Astronomy in Focus
XXXIInd IAU General Assembly, August 2024
Diana M. Worrall, ed.
doi:10.1017/S1743921325000122

Ilifu: A Cloud Computing System Supporting Astronomy and Bioinformatics

Rob Simmonds

Inter-university Institute for Data Intensive Astronomy
South Africa

Abstract. This paper describes the ILIFU cluster which is an OpenStack based cloud system located at the University of Cape Town. It is operated by the Inter-university Institute for Data Intensive Astronomy and mainly supports astronomy and bioinformatics research. In terms of Astronomy the main data processed is from the MeerKAT and MeerLicht telescopes that are both based in South Africa.

Keywords. OpenStack, IaaS cloud, Astronomy, Bioinformatics

1. Introduction

The ILIFU cluster was established in 2016 and is operated by the Inter-university Institute for Data Intensive Astronomy (IDIA) which is an institute that spans three universities; the University of Cape Town (UCT), the University of the Western Cape (UWC) and University of Pretoria (UP). The cluster was built using the management nodes from an existing OpenStack (https://openstack.org/) system that was constructed to explore the use of an OpenStack based Infrastructure of a Service (IaaS) cloud system to support university research. This meant the system could be brought into production very quickly as it just needed the new network switches, the compute and storage nodes added to the working OpenStack system. The name ilifu means Cloud in the local Xhosa language and the system is now used primarily to support astronomy and bioinformatics research.

The rest of this paper is laid out as follows. Section 2 explains the hardware infrastructure and how the system is connected to the research networks. Section 3 explains the middleware employed while Section 4 describes the core software that is run. Section 5 then explains the IDIA science gateway that was added to the system in 2023. Section 6 describes the main uses of the system and the paper is concluded and future work is discussed in Section 7.

2. Hardware

The ilifu cluster currently has over 4000 processor cores with most nodes having 256 GB of RAM and 32 cores. Some nodes have 512 GB and one has 1.5 TB of RAM. The large memory node is mainly used for remote visualisation and testing of Virtual Reality (VR) streaming. The cluster has over 13 Petabytes of raw storage which is mainly provided by hard drives, but does include SSDs to provide backing store for virtual machines (VMs) and provide POSIX shares used mainly to support visualisation. The main cluster interconnect is a mix of 50 Gb/s and 100Gb/s Ethernet using Mellanox SN2700 switches, with the storage servers connected at 100 Gb/s and the most of compute servers connected at 50 Gb/s. The cluster is connected to the SANReN (https://saran.ac.za) research network with a 100 Gb/s link that then connects to a 400 Gb/s link within South Africa

© The Author(s), 2026. Published by Cambridge University Press on behalf of The International Astronomical Union. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S1743921325000122 Published online by Cambridge University Press

and out to Europe. The network uplink connects to SANReN at the main point of presence in Cape Town and is not shared with any of the UCT campus traffic.

3. Middleware

ILIFU utilises OpenStack IaaS middleware that allows us to run VMs or bare-metal deployments utilising the OpenStack Ironic subsystem. Recently we have also been experimenting with Kubernetes (K8s) (https://kubernetes.io/) as that is an interesting way of supporting interactive applications and for deployment in general, in particular for micro-service architecture applications. We have not promoted K8s for production use yet, since we are still learning the best ways to have it interact with our other low level environments. We use Ceph to provide storage. This provides ephemeral storage to back VM block storage to create volumes (mainly on SSDs), Object Storage for supporting projects with large numbers of small files and CephFS shares for creating large cluster file-systems. All the storage is configurable from the OpenStack Horizon dashboard enabling users to create their own storage entities and this also makes it very fast for the support team to create new computing and storage environments when needed.

Some projects have long-lived cluster environments within ilifu which they maintain themselves using the OpenStack tools. Examples of this is the cluster that is used to process the data from the MeerLicht (https://science.uct.ac.za/meerlicht) telescope which is an optical telescope that tracks the MeerKAT (Jonas 2016) telescope's observations. This functionality is also employed by some of the bioinformatics projects to support custom environments and used to add custom environments to support specific workshops and hackthons. The latter was done for the FM7 hackathons held during the 2024 IAU General Assembly in Cape Town (see FM7 summary, this volume).

