

PROviding Computing solutions for ExaScale ChallengeS

Unlocking the LOFAR LTA

Hanno Spreeuw, Souley Madougou, Ronald van Haren, Berend Weel, Adam Belloum, Jason Maassen

24th Sep 2019, eScience conference – EINFRA, San Diego



his project has received funding from the European Union's Horizon 2020 research and innovation ogramme under grant agreement No 777533

Ŵ

netherlands













PRÖCESS LOFAR: Low Frequency Array



LOFAR radio telescope – is a "distributed software telescope" consisting of ~88.000 antennas in ~51 stations scattered over Europe. It produces up to **1.6 TB/s of raw data**, processed real time to combine the signals of multiple antennas (correlation). **This results in up to 35 TB/h of intermediate data** (visibilities) which is stored for further analysis.

PRÖCESS LOFAR: Key Science Projects



Slide courtesy of Roberto Pizzo (ASTRON). http://www.tauceti.caltech.edu/science-at-low-frequencies-2016/slides/A/02_Pizzo.pdf

PRÖCESS LOFAR LTA: Long Term Archive



	но	ME SEA	ARCH DATA	BROWSE PROJEC	CTS HELP										
						Ob	servation	1 to 100) (show edit colur 2 3 4 5 6 7	ing 100 ^{nns} 8 9 10 11	of total 3:	2940) -			
#		Project	Release Date	Instrument Filter	Channel Width [MHz]	Bits per Sample	Number Of SubArray Pointings	Duration [s]	Nr Stations Core	Nr Stations Remote	Nr Stations International	Number Of Stations	Number Of Correlated DataProducts	Number Of BeamFormed DataProducts	Creation Dat
1	×	LC10_024		110-190 MHz	vidth [MHz]	Sample	Pointings	[s] [1]	Core	Remoté	International	Stations	DataProducts	DataProducts	Creation Dat
2	×	LC10_024		110-190 MHz	0.000000	8	1	951.0	24	14	12	50	243	0	2018-10-26 11:56:44
2	*	LC10_024		110-190 MHz	0.000000	8	3	14760.0	24	14	12	50	486 / 487	0	2018-10-26 10:07:54
3	×	LC10_024		110-190 MHz	0.000000	8	1	951.0	24	14	12	50	243	0	2018-10-20 08:59:42
4	×	LT10_010		110-190 MHz	0.000000	8	1	950.0	24	14	12	50	243	0	2018-10-25 13:23:14
5	×	LT10_010		110-190 MHz	0.000000	8	3	29170.0	24	14	12	50	486 / 487	0	2018-10-25
6	×	LT10_010		110-190 MHz	0.000000	8	1	951.0	24	14	12	50	243	0	2018-10-25
7	×	LC10_018		110-190 MHz	0.000000	8	1	960.0	24	14	11	49	244	0	2018-10-25
8	×	LC10_018		110-190 MHz	0.000000	8	7	29180.0	24	14	11	49	388	0	2018-10-25
9	×	LC10_018		110-190 MHz	0.000000	8	1	951.0	24	14	11	49	244	0	2018-10-25
10	× (DDT10_006	2019-10-22	110-190 MHz	0.000000	8	1	3161.0	23	0	0	23	0	240	2018-10-20
11	× (DDT10_006	2019-10-22	110-190 MHz	0.000000	8	1	641.0	23	0	0	23	0	240	2018-10-20
12	×	LC10_015		110-190 MHz	0.000000	8	1	2141.0	21	0	0	21	0	1	2018-10-25
13	×	LC10_007		30-90 MHz	0.000000	8	2	29170.0	24	13	0	37	488	0	2018-10-2
14	× (DDT10_006	2019-10-22	110-190 MHz	0.000000	8	1	641.0	23	0	0	23	0	240	2018-10-20
15	×	DDT10_006	2019-10-22	110-190 MHz	0.000000	8	1	3161.0	23	0	0	23	0	240	2018-10-2
16	×	DDT10_006	2019-10-22	110-190 MHz	0.000000	8	1	640.0	23	0	0	23	0	240	15:55:27 2018-10-2

LOFAR observations are stored in the long term archive (LTA) which is distributed over Amsterdam, Juelich and Poznan. It currently contains ~43 PB of data and grows with 5 to 7 PB/year. Data is stored mostly on tape, using hard disks for temporary storage (dCache). A web interface is available to request data to download. Processing needs to be done by the astronomers themselves on local infrastructure \rightarrow this is the hard part!

