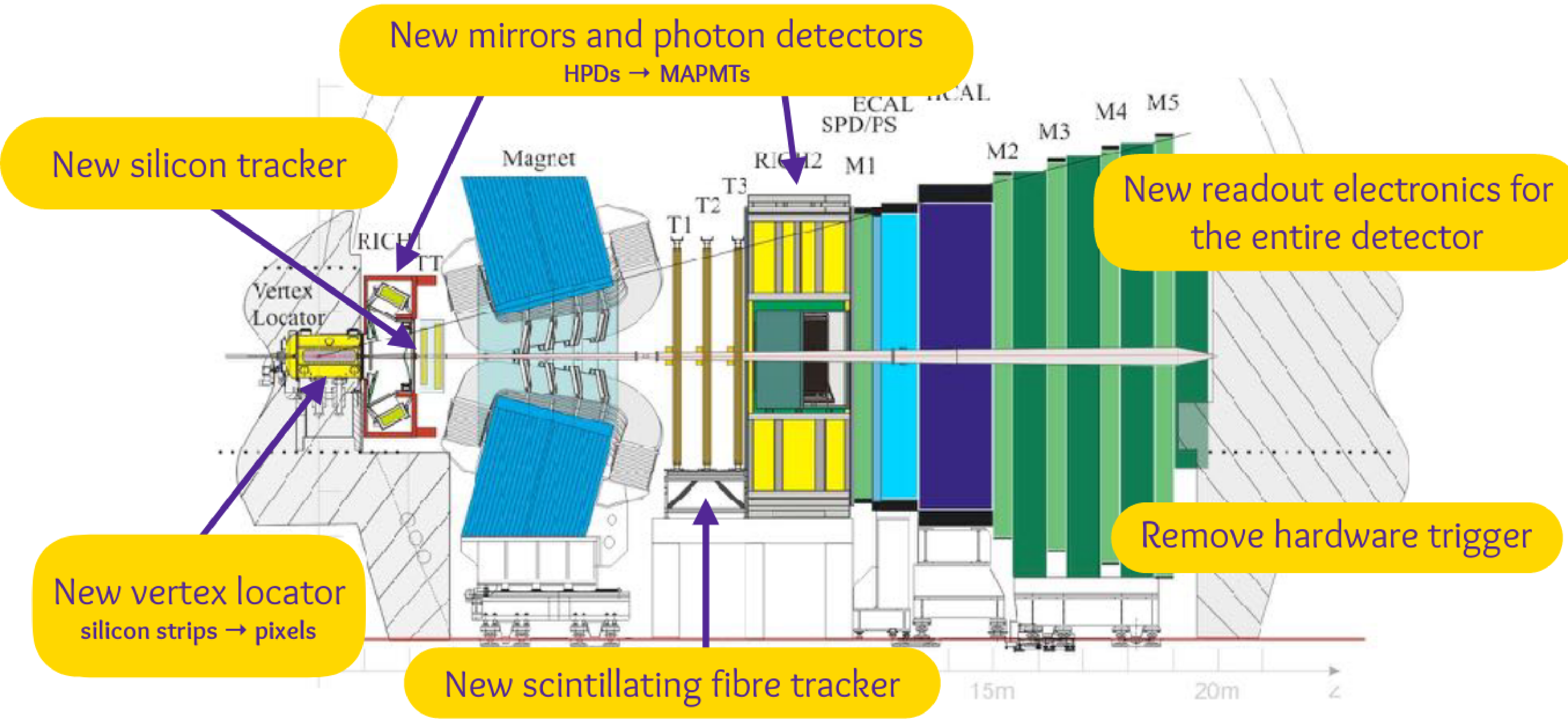
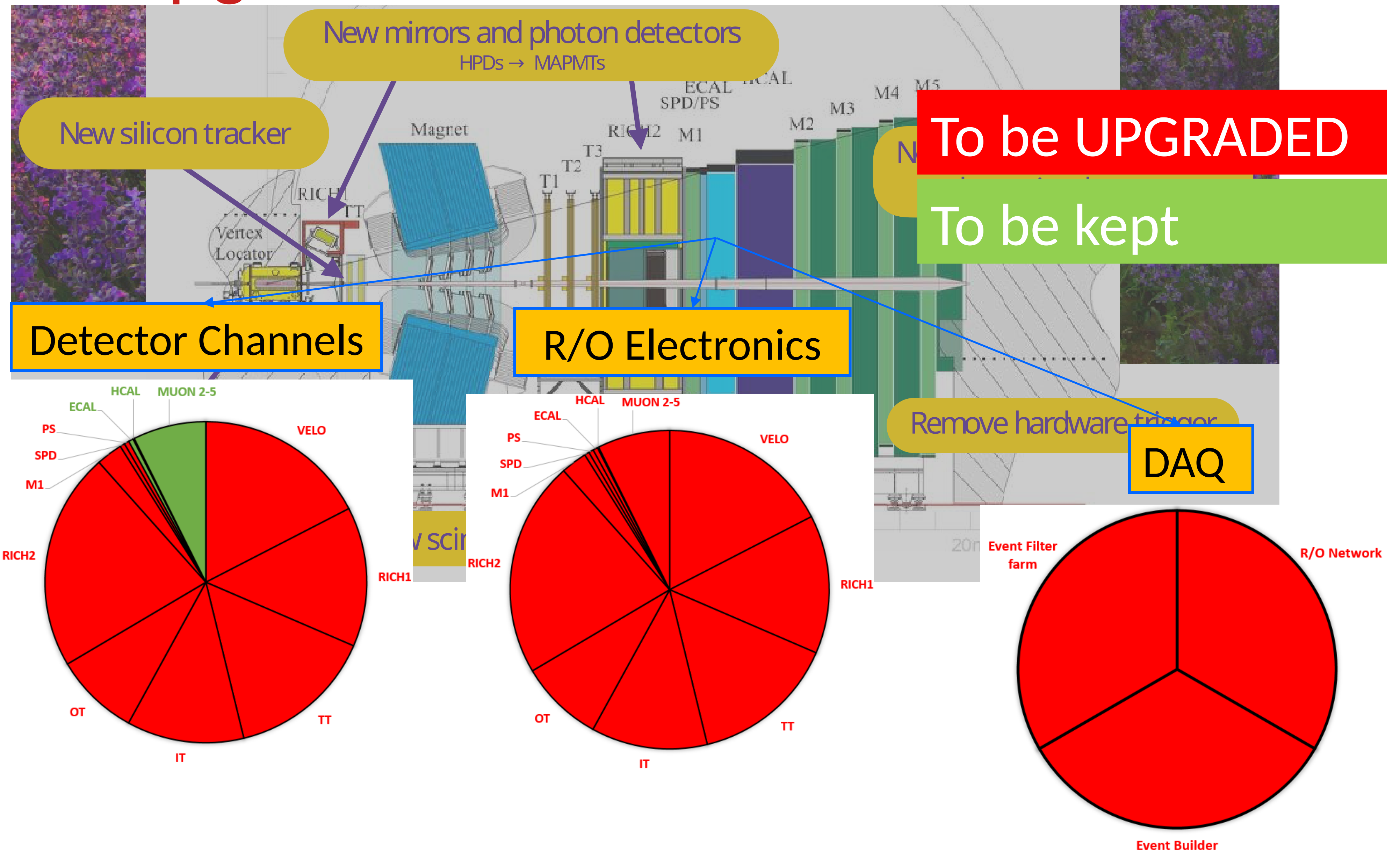
LHCb Run3 computing model