Authentication and authorisation are handled by KeyCloak (https://keycloak.org). This system enables us to use local passwords, or preferably to use federated identity providers that can be accessed through SAFIRE (https://safire.ac.za/), EduGain (https://edugain.org/) and EGI's Check-in (https://egi.eu/service/check-in/). It also provides Single Sign-on (SSO) to all our web based applications when they are accessed through the IDIA Science Gateway (see Section 5). Enabling users to authenticate using external identity providers means that support staff no longer need to reset passwords for users. This is something that took time for them to do and also could take time for them to get to if the service was required at a weekend. Most institutional password management systems provide a way for users to reset their passwords without needing assistance, so we leave that to those systems to deal with.

4. Core Software

A large amount of software is provided on ilifu, some of which is specific to domains such as CASA (https://casa.nrao.edu) based astonomy pipeline (Collier et al. 2021) developed by IDIA that is used to process most of the MeerKAT data that arrives at ILFU. As far as possible application software is managed in Singularity containers enabling users to access a large number of applications and often multiple versions of the software. Other software, such as JupyterHub (https://jupyter.org/hub) support many domains, in this case by supporting multiple kernels that then can run specific notebooks for specific tasks.

Interactive web based applications that are accessible directly from external networked devises include the CARTA astronomy visualisation system (http://cartavis.org) and Gen3 (http://gen3.org) bioinformatics system. Note that JupyterLab that starts JupyterHub instances is also available in this way.

https://doi.org/10.1017/S1743921325000122 Published online by Cambridge University Press

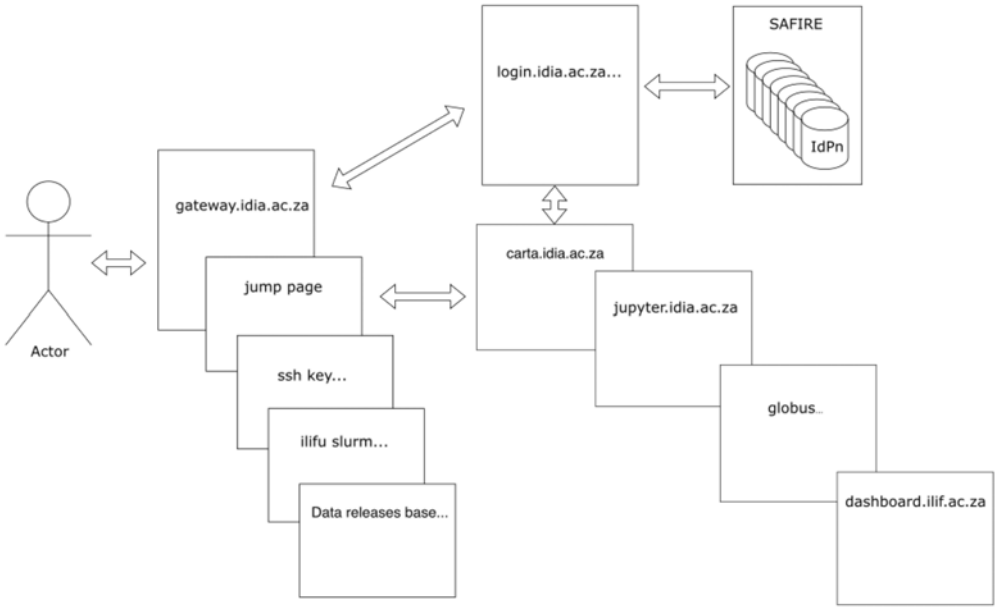


Figure 1. High level design of the IDIA Science Gateway.

5. IDIA Science Gateway

As part of a project funded by The Data Intensive Research Initiative of South Africa (https://dirisa.ac.za), IDIA created a science gateway to access ILIFU. The aim was to create a portal that would provide a single sign-on environment to best-in-class web applications with unified branding. This aim was based on previous experience creating the CyberSKA (Kiddle 2011) portal that was run in Canada and South Africa for over fifteen years which was created using a portal toolkit. The problem with that was that many of the components needed to written and maintained by the team that created the portal. Since CyberSKA used many software layers, maintaining the components was very time consuming, and was necessary when security issues were discovered somewhere in the stack. Seeking funding for code maintenance is more difficult than gaining funding for new tools, so we wanted to limit the number of components that needed to be maintained by ourselves.

