

# Data Analysis in the LOFAR LTA



Hanno Holties<sup>1</sup>, Ger van Diepen<sup>1</sup>, Dennis van Dok<sup>2</sup>, Fokke Dijkstra<sup>3</sup>,  
Marcel Loose<sup>1</sup>, Adriaan Renting<sup>1</sup>, Coen Schrijvers<sup>4</sup>, Willem-Jan Vriend<sup>5</sup>

1) ASTRON 2) NIKHEF 3) RUG/Donald Smits Center for Information Technology 4) SARA 5) RUG/Kapteyn Astronomical Institute



www.astron.nl

## The LOFAR Telescope

The LOw Frequency ARray is a multi-purpose sensor array. Its main application is astronomy at low frequencies (10-250 MHz). Its heart is in the Northeast of the Netherlands but LOFAR stations are spread over the whole of Europe.



Radio signals are collected and pre-processed at each station and sent to the Central Processing (CEP) facility. At CEP, an IBM BG/P supercomputer takes care of real time processing while two linux clusters are available for further data analysis, each with approximately one hundred nodes and a total of 2.5 Petabyte storage and 25 Tflops of compute power.

CEP does not have sufficient capacity to provide long term storage for LOFAR data or to carry out all processing for the generation of scientific data products. This is provided by the LOFAR LTA.

## Processing Framework

The processing framework is powered by Astro-Wise. It integrates a central catalog and client user interfaces with LOFAR processing pipelines running on Grid based and University HPC clusters.

### ( Control

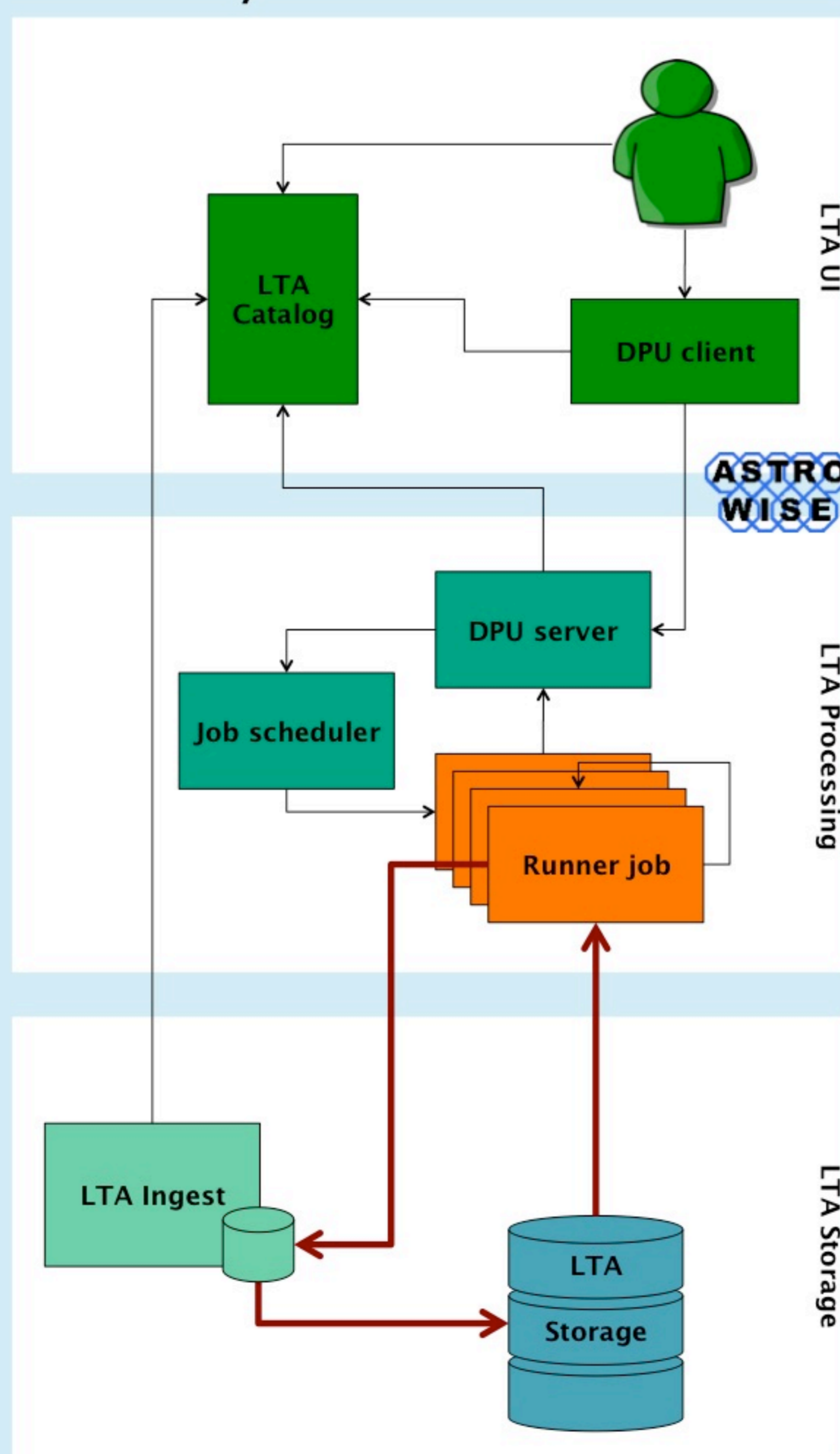
- ( User defines and initiates workflow
- ( DPU client sends jobsript to server
- ( Catalog: security and persistence

