Accessing NWC SAF buckets with kerchunk, virtualizarr & icechunk

Applied to NWC SAF EWC buckets

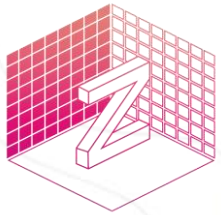
GOBIERNO DE ESPAÑA
VICEPRESIDENCIA TERCERA DEL GOBIERNO
MINISTERIO PARA LA TRANSICIÓN ECOLÓGICA Y EL RETO DEMOGRÁFICO

aemet Agencia Estatal de Meteorología

EUMETSAT NWC SAF

Accessing NWC SAF CF buckets with kerchunk, virtualizarr & icechunk.

Short guide for: `https://gitlab.aemet.es/jllisov/vzarr.git`

The project is a set of notebooks teaching how to build virtual data cubes from NWC SAF CF buckets.

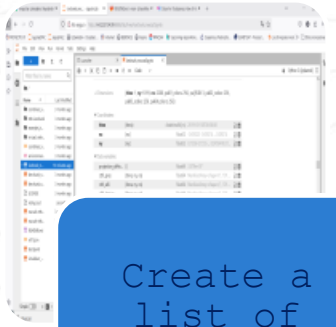
`Virtual access implies do not duplicating the files to build the datacube.`

The lessons use the buckets containing NWC SAF GEO reformatted outputs on EWC.
The format of the files is the precursor for MTG format of the NWC SAF.
This bucket is public inside the EWC (***"nwc-saf.0-degree.level-2-cf"***)

`The full lessons can be executed only on EWC.`

Once cloned the project and created the **Vzarr** conda environment you should activate the environment and launch the jupyter lab session with:
jupyter lab --ip 0.0.0.0 --port 8888

Workflow:



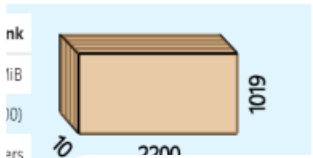
Create a list of virtual file references

```
virtual_ds = xr.concat(
    virtual_datasets,
    dim="time",
    coords="minimal",
)
```

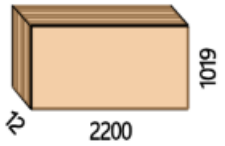
Concatenate with xarray

```
# write the virtual dat
virtual_ds.virtualize.t
session.commit("Wrote f
```

Store as an Icechunk store



Open with Dask Access



Append new files

Features

Using Virtualizarr implies do not duplicating the files to build the datacube.

The file colection is accessed via xarray.

Data Access:

