# Online and offline QC in ALICE

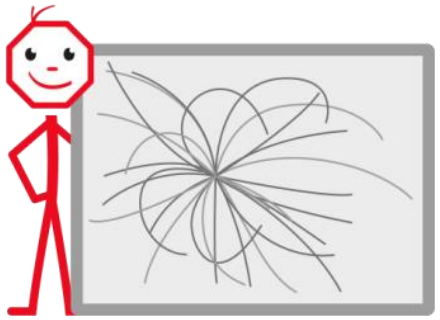
Barthélémy von Haller

CERN

3.12.2018

# Outline

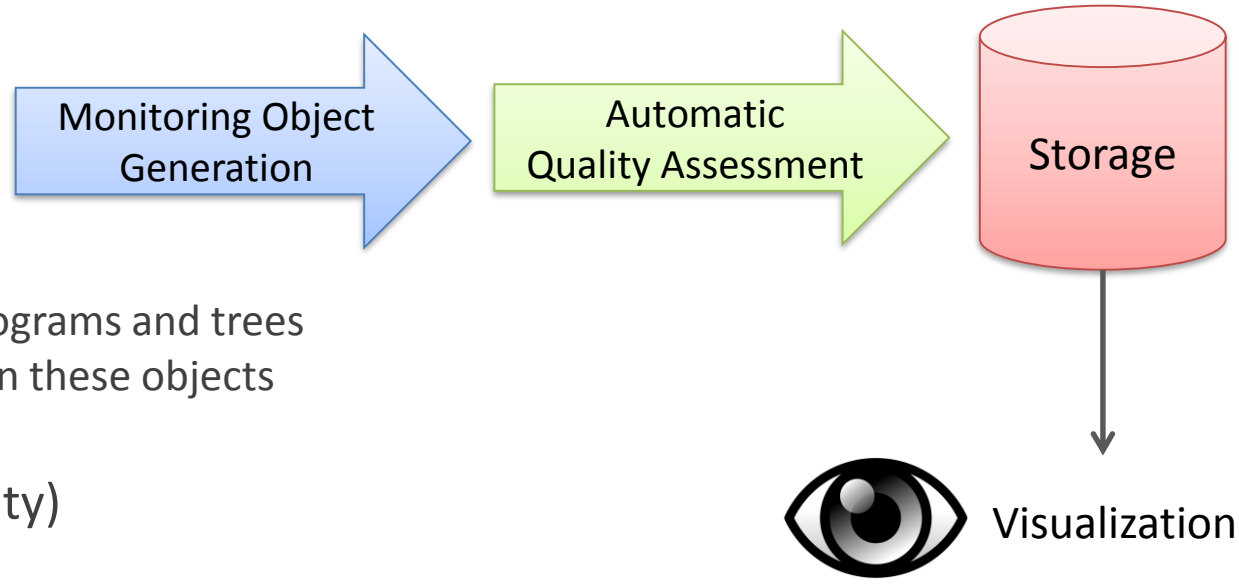
- DQM, QA and QC
  - Definitions
- Run 2
  - Brief overview
- Run 3
  - Overview (O2, principles, organization)
  - Architecture
- A word about ML techniques for QC in Run 3
  - Applicability, challenges and strategy



# Definitions

## DQM, QA and QC

- **Feedback** on the quality of data
  - Online (DQM – Data Quality Monitoring)
    - Make sure to **record** high quality data
    - Identify and solve problem(s) early
  - Offline (QA – Quality Assurance)
    - Make sure to **analyze** high quality data
    - Identify high quality runs
  - QC = (DQM + QA) * Run 3/4
- Involves (at the very least)
  - Analysis by user-defined algorithm
    - Production of monitoring objects such as histograms and trees
    - Assessment of the quality of the data based on these objects
  - Storage of monitoring data
  - Visualization (+ human assessment of quality)



# Run 2 : DQM & QA

- Clear boundary between offline and online
  - Different “frameworks” (AMORE vs AliRoot QA vs offline QA scripts)
  - Different output in separated storages
  - Different lifetime : DQM data available for 2 weeks vs forever in QA
- Really ?
  - AliRoot QA can be used within AMORE
  - Offline QA can be run within HLT
  - HLT QA output imported in AMORE
  - HLT QA output sent to Overwatch

# Run 3

QC

- Online / Offline distinction is gone in $O^2$
  - Unique framework for both DQM and QA
- No Undo/Redo in O2
  - Increased importance of DQM and QA

- *Data Quality Control and Assessment* (QC)

- Long definition :
  Framework and infrastructure for all aspects related to software aimed at identifying possible issues with the data itself, and indirectly with the processing done synchronously and asynchronously (e.g. reco and calib).