### ( Processing

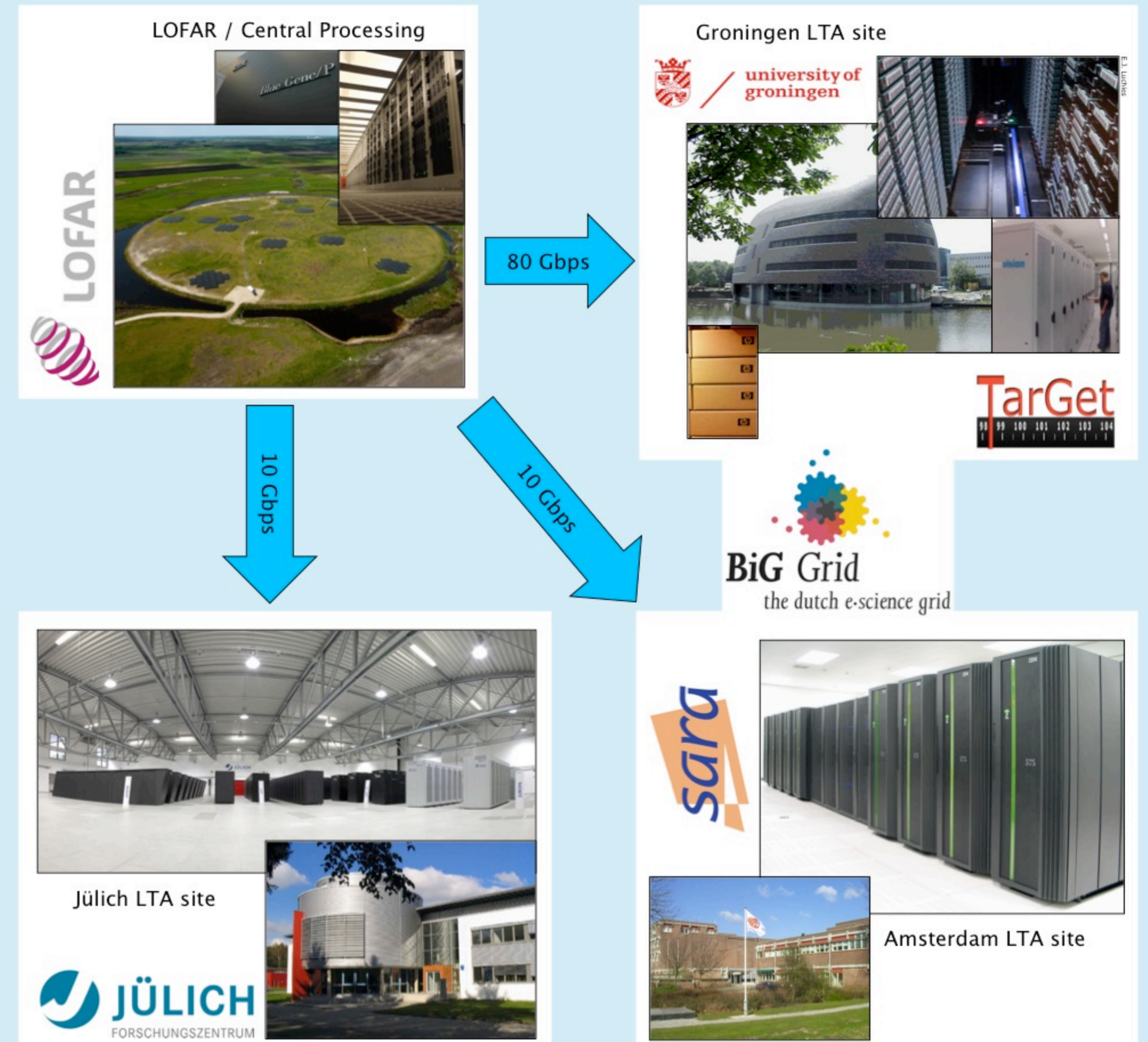
- ( DPU server submits pilot jobs
- ( Multiple schedulers supported
- ( Runner jobs contact server for work
- ( Retrieve data at start
- ( Final output ingested

### ( Storage

- ( Random access requires customization
- ( Streaming processing possible
- ( Special area to store data before ingest



