Selecting data for top studies

Tommaso Lari

In Milano, we have always run the analysis on AODs.

It works fine, but if everybody wants to do the same when we will have 100M events per stream there might be problems (see UAT test).

Even if it works, preselection of interesting events allows to get results from the analysis faster. If the preselected samples is small enough it might even be stored locally.

The obvious solution is

- Select (loosely) events of interest, save them (DPD). This is done once for all the analysis sharing the same preselection.
- Run on DPD, produce ntuples.

If primary or group DPD are suitable they can be used of course.

Suitable: produced promptly, selection not to loose (too many events to run on) or tight (constrain on analysis, control samples, ...).

In this talk I make some considerations on:

- Options for 900 GeV collisions
- Options for first few weeks/pb^{-1} at 7000 GeV.

I do not see any need now to discuss DPD selection late in 2010, we will have time to discuss this based also on experience with early data.

900 GeV collisions - reminder

Possibly 10 milion events expected (the first dozen or so taken yesterday).

I have runned several times in last week over 3 x 10M simulated 900 GeV collisions in FZK without any problem. However

- Producing ntuples in half an hour on a local sample is better than doing it in one day on the grid.
- Most (99+ per cent) of events are useless if one is interested, say, in events with any jet ($p_T > 15$ GeV), OR a (fake) egamma candidate, OR a muon.

Some preselection would be useful. Also a technical test of DPD production.

Centrally produced options

- Data shuld be streamed. The L1Calo stream should be a few Hz only and contain all events of interest. Not obvious wether any of yesterday's few collision have jets. One may download the 537 L1Calo events to check.
- DESD_CALOCOMM will contain nearly full ESD information for all events with at least one jet of 10 GeV (AntiKt4H1Tower, EM scale). Not produced yesterday.
- DESD_EGAMMA will contain full ESD information for all events with any egamma candidate. Not produced yesterday.

Private option

Both a backup and technical test.

In Milano (Lidia Dell'Asta) we are testing the option of selecting interesting events with TAG. In this phase it's very experimental (we will run on both TAG selected and inclusive sample and compare). But if you are interested, we can organize the production of common skimmed samples.

An other option would be to run a DPDMaker on all events and dump those of interest.

What Lidia has done

She produced TAG datasets for simulated 900 GeV data (for real data these will be produced at Tier-0). Produced with 15.5.1, Tier-0 TAGS will use 15.5.3 using for example a different jet collection.

She then selected events with a TAG query and produced a skimmed file of AOD out of 10 milion initial events:

- p_T(leading jet)$>$ 15 GeV, 11812 events selected
- $MET > 10$ GeV: 4303 events selected
- $N_\tau >= 1$: 459 events selected

Disclaimers: numbers for simulated collisions. May be different with data. Cosmics may give larger rates. If the TAG were not produced with your favorite jet algorithm, you will get a biased sample - AntiKt4 and Cone7 can have very different p_T in such a low range.

Common selection

It should not be difficult to decide on a selection of interest for all the groups doing top analysis. What about selecting all events with at least one 10 GeV jet OR an egamma candidate OR a muon candidate and save them ?

In Milano we volunteer to do it via TAG. Any volunteer to try it via DPDMaker ?

Selection for 7 TeV collisions - first pb-1

Start with 900 GeV cuts on first collision, tighten them as luminosity (and our understanding of detector) increases. Possible status at the end of pilot run (rates from a quick look at JF17).

- On egamma stream: 1 loose electron $p_T > 15$ GeV AND 1 jet $p_T > 15$ GeV AND (medium electron OR missing energy OR more jets OR prescale). First two conditions only: about 1 milion events.
- On muon stream: 1 muon $pT > 15$ GeV, 1 jet $p_T > 15$ GeV (about 100k events)

The point is, I think we can easily accomodate a selection which suits everybody.