Continuous and fast calibration of the CMS experiment

Design of the automated workflows and operational experience
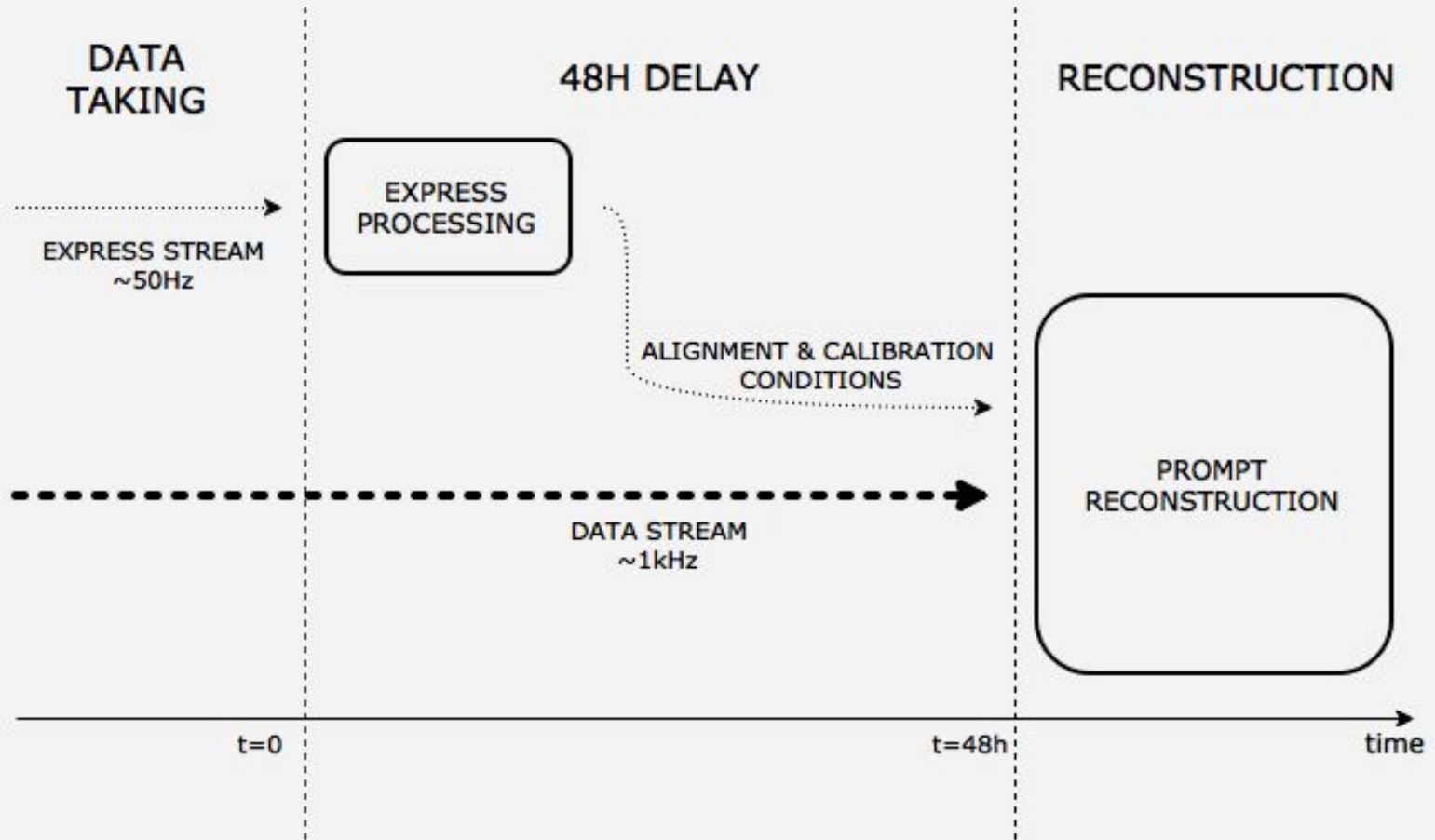
Piotr Oramus

CERN, AGH University of Science and Technology

on behalf of

CMS Collaboration

Prompt calibration concept and data processing

# Alignment & Calibrations run promptly

What we do:

- fit of the luminous region position and width (using tracks)
- identification of Silicon Strip tracker problematic channels
- determination of charge gains of the Silicon Strip tracker
- track-based alignment of Silicon Pixel inner tracker "large structures"

When done promptly:

- efficient online event selection by HLT (Higher Level Trigger)
- ready for analysis datasets within a few hours from data taking

