

UNL HTCondor Cluster
Scheduler Address: [tis://oksana-2eshadur-40cern-2ech.dask.coffea.casa:8443](https://oksana-2eshadur-40cern-2ech.dask.coffea.casa:8443)
Dashboard URL: <https://cmsaf-jh.unl.edu/user/oksana.shadur@cern.ch/proxy>
Number of Cores: 12
Memory: 18.73 GB
Number of Workers: 3
Minimum Workers: 1
Maximum Workers: 100

<> SCALE SHUTDOWN

Code

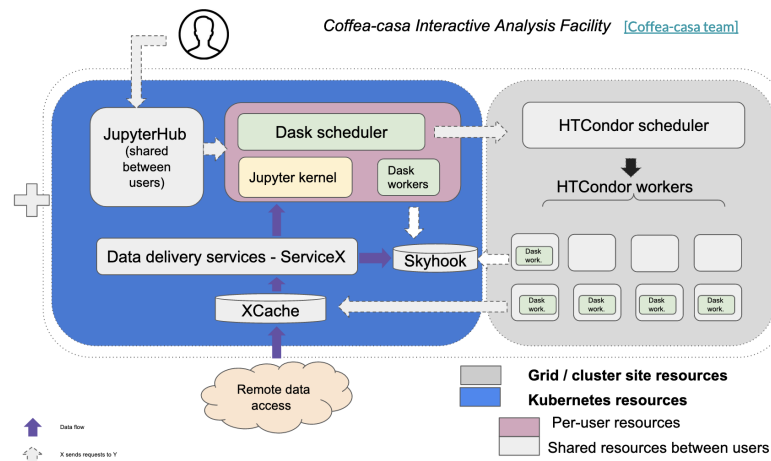
 git

Python 3 ●

```
[1]: %matplotlib inline
import glob
import os
import numpy as np
import numba as nb
import awkward as ak
import matplotlib.pyplot as plt
import uproot
from uproot_methods import TLorentzVectorArray
from coffea import hist
from coffea.analysis_object
from cyclical import cyclical
import coffea.processor as
```

The diagram illustrates the integration of a data catalog (fileset) with a data lake architecture. The fileset contains paths like 'root://tHq:', 'root://tHW:', 'root://ttW:', and 'root://ttWJets:'. These paths are mapped to a data lake structure. The data lake includes components like 'JupyterHub (shared between users)', 'Dask scheduler', 'Jupyter kernel', 'Dask workers', 'Data delivery services - ServiceX', 'XCache', and 'Remote data access'. A pink arrow points from the fileset to the data lake, indicating the mapping process. A yellow circle labeled 'datadb:' is also shown, representing the metadata database.

reality!



naive...

