      mix slwrk-[201-202,205,209]
JupyterSpannerONLY up infinite  4      alloc slwrk-[206-208,210]
JupyterSpannerONLY up infinite  2      idle  slwrk-[203-204]

jeremy@slurm-login:~$ sbatch compute job.sh
```



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


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

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

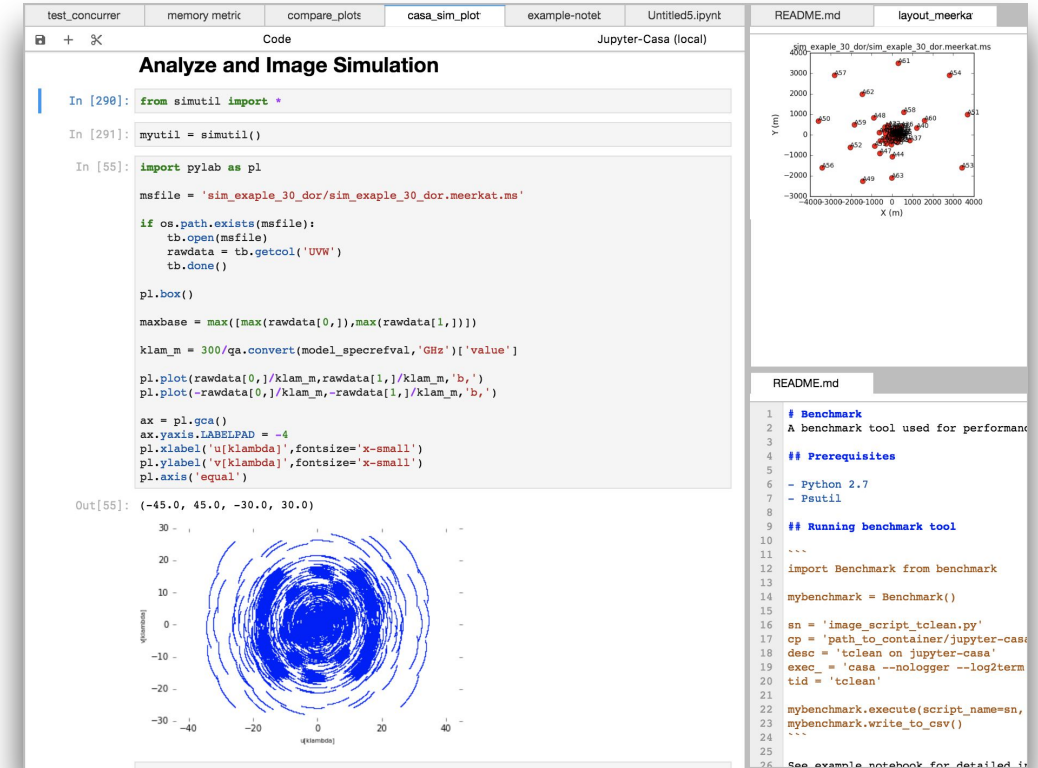
* Canonical Livepatch is available for installation.
- Reduce system reboots and improve kernel security. Activate at:
https://ubuntu.com/livepatch

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]
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

Jupyterlab



The screenshot shows the JupyterLab interface with a code editor on the left and a terminal on the right. The code editor displays a Python script for analyzing and simulating image data. The terminal shows the output of the script, which includes a plot of the data.

Code Editor:

```
Analyze and Image Simulation

In [290]: from simutil import *
In [291]: myutil = simutil()
In [55]: import pylab as pl

msfile = 'sim_exaple_30_dor/sim_exaple_30_dor.meerkat.ms'

if os.path.exists(msfile):
    tb.open(msfile)
    rawdata = tb.getcool('UVW')
    tb.done()

pl.box()

maxbase = max([max(rawdata[0,]),max(rawdata[1,])])

kiam_m = 300/qa.convert(model_specrefval,'GHz')['value']

pl.plot(rawdata[0,]/kiam_m,rawdata[1,]/kiam_m,'b,')
pl.plot(-rawdata[0,]/kiam_m,-rawdata[1,]/kiam_m,'b,')

ax = pl.gca()
ax.yaxis.LABELPAD = -4
pl.xlabel('u[klambda]',fontsize='x-small')
pl.ylabel('v[klambda]',fontsize='x-small')
pl.axis('equal')
```

Terminal Output:

