Workflows and Data Management

**Workflow and DM**
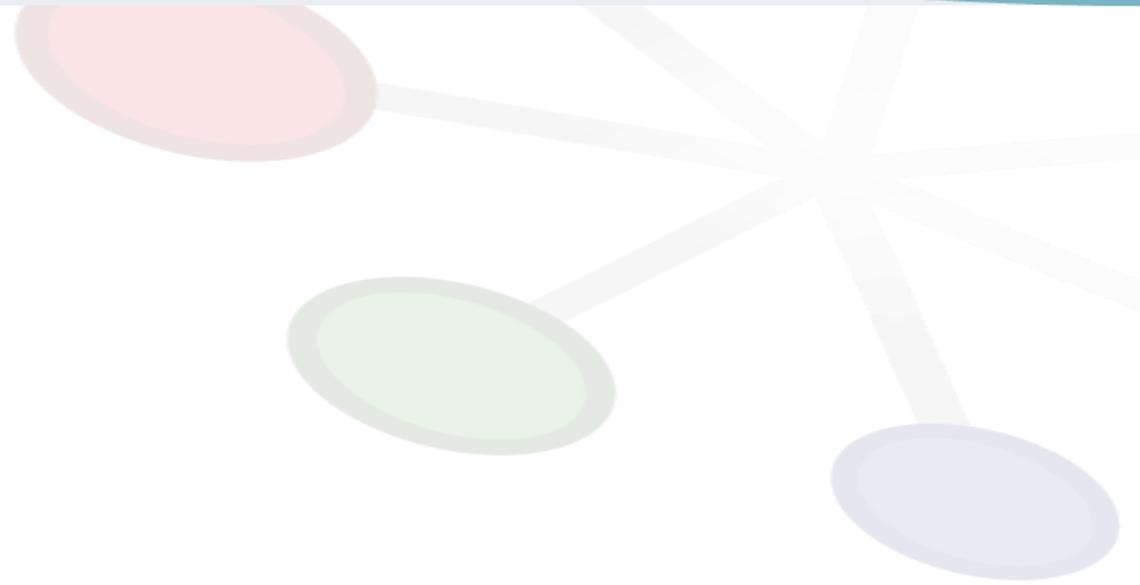
- ○ **LHCb major upgrade is for Run3 (2020 horizon)!**
  - □ **Luminosity x 5 (2 10^{33})**
    - ☞ **Lumi levelling: higher pileup, from 1.1 to 5.5**
  - □ **Trigger rate... x 5 (at least, dominated by charm physics)**
  - □ **RAW data size x 2 (pileup)**
  - □ **Online reconstruction = offline reconstruction**
    - ☞ **Allows direct analysis from online data (TURBO stream)**
      - ♎ TURBO data format is directly analysis data (no RAW!)
  - □ **Output from DAQ:**
    - ☞ **Any linear combination from TURBO data to full reconstruction output (reco + RAW)**
    - ☞ **Use year "n" data to tune TURBO for year "n+1" !**
  - □ **Throughput between 6 and 10 GB/s (GPDs of today)**
  - □ **Trigger (SW only) == offline selection**
    - ☞ **Stripping and streaming are no longer effective (all events are for physics!)**

Online Calibration & Alignment

Workflow and DM

○ **Novel concept of detector alignment & calibration done in between the two stages of HLT processing**

  □ **Successfully exercised in 2015**

  □ **Part of online reconstructed events immediately available for user and**

  □ **Enabling HLT2 proces for better signal yiel**

  □ **Same constants used offline processing**
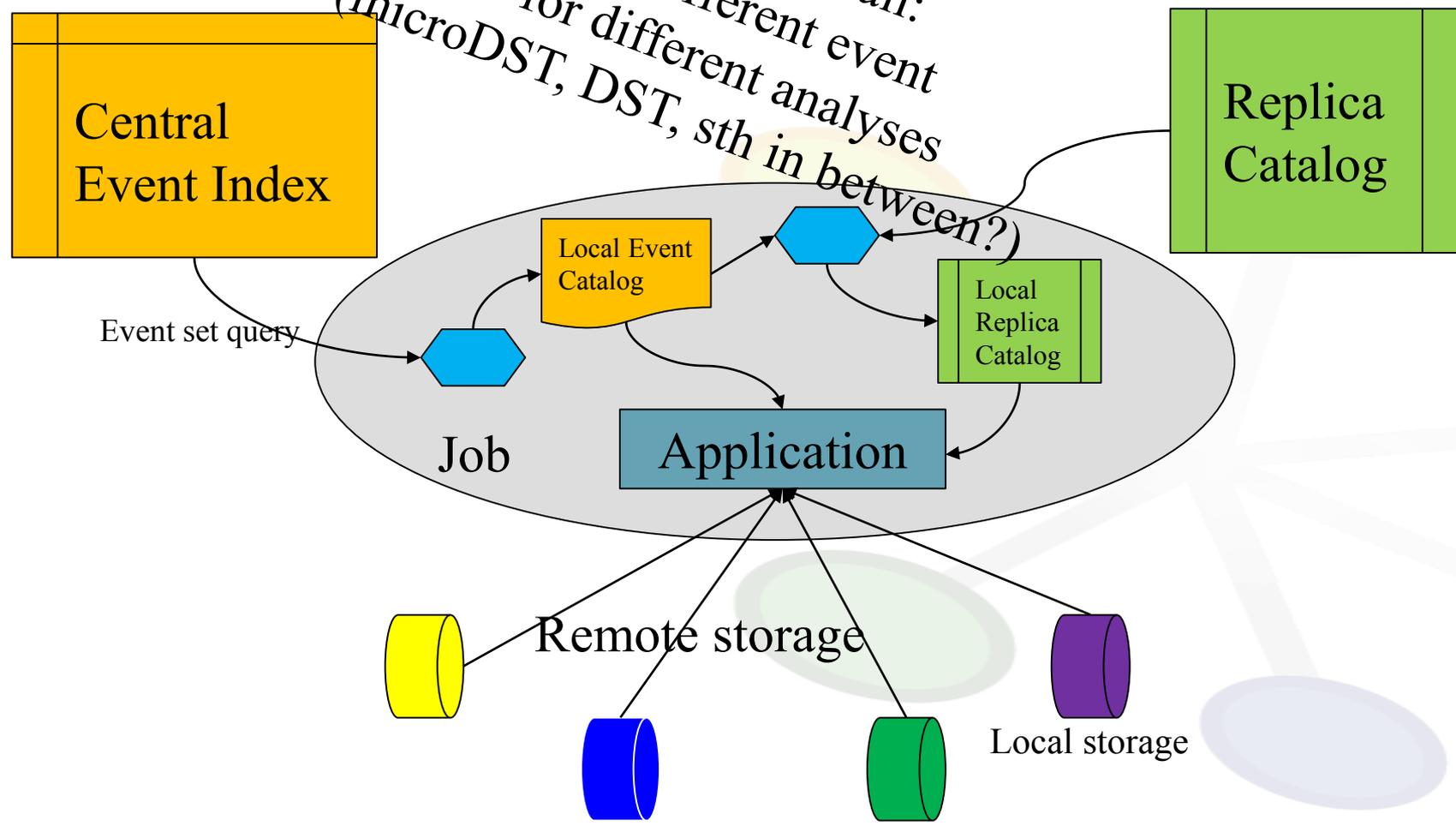
  □ **Concept will be furth exploited in Run 3**

