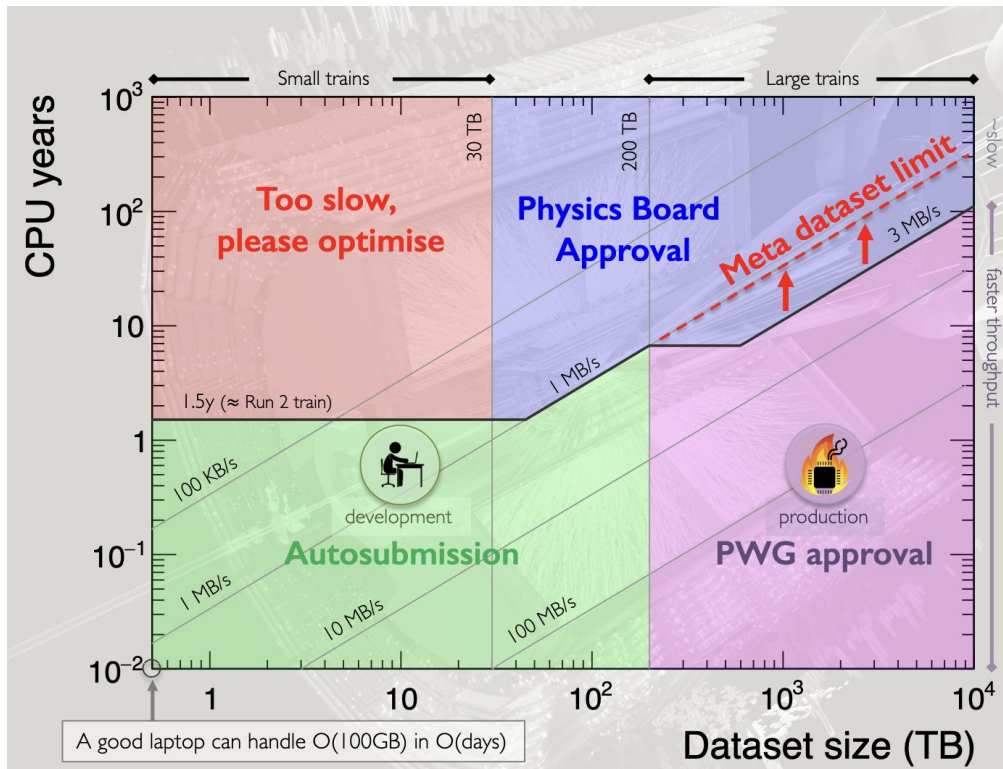# Hyperloop policy and DQ stream datasets

C. Jahnke, L. Massacrier, L. Micheletti
based on the contributions/inputs from O2/DQ members

Hyperloop policy

From D. Chinellato at PB

Running (slow) trains on large datasets require large resources and PB approval:

- Important to optimize the code (CPU, throughput) to reduce resources
- Mutualize the resources among analyses → **Use of derived datasets produced by the PWG (should be the standard procedure to run analyses)**
- Current hyperloop policy favors the usage of metadatasets agreggating runs from several periods

First DQ derived datasets produced for QM2023 analyses (based on unskimmed data): Muons

- ❏ Productions prepared for QM2023 preliminaries
- ❏ Used for muon analysis but **they contain also electrons (not slim dataset!)**
- ❏ The production is done using both the **ppFilter** and the **tableMaker**

  - ➢ **d-q-filter-p-p-task** ([link to the configuration](#))
    - ○ cfgBarrelSels: jpsiO2MCdebugCuts7_Corr::2
    - ○ cfgMuonSels: MuonLow::1

  - ➢ **table-maker** ([link to the configuration](#))
    - ○ cfgBarrelTrackCuts: jpsiO2MCdebugCuts7_Corr, kaonPID
    - ○ cfgMuonCuts: muonQualityCuts, matchedGlobal

  - ● **LHC22o**: DQ_LHC22o_pass4_muon_electron ([link](#))
  - ● **LHC22m**: DQ_LHC22m_pass4_muon_electron_new ([link](#))
  - ● **LHC22r**: DQ_LHC22r_pass4_muon_electron ([link](#))
  - ● **LHC22t**: DQ_LHC22t_pass4_muon_electron ([link](#))

First DQ derived datasets produced for QM2023 analyses (based on unskimmed data): Electrons

- The motivation is to have a slim data of 2022 pp at 13.6 TeV collision for the QM preliminary results preparation
- Used the **o2-analysis-dq-filter-pp** wagon plus the **o2-analysis-dq-table-maker**