```
Out[55]: (-45.0, 45.0, -30.0, 30.0)
```

Plot:

The plot shows a 2D distribution of data points in the u-k and v-k plane. The axes are labeled 'u[klambda]' and 'v[klambda]'. The data points are concentrated in a central region, forming a circular pattern.

Terminal:

```
1 # Benchmark
2 A benchmark tool used for performance
3
4 ## Prerequisites
5
6 - Python 2.7
7 - Psutil
8
9 ## Running benchmark tool
10
11 ...
12 import Benchmark from benchmark
13
14 mybenchmark = Benchmark()
15
16 sn = 'image_script_tclean.py'
17 cp = 'path_to_container/jupyter-casa'
18 desc = 'tclean on jupyter-casa'
19 exec_ = 'casa --nologger --log2term
20 tid = 'tclean'
21
22 mybenchmark.execute(script_name=sn,
23 mybenchmark.write_to_csv()
24
25
26 See example notebook for detailed i
```

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
- \pm 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

ilifu

Jupyterlab



ilifu

ilifu

jupyter
login

Sign in

Username:

jeremy

Password:

.....

Sign In

• <https://jupyter.ilifu.ac.za>

Spawner Options

Select a job profile:

Minimum Node - 1 core, 4 GB, 72 hours

Spawn

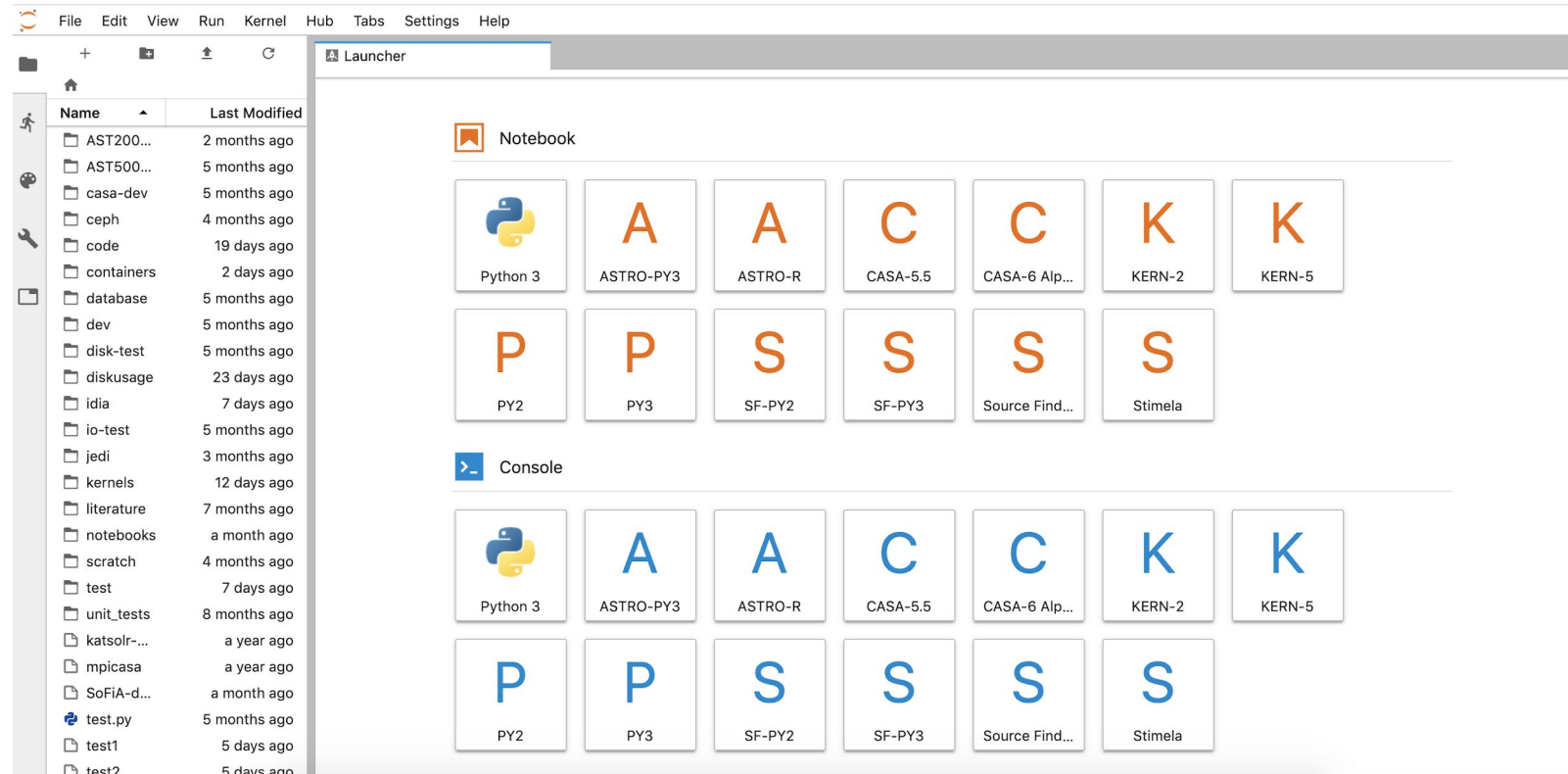
• <https://jupyter.ilifu.ac.za>

