The ECoM2x Process

Frank Wuerthwein
UCSD/SDSC

Joint US Atlas & US CMS facilities R&D meeting

March 17th 2020

The ECoM Process (I)

- Every few years the CMS Spokes person charges a task force to provide advice on the **E**volution of the **Co**mputing **M**odel, or simply ECoM.
- The task force includes thought leaders from Physics, Trigger, Software, Computing, Run coordination and data validation, in short all areas possibly affected by the computing model, broadly defined.
- Typically two dozen members for the task force.

The ECoM Process (II)

- The resulting report is in principle delivered directly to the spokes person, and they decide how much if anything they want to make public and/or follow up on.
- In practice, this has been a very open process for the entire collaboration to build consensus on future directions.
- And especially for s&c, recommendations are often followed up already while the report is being finalized.

Example ECoM17

- ECoM17 was concluded in 2017 with the scope of providing advice on the remainder of Run2, and the entirety of Run3.
  - Increase HLT trigger rate to not compromise the “golden final states” of H, W, Z, and top.
  - Create a 2-tier MC production to benefit from beyond pledge computing capacity.
  - Impress upon the generator community to develop NLO generators that do not have negative event weights.
  - Create a “MicroAOD” that is at least x10 smaller than MiniAOD, and satisfies the needs of at least 50% of all analyses.
  - Develop pre-mixing for HL-LHC as a high priority R&D to reduce HL-LHC MC production costs.
  - Start R&D towards a framework to operate across multiple heterogeneous hardware architectures, accelerators and single die CPU/FPGA or alike.
  - Engage at least one other experiment in discussion about a shared data management system.

This is just a sample of recommendations that have since been followed up on.

ECoM2x

- Started in Spring 2018, and lasted two years.
  - Unusually broad charge and long lasting.

"We are forming a new task force, ECoM2x, to study how to adapt the so far successful CMS computing model to the expected demand of the HL-LHC data taking. The task force will include members from Physics Coordination; Software and Computing, including representatives of our tiered centers; Physics Performance and Datasets (PPD); Trigger, both Level 1 and High Level Trigger; Upgrade; Run coordination and representatives of CMS users/analysts. It will also include, as ex-officio members, a representative of the Computing Resource Board (CRB), the Collaboration Board, the Finance Board and the Spokespersons Team."

  - Chaired by Tommaso Bocalli & fkw

Split into 7 working groups

- WG1: *Technology tracking and expectations from industry (H. Newman).*
- WG2: *LHC and CMS modelling parameters. This includes general status of event sizes, cpu requirements of HL-LHC software today, premixing. Therefore, it aims to describe where we are today, and where we think we need to get to as a goal to make reasonable budget assumptions (D. Lange).*
- WG3: *Physics choices and their impact. HLT rate, (re)processing model (prompt vs scouting vs parking). Definition of analysis data tiers. Definition of benchmark analyses. Physics impact of budgetary constraints on things like tracking (higher Minimum pT cut etc.), HGCal granularity in reconstruction, HLT rate, ... (M. Pierini).*
- WG4: *CMS SW: frameworks, access to heterogeneous computing, architecture unaware programming (A. Bocci).*
- WG5: *CMS SW: algorithms for Run-4. Identification of resource critical algorithms / parts. Estimates of utilization of GPUs / accelerators / ... Impact of generators and Simulation / fast simulation (F. Pantaleo).*
- WG6: *Facilities and distributed computing. Data model, data lake, T0/HLT integration. HPC integration. Analysis facilities (J. Flix).*
- WG7: *R&D in CMS/HSF/WLCG/Industry/Countries (M. Girone).*

Report finalized last week.

ECoM2x Report

- This is not a public document, and I can thus not report on it here.
- It will (most likely) inform the public documents to be written for LHCC review later this year.
- The following are my opinions, rather than the conclusions of the report. So take them as such.
- I think these are (mostly) neither controversial nor surprising.

Fkw's opinions (I)

- Computing cost overruns pose a significant risk to HL-LHC operations that could prevent CMS from fully exploiting the physics potential of the HL-LHC
  - Primary cost drivers are reconstruction CPU and disk space for hosting data for analysis.
- Significant R&D is necessary to mitigate those risks.
  - Hardware Accelerators and vector units on CPUs provide significant cost reduction opportunities
  - Algorithms that work better with this hardware provide significant cost reduction opportunity
    - Kalman Filter is not such an algorithm, ML inference typically is, but so are other, more conventional options.
  - Much more aggressive use of tape and caching provides significant cost reduction opportunities.
    - Smaller fraction of our data needs to stay on disk for less time.
    - More aggressive, carefully planned use of networks can save disk space.

Fkw's opinions (II)

- Significant changes in how we traditionally have operated need to be considered to reduce costs.
  - No more prompt reco of the full dataset.
    - Growth in luminosity is slow, thus urgency to immediately process all the data is not warranted by the physics.
  - Algorithm and object selections can no longer be developed on the full HLT output bandwidth.
    - Only small fraction of total # of events can be on disk in AOD.
  - Retire the notion that we can afford x2-3 more MC events than we have data events taken, and arrive at an MC needs model that is linear in luminosity rather than # of events taken.
    - This avoids the close to x10 factor of increase in HLT output bandwidth when scaling up MC needs from Run 2 to Run 4.

Fkw's opinions (III)

- We can not afford to analyze the full HLT output rate in AOD, or even RAW or RECO formats.
  - Year 1 of HL-LHC data taking needs to focus on commissioning not just the detector and trigger but also the lower size data formats (MINI and NANO) with a carefully chosen subset of the full HLT rate.
    - This is much more important than doing physics with the early HL-LHC data because not doing it will prevent us from doing physics in year 2 and beyond, when there is enough luminosity accumulated to make doing physics with HL-LHC data worthwhile.

Summary & Conclusions

- I personally am convinced that the HL-LHC physics program can be accomplished within traditional budget guidance.
- However, doing so requires:
  - Some of the ongoing R&D to provide significant cost reductions.
  - Some social engineering to change habits of the collaboration.

Comments & Questions