SWOT HELPDESK

Accelerate access and processing of SWOT data

R. CHEVRIER, A.-S. TONNEAU, G. DIBARBOURE

Contents

Contents

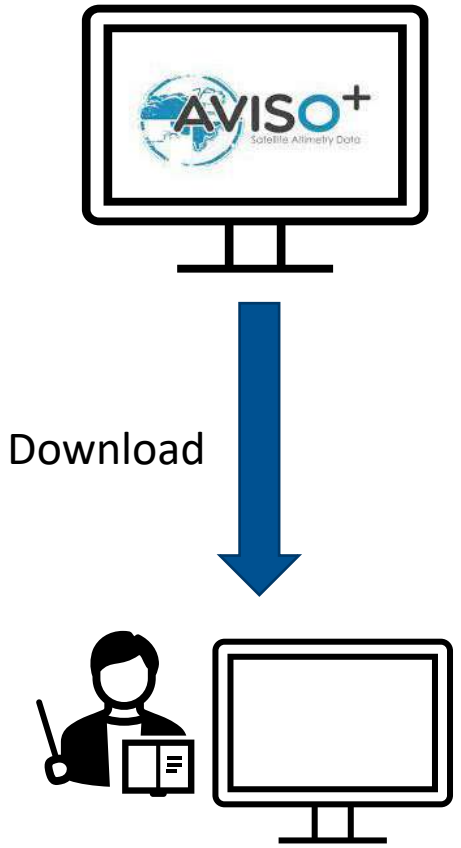
SWOT LR KaRIn ocean products

Product	Grid	Volume		Availability
		/day	/year	
L1B_LR_INTF	KaRIn Metering Structure Frame (KMSF)	1 TB	365 TB	HPC CNES (on demand)
L2_LR_SSH Basic	2 km geographically fixed along-track grid	1 GB	365 GB	HPC CNES / PODAAC / AVISO
L2_LR_SSH WindWave	2 km geographically fixed along-track grid	1 GB	365 GB	HPC CNES / PODAAC / AVISO
L2_LR_SSH Expert	2 km geographically fixed along-track grid	3 GB	1 TB	HPC CNES / PODAAC / AVISO
L2_LR_SSH Unsmoothed	~250 m native (center-beam) along-track grid	40 GB	15 TB	HPC CNES / PODAAC / AVISO
L3 KaRIn	2 km geographically fixed along-track grid	200 MB	73 GB	HPC CNES / AVISO
L4 KaRIn	Regular 2D-grid 1/8°	<100 MB	<36 GB	HPC CNES / AVISO

Classical data access (download + local work)

AVISO offers the following

- Data is available on AVISO : L2 (2km and 250m), L3 and L4
- Support on products format, content and algorithms
- Opentoolbox for collaborative work on swath *[NEW]*
  - ✓ Download AVISO / PODAAC / Copernicus …
  - ✓ Extraction
  - ✓ Statistics
  - ✓ Diagnostics : derivatives, currents, strain, vorticity
- Visualization via SeeWater Aviso+ *[NEW]*

