ilifu Online Training

Jeremy Smith

Ilifu User Training Workshop – Basic Training

30 March 2021

Topics

- Support channels
- Introduction to the ilifu research facility services
- Directory structure
- Software Environment
  - Singularity containers
  - Modules
- Using JupyterHub
- Introduction to SLURM
  - Submitting a job on SLURM
  - Interactive sessions on SLURM
- Best Practices

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
  http://www.ilifu.ac.za/il/accessing-facilities/training

ilifu Research Facility

Cloud-based infrastructure for data-intensive research

- Support variety of different scientific projects and requirements
- Data management: storage, transfer
- Flexible compute environment

Computing environment

- Cluster & Job Scheduler
- JupyterHub service – development environ.
- Containerised software environment
- Other services: data transfer, CARTA

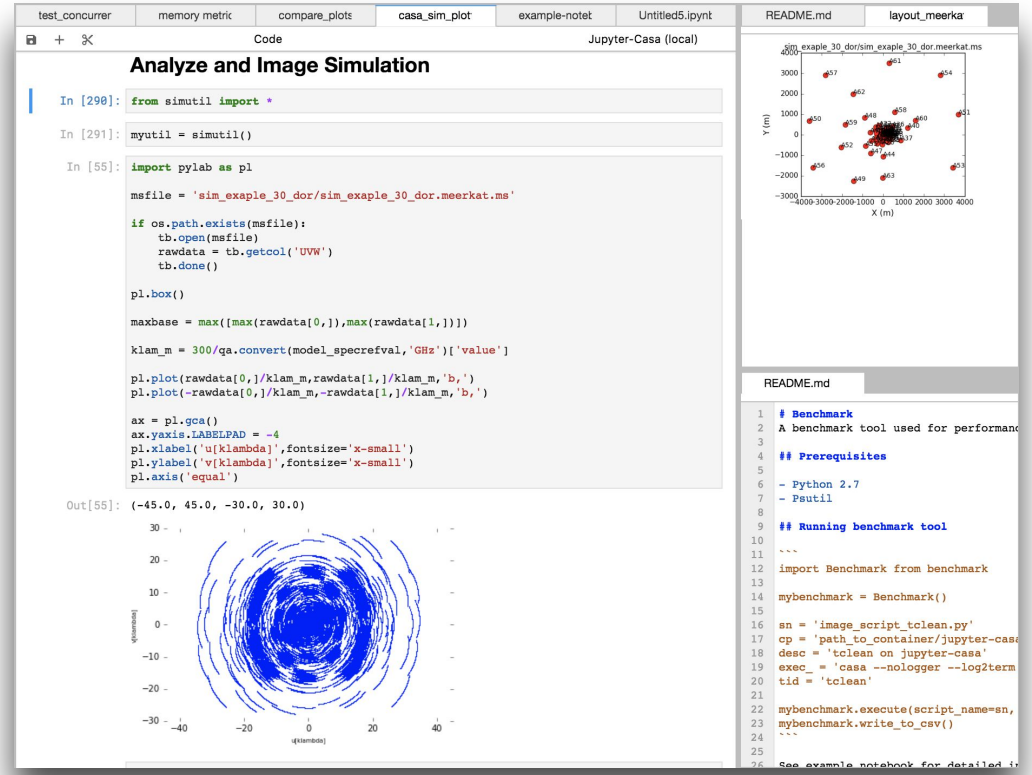
Computing environment - interface

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`

JupyterHub



https://jupyter.ilifu.ac.za

Directory Structure

- Common areas:
  - /users - only 40 TB 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.**
  - Scratch storage:
  - /scratch/users or /scratch3/users

    - storage space for processing, 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.

