

Resources Needed for MC

Bradley Yale

04/12/2017

Space needed for a typical wab-beam-tri sample
(100 recon files, 270k events = barely enough statistics):

- **Generated lhe/stdhep** (hundreds of files per process): 1.5-200 MB/file
 - beam is the largest, but least-frequently needs to be tweaked
- **SLIC** (10's of thousands of files): ~430 MB/file
- **Readout** (hundreds of files): ~20 MB/file
- **Recon** (hundreds of files): ~80 MB/file
- **DSTs** (hundreds of files): ~15 MB/file

*Do we ever re-use
stored SLIC files?*

Total: ~5 TB

Total without SLIC: ~20 GB

(why SLIC is kept on tape)

Multiply this by # of Detectors

Time needed to generate a typical wab-beam-tri sample (100 recon files, 270k events = barely enough statistics):

- lhe/stdhep generation:
 - 100 beam files (most run in parallel): ~3 hours (Mollers are similar)
 - MadGraph (10,000 events): 10-30 min.
 - SLIC: 500,000 events processed /file
 - ~ 1 hour per job, 10,000 jobs submitted at once (~1GB used/job)
 - For SLIC, the bottleneck comes from the very large number of jobs required to get enough statistics
 - 10's of 1000's submitted, but 100's run at once (2 GB allocated)
 - Requested resources could be lowered to 1.5 GB safely, but still depends on available nodes
 - Readout takes 100x fewer jobs, similar resources, but far more actual run time (>24 hrs.)
 - For reference, "1 day of beam time" (wab-beam-tri):
 - ~100 million SLIC jobs (1ms of beam/file) + 1 million readout + 1 million recon (80 TB cached)
- What makes readout so slow?*

Time needed to generate a typical wab-beam-tri sample (100 recon files, 270k events = barely enough statistics):

- Readout: 100 SLIC files (500k events each) readout/file
 - ~ 25 hours per job, the most time consuming step
 - Can be reduced by filtering events first, but a lot of events to process
 - Time could also be reduced by 10to1 readout
 - Either way, 100x more recon = 1 million more SLIC jobs
- Since readout only triggers on $\sim 4 * 10^{-5}$ events from what is generated, recon only takes another hour after readout
- Yes, simulating the full beam is incredibly inefficient as it stands.

One Possible Improvement (feel free to add)

- Making “tritrig-wab-beam-tri” can improve statistics, by having a trident in every readout event
 - Done by merging spaced tritrig events with beam overlay after SLIC, before readout
 - The most efficient way of creating trident events with beam overlay (by construction)
 - Not sure if the beam time is preserved though, since inserting tridents based on trigger rate instead of using beam bunch structure (current) + cross section.
- Currently, I use MadGraph5 tridents for wab-beam-tri, because the cross section is unstable for an ESum cut looser than 50 MeV.
 - This means that “unbiased” (uncut) tridents may not be possible to make
- So “MG5” tritrig-wab-beam-tri would consist of “wab-beam” readout simultaneously with the MG5 tritrig currently used in wab-beam-tri

By the way

- Lots of requested 1.056 GeV Layer0 wab-beam-tri is still being made.
- 100 beam seconds (~2.7M recon events) is the goal by next week.

• Please use it...

• Requests for MC jobs should go to Takashi, who will prioritize & send the jobs on to Bradley.

• Using pulser data for background could speed up this process for beam energies where we have data.

- beam-MC "overlay" of events is not trivial. - Thoughts?