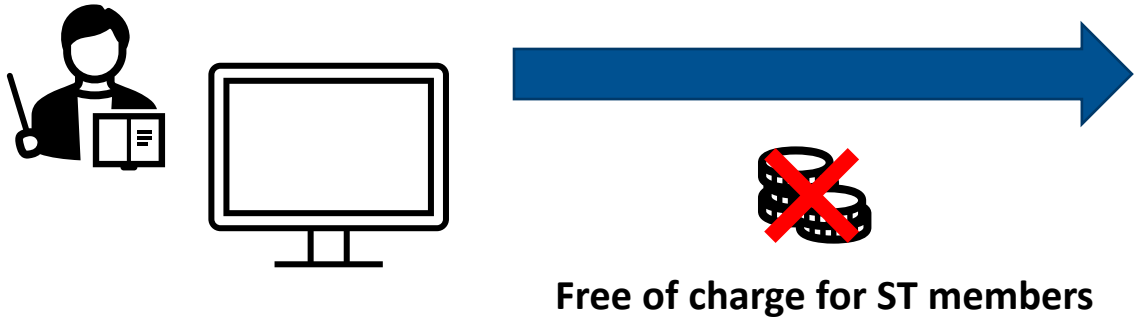


Process-centric : bring the data to the algorithm

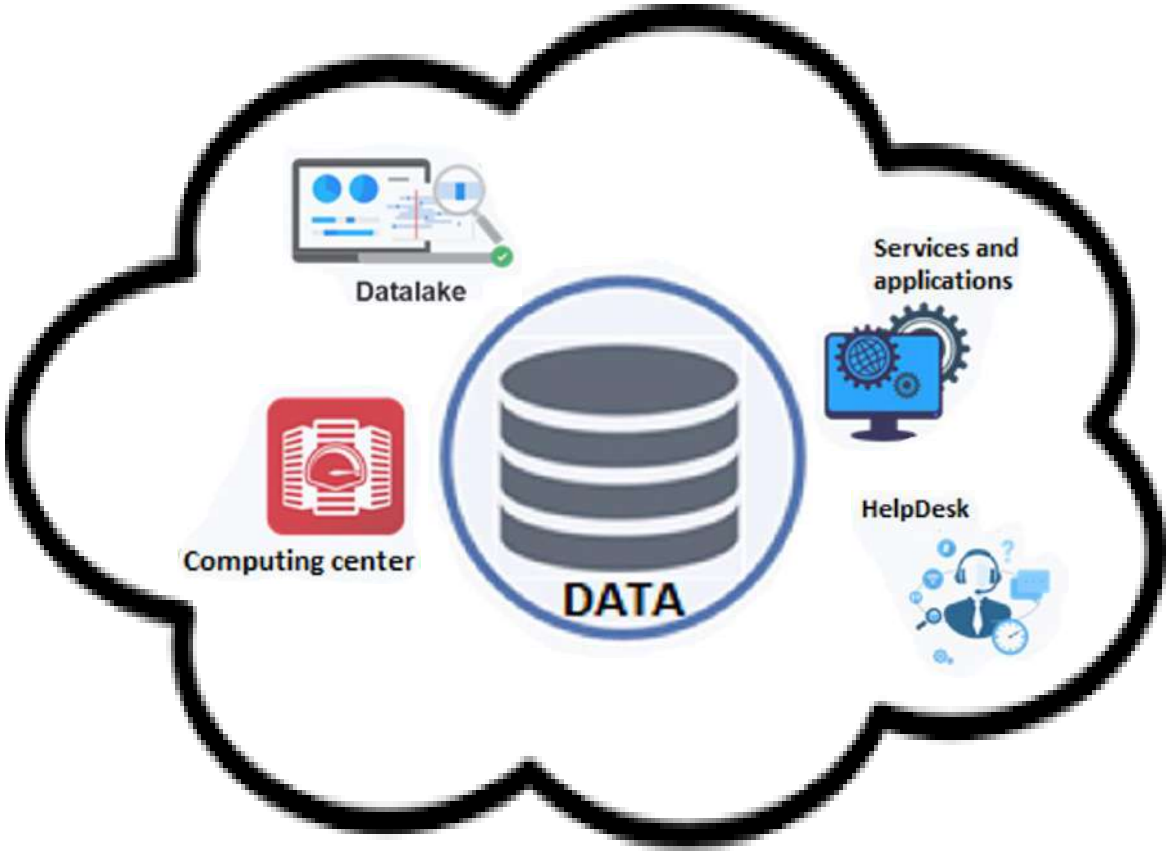
New form of data access (working remotely in the cloud)

AVISO offers the following

- Free hosting of ST Projects on CNES infrastructure in **Toulouse**
- High processing power connected to data
- Integrated SWOT data, tools and dedicated support
- External data for your research (e.g. Copernicus)



Free of charge for ST members



Data-centric: bring the algorithm to the data

Local VS Remote: pros and cons

	Criteria	CNES cloud	Local work
Invest time	Work method		
	Project setting up		
Save time	Cost		
	Computation power		
	Access to SWOT data		
	Multi-mission & multi-temporal analysis		
	SWOT dedicated tools		
	Software support		
	Contact with SWOT experts		

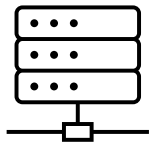
Investment needed

- Big
- Medium
- Light

Contents

Simplified HPC architecture

Processing power



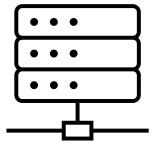
Interactive sessions

Small: 1 core – 8 GB RAM

Medium: 4 cores – 32 GB RAM

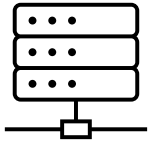
Large: 8 cores – 64 GB RAM

XLarge: 16 cores – 128 GB RAM



Batch nodes
CPU

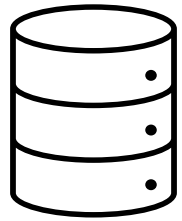
16 000 cores – 115 TB RAM



Batch nodes
GPU

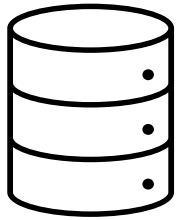
48 GPU (A100, A300, V100)

Storage capacity



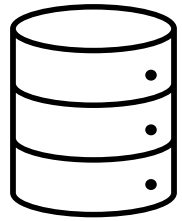
Project space (GPFS)

5-10 TB



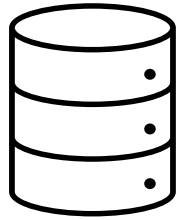
Shared Datalake (S3)

70 PB



User Space (NFS)

10 GB



Scratch (GPFS)

