EPJ Web of Conferences **214**, 03045 (2019) https://doi.org/10.1051/epjconf/201921403045
CHEP 2018

Towards a computing model for the LHCb Upgrade

Concezio Bozzi[1,2,*] and *Stefan* Roiser[1,**] on behalf of the LHCb collaboration

[1]European Organization for Nuclear Research (CERN), Geneva, Switzerland
[2]INFN Sezione di Ferrara, Ferrara, Italy

Abstract. The LHCb experiment will be upgraded for data taking in the LHC Run 3. The foreseen trigger output bandwidth trigger of a few GB/s will result in data sets of tens of PB per year, which need to be efficiently streamed and stored offline for low-latency data analysis. In addition, simulation samples of up to two orders of magnitude larger than those currently simulated are envisaged, with big impact on offline computing and storage resources. This contribution discusses the offline computing model and the required offline resources for the LHCb upgrade, as resulting from the above requirements.

1 Introduction

The LHCb experiment will undergo a major upgrade of the detector [1] and its software and computing infrastructure [2–4] for the upcoming Run 3 of the LHC and beyond. New paradigms such as the calibration and alignment of the detector during the online selection [5], and the Turbo stream [6] have been successfully tested during Run 2 and will be further exploited during Run 3. The output bandwidth of the trigger will increase during Run 3 and therefore put additional strain on handling the offline resources in the distributed computing environment of the experiment.

This paper presents a preview of the computing model for the LHCb upgrade. The evolution of the data processing model from Run 2 to Run 3 is described in Section 2. The logical workflows to be implemented for Run 3 and the handling of the resource needs are described in Section 3. A summary is provided in Section 4. A more detailed discussion of these aspects, including the offline computing resource requirements for Run 3 and beyond, can be found in Ref. [3].

2 Evolution of the data processing

The online and offline data processing model of LHCb has evolved during Run 2 with the introduction of new concepts.

- The high-level software trigger has been split into the two stages, HLT1 and HLT2 [7], with the buffering of data between those two. The HLT1 runs synchronously during LHC collisions and brings the event rate down from 1 MHz to 100 kHz. The HLT2 processes data performing a full reconstruction of the events, and reduces the trigger rate to around

*e-mail: concezio.bozzi@cern.ch
**e-mail: stefan.roiser@cern.ch

10 kHz. If enough computing resources are available, the HLT2 is run also during LHC fill periods. Otherwise, data are processed during inter-fill and other periods without beam interactions, e.g. technical stops.

- A disk buffer space has been introduced for the online processing. This space gathers events after the HLT1 processing.
- The final calibration and alignment of the detector is performed from the data gathered on the disk buffer after HLT1 processing [5]. No second calibration and alignment of data has been done during Run 2 data processing on offline resources. As a consequence, the reconstructed data after HLT2 processing achieves the same quality as the offline reconstructed ones. This also removes the need of re-reconstructing data offline.
- The final reconstruction is already available in the HLT. Therefore, the final reconstructed physics objects can be directly streamed out of the HLT and the corresponding RAW data information needed for the reconstruction can be fully or partially removed. This is called the *Turbo* stream. For this stream, full re-reconstruction is not possible anymore after the online processing. Another feature of the Turbo stream is *selective persistency*, *i.e.* the possibility of saving information related to signal processes only, or a combination of signal and other information of interest, instead of the full event. Two more streams, the *Full* and *Calibration* streams, are exported after HLT processing and reconstructed on offline resources. The Full stream contains events for analysis work while the Calibration stream is used to verify the quality of the online and offline processed data, to isolate data control samples that are used to correct simulation, and to study systematic uncertainties in physics measurements.

The lowest-level hardware trigger (L0) that performs a pre-selection before HLT1 will be removed in Run 3. Therefore, the full 30 MHz inelastic collision rate is passed directly into the high-level software trigger. As a consequence, the trigger efficiency for most of the physics channels will increase by at least a factor 2. Also, the instantaneous luminosity will increase by a factor 5 to $2 * 10^{34}\ cm^{-2}\ s^{-1}$. Given the high signal purity of the trigger selections, the signal rate will scale accordingly. Moreover, due to the increase of event multiplicity, the raw event size of the Full and Calibration streams will increase by a factor 3. On the contrary, the event size in the Turbo stream, currently one-half of that of the Full stream, is expected not to change, due to the selective persistency mechanism mentioned above. In Run 2, roughly 30% of the trigger lines are sent to the Turbo stream. If this stays the same, then all the above changes in the data processing environment result in an increase of the trigger output bandwidth from Run 2 to Run 3 of a factor 25, i.e. from 0.7 GB/s to 17.5 GB/s. In order to keep the trigger output bandwidth under control, the majority of events in the upgrade will need to be processed through the Turbo stream. However, part of the LHCb beauty, electroweak and high p_T physics programmes will require inclusive trigger selections and the flexibility to develop new analysis ideas, that are best accommodated by saving data in the Full stream. It is found that the optimal point that does not sacrifice any of the physics program of LHCb results in about 70% of the event rate saved in the Turbo stream and 30% in the Full and Calibration streams. This corresponds to an HLT output bandwidth of 10 GB/s.