Spawner Options

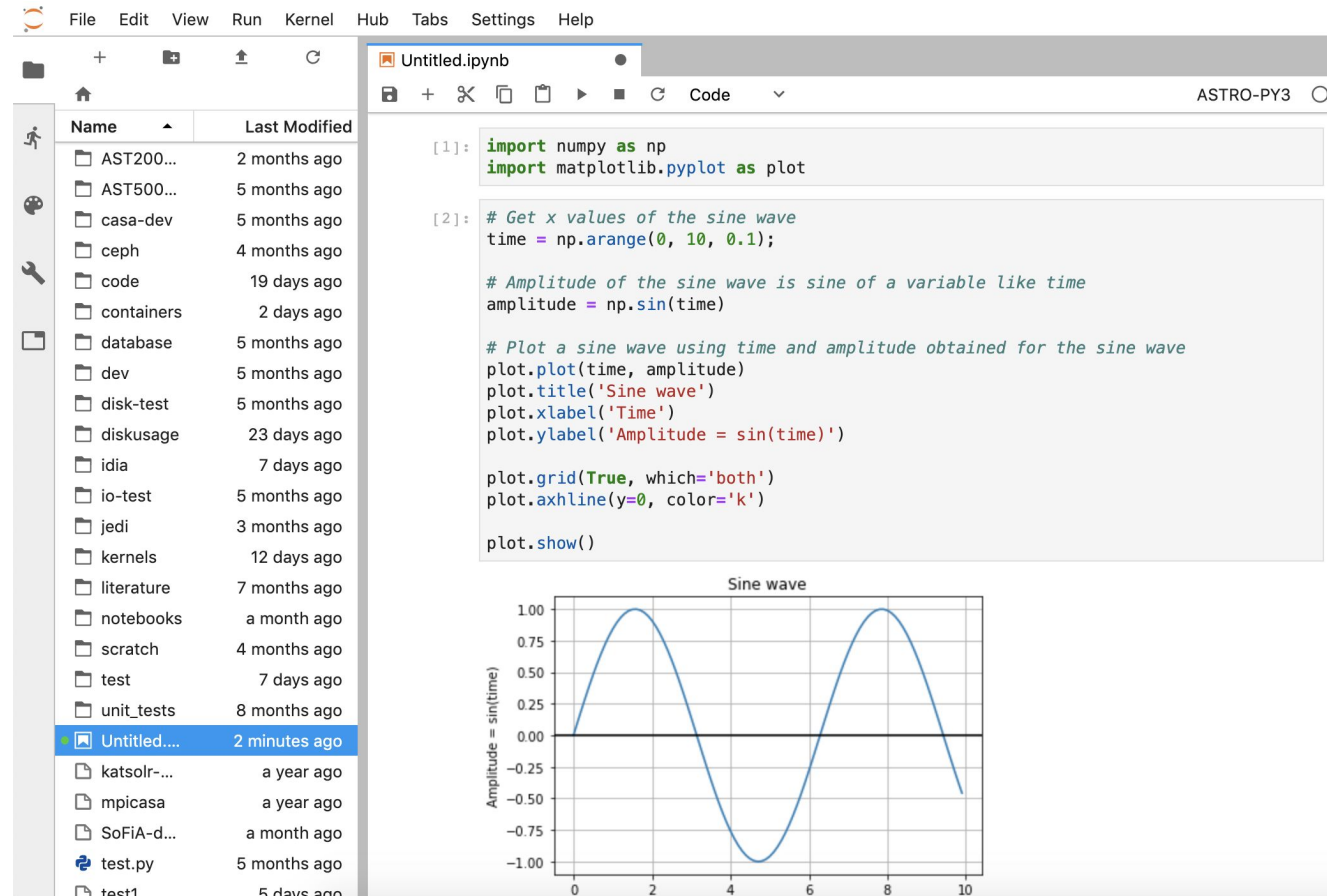
Select a job profile:

- ✓ Minimum Node - 1 core, 4 GB, 72 hours
- Small Node - 2 core, 16 GB, 72 hours
- Medium Node - 4 core, 32 GB, 72 hours
- Large Node - 8 core, 59 GB, 72 hours**
- Half-Max Node - 16 core, 118 GB, 72 hours
- Max Node - 32 core, 236 GB, 72 hours

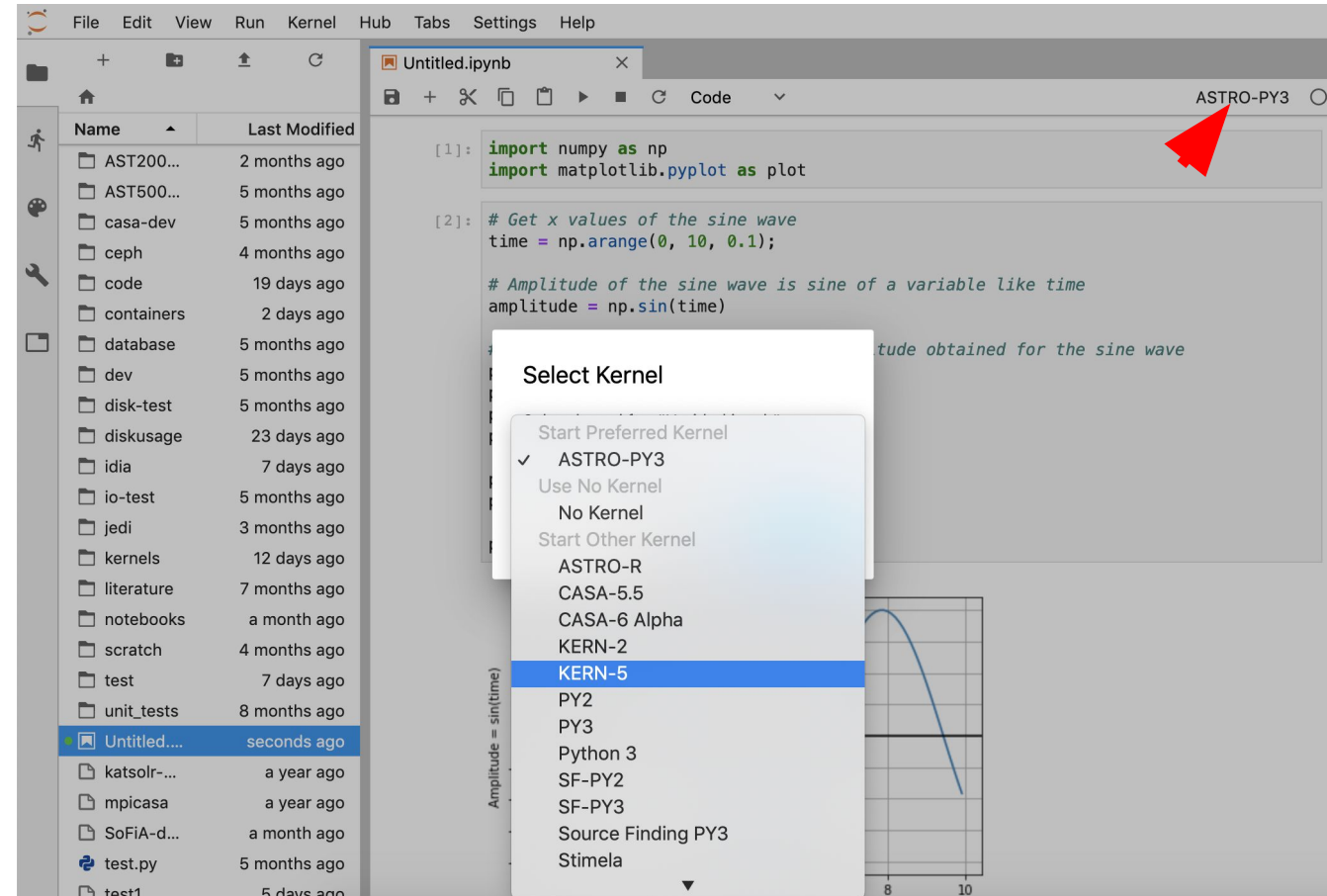
• <https://jupyter.ilifu.ac.za>



• <https://jupyter.ilifu.ac.za>



• <https://jupyter.ilifu.ac.za>



• <https://jupyter.ilifu.ac.za>

User Documentation – docs.ilifu.ac.za

Type to search

ilifu

Getting Started

Request access

SSH keys

Accessing the ilifu services

Change your password

Directory structure

Software containers

Submit a job on SLURM

Using JupyterLab

Getting help

Best Practices

Next Steps

Technical Documentation

System Specifications

Best practices

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>

ilifu call for data intensive research projects in astronomy or bioinformatics (Deadline 1 Oct 2019)

The ilifu consortium invites proposals for data-intensive research projects that make use of DIRISA-supported resources on the ilifu cloud computing facility.

Deadline : 1st October 2019