50 GB

until 2TB during few days

CPU : Central Processing Unit

Cheap and computations historically done on this hardware

GPU: Graphical Processing Unit

Used for machine learning, interpolation… but expensive

GPFS : General Parallel File System

High performance file system made for distributed architectures

S3: Simple Service Storage

Optimized for availability and scalability

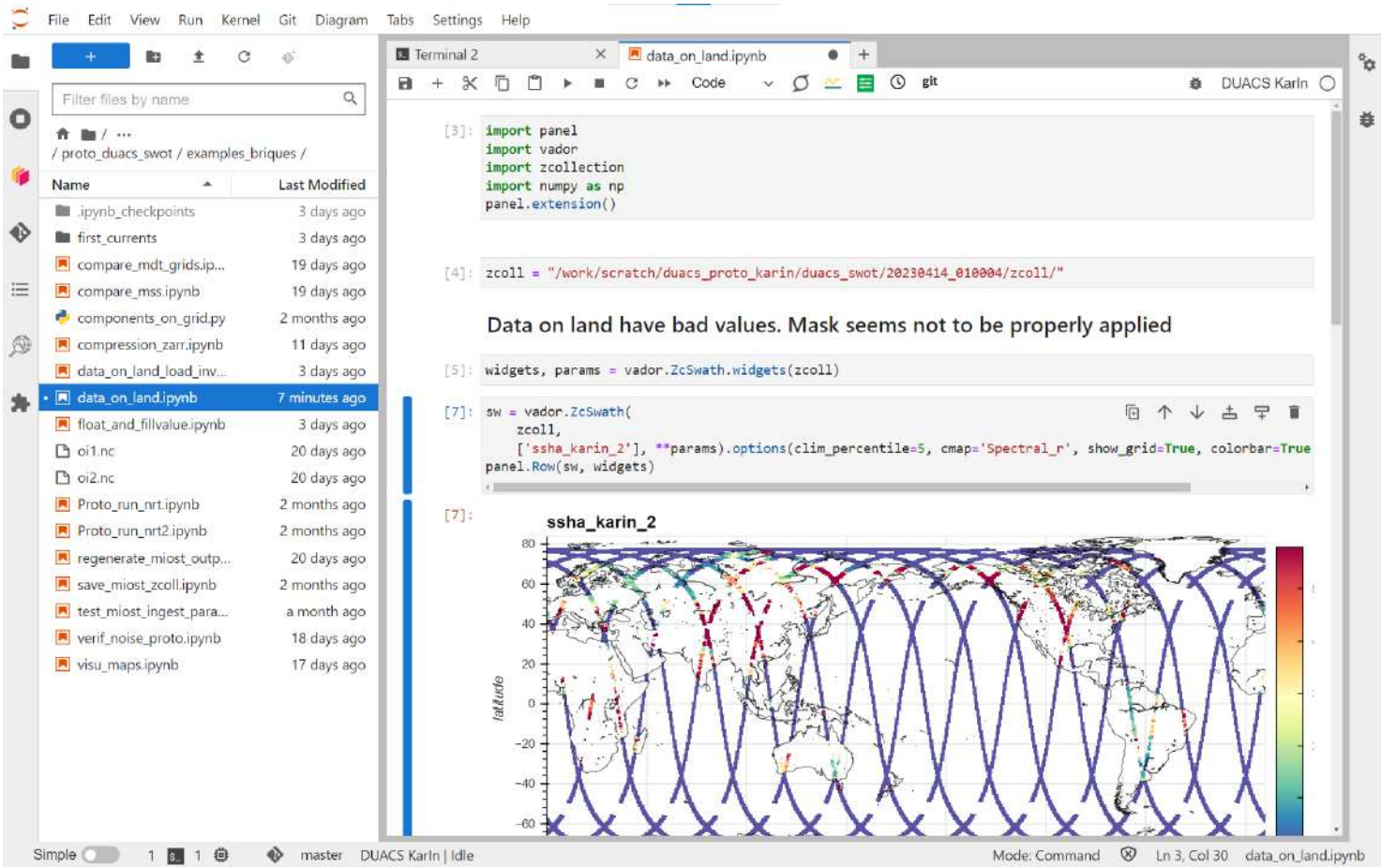
Work methods to work on CNES cluster

JupyterLab

https://jupyterhub.cnes.fr/

Access to a set of tools :

- Complete access through a web browser
- File explorer
- Python Notebooks to work on SWOT data
- Examples notebooks and tutorials dedicated to SWOT data
- Documentation



More information:
https://www.dataquest.io/blog/jupyter-notebook-tutorial/
https://jupyter-notebook.readthedocs.io/en/latest/notebook.html

Other methods to work on CNES cluster

Remote desktop

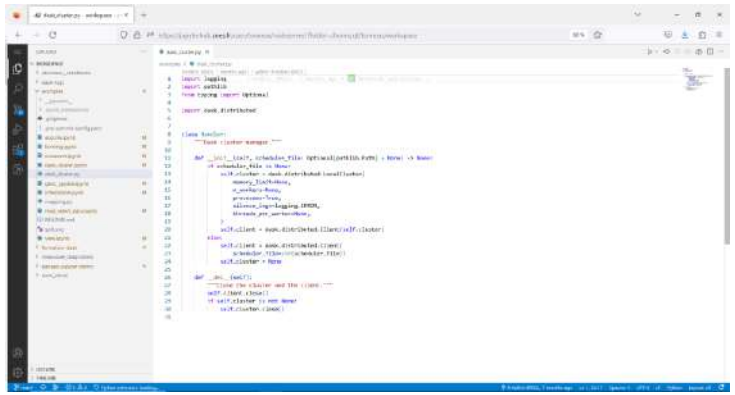
Very simple access:

