Astronomical Data Analysis Software and Systems XXVIII
ASP Conference Series, Vol. 523
P.J. Teuben, M.W. Pound, B.A. Thomas, and E.M. Warner, eds.
© 2019 Astronomical Society of the Pacific

DALiuGE/CASA Based Processing for the Extragalactic HI Observations with FAST

Vyacheslav Kitaeff,[1] Ming Zhu,[2] Lister Staveley-Smith,[1] Rodrigo Tobar,[1] Kevin Vinsen,[1] Andreas Wicenec,[1] and Chen Wu[1]

[1]*The International Centre for Radio Astronomy Research,*
The University of Western Australia, Australia

[2]*National Astronomical Observatories*
Chinese Academy of Sciences, Beijing, China

Abstract. We present a prototype for the spectral-line data reduction pipeline based on the graph-based execution framework DALiuGE, and the CASA single-dish spectral-line package. The pipeline has been designed for the drift-scan mode of FAST multi-beam radio telescope targeting extra-galactic HI observations.

1. Introduction

The Five-hundred-meter Aperture Spherical radio Telescope (FAST) is planning a multi-beam multi-purpose survey that includes an extragalactic HI survey (Li et al. 2018).

We wanted to develop a pipeline that would be a flexible, scalable, and overall future-proof option to reduce FAST-HI extragalactic survey data. We started from investigating a few established options, including LiveData/Gridzilla, ASAP, proprietary code development, and CASA. We've selected the CASA (McMullin et al. 2007) single dish spectral line package that has been recently refurbished into a new more compact interface starting from version 5.0, along with the DALiuGE (Wu et al. 2017) execution framework. We have also integrated Next Generation Archive System (NGAS) software for data management purposes (Wu et al. 2013).

2. Rational for the selection of software packages

DALiuGE:. The Data Activated Liu Graph Engine (DALiuGE) developed by ICRAR is an execution framework for processing large astronomical datasets at a scale required by the SKA1 (Wu et al. 2017), (`https://github.com/ICRAR/daliuge`). It includes an interface for expressing complex data reduction pipelines consisting of both data sets and algorithmic components and an implementation run-time to execute such pipelines on distributed resources. By mapping the logical view of a pipeline to its physical realization, DALiuGE separates the concerns of multiple stakeholders, allowing them to collectively optimize large-scale data processing solutions in a coherent manner. The execution in DALiuGE is data-activated, where each individual data item autonomously triggers the processing on itself. Such decentralization also makes the execution framework scalable and flexible.

NGAS: The Next Generation Archive System (NGAS) is a feature rich, archive handling and management system (`https://github.com/ICRAR/ngas`). In its core it is a HTTP based object storage system. It can be deployed on single small servers, or in globally distributed clusters. It is possible to run more than one server on a single host and it is possible to run many servers across hundreds of nodes as well as across various sites. It also allows mirroring the sites running independent NGAS clusters or running multiple clusters against a single database. The data can be archived and retrieved programmatically with data integrity being checked via various checksum methods. NGAS has a high customization via user-provided plug-ins. The standard distribution of DALiuGE included NGAS drop.

CASA: Although, CASA had been developed with the primary goal of supporting the data post-processing needs of the next generation of radio astronomical telescopes such as ALMA and VLA (McMullin et al. 2007), (`https://casaguides.nrao.edu/index.php/Main_Page`), the package can process both interferometric and single dish data. The CASA infrastructure consists of a set of C++ tools bundled together under an iPython interface as a set of data reduction tasks. This structure provides flexibility to process the data via task interface or as a Python script as a module in DALiuGE. In addition to the data reduction tasks, many post-processing tools are available for even more flexibility and special purpose reduction needs. The Single Dish tool was initially developed by CSIRO based on the ASAP software package. Beginning from release 5.0.0 the development is driven by ALMA. *Sdcal* function contains calibration modes that make CASA suitable tool for the planned drift-scan HI survey with FAST.