Christophe Haen
Concezio Bozzi
ISGC 2023

The upgraded LHCb detector for Run 3-4



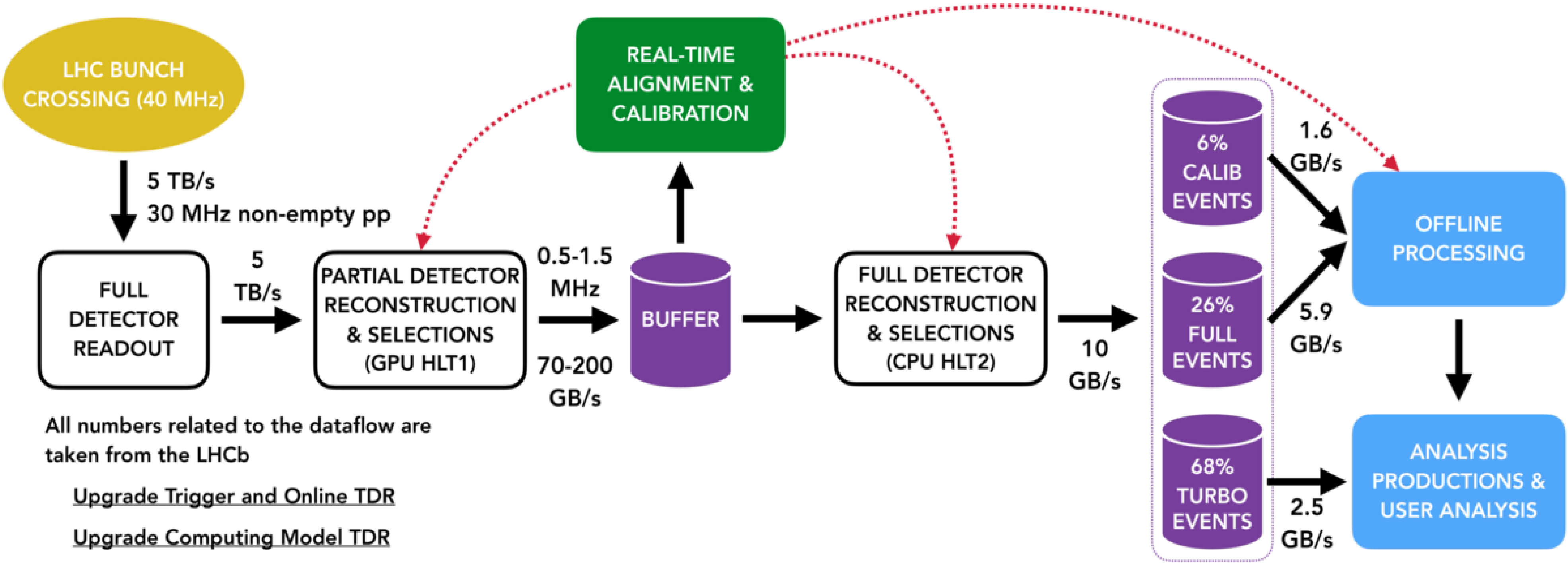
The upgraded LHCb detector for Run 3-4



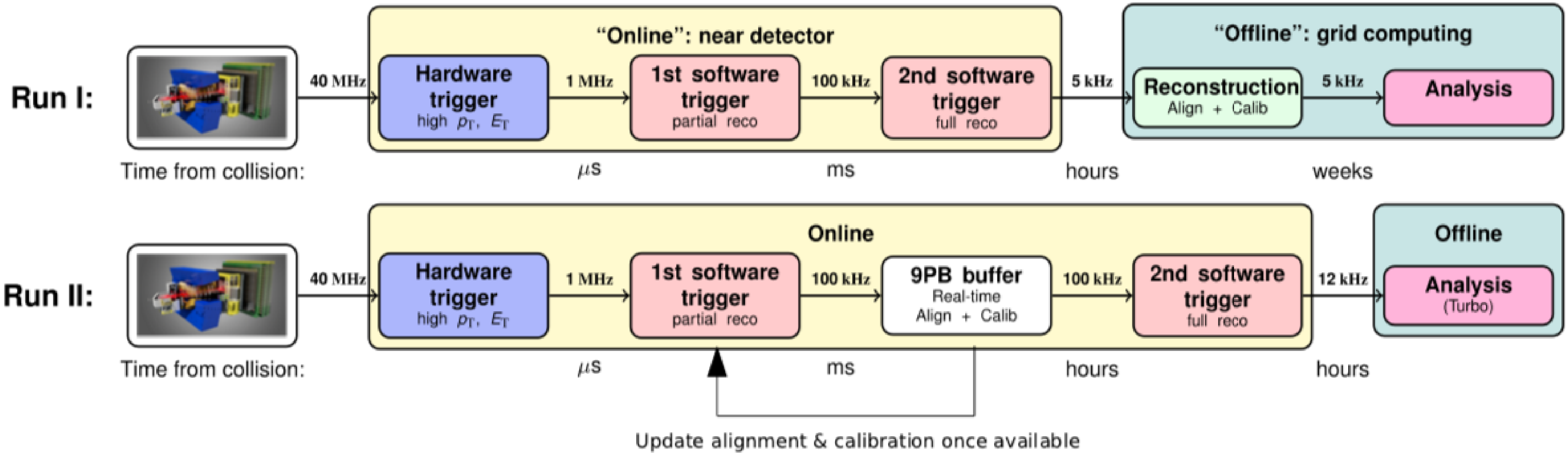
A big challenge in data handling

- Major expansion of LHCb physics programme through:
  - 5-fold increase in instantaneous luminosity
    - 4×10^{32} to 2×10^{33} $cm^{-2}s^{-1}$
  - Full software trigger at 30MHz inelastic collision rate
    - Factor 2 increase in trigger selection efficiency
- Order of magnitude increase in physics event rate to storage
- Pile-up increase
    - Factor 3 increase in average event size
- 30x increase in throughput from the upgraded detector
  - Without corresponding jump in offline computing resources

Outline


LHC BUNCH CROSSING (40 MHz)
5 TB/s
30 MHz non-empty pp
FULL DETECTOR READOUT
5 TB/s
PARTIAL DETECTOR RECONSTRUCTION & SELECTIONS (GPU HLT1)
0.5-1.5 MHz
70-200 GB/s
BUFFER
REAL-TIME ALIGNMENT & CALIBRATION
FULL DETECTOR RECONSTRUCTION & SELECTIONS (CPU HLT2)
10 GB/s
6% CALIB EVENTS
26% FULL EVENTS
68% TURBO EVENTS
1.6 GB/s
5.9 GB/s
2.5 GB/s
OFFLINE PROCESSING
ANALYSIS PRODUCTIONS & USER ANALYSIS
All numbers related to the dataflow are taken from the LHCb
Upgrade Trigger and Online TDR
Upgrade Computing Model TDR