https://jupyterhub.cnes.fr/user/login/desktop/vnc.html/

Command line

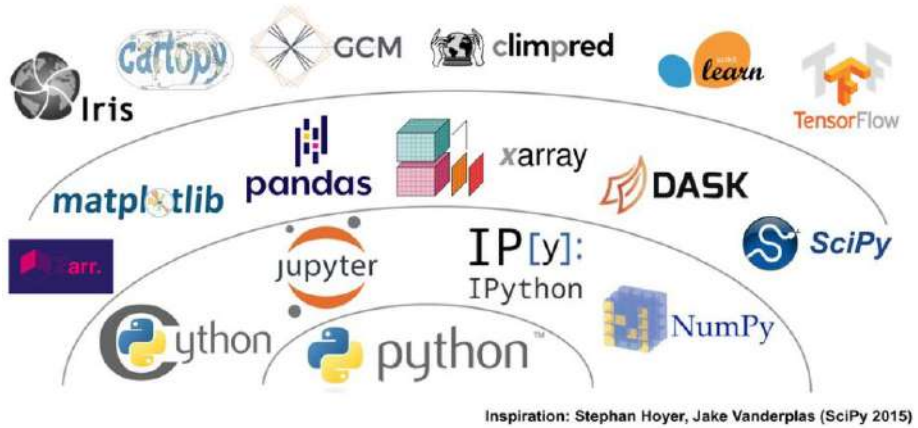
```
ssh -Y login@hal.cnes.fr
```

IDE - VSCode



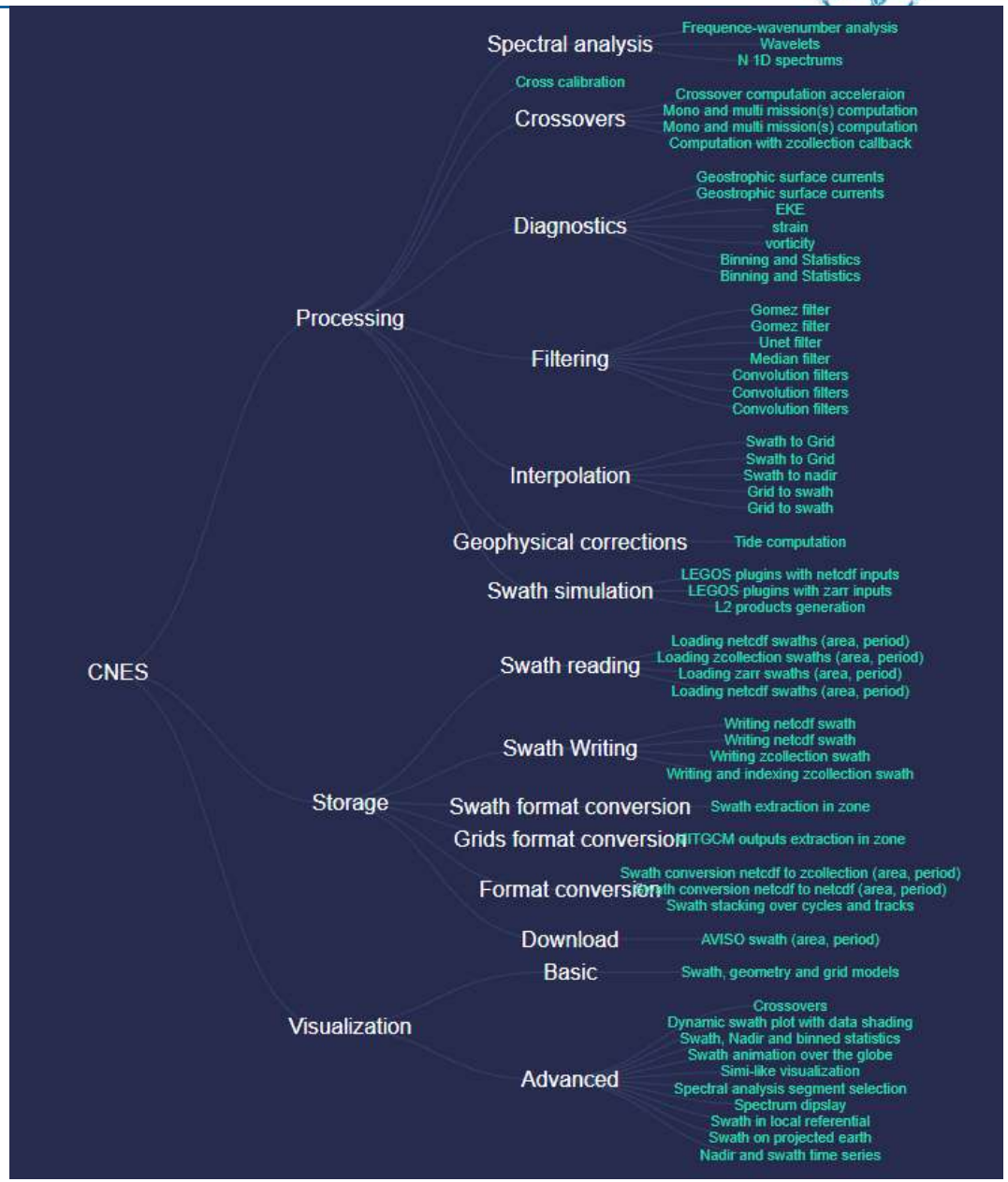
Tools – Libraries and use cases

Python PANGEO Stack : pangeo.io
(research-oriented python environment)

