Production meeting

Merging update

Merging

We need this to deal with:

- root-tuple-virtual outputs form keepup
- CAF's

- Need to merge the files and their metadata

- Interactive scripts are in data-mgmt-ops/utilities/merging
- Currently branch `HMS-Jul24`

Major scripts

- ProdChecker.py in utilities tells you if runs are complete
- submitMerge.py calls mergeRoot.py and mergeMetaCat.py
- Can then use that information to do batch submissions
- Here is some basic setup:

```
git clone https://github.com/dune/data-mgmt-ops.git
cd data-mgmt-ops
git checkout HMS-July24 # for now
cd utilities/merging
mkdir logs
source setup_local.sh # ignore the stuff with my name on it
mkdir /pnfs/dune/scratch/users/$USER/tars
mkdir /exp/dune/data/users/$USER/tars
mkdir /pnfs/dune/scratch/users/$USER/merging
```

First time

```
python submitMerge.py --run 28005 --version v09_91_02d01 --maketar
```

Will run on run 28005 and put the tarball in

```
/pnfs/dune/scratch/users/$USER/tars
```

And output will go to:

```
/pnfs/dune/scratch/users/$USER/merging/run0000028085_000000_20240821182539-local
```

The 000000 is a skip # case you want to do runs in chunks

```
usage: submitMerge.py [-h] [--detector DETECTOR] [--chunk CHUNK] [--nfiles NFILES] [--skip
SKIP] [--run RUN]
                      [--destination DESTINATION] [--data_tier DATA_TIER] [--file_type
FILE_TYPE]
                      [--application APPLICATION] [--version VERSION] [--debug] [--maketar] [--
usetar USETAR]

optional arguments:
  -h, --help                 show this help message and exit
  --detector DETECTOR        detector id [hd-protodune]
  --chunk CHUNK              number of files/merge
  --nfiles NFILES            number of files to merge total
  --skip SKIP                number of files to skip before doing nfiles
  --run RUN                  run number
  --destination DESTINATION.destination directory
  --data_tier DATA_TIER      input data tier [root-tuple-virtual]
  --file_type FILE_TYPE      input detector or mc, default=detector
  --application APPLICATION  merge application name [inherits]
  --version VERSION     software version for merge [inherits]
  --debug               make very verbose
  --maketar             make a tarball
  --usetar USETAR       full path for existing tarball
```

Submits batches

- You can start at arbitrary points by using –skip and change the chunk size as well.
- Right now it submits jobs in batches of 20 merges until –nfiles is hit.
- If chunks are 50 input files, that means 1000 files/job
- Some runs have > 20,000 files

Outputs are:

```
hd-protodune_detector_run0000028078_physics_standard_reco_stage2_calibration_protodunehd_keepup_root-tuple-virtual_merged_skip000000_lim000050_20240821T233041.root

hd-protodune_detector_run0000028078_physics_standard_reco_stage2_calibration_protodunehd_keepup_root-tuple-virtual_merged_skip000000_lim000050_20240821T233041.root.json

run_28078_0_50_root-tuple-virtual.log
```

Issues

- Missing files:
  - Small # of files are missing from production
  - Some files do not make it to fnal and only files at fnal can be merged easily.
- Rucio failures
  - Asks for 50 files at once
  - Sometimes gets an error – quits there.
- Once we can store files and mark files as merged/stored in metacat, we can run recovery on the missing files.

Next step

- Store/declare a couple runs to metacat/rucio
- I will then work on methods for tagging input files as merged
- Can then then make the merger refuse to merge files that are already done and do cleanup
- Make datasets and distribute
- Can then retire the intermediate files from metacat/rucio