  - /software
- Remaining storage is separated by group:
  - IDIA, CBIO, Ilifu (DIRISA projects)

Directory Structure

IDIA structure:

- /idia/users - user's private workspace, may store data products that are not ready to move to shared project space (limited to 10TiB per user)
- /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
- /idia/software/containers

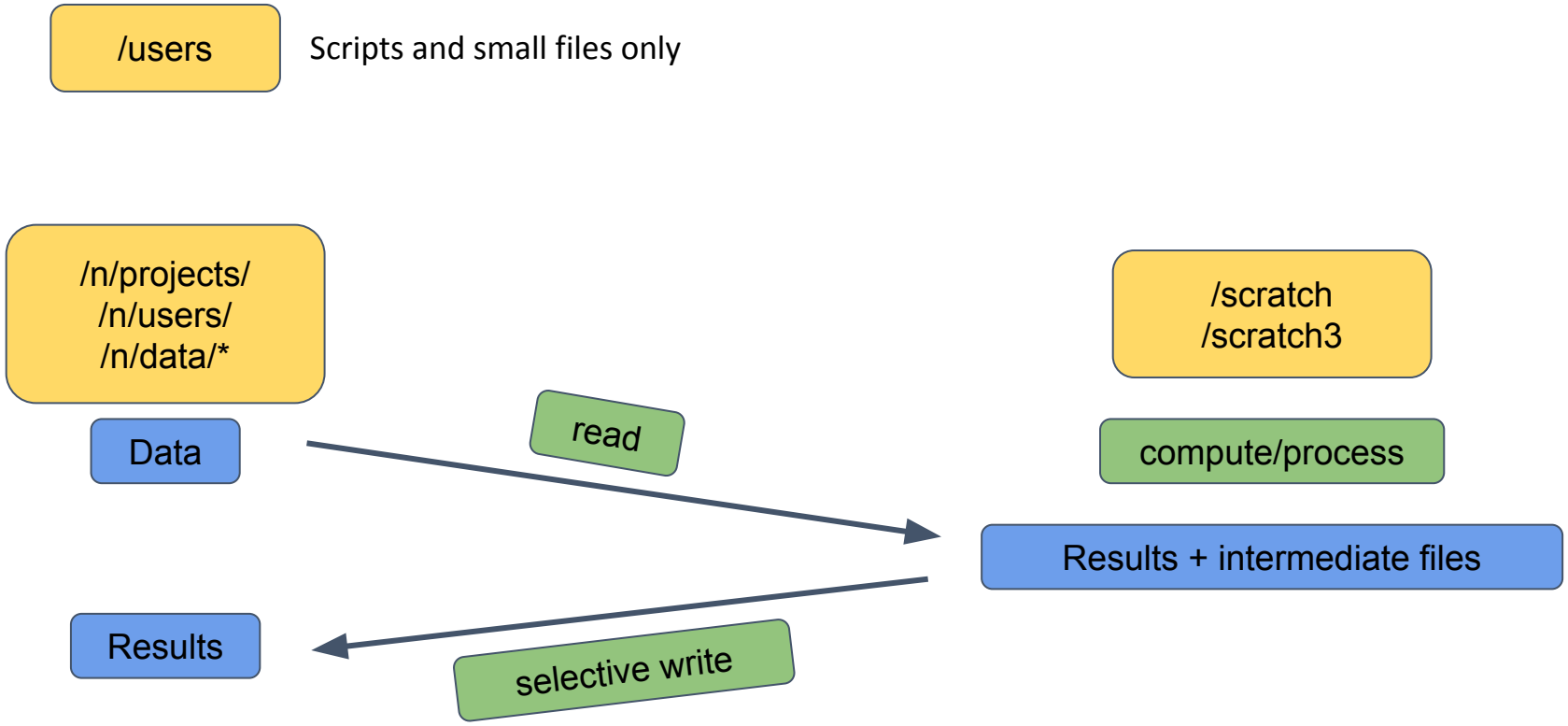
Directory Structure

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 - workflow



*/n/data generally read-only

Singularity containers

- 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

Supported Containers:

