DAOD & D3PD experience in the SUSY WG from a PAT perspective

David Côté (CERN)

Software & Computing Workshop

7th April 2011

On behalf of the SUSY Software & Production team.
Warm thanks to Attila Krasznahorkay! :-)

SUSY software & production

- main data formats:
  - SUSYD3PDMaker (E_{tmiss} sub-group)
    - to a smaller extent: DAOD
  - DESD (RPV/LL sub-group)
- production entirely done with ProdSys
- organization: team of ~15 people
    - started only recently (February 2011)
  - main areas of activity:
    - production of samples
      - Katarina Pajchel is our Grid expert
      - huge help from Junji Tojo (not formally in the group)
    - D3PDMaker software
    - data reduction
    - NTUP validation

I will focus on PAT aspects in this talk. ADC aspects were covered this morning. Also, I will mostly concentrate on the (larger) E_{tmiss} sub-group, for simplicity.

Manpower (random order)

- D3PDMaker software
  - J. Goodson, L. Heelan, J. McFayden, A. Olariu, D. Côté
- Production & monitoring
  - K. Pajchel, B. Gjelsten, T. Müller, C. Ohm, J. Sundermann
- Data reduction
  - L. Ancu, L. Heelan, L. Marti, D. Côté
- NTUP validation
  - C. Adam-Bourdarios, A. Kravchenko, R. Brunelière, J. Lorenz, S. Mahmoud, S. Becker, D. Côté

18 names on this slide! (large variation of time fraction)
- large… but not compared to everyone doing their own productions

Development phase right now: more manpower required temporarily(?)
- planning on 2.5 FTE for an eventual steady-state

General strategy

- Common NTUP_SUSY with aggressive skimming and slimming of the content
  - keep data volume manageable for local analysis
- …but fast turn-around to fix/add things
  - using DAOD as anchors for quicker re-processing
- Note: the D3PDMaker is our main analysis tool, but support for athena code remains mandatory
  - e.g. precise object definition from CP group
    - needed for AOD-based vs NTUP-based validation
    - needed for DAOD skimming from offline objects
  - also: athena experience is required for D3PDMaker development and support

Data flow

data: AOD→DAOD (once)
DAOD→NTUP_SUSY (multiple times)

MC: AOD→NTUP_SUSY (multiple times)

- skimming-bookkeeping machinery works nicely!
  - great to see this finally used full-steam by the community! ;-)
    - one bug found and fixed in CutFlowSvc
  - full support of D3PD
    - has been useful already in the SUSY group
  - virtual cuts activated for MC

Data volume

- Current SUSY disk usage: ~250 / 350 TB
  - production 06-20 (reference)
    - NTUP: 51 TB for data, 27 TB for MC
  - production 09-04 (recent)
    - NTUP: 2.8 TB for data (~5%), 8.6 TB for MC (>30%)
    - DAOD*: 42.2 TB (!)
      - DAOD_SUSYJETS: 18.1 TB, skim rate: 30%
      - DAOD_SUSYEGAMMA: 19.2 TB, skim rate: 35%
      - DAOD_SUSYMUONS: 4.9 TB, skim rate: 9%
- Current DAOD tuning uses too much disk
  - possible solutions:
    - DAODM, also good for signal MC
    - Tigher cuts in DAOD using offline objects
    - DAODs only for less inclusive analyses
      - no DAOD for inclusive 0/1-lepton
    - re-run from AOD at each processing
    - use TAG(?)

*2011 expectations: 1.5-2x more than 2010 (Egamma 25%, Muons 25%, JetTauETmiss 40%)

How faster is it to run from DAOD?

- Compared p428 *vs* p494, Muons stream, run 162623
  - p428 (input AOD, 3863k events)
    - 160 jobs
    - median job duration: ~170 minutes
    - total processing time: 27200 minutes
    - end-user time for all jobs: 15:12 - 23:10 (22.02.2011)
  - p494 (input DAOD_SUSYMUONS, 760k events)
    - 27 jobs
    - median job duration: ~80 minutes
    - total processing time: 2160 minutes
    - end-user time for all jobs: 13:07 – 15:36 (06.04.2011)
  - p494/p428 ratios:
    - input skimming: ~20%
    - number of jobs: ~17%
    - processing time: ~8%
    - end-user time for all jobs: ~33%

Strategies: NTUP_SUSY validation

- Two aspects: technical usability and quality of content
  - both tested automatically on nightly builds
- Technical (ATN framework): run AOD→NTUP_SUSY and report any ERROR
  - for MC and for data with/without skimming
    - implemented and running on all important releases
- Content (RTT framework): use AOD as reference and compare cut-flow tables from (d)AOD-based *vs* NTUP-based scripts
    - implementation ongoing in new SUSYTools package
    - AOD-based analysis will share AOD validation software