## LOFAR Long Term Archive



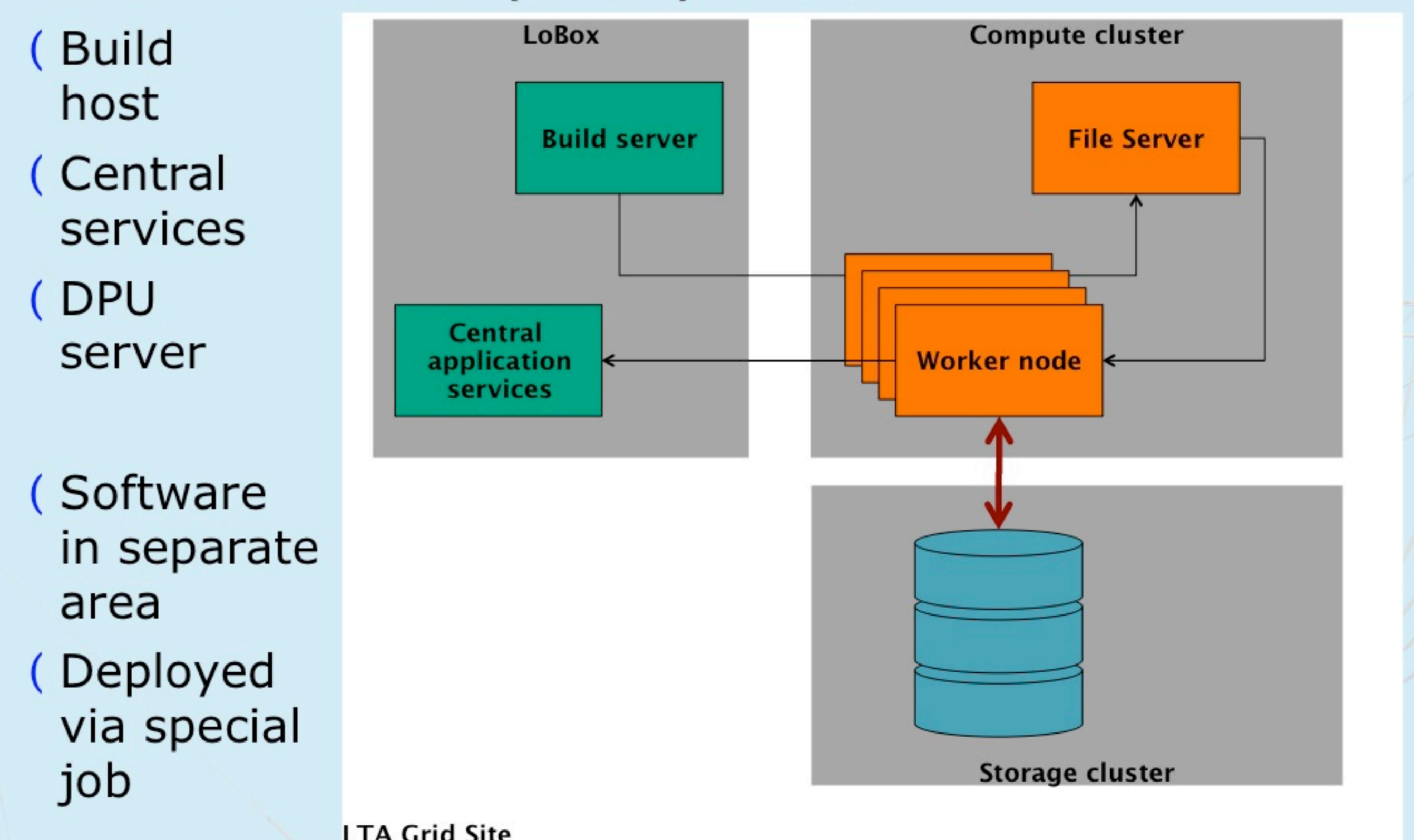
The LOFAR Long Term Archive (LTA) provides integrated services for data analysis as well as long term preservation of astronomical datasets and their provenance. By unlocking the power of international HPC infrastructure, potential science output is increased dramatically.

Challenges for processing framework:

- ( Distributed
  - ( Custom central (file) catalog needed
  - ( Heterogeneous
  - ( Data locality matters
- ( Shared environment
  - ( Little control over compilers, libraries, paths, etc
  - ( Obtaining access to sufficient nodes for large jobs challenging
  - ( Special solutions required for e.g. databases & application services
- ( Grid infrastructure
  - ( Not well tuned to few large parallel jobs; no standard MPI support
  - ( Strict A&A
  - ( No direct access to storage

## Build environment (Grid)

More control over software environment provided via LOFAR VO box (LoBox) on site.



Netherlands Institute for Radio Astronomy



ASTRON is part of the Netherlands Organisation for Scientific Research