Run1 + Run2 trigger

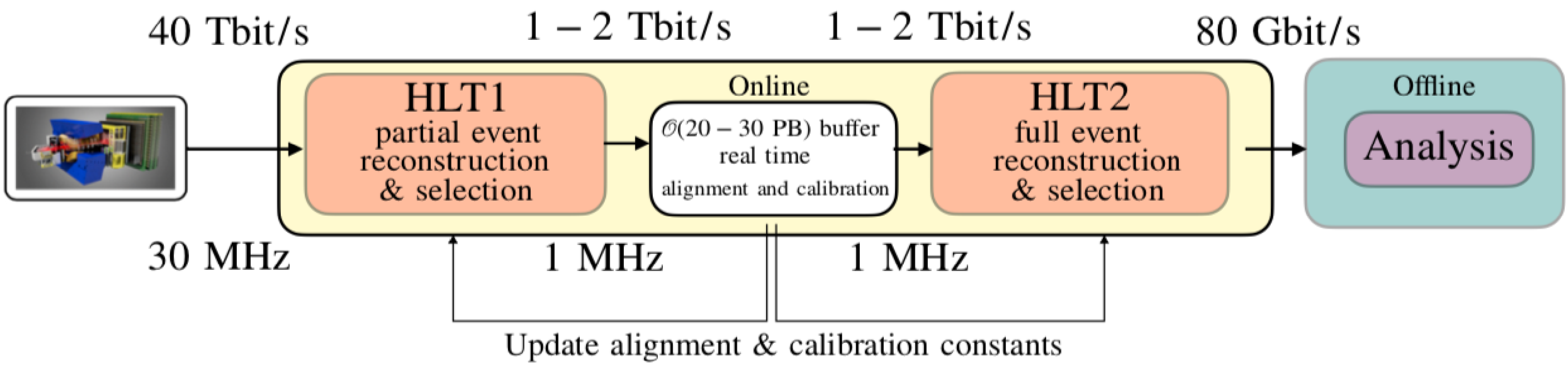


- Hardware trigger: based on muon detectors and calorimeters

Run 2

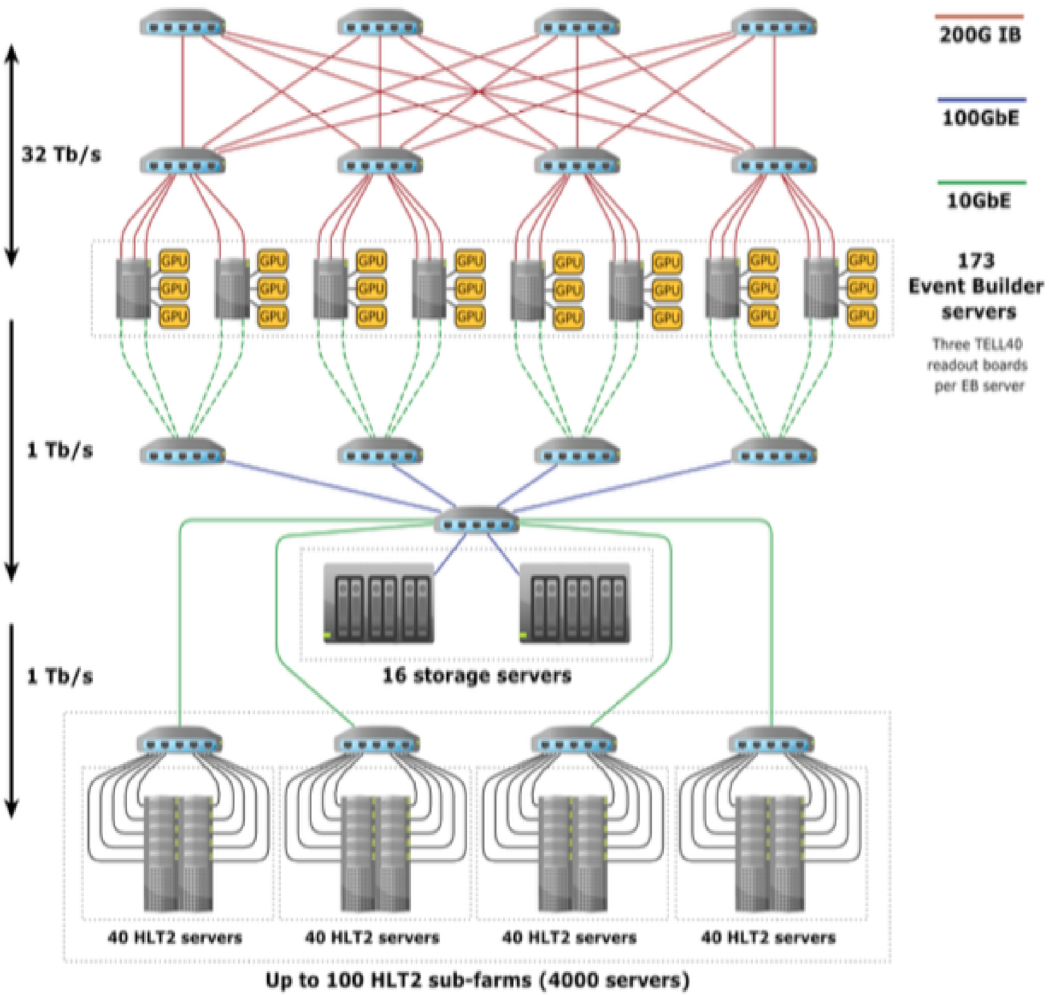
- Data buffered in between two software trigger stages
- Allows for real-time alignment and calibration Offline-quality reconstruction within the trigger

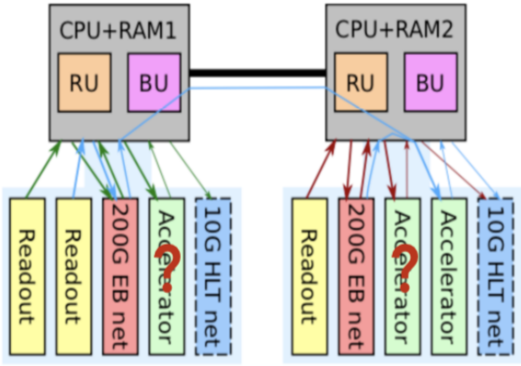
Run 3 trigger



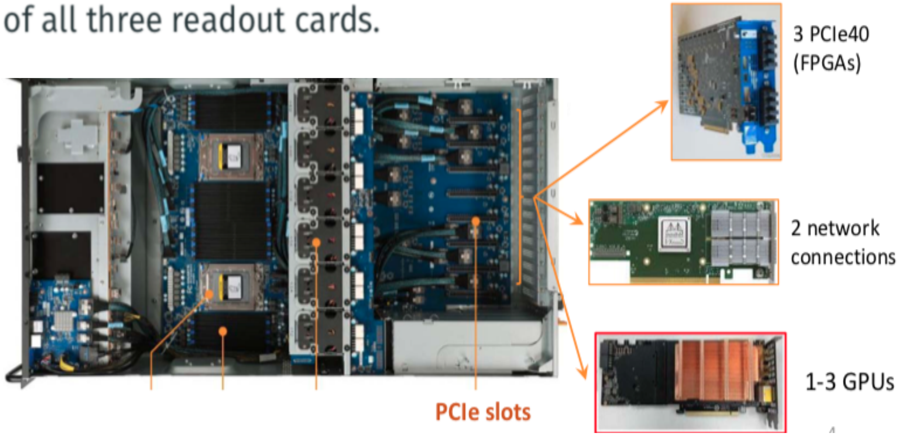
- Remove hardware trigger
- Two stage software trigger:
  - HLT1 on GPU
  - HLT2 on CPU
- Read-out at 40 MHz (bunch crossing rate)

Practical implementation





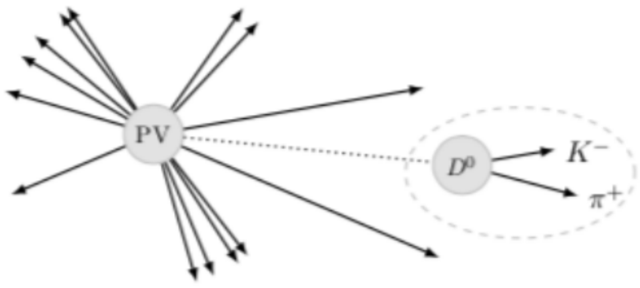
GPU-equipped event builder PC, with traffic of all three readout cards.

