LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Astronomische Gesellschaft
German Astronomical Society

Data-heavy Astro processing:
deploying PUNCH jobs on compute clusters with attached parallel filesystems

Daniel Hernandez-Lang

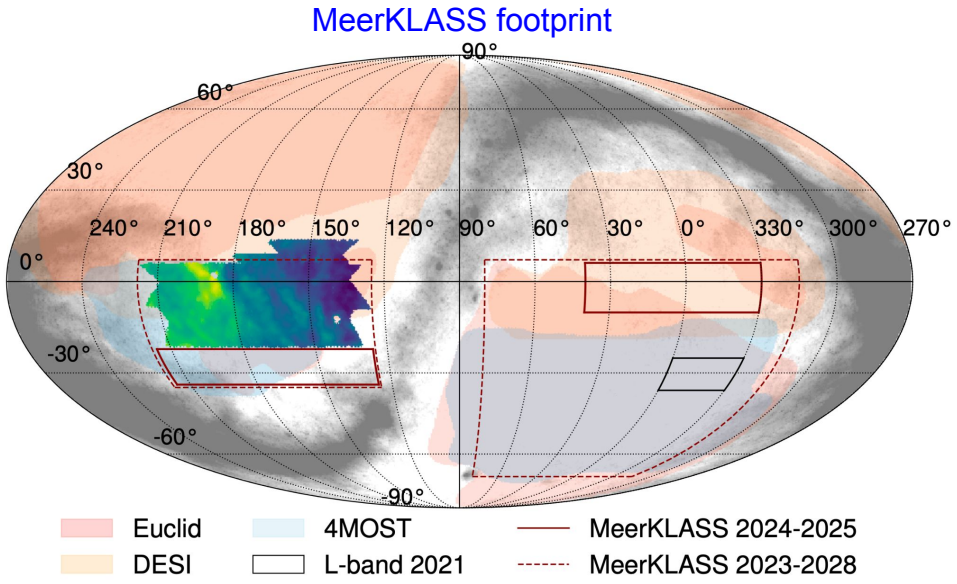
In collaboration with: Sarvesh Mangla, Joseph Mohr, Kristof Rozgony, Benoit Roland, Manuel Giffels and Stephan Hachinger

The use case

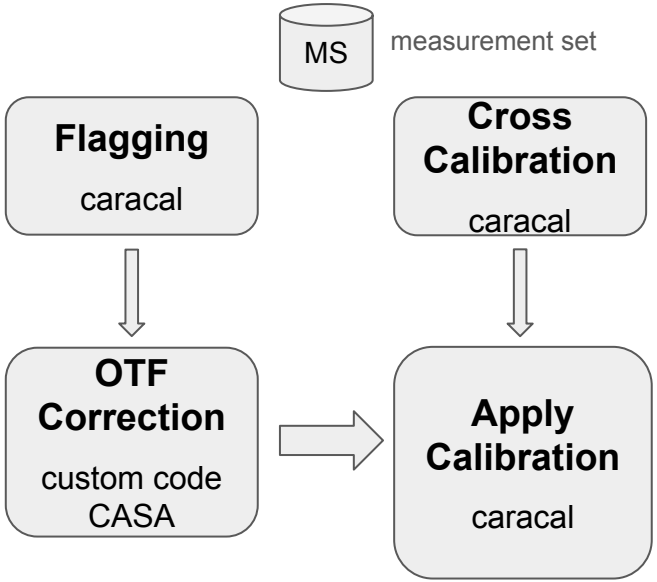
- Large data volume/need for intensive data access → Test storage
- Computationally intensive/need for computation resources → Test computing
- Complex workflow → Test workflow management
- Delivering results/making results available through PUNCH → Metadata section and science reproducibility

MeerKLASS survey

- Observed with MeerKAT array
- 300 deg^2 in L band available.
- 10,000 deg^2 in UHF band
- ~4 PB of raw data
- 2,000 hr of observations.
- Science goals:
  - HI intensity mapping for BAO
  - Continuum imaging of AGN, galaxies
- Intensity Mapping requires scan-mode observing
- New software enables On The Fly (OTF) interferometric imaging.