# Run 3

## $O^2$ Organization

- Work Packages (formerly Computing working groups)
- WP7 is dedicated to QC
- Joint meetings with DPG-QA-tools
  - Wednesday at 14:30
- Everyone can join
  - To discuss the matter
  - Help design or implement
  - Get informed

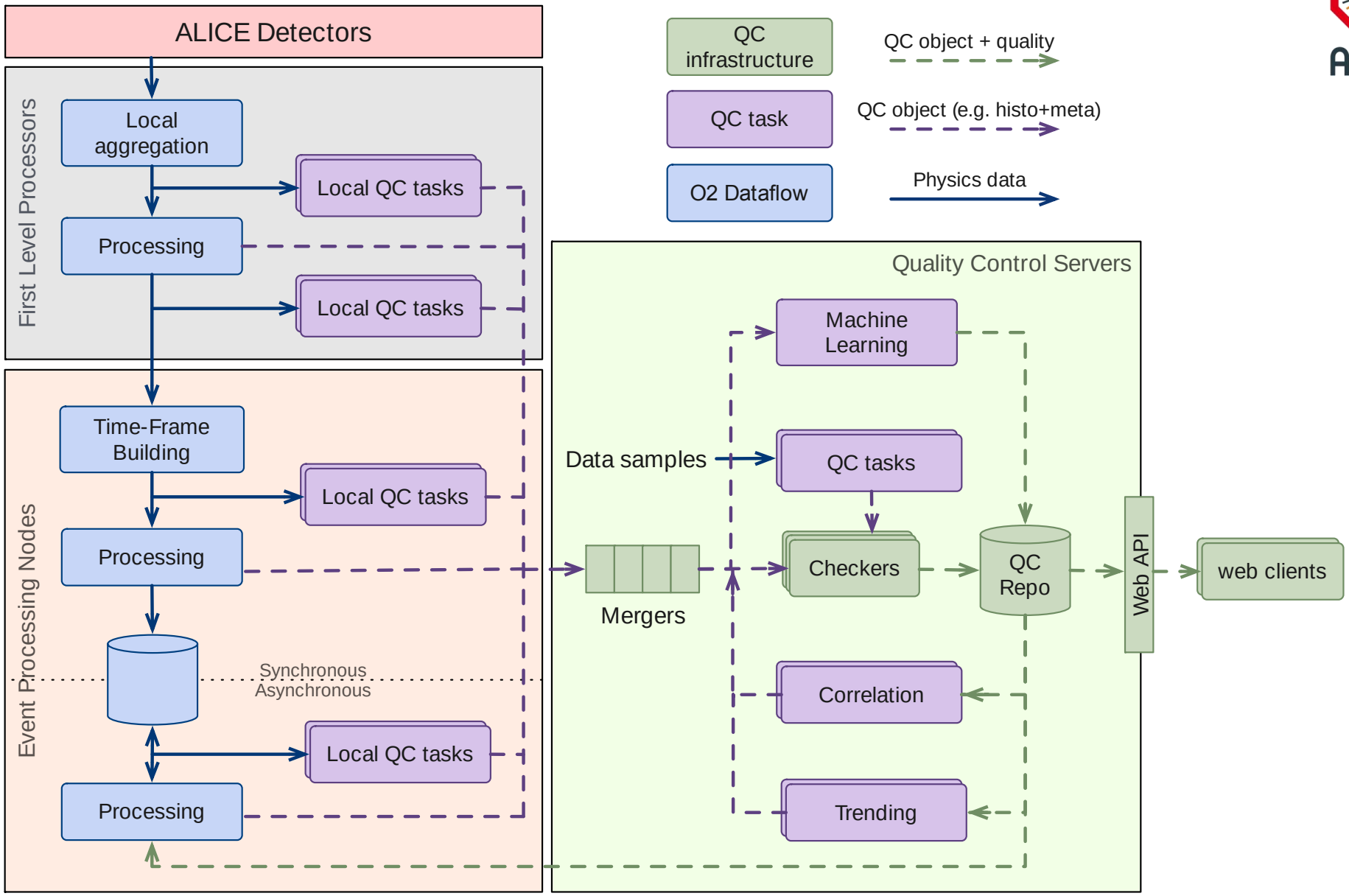
! Warning !

Big schema ahead

Stay close to the guide !

