CPPM

Marlon Barbero, Arnaud Duperrin

ATLAS France CAF - ATLAS France users meeting
21st November 2023
https://indico.in2p3.fr/event/30149/

CPPM Team

- Composition of the team
  - **team: 4 EC, 10 CNRS, 1 IR Computing, 7 PhD (1 cotutelle), 2 postdocs**
  - **analyses/activities**
    - **Di-Higgs boson:** transition of efforts to di-Higgs analyses
      - HH→bbɣɣ: finalization full Run 2 BSM interpretation in X→SH→bbɣɣ (1 PhD), starting Run 3 SH (1 PhD) and ZH→bbɣɣ measurement
      - HH→bbττ: new effort in link with ANR Dive (1 PhD, 1 postdoc)

      → T. Strebler appointed HH Framework coordinator (with Louis D'Eramo)
    - **other analyses**: H^{++} multilepton, SUSY RPV multi-bjet, ttHML

- Involvement in computing
  - **Local T2 CPPM:** 1 IR (0.5 FTE - Edith Knoops) + 1 CNRS (0.05 FTE – A.D.) + help from an engineer from computing service (Carlos Carranza) + collaboration with LHCb (Andrei Tsaregorodtsev)

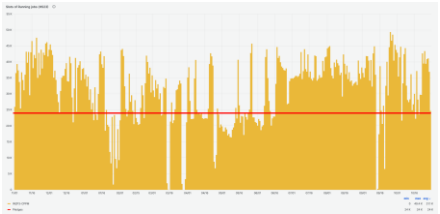
- Involvement in software
  - Run 3 pixel software
  - Releases: MC production, reconstruction, validation, git merge request reviewer
  - CP (Run 3 and Run 4): trigger, b-jets, Egamma, tracking

CPPM in categories "Software" or "Analysis support" or "Trigger"

from ATLAS OTP Institution Report - 2023 - Class 1, Class 2, Class 3 *(x2 for many activities since only field for first semester atm)*

First Name	Last Name	Activity	System	Task	Alloc
Elemer	Nagy	Trigger	General Tasks	Bjet Software and Performance	0.60
Thomas	Strebler	Computing/Software	ID gen	Common Tracking Software	0.20
Jozsef	Toth	Computing/Software	General Tasks	Validation of software release	0.20
Fares	Djama	Computing/Software	PIXEL	Software Development/Maintenance and Physics Performance	0.15
Grigore	Tarna	Analysis Support	General Tasks	Performance Studies - Egamma	0.15
Lorenzo	Feligioni	Trigger	General Tasks	Bjet Software and Performance	0.10
Thomas	Strebler	Analysis Support	General Tasks	Performance Studies - Flavour Tagging	0.10
Thomas	Strebler	Computing/Software	General Tasks	GIT merge request review, Level 1	0.10
Thomas	Strebler	Computing/Software	General Tasks	Reconstruction	0.10
Thomas	Strebler	Computing/Software	Upgrade	ITK - Performance Studies	0.10
Timothee	Theveneaux-Pelzer	Analysis Support	General Tasks	Generator Software	0.10
Arnaud	Duperrin	Analysis Support	General Tasks	Internal Software	0.08
Thomas	Strebler	Analysis Support	General Tasks	Performance Studies - Tracking CP	0.05
Thomas	Strebler	Computing/Software	General Tasks	GIT merge request review, Level 2	0.05
Arnaud	Duperrin	Computing/Software	General Tasks	Reconstruction	0.03
					2.11

Computing resources in 2023-2024



- Grid resources Tier 2-ATLAS pledge 2024
  - **Storage** = 2 200 TB (2200 TB pledged in 2023 → **+0%**)
  - **Computing** = 24 000 HS06 (24k HS06 pledged in 2023 → **+0%**)

  ⇒ due to lack of funding visibility, **T2 pledged resources were not increased in past years**
    - + problematic external low network connection (limited to 10 Gb/s… often saturating – 100 Gb/s in 2024 ??)
    - + cost of electricity etc.