Workflows running at CERN Tier0

- Tier0 computing farm runs automated calibration and alignment workflows
- Computing alignment and calibration conditions on run#N allows to apply them on the same run#N in prompt reconstruction
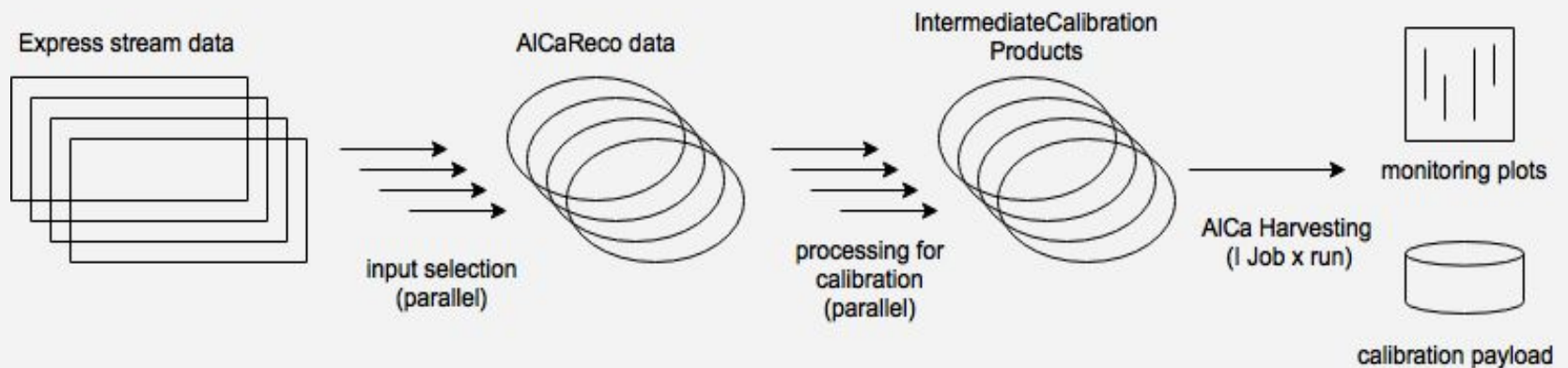
An approach to multi-run harvesting

- Calibration algorithms become more complex
  - More statistics needed
  - Single runs not sufficient

- Idea: merge runs to have more data
  - selected runs have to have the same properties (dataset, magnetic field, Tier0 computing environment etc.)

- Run AlCa Harvesting step over multi-runs
  - when still too narrow extend multi-run and repeat
  - perform following uploads

Multi-runs assembling

- Discover new data when it is available through Run Registry (an external service)
- Get further information about the data from DBS (Dataset Bookkeeping System) and Tier0 API
- For runs with the same properties create exactly one multi-run
  - If cumulative number of events exceeds some threshold then the multi-run is ready to be processed
- Complications
  - Processing does not have to produce the alignment and calibration constants - multi-run has to be extended then
  - Runs data differ in length - do not always arrive in the same order as produced by detector

The State Machine

- Multi-run processing flow control

- Supports implementing resilience by:
  - Retrying processing in case of failures
  - Easy monitoring of the process
 - Running independent multi-run computations in parallel
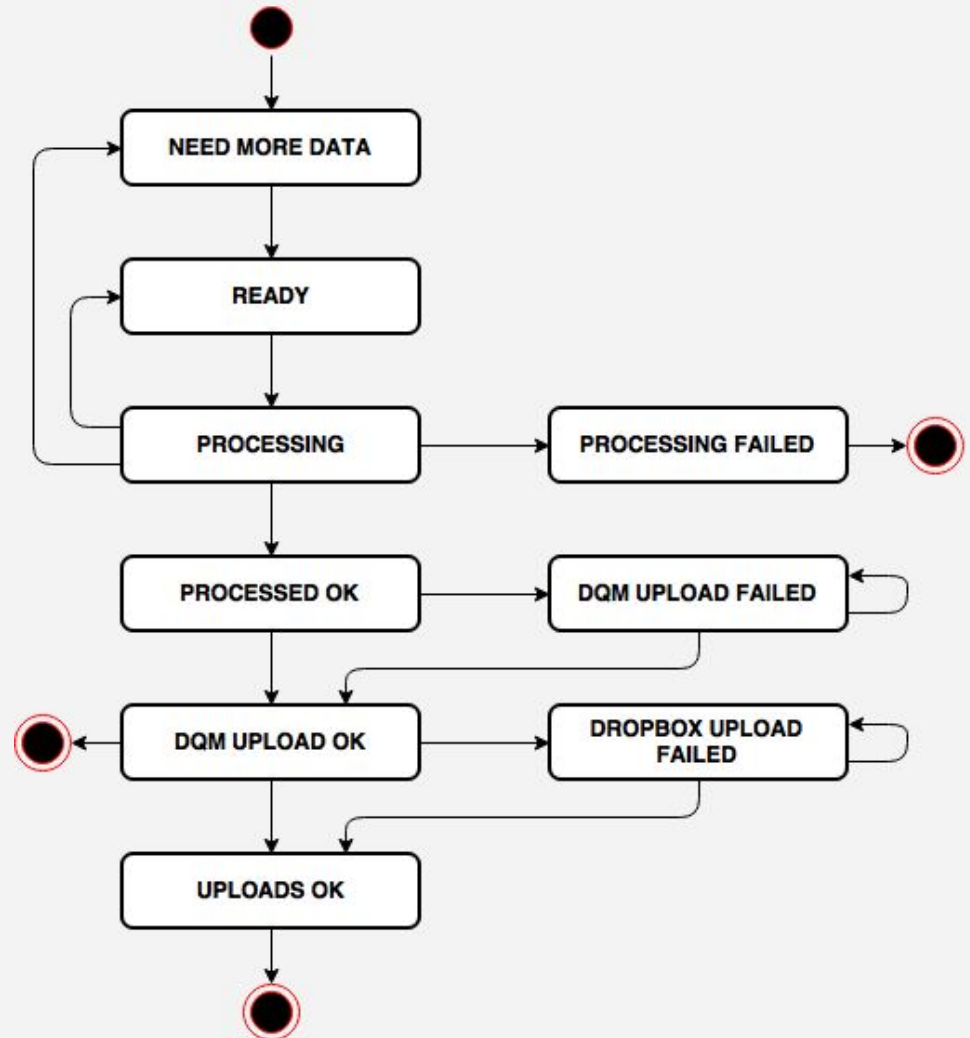
Multi-run harvesting architecture

