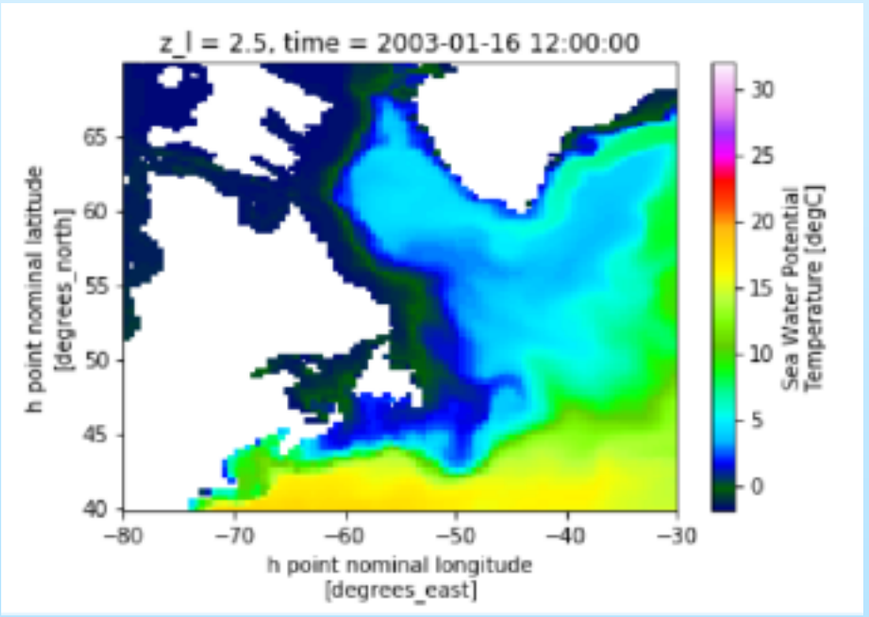
xarray: “label-aware” arrays

- Philosophically similar to netcdf data model
- Dataset = set of DataArrays
- Datasets can be build from N files
- DataArrays have labelled dimensions/coords
- We can use methods working on these labels

```
ds['thetao'].sel(xh=slice(-80,-30), yh=slice(40,70),
                 z_l=2.5, time='2003-01').plot(vmin=-2, vmax=32,
                                               cmap='gist_ncar')
```

```
clim = ds.mean(dim='time')
```



- xgcm: adds staggered grid awareness to xarray

dask: lazy, parallel and OOC

- xarray runs either numpy or dask under the hood
- if chunks are specified, then dask is the backend
- dask operates in lazy mode, numpy in eager mode
- dask build graph of operations, delays execution
- dask only executes when data is requested (plot,…)
- execution is multi-threaded on cluster (local, k8s, jobqueue)
- can handle dataset size larger than memory (OOC)

```
from dask.distributed import Client, LocalCluster
cluster = LocalCluster()
client = Client(cluster)
client
```

Client

Scheduler: tcp://127.0.0.1:63195
Dashboard: http://127.0.0.1:63196/status

Cluster

Workers: 4
Cores: 8
Memory: 17.18 GB