

more Gino use cases

There are 2 more use cases I expect we will need to handle:

- splitting input files
 - eventually we will get ~1.5 GB files from the MOC for each downlink, and will need to separate them into many files to feed to the estimate 75 CPUs needed to process the downlink in ~1 hour. It is possible we will receive the 1.5 GB in multiple files - unknown yet, but see next bullet anyway.
 - by the time we get ~4 towers integrated this coming year, keeping the raw data files small enough so that the recon files don't get out of sight will become a problem. They will be taking 10 minute or less runs!
 - so we need a mechanism to expand out a run's input file to go to multiple nodes and then regroup them all at the end, I assume allowing a run to be a logical entity whose datasets could be lists of files. Further complication would be in handling subsequent steps in the pipeline for that run, and whether external pipelines make use of a dataset.
- MC concatenation of files
 - once an MC run is done, we gather up the myriad ntuple files, run a pruning mechanism on them (for the MC I just ran I weeded out events with no tracks) and then concatenate them to a more handleable set of files (by user request <~ 200 MB or so).
 - it would be nice if these could both be done automatically in the pipeline and be recorded. I am doing both manually. A task would have to have something like summary properties where you could make outputs based on a selection of all the input datasets in the task, and record the outcome. And again, the output dataset could be a list of files.