3. FAST-HI pipeline

The prototyped pipeline provides six modules:

1. *FASTcal* – based on *sdcal* that implements a single-dish data calibration scheme similar to that of interferometry, i.e., generate calibration tables (caltables) and apply them.

2. *FASTimaging* – mapping Tsys and Tsky calibrated data onto an image grid.

3. *FASTflagging* – based on *dataflag* that flags an MS or a calibration table.

4. *FASTMLflagging* – convolutional neural network inference automatic RFI flagging.

5. *FASTbaseline* – based on *sdbaseline* that performs baseline fitting/subtraction for single-dish spectra.

6. *FASTexportfits* – converts a CASA image to a FITS file in accordance with FITS 3.0 standard.

Each processing module can have any number of configuration files. If a module is executed without specifying one, it will try using a default configuration. If it cannot find the default configuration, it will create one. The *conf* directory contains some configuration files. These files are not the default configurations, but rather, those that were used during testing on ICRAR test system, and likely would need to be modified when the tests are done on a different system using different datasets.

There are three processing modes currently available and prototyped in DALiuGE logical graphs: real-time calibration, imaging, and reprocessing.

The real-time calibration pipeline assumes that the observation dataset is copied into NGAS archive as soon as it becomes available at the telescope. This will trigger deployment of calibrate_pipeline that will calibrate the observation and copy the resulting datasets into an archive.

The imaging pipeline provides gridding and imaging for selected observations. The configuration file contains a list of observations to be imaged. The image is produced as a measurement set and then exported as FITS cube.

The reprocessing pipeline combines flagging, calibration and imaging step from the archive. This is computationally extensive mode that normally requires a compute cluster.

CASA dataflag provides the algorithms to flag RFI. However, as part of the design we've included an option to use flagging based convolutional neural networks algorithm using PyTorch as a drop in DALiuGE. This technique is promising in characterization of the RFIs that are specific to the telescope in a location, therefore can be more accurate than signal processing based techniques in flagging difficult cases of RFI that have a broadband continues nature of the signal.

4. Summary

We developed a prototype of the data reduction pipeline for the planned FAST HI extragalactic survey that is scalable, extendable, and simple to use and develop further. Commissioning the FAST telescope with 19 beam receiver will allow testing the software on real data, and training machine learning based RFI flagging in near future.

References

Li, D., Wang, P., Qian, L., Krco, M., Dunning, A., Jiang, P., Yue, Y., Jin, C., Zhu, Y., Pan, Z., & Nan, R. 2018, IEEE Microwave Magazine, 19, 112

McMullin, J. P., Waters, B., Schiebel, D., Young, W., & Golap, K. 2007, in ADASS XVI, edited by R. A. Shaw, F. Hill, & D. J. Bell, vol. 376 of ASP Conf. Series, 127

Wu, C., Tobar, R., Vinsen, K., Wicenec, A., Pallot, D., Lao, B., Wang, R., An, T., Boulton, M., Cooper, I., Dodson, R., Dolensky, M., Mei, Y., & Wang, F. 2017, Astronomy and Computing, 20, 1. URL `http://www.sciencedirect.com/science/article/pii/S2213133716301214`

Wu, C., Wicenec, A., Pallot, D., & Checcucci, A. 2013, Experimental Astronomy, 36, 679. `1308.6083`

ADASS
6-10 October 2019 XXIX
GRONINGEN - THE NETHERLANDS
www.adass2019.nl
SCAN THIS QR CODE TO VOTE FOR TOPICS AND SUGGESTIONS FOR ADASS XXIX or go to http://tiny.cc/ADASSvote
THE RESULTS WILL BE SHOWN LIVE AT THIS MONITOR AT DIFFERENT MOMENTS THROUGHOUT THE MEETING.
ADASS
6-10 October 2019 XXIX
GRONINGEN - THE NETHERLANDS
www.adass2019.nl
ADASS
6-10 October 2019 XXIX
GRONINGEN - THE NETHERLANDS
www.adass2019.nl
Scan me

Dutch bribes (Photo: Peter Teuben)