

TAG Discussion

TAG

- PAT's Problem Child:
 - Lots of effort and promise... but effectively unused by physics community
 - Event Picking used by Data Prep and Combined Performance
 - Perpetually behind in validation, problematic for production, and inconsistent with newest object selections.
- Strategy, let's make the physics case:
 - New effort to demonstrate performance/disk benefits of TAG, using SUSY DPD making as concrete example (Jack Cranshaw, Louise, David C.)
 - Some groups are producing DAODs to speed up D3PD iterations... but then they fill disks with DAODs...TAGS could eliminate this problem.
 - Plan to simplify TAG building by using the D3PD framework.
 - Core elements will be implemented by Attila by May...
 - Need manpower for actual reimplementation of TAG via D3PD
 - Perhaps this person can then be the TAG content gatekeeper
 - New Manpower in TAG Validation (Wolfgang Ehrenfeld, James Dassoulas)
 - DPD Production Plan: Rebuild TAGs with the DPD so it is up-to-date with newest/greatest selections and DPD streams can be monitored/reproduced via the TAG
- Discussion about TAG today...

Making the Physics Case for TAG

- Needs to happen from 2 directions:
 - TAG Building: Better content, validation, and production (coupled to the TAG rewrite and merger with DPD production)
 - TAG Event Picking: Demonstrate both performance gain and the ease of use in ProdSys/DA
- Recall that without TAG, people get most performance boost from skimming by producing intermediate DPDs
 - Examples today are DAODs in SUSY or or very large D3PDs in Jet group (skimmed on GRID).
 - Usually means that there is an additional step required to make the DAOD/ D3PD before you can take advantage of skimming... so TAG can help eliminate this step.
 - But the main problem is the space for storing the intermediate DPDs... this is where the TAG can really help

Making the Case via Event Picking

- Really we need to compare performance and disk requirements between:
 - running on AOD, skimming on input w/o TAG
 - running on pre-skimmed DAOD and fairly accounting for the DAOD production time
 - running on AOD pre-skimming using TAG DB
 - running on AOD pre-skimming using TAG Files
- Note that we need to make the case for Event Picking first, then the case for DB over TAG files.
- Must be done on the GRID
 - Properly account for file staging/touching
 - Evaluate both functionality and usability of DB and Tag Files in both DA and ProdSys
- We should scan the skim rate space (let say 1 to 10%) to find the performance boost sweet spot.
 - For now, more relevant for the non-SUSY DPD prod case... but we can emulate it with SUSY DPD prod.
- SUSY DPD Production ideal test case:
 - Jobs are ready... just need to try it with the TAG.
 - Impending DAOD disk space problem...
 - Skimming via Triggers, which the current TAG supports.
 - All we need to do is show it works...
 - It is immediately useful and give us a place where TAG is used for analysis.

Moving TAG to DPD Production

- Motivations:
 - If we set TAG bits corresponding to DPD streams, we can
 - quickly reproduce a stream
 - monitor overlaps
 - We can keep the TAG consistent with newest AODFix, Object Cleaning/Selection
 - Most practical way to get better usage of TAG Physics word.
- Will we still need to produce the TAG in Tier 0?
 - We should be aware that TAG content optimized for DP/CP would be different than TAG content optimized for Physics
 - eg: DP/CP prefer all Jets marked as good, bad, ugly... Physics would prefer Good Jets only to make queries like `good J1_pT>100, good J2_pT>50, ...` easy.
 - Possibilities:
 - No TAG from Tier 0 or DPD Prod Tag replaces tags...
 - Then we need to optimize/augment content to meet both DP/CP and Physics Use Cases

TAG Validation

- TAG Validation machinery is decoupled from TAG production, building, or the rewrite.
- Really 2 aspects:
 - RTT: Does it run and give some content
 - Validation Group: is that content correct
- What are reasonable content validation questions?
 - The content hasn't changed since last time?
 - Need to keep track of CP changes as well as TAG content evolution.
 - Is the content consistent with AOD?
 - Should be compared to AOD after the AODFix
 - AOD doesn't have object cleaning/selection so ...
 - Essentially this means that the validation code has to independently reproduce the TAG. It's like having 2 versions of the TAG that we maintain.
- TAG Validation is both different in nature and more challenging than most other validation efforts... requires a dedicated person!

TAG Building Rewrite

- Motivation:
 - TAG is really a form of D3PD!
 - Current TAG building is very proprietary and “hard-wired”...
 - Maintaining TAG content will be easier to delegate to CP groups if we use the D3PD framework to build the TAG.
 - Really moving TAG content definition from C++ to Python.
 - Another way to encourage better usage of TAG Physics word.
- We need to support the current TAG while rewriting the TAG
 - At first look the Current TAG content is probably unsuitable for the broad physics community
 - But I suspect that Trigger-based skimming is the only sensible thing now... and this is fine in the current TAG.
 - We unfortunately do not have a TAG content manager... we are addressing the few request on case-by-case basis...
 - My Personal view (open to discussion): if the current TAG can be used for physics now (eg via Triggers) then great... but effort put into major content changes to current TAG in order to meet physics is better direct towards to the rewrite.
 - We should try to meet DP/CP content request changes... but this is already fixed because for now the TAG is produced in Tier 0 and we won't move to combined TAG/DPD production until after the rewrite.

TAG Building Rewrite

- Really 2 steps:
 - Upgrade D3PD Maker:
 - Backend...
 - Flattening (no vectors in TAG)/Max objects
 - Need detail level corresponding to TAG content.
 - Re-implement the TAG:
 - Opportunity to optimize TAG content for physics
 - Will use standard tools/sequences/containers, so this isn't a hugely difficult task.
 - Rewrite is closely related to the validation
 - Validation is basically testing TAG against hand-written version of TAG... so the Validation expert knows how to re-code the TAG.
 - New TAG needs to be validated against the old anyway.
 - Ultimately the CP/Physics groups will be maintaining their portion of the TAG building software and the associated content...
 - so they should contribute the rewrite from the beginning.

TAG Manpower

- The wish-list:
 1. Event-picking Support from both DB and PANDA/Ganga- Jack/QZ?
 2. TAG Building Developer (possibly short-term)
 3. TAG Content Manager
 4. TAG Validation- Dimitris -> Wolfgang/James
- Very hard to get help... but we have now: 1 and 4.
 - Being practical, we need combine
 - 2+3: 2 is probably mostly short term, so preferred
 - 3+4: 4 is very involved, so not preferred.
 - We need at least one more person working on TAG.

Other Questions

- Train: Can we update TAG every 2 weeks?
- Can we update just certain variables in the DataBase?
- How are Version handled?
- Book-keeping: how do we know we ran on everything?
- Can we get the TAG input LB list for a ELSSI query?
- Are there any differences between Event Picking in DA vs ProdSys?