Introduction

Federico Meloni (DESY), Matthew Wing (UCL)

LUXE computing kick-off
31/08/2022

Miscellanea

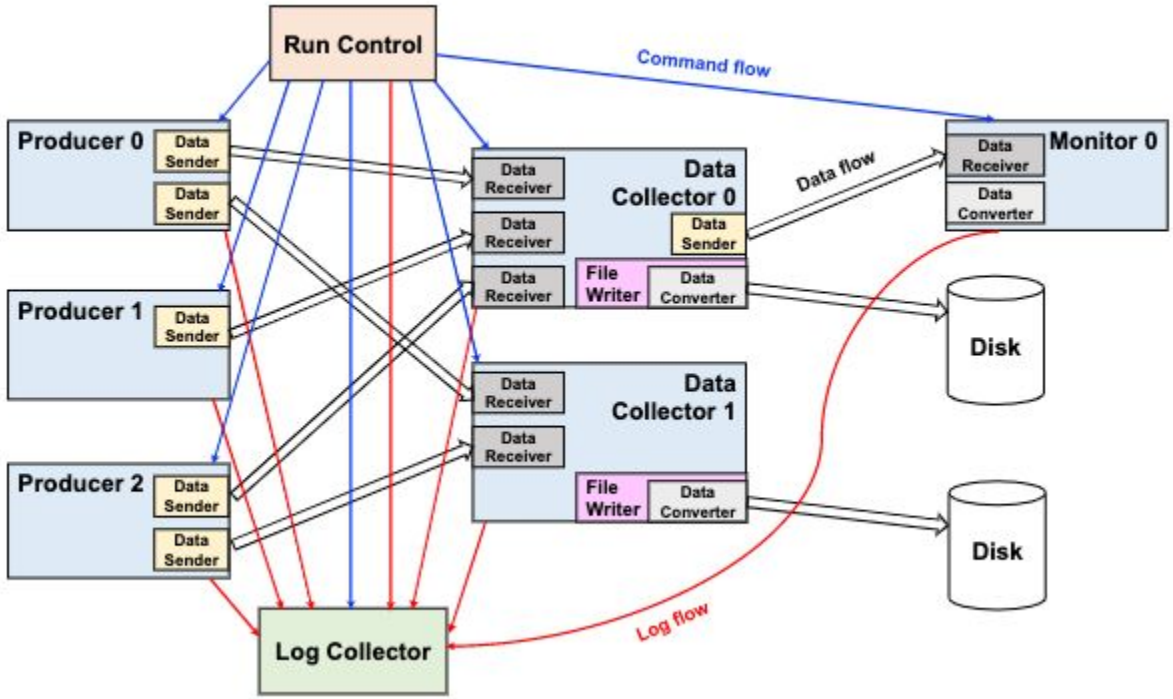
In case you haven't subscribed yet, please sign-up to our dedicated mailing list: luxe-daq-computing@desy.de

The meetings will live under: https://indico.desy.de/category/913/
Plan to continue alternating DAQ and computing-focused meetings.

- Next computing-focused meeting 28/09
- We might swap slots as needed

Do we need anything else?

