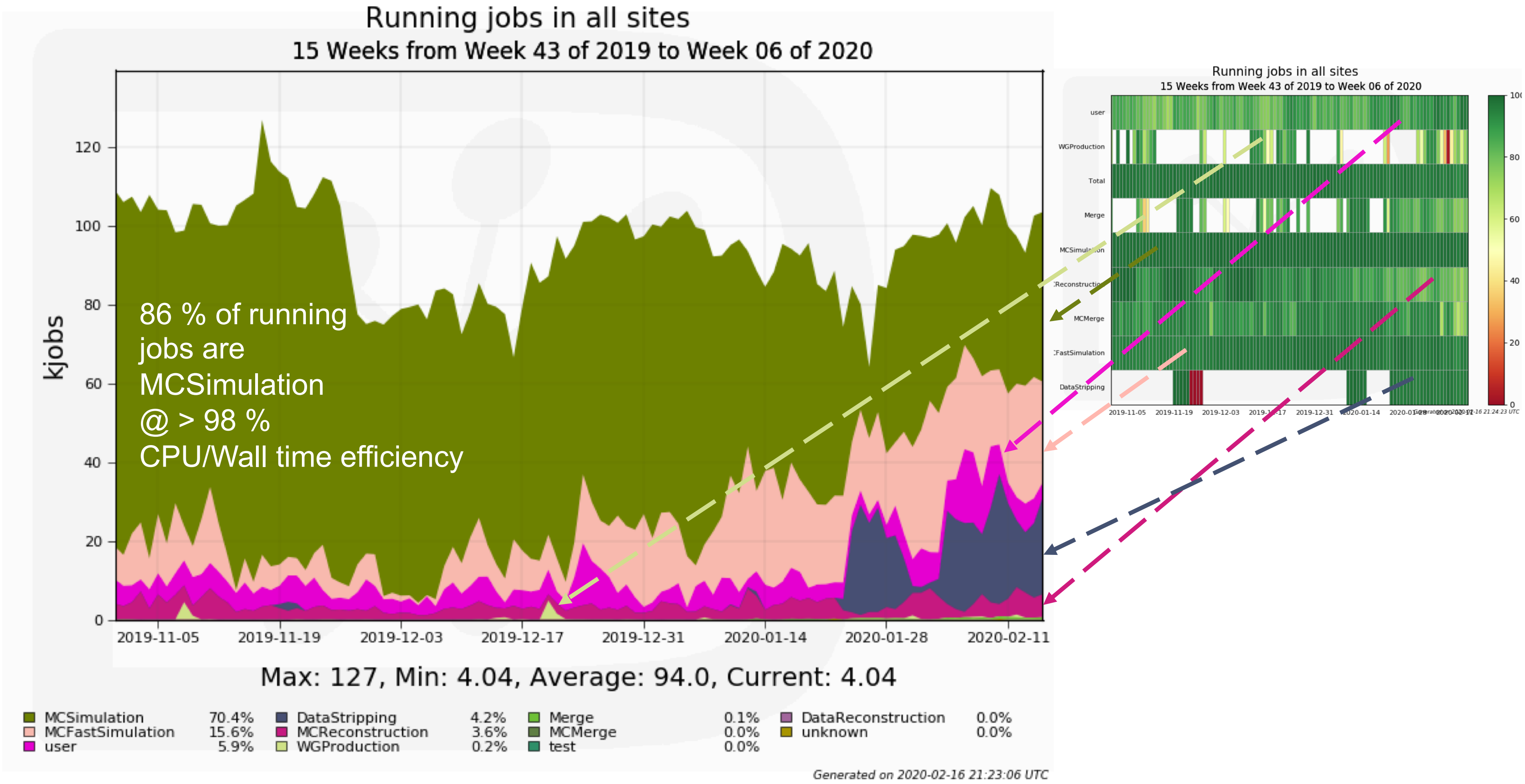
LHCb Software & Computing

Concezio Bozzi

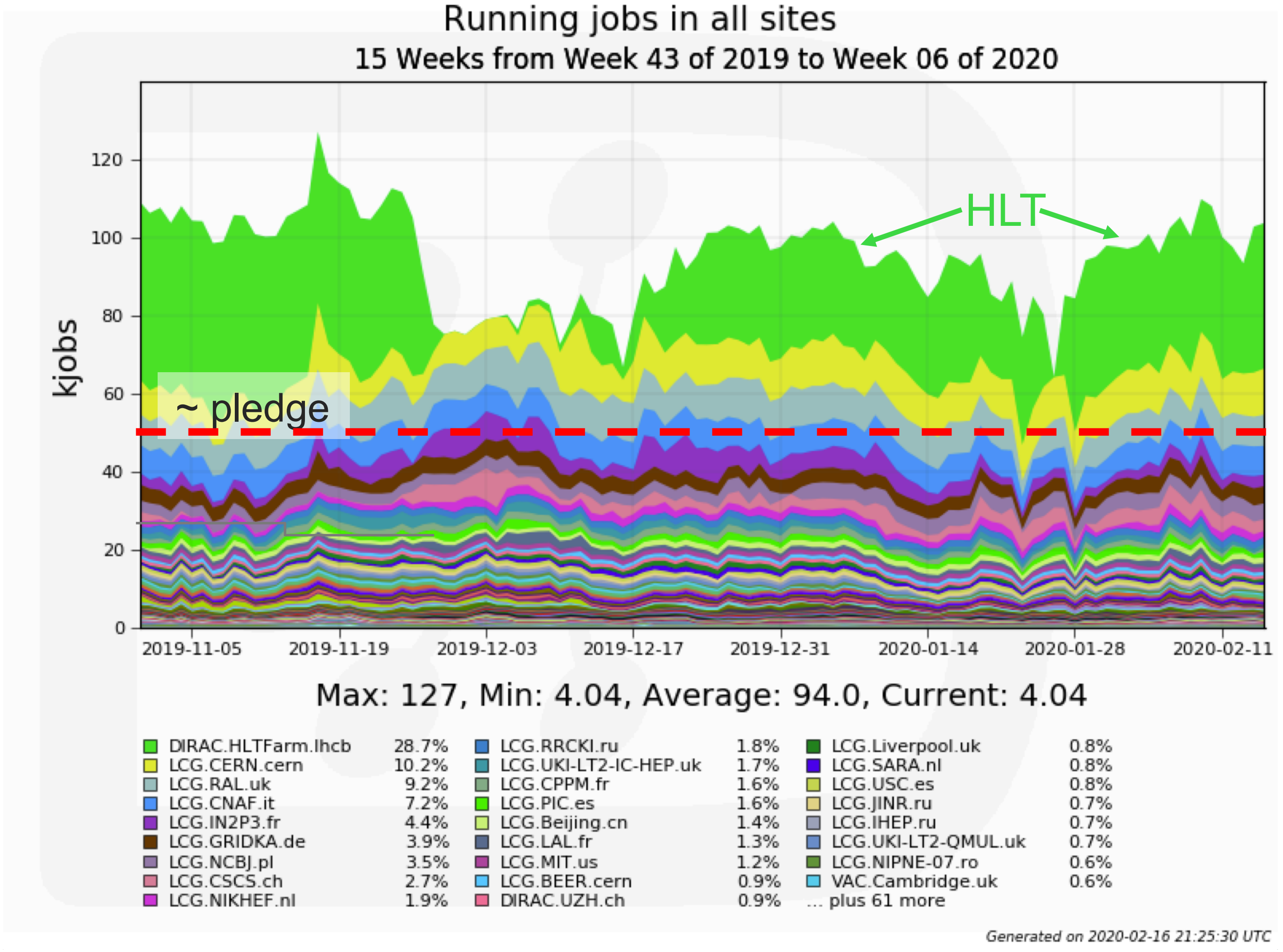
WLCG/LHCC Referees

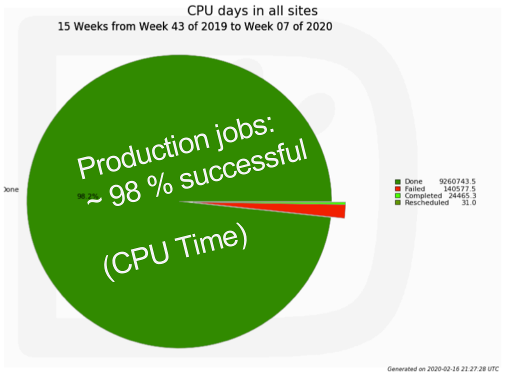
18 February 2020

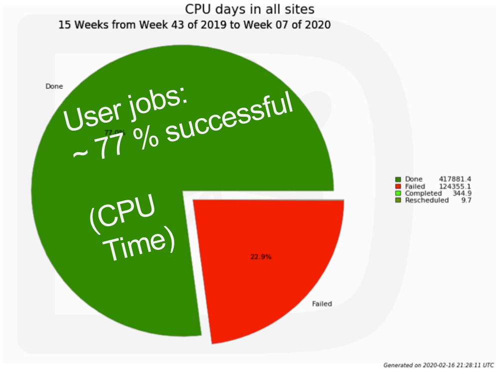
Operations



Running jobs in all sites
15 Weeks from Week 43 of 2019 to Week 06 of 2020
kjobs
0
20
40
60
80
100
120
2019-11-05
2019-11-19
2019-12-03
2019-12-17
2019-12-31
2020-01-14
2020-01-28
2020-02-11
86 % of running jobs are MCSimulation @ > 98 % CPU/Wall time efficiency
Max: 127, Min: 4.04, Average: 94.0, Current: 4.04
MCSimulation 70.4%
