Data Quality Monitoring in Run 2 and Run 3

B. von Haller

CERN

24.11.2016

Outline

- DQM definition
- Run 2 (2015-2018)
  - AMORE
  - Event Display
- Run 3 (2021-2024?)
  - ALICE
  - O^2
  - QC

Quite some details.
Pick whatever you need
and ignore the rest !

Online Data Quality Monitoring

Definition

- **Online** feedback on the quality of data
- Make sure to take and record high quality data
- Identify and solve problem(s) early
- Data Quality Monitoring (DQM) involves
    1. Online Gathering of data
    2. Analysis by user-defined algorithm
    3. Storage of monitoring data
    4. Visualization

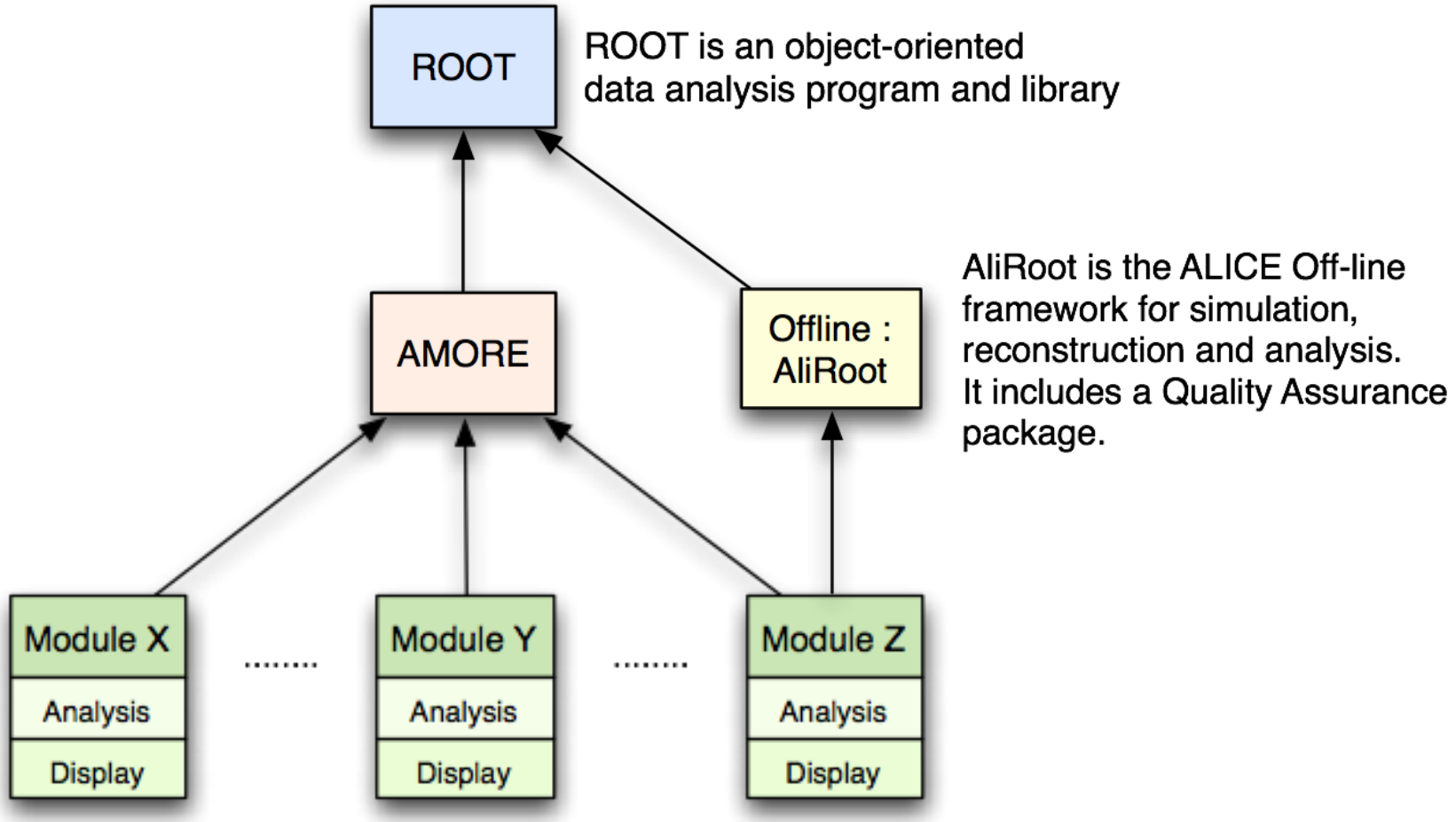
- The offline counterpart is usually called *Quality Assurance (QA)*

Run 2

- 2015-2018
- ALICE Systems are stable
- No more development (in general), only maintenance
- DQM framework is called AMORE (Automatic MOnitoring Environment)
  - It is not really automatic though, mostly due to
    - complexity of determining the rules to say whether a plot is good or bad
    - general reluctancy not to rely on humans for this task