- CASA
- KERN suite
- Astronomy container (ASTRO-PY3.simg)
- Python 2.7, Python 3.6, R
- Project containers:
  - MeerLICHT,
  - Simba
  - HI intensity map

- /idia/software/containers; /cbio/soft; /ilifu/software/containers

Singularity containers

Open container as an interactive shell:

```
singularity shell /path/to/container
```

Example:

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

Run a script/workflow using a container environment:

```
singularity exec /path/to/container <software> <script/input_params>
```

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

Software environment - modules

- module avail

```
~$ module avail

---------------------------------- /software/modules/common ----------------------------------------------------------
   LAPACK/3.9.0                    anaconda3/login.old            drmaa/1.1.1                openBLAS/0.3.9        perlbrew/perlbrew
ruby/2.6.6
   R/RStudio1.2.5042-R4.0.0        anaconda3/login                homebrew/2.4.13            openmpi/2.1.6         python/2.7.18
   R/RStudio1.2.5042-R4.0.4        anaconda3/2020.07       (D)    java/jre-1.8.0_261         openmpi/3.1.6         python/3.7.7
   R/3.6.3                         cuda/10.0.130_410.48           java/openjdk-14.0.1 (D)    openmpi/4.0.3         python/3.8.2
   R/4.0.0                         cuda/10.1.243_418.87.00        julia/1.5.3                openmpi/4.0.5         python/3.8.3
   R/4.0.2                         cuda/10.2.89_440.33.01         maven/3.6.3                openmpi/4.1.0   (D)   python/3.8.6
   R/4.0.3                  (D)    cuda/11.0.2_450.51.05   (D)    mono/6.8.0.123             perl/5.33.0           python/3.9.0      (D)

---------------------------------- /software/modules/astro -----------------------------------------------------------
   casa/5.7.0    casa/5.7.2-4    casa/6.1.0-118-monolithic    casa/6.1.2.7-pipeline    casa/6.1.2.7-modular (D)    pybdsf/1.9.2

---------------------------------- /software/modules/bio -------------------------------------------------------------
   bcbio/1.2.3         canvas/1.40.0.1613    genomestrip/2.00.1958    plink/2.00a2.3    prsice-2/2.3.1d    treePL/homebrew
vep/singularity
   bcftools/1.10.2     gemini/gemini         htslib/1.10.2            popgen/0.1        samtools/1.10      vcftools/0.1.16    vep/101.0
(D)

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

  Where:
   D:  Default Module
```

Software environment - modules

- module avail

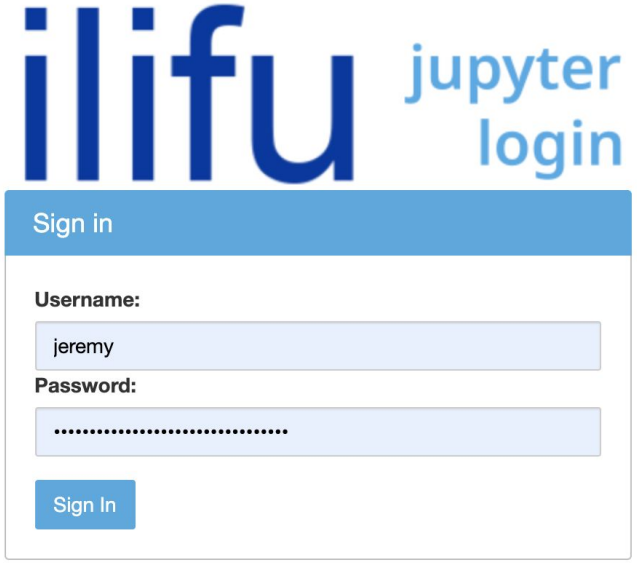
- module help <module>

```
~$ module help python

--------------------------- Module Specific Help for "python/3.9.0" ---------------------------------------
This module configures Python 3.9.0 for use
```

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

JupyterHub

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



Demo time