

RapidXfer: Proposed Data Transfer Framework for Square Kilometre Array

- Dr Priyaa Thavasimani, Prof Anna Scaife,
Jodrell Bank Centre for Astrophysics,
The University of Manchester



Introduction to SKA

- World's largest radio telescope project – sites Australia (ASKAP) and South Africa (MeerKAT) - integrated into Phase 1 of SKA.
- The volume of telescope data is vast.
- data rate as approximately 23 Tb/s from the antennas to the correlator, 14 Tb/s from the correlator to the HPC (provided by SDP).
- Expected volume of data to partner countries – 600 PB / year
- 15 member countries.

What is MeerKAT?



- Karoo Array Telescope
- radio telescope consisting of 64 antennas in the Northern Cape of South Africa, inaugurated in 13 July 2018.
- a precursor for the SKA

MeerKAT Data transfer from IDIA to IRIS

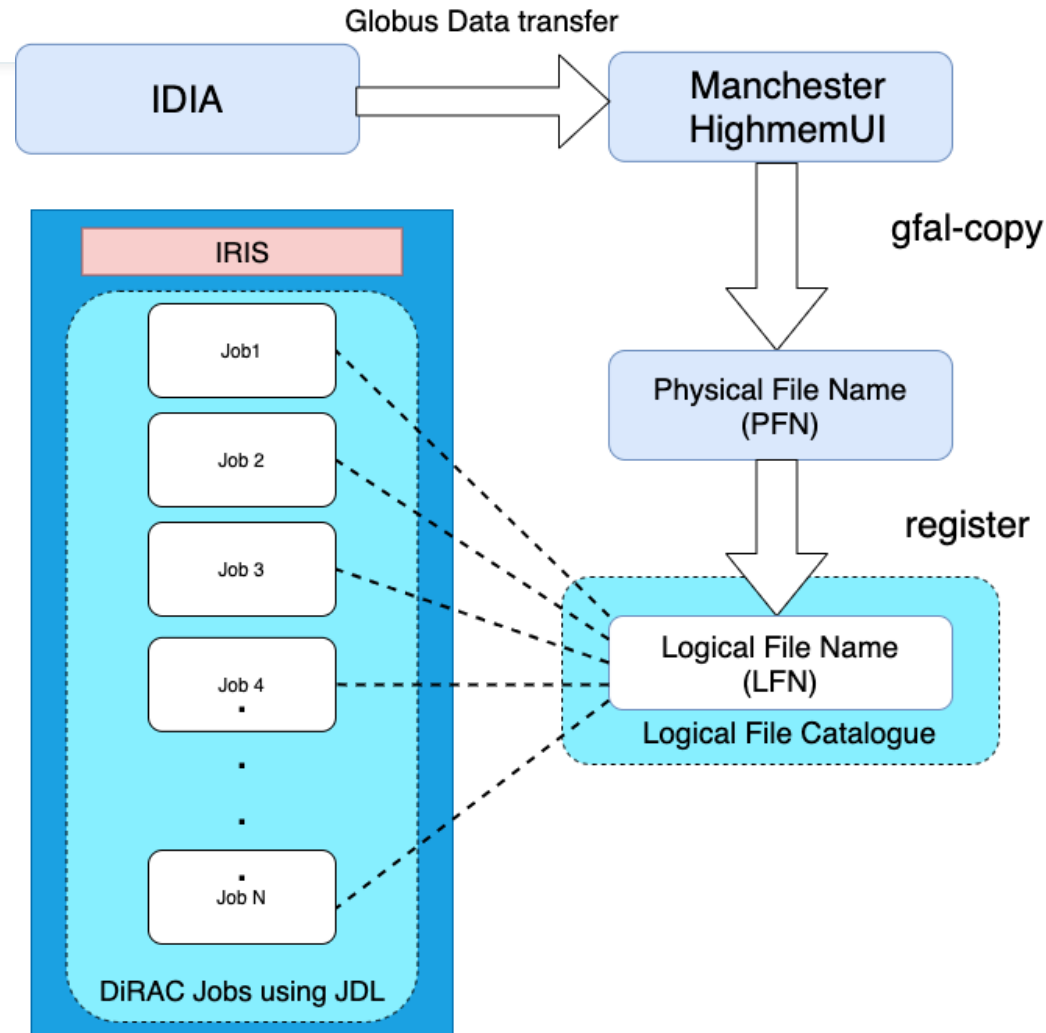
- IDIA (Inter-University Institute for Data Intensive Astronomy) intends to provide researchers from South Africa with cloud resources to process Astronomical Big Data.
- IRIS funded by the STFC (one of research council in UKRI) - create and develop the digital research infrastructure - computing related resources needed to support science.
- Datasets (of 1.3 TB approx each) to be transferred from IDIA to IRIS.
- Globus online - secure, reliable service, enables you to share and transfer files faster than traditional scp and rsync transfers.



Containerised DiRAC Environment on IDIA

- SKA UK is currently using WLCG DMS/WMS services in particular DiRAC.
- Containerised DiRAC Environment on IDIA
- `dirac-dms-add-file` functionality facilitates transfer of smaller files.

RapidXfer - Data Transfer and Processing Framework



Globus Transfer of 1.33 TB Data

The screenshot displays the Globus Transfer web interface. On the left is a dark blue sidebar with navigation icons for File Manager, Bookmarks, Activity (highlighted), Endpoints, Groups, Console, Account, Logout, and Help. The main content area shows a transfer titled "Ilifu DTN to ManchesterUI" with a green checkmark and the status "transfer completed". Below the title are tabs for "Overview" (selected) and "Event Log". The "Overview" tab displays the following details:

| | |
|-------------|--|
| Task Label | Ilifu DTN to ManchesterUI |
| Source | Ilifu DTN ⓘ |
| Destination | ManchesterUI ⓘ |
| Task ID | 109394ac-a403-11ea-9a3a-0255d23c44ef |
| Owner | Priyaa Thavasimani (0000-0002-4846-1661@orcid.org) |
| Condition | SUCCEEDED |
| Requested | 2020-06-01 01:26 pm |
| Completed | 2020-06-02 02:34 pm |

Transfer Settings

- verify file integrity after transfer
- transfer is not encrypted
- overwriting all files on destination

Summary statistics are shown in a box on the right:

| | |
|------------|-------------------|
| 79 | Files |
| 14 | Directories |
| 1.33 TB | Bytes Transferred |
| 14.78 MB/s | Effective Speed |
| 0 | Skipped |

[View debug data](#)

Data Transfer Speed from IDIA to IRIS (Approximate Calculation)

- (Globus transfer) 1 TB = 18.79 hours
Gfal-copy = 0.829 hours
Total = 19.56 hours
Total Data rate: 14.20 MB/sec

Note: Represents average repeatable performance, i.e. more than one transfer.

Ongoing Work:



- Some of our 'Data Transfer' tests show Rucio have achieved a 3MB/s sustainable transfer rate.
- Applicability for larger sample datasets.
- Comparing gfal-copy with 'add-File' functionality of Dirac.
- Comparison or Integration with Rucio.

Thank you!

References:

- [1] [SKA, "The ska project."](#) .
- [2] [J. B. C. for Astrophysics, "The square kilometre array \(ska\)."](#).
- [3] ["Inter-university institute for dataintensive astronomy."](#)
- [4] [AENEAS, "Ska regional centres."](#)
- [5] ["Rucio - scientific data management,"](#)
- [6] [STFC-IRIS.](#)
- [7] [Globus Online.](#)