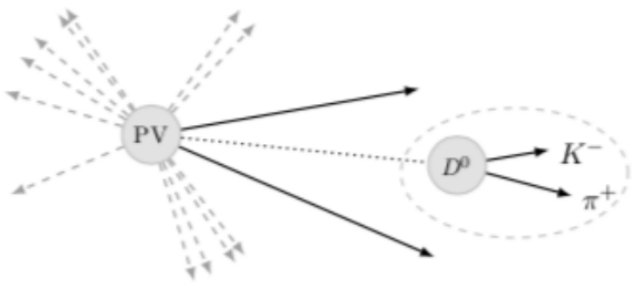


LHCb-TDR-018

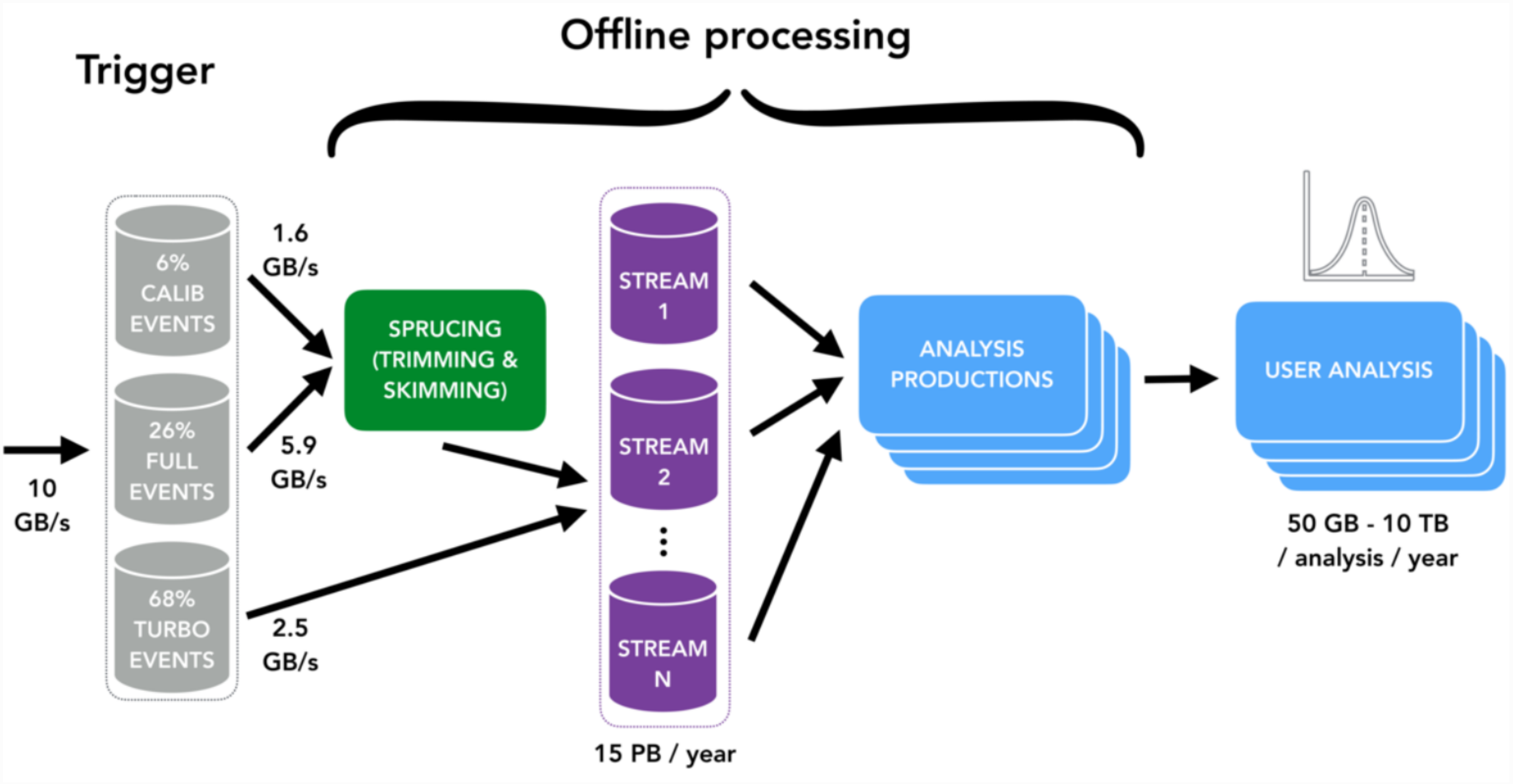
Data streams and dataflow

- **FULL (left):** «classic» stream, all reconstructed objects in the event. Needs central «slimming and skimming» for subsequent physics analysis
- **TURCAL:** calibration stream, with both reconstruction output and (some) RAW banks(performance studies.
- **TURBO (right):** Write only the interesting part of the event. Data ready to be analysed, no further processing needed





Offline processing



Real data distribution

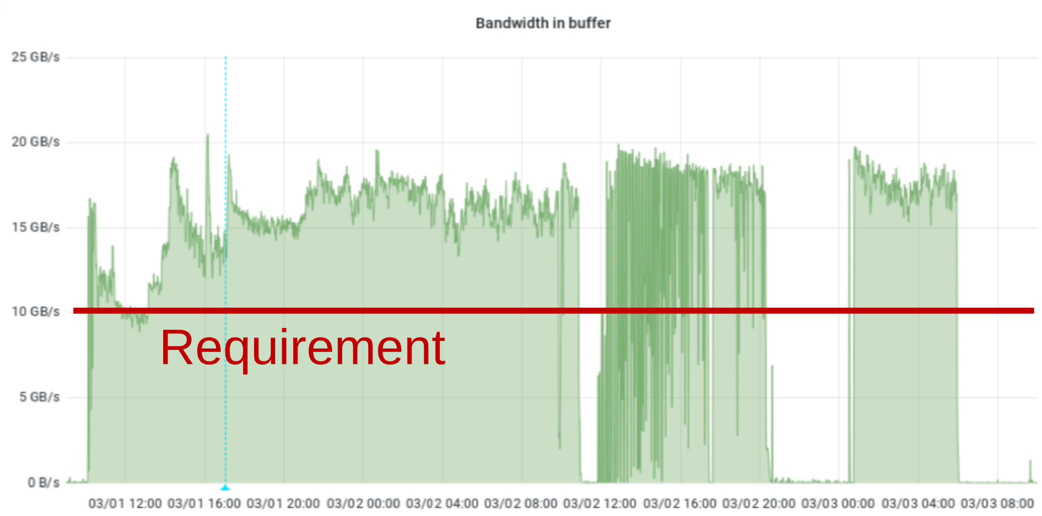
- Data distribution model quite simple
- Jobs run where data is
  - Mostly at Tier0 and Tier1s
- Number of sites with data relatively small
  - 1 T0, 7 T1s, 14 T2-Ds
- Well-balanced CPU and disk resources
- No need for caches, pre-placement, etc
- Little impact on WAN other than dataset replication (2 copies)