Beyond the DAQ

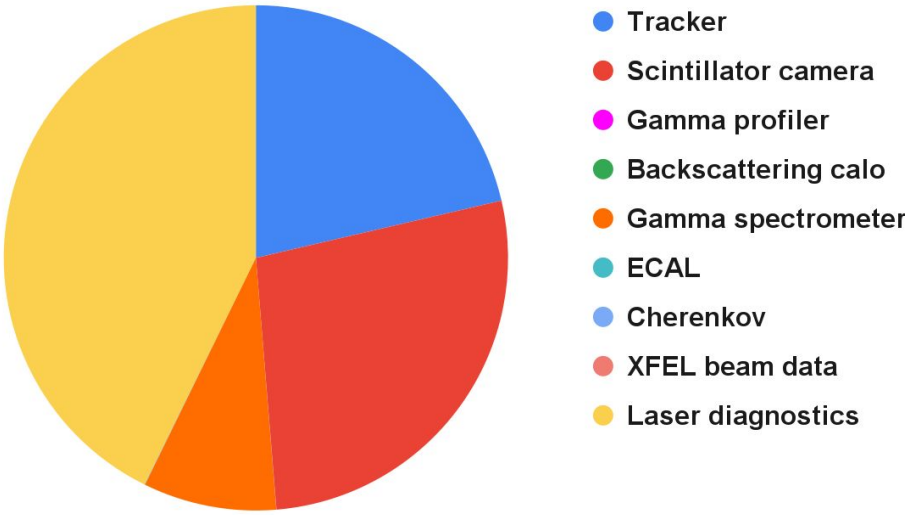


In these meeting(s) we will discuss:

- Organisation of computing infrastructure (disk/cpu)
- Related software infrastructure
- Data EDM for RAW formats

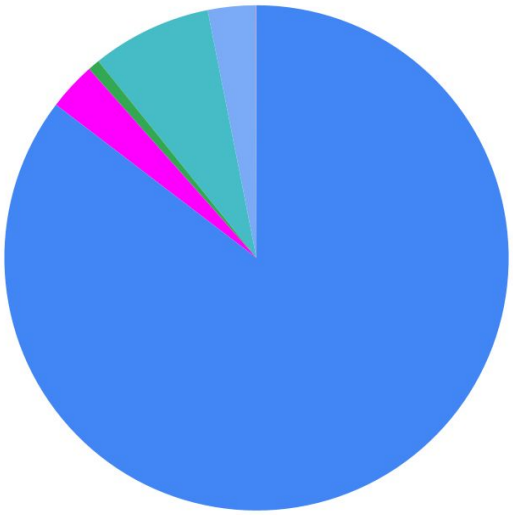
Raw event size

Signal event size (e-laser)



50 MB

Background event size (e-laser)



~1 MB

The largest contributions come from the laser diagnostics and camera images.

Every camera image is assumed to be stored in raw format.

- Need to investigate impact of lossless compression.

No zero suppression assumed for calorimeter and Cherenkov detectors (but will be there).

Storage resources

For robustness and accessibility, we'd like to use **grid resources for the storage of the data**, simulated data and reconstructed event samples.

- To mitigate the risk of data loss, we should keep at least one replica of everything.

Assuming 10^7 s of data-taking time per year, we can estimate the disk space needs.

- Keep 2 copies of reconstructed data (N and N-1 reprocessing versions)
- Assume 1:1 ratio between simulation and data

Assume to use only robust disk-based and tape grid resources.

Type of data	Data Size
Raw data (filled BX)	0.46 PB
Raw data (empty BX)	0.01 PB
Reconstructed data	0.40 PB
Simulated data (1:1 filled BX)	0.54 PB
Total	1.42 PB

CPU resources

The simulation of a signal bunch crossing can take **up to 7 10^6 s**.

- 4.5 particles per second with hadronic physics list (QGSP_BERT_HP)
- Up to 3.5 10^7 particles per signal event

Assume 1:1 sim events (not distributed uniformly) and max sim time → 2 10^{11} CPUh/year

- Likely grossly overestimated
- Need to base this estimate on planned data taking modes and ξ values

Digitisation not considered yet (but might be sizeable until we develop parameterised response functions).

A combined reconstruction software is not available yet, but it is expected that the average per-event processing time will be of the order of one minute.

- **9 10^5 CPU hours per year** would be required to reconstruct all the data from the detector (twice) and the simulated bunch crossings

Today and the next meetings

Goal: use this meeting to get in touch with each other and learn about needs/news from interested group.

We'll use the collected material to organise the upcoming meetings.