- Tier0 runs all the calibration and alignments workflows (including CPU-intensive jobs) while the multi-run datasets are used only for computing AlCa Harvesting step in a minimal environment prepared on top of the CMSSW and Tier0 implementation
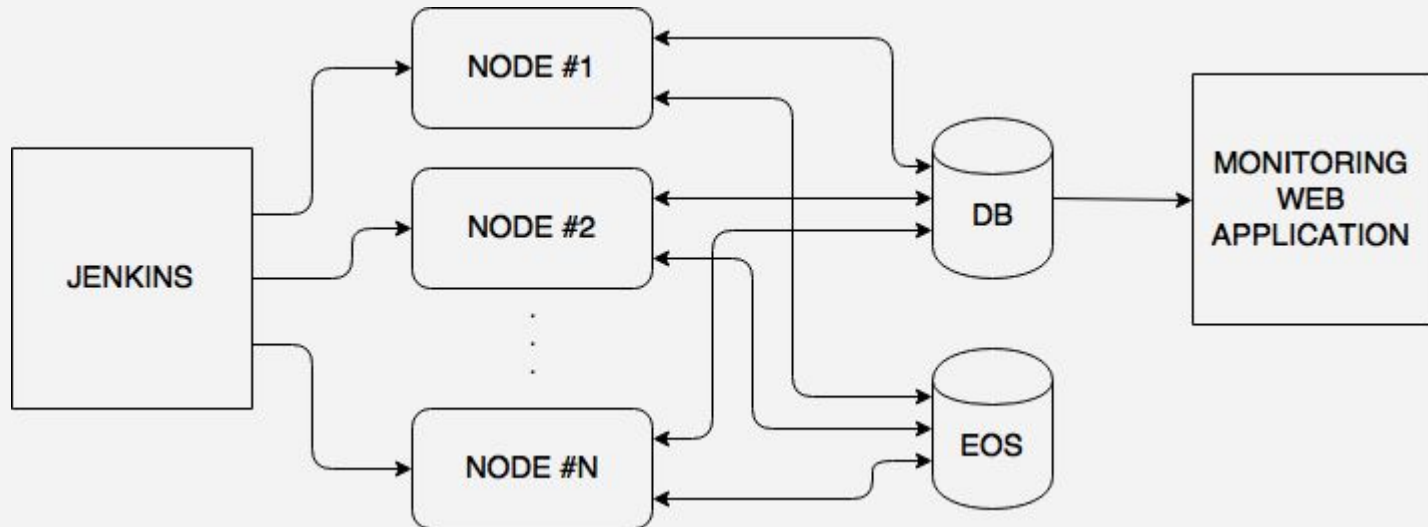
Monitoring infrastructure - *pclmon* (run-based workflows)
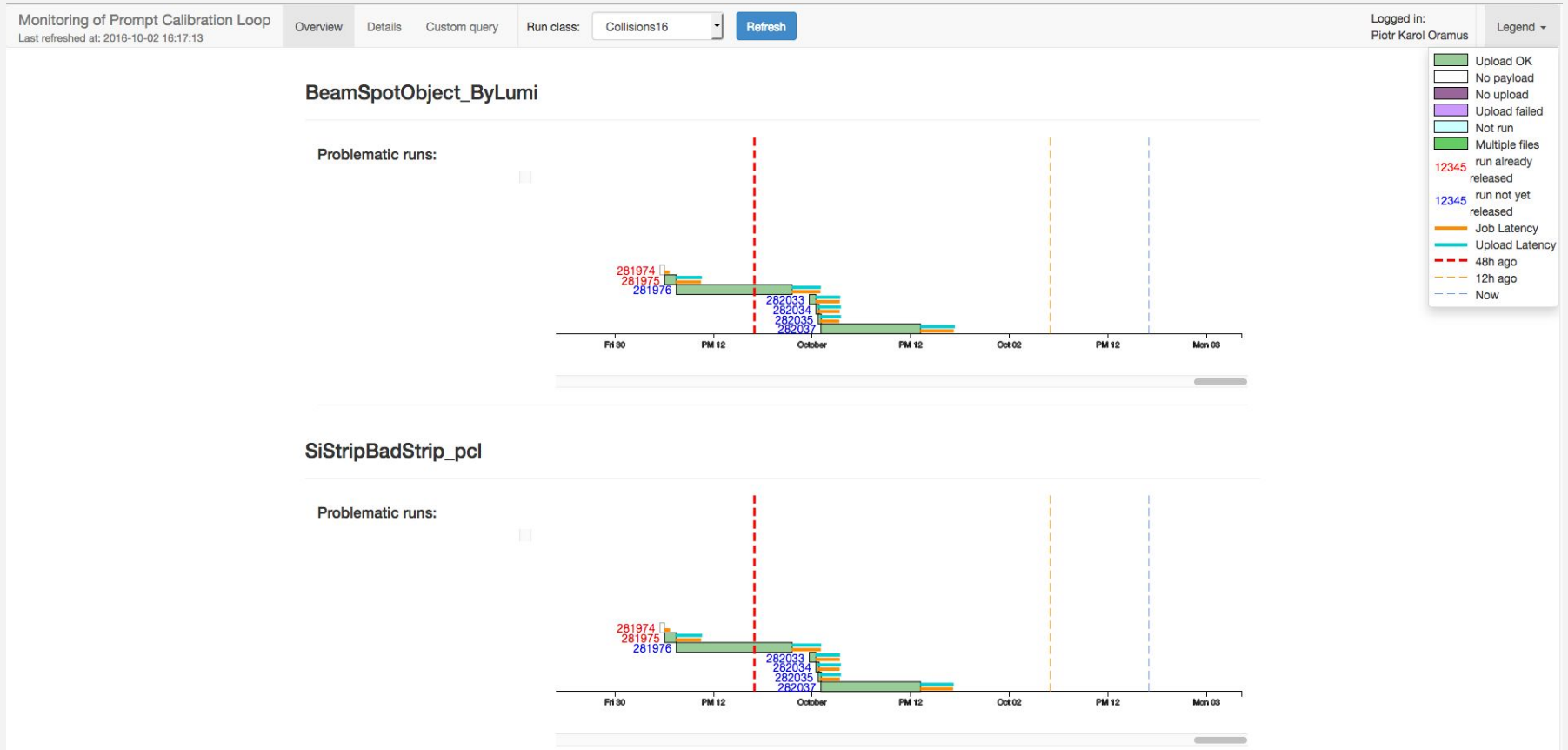
Monitoring infrastructure - multi-run based workflows