The principal investigator / project lead must be a researcher at an ilifu partner institution (University of Cape Town, University of the Western Cape, Cape Peninsula University of Technology, Stellenbosch University, Sol Plaatje University and South African Radio Astronomy Observatory). Participation in projects and access to facilities is open to all collaborators regardless of affiliation.

Proposals in any area of data intensive astronomy and bioinformatics research are welcome.

Collaborative proposals involving multi-disciplinary teams from multiple partners are allowed and encouraged.

Type to search

ilifu

Getting Started

Request access

SSH keys

Accessing the ilifu services

Change your password

Directory structure

Software containers

Submit a job on SLURM

Using JupyterLab

Getting help

Best Practices

Next Steps

Technical Documentation

Directory structure

The following describes the directory structure as related to the ilifu data policies.

Users' home directories are located in `/users/`. This directory is for users' scripts. No large files should be stored in `/users/`. It is recommended that users make use of a repository such as Github to backup scripts and files.

`/scratch/` is the primary directory for data processing. `/scratch/` is for short-term storage only. Only temporary data required for processing should be copied here. After processing, data products should be moved to the relevant project or group directory, and intermediate data products should be removed immediately.

`/scratch/users/<username>` is a users' specific directory for data processing.

`/scratch/projects/<project_name>` is for data processing related to a project where multiple users will be involved in the data processing. `/scratch/projects/` directories are created on request.

There are a number of groups within the ilifu cloud computing community, including IDiA, CBIO and SANBI among others. For medium- and long-term storage, the directory structure has been broken down by group.

IDiA directory structure

NOTE: we're in the process of moving to the following directory structure. During this process some directories will still exist at `/data/`. We will contact project groups directly as project folders are moved to the new directory structure.

- `/idia/` - the base directory for all IDiA related projects.

ilifu support team: support@ilifu.ac.za