Write tests:
CERN disk → T1 disk → T1 tape

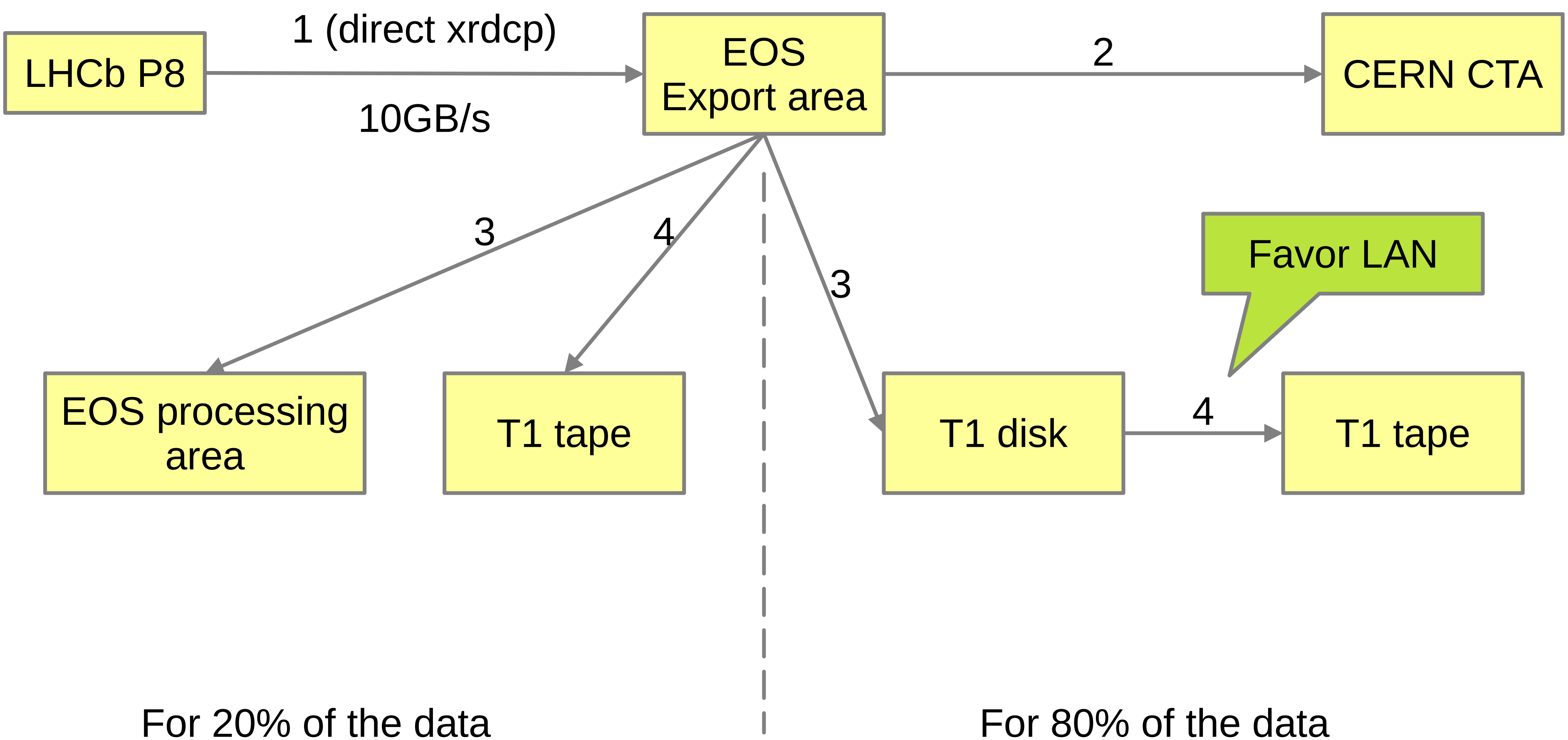
Site	expected Speed (GB/s)
CERN	11
CNAF	1.72
GRIDKA	2.23
IN2P3	1.25
NCBJ	1.32
PIC	0.2
RAL	2.96
RRCKI	0.25
SARA	1.07