AMORE

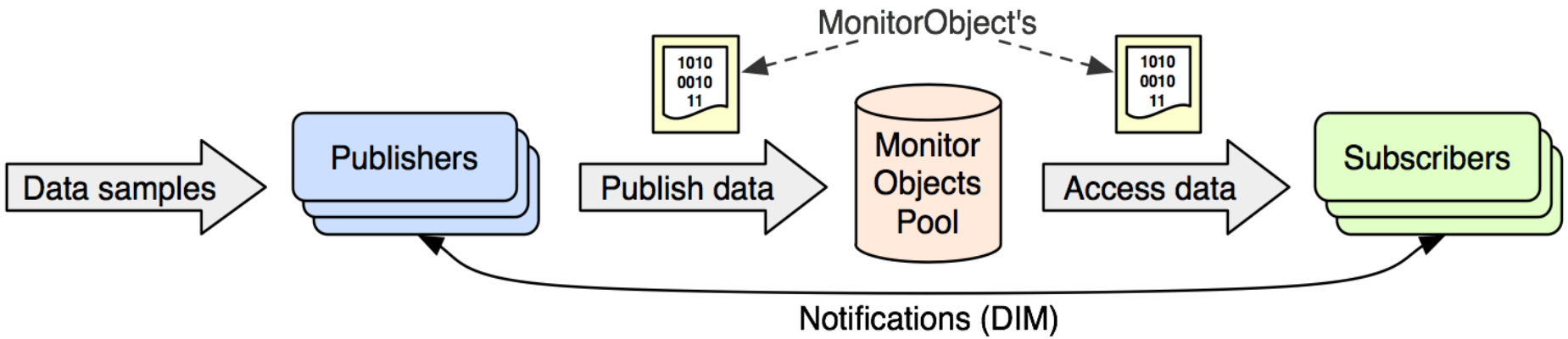
DQM framework



AMORE

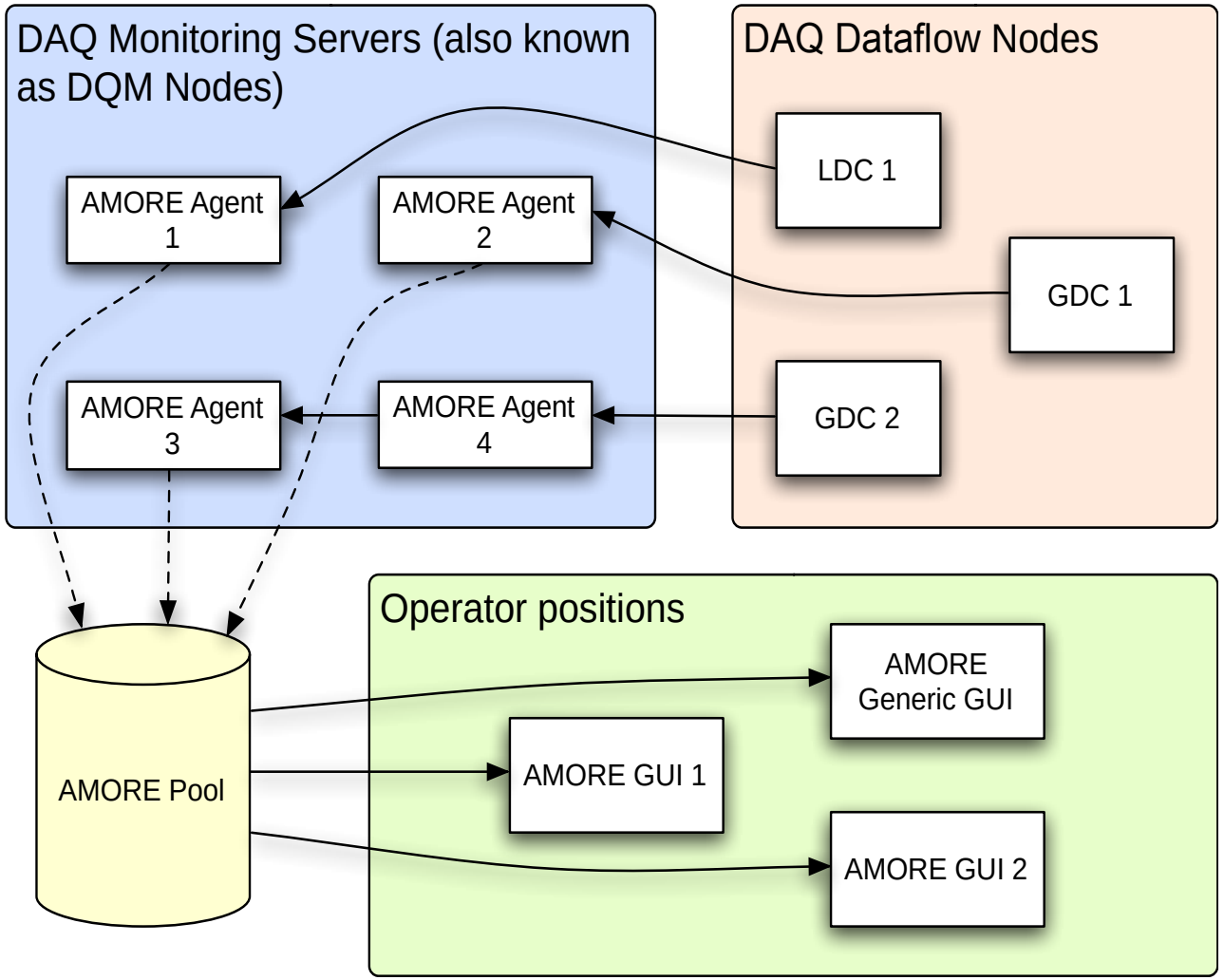
Architecture

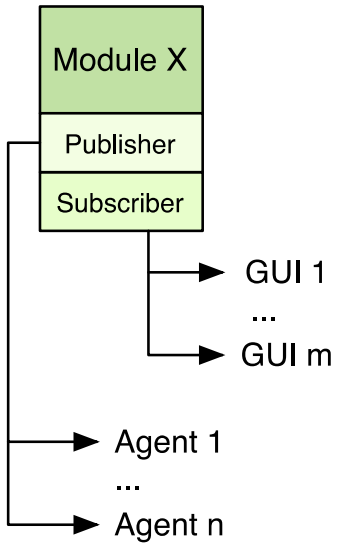
- Publisher – Subscriber paradigm
- Database used for the data pool
- Published objects are encapsulated into « MonitorObject »
- Plugin architecture using ROOT reflection
- Modules are dynamic libraries loaded at runtime