On-The-Fly Pipeline overview

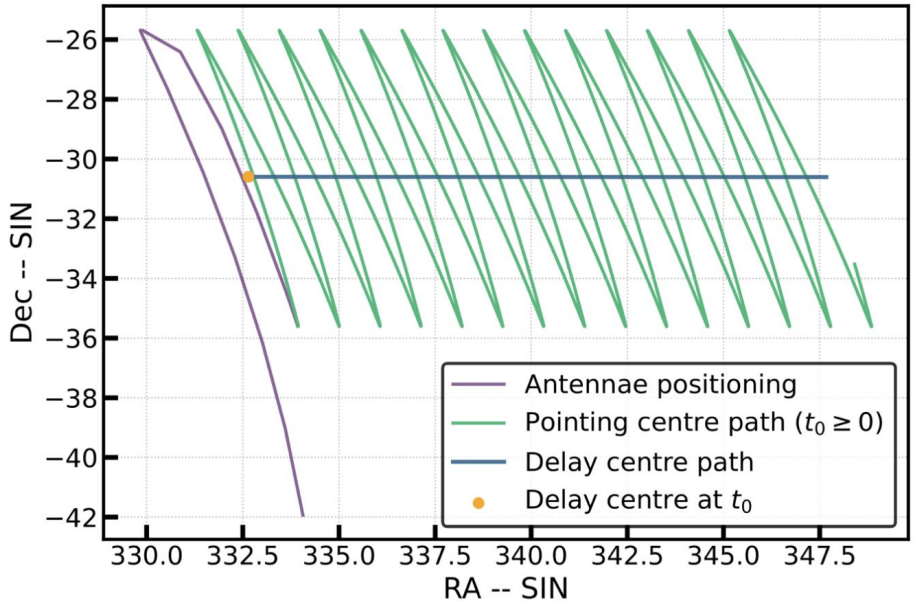


Data acquired in scans of 300 deg^2 each, with each scan lasting 2 hr.

Each 300 deg^2 sky region is scanned ~40 times.

