Feedback on computing upgrade from B&Q

B&Q working group

Data format in between DST and muDST

Q: What is the minimal set of information you need for analysis that currently are on DST? As an example, often isolation variables are calculated within a cone and what lies outside of this cone is not used at all and could be ignored. We would like to have some idea of how this cone can be defined and a quantitative estimate of the "size" of it.

A: We are investing some time in these studies because already since 2017 we would like to have some Turbo lines with this format. To exploit this format, we should keep in mind that we need to maintain "data mining" capabilities, especially when reaching new luminosity levels. We should be careful not to make cuts which will limit scope of data mining. The cone seems a bit dangerous for spectroscopy analyses. Sebastian is now studying the effect of it. As another example, doing channels with K0s may become more interesting. Preserving all tracks making Ks and Lambda candidates in addition to a cone or near-the-same-PV criterion may be useful. But there are also other options under consideration, e.g. removing tracks belonging to far-away PVs may be an effective way to reduce size without losing any of detached tracks. Yanxi, Marco, Sebastian have already expressed interest in the developments and contributed to the first validation of Turbo 2.0. In the event of an excess of person power in our WG we will be happy of addressing newcomers to this important item!

Turbo stream

Q: How many analysis in your WG can not be moved to turbo?

A: The current status of Turbo, with a trigger line per decay mode, is not sustainable to move all of the J/psi analyses to Turbo. Further developments of Turbo allowing inclusive selections, possibly with the opportunity of storing some row banks (muons?), would make it more appealing. Besides, it should be considered that opportunities for data mining and searches for rare states will benefit from charm hadron produced in the b decay thanks to a trigger strategy reducing the importance of the hardware stage, effectively increasing the selection efficiency of hadronic b decays. The study of techniques to allow for searches has therefore an interest beyond the inclusive dimuon selections.

Centralised ntuple production

Q: Would you be happy if ntuples could be produced every couple of weeks centrally by the production team?

A: This would be great for 95% use cases. Still, it cannot replace user access to DaVinci data for cases where the algorithms are not committed and released. It is important to offer users test samples large enough to test the nTuple production scripts with some statistical significance. The problem with this strategy is that it requires strict code review policies to avoid that my script breaks all of the others running together, and on the other hand a number of scripts of of the same order of magnitude as stripping lines should be expected. This risks to overload the maintainers/code reviewers making the user production (that I hope you don't want to deny) more effective. An automatic exclusion of failing scripts, at run-time, could be investigated.

Event index

Q: The idea is that events are indexed according to trigger/stripping line they fired and everything is put into a sort of database. The users will make a query to the database which will return a list of events. The gain from this approach is a much faster access to the events and allows an optimisation of data replication. We would need input from the WG to unserstand what information should go into the index (e.g. Stripping lines, Trigger lines, multiplicities etc)

A: It would be also a very nice improvements. To the obvious items you listed (Stripping lines and Trigger lines) I would certainly add the number of each CommonParticle. If I want to run on all the lines with a J/psi and a D*, I wouldn't like to read all the events triggered by a J/psi and check if there is also a D*, when I could simply query the index with a nice "SELECT * WHERE TriggerLine == 'Jpsi' AND nDstar > 0".