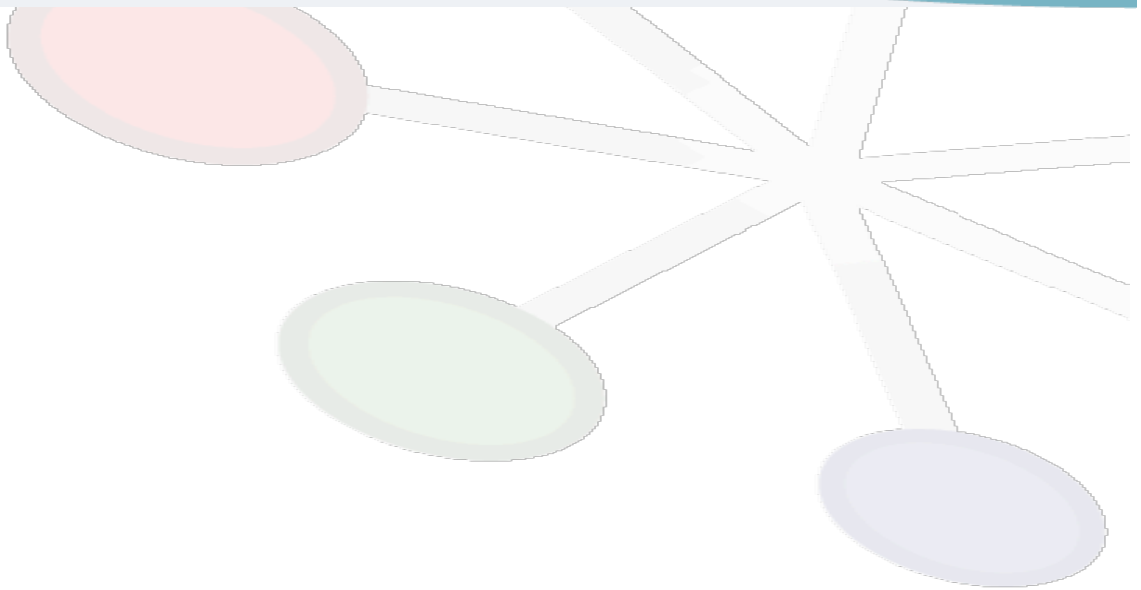
# LHCb Computing

## LHCb computing plans for Run 3

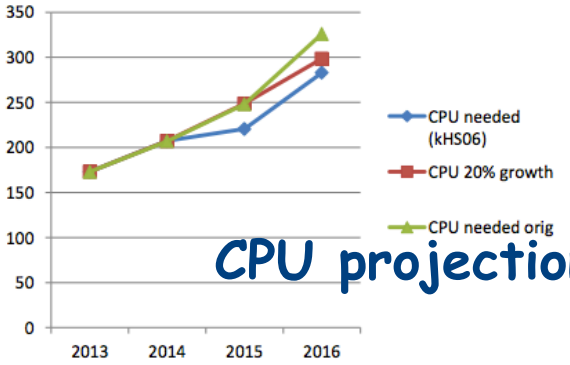


# Towards the LHCb Upgrade (Run 3, 2020)

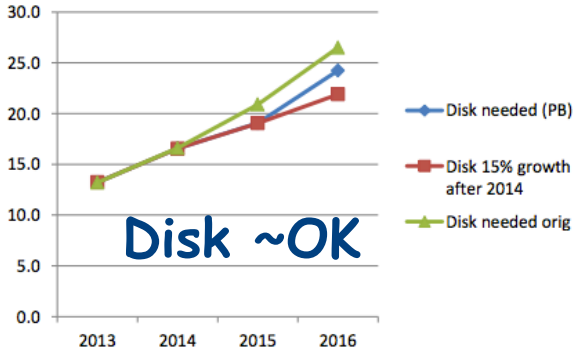
- We do not plan a revolution for LHCb Upgrade computing
- Rather an evolution to fit in the following boundary conditions:
  - Luminosity levelling at $2 \times 10^{33}$
    - Factor 5 c.f. Run 2
  - 100kHz HLT output rate for full physics programme
    - Factor 8-10 more than in Run 2
  - Flat funding for offline computing resources
- Computing milestones for the LHCb upgrade:
  - TDR: 2017Q1
  - Computing model: 2018Q3
- Therefore only brainstorming at this stage, to devise model that keeps within boundary conditions

# Comparison with "flat budget"

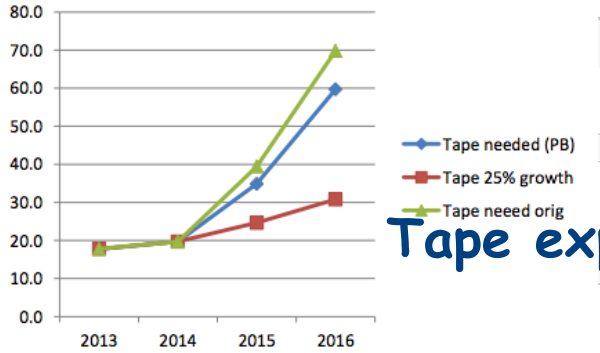
- Definition of flat budget: same money will buy
  - 20% more CPUs
  - 15% more disk
  - 25% more tape