OTF corrected Visibilities

DDFacet

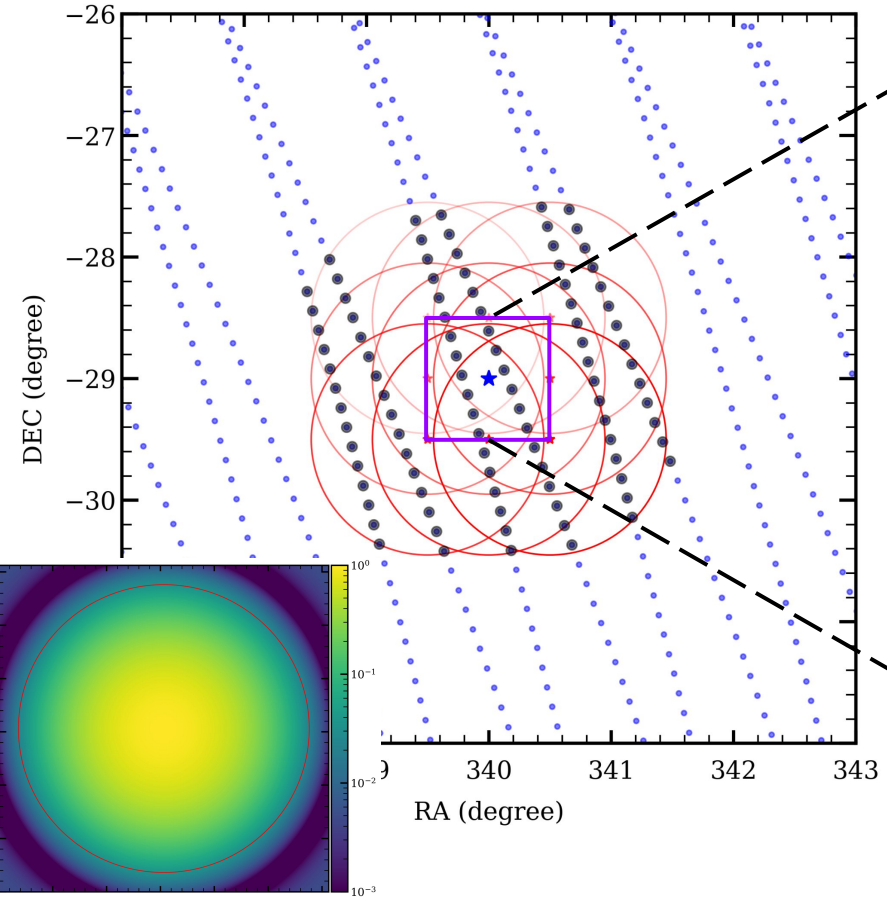


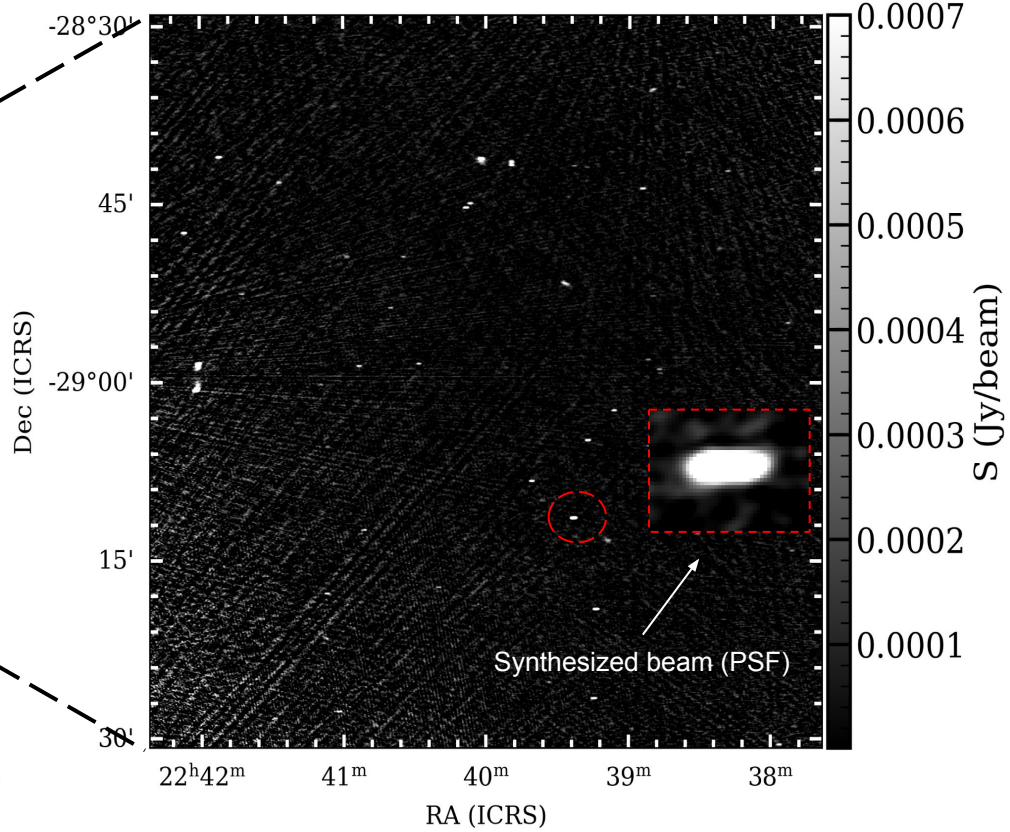
Antenna pointing and delay setup for MeerKAT scans: array scans back and forth in azimuth at fixed altitude, while the Earth rotates new sky into view

