URSI AP-RASC, Sydney, Australia, 17 – 22 August 2025

The High-performance Processing pipeline for the ASKAP Telescope

Matthew T. Whiting*(1) and Wasim Raja(1)
(1) CSIRO Space & Astronomy, PO Box 76, Epping NSW, 1710 Australia

1 Extended Abstract

The ASKAP Telescope [1] is a cm-band synthesis imaging telescope operated by CSIRO at Inyarrimanha Ilgari Bundara, the CSIRO Murchison Radio-astronomy Observatory in Western Australia. Its defining feature is its wide field-of-view, made possible by the innovative phased-array-feed (PAF) receivers. This field-of-view, while enabling a range of large-area survey science projects, results in very large data rates coming out of the correlator. This requires high-performance computing to create the calibrated images and catalogues quickly enough to keep up with the incoming data rate, and allow the data products to be deposited in the CSIRO ASKAP Science Data Archive (CASDA) for use by astronomers.

The data processing that creates the science products is orchestrated by the ASKAP Pipeline, a scripted workflow that leverages the Slurm workload manager to execute all the required calibration, imaging and source-finding tasks. These tasks are realised through the custom-written calibration & imaging package ASKAPsoft, which has been designed to address challenges presented by the scale of data ASKAP produces. One of the key requirements of the Pipeline is that a given observation must be processed in near-real-time in order to keep up with the incoming data rate, and this has resulted in highly-parallelised processing tasks.

The ASKAP Pipeline is operational, with regular survey observing resulting in large amounts of data being deposited into the CASDA archive and made publicly available. Data rates into the archive can be around 70 TB per week. The Pipeline is nearly autonomous - most observations are processed in a completely hands-off fashion, with only a small fraction requiring manual attention (perhaps due to data quality issues). It is a demonstration of what can be achieved through a large and complex supercomputing workflow, and provides important lessons for planning of even larger workflows expected for future facilities.

This talk will provide descriptions of the key design decisions that shaped the workflow and allowed it to scale to its present capacity. It will describe the supercomputing setup used at the Pawsey Supercomputing Centre, including the different types of jobs and their varying requirements and impacts on the system. The talk will reflect on lessons learned from managing the large data rates and consequent high I/O, and from scaling up the processing and managing capacity constraints on different systems. The talk will also take a forward look, examining potential upgrades and likely considerations for future large-scale processing for facilities such as the SKA.

References

[1] A. W. Hotan et al., “Australian square kilometre array pathfinder: I. system description”, *PASA*, **38**, pp. e009–e039, 2021, doi:10.1017/pasa.2021.1