URSI AP-RASC 2019, New Delhi, India; 09 - 15 March 2019

The ASKAP and its big data challenge

M. A. Voronkov*[(1)], D. Mitchell[(1)], S. M. Ord[(1)], M. T. Whiting[(1)], W. Raja[(1)], M.H. Wieringa[(1)], E. Bastholm[(2)] and J. C. Guzman[(2)]
(1) CSIRO Astronomy & Space Science, PO Box 76, Epping, NSW 1710, Australia, http://www.atnf.csiro.au
(2) CSIRO Astronomy & Space Science, PO Box 1130, Bentley, WA 6102, Australia

The Australian Square Kilometre Array Pathfinder (ASKAP) is a new generation low frequency (700-1800 MHz) interferometer capable of observing an instantaneous field of view of 30 square degrees. Such a wide field of view is realized using the Phased Array Feed (PAF) technology that allow forming multiple beams on the sky. The telescope is already in the early science phase with 28 of the 36 antennas routinely available for observations at the time of writing. The hardware has been installed in all antennas and the full array is nearing completion. It is capable of generating about 200 Tb/day of raw data and producing spectral cubes up to 70 Tb in size per observation (if all the data are used; the current plan is to sacrifice the spatial resolution). This fact makes ASKAP an excellent pathfinder and test-bed for the big data challenge which the Square Kilometre Array (SKA) project is facing.

To meet these high-performance computing (HPC) challenges we have taken measures at several levels of ASKAP's design, including hardware, processing software and operational approaches. In particular, the unique mount of the telescope and on-dish radiators are employed to stabilise the beam patterns on the sky reducing the need to deal with time-variable direction-dependent calibration effects during data reduction. Also, some commissioning effort has been directed towards understanding the effects normally absorbed into self-calibration solutions as well as to control the beam shapes. This allowed us to reduce the required cadence of gain calibration and simplify the imaging process, thus relaxing the overall compute requirements.

Nevertheless, the initial plan was to reduce the data in near-realtime and dispose the raw data shortly after processing. To achieve this, we designed a specialised suite of software called ASKAPsoft. In contrast to a general purpose synthesis imaging software like CASA, ASKAPsoft targeted very closely the ASKAP's processing model and has been designed up front for distributed processing in the HPC environment with minimal overheads. The early science phase revealed the need for a greater flexibility on the data reduction part, at least in the short term. Therefore, the original "one size fits all" processing pipeline has been deferred. Instead, we use a suite of scripts wrapping around the original ASKAPsoft tasks to achieve offline processing on the dedicated Cray XC30 "Galaxy" supercomputer at the Pawsey Supercomputing Centre in Perth, Western Australia. This approach enables more experimentation with the imaging algorithms and their parameterisation for the price of reduced operational efficiency.

The main factors driving the development of the imaging algorithms and the data distribution framework are the available memory (both total and per core), memory bandwidth and disk I/O. In the course of several years, we tried different approaches and converged on partitioning data both per beam and by splitting the bandwidth. This approach allowed us to achieve about 50% core utilisation on the current platform which can be further improved with the hardware upgrade. We review the current state of the data reduction software, challenges encountered, both in terms of the scaling in the parallel environment and algorithmically, as well as future plans and potential reuse of the ASKAP's software for the early SKA. In particular, the processing model of ASKAP bears some similarities to that of the SKA-MID enabling straight forward reuse and the case of early SKA-LOW can be catered for with reduced efficiency.