During scan, MeerKAT correlator integrations are stored every 2s, creating about 3000 snapshots

DDFacet Selection Criteria

To build sky patch (1 deg^2 in figure below), include all 2s snapshots that overlap that patch (~100 per scan)





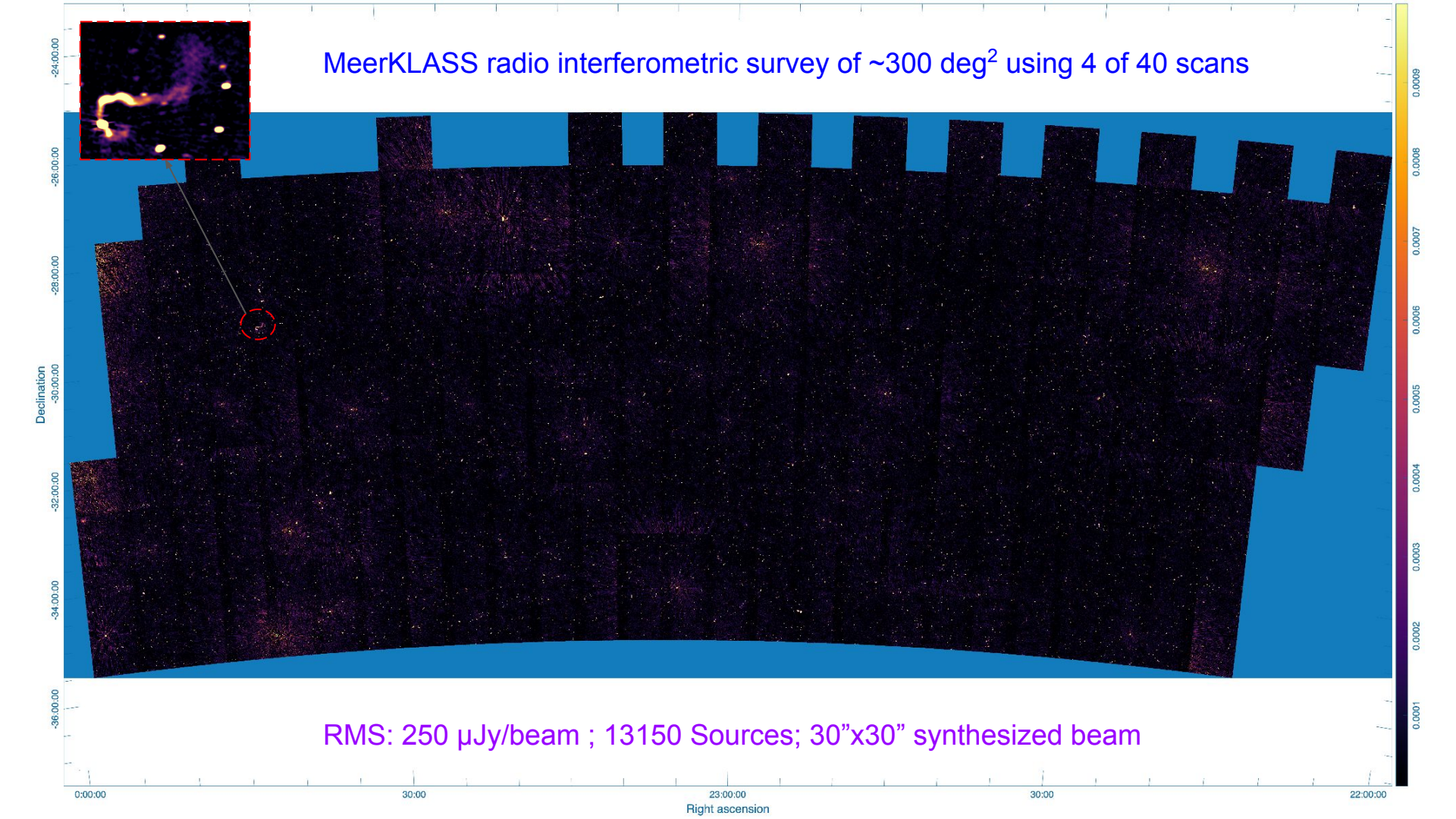
Requirements to image 1 deg^2 sky patch scanned 4 times

