



Observatoire de Paris



A distributed data-mining software platform for extreme data across the compute continuum



Radio imaging cloud workflow for multiband transient source reconstruction

J. N. Girard (1,2), F. Nammour (3), B. Cecconi (1,2) (1) ORN, Observatoire de Paris, Université PSL, Univ. Orléans, CNRS (2) LIRA, Observatoire de Paris, Université PSL, Sorbonne Université, CNRS (3) Ekinox, 26 boulevard d'Argenson, 92200 Neuilly-sur-Seine

Abstract

New and large radio interferometers such as the upcoming SKA (Square Kilometre Array) deliver massive and ever increasing amount of raw data for astrophysical source observations. For this purpose, it is crucial to have a workflow that enables efficient transferring and processing data, in a compute continuum context.

PSL

EXTRACT Project

Extreme Data Cloud . Volume, Velocity and Variety (3V challenge) Complexity and Diversity Sparse, Missing, Insufficient Data with Extreme Variability Dispersity of Data Sources

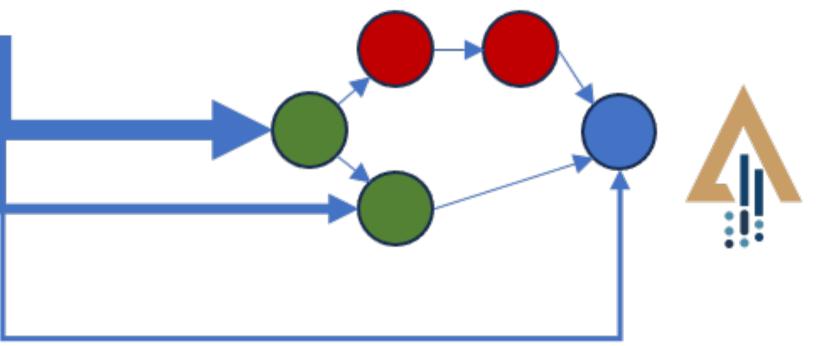
Compute Continuum	
HPC	Data Mining Software Platform
Edge	 Data infrastructures and AI &

All the components should be seamlessly orchestrated while enabling the development, test and run **custom workflows**. Data can be hosted on various sites and the computing ressources can be spread country-wide and even internationally (e.g. Future SKA Region centers). In the context of the **EXTRACT** project, the development of interconnections between facilities allows a seamless usage of data centers, HPC and computing clusters on the cloud.

We present two use cases: a demonstrator of the workflow design implemented between facilities, and a new deep learning imager for radio interferometric imaging to reconstruct radio transient sources using this framework (remote data, remote registry, HPC ressources using, etc)

Big-data frameworks

- Data-driven orchestration
- Interoperability



Current situation

No existing compute continuum workflow that handles massive & multi-site data No satisfying method to process reconstruct huge dynamic data flows Increasing amount of astronomical observations data rate, e.g. 20 Tb/s for SKA-Mid [1]



