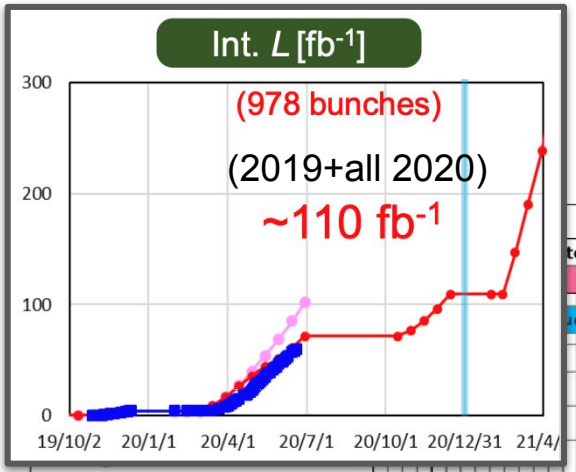
Data processing remarks for 2020c

Data Production Workshop
08/09/2020

Stefano Lacaprara, Marco Milesi
INFN Padova, Uni Melbourne

Estimate for 2020c data taking

Expect ~ 40 fb^{-1} during 2020c. Total ~110 fb^{-1} for all 2019 + 2020



Start up
- Vacuum scrubbing (40 Ah)
- Recovery to 2020b condition
- Injection study

- Squeezing β_y^* to 0.8 mm, 0.7 mm, 0.6 mm
- Injection study/tuning
- Background study
- Beam-beam / optics study
- Other hardware/software studies

- Machin studies if necessary
- Optics study

Stefano Lacaprara, INFN Padova - Marco Milesi, University of Melbourne

Phase III - Input RAW data size / fb^{-1}

Campaign	Int. lumi [fb-1]	RAW data size [TB / fb-1]							
		ALL events - no filter	ALL events - w/ filter [*]	bhabha	bhabhaecl	gamma_gamma	hadron	mumu_2trk	cosmic
proc11 (e8, filter on)	1.72	---	4.7	1.7	---	0.2	0.3	0.1	---
bucket10	10.4	12.8	3.8	3.3	1.4	0.6	0.3	0.1	0.2
bucket11	12.7	12.3	3.7	3.3	1.5	0.6	0.4	0.1	0.2
Bucket12 (off-res)	2.4	13.1	3.9	3.9	1.7	0.6	0.3	0.1	0.1
bucket13	5.0	11.2	3.4	3.8	1.5	0.6	0.4	0.1	0.1
bucket14	9.7	13.0	3.9	3.5	1.5	0.6	0.4	0.1	0.1
bucket15	19	13.2	4.0	3.5	1.5	0.6	0.3	0.1	0.1

[*] Late 2020b has 25-30% retention.
Was 40% in proc11

RAW HLT hadron takes ~ 0.3 TB/fb^{-1}. NB: none of HLT skims is prescaled.

W/ HLT filter on, expect + 8-10% disk usage per fb-1 (evt. duplication) to store RAW HLT hadron skims on disk.

Stefano Lacaprara, INFN Padova - Marco Milesi, University of Melbourne

Phase III - T_{CPU} (HLT skims)

NB: stats from local KEKCC processing → expecting no much change on grid

Campaign	Int. lumi [fb-1]	Average T_{CPU} / evt [ms]				
		bhabha	bhabhaecl	gamma_gamma	hadron	mumu_2trk
bucket10	10.4	1.4	1.4	1.3	1.9	1.7
bucket11	12.7	1.6	1.6	1.4	2.2	2.0
bucket14	9.7	1.6	1.6	1.4	2.3	2.0

Campaign	Int. lumi [fb-1]	Average T_{CPU} / job [h]				
		bhabha	bhabhaecl	gamma_gamma	hadron	mumu_2trk
bucket10	10.4	2.1	2.1	1	7.1	4.4
bucket11	12.7	2.1	2.1	1	7.0	4.4
bucket14	9.7	2.1	2.1	1	7.0	4.3

Stefano Lacaprara, INFN Padova - Marco Milesi, University of Melbourne

Phase III - T_{CPU} (all events, grid)

Campaign	Int. lumi [fb-1]	Average T_{CPU} / evt [ms] (RawProcessing)	Average T_{CPU} / job [h] (RawProcessing)
bucket10	10.4		2.7
bucket11	12.7		2.8
bucket13	9.7		3.6