Cuts for skimmed tree creation

➤ p_T > 1 GeV/c
➤ |η| < 0.9
➤ TPCncls > 70
➤ TPCchi2 < 4
➤ SPDany

➤ -4 < TPC$n\sigma^e$ < 4
➤ TPC$n\sigma^p$ > 2.5
➤ TPC$n\sigma^\pi$ > 2.5

⬉ **Derived data**

Store	Binding	Description
☑	ReducedEvents	REDUCEDEVENT
☑	ReducedTracks	REDUCEDTRACK
☑	ReducedEventsExtended	REEXTENDED
☑	ReducedTracksBarrelPID	RTBARRELPID
☐	ReducedMuons	RTMUON
☐	ReducedMuonsCov	RTMUONCOV
☐	ReducedMuonsExtra	RTMUONEXTRA
☑	ReducedEventsVtxCov	REVTXCOV
☑	ReducedTracksBarrel	RTBARREL
☑	ReducedTracksBarrelCov	RTBARRELCOV

link of the TPC post-calibration maps:
Users/z/zhxiong/Postcalib/pass4/apass3BB

- The output with corresponding train numbers, the total output is around 75G for analysis, it can be run locally.
  - LHC22m pass4 (60.61B events) / train number 104249/ total size of the output. 67 G (approval from PWG only)
  - LHC22o pass4 (312.60B events) / train number 103617/ total size of the output. 452 G (need the approval from PWG and PB)
  - LHC22r pass4 (75.83B events)) / train number 104251/ total size of the output. 87 G (approval from PWG only)
  - LHC22t pass4 (75.62B events)) / train number 104250/ total size of the output. 87 G (approval from PWG only)

Derived dataset campaign: <u>short term future plans</u>

- 2022 data were re-organized according to the interaction rate:

    - LHC22_pass4_lowIR: https://alimonitor.cern.ch/hyperloop/view-dataset/560
    - LHC22_pass4_highIR: https://alimonitor.cern.ch/hyperloop/view-dataset/582
    - LHC22o_pass4_minBias:https://alimonitor.cern.ch/hyperloop/view-dataset/433

- <u>New streaming campaign in preparation with the same logic of QM2023 preliminaries</u>

- The <u>pass4</u> reconstruction is still affected by some issues (φ angle bug, misalignment of MFT, …) so a new reconstruction will be needed as soon as the 2022 (skimmed) data will be reprocessed

- The same streaming procedure will be performed for 2023 data once new reconstruction will be available

Discussion on definition of DQ streams for future analyses on skimmed data

- Main strategy :
    - PB approval needed only to produce the centralized derived datasets
    - Produce outputs/trees that analyzers can run on (frequently if needed) without PB approval, but with simple PWG approval
        - Requires a good preparation of the streams to have all the needed tools included for most analyses, keeping at the same time the streams at a reasonable size for efficient running

Discussion on definition of DQ streams (pp)

For the moment two types of stream have been discussed:

- An "inclusive stream" :
    - Keep all triggered OS* dilepton (dimuon+dielectron) events and all tracks in the events after simple quality criteria (keep covariances as well for 2ndary vertexing)
    - Store global event properties (multiplicity from several estimators), luminosity counters…
    - Stream needed for forward/mid rapidity correlation, quarkonium versus multiplicity….

* in the skimmed pp data we filter on OS dileptons (--> LS heavily downscaled due to this selection)

- A "dimuon stream" :
    - Keep all triggered OS dimuon + global event properties (no barrel track information)
    - Smaller stream for basic forward y quarkonium analyses (production, prompt/non prompt charmonia, Bc)

- Additional dielectron stream (might) be useful for basic quarkonium production analyses at mid-y

→ Need studies of the output size versus variables included in the stream (contact barrel : Leopoldo, muons : Emilie)

Discussion on definition of DQ streams (PbPb) - first thoughts

- A large fraction of events contain (OS, LS) dileptons in PbPb, further reduction needed with respect to pp selections from filterpp task
- In "inclusive stream", non-dilepton information (ie. other event tracks properties) needs to be largely reduced (to few variables)
- dedicated muon stream (dimuons+global event properties, no barrel tracks) and barrel streams (dielectron + global event properties for inclusive analyses, stream with more information for mid-y/forward-y correlations but with stringent cuts on the candidates….) could also be foreseen

→ Need studies to define the tracks/event cuts, the output size versus variables included in the stream (looking for contact people)