AMORE

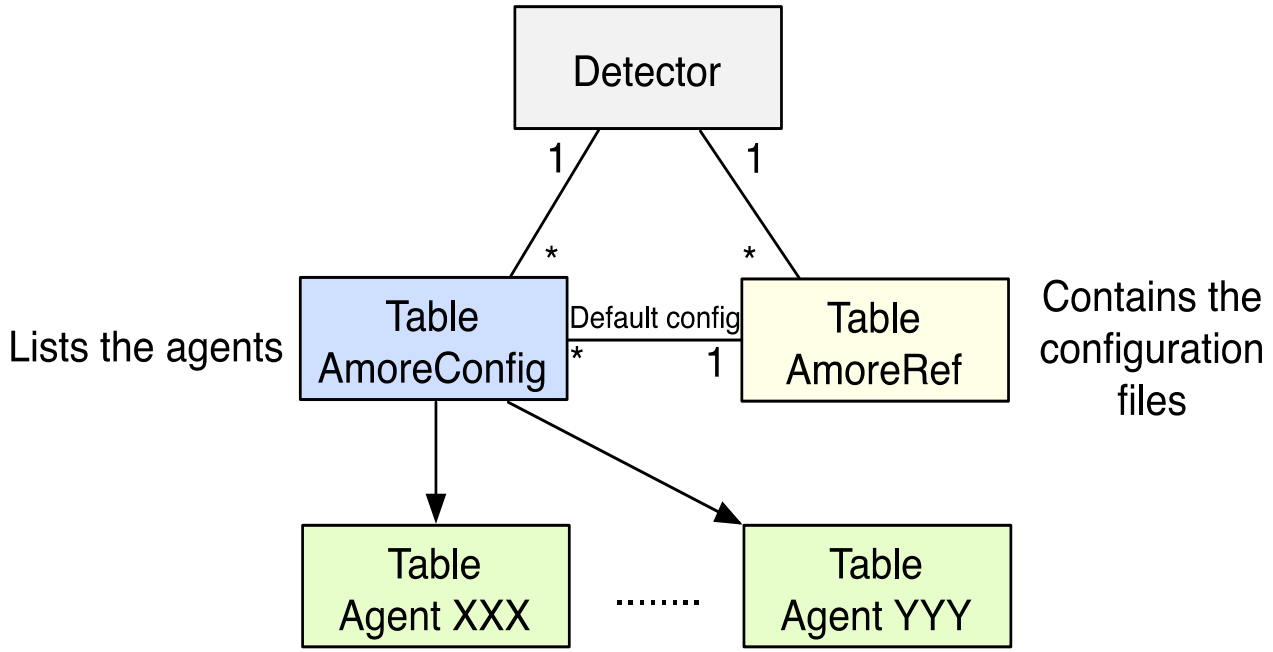
Data flow





The pool

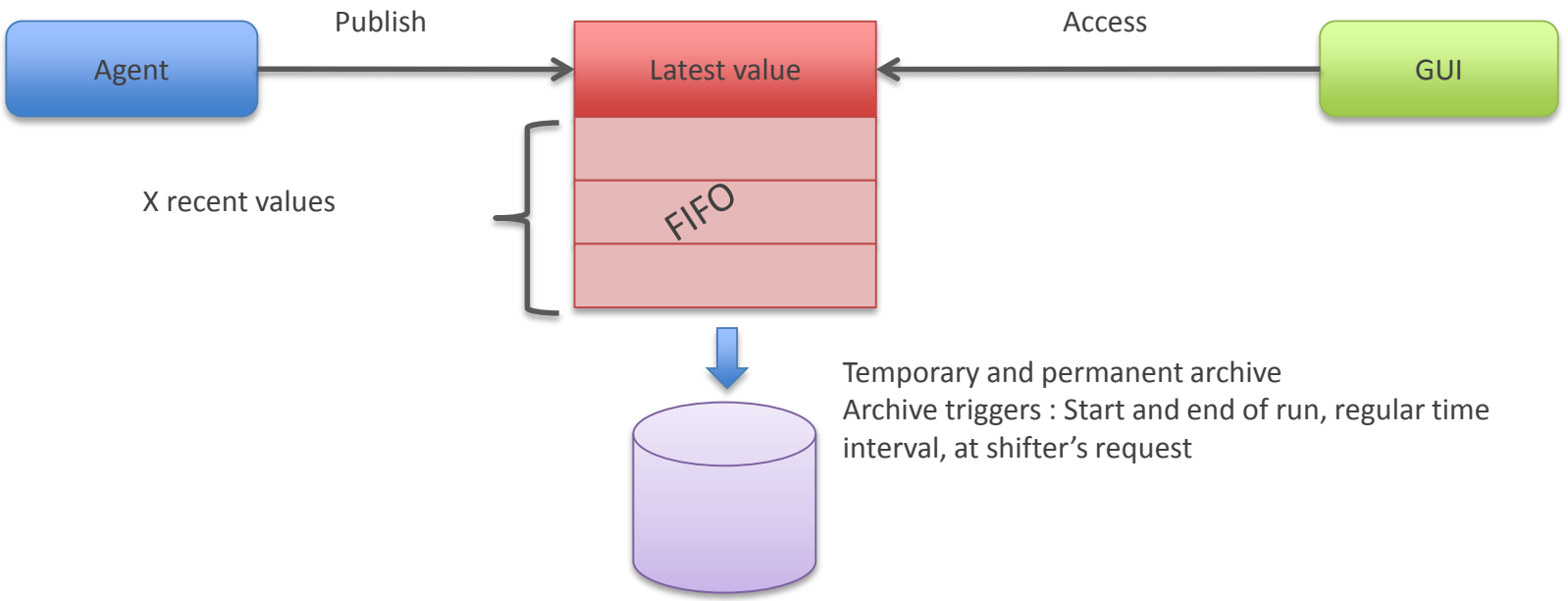
- Current implementation based on a database
- MySQL : light-weight, reliable, freely distributable



AMORE

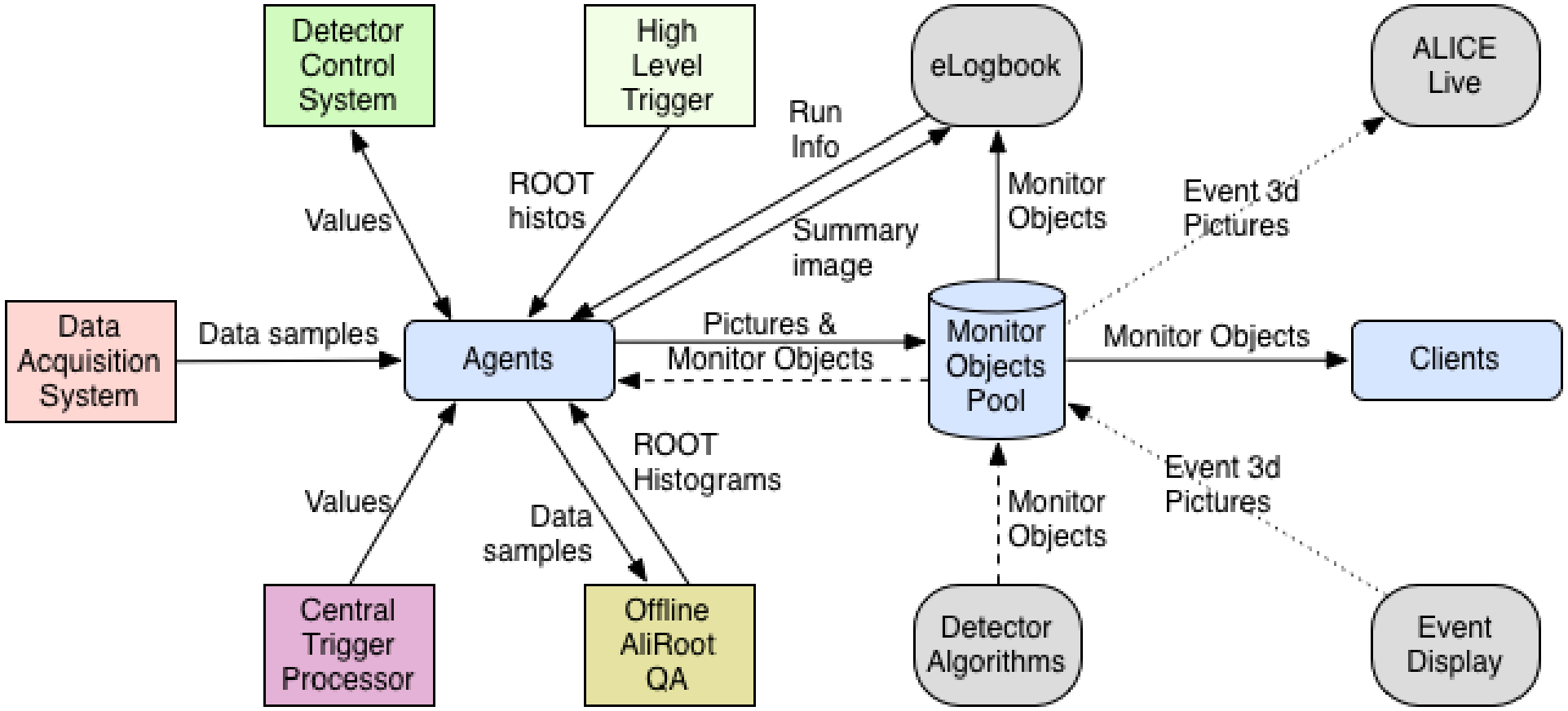
Archives

- Short-term history : First-In First-Out (FIFO)
- Long-term archives : At end of run, regular intervals, and users' request



