



Batch processing pipeline development in the SKA Data Processing ART: Updates from team DHRUVA

Ruta Kale^{*(1)}, Abhirup Datta⁽²⁾, Nimalan M.⁽³⁾, Justin Jose⁽³⁾, Maneesh Sutar⁽³⁾, Swethaa A.⁽³⁾, Manikanta Addhanki⁽³⁾, Nitin Rajkumar⁽³⁾, Suman Chintapatla⁽⁴⁾, Yogesh Wadadekar⁽¹⁾ and Yashwant Gupta⁽¹⁾

(1) National Centre for Radio Astrophysics, T. I. F. R. , Pune, India; email*: ruta@ncra.tifr.res.in

(2) Indian Institute of Technology Indore, Indore, India

(3) Thoughtworks

(4) Persistent Systems Limited

The SKA data processing presents a significant challenge due to the unprecedented data volumes and the requirements of real-time processing. The batch processing pipelines play a role once the data are recorded and are ready to be processed to produce science-ready images. These involve pre-processing the calibrator and target source data, obtaining calibration solutions, applying them to the data, and then carrying out imaging and self-calibration. As a part of the SKA Data Processing Agile Release Train (DP-ART), team DHRUVA, based in India, has been contributing over the period of Planning Increments (PI) 23, 24 and 25 (June 2024 - present) and continuing. The team consists of a combination of radio astronomers and software professionals from industry partners, namely, Thoughtworks and Persistent Systems Limited.

The team has delivered the first “MSv4” compatible spectral line imaging pipeline over the PIs 23 and 24. The MSv4 is the next-generation measurement set data format¹ under development, that can be efficiently processed in a distributed manner. The spectral line imaging pipeline performs continuum subtraction and imaging of the spectral line cube. The work involved the development of a flexible pipeline framework and the individual processing steps. The team also demonstrated the first batch processing for the Science Data Processor (SDP) using spectral line imaging pipeline as a test case. The team collaborated with other teams within the DP-ART to put together the first end-to-end batch continuum imaging pipeline. The batch pre-processing, instrumental calibration and self-calibration pipelines were included in a single pipeline framework to demonstrate an end-to-end run on simulated SKA-Low data. In future, the team foresees working on adding more capabilities to the instrumental calibration pipeline and taking forward the end-to-end continuum imaging pipeline incorporating the improvements in the individual pipelines by the DP-ART teams.

¹ https://xradio.readthedocs.io/en/stable/measurement_set_overview.html