The LOFAR LTA distributed information system for Radio Astronomical data

Authors
Hanno Holties[1]

Affilitations
1) ASTRON, Oude Hoogeveensedijk 4, 7991 PD Dwingeloo, the Netherlands
holties@astron.nl

Keywords
Radio Astronomical Instrument, Distributed Information System, Grid, Large Data Volume, Bandwidth On Demand

Abstract
The Low Frequency Array (LOFAR) is a large distributed radio telescope that observes at frequencies from 30 MHz to 240 MHz. LOFAR combines phased array antenna stations in Germany, the UK, France, and Sweden with forty stations in the Netherlands to form a radio telescope aperture synthesis array with a diameter of 1000 km and a total collecting area of up to a square kilometer. When LOFAR is fully operational, it will produce observation data at up to 80 Gbps. The archive of science data products is expected to grow by up to five petabyte per year.

The LOFAR Long Term Archive (LTA) is a distributed information system created to store and process the large data volumes generated by the LOFAR radio telescope. It currently involves sites in the Netherlands and Germany. For astronomers, the LOFAR LTA is the principal interface not only to LOFAR data retrieval and data mining but also to processing facilities for this data. Each site involved in the LTA provides storage capacity and optionally processing capabilities. To allow collaboration with a variety of institutes and projects, the LOFAR LTA merges different technologies (EGI, global file systems, Astro-WISE dataservers). Well-defined interfaces ensure that to both the astronomer and the LOFAR observatory the LTA behaves as a coherent information system. Access and utilization policies are managed via the central LOFAR identity management system that is designed to allow federation with organizational user administrations.

The network connecting the LTA sites in Groningen, Amsterdam and Juelich consists of 10 Gbe connections that are shared with LOFAR station connections and with the European eVLBI network. The bandwidth between the sites is sufficient for regular one-time LTA data transports but to allow transparent processing within the LTA it will need to grow to 60 – 80 Gbps bandwidth. This bandwidth will enable two major new processing modes: 1) Streaming of realtime or buffered observation data to a remote HPC system. 2) Streaming of stored data from one LTA site to a compute cluster located at another site. With these modes an optimal utilization of storage and processing facilities can be

realized: If additional processing capacity is required for a given observing mode or for large-scale data processing, existing capacity at partner institutes can be brought in without having to store (multiple copies of) datasets before processing can commence. For LOFAR datasets, which can grow up to hundreds of Terabytes, this will be essential.

Besides the deployment of a large bandwidth network, several issues will need to be addressed:

- Availability of storage, network, and processing capacity will have to be scheduled in a well-coordinated way (this will be both a technical and an organizational challenge).
- Scalable and sustainable storage systems are required for data buffering that can support simultaneous reading and writing at network speeds.
- Processing frameworks will be required to handle streaming data and long distance network issues such as increased latencies and (temporary) network failures.

The philosophy and architecture of the LOFAR LTA is applicable to any multi-petabyte data system for international science projects, not in the last place to the future Square Kilometer Array (SKA) radio telescope. In this paper an overview of the design and the current state of the LOFAR LTA as well as the requirements and plans for the development of remote processing capabilities are presented.

Vitae

Dr. Hanno Holties is system engineer within the ASTRON Radio Observatory Divisio. He has a wide experience in software development projects for both the WSRT and LOFAR systems. Since 2008 he is project manager of the development programme for the LOFAR Long Term Archive information system. This system is designed to handle the multi petabyte yearly growth of the LOFAR data products. It includes storage and processing facilities distributed over locations in the Netherlands and Germany and provides a central catalog and user interface that can be used by astronomers world wide to work with data products stored within the LTA.