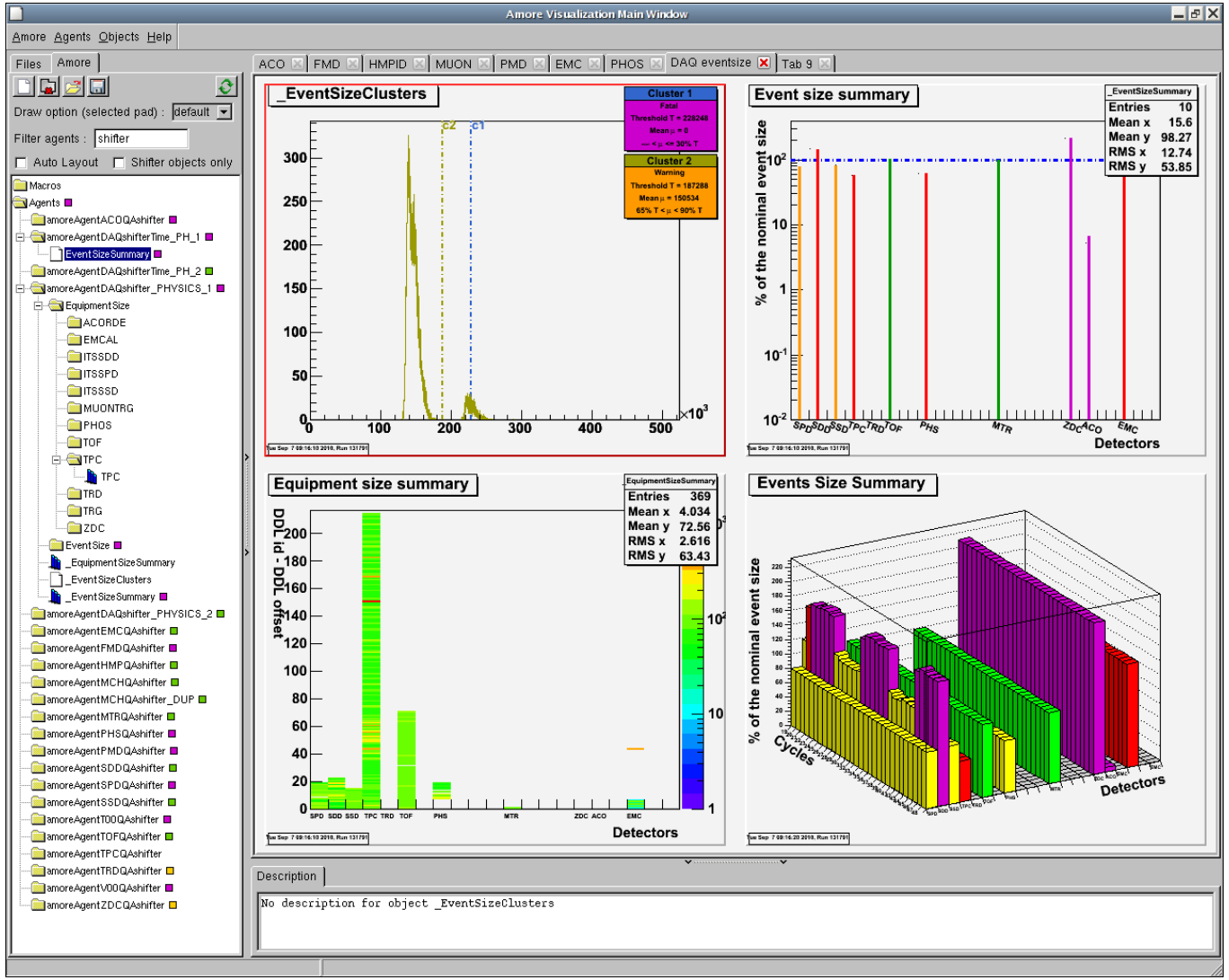
AMORE

Data providers



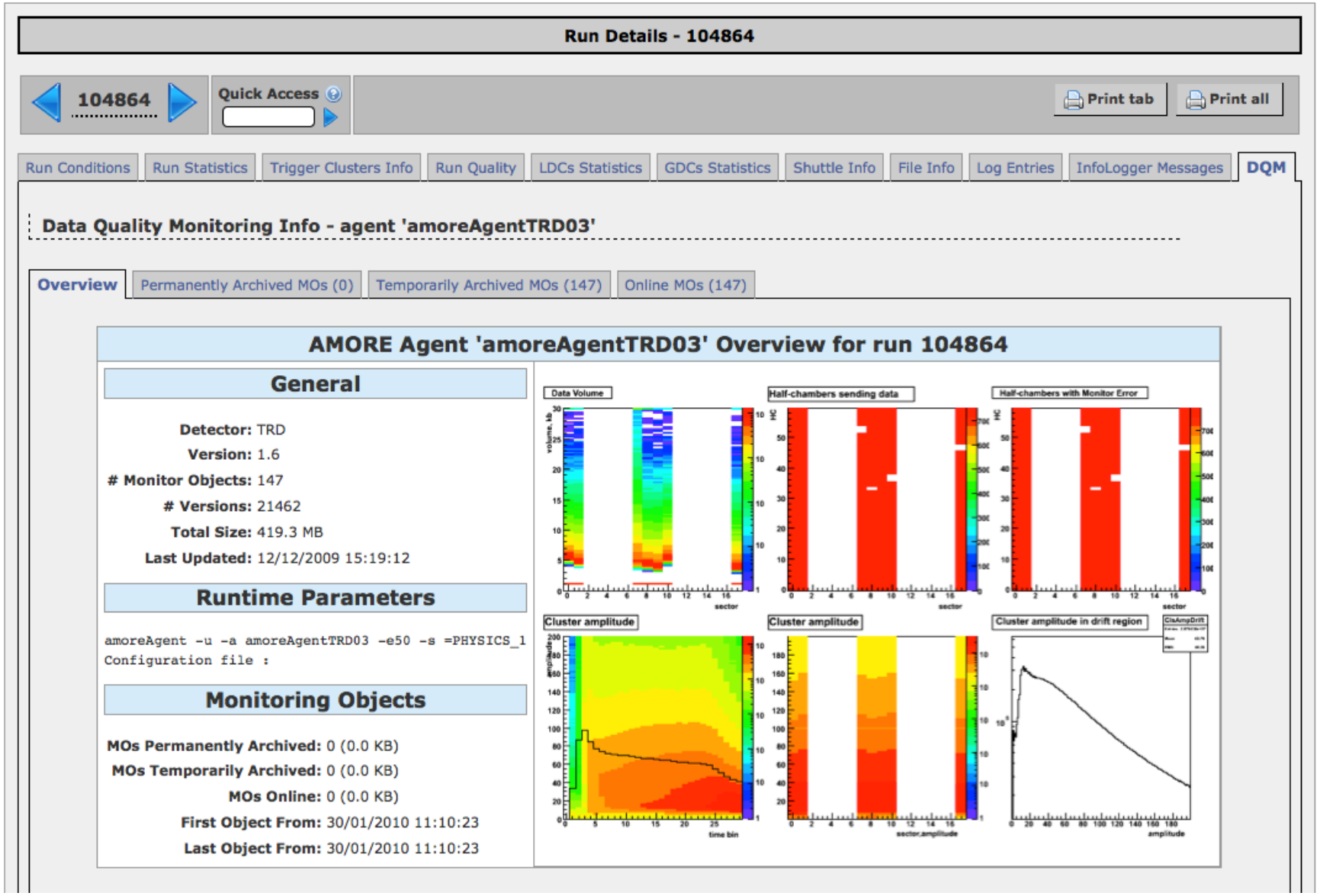
AMORE

GUI


Amore Visualization Main Window
Amore Agents Objects Help
Files Amore
Draw option (selected pad) : default
Filter agents : shifter
Auto Layout Shifter objects only
Macros
Agents
amoreAgentACOQAshifter
amoreAgentDAQshifterTime_PH_1
EventSizeSummary
amoreAgentDAQshifterTime_PH_2
amoreAgentDAQshifter_PHYSICS_1
EquipmentSize
ACORDE
EMCAL
ITSSDD
ITSSPD
ITSSSD
MUONTRG
PHOS
TOF
TPC
TPC
TRD
TRG
ZDC
EventSize
_EquipmentSizeSummary
_EventSizeClusters
_EventSizeSummary
amoreAgentDAQshifter_PHYSICS_2
amoreAgentEMCQAshifter
amoreAgentFMDQAshifter
amoreAgentHMPQAshifter
amoreAgentMCHQAshifter
amoreAgentMCHQAshifter_DUP
amoreAgentMTRQAshifter
amoreAgentPHSQAshifter
amoreAgentPMDQAshifter
amoreAgentSDDQAshifter
amoreAgentSPDQAshifter
amoreAgentSSDQAshifter
amoreAgentT00QAshifter
amoreAgentTOFQAshifter
amoreAgentTPCQAshifter
amoreAgentTRDQAshifter
amoreAgentV00QAshifter
amoreAgentZDCQAshifter