The increase in output bandwidth and also the number of collected events during the upgrade has several consequences on the offline resource needs. The requirements both for disk and custodial storage will increase, as the storage needs are dominated by the data from the detector. The data processing on offline resources will be reduced, as the majority of events are fully reconstructed during online processing. On the other hand, the CPU needs for Monte Carlo simulation will increase, as the number of simulated events scales with the collected data during Run 3.

3 The Run 3 data processing model

The interaction with distributed computing resources will continue to be handled by the LHCbDIRAC framework [8] which is an extension of the DIRAC grid interface [9] that is already being used by several experiments. This framework allows the efficient handling of workflow and data management and provides a single point of interaction to all kinds of distributed computing and storage infrastructure.

As discussed in the previous section three streams, possibly divided into sub-streams, will be exported after online processing (see Fig. 1). As in Run 2, the Turbo stream will be converted offline from the LHCb-internal MDF format into ROOT [10] format and luminosity information will be added. This processing will consume less than 1% of the available CPU resources. In addition the Turbo data will be kept in two copies on custodial storage. The full and calibration stream will also be kept in two copies on custodial storage. In addition a *stripping* of these two streams will be performed when written to disk. The stripping may perform either a reduction of the event content, leading to an event size reduction up to the size of the Turbo event, or event selection, or both. This processing reduces by a factor 3 the throughput to disk compared to tape for those two streams. A re-stripping of this data is foreseen at the end of a data taking year, and also during longer shutdown periods.

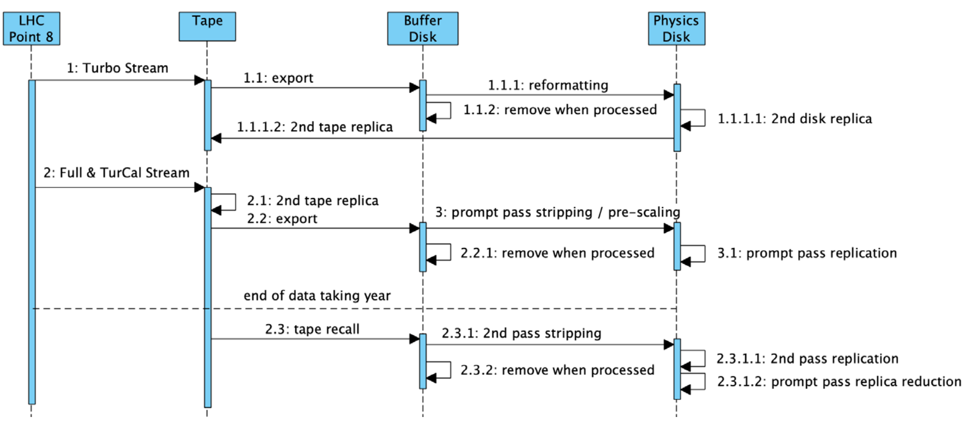

Figure 1. Logical workflow of the offline data processing in Run 3

As also discussed in the previous section, the amount of simulated events will increase in Run 3. The Monte Carlo simulation is split into several steps (see Fig. 2): event generation and simulation of the detector response as the first step, the remaining data processing, i.e. digitization, reconstruction and stripping as the second step, and a merging of the output data as the last step. The first step of this workflow usually consumes around 95% of the total CPU time. The first step of the simulation work flow will start without the need for additional input data such as underlying events. This implies the full simulation of the increased event multiplicity, but it allows to deploy the simulation work flow without any additional data management operations for the input data handling, and consequently deploy it easily at all kinds of compute platforms such as opportunistic resources, high performance compute centers or cloud resources.