CPU projections ~OK to >2016




Disk ~OK

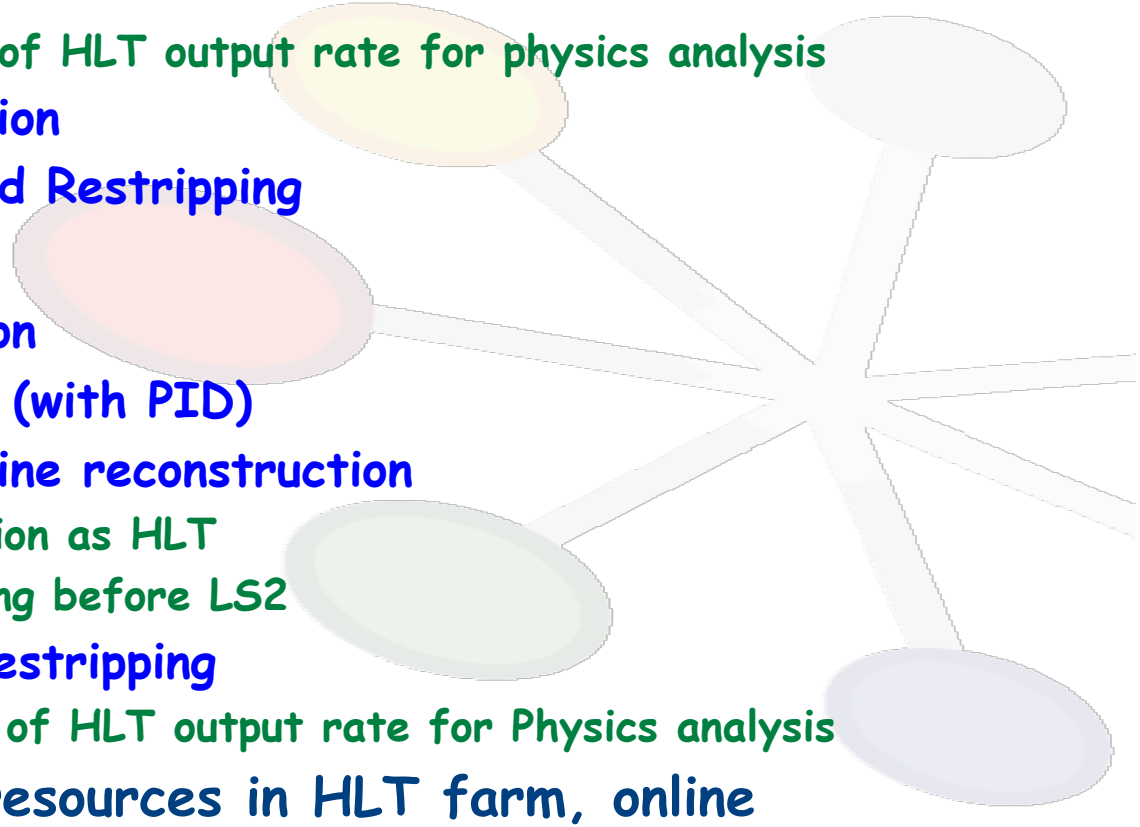



Tape explodes

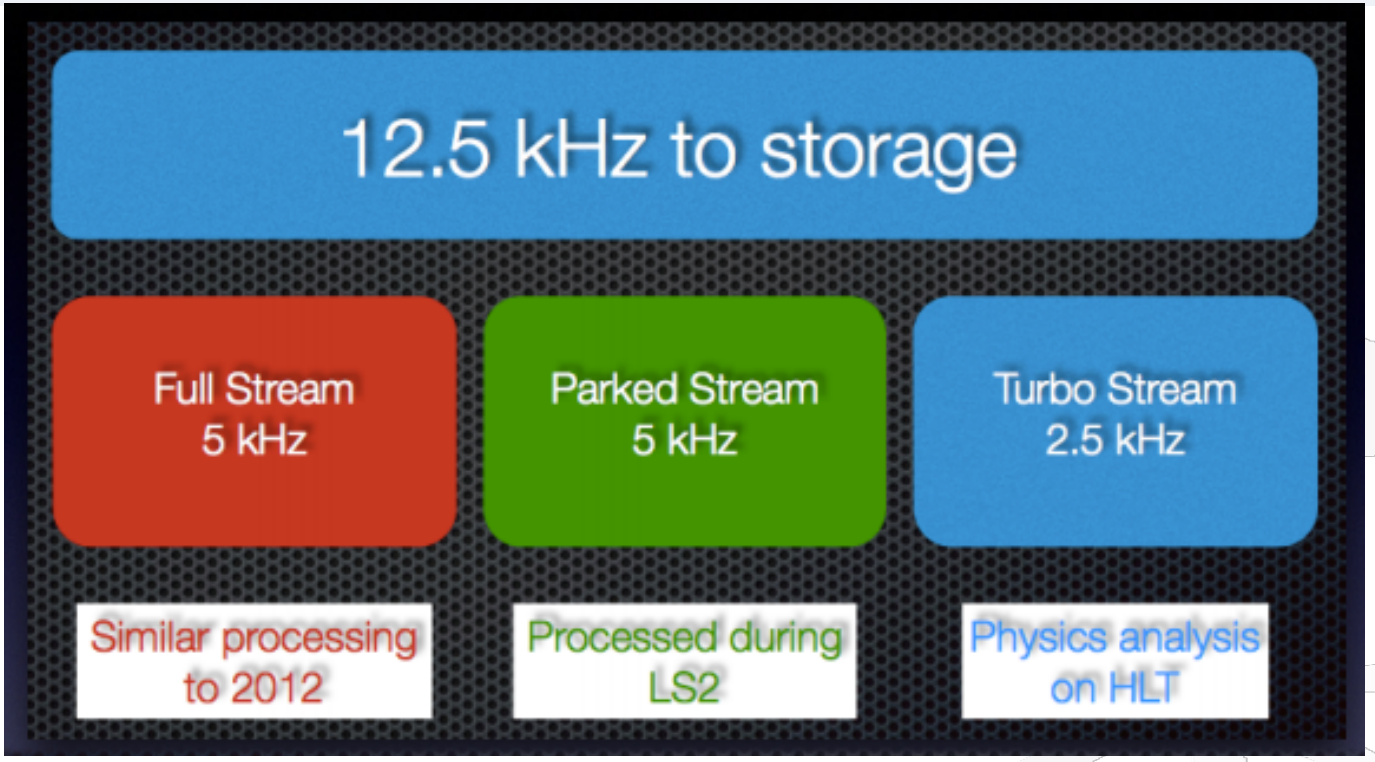
- Tape requirement driven by:
  - Two copies of RAW
    - Incompressible, but ~never accessed
  - One copy Reconstruction output (FULL.DST)
- Flat funding cannot accommodate order of magnitude more data expected in Run 3 - need new ideas

# Evolution of LHCb data processing model

- Run 1:
  - Loose selection in HLT (no PID)
  - First pass offline reconstruction
  - Stripping
    - selects ~50% of HLT output rate for physics analysis
  - Offline calibration
  - Reprocessing and Restripping
- Run 2:
  - Online calibration
  - Deferred HLT2 (with PID)
  - Single pass offline reconstruction
    - Same calibration as HLT
    - No reprocessing before LS2
  - Stripping and Restripping
    - Selects ~90% of HLT output rate for Physics analysis
- Given sufficient resources in HLT farm, online reconstruction could be made ~identical to offline



# Run 2: Reconstruction streams




- Full stream: prompt reconstruction as soon as RAW data appears offline
- Parked stream: safety valve, probably not needed until 2017
- Turbo stream: no offline reconstruction, analysis objects produced in HLT
  - Important test for Run3

# TurboDST: brainstorming for Run 3

- In Run 2, Online (HLT) reconstruction will be very similar to offline (same code, same calibration, fewer tracks)
  - If it can be made identical, why then write RAW data out of HLT, rather than Reconstruction output?

- In Run 2 LHCb will record 2.5 kHz of "TurboDST"
  - RAW data plus result of HLT reconstruction and HLT selection
  - Equivalent to a microDST (MDST) from the offline stripping
  - Proof of concept: can a complete physics analysis be done based on a MDST produced in the HLT?