MCFastSimulation 15.6%
user 5.9%
DataStripping 4.2%
MCReconstruction 3.6%
WGProduction 0.2%
Merge 0.1%
MCMerge 0.0%
test 0.0%
DataReconstruction 0.0%
unknown 0.0%
Generated on 2020-02-16 21:23:06 UTC
Running jobs in all sites
15 Weeks from Week 43 of 2019 to Week 06 of 2020
user
WGProduction
Total
Merge
MCMerge
DataStripping

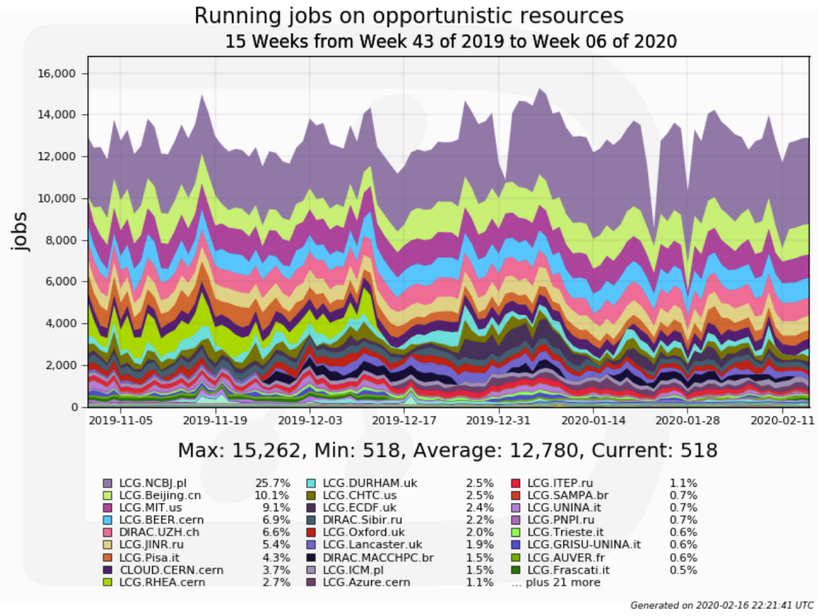

Running jobs in all sites
15 Weeks from Week 43 of 2019 to Week 06 of 2020
kjobs
0
20
40
60
80
100
120
~ pledge
HLT
2019-11-05
2019-11-19
2019-12-03
2019-12-17
2019-12-31
2020-01-14
2020-01-28
2020-02-11
