

haddCondor

Justin Anguiano



University of Kansas

April 12, 2021

Overview

haddCondor - a batch tool to accelerate combining ROOT files

- Combines ROOT hadd with HTCondor and DAGman
DAGman - Directed Acyclic Graph Manager
- Iteratively adds together ROOT files and their outputs until all files have been combined into a single file
- C++ code to automatically generate submission files and job dependencies
- DAG jobs automatically launch when dependencies are fulfilled

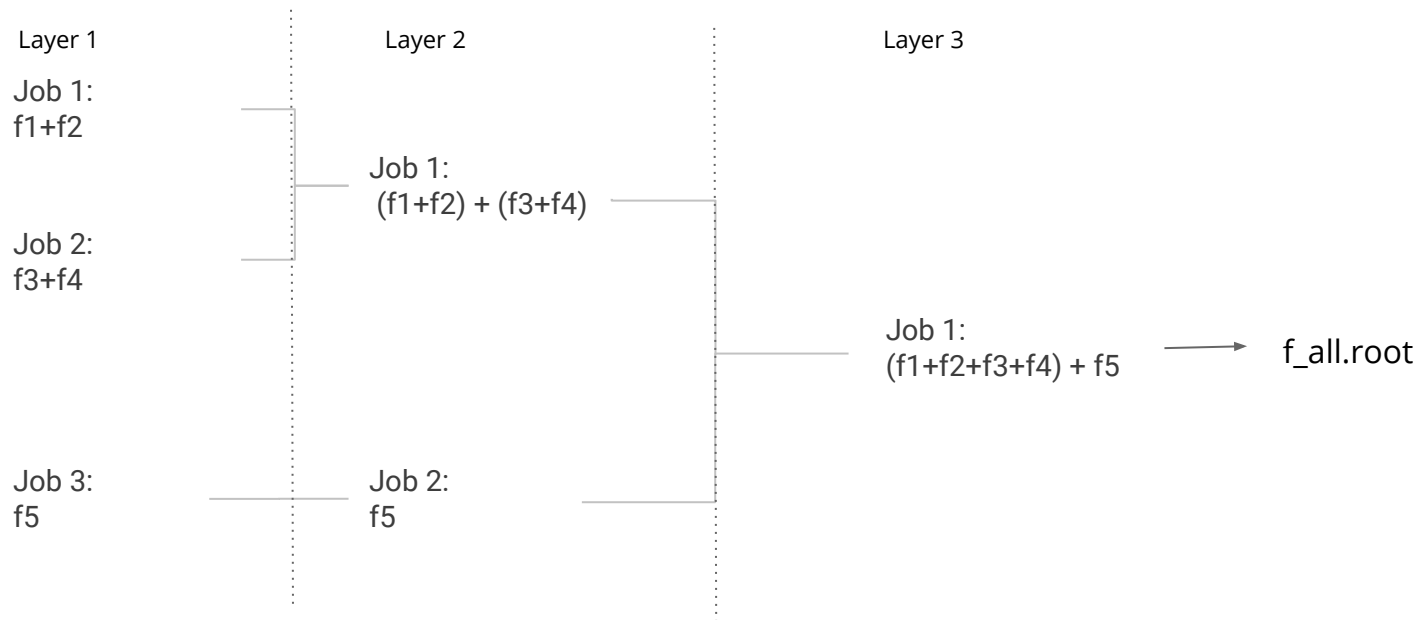
Example

Suppose we have 5 files to combine, we choose to combine N files per condor job

Input Files: f1.root, f2.root, f3.root, f4.root f5.root

Target: f_all.root (hadd of all root files)

Let $N=2$



Jobs proceed by layer, with preceding layers output being the input to the proceeding layer

Workflow

- Create an Input file list
- Run haddcondor.x
 - Generates workspace with directories for rootfiles, submit scripts, and logs
 - Creates a .dag file which maps out job dependencies
- Do a condor_submit_dag dagfile.dag
- Hadded file is written to /root/ directory

Performance:

Tested on BFI files with TChiWZ and ++bkg, 74 root files at ~1.8MB each

“ T2tt and ++bkg, 167 root files at ~ 2MB each

- Interactive hadd with all files : (TChiWZ) 40 minutes | (T2tt) 3 hours 40 minutes
- haddCondor with nCombine = 2,9 : (TChiWZ) 10 minutes | (T2tt) 16 minutes

Links + Other

Get this tool at : <https://github.com/lphsx/haddCondor>

- Detailed example and build information in readme
- DAG may break with a low nCombine and a high number of jobs per layer
 - Appears to have trouble parsing dag file with a long list of parents /children
 - Solve by just increasing nCombine to reduce the number of jobs
 - T2tt failed with 167 files and nCombine=2, works with nCombine = 9
- Intermediate files currently not deleted, need to be deleted manually
- Picky formatting, make sure to format lists and paths correctly
- Hadd speed can be further accelerated by adding -j flag to do local parallelization
 - Could boost performance in general, but, requires requesting more resources
 - -j flag is untested