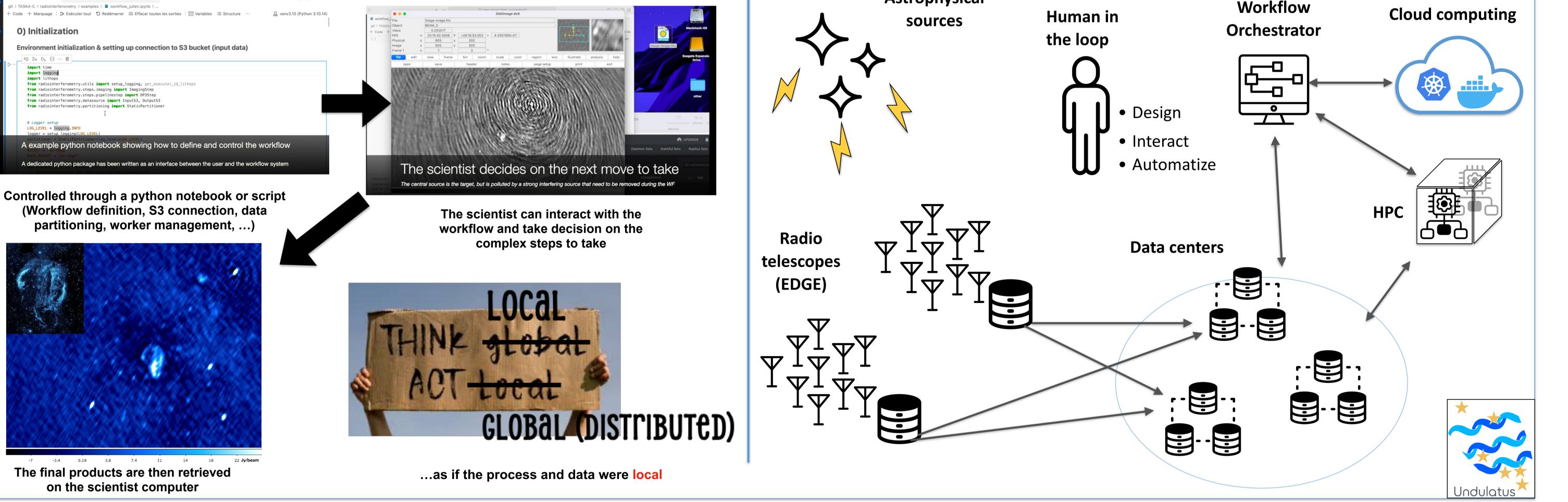


TASKA **Transient Astrophysics with an SKA pathfinder**

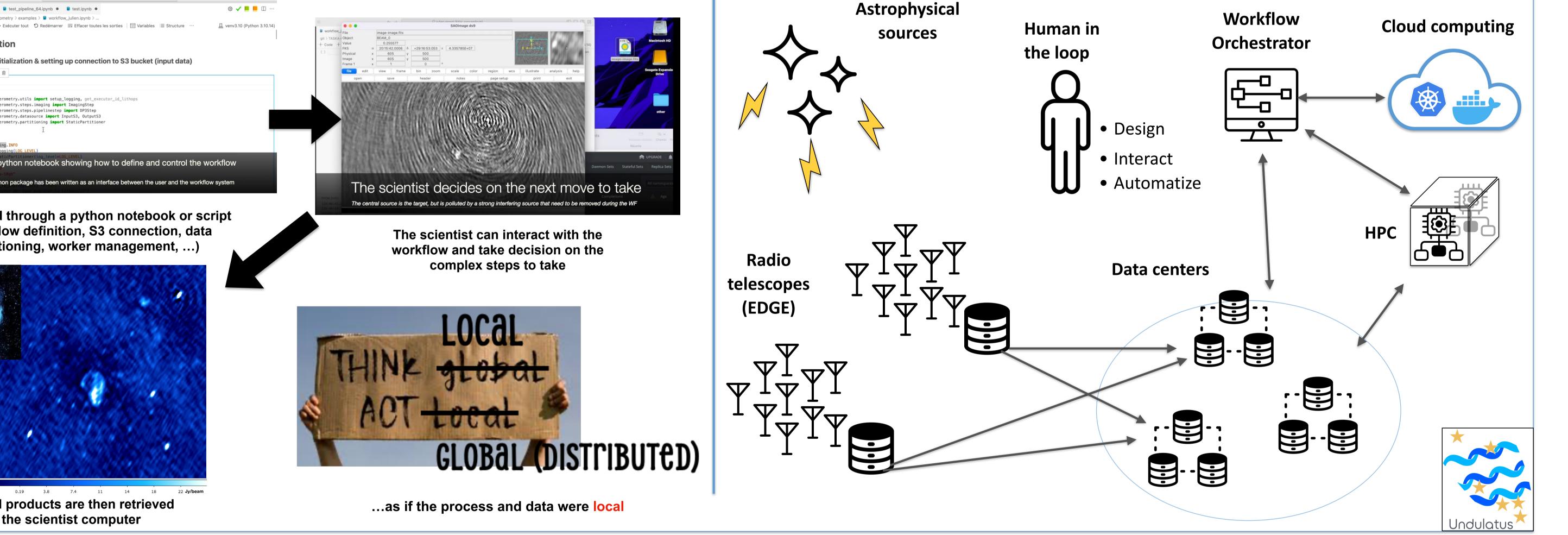
Workflow management from the user side

workflow_julien.ipynb • 🔋 test_pipeline_64.ipynb • 🔋 test.ipynb •	🌐 🗸 📕 📒 🗆 …
git > TASKA-C > radiointerferometry > examples > 🕒 workflow_julien.ipynb >	
🔘 🕂 Code 🕂 Marquage ⊳ Exécuter tout 🖱 Redémarrer 🗮 Effacer toutes les sorties 🗔 Variables 🗮 Structure \cdots	🚊 venv3.10 (Python 3.10.14)
O) Initialization	