Max: 127, Min: 4.04, Average: 94.0, Current: 4.04
DIRAC.HLTFarm.lhcb 28.7%
LCG.CERN.cern 10.2%
LCG.RAL.uk 9.2%
LCG.CNAF.it 7.2%
LCG.IN2P3.fr 4.4%
LCG.GRIDKA.de 3.9%
LCG.NCBJ.pl 3.5%
LCG.CSCS.ch 2.7%
LCG.NIKHEF.nl 1.9%
LCG.RRCKI.ru 1.8%
LCG.UKI-LT2-IC-HEP.uk 1.7%
LCG.CPPM.fr 1.6%
LCG.PIC.es 1.6%
LCG.Beijing.cn 1.4%
LCG.LAL.fr 1.3%
LCG.MIT.us 1.2%
LCG.BEER.cern 0.9%
DIRAC.UZH.ch 0.9%
LCG.Liverpool.uk 0.8%
LCG.SARA.nl 0.8%
LCG.USC.es 0.8%
LCG.JINR.ru 0.7%
LCG.IHEP.ru 0.7%
LCG.UKI-LT2-QMUL.uk 0.7%
LCG.NIPNE-07.ro 0.6%
VAC.Cambridge.uk 0.6%
... plus 61 more
Generated on 2020-02-16 21:25:30 UTC