ACO FMD HMPID MUON PMD EMC PHOS DAQ eventsize Tab 9
_EventSizeClusters
Cluster 1
Cluster 2
Event size summary
_EventSizeSummary
Entries 10
Mean x 15.6
Mean y 98.27
RMS x 12.74
RMS y 53.85
% of the nominal event size
Detectors
Equipment size summary
EquipmentSizeSummary
Entries 369
Mean x 4.034
Mean y 72.56
RMS x 2.616
RMS y 63.43
DDL id - DDL offset
Events Size Summary
Cycles
Description
No description for object _EventSizeClusters

AMORE

Logbook access



Control and Configuration in production

- Set of web tools to let admins, experts and shifters
  - Start/stop agents
  - Configure agents
  - Release agents

Development setups

- Detectors are encouraged to install their own development and test setup
- Guest machines are available as well
- It runs on SLC6 (<- SL6 <- RHEL6) only
- Distributed as RPMs
- https://alice-daq.web.cern.ch/products/amore-setup

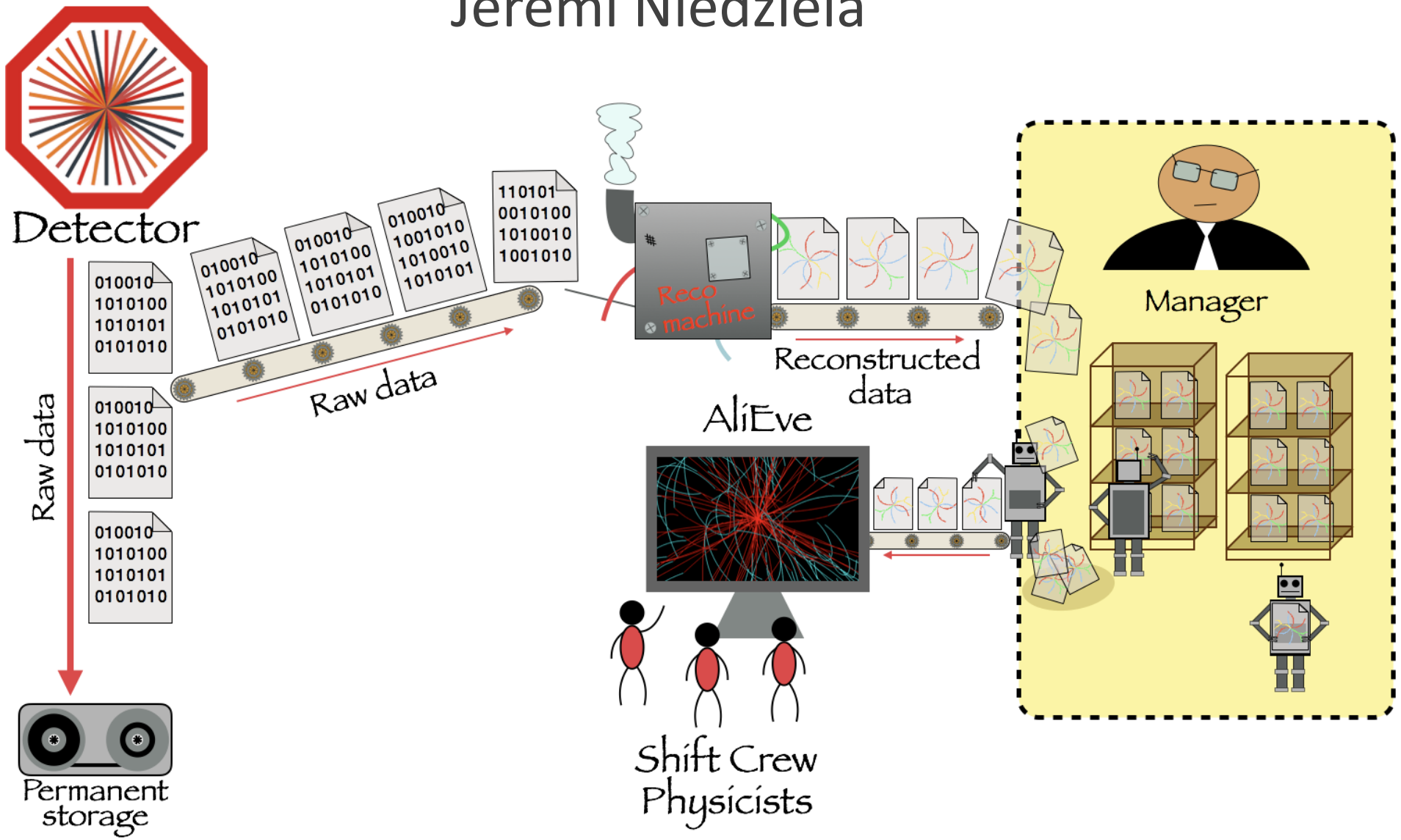
Documentation

- Entry point : https://alice-daq.web.cern.ch/products/amore
- Very extensive (almost exhaustive but who dare says that ?)

