PUNCH4NFDI --- FZJ IN TA5

21 FEBRUARY 2022

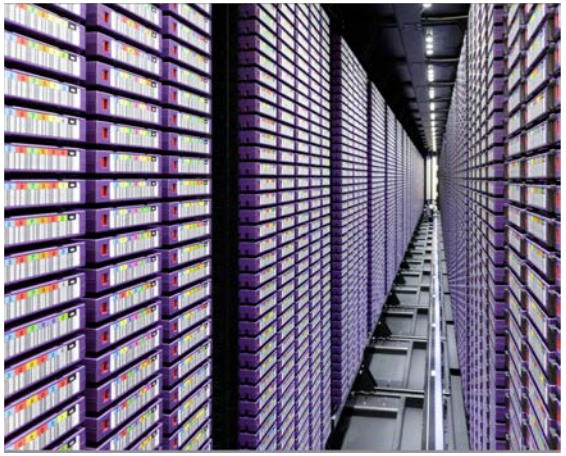
SUSANNE PFALZNER

JÜLICH SUPERCOMPUTING CENTRE

High Performance Computing meets Data Science

Jülich Supercomputers

Jülich LOFAR station



LOFAR Archive

TA 3
Computing and machine learning
TA 5

½ PhD position

TA 2
Storage and data provision
TA 5

½ postdoc position for 3.75 years

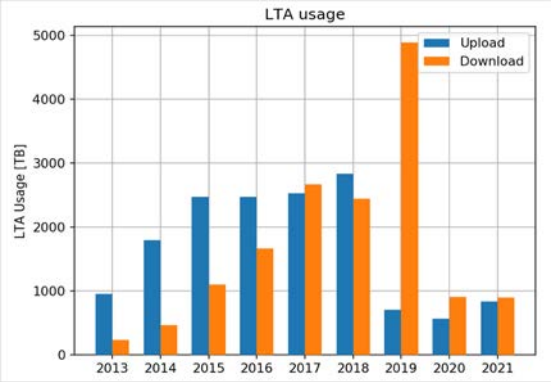
JÜLICH Forschungszentrum

STORAGE OF BIG DATA – THE LOFAR LONG-TERM ARCHIVE

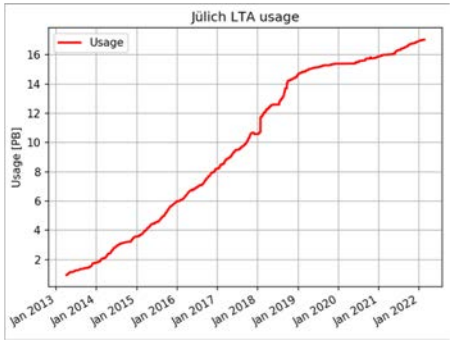
LTA consists of 4 components:

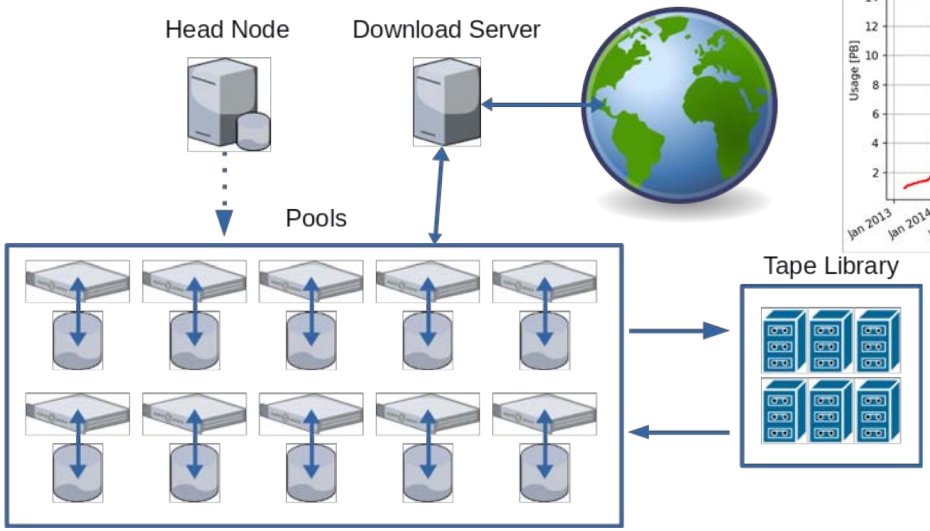
- Head node to control everything
- Pools as a temporary storage for up- and downloads
- Download Server from where end-users can download data
- Tape library for permanent “cold” storage

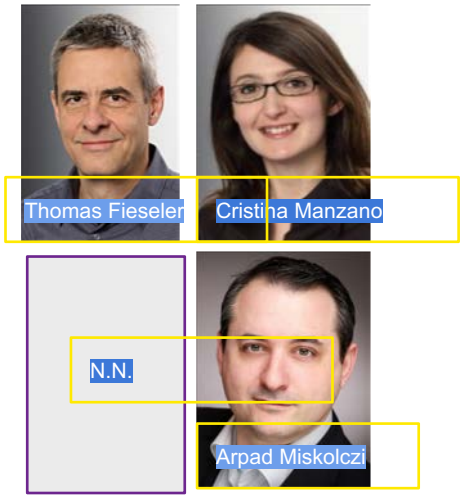
Yearly upload/download statistics of the LTA



- Continuous rise with ~2 PB/year until 2019, when data started to be compressed and Poznań LTA started ingesting data.
- Increased downloads in 2019 due to LoTSS release.