Workflow and DM
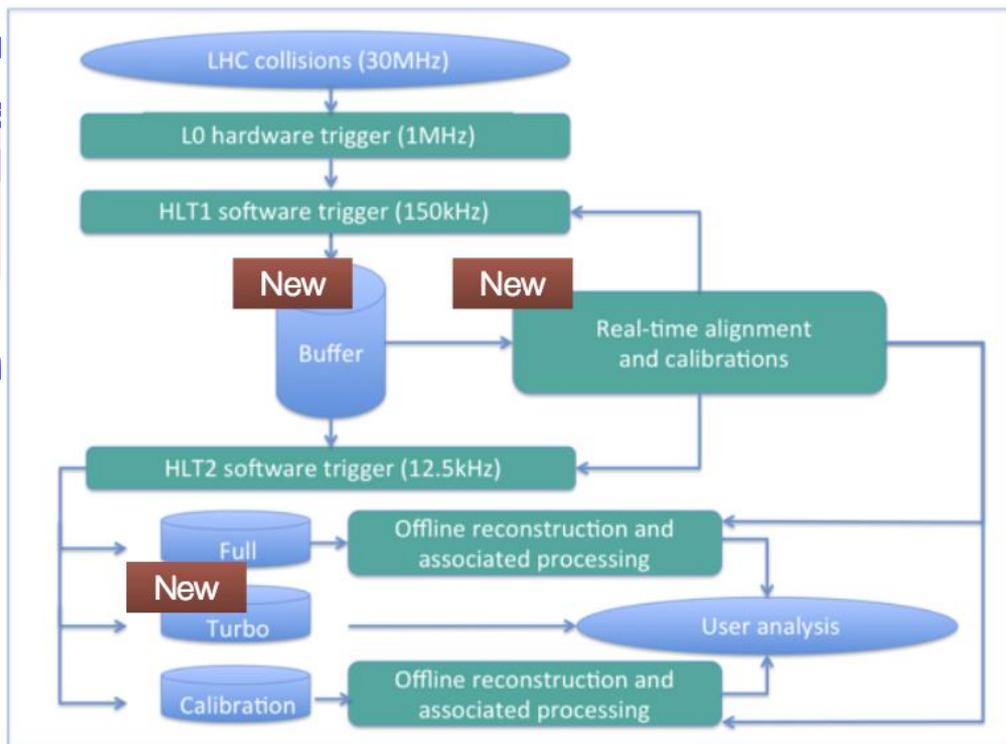
- ○ **Using event indices for analysis**
  - □ **Replace "stripping + streaming" with "selection + indexation"**
    - ☞ **Because stripping retention will be high (more selective trigger)**
  - □ **Event set query to central (or local) index**
    - ☞ **Download a local event collection (i.e. direct access addresses)**
  - □ **Random access to local or remote data**
    - ☞ **Using a local replica catalog (Gaudi Federation)**
- ○ **R&D can start now (2016/17) for:**
  - □ **Setting up train analyses**
    - ☞ **framework similar to stripping**
  - □ **Data indexing**
    - ☞ **Select technology (central vs distributed, DB vs files)**
    - ☞ **Index content to be defined**
    - ☞ **Event set queries to be defined for jobs**
  - □ **Optimizing random access through ROOT**
- ○ **Not to forget for analysis data access:**
  - □ **Network bandwidth is not all: disk spindles is equally important**

Workflow and DM

One size does NOT fit all: we will have different event formats for different analyses (microDST, DST, sth in between?)

Central Event Index

Replica Catalog

Local Event Catalog

Local Replica Catalog

Event set query

Job

Application

Remote storage

Local storage