Run 2

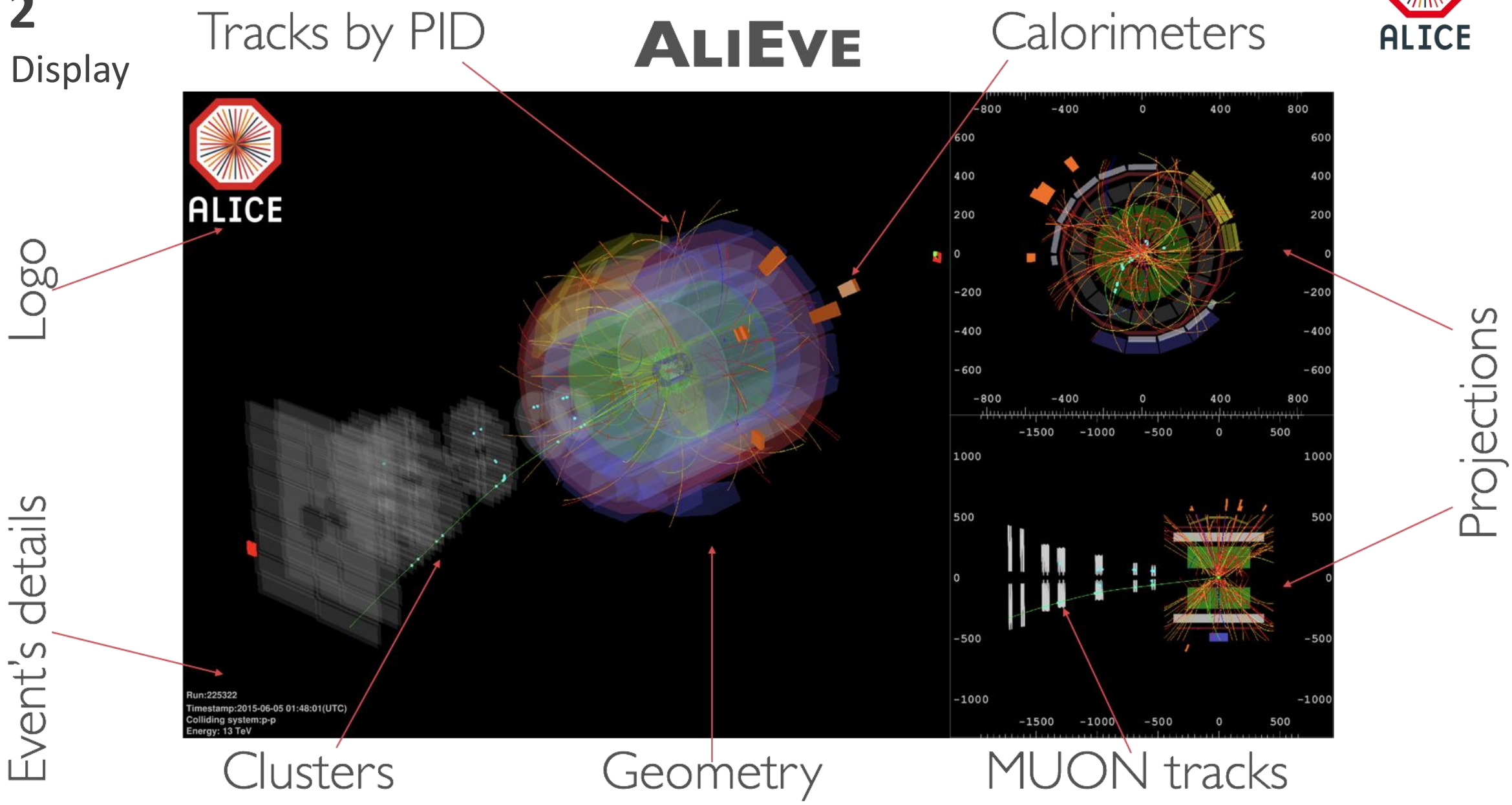
Event Display

Jeremi Niedziela



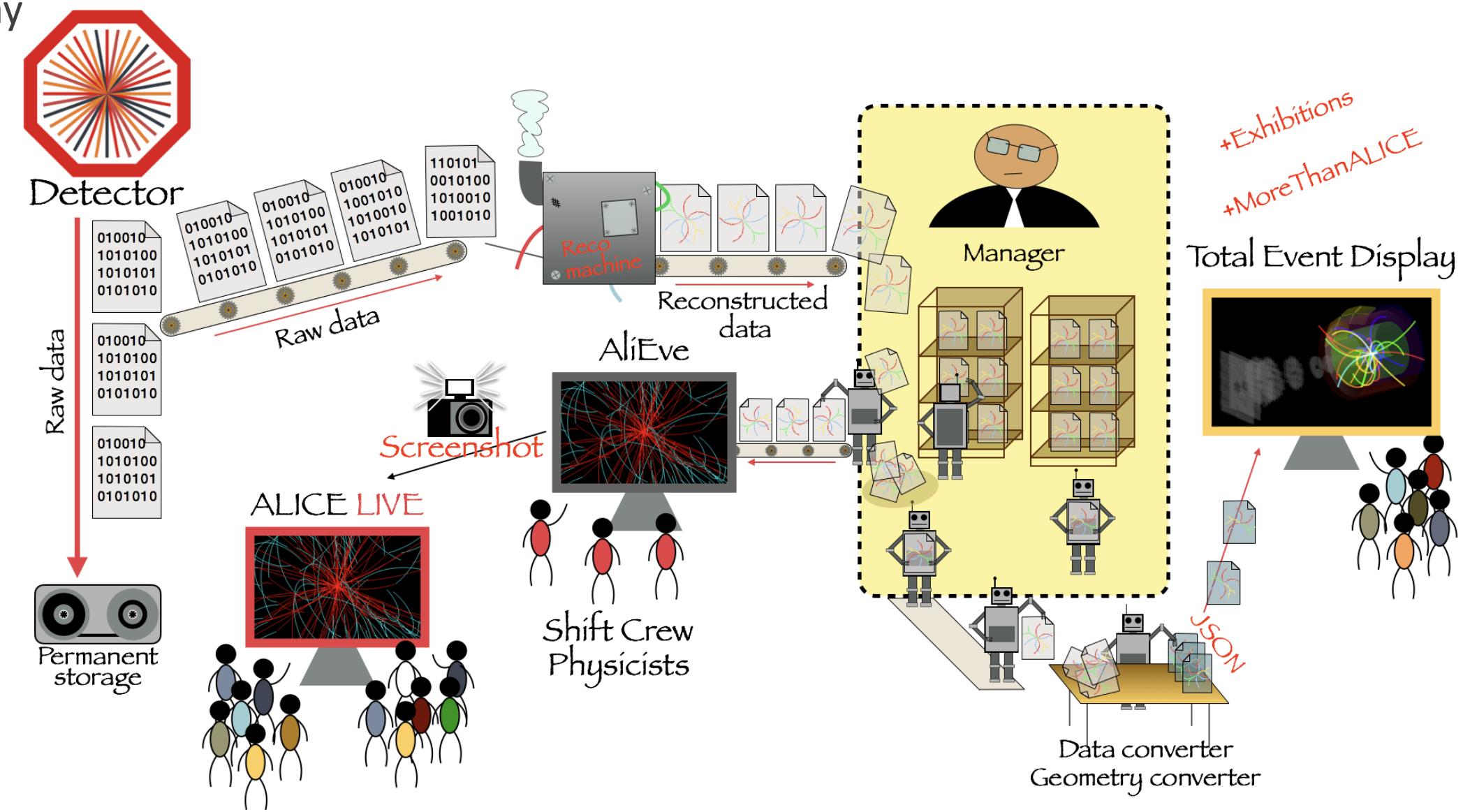
Run 2

Event Display



Run 2

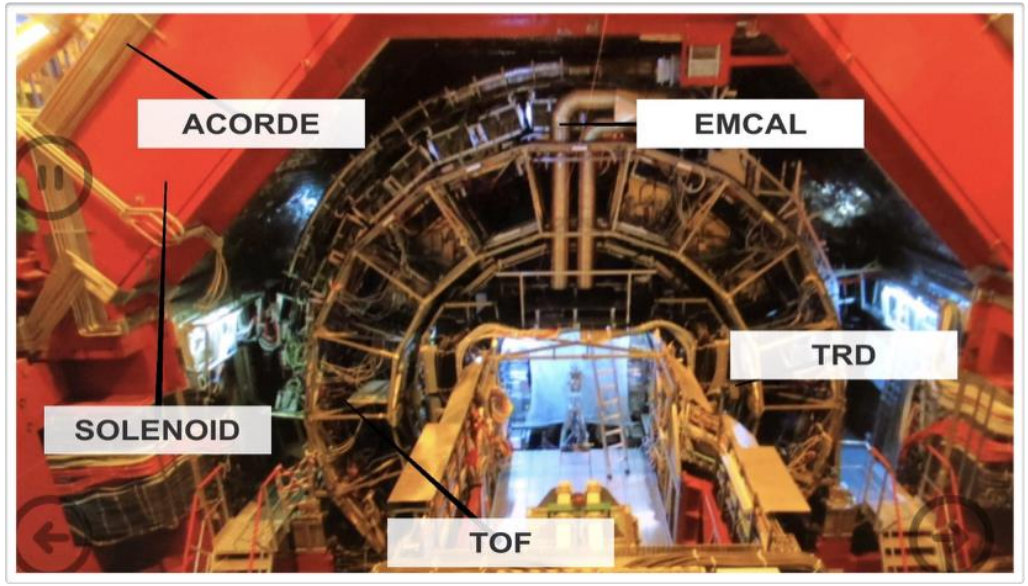
Event Display



Run 2

More Than ALICE

- AR + VR



Run 3

Context

- 2021-2024
- ALICE major upgrade during LS2
- LHC upgrade as well

➔ 100 times more data than Run 1

➔ Physics goals make triggering technics inefficient