Phase III - Effective turnaround time

Campaign	Int. lumi [fb-1]	HLT skim (local) - N_{cores} = 2500		All Events (Grid) - $\langle N_{cores} \rangle$ = 2000 (peak ~5000)			
		All HLT skims	HLT hadron	RawProcessing		RawProcessing + Merge	
		$T_{turnaround}$ [days/fb^{-1}]		$T_{turnaround}$ [days]	$T_{turnaround}$ [days/fb^{-1}]	$T_{turnaround}$ [days]	$T_{turnaround}$ [days/fb^{-1}]
bucket10	10.4	0.4	0.1	9	0.9	17	1.6
bucket11	12.7	0.3	0.1	10	0.8	14	1.1
bucket13	5.0	0.5	0.1	8 **(had+udst 0.7)**	1.6 **(had+udst: 0.14)**	15 **(had+udst 13)**	3.0 **(had+udst 2.6)**
bucket14	9.7	0.3	0.1	28 [d] going [*][**]		28 [d] going [*][**]	

NB: bucket13 and 14 first attempt at processing simultaneously:

- RAW ALL events → mDST(all)
- RAW HLT hadron → mDST(hadron) + D* uDST ("systematic" skim)

Main reason for delays on grid production is Merge step

[*] NB: KEKCC replacement, BelleDIRAC migration, post ICHEP holidays…

[**] RawProcessing stuck on DDM (due to KEKCC replacement?)

Data processing schema for 2020c

Two-fold goal:

1. Prioritise fast processing of RAW HLT **unprescaled** hadron inputs (~8% retention)
   a. Most of *physics* (B-phys) starts from it.
   b. Some disk overhead (~10% duplication), but less stress for staging.
   c. NB: at current stage, unfeasible to duplicate data for other HLT skims.
2. Drive users towards uDST (i.e., analysis/systematic skims), or at least HLT-skimmed mDST → (much) smaller and faster to analyse.

NB: this does not cover calibration plans!

Data processing schema for 2020c

Extending “bucket14” schema, we propose a ***staged* processing** of:

1. **I:** RAW HLT hadron - **O:** *simultaneous* production of mDST(hadron) + *few* compatible uDST (syst.)
2. **I:** RAW ALL events (filter on) - **O:**
   a. Baseline: production of mDST(all) **only**
   b. Advanced: *simultaneous* production of mDST(all) + *few* mDSTs(low retention* HLT skims), eg 4lep, or similar needed by phys perf.
      i. 👍Reduce CPU usage and disk staging time. 👎Increase merging complexity.
      ii. Production of mDST selected w/ different HLT skim flags not an issue for input staging: same input.
      iii. Output (mDST) event duplication and CPU overhead shouldn’t be a problem (for now).
      iv. HLT tau skims have too large retention rate (~40%) to fit in current scheme.
3. Production of uDSTs (not already in step 1) to follow completion of 1. (and 2.)
   a. Can either use mDST(hadron/...) or mDST(all) as input if appropriate to speed up.

* < 20%

No more Unofficial for 2020c (sorry, but not sorry). Any objections?

Stefano Lacaprara, INFN Padova - Marco Milesi, University of Melbourne

Question time

- Is this feasible from computing's point of view?

Campaign processing plan for 2020:

- Baseline DS for Moriond 2021:
  - proc11 (2019, e7-8-10) + prompt (2020a-b, e12, buckets 9-15) → release-04, ~75 fb^{-1}
    - Currently, only bucket15 missing…
    - MC13a, MC13b
- Re-processing of bucket15 (or part of), e12 w/ updated calibration → release-05
  - Validate against same bucket in release-04
- 2020c prompt processings → release-05.
  - Available for Moriond for Phys Perf
- Full 2019+2020 reprocessing (proc12) at EOY → release-05:
  - Depending upon calibration plans.
  - Depending upon MC14 readiness:
    - MC14a (run-independent) can start early October w/ next minor release.
    - MC14b (run-dependent) requires BG staging on grid.
  - Unlikely to be ready for Moriond...

Lessons from 2020a-b

- Production of mDST + (single) uDST from RAW HLT hadron is working:
  - Some bits to be refined (final path on the grid, metadata, registration on DS, …)
  - Merge step is (as usual) the most critical part. Hoping for Rucio to help...
- Leave more complex production workflows for next year:
  - Eg: produce simultaneously *all* possible uDSTs from HLT hadron…
  - Likely introduction of significant extra burden for DC.
- Staging of large datasets works, but need to be coordinated well w/ DC.
  - Well defined schedule will help.
  - Automation would be very welcome! (for staging and un-staging)
- Often grid issues require direct manual intervention from DC experts:
  - They must be praised for being always fast-acting and helpful, but they’re human beings after all :), and extremely manpower-limited.
  - DP managers have limited control and little-to-no expertise about subtleties of DC infrastructure (DIRAC).
  - Ideally it would be very useful to have a DP-DC liaison to help tools development and maintenance.

CPU usage at BNL

Up to mid July