CPU days in all sites
15 Weeks from Week 43 of 2019 to Week 07 of 2020
Production jobs:
~ 98 % successful
(CPU Time)
Done 9260743.5
Failed 140577.5
Completed 24465.3
Rescheduled 31.0
Generated on 2020-02-16 21:27:28 UTC


CPU days in all sites
15 Weeks from Week 43 of 2019 to Week 07 of 2020
Done
User jobs:
~ 77 % successful
(CPU Time)
22.9%
Failed
Done 417881.4
Failed 124355.1
Completed 344.9
Rescheduled 9.7
Generated on 2020-02-16 21:28:11 UTC

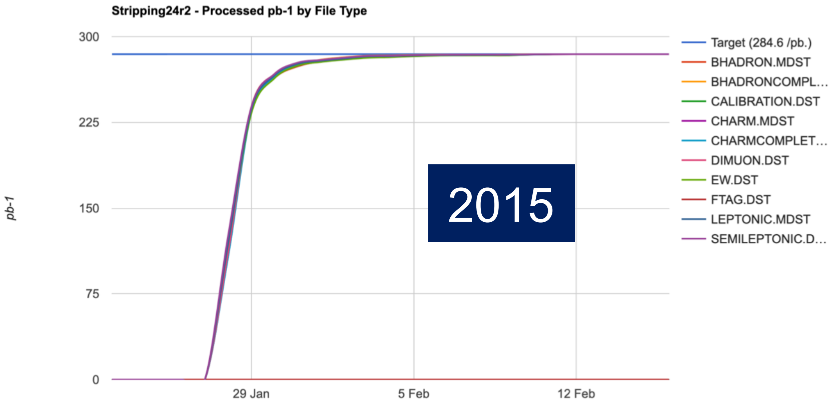
Opportunistic resources and HPC centers

- O(10%) contribution to the total
  - Mostly WLCG sites not pledging to LHCb
  - Other contributions: BEER, UZH, clouds, Condor HTC, Sibir, SDumont HPC (Brazil)
- Still working on expanding pool of opportunistic resources
  - DIRAC being updated for running on worker nodes with many-core and multi-process / multi-threaded applications
  - Simulation application Gauss ported to ARM, validation tests running on LHCbPR system

