CHEP 2018
23RD INTERNATIONAL CONFERENCE ON
COMPUTING IN HIGH ENERGY AND NUCLEAR PHYSICS
9-13 July 2018
National Palace of Culture
Sofia, Bulgaria

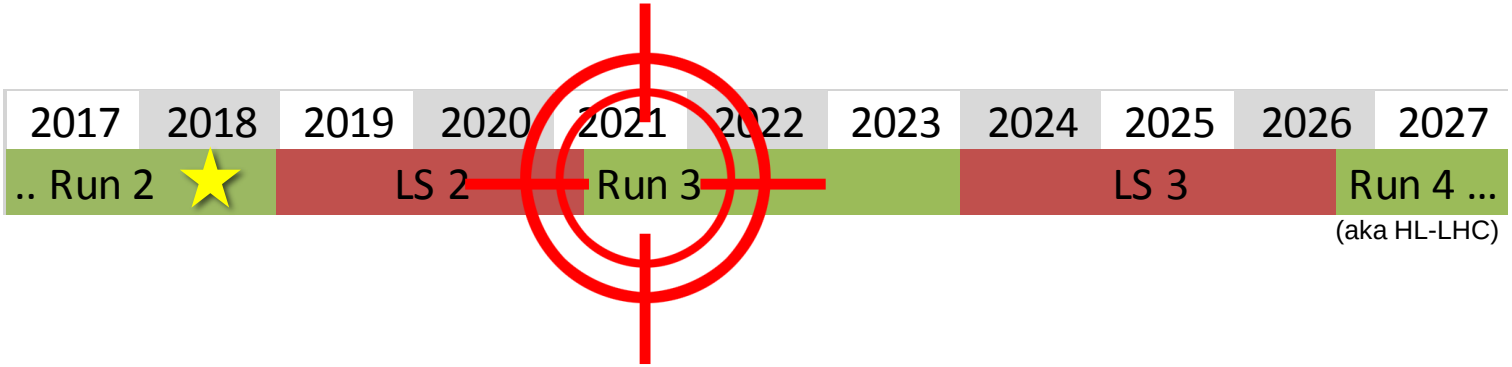
Towards a Computing Model for the LHCb Run 3 Upgrade