Figure 1 shows the basic design diagram for the gateway. This was created quickly before an IDIA group meeting on a Monday morning. The inital version of the gateway was working one week later and exposed to the end users in production form three weeks after this. That compares to years of work that went into creating the CyberSKA gateway and shows the power of the tools used to create the gateway.

A few of the more simple components were created by our support team using the Django (https://www.djangoproject.com) portal framework. These included the Landing Page, an SSH public key management page and a Usage Information Page. The purpose of the SSH public key management page was to enable users to add new public keys if there was any problem with their current key pair. This was a task previously requested from the support team and happened quite often.

Other applications are accessed via an OpenOnDemand (https://openondemand.org) instance that provided fast setup and easy configuration (see Figure 2). This was the major key to being able to setup the IDIA Science Gateway so quickly. Of the best in class applications JupyterLab is exposed from the OpenOnDemand Dashboard. Also exposed

https://doi.org/10.1017/S1743921325000122 Published online by Cambridge University Press

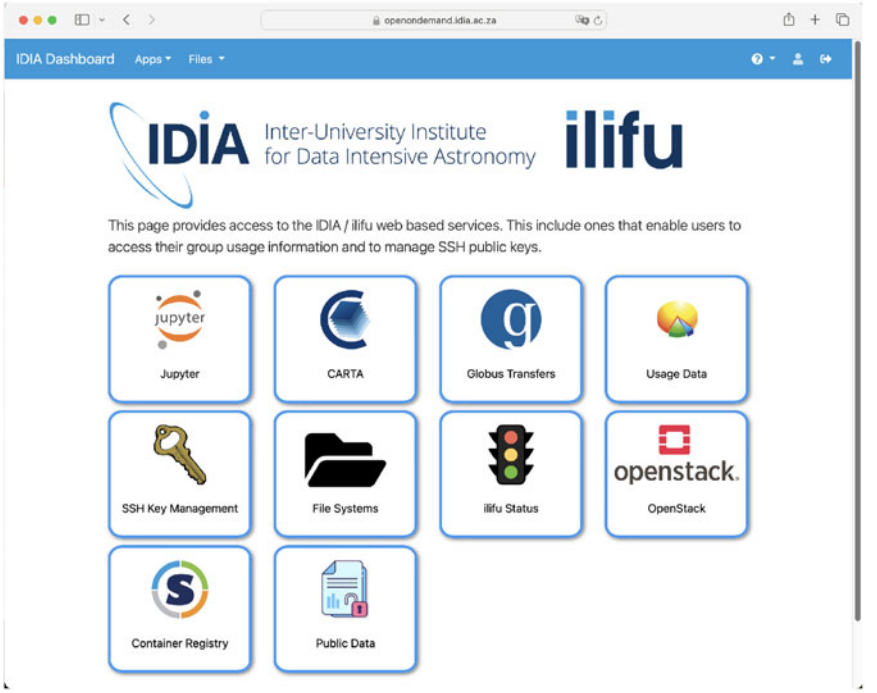


Figure 2. The IDIA Science Gateway application launch page.

in this way are the CARTA (https://cartaviz.org) visualisation system, the Globus (https://globus.org) Wide Area Network transfer client interface and the OpenStack Horizan Dashboard that is used for configuring the computing environments. Also this year we added access to the Singularity Registry (https://singularityhub.github.io/sregistry) that is used to maintain the Singularity containers that provide the majority of software to the ILIFU cluster and allows containers to be downloaded for use on other systems.

The Data Release pages are exposed using the Wagtail (https://wagtail.org) Content Management System (CMS) that is often used with Django. This is exposed from the Front Page, for access to only public data, and from within OpenOnDemand to provide access to all the released data.

6. Users

Five of the eight MeerKAT telescope Large Survey projects are run on ilifu. These include:

- LADUMA (Baker et al. 2018) (Deep neutral hydrogen)
- MIGHTEE (Jarvis 2016) (Deep continuum imaging of the early universe)
- MeerKAT Fornax Survey (Su 2021) (Deep HI Survey of the Fornax cluster)
- MHONGOOSE (de Blok de Blok) (targeted nearby galaxies HI)
- ThunderKAT (Fender 2016) (exotic phenomena, variables and transients)

Research associated with 77 MeerKAT Open Time projects and over 20 Director Discretionary Time projects have also been supported on ilifu to date.