➔ PbPb interactions at 50kHz -> too fast for TPC -> continuous readout

Run 3

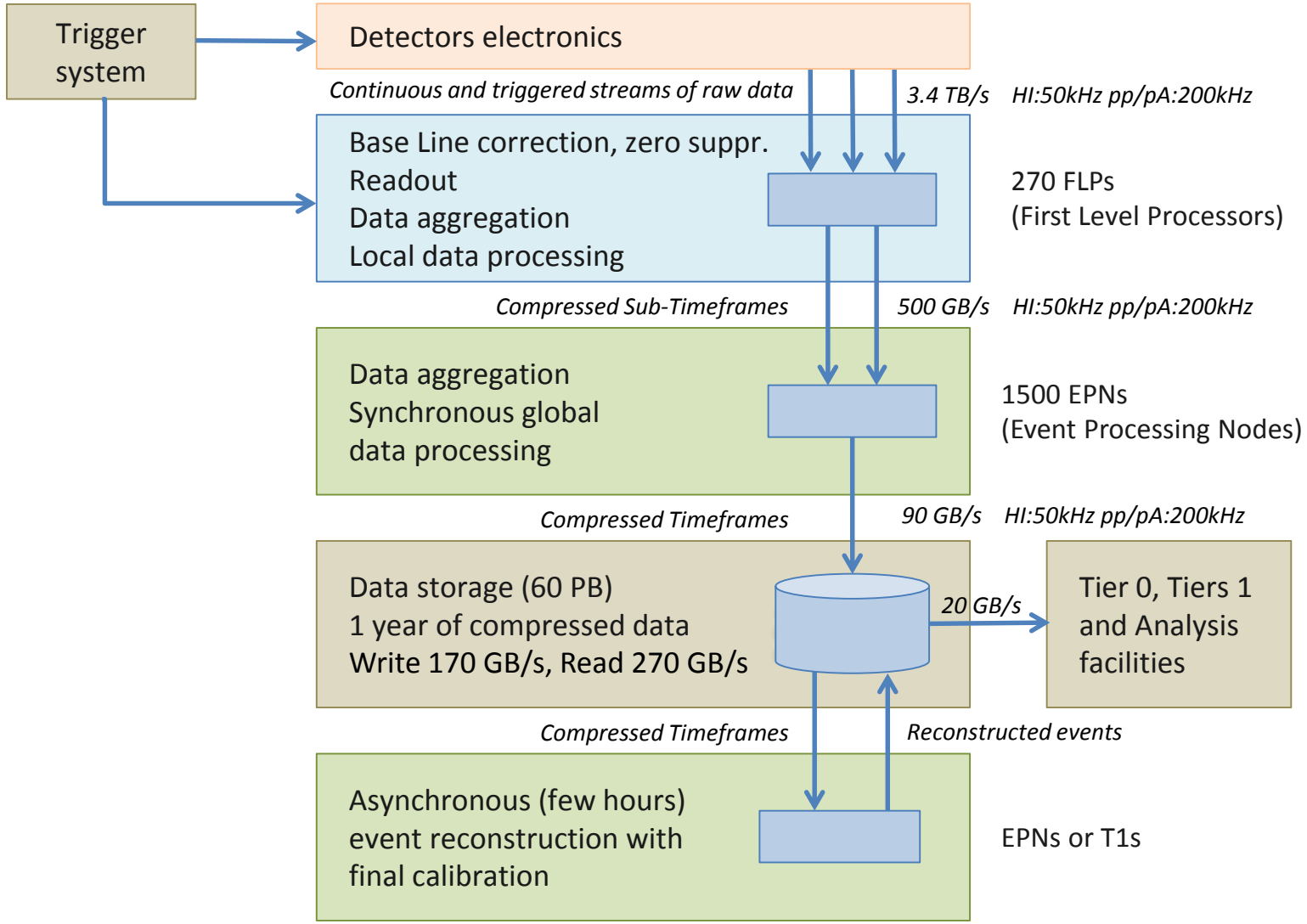
O^2

Requirements make it necessary to design a new computing system characterized by

- The readout of all interactions
- The intelligent compression of this large amount of data by reconstructing and calibrating it online
- One common online and offline computing system called **O^2**