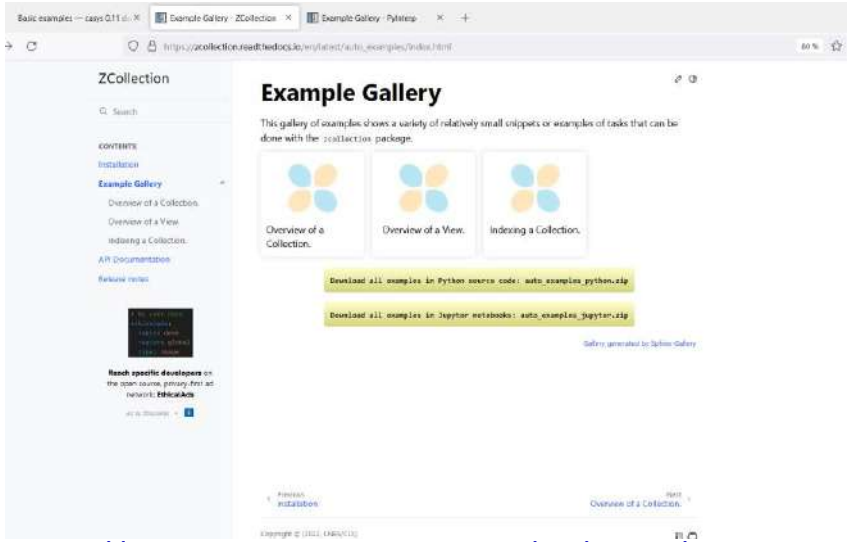


SWOT-specific additions to the PANGEO stack

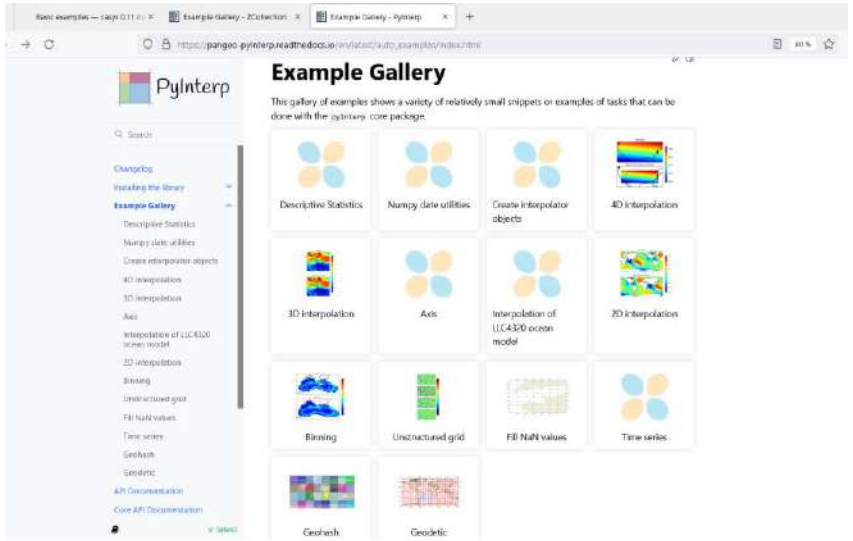
- Zcollection
- Swot Calval
- Pyinterp
- Casys
- Vador
- Ocean tools



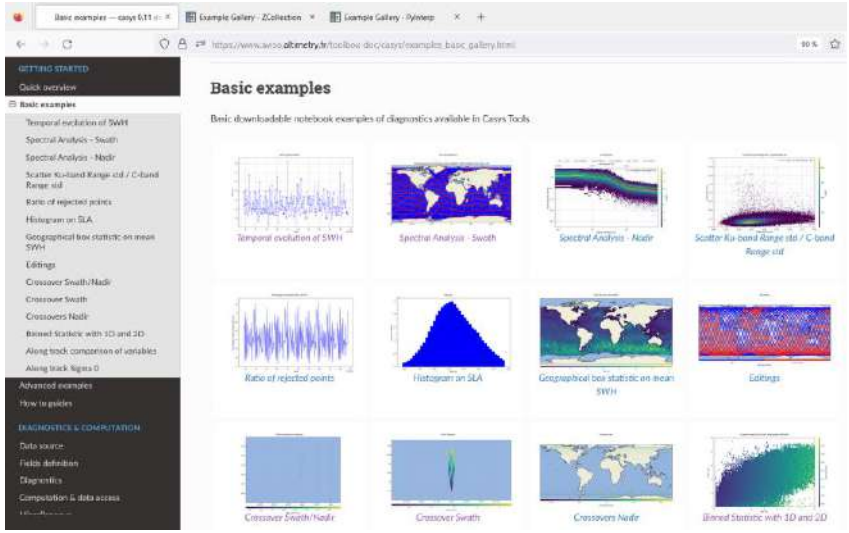
Tools – A lot of use case examples available