    - i.e. no offline reconstruction
      - no offline realignment, reduced opportunity for PID recalibration
    - RAW data remains available as a safety net
  - If successful, can we drop the RAW data?
    - HLT writes out ONLY the MDST ???

- Currently just ideas, but would allow a 100kHz HLT output rate without an order of magnitude more computing resources.

# Simulation

- LHCb offline CPU usage is dominated by simulation
  - Already true in Run 2: simulation >60% of CPU needs in 2016
    - Many measurements start to be limited by simulation statistics
- Simulation suited for execution on heterogeneous resources
  - Pursue efforts to interface Dirac framework to multiple computing platforms
    - Allow opportunistic and scheduled use of new facilities
  - Extend use of HLT farm during LHC stops
- Several approaches to reduce CPU time per event
  - Code optimisation, vectorisation etc.
    - Contribute to and benefit from community wide activities, e.g. for faster transport
  - Fast simulations
    - Not appropriate for many detailed studies for LHCb precision measurements
    - Nevertheless many generator level studies are possible
  - Hybrid approach
    - Full simulation for signal candidates only
    - Fast techniques for the rest
      - e.g. skip calorimeter simulation for out of time pileup
- To avoid being limited by disk space
  - Deploy MDST format also for simulated data

# Conclusions

- LHCb event output rate will be an order of magnitude larger in Run 3 (2020)
- Currently brainstorming on ideas for reducing data rate without reducing physics reach
  - Run 2 as a test bed
- Computing efforts concentrated on
  - Code optimisation
  - Opportunistic use of diverse resources