PRÖCESS LOFAR LTA: Long Term Archive



- Manage data growth: compression tool (Dysco) in production since September 2018
- More than a factor of 10 between proprietary vs non-proprietary data downloads: 60 PB vs 5 PB
- Need to make (science-ready) data accessible: LOFAR efficiency + advanced pipelines in production + exposing data

Slide courtesy of Roberto Pizzo (ASTRON). https://www.astron.nl/lofarscience2019/Documents/Monday/LUM_Pizzo.pdf

PRÖCESS Problem: Big Data & Complex Software

The data size for a single observation tend to **be large**, **10's of TBs**, which are non-trivial to download and handle.

There is a lot of software available to process LOFAR observations. Usually several packages are needed for the different aspects of calibration and imaging. These often require expert knowledge to install and operate.

The combination of these two makes it **hard** for **astronomers** to make optimal use of the archive.

~	0.0			
C 0	(i) 🔒 GitHu	b, Inc. (US) https://github.com/search?q=LOFAR	♥ ☆	⊻ II\ 🛈 📽
OFAR		Pull requests Issues Marketplace Explore		≵ +-
Repositories	133	133 repository results		Sort: Best match -
Code	34K			
Commits	15K	lofar-astron/factor	Python	★ 14
Issues	724	Facet calibration for LOFAR		
Marketplace		GPL-2.0 license Updated 23 days ago		
Topics	2			
Wikis	78	CFAR solutions tool	 Python 	★7
Users	19	GPL-3.0 license Updated 2 days ago		
Languages Python C++ Shell Jupyter Notebook	78 12 7 6	pelican/pelican- <i>lofar</i> LOFAR single station processing pipeline using PELICAN Updated on Dec 17, 2016	• C++	★ 13
C CSS HTML Makefile Java	4 3 2 2 1	varenius/lofar-lb Collection of LOFAR long baseline software Updated on Mar 24, 2017	● TeX	★ 4
JavaScript	1 sheet	Iofar-astron/<i>Iofar</i>-deploy <i>LOFAR</i> software stack deployment scripts and Dockerfiles for R&D purposes GPL-3.0 license Updated on Jun 12	Shell	* 6
		lafor actron// OEAP Contributions	Dallar	



Mature, modular, generalizable open source solutions for user friendly exascale data.



PRÖCESS Reusable PROCESS services



PRÖCESS Step 1: Stage in the data



PRÖCESS Step 2: Launch the processing workflow



PRÖCESS Step 3: Stage out results



PRÖCESS Effect of Processing



RA (J2000)

RA (J2000)