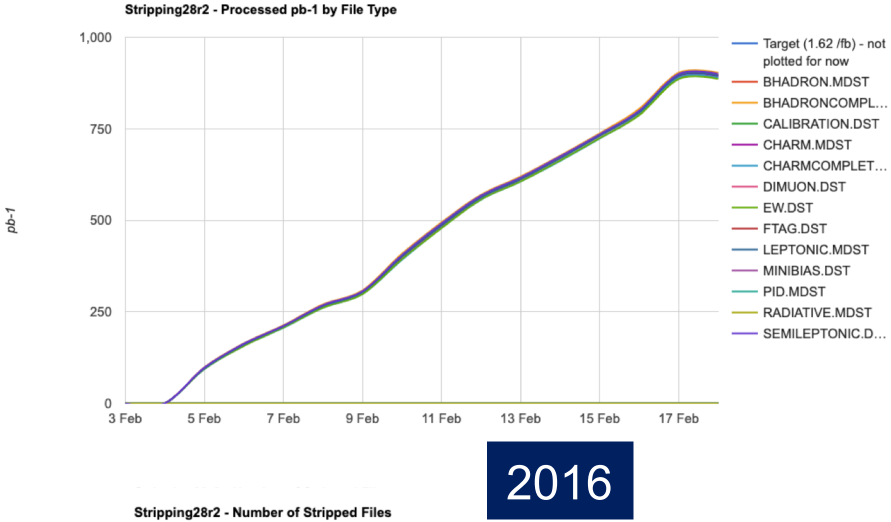


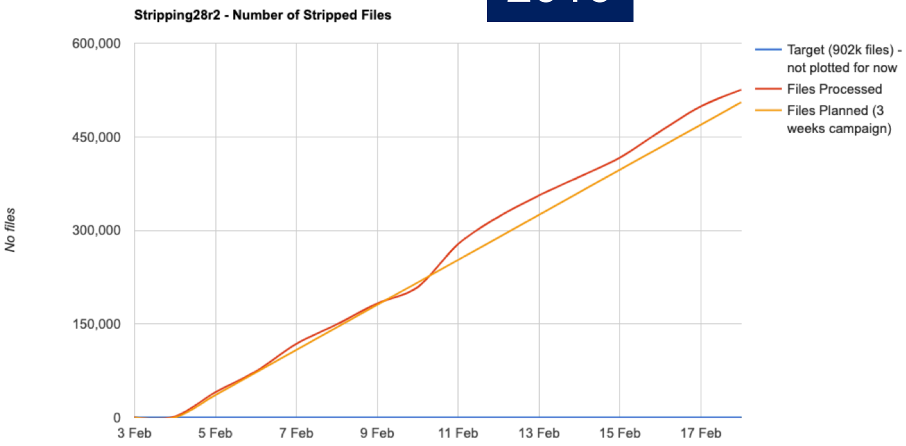
Data productions

Legacy stripping campaigns for Run1+ 2015+2018 complete. Now running 2016, then 2017

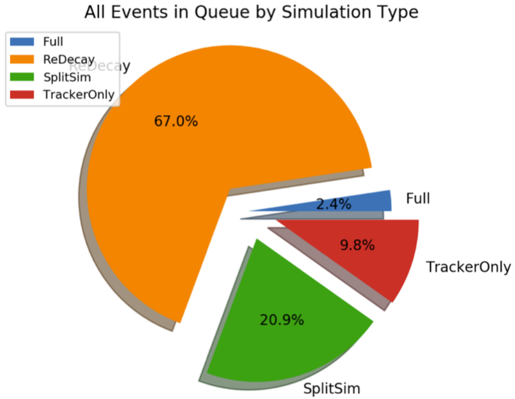


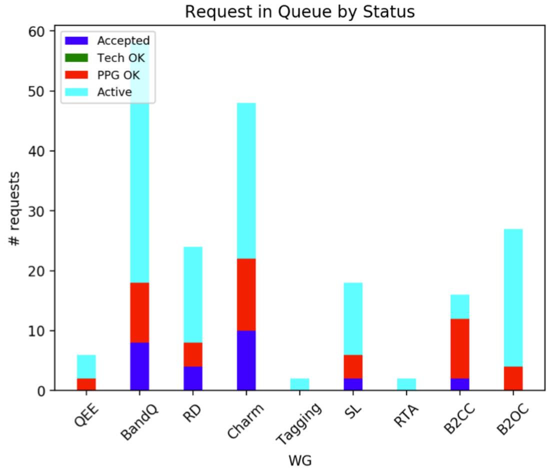
- Massive data recalls from tape systems
- Plan to finish processing of data by this spring