Computing resources

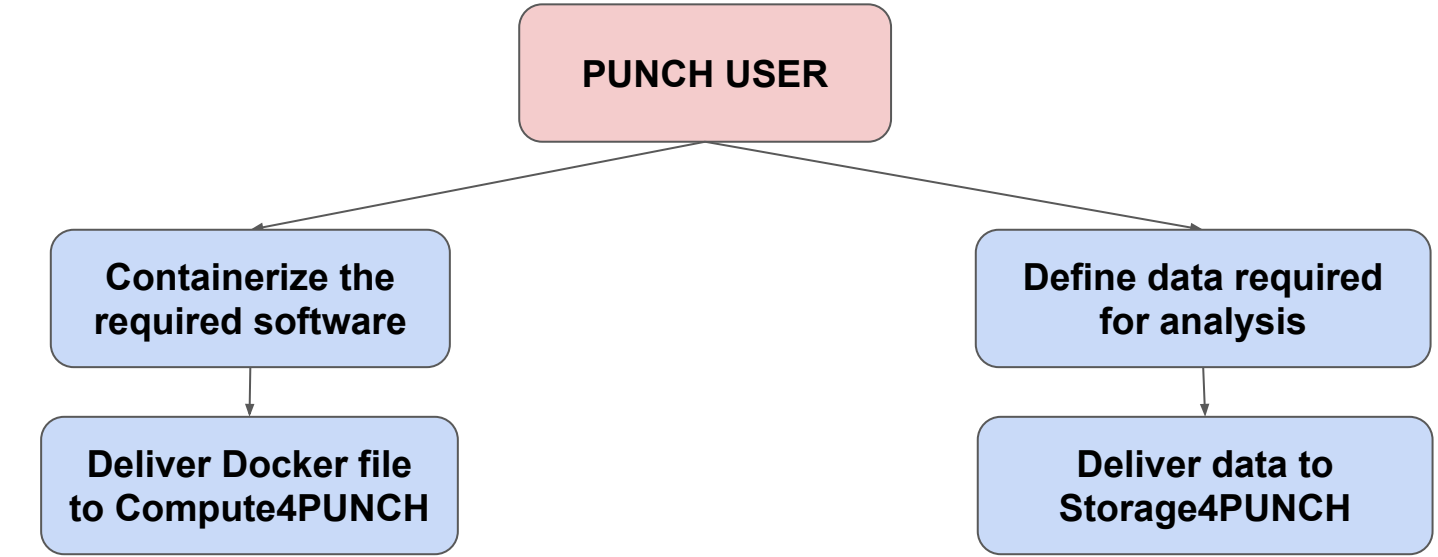
- 128 cores (single node)
- 500 GB of RAM
- Runtime ~7-8 hours

Disk resources

- Each 2s snapshot is ~200 MB
- ~400 snapshots are used to build 1 deg^2 skypatch → 80 GB
  - Each sky region will be re-scanned 40 times → 800 GB input data
- Intermediate data products 4-5 times larger → 400 GB to 4 TB
- Final output images and catalog from 1 deg^2 → 5 GB


MeerKLASS radio interferometric survey of ~300 deg^2 using 4 of 40 scans
RMS: 250 μJy/beam ; 13150 Sources; 30”x30” synthesized beam
Declination
-24:00:00
-26:00:00
-28:00:00
-30:00:00
-32:00:00
-34:00:00
-36:00:00
0:00:00
30:00
23:00:00
30:00
22:00:00
Right ascension
0.0009
0.0008
0.0007
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001

