

RapidXfer - Proposed Data Transfer Framework for Square Kilometre Array

Tuesday, 3 November 2020 14:20 (15 minutes)

Square Kilometre Array will be the largest radio telescope, which comes with its huge data challenges [1]. SKA's host sites are in South Africa and Australia[2]. Each of the host sites is estimated to produce data at different rates. Very high-performance central supercomputers (one in South Africa and another in Australia) process the extremely voluminous data produced by the SKA. The initial data products [3] that are generated by the SKA's Science Data Processors (SDP) are not suitable for immediate imaging. The Data delivery architecture [4], facilitates the transfer of data SKA-SA (SKA South Africa) CHPC (Centre for High-Performance Computing) to the IDIA [5] Regional Science Data centres using the dedicated Globus endpoint. The partially preprocessed data are sent to SKA Regional Centres around the world for further processing. SKA Regional Centres play a key role in the transfer of data from SKA's sites to CERN's Tier 1 sites and further to other Tier 2 sites. SRC forms an intrinsic part of SKA operations [6], it's model is still at its infancy. The Rucio[7] provides a generic scalable approach to transfer data for high-energy physics experiments and it is still being evaluated for SKA. We propose our Rapid Data transfer framework "RapidXfer", which is a solution that we are currently using to transfer from MeerKAT IDIA to DiRAC's Logical File Catalogue (LFC) for further processing. The framework "RapidXfer" makes use of Globus online transfer through a dedicated Globus endpoint. "Grid File Access Library" shortly called "gfal" is used to transfer from high memory machines to Physical storage where each file is given a "Physical File Name", then it is registered in DiRAC's "Logical File Catalogue"[8]. This DiRAC's register helps to make as many replicas as we need depending on the preferable Storage Element. The "RapidXfer" framework reduced the time for data transfer from South Africa's SDP to the IRIS machine [9] to half compared to traditional SCP transfer and the direct file transfer to LFC from IDIA (uses DiRAC's "dirac-dms-add-file" feature). Our future work focuses on transferring different-sized MeerKAT datasets to evaluate its efficiency and scalability.

Primary authors: Dr THAVASIMANI, Priyaa (The University of Manchester); Prof. SCAIFE, Anna (The University of Manchester)

Presenter: Dr THAVASIMANI, Priyaa (The University of Manchester)

Session Classification: Data transfer workshop - Part 1