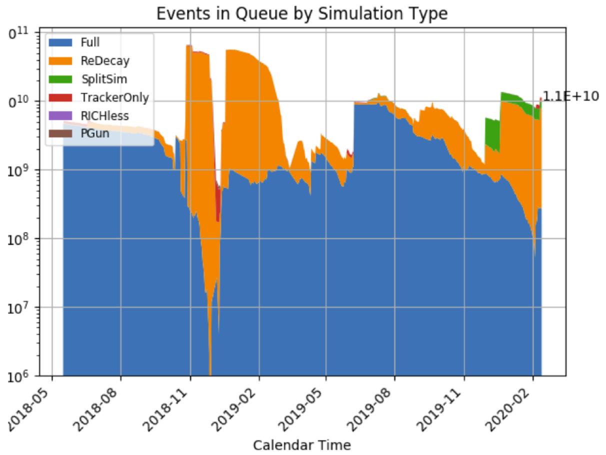




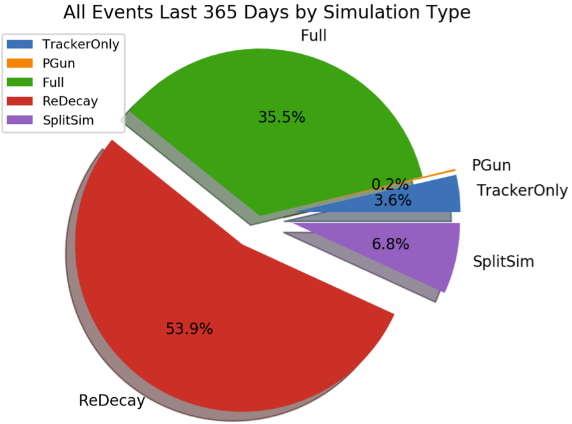
MC productions

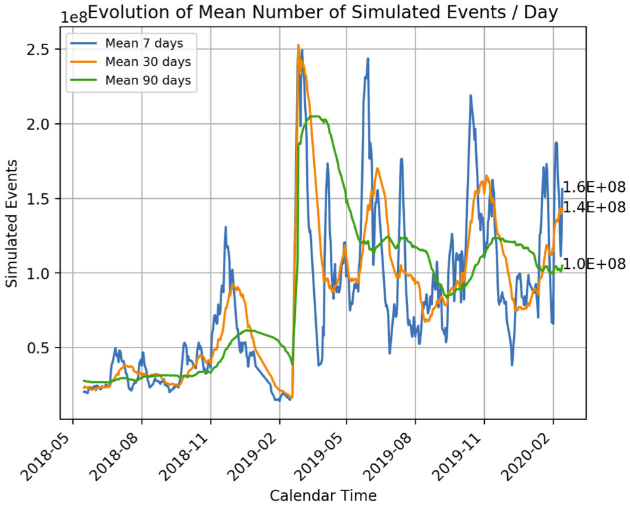






- Dominated by fast simulation
- Most productions are filtered before being written to disk

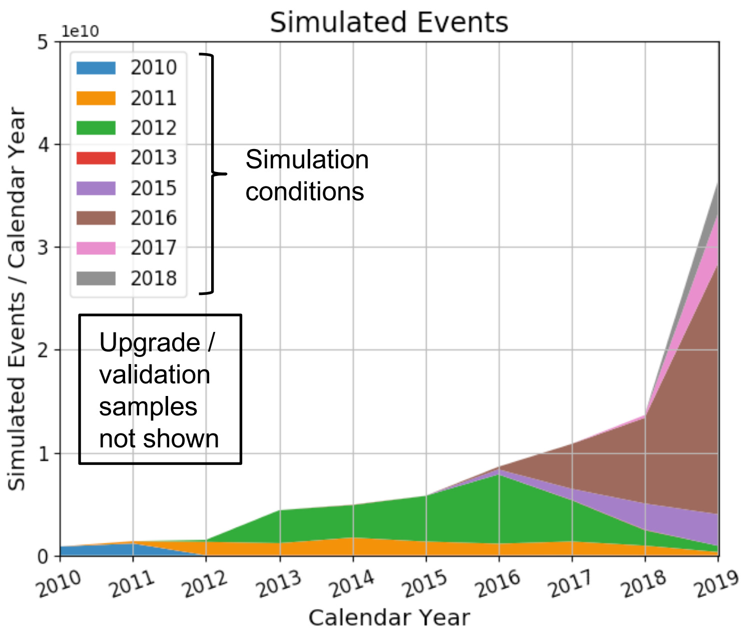




MC productions

- Factor ~4 increase in number of events produced in 2019 wrt previous years
- CPU work increased only by 20% per year

Year	Simulated events (10^9)	Stored events (10^9)	Ratio	CPU work kHS06.y	CPU per event kHS06.s	Disk TB
2017	10.3	4.2	40.3%	817	2.50	640
2018	12.0	3.0	25.3%	1009	2.65	550
2019	45.0	6.9	15.2%	1290	0.90	1110



Run3 preparation

Offline computing requests for 2021

- Requests being prepared for the C-RSG
- Following model presented in the LHCb Upgrade Computing Model TDR (CERN-LHCC-2018-014; LHCb-TDR-018), adjusted for
  - LHC schedule (following LHC program coordination meeting of January 20th 2020)
  - LHCb data taking plans (see here)
    - Commission detector during LHC beam commissioning
    - Start data taking with «commissioning μ» (~1.1) during LHC intensity ramp-up
    - Perform detector tuning, calibration&alignment, monitoring, while LHC increases the number of bunches
    - Move gradually to «nominal μ» (~7) afterwards
    - Take data during Pb-Pb Ion run (configuration details TBD)
  - Additional resources required for heavy-ion and fixed-target collisions