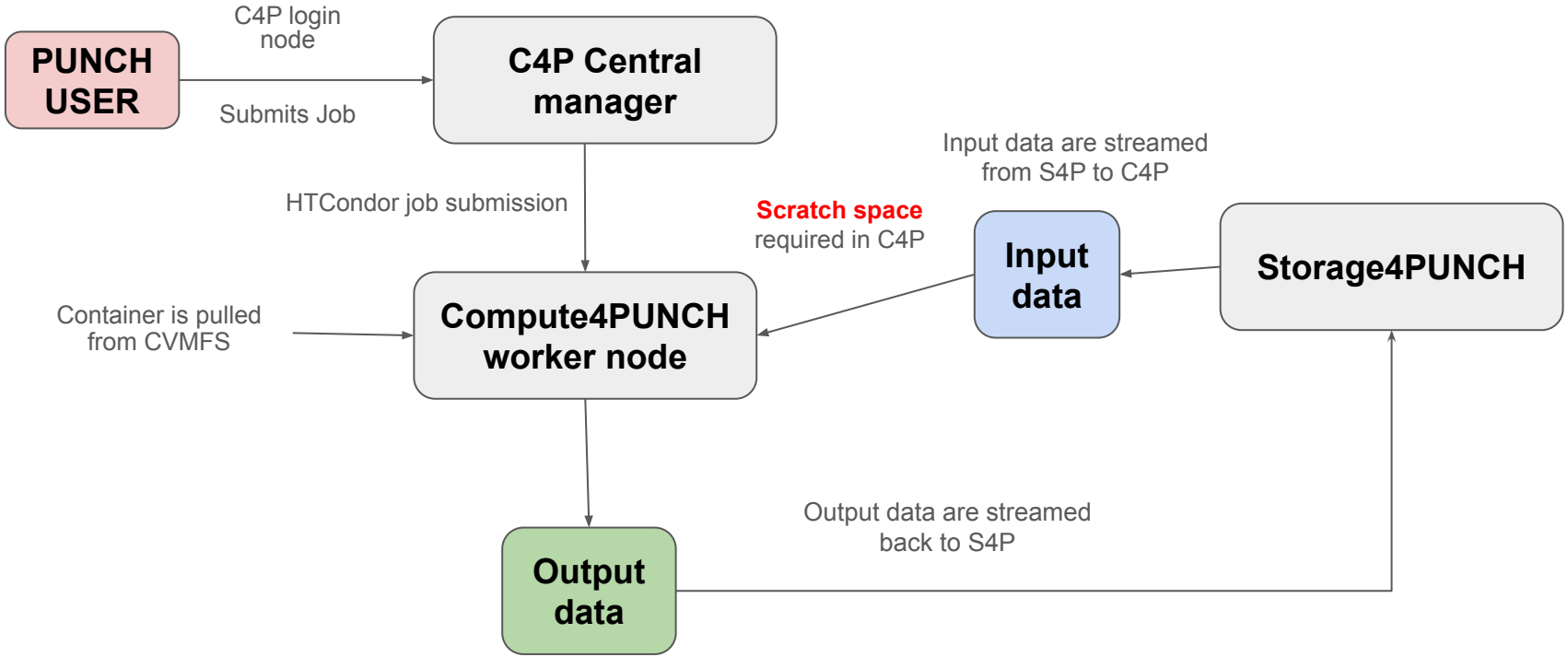
Usual PUNCH analysis framework: Preparation