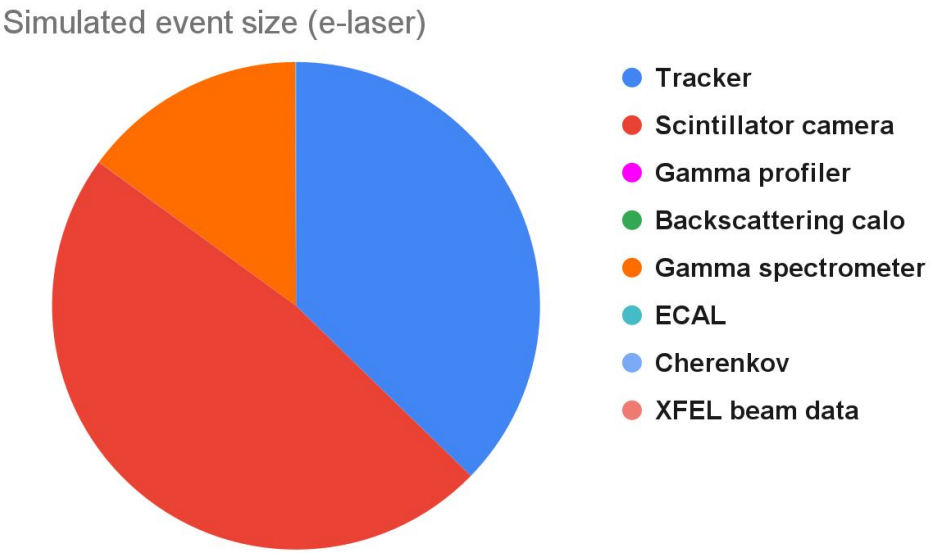
Some items already on the list:

- Deeper dive into EDM / connections with DAQ
- Tools for data management (e.g. Rucio@DESY)
- …

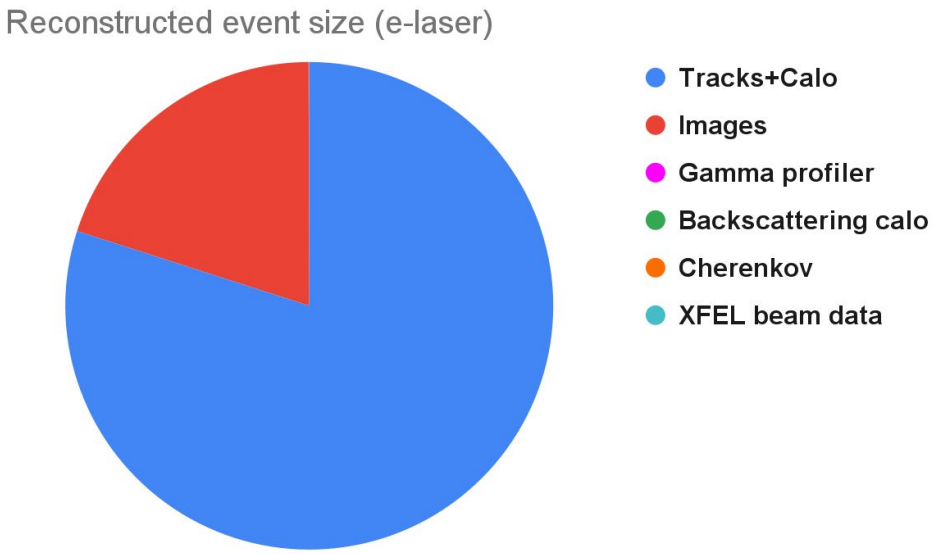
Please feel free to bring up any other topic you'd like to see discussed.

Thanks!

Simulated and reconstructed event size



30 MB



20 MB

Note: these numbers are more speculative and should be rather seen as “initial targets”.

The estimates assume:

- Simulation and reconstruction don’t include laser diagnostics
- Camera images and laser diagnostics are strongly reduced in size in reconstructed data (down to ~1 million floats)
- Truth record in simulated events can be kept under control