  - How to finance our T2 in future ? (CPER is over)
  - **CPPM lab direction (Cristi) supports the development** of the T2 grid infrastructures for the new protocol
  - Several **engineers** from computing services involved in T2 infrastructure **will retire** in a near future

- Other "grid" resources (but non pledged resources)
  - **Storage** = 245 TB on LOCALGROUPDISK

- Other local (lab) resources (i.e. whatever is non grid)
  - 4 local-lab servers (263 TB, 204 cores) for ATLAS-CPPM activities (2 have expired warranty)
  - 1 new "small" server in 2023 for trigger NN LAr activities (AIDAQ)
  - CPPM has some GPU for developments/tests (not prod) shared among groups (not used by Atlas this year)
  - moving towards a **cloud computing** model at CPPM (i.e. no more local group servers) to mutualize computing resources among groups

Example of Analysis and needs (1)

- SH→bbyy : based on full Run 2, release 21 (finalisation, ATLAS circulation soon)

→ **we contribute to all analysis steps (MxAOD prod to limits)**

→ **model:**

- DAOD HIGGSD1→ MxAOD (data + MC nominal + MC syst)
  - Mini MxAOD are smaller calibrated DAOD produced by SH analysis group using the H→γγ Higgs analysis framework
  - 10 TB for MC, 0.5 TB for signal, 0.25 TB
  - Produced on the GRID and stored on local group disks from institute + eos
  - time to process: two weeks (mainly for the syst prod) by a group of 5 people in // but can take up to one month for full completion
- MxAOD → ntuples (300 GB)
  - CERN batch
  - Ntuples are stored directly on eos except for the yy+jets systematics samples which are too large and had to be stored on local grid institute disks
  - time to process: 15 minutes for nominal, can take up to a few hours for systematics.
- ntuples → fits (3 GB)
  - PNN histograms are produced on CERN batch and stored in eos
  - Fits are performed on afs using Condor to speed them up for the ~200 points (only available on afs)

→ **Good points/Difficulties/Needs/Expectation**:

- Software strategy to be revisited for Run 3 in link to have a general HH Framework for all final states

Example of Analysis and needs (2)

- Analysis Run 2 + 3 bbtautau (new involvement since Spring 2023)

→ **software and computing contribution**:
  - EasyJet framework development + maintenance for ntuple production

→ **model**: DAOD_PHYSLITE → ntuple
  - Post-processing steps on ntuple anticipated but TBD
  - Full MC20 + MC23 + Data datasets to be ultimately processed

→ **time to process**: Unknown

→ **where this analysis is mostly performed**: Grid + local

→ **good points/difficulties/needs/expectations**:
  - New common framework shared between all HH analyses
  - Central support from HH group, in link with AMG
  - Shared central developments + consistency in view of future combinations

Example of Analysis and needs (3)

- b-tagging ITk
  - **S&C contributions**: all steps from prod to performance
  - **Model**:
    - FTAG1 derivation samples → HF5 ntuples using the FTAG framework (training dataset dumper)
    - Training Graph Neural using the Salt FTAG framework
    - Evaluate the performance of the trained algorithm in Umami or Puma (note: Root disappeared from whole chain…)
  - **Size**: 50M ttbar semilep + 50M ttbar dilep + 30M Z'
  - **Time to process**: one week for FTAG1 (few times per year), days for HF5, days for training, few hours for performance
  - **Where**:
    - on the grid (FTAG1, HF5), GPU for optimal trainings (lxplus or other local cluster)
    - Storage of HF5 on T2 local group disk or local
    - Analysis on local servers
  - **Comments**: all process has to be repeated several times per year and the software environment is evolving (but well documented and maintained by FTAG group)

Conclusion

- CPPM team:
  - ~6 physicists (not full time) + ~3 PhD students + 1 postdoc **involved in physics analyses**
  - **no more full time informatic IT support** since Manu's departure in 2020 (Edith helping at 50% but mainly for the T2 Grid activities)
- Computing:
  - **T2 CPPM:** running smoothly but uncertainty on financing + electricity cost + RH
- Software:
  - **software** involvement steadily **decreasing**
  - **ramping up on HH** analyses for Run 3&4 as main CPPM topic
    - Ex software: HH Framework coordinator
    - Ex ML:
      - b-tagging for Run 4 based on GNNs using new ITk detector
      - $H \rightarrow \tau\tau$ boosted tagging with GN2X