Compute4PUNCH

- Validates/builds container
- Posts to CVMFS for use

Usual PUNCH analysis framework: Analysis Job

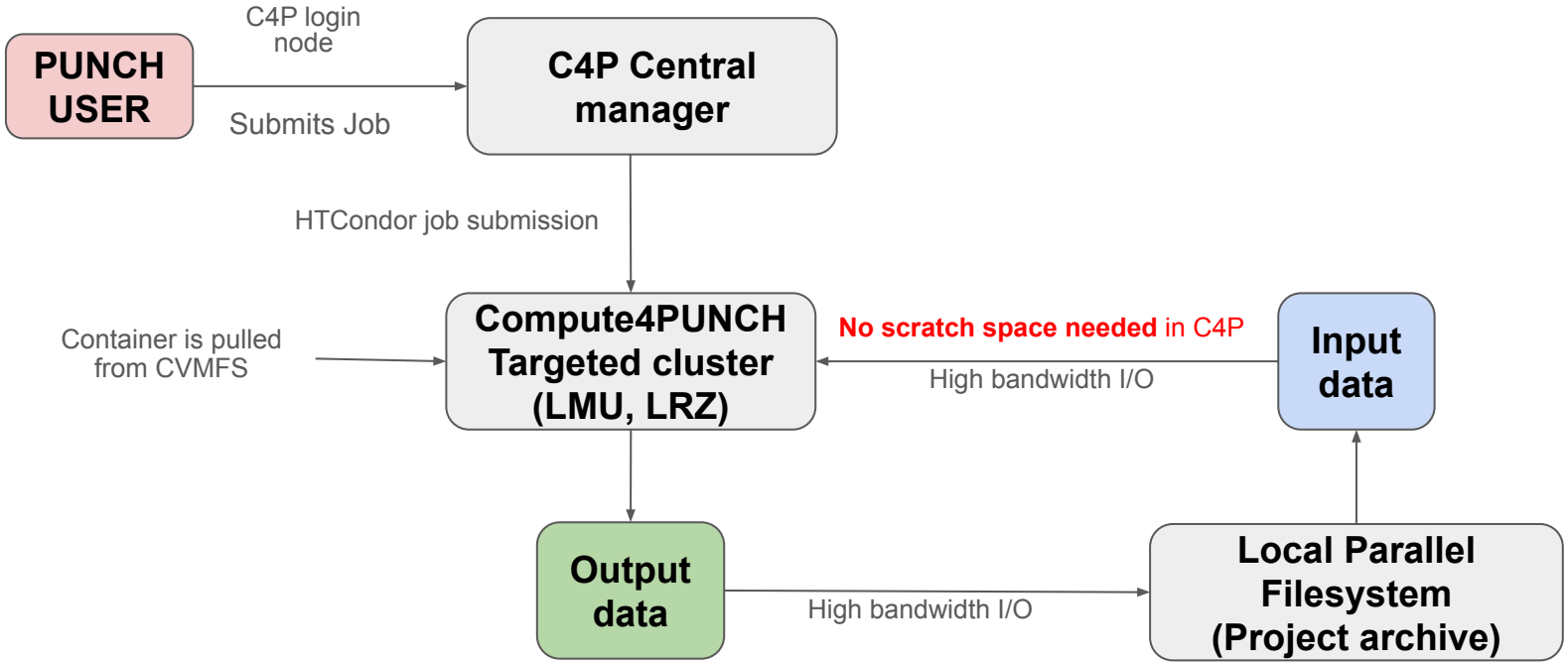


PUNCH extension: data pre-staged on local parallel filesystem

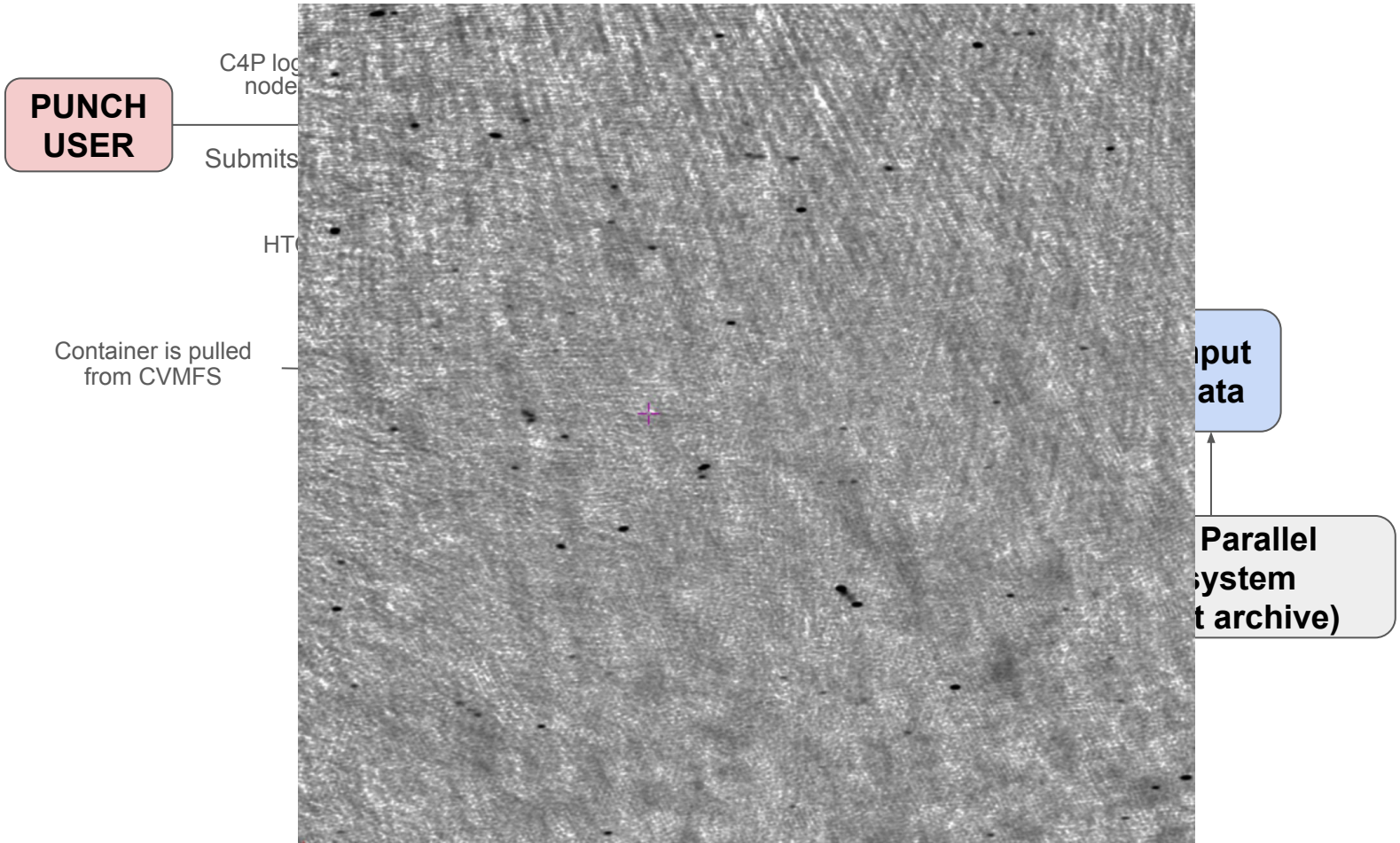
"take the computing to the data"

- Integrate target compute cluster (HPC or group resources) into PUNCH
- Input data pre-staged as archive on local parallel filesystem
- Output data can be stored locally on same filesystem
- Many jobs can run in parallel, I/O bandwidth to disk higher than network I/O
- Compute platforms tested so far for a few test users
  - LMU group compute cluster
  - HPC cluster at LRZ (C2PAP cluster)

PUNCH extension: Analysis Job on Target Cluster



PUNCH extension: Analysis Job on Target Cluster



Conclusions

- MeerKLASS imaging provides a data intensive test case for PUNCH tools.
  - Radio imaging jobs require 80-800GB input data each, and there are 1000’s of them

- → PB-scale data sets pre-staged on local parallel filesystem that is direct connected to C4P target cluster (no streaming over network)

- PUNCH extension tested on LMU compute cluster and HPC cluster at LRZ

- Enables data heavy processing in limit of non-streamable data types (e.g. FITS files in astronomy) and analysis jobs with large input datasets