The ilifu cluster was partly funded by the Cbio (Fender cbi) biochemistry group at UCT. There are currently about 140 active bioinformatics users and current projects include H3ABioNet (cbi H3A) and eLwazi (https://elwazi.org/), an African led open data science platform that provides an interactive environment to apply data science

https://doi.org/10.1017/S1743921325000122 Published online by Cambridge University Press

techniques to diverse datasets for novel health discoveries. IDIA staff have worked on updating the Gen3 (http://gen3.org) bioinformatics platform to support eLwazi in the OpenStack middleware environment.

7. Summary and Future Work

The ilifu OpenStack cluster that was established in 2015 supports over 550 regular users working mainly in Astronomy and Bioinformatics. The IDIA Science Gateway has been added that enables signon from federated identity providers, and in conjunction with KeyCloak, provides SSO within the cluster. The use of the OpenStack middleware has enabled a very flexible environment, where custom clusters with their own storage connected can be created quickly. This provides the ability to support projects that would be difficult to support in a conventional high performance computing environment without far more interaction for the cluster's support team and without delays caused by needing lower level configuration.

Currently for users of MeerKAT data to get the data to ilifu they need to log into a portal at South African Radio Astronomy Observatory (SARAO) (https://sarao.ac.za) and register the need for the transfer. We would like to streamline that so they can either do that directly from the IDIA Gateway, or SSO from the IDIA Gateway to SARAO register the transfer. We are currenly looking at how we could best use data replication tools, in particular RUCIO (https://rucio.cern.ch) to better schedule the movement of data between SARAO, IDIA and possibly other sites such as Centre for High Performance Computing (https://chpc.ac.za) in Cape Town. We see the need for that to avoid the case where we have the same raw data available on multiple systems that could be accessed quickly and avoid users wanting to keep copies locally due a lack of clarity on when data will be fetched. This will be aided by planned changes to the way that SARAO stores data, but improving automation in this area will help in any case.

Other changes to ilifu itself could be driven by changes in the applications being run. For example, IDIA is currently looking at changing the design of the backend of the CARTA visualisation system from a monolithic to a micro-services architecture that could provide greater scalability on large clusters and supporting that well will require some changes to the ilifu middleware.

We would also like to federate more functions with more sites to enable users to move quickly from one cluster to another, or just offload tasks across clusters more simply. We will explore funding options to enable this.

References

Baker A. J., et al., 2018. In *American Astronomical Society Meeting Abstracts*. No. 231.

de Blok W. J. G. et al. "MHONGOOSE: A MeerKAT nearby galaxy HI survey". In *Astronomy and Astrophisics.*, 10.1051/0004-6361/202348297.

CBio web site, https://health.uct.ac.za/computational-biology.

Collier J. D., et al., 2021. "The IDIA PROCESSMEERKAT pipeline: Fast CASA Processing on a Cloud-Based HPC Cluster". In *2021 XXXIVth General Assembly and Scientific Symposium of the International Union of Radio Science (URSI GASS).* pp 1–4, 10.23919/URSIGASS51995.2021.9560276.

Fender R. et al., 2016. "ThunderKAT: The MeerKAT Large Survey Project for Image-Plane Radio Transients". In *MeerKAT Science: On the Pathway to the SKA.*, 10.48550/arXiv.1711.04132.

H3ABioNet web site, https://www.h3abionet.org/index.php/nodes/cbio.

Jarvis K. et al., 2016. "The MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) Survey". In *Proceedings of MeerKAT Science.*, 10.48550/arXiv.1709.01901.

Jonas J., 2016. "On the Pathway to the SKA". In *Proceedings of MeerKAT Science.*

https://doi.org/10.1017/S1743921325000122 Published online by Cambridge University Press

Kiddle K. et al., 2011. "CyberSKA: an on-line collaborative portal for dataintensive radio astronomy". In *Scince Gateway Workshop*. pp 65–72, 10.1145/2110486.2110496.

Su A. H. et al., 2021. "The Fornax Deep Survey with VST". In *Astronomy and Astrophysics.*, 10.1051/0004-6361/202038015.

https://doi.org/10.1017/S1743921325000122 Published online by Cambridge University Press