Core software

- Continuing progress on the core software framework
- The Gaudi framework can handle both the current LHCb «DetDesc» geometry/condition system and the new one based on DD4hep.
- Development is ongoing for the migration of the sub-detectors and their software to DD4hep.
- Software hackathons continue to be a successful initiative for code development

Network requirements

- Shown at LHCOPN/LHCONE workshop on January 13th:
- The LHCb data distribution model is very simple and it will not change
- The throughputs measured in Run2 scale by appropriate factors
- Total ballpark estimate, with O(20%) contingency:

20GB/s on average, Run3 ~ 10x Run2

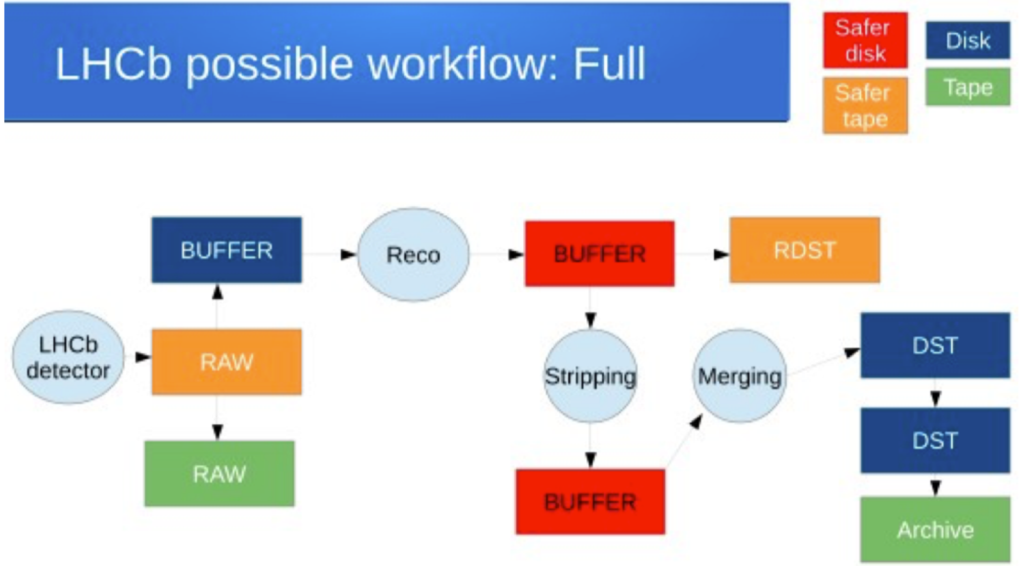
Data type	Source	Destination	Run2 (GB/s)		Run3/4 (GB/s)	
			average	peak	average	peak
Primary datasets	T0	T1	0.6	1.7	10	20
Stripping output	T0, T1	T0, T1, T2D	0.2	0.9	1.7	7.5
Monte Carlo	All sites	T0, T1, T2D	0.6	2	5	10
TOTAL			**1.5**		**17**	

- Not to be overlooked: ~4GB/s aggregated tape reading for (E)YETS re-processing

DOMA QoS: LHCb input

- Not enough LHCb efforts available to substantially contribute

- Feedback given at recent DOMA QoS workshop
  - Favour site uniformity within the same tier level, e.g.
    - do not bound experiment workflow to a site hardware tender
  - Be pragmatic e.g.
    - Limited number of storage classes, manually manageable
    - Same classes within a tier level