PromptCalibProdSiPixelAli

	ID	State	Dataset	Events
+	293798	Processing	/StreamExpress/Run2016H-PromptCalibProdSiPixelAli-Express-v2/ALCAPROMPT	2168478
+	263850	Processing failed	/StreamExpress/Run2016H-PromptCalibProdSiPixelAli-Express-v2/ALCAPROMPT	4371469
+	258321	DQM upload OK	/StreamExpress/Run2016H-PromptCalibProdSiPixelAli-Express-v2/ALCAPROMPT	121061
+	254975	DQM upload OK	/StreamExpress/Run2016H-PromptCalibProdSiPixelAli-Express-v2/ALCAPROMPT	754431
+	236331	Need more data	/StreamExpress/Run2016G-PromptCalibProdSiPixelAli-Express-v1/ALCAPROMPT	3454511
+	233651	DQM upload OK	/StreamExpress/Run2016G-PromptCalibProdSiPixelAli-Express-v1/ALCAPROMPT	616751
+	227980	DQM upload OK	/StreamExpress/Run2016G-PromptCalibProdSiPixelAli-Express-v1/ALCAPROMPT	919996

Current status and future plans

- Currently

  - Run-based workflows computed on Tier0 farm

  - Multi-run workflows computed in pre-production environment

  - Monitoring infrastructure established

- In the future

  - Testing of new multi-run framework

  - Further development of monitoring infrastructure

  - Adding more calibrations to the multi-run system in RunII

Piotr Oramus

CERN

piotr.oramus@cern.ch