ADIOS improves parallel I/O for Square Kilometer Array (SKA)

Achievement:
- The Square Kilometer Array, or SKA, is a global next-generation radio telescope project
- The full SKA will use a million antennas to enable astronomers to monitor the sky in unprecedented detail and survey the entire sky much faster than any system currently in existence
- Data produced by SKA will be used to verify cosmological simulations produced by DOE codes, such as HACC
- The data can be extremely large, over 25 exabytes/year
- ADIOS is one of DOE’s premier I/O frameworks to reduce the cost of I/O for leadership science
- The integration of ADIOS into the SKA pipeline reduced the processing time by over 10X

Significance and Impact:
- One of SKA’s greatest challenges is in the ability to move, process, and store data, without losing information
- ADIOS has allowed the SKA data processing framework to utilize more I/O bandwidth, allowing SKA to grow to a larger number of antennas

Research Details:
- AdiosStMan, the storage manager for CASA, is part of the SKA Science Data Processor module work, where ADIOS is used as the storage backend
  - CASA is one of the most widely used radio astronomy data processing software packages
  - AdiosStMan shows excellent scalability for table based data processing

Sponsor/Facility: Work was performed at the University of Western Australia, SKA Perth headquarters, and in concert with the Oak Ridge National Laboratory

PI and affiliation: ??

Team: R. Wang, C. Harris, A. Wicenec


Overview:
In this paper, we investigate the Casacore Table Data System (CTDS) used in the casacore and CASA libraries, and methods to parallelize it. CTDS provides a storage manager plugin mechanism for third-party developers to design and implement their own CTDS storage managers. Having this in mind, we looked into various storage backend techniques that can possibly enable parallel I/O for CTDS by implementing new storage managers. After carrying on benchmarks showing the excellent parallel I/O throughput of the Adaptive IO System (ADIOS), we implemented an ADIOS based parallel CTDS storage manager. We then applied the CASA MSTransform frequency split task to verify the ADIOS Storage Manager. We also ran a series of performance tests to examine the I/O throughput in a massively parallel scenario.