https://zcollection.readthedocs.io/en/latest/auto_examples/index.html

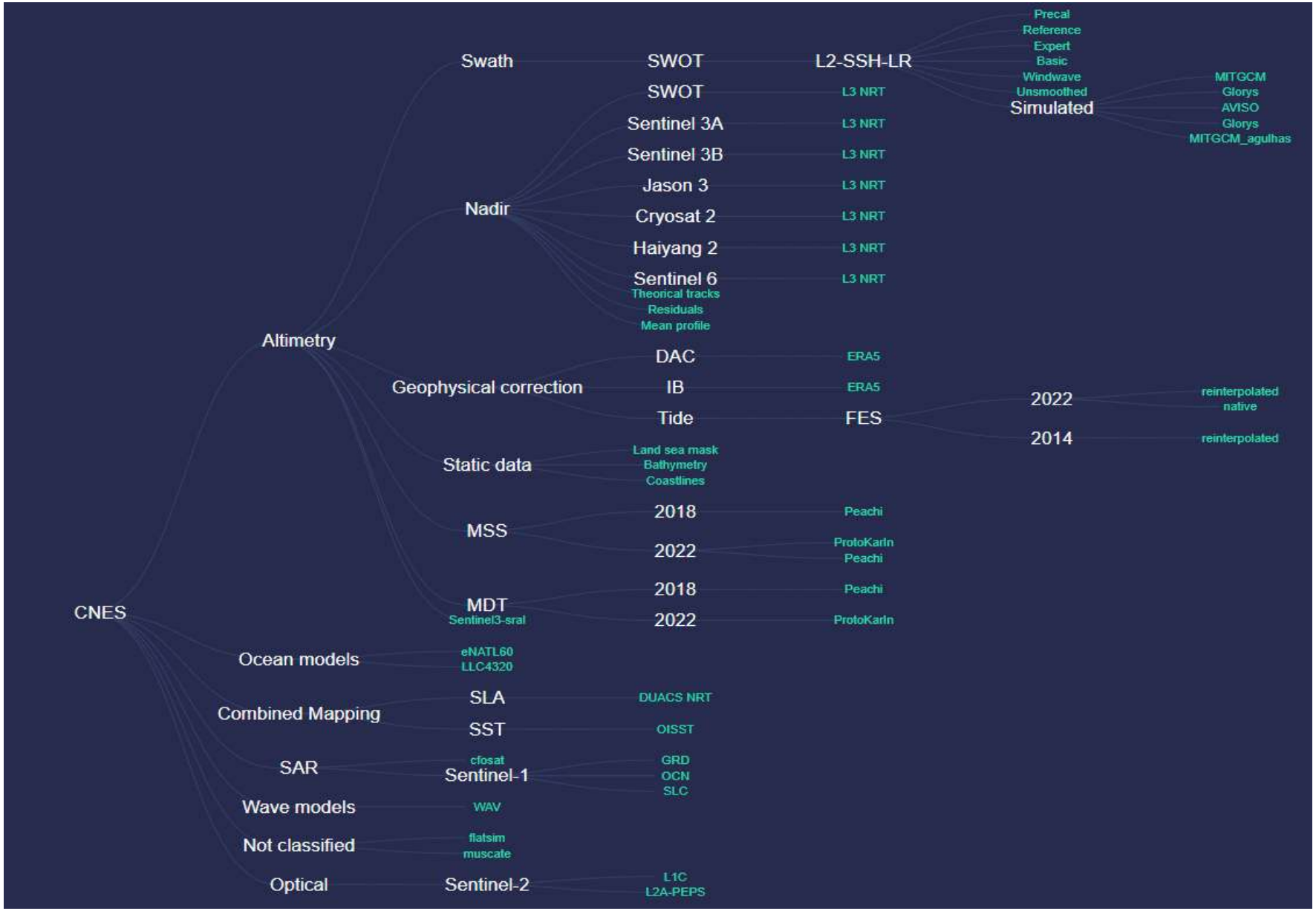


https://pangeo-pyinterp.readthedocs.io/en/latest/auto_examples/index.html



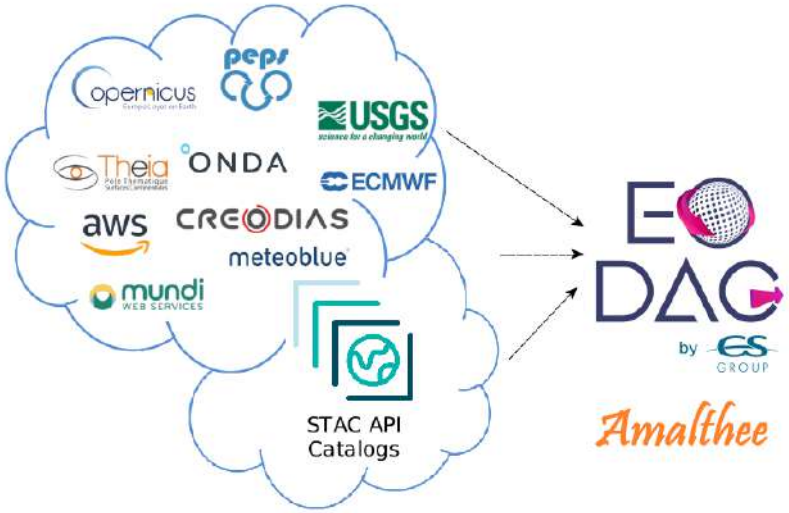
https://www.aviso.altimetry.fr/toolbox-doc/casys/examples_basic_gallery.html

Data already on the CNES cloud



External data retrieval – Data from other catalogs

External data repositories



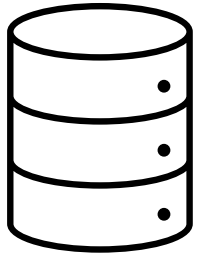
EODAG by ES GROUP

Amalthee

Catalog requests and download tools

Occasional or regional download
Small volumetry

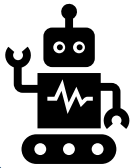
User

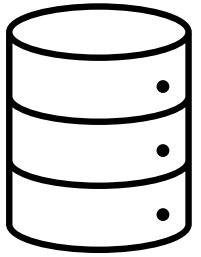


Project Space

Regular download
High volumetry

CNES Helpdesk



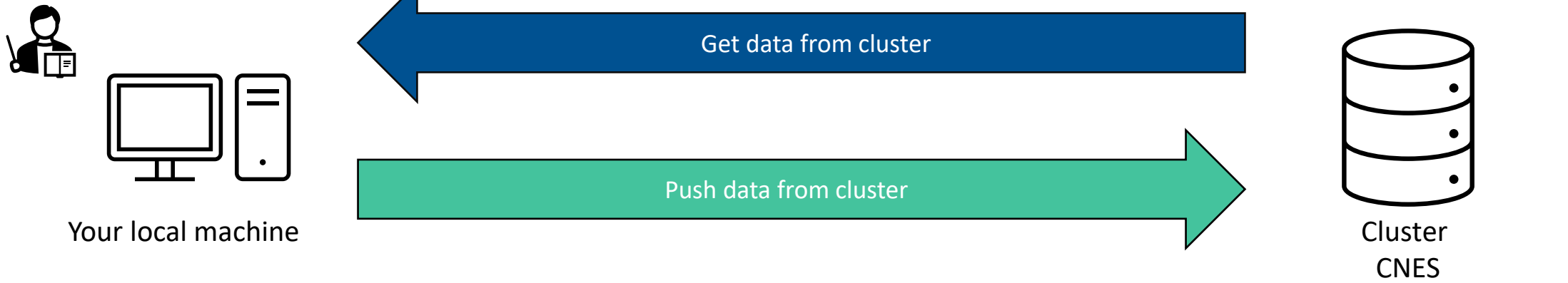


Shared Datalake

External data retrieval – what about your own data?