# Run 3

## Architecture

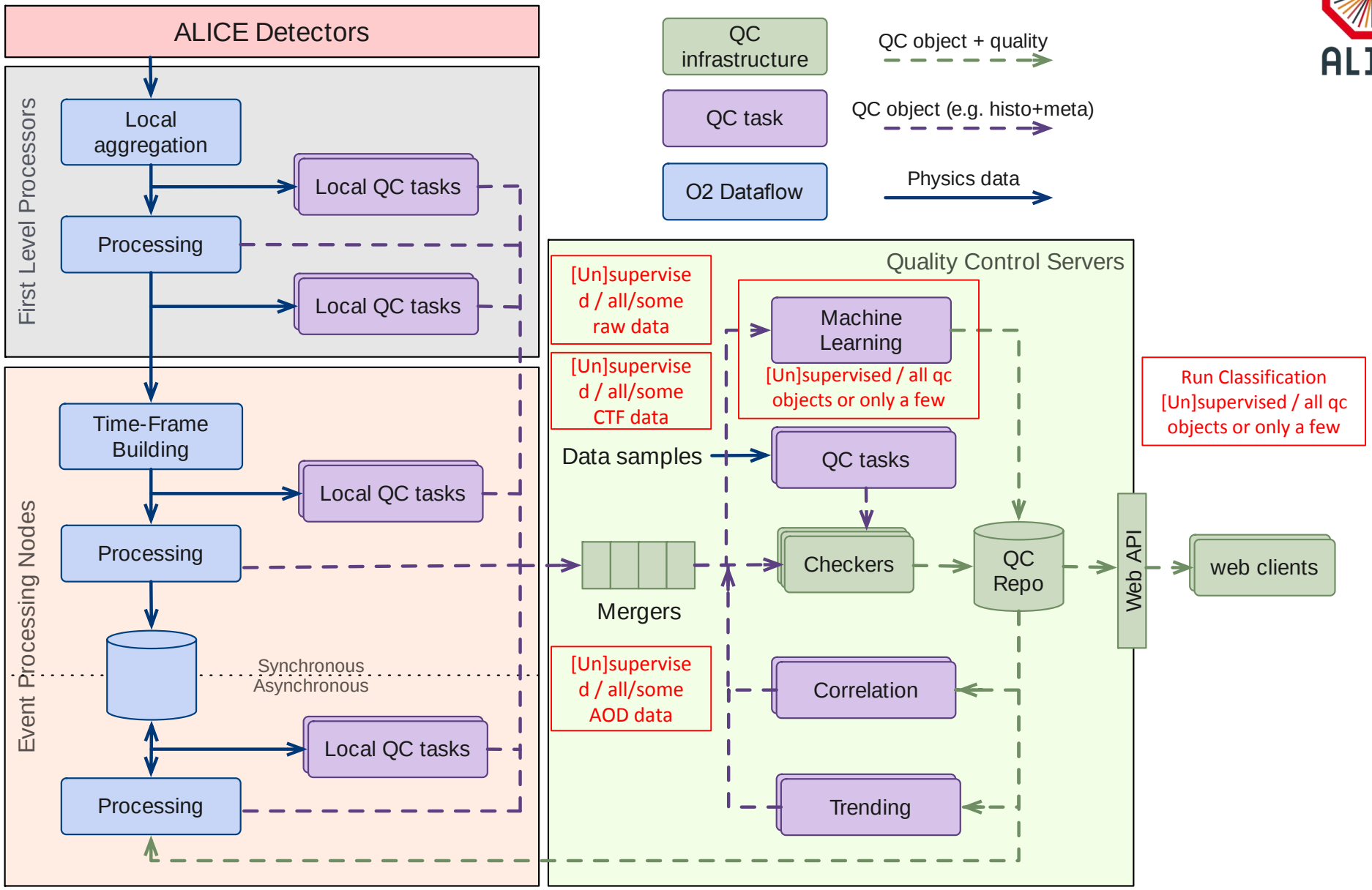


# Machine Learning in QC (1)

- Applicability
  - Probable (cf other experiments and common sense) … but to be proven !
  - Where ? How ?
    - *instead* of all QC tasks or specific one(s) ?
    - *instead* of all QC checkers or specific one(s) ?

# Run 3

## Architecture



# Machine Learning in QC (2)

- Challenges (just a few)
  - (Too) many possibilities, risk of dispersion, redundant efforts
  - No data, and no real data before Run 3
  - Data format should be digestable by ML
  - Labels (flags) should be reliable
  - We need to prove that results are correct
  - Not fall for the buzz of the moment

("blockchain-backed ML using Quantum Computers in the clouds")

# Machine Learning in QC (3)

- Strategy
  - Full “traditional” QC system ready for Run 3
    - Necessary for comparison and for the start of Run 3
  - Allow and nurture ML alternatives