Read tests
T1 tape → T1 disk

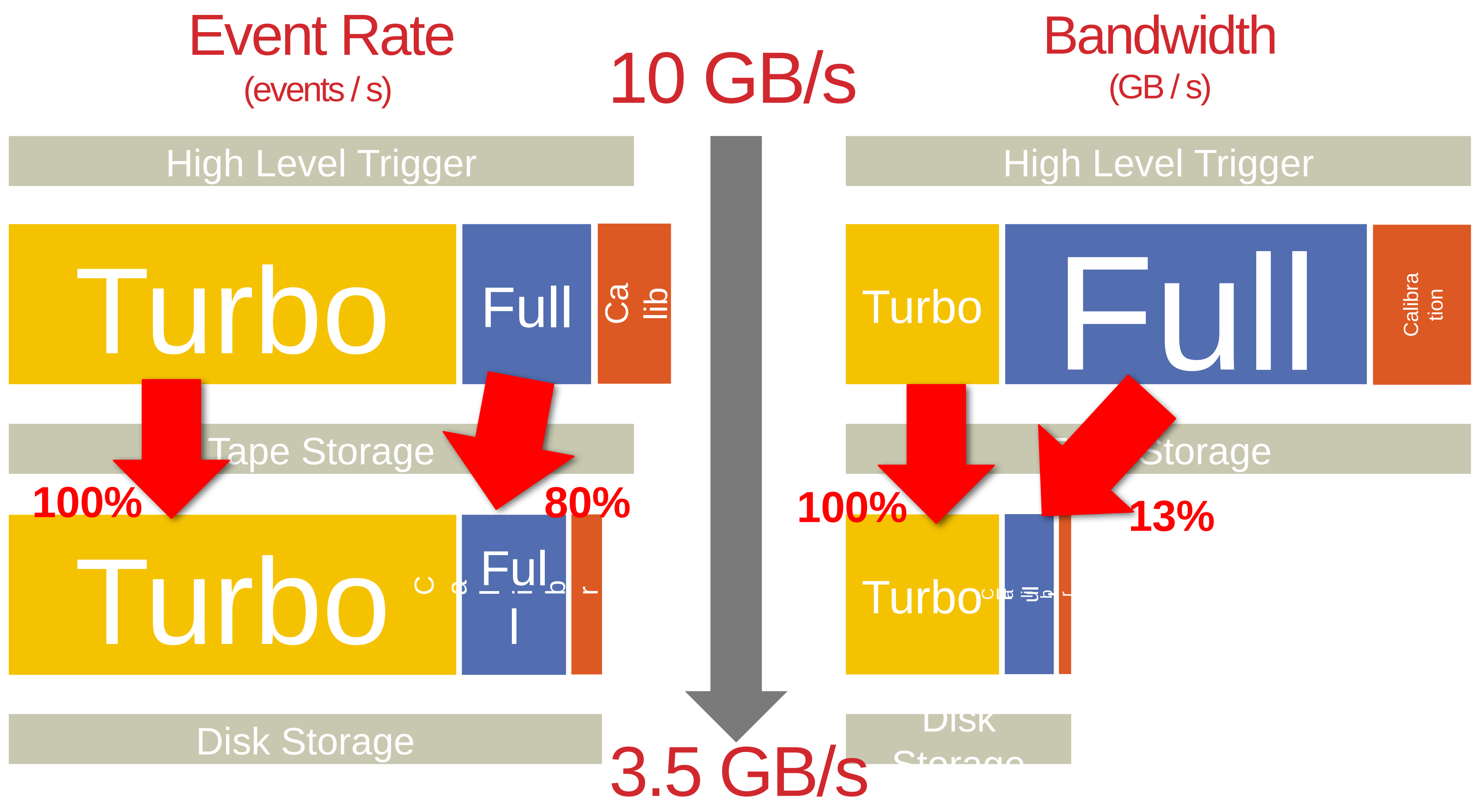
Site	expected Speed (GB/s)
CERN	1.90
CNAF	1.35
GRIDKA	1.36
IN2P3	0.98
NCBJ	0.91
PIC	0.17
RAL	1.93
RRCKI	0.21
SARA	0.74



Real Data distribution (most common workflow)