Concezio Bozzi[1,2], Stefan Roiser[1] on behalf of the LHCb Collaboration

[1]CERN, [2]Universita e INFN, Ferrara

23rd CHEP Conference, Sofia, Bulgaria

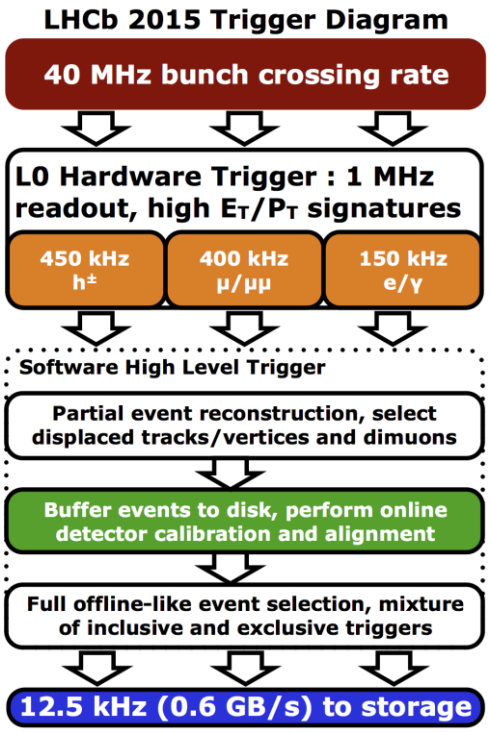
12 July 2018

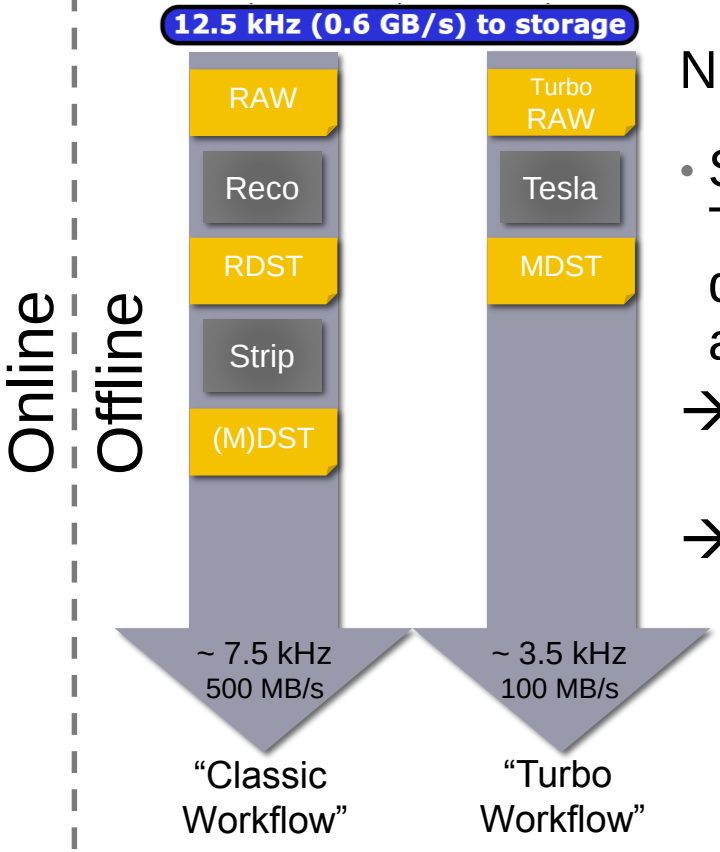


- Run 3 is the big step in the upgrade programme of the LHCb detector
  - Introducing major changes in sub-detectors, trigger and computing model
    - In comparison the upgrade to Run 4 (HL-LHC) will feature only minor changes

Processing Workflow: Run 2 (2015 – 2018)

[1] A. Dziurda





New in Run 2:

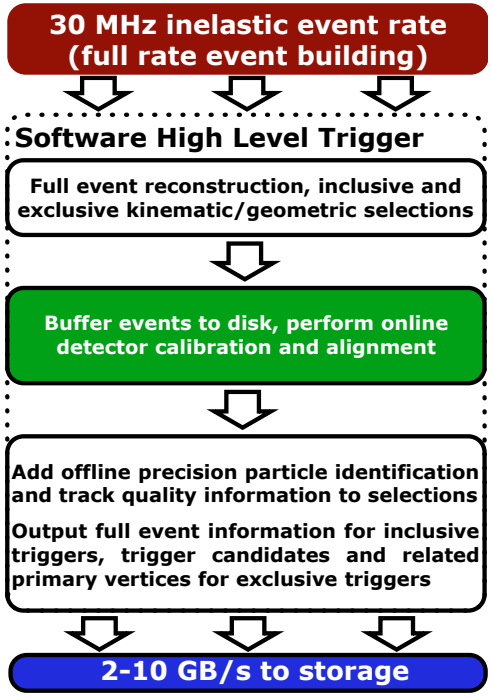
- Split High Level Trigger with final detector alignment and calibration

→ Offline quality reco in the HLT

→ Fully reco'ed smaller "Turbo" physics objects containing only necessary info

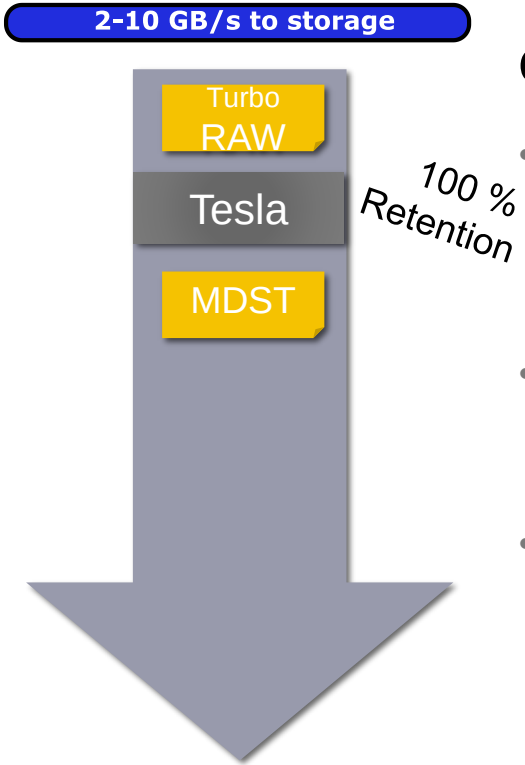
Processing Workflow: Run 3 (2021 – 2023)

[2] R. Matev
[3] S. Ponce



Online

Offline



"Turbo Workflow"

Changes in Run 3:

- Increase of luminosity and mu (x5)
  → Increase in offline simulation needs
- Remove HW trigger
  → Increase (x2) of trigger efficiency
- Vast majority of events in Turbo mode
  → Big increase in amount of collected signal information

LHCb Trigger

Today

Run 3

Monte Carlo Simulation

- Usage of grid resources for simulation will grow from ~ 80 % to ~ 90 %
- Despite pileup increase the workflow will stay same as Run 2
  - First step, no input data, w/ generator and detector response (95 % of work)
  - Second step with digitization and reconstruction
  - Optionally event filtering can be applied