import loggin



Undulatus workflow

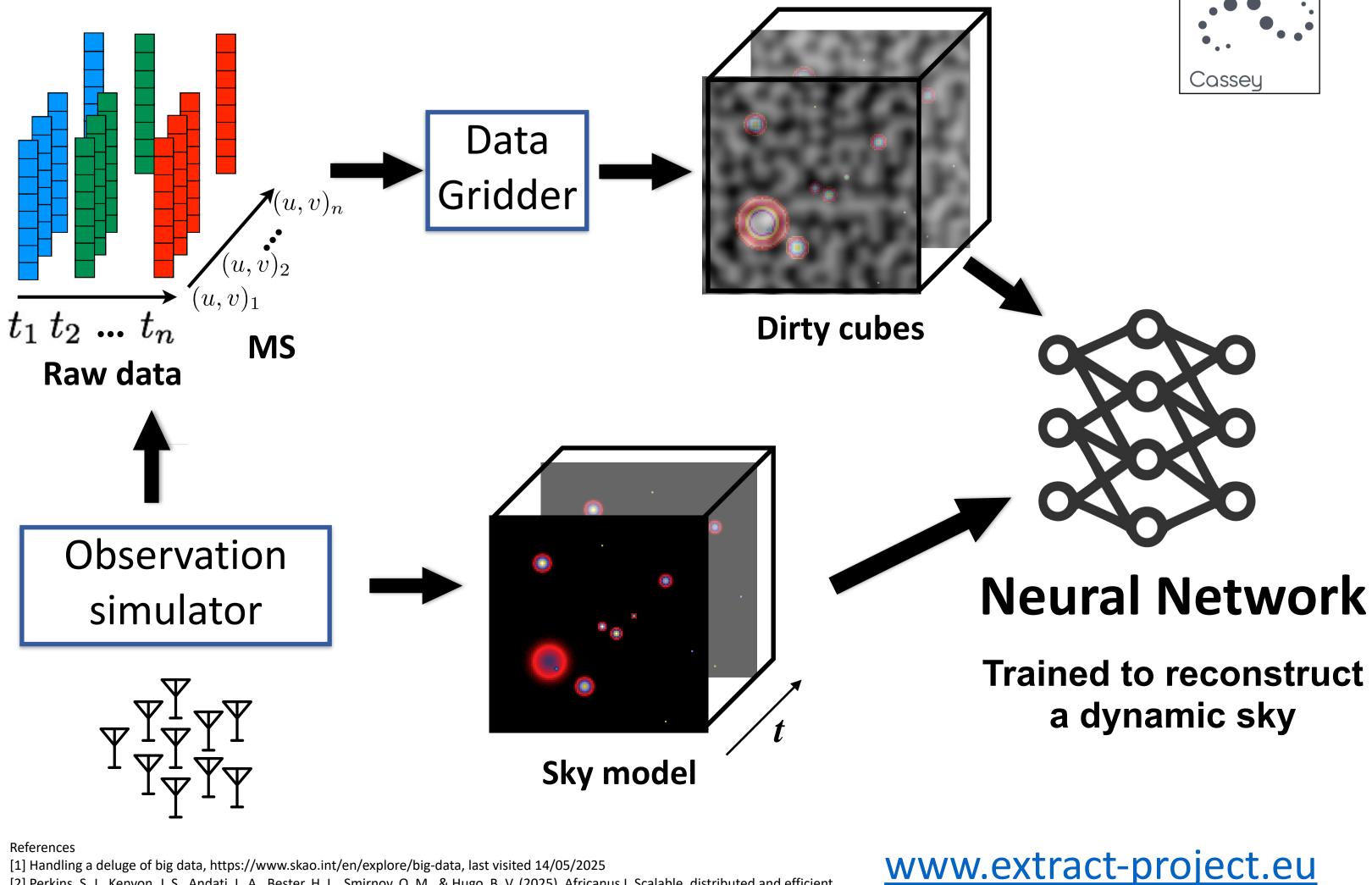


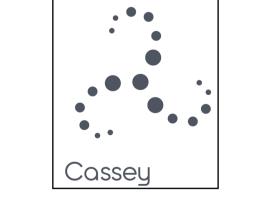


Cassey imager and training pipeline

[2] Perkins, S. J., Kenyon, J. S., Andati, L. A., Bester, H. L., Smirnov, O. M., & Hugo, B. V. (2025). Africanus I. Scalable, distributed and efficient

radio data processing with Dask-MS and Codex Africanus. Astronomy and Computing, 52, 100958.





Noise>						
t=27						
t=28				-		
t=29						
t=30	10000	100000	Sec. 87 No.	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		

extract_eu_proj

- At low SNR, classical imagers have problematic reconstruction and cannot leverage temporal structures in signals
- The network we chose is able to capture the coherent time structure of the signal

Realistic simulation of observations (source)

built using Codex Africanus library [2]

Currently, the network is trained for one specific

visibility, instrumental PSF, data rate and format)

Future works & next steps

* * * * ** @extract-eu-project

telescope

Extended method to emission transients

Public release of open-source code

<u>Contact:</u> fadi.nammour@obspm.fr

Test case on real data

The EXTRACT Project has received funding from the European Union's Horizon Europe programme under grant agreement number 101093110