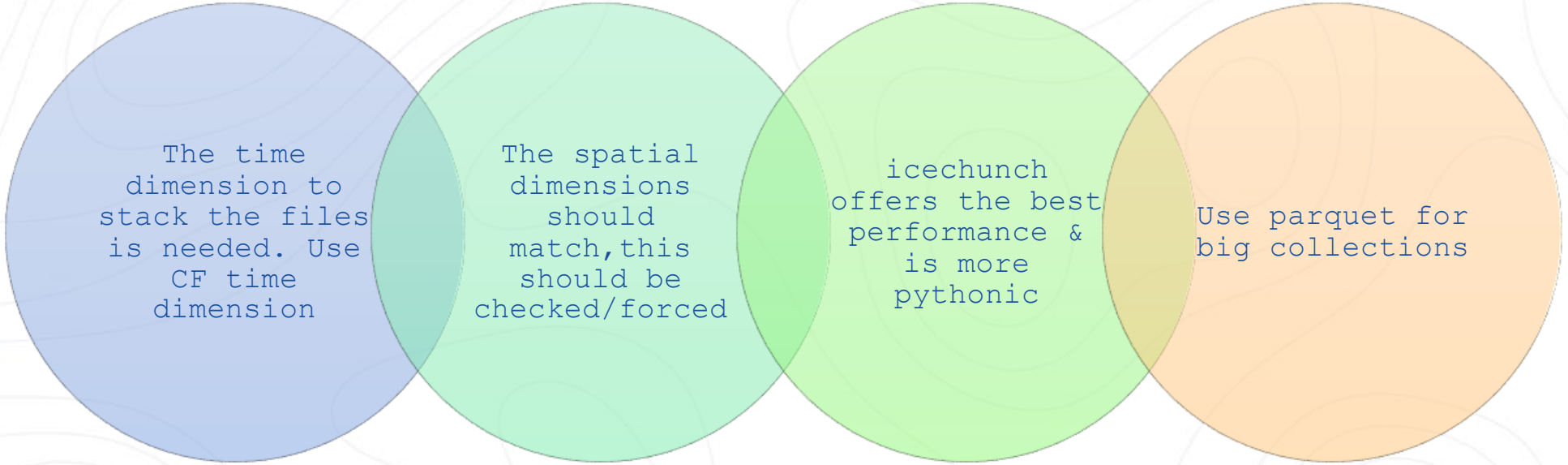
xarray datasets

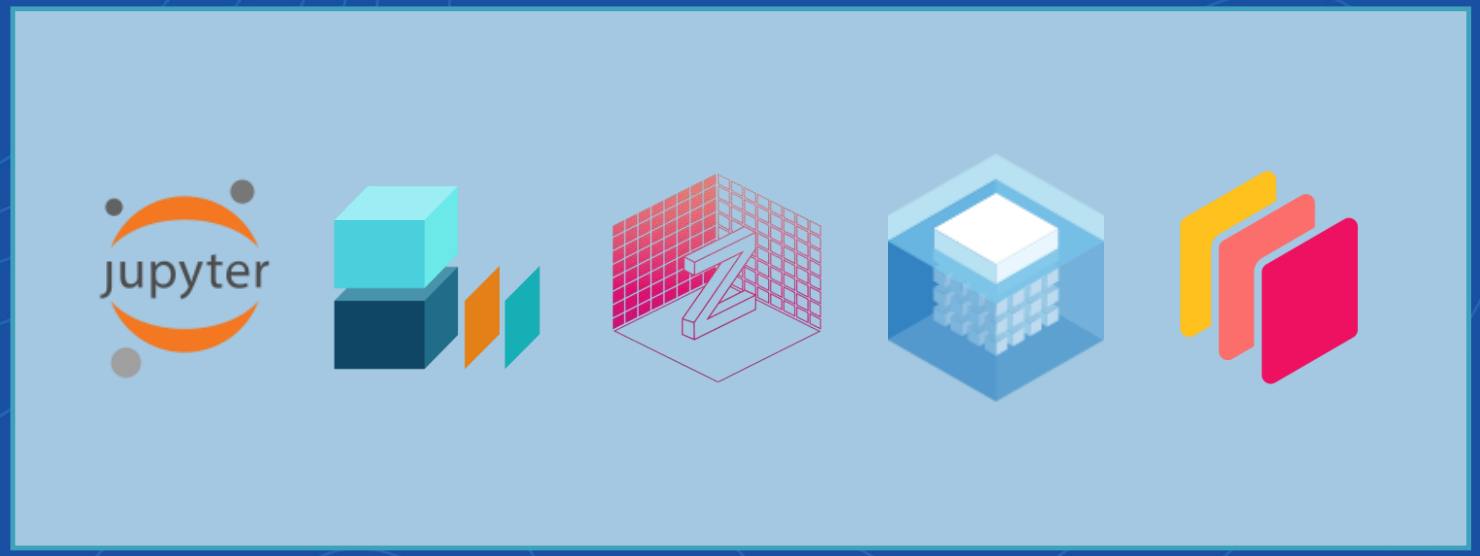
- *Dask arrays*
- Lazy/graph/memory

The files are not duplicated.

An index file is build (json/parquet)

Lessons learned





GOBIERNO DE ESPAÑA

VICEPRESIDENCIA TERCERA DEL GOBIERNO

MINISTERIO PARA LA TRANSICIÓN ECOLÓGICA Y EL RETO DEMOGRÁFICO

aemet Agencia Estatal de Meteorología

EUMETSAT NWC SAF