The storage needs for Monte Carlo simulation will be mitigated by moving the majority of the output in a micro-DST format, which is a factor 20 smaller than the currently used DST format. The increased CPU needs will be reduced by increasing the throughput via various fast Monte Carlo simulation techniques:

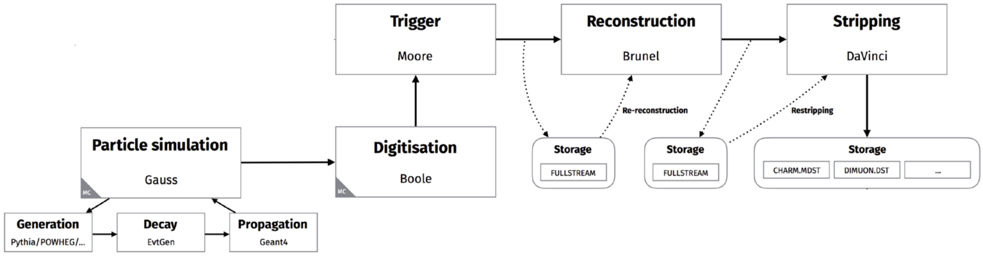

Figure 2. Logical workflow of Monte Carlo simulation in Run 3

- Re-decay [11], simulating the background information once and injecting the signal simulation multiple times
- Switching off various sub-detectors such as the Ring Imaging Cherenkov detector (RICH)
- Simulating only the tracking sub-detectors
- Libraries to provide immediate responses for individual sub-detectors such as the calorimeters.

The various fast simulation techniques provide improvements in the event throughput between 30% and a factor 20. In addition to fast simulation also a fully parametric simulation based on Delphes [12] is being prepared which increases the throughput further. It will be possible to combine various fast simulation models allowing to profit from an aggregated speedup provided by the various techniques. As in real data processing, also the simulation framework will be moved to the new multi-threaded Gaudi framework [2] and also profit from improvements from vectorisation and memory management on modern CPU architectures.

The majority of user analysis on distributed computing will be moved from individual jobs submitted by the users to centrally-organized working group productions [2]. These productions are handled by the LHCb computing operations team and via the LHCbDIRAC grid interface, allowing for a better and more efficient handling of those data selections. In addition a train model where several analysis selections are combined into a single workflow is being evaluated. Analysts will nevertheless have the possibility of using already existing tools such as Ganga [13] to submit individual productions for e.g. validation and prototyping.

4 Summary

With a major upgrade of the LHCb detector towards LHC Run 3 and beyond, also the data processing and consequently the computing model of the experiment undergoes major changes. In this paper we briefly presented these changes and how the computing infrastructure will cope with them. The major drivers for increased resource needs are explained, as well as the measures to keep them under control. Reference [3] contains more detailed discussions of all the above topics, as well as the associated offline computing resource requirements.

References

[1] LHCb collaboration, *Framework TDR for the LHCb Upgrade: Technical Design Report*. LHCb-TDR-012

[2] The LHCb Collaboration, *Upgrade Software and Computing Technical Design Report*. CERN-LHCC-2018-007. LHCB-TDR-017

[3] The LHCb Collaboration, *Computing Model of the Upgrade LHCb experiment Technical Design Report*. CERN-LHCC-2018-014, LHCb-TDR-018

[4] The LHCb Collaboration, *LHCb Trigger and Online Technical Design Report*. CERN-LHCC-2014-016, LHCb-TDR-016

[5] G. Dujany, B. Storaci, J. Phys. Conf. Ser. **664**, 082010 (2015)

[6] R. Aaij et al., Comput. Phys. Commun. **208**, 35 (2016)

[7] M. Frank, C. Gaspar, E. v.Herwijnen, B. Jost, N. Neufeld, J. Phys. Conf. Ser. **513**, 012006 (2014)

[8] F. Stagni, P. Charpentier, R. Graciani, A. Tsaregorodtsev, J. Closier, Z. Mathe, M. Ubeda, A. Zhelezov, E. Lanciotti, V. Romanovskiy et al., J. Phys. Conf. Ser. **396**, 032104 (2012)

[9] F. Stagni, A. Tsaregorodtsev, L. Arrabito, A. Sailer, T. Hara, X. Zhang (DIRAC consortium), J. Phys. Conf. Ser. **898**, 092020 (2017)

[10] R. Brun, F. Rademakers, Nucl. Instrum. Meth. **A389**, 81 (1997)

[11] D. Müller, *Measurements of production cross-sections and mixing of charm mesons at LHCb*, Ph.D. thesis (2017), CERN-THESIS-2017-257

[12] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, M. Selvaggi (DELPHES 3 Collaboration), JHEP **02**, 057 (2014)

[13] R. Currie, U. Egede, A. Richards, M. Slater, M. Williams, J. Phys. Conf. Ser. **898**, 052032 (2017)