ALICE O²

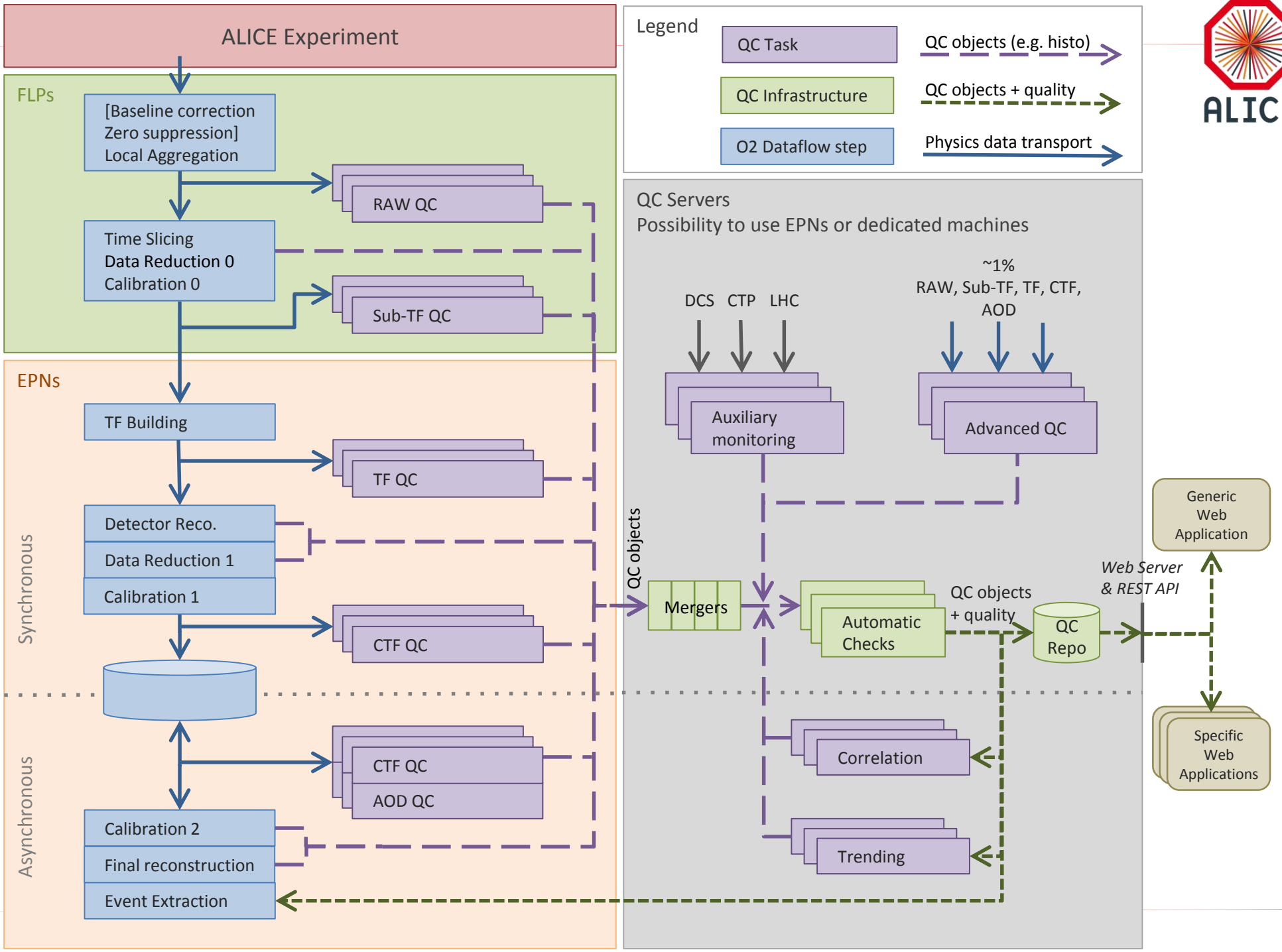
Dataflow



CWG9

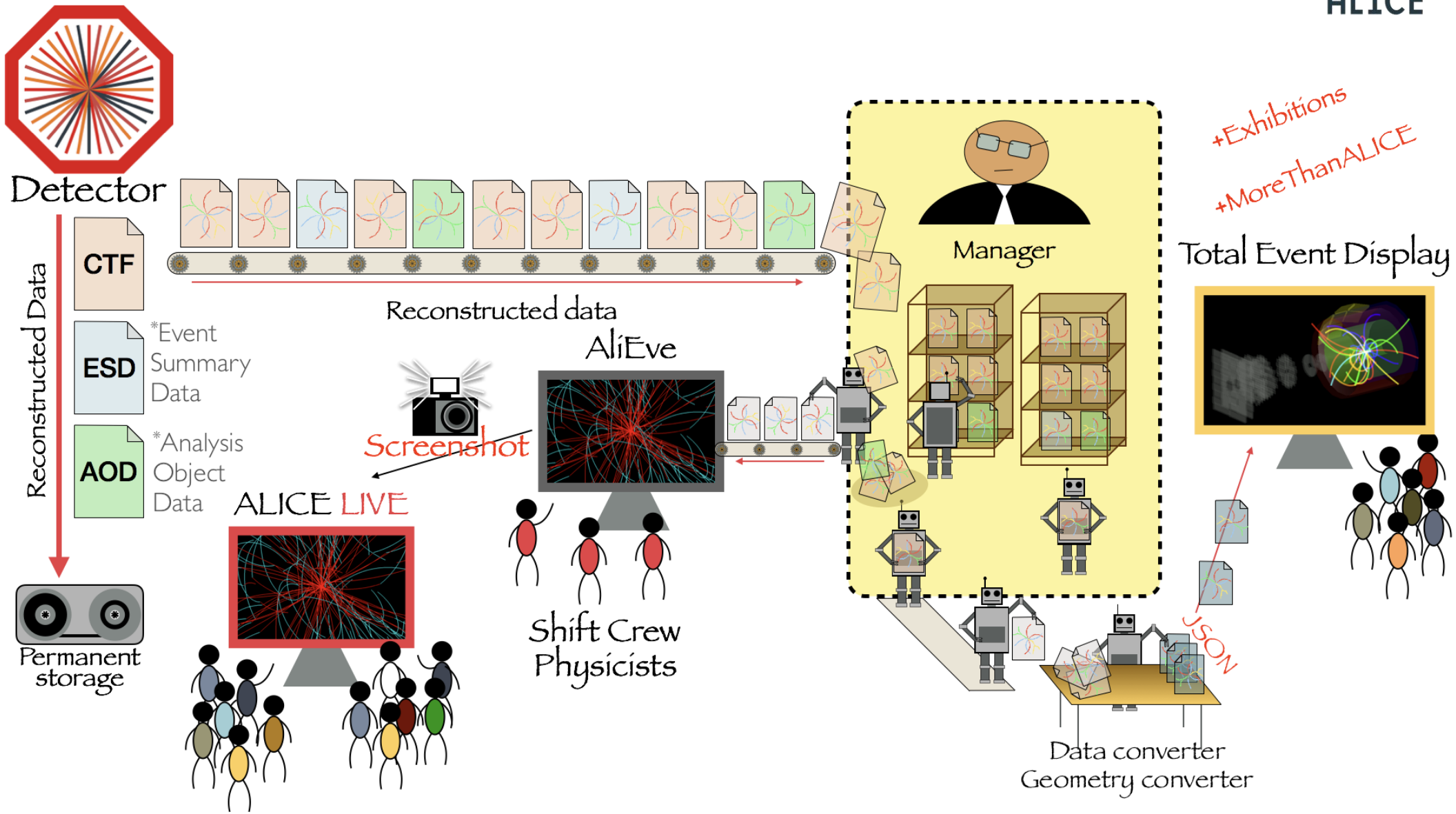
- **Online** DQM, **offline** QA and Visualization
  - No more distinction !
  - ”Data Quality Control and Assessment” -> **QC**
- 15 people from 7 institutes
  - We try to include all interested parties and stakeholders
- Tasks
  - Requirements
  - Design and architecture (->TDR)
  - QC framework prototype and benchmark
  - Merging infrastructure development and benchmark
  - Detectors survey (v2, ongoing)

QC
Architecture



Run 3

Event Display



Run 3

Event Display

- Starting from reconstructed Time Frame.
- Manager has to deal with new data types.
- AliEve must be able to visualise new data types.
- Data converter should be adapted to new data types.
- New geometry has to be prepared.
- Geometry converter should be tested with the new geometry.
- Distinction between events in CTFs not so obvious.

Machine/Deep Learning

- Aim at Run 3, practice in Run 2
- Goal : identify automatically patterns, trends and abnormalities
  - Raise alarms on possible issues with the data
- In addition to the standard deterministic checks written by physicists