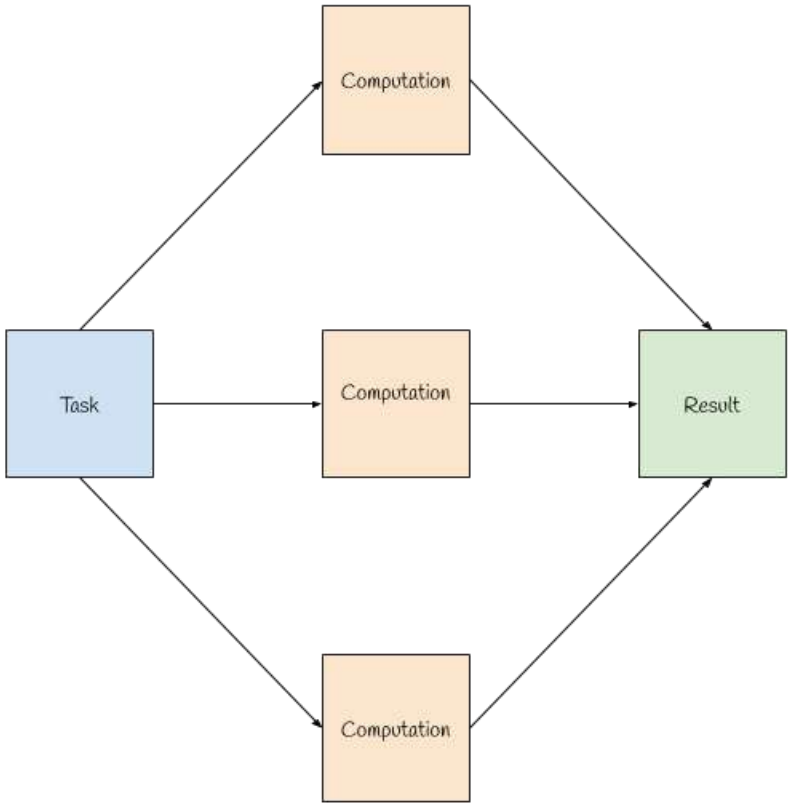
Supported protocols : SCP, SFTP

Tools :



Exploiting cloud capabilities: parallel computing made simple

Parallel computing available with a few lines of python (or even implicitly)




```
import xarray as xr

data = "/path/to/data"
ds = xr.open_mfdataset(data, combine="nested", concat_dim="time")
ds.mean("time").compute()
```

A fair number of Python libraries have partial or total integration of the underlying parallelization framework :

```
import xarray as xr
import dask_jobqueue

# Launching the cluster
cluster = dask_jobqueue.PBSCluster(
    cores=2, processes=8, memory="10GB",
    walltime='01:00:00', interface="ib0")
cluster.scale(1)

data = "/path/to/data"
ds = xr.open_mfdataset(, combine="nested", concat_dim="time")
ds.mean("time").compute()
```

Exploiting cloud capabilities: parallel I/O for fast data access

SWOT products available in their native netCDF format and in a new ZARR format (parallel and faster data access)

	Non-parallelized	Parallelized
Netcdf	90 m	33 m
Zarr	2 m 15 s	**1 m 20 s**

Mean over >10 000 netcdf

```
import xarray as xr
data = "/path/to/netcdf"
ds = xr.open_mfdataset(data, combine="nested", concat_dim="time")
ds.mean("time").compute()
```