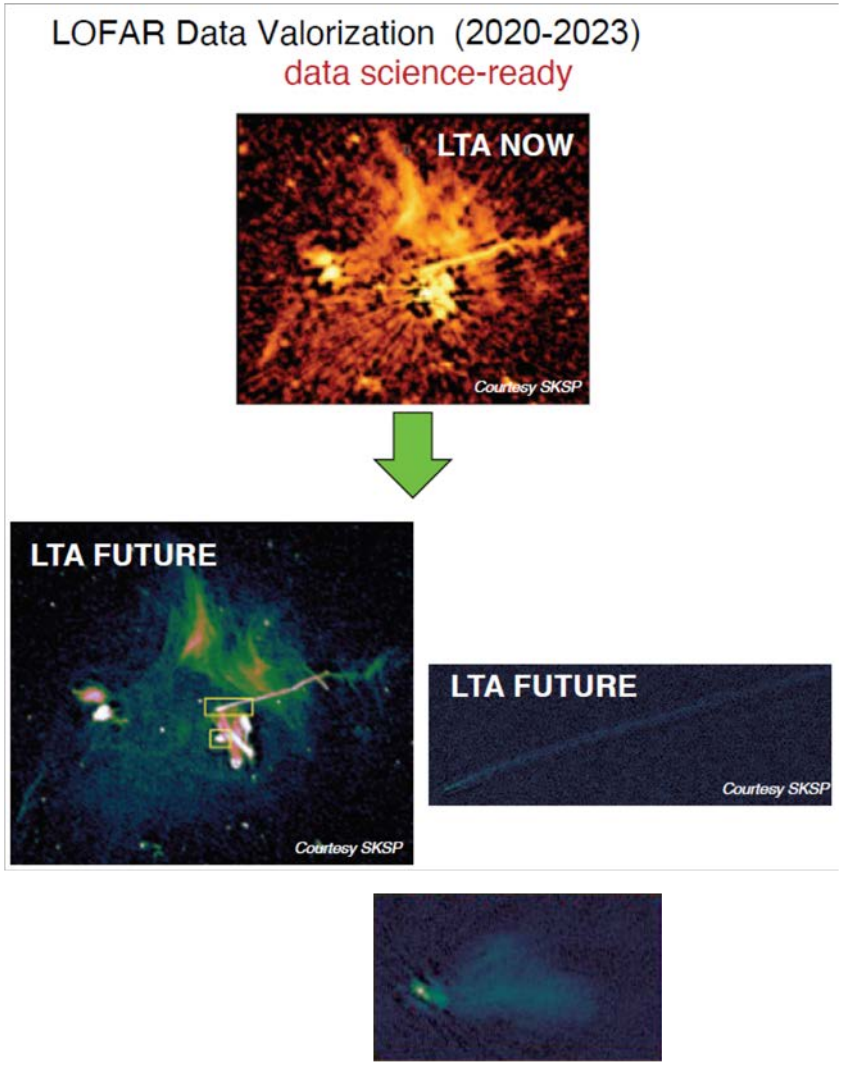






Growth of the LTA over time

... WHEN DATA GET EVEN BIGGER



… innovative storage technologies and approaches are needed

- Faster storage, ie. SSD drives
- Different storage approaches, i.e. GPFS, Object storage
- Higher capacity tape drives
- Data specific compression
  - i.e. LOFAR implements DYSCO compression specifically developed for LOFAR data -> compresses data by a factor of up to 6 while only increasing noise by ~1%

INDENTIFICATION OF CLUSTERS ON DISTRIBUTED NODES

Use case: Idenfication of young star clusters

Aim: Perform task for entire GAIA data set
Problem: Huge number of stars

Method: Machine learning
Calculation on different nodes

Solution: Use HeAT tool

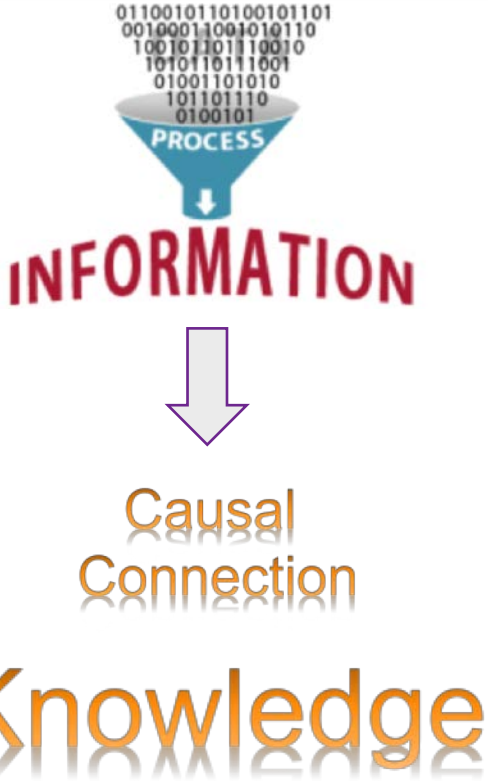
Challenge: Identify objects with information stored on different nodes

Especially hard: Young clusters are highly dynamic requires backtracking over different nodes

TA 3
Computing and machine learning
TA 5

Project start: April 2022
Position in TA 5: 0.5 PhD position

MACHINE LEARNING AND KNOWLEDGE GAIN



MACHINE LEARNING DOES NOT AVOID
HUMAN MISTAKES, BUT CAN HARDWIRES THEM

MODEL ASSUMPTIONS
CAN CORRUPT DATA ANALYSIS

DATA CHALLENGE REQUIRES
LEARNING FROM OUR MISTAKES NOW

OVERCOMING ONES PREJUDICES PAYS

No positions funded by PUNCH
Contribution completely in kind