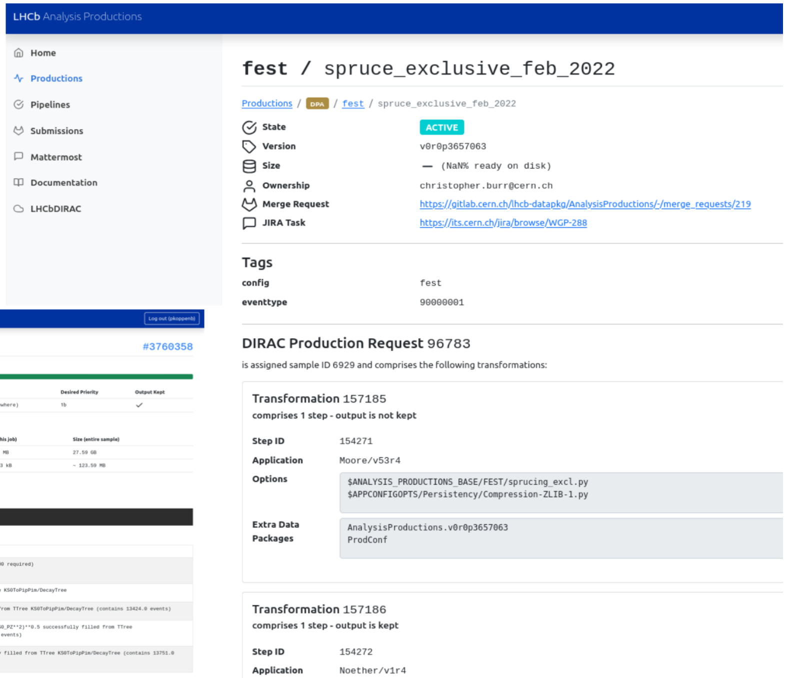


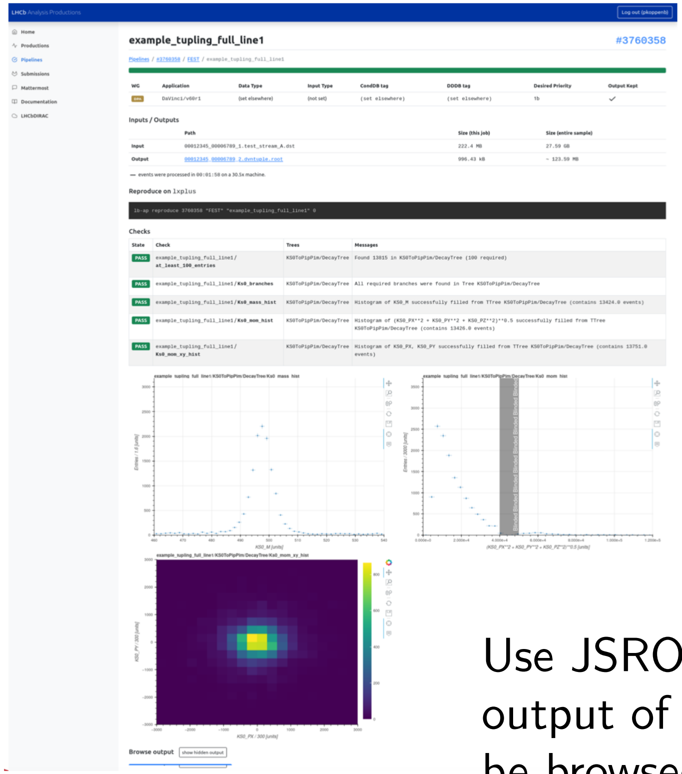
Sprucing


Event Rate
(events / s)
10 GB/s
Bandwidth
(GB / s)
High Level Trigger
Turbo
Full
Ca lib
Tape Storage
100%
80%
Turbo
Full
Disk Storage
High Level Trigger
Turbo
Full
Calibration
Storage
100%
13%
Turbo
Disk Storage
3.5 GB/s

Analysis productions

- Support user processing of data and simulation using the DIRAC transformation system
  - Based on gitlab: code review + tests in CI prior to submission
  - User do not need to monitor GRID jobs
  - Job details / configuration / logs automatically preserved in LHCb bookkeeping / EOS
  - Automated error interpretation / advice
  - Intuitive web interface for requesting / testing / browsing outputs

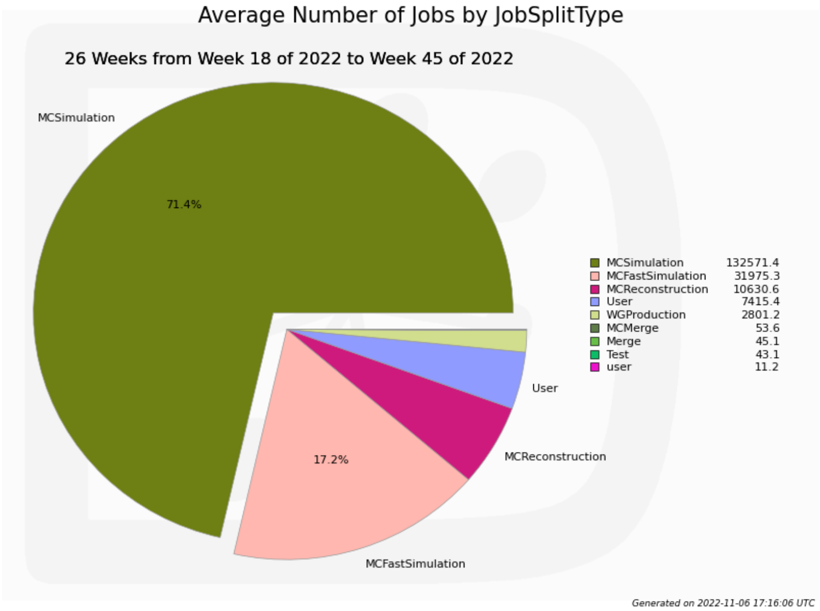




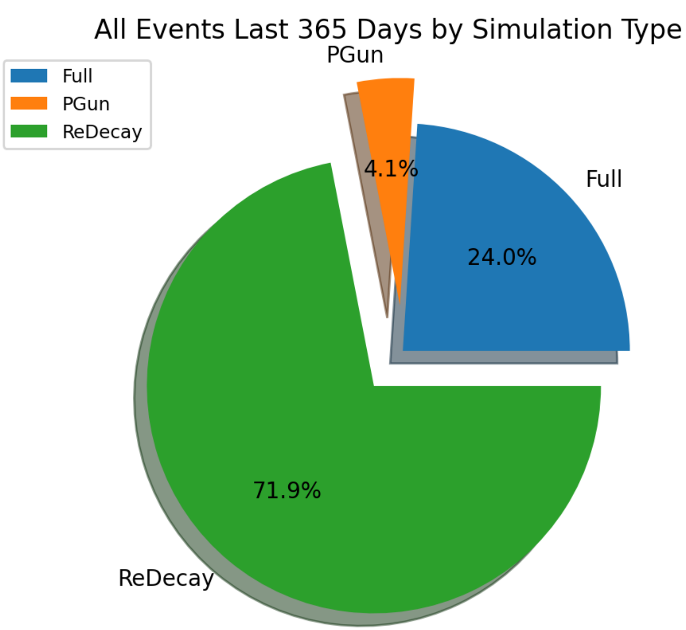
Use JSROOT for allowing the output of test productions to be browsed.

Monte-Carlo

- Computing work dominated by MC production (97%): Ideal candidate for CPU optimization
  - **Full –** full Geant4 detector simulation
  - **PGun –** single signal particle spawned with kinematics configured to follow distribution (no full pythia event) Factor 50 speed increase
  - **ReDecay –** re-use the underlying event but generate and simulate new signal decays every time Eur. Phys. J C78 (2018) 1009 Factor 10-20 speed increase
  - **TrackerOnly** simulation – Factor 10 speed increase
  - **SplitSim –** only simulate full event if required condition is passed e.g. if a photon converts to e+e- Speed up depends on condition
  - Investigating parametric simulation (Lamaar)
- Moving towards Analysis Productions style model with code review+CI prior to submission



Detailed



Summary

- 30x larger data volume from detector
- Full software trigger, aggressive triggering strategy, filtering and heavy use of Turbo stream (selective persistency)
- Offline sprucing reduces the size even further
- Data distribution optimized
  - Favour LAN over WAN
- Analysis productions
  - Bottom-up approach, collecting use cases towards a more structured activity
- CPU offline resources dominated by simulation production
  - Fast simulation significantly mitigates requirements