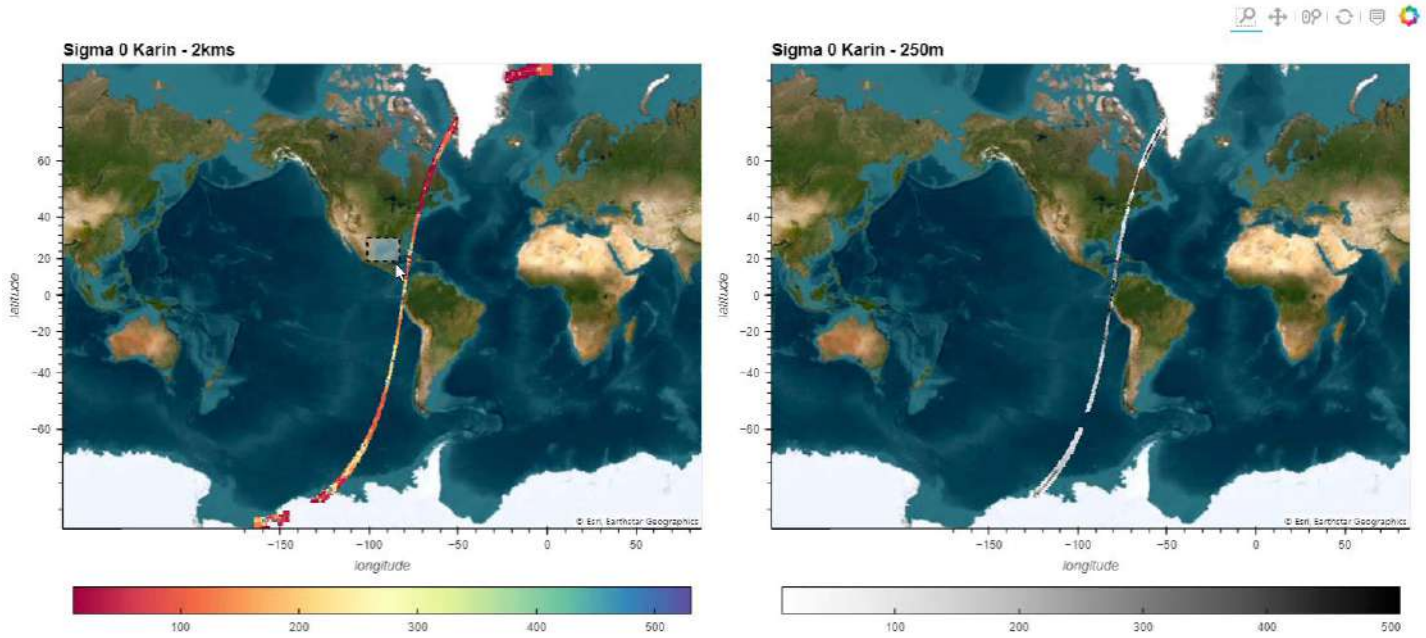
Mean over a zarr store

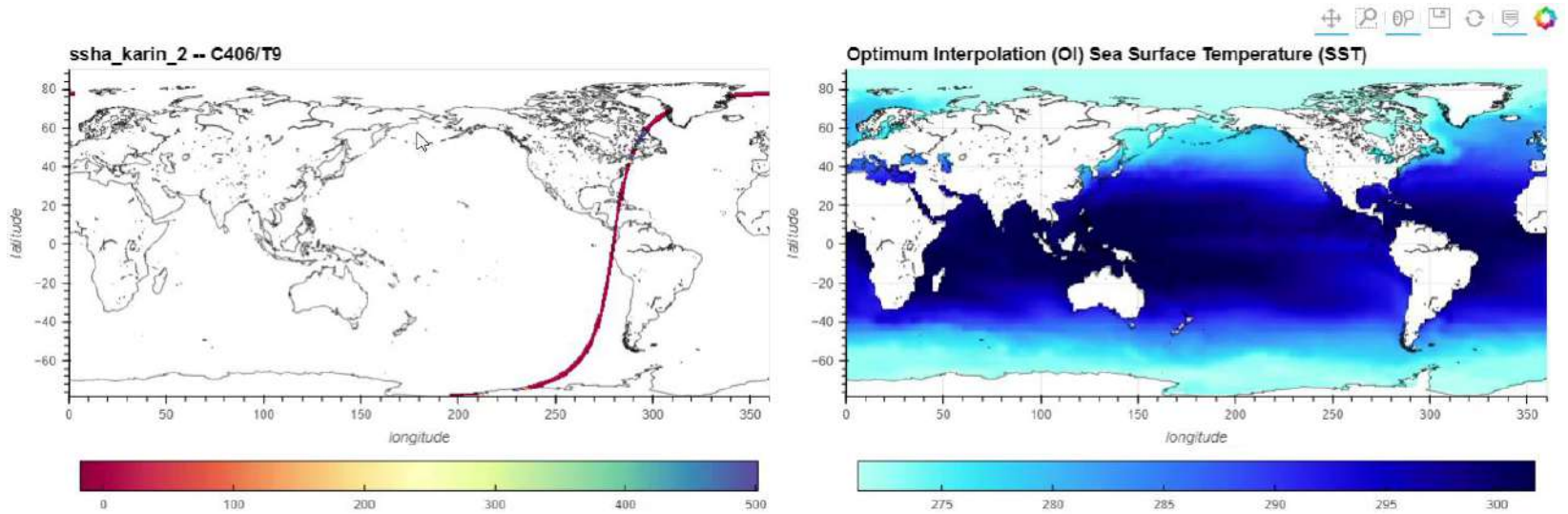
```
import xarray as xr
data = "/path/to/zarr"
ds = xr.open_zarr(data)
ds.mean("time").compute()
```

Mean over a zcollection

```
import swot_calval.io
data = "/path/to/zcollection"
zc = swot_calval.io.open_collection(data)
zds = zc.query()
ds = zds.to_xarray()
ds.mean("time").compute()
```

Fast interactive data visualization





SWOT HelpDesk - we are here to help you

Demonstration

- Experiments on simulated data

CNES cloud access request

Getting started

- Cluster working methods
- HPC training
- Tools upskilling
- SWOT data training
- Documentation, examples and tutorials

Project set up

- Project's workspace creation
- Accounts creation
- ACL (Access Control List) management
- Data retrieval
- Code installation / coding environment installation

Work on your project

- HPC support
- Code optimisation
- FAQ & HelpDesk

Helpdesk Service mutation

- Export or archive your codes and data
- Support will keep helping you during this phase
- **Nothing is lost !**

Contents

Which SWOT products ? type, region, period...

Which kind of external data ? sensor type, region, supplier...

What kind of tools ? SWOT specific, cloud-specific, generic...

What kind of hardware resources ? Cpu, disk

Does the AVISO offer make sense for you?

How can we make it better for you?

Summary

What we can do for you

- Give access to processing power
- Take care of setting up an efficient coding environment
- Explore the SWOT ocean swath with few line of codes
- Decrease the time you spend developing SWOT-adapted tools
- Accompany you during the important steps of your project

Contact

rchevrier@groupcls.com

atonneau@groupcls.com

How to apply

Send a mail with a short project description to our contact mail