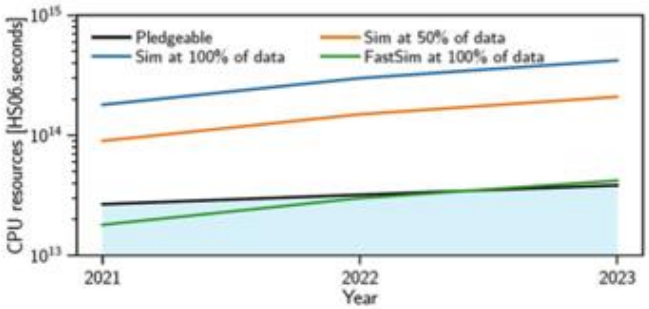
→ 1st step of simulation can be easily deployed on all resources

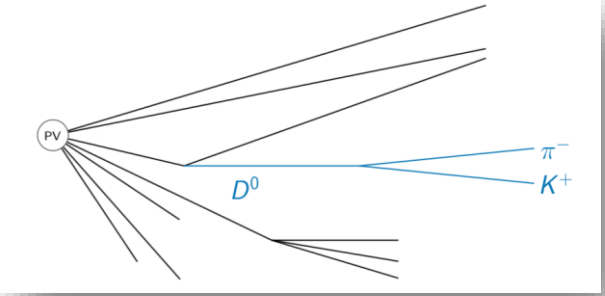
[4] D. Muller

- Simulation currently being moved to multi-threaded framework → reduction in memory usage opens new resource usage scenarios

Fast Monte Carlo

- CPU simulation needs scale with recorded integrated luminosity
  - Continuing with full simulation only not possible
- Working on fast Monte Carlo to keep resource needs under control
  - Several options already in production: ReDecay, RICHless, TrackerOnly, particleGun providing in general speedups between factor 5 and 10
  - More fast simulation workflows in preparation: Calo shower libraries and full parametric simulation based on Delphes,





ReDecay, inject signal information multiple times

[5] B. Siddi
[6] M. Rama

Data Processing & User Analysis

[7] F. Stagni

- Classic offline data reconstruction and stripping (streaming / skimming / slimming) reduced to bare minimum
- Main data processing workflow is turbo processing
  - i.e. convert online (LHCb specific) to offline (ROOT) format and streaming
  - In Run 2 this turbo workflow accounts for 0.1 ‰ of the grid work

- User analysis will move from individual to centrally organized data selections
  - Possibility to increase I/O by aggregating multiple selections (train model)

Storage Requirements

- Storage needs are driven by data of HLT output bandwidth
  - Tape needs incompressible, while mitigations possible for disk
    - E.g. parking scenarios are considered but introduce additional operational costs for the experiment and infrastructure costs for sites
- MC simulation output data format mostly migrated to m(icro)DST format with small contribution to needs introducing a size reduction of factor 20

- LHCb relies on a small amount of sites with disk storage:
  - T0 + 7 T1s + 13 T2s with minimum size requirements especially for T2s
  - Data caching especially on ”small disk sites” is not a major use case

Data Movement

- Introduce multiple streaming layers to keep data set size under control
  - O(10) streams from Online, O(100) streams Offline
  - Expect on average 500 TB per data set / data taking year
  - In case of parking these need to be staged in due time, O(days)

- Throughput to/from tape systems will increase by several factors

- WLCG/DOMA initiative welcome to possibly further reduce costs
  - Especially optimizations on the timescale of Run 3

Resource Needs Based on Computing Model

- Basic assumptions are
  - 10 GB/s HLT output bandwidth with almost all data in Turbo stream
  - Run 2 simulation ramping down between 2021 and 2023
  - Percentage of fast and parametric simulated events gradually increasing
  - Most simulation output data in mDST format
  - ”Commissioning year” in 2021, full data taking in 2022, 2023
- Resource needs on average above extrapolations for the Run 3 years
- Working on additional scenarios with more mitigation measures

Summary

- LHC Run 3 is the major change of the LHCb upgrade programme
  - Resulting in a step up in offline resource needs for the experiment
- Several mitigation measures to keep resources under control are in place / under development
  - Turbo stream, fast simulation, mDST format, central analysis
- A computing model document will be submitted to the LHCC by the end of this year

Thank you for your attention !!!

[1] A. Dziurda et al., LHCb full-detector real-time alignment and calibration: latest developments and perspectives
[2] R. Matev, A 30MHz software trigger for the LHCb upgrade
[3] S. Ponce et al., The core software framework for the LHCb upgrade
[4] D. Muller and D. Popov, Adopting new technologies in the LHCb Gauss simulation framework
[5] B. Siddi and A. Davis, A fully parametric option in the LHCb simulation framework
[6] M. Rama and G. Vitali., Fast calorimeters options in the LHCb Gauss simulation framework
[7] F. Stagni, LHCb and DIRAC strategy towards the LHCb upgrade