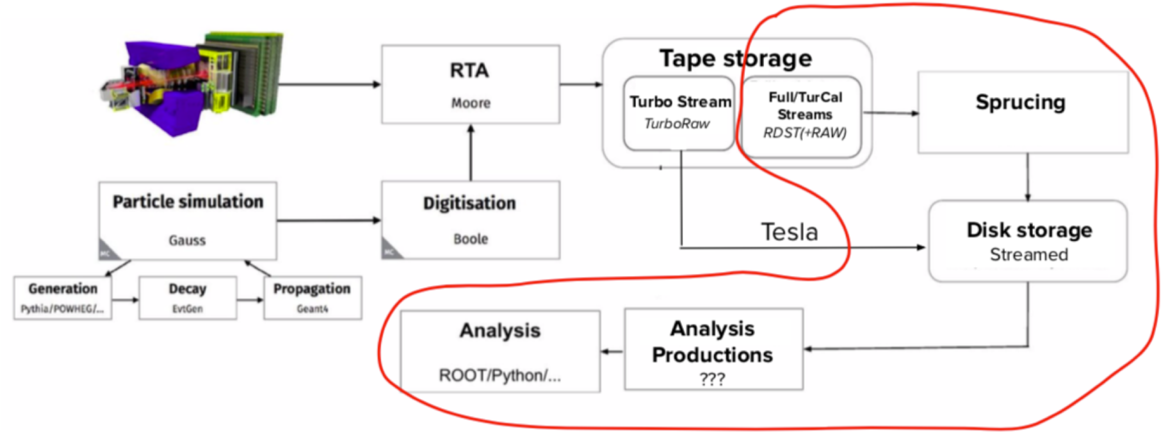


- Very good initiative!
  - Opening the door to a lot of innovations, improvements and cost savings
  - Concrete and realistic deliverables

Offline Analysis

- Increase of offline data volume creates challenges in offline data handling and processing for physics analysis
- Offline Analysis Task Force created with a mandate to identify scope, deliverables and organisation of a potential *offline analysis project for Run 3*
- Recommendations provided in document, circulated to the collaboration and the CB
- Now: developing proposals agreed by the collaboration, with institutional commitments to build and maintain the offline analysis infrastructure
- Operation phase: start of Run3 onwards

Offline Analysis



OATF has identified five areas that could permit work packages for offline analysis

WP1: Sprucing - centralised offline data selection/streaming for data that cannot go (initially) to TURBO stream.
WP2: Analysis productions - an upgrade to the WG productions of Run 2.
WP3: Offline Analysis Tools - review/development of existing tools (DV/TupleTools etc.), offline production framework to provide data suitable for various analysis tools (calibration/momentum scaling etc.), analysis preservation...
WP4: Heterogeneous computing resources - co-ordinate the use of GPUs and multi-CPU-cores for analysis-level applications within the collaboration (eg. amplitude fits in TensorFlow)
WP5: Run 1 and Run 2 legacy stripping

Offline Analysis: how to run the jobs

0. **Car** 🚙: User jobs
   - Most flexible but you have to drive yourself and you might not be a good driver
   - Makes bad use of human and compute resources
1. **Taxi** 🚕: Individual production requests per output file
   - Very flexible but there needs to be enough qualified drivers available (operations team)
   - Smaller input streams of data are preferable
2. **Train** 🚂: Run processes sequentially in a single DIRAC job
   - Very flexible but less efficient use of computing resources
3. **Bus** 🚌: Create all "DecayTreeTuples" in the same application
   - Very efficient with resources but getting there might be less pleasant

Vehicle choice will be transparent to analysts submitting productions

Decision path towards HLT1 technology

- More details in Renaud LeGac's talk at the LHCb-LHCC referee meeting this afternoon
- Management setup an internal review committee to prepare a decision on whether or not to use GPUs for HLT1. Review will be based on:
  - «TDR-like» document with technical details, integration with online and offline systems, physics performance, project organization, schedule, maintainance and operations
  - Comparison document with the x86 baseline solution
  - Discussion on costs and benefits
  - Feedback from Online and Computing projects
- Review committee provides advise, remarks and recommendations, and advises the Technical Board on decision
- *After the revision by and upon green light from the Review Committee, the documentation will be circulated to the collaboration a first time, will undergo a second revision incorporating feedback and remarks from the collaboration and undergo a second circulation before the final scrutiny and recommendations by the TB and the final ratification by the CB.*
- TDR-like document ready, comparison document expected by the end of February, cost model for the two implementations in preparation
- A set of recommendations is expected from Review Committee by the end of March

Executive Summary

Smooth and efficient usage of computing resources™

- ~2x above WLCG pledged resources
  - O(10%) from opportunistic resources
- Working on expanding opportunistic pool
- CPU work dominated by MC simulation
- Fast simulations taking over full simulation
  - As a consequence, ~4x more events produced in 2019 with only a modest increase of CPU work
- Completing legacy stripping of Run1+Run2 datasets

Preparing for Run3

- Offline computing requests for 2021 being sent to the C-RSG
- Network requirements shown to LHCOPN/LHCONE
- Defining requirements for storage Quality of Service
- Working on detector description and access to conditions data
- Offline Analysis is the next big topic to attack
- Decision path defined towards HLT1 technology choice