IEEE eScience conference 2019, San Diego

PRÖCESS Overview of Hardware Locations



IEEE eScience conference 2019, San Diego

PRÖCESS Video Demonstration 1/3

			localhost:2015	(ال		(A)(0)(7
The go proces astron Throug hardw the po	bal of this use case is to simplify the ssing pipeline from there. For this we omical observation data. Igh this portal, the astronomer must are infrastructure available in the p	e processing of archived data. Ar ve need an easy to use, flexible, of the able to browse through the a roject. Data should then be trans	2: SKA/L stronomers should be able to se efficient and scalable workflow in available datasets and available of ferred from the LTA to the proce	Lect a dataset on a portal, select infrastructure for processing of ex- workflows, and launch processin ssing infrastructure, processed,	a workflow, and then launch the ktremely large volumes of g directly from there to the and the results made available in	
	STARTTIME -	ENDTIME -		DECLINATION	NR_SUBBANDS	
	\$	\$	\$	\$	< ‡	
	Enter STARTTIME	Enter ENDTIME	Enter RIGHTASCENSI	Enter DECLINATION 3	3	
6	2012-02-04T18:48:00.000Z	2012-02-04T18:48:00.000Z	111.42857125	90	1.	
6	2012-02-18T09:03:00.000Z	2012-02-18T09:03:00.000Z	8.57142875	90	1	
6	2013-04-01T02:06:01.000Z	2013-04-01T02:06:01.000Z	243.529411667	90	1.	
6	2012-02-10T23:33:03.000Z	2012-02-10T23:33:03.000Z	159	90	1	
6	2012-02-10T23:48:01.000Z	2012-02-10T23:48:01.000Z	159	90	1	
6	2012-02-11T05:48:01.000Z	2012-02-11T05:48:01.000Z	273	90	1	
6	2012-02-20T06:03:07.000Z	2012-02-20T06:03:07.000Z	225.714285833	90	1	
	2012-02-10720-48-00.0007	2012-02-10720:48:00.0007	159	90	1	

PRÖCESS Video Demonstration 2/3

Waiting: 1 Running: 0 Successful: 1 Errored: 0 All Jobs Delete All Jobs Filter: P23ing Started: 2019-05-03 10:28:15 Success P23ing Started: Waiting	Waiting: 1 Running: 0 Successful: 1 Errored: 0 All Jobs Delete All Jobs Filter: P23ing Started: P23ing Started: Waiting	Xenonflow Admin	Server Status:				
All Jobs Delete All Jobs Filter: I I I I I I I I I I I I I I I I I I I	All Jobs Delete All Jobs Filter: I I I I I I I I I I I I I I I I I I I		Waiting: 1	Running: 0	Successful: 1	Errored: 0	
All Jobs Delete All Jobs Filter: P23img Started: 2019-05-03 10:28:15 Success P23img Started: Waiting	All Jobs Delete All Jobs Filter: III I I I I I I I I I I I I I I I I I						
Delete All Jobs Filter: P23img Started: 2019-05-03 10:28:15 Success p23img started: Waiting	Delete All Jobs Filter: Image: Control of the second se	All Jobs					
Filter: III I I I I I I I I I I I I I I I I I	Filter: III I I I I I I I I I I I I I I I I I	Delete All Jobs					
III Image: Constraint of the second se	II Image: Started: P23img Started: Started: Waiting	Filter:					
Filter: p23img Started: 2019-05-03 10:28:15 Success p23img Started: Waiting	Filter: p23img Started: 2019-05-03 10:28:15 Success p23img Started: Waiting	II 🌣 🗸 🛇 X 9 A					
p23img Started: 2019-05-03 10:28:15 Success p23img Started: Waiting	p23img Started: 2019-05-03 10:28:15 Success p23img Started: Waiting	Filter:					
Started: 2019-05-03 10:28:15 Success p23img Started: Waiting	Started: 2019-05-03 10:28:15 Success p23img started: Waiting	p23img					
Success p23img Started: Waiting	Success p23img Started: Waiting	Started: 2019-05-03 10:28:15					
p23img Started: Waiting	p23img Started: Waiting	Success					
Started: Waiting	Started: Waiting	p23img					
Waiting	Waiting	Started:					
•		Waiting					
		•					
		L					

PRÖCESS Video Demonstration 3/3



PRÖCESS Conclusions

- There is a need for <u>user-friendly access</u> to the LTA's data and subsequent processing.
- Deliverance of a demo using <u>reusable services</u>.

• Future work:

- Making workflow processing possible on several clusters connected via <u>high-bandwidth networks</u>.
- Further acceleration of the processing pipeline through parallel computing (direction dependent).





"Now I can finally spend my time on actual science!" Dr. Hanno Spreeuw



PROviding Computing solutions for ExaScale ChallengeS

Unlocking the LOFAR LTA

Hanno Spreeuw, Souley Madougou, Ronald van Haren, Berend Weel, Adam Belloum, Jason Maassen

24th Sep 2019, eScience conference – EINFRA, San Diego



his project has received funding from the European Union's Horizon 2020 research and innovation ogramme under grant agreement No 777533

Ŵ

netherlands











