QA tools developments

Jacek Otwinowski

(for the DPG QA tools and WP7 groups)

Outline

Main goal is to develop a common set of tools to be used in Run2 and Run3.

- Software validation
- Offline QC tools in Run2
- OVERWATCH - Online QC in Run2
- QC in Run 3
- Outlook

Developments are discussed at the QA tools and WP7 joint meetings (Wednesdays 2:30PM)
alice-dpg-qa-tools@cern.ch
alice-o2-wp7@cern.ch

Software validation with reference RAW and MC data

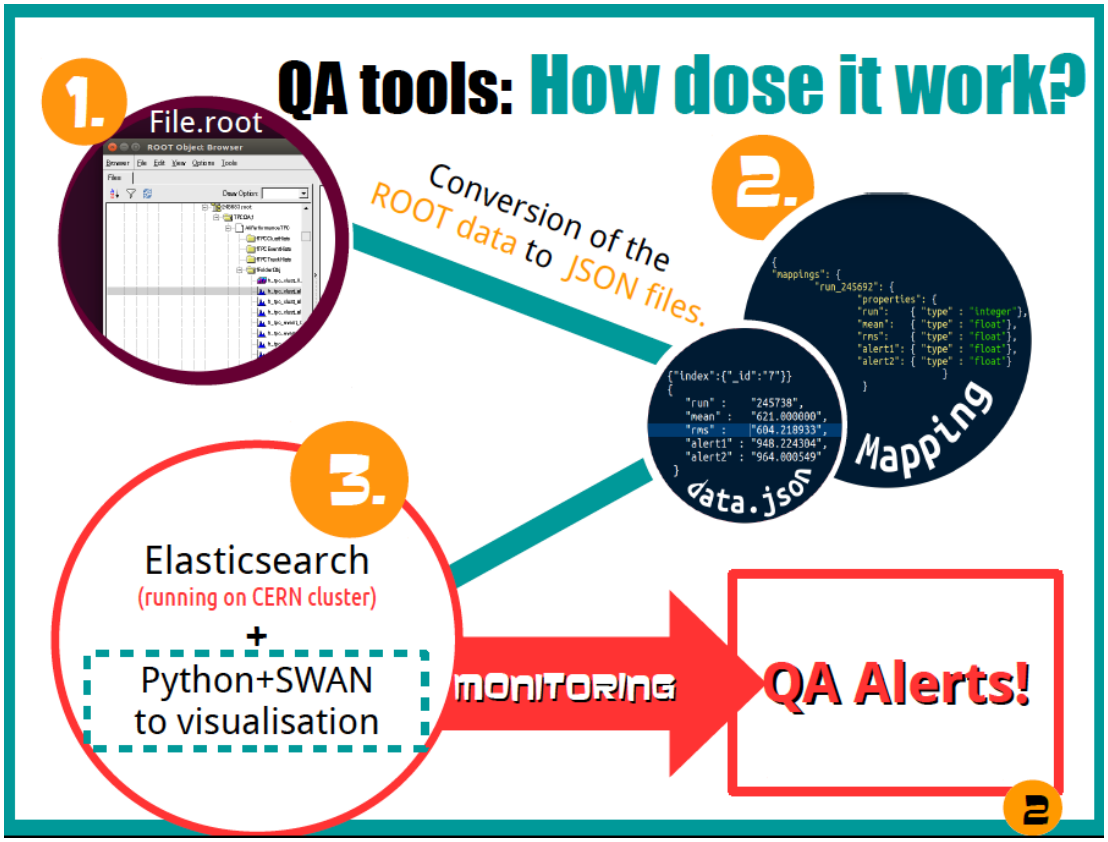
- Full processing chains
  - Data: Calibration, reconstruction, QA
  - MC: Performance generation, Geant3 transport, reconstruction, QA
  - Run for each ALICE software build
- Execution system (implemented, tested) – Matteo Concas/Dario Berzano
  - https://pypi.python.org/pypi/alien-jdl2makeflow
- Performance MC generator (under tests) – Marian Ivanov/Jacek Otwinowski
  - https://alice.its.cern.ch/jira/browse/ATO-245
- Completeness checks (to be implemented)
- Analyze RAM/CPU usage (to be implemented)
- Analyze QA output – trending (to be implemented)

Offline QC tools in Run2

Iwona Sputowska, Marian Ivanov, Jacek Otwinowski,
Boris Rumantsev, Christian Bourjau

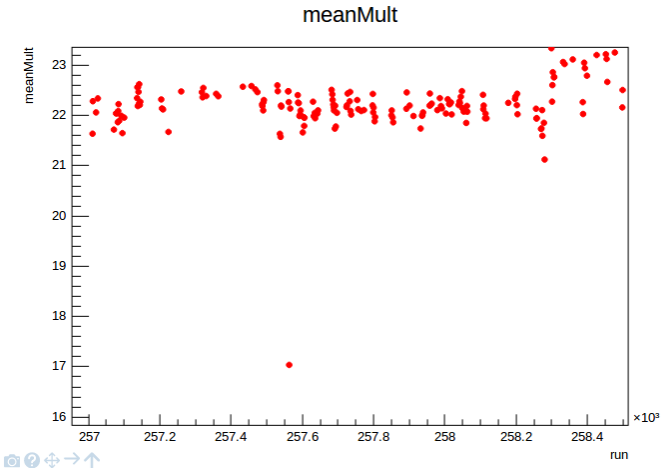
- Detector, trigger, calibration, tracking, PID and analysis QA
- QA based on Elasticsearch (under tests)
  - https://alice.its.cern.ch/jira/browse/ATO-372
- ROOT tree based DB (TPC QA generalization) (implementation ongoing)
  - Time series support for ROOT tree based DB
  - https://indico.cern.ch/event/578479
- Data samples comparison based on trending information (implementation ongoing)
  - https://indico.cern.ch/event/578479
- Data samples comparisons based on the parameterization maps (implementation ongoing)
  - https://alice.its.cern.ch/jira/browse/PWGPP-163
- GUI / Visualization / dashboards (implementation and tests ongoing)
  - SWAN/Jupyter notebook
  - TPC QA WEB interface
  - Dashboards (hackathon student project) - Christian Bourjau
  - https://indico.cern.ch/event/665974

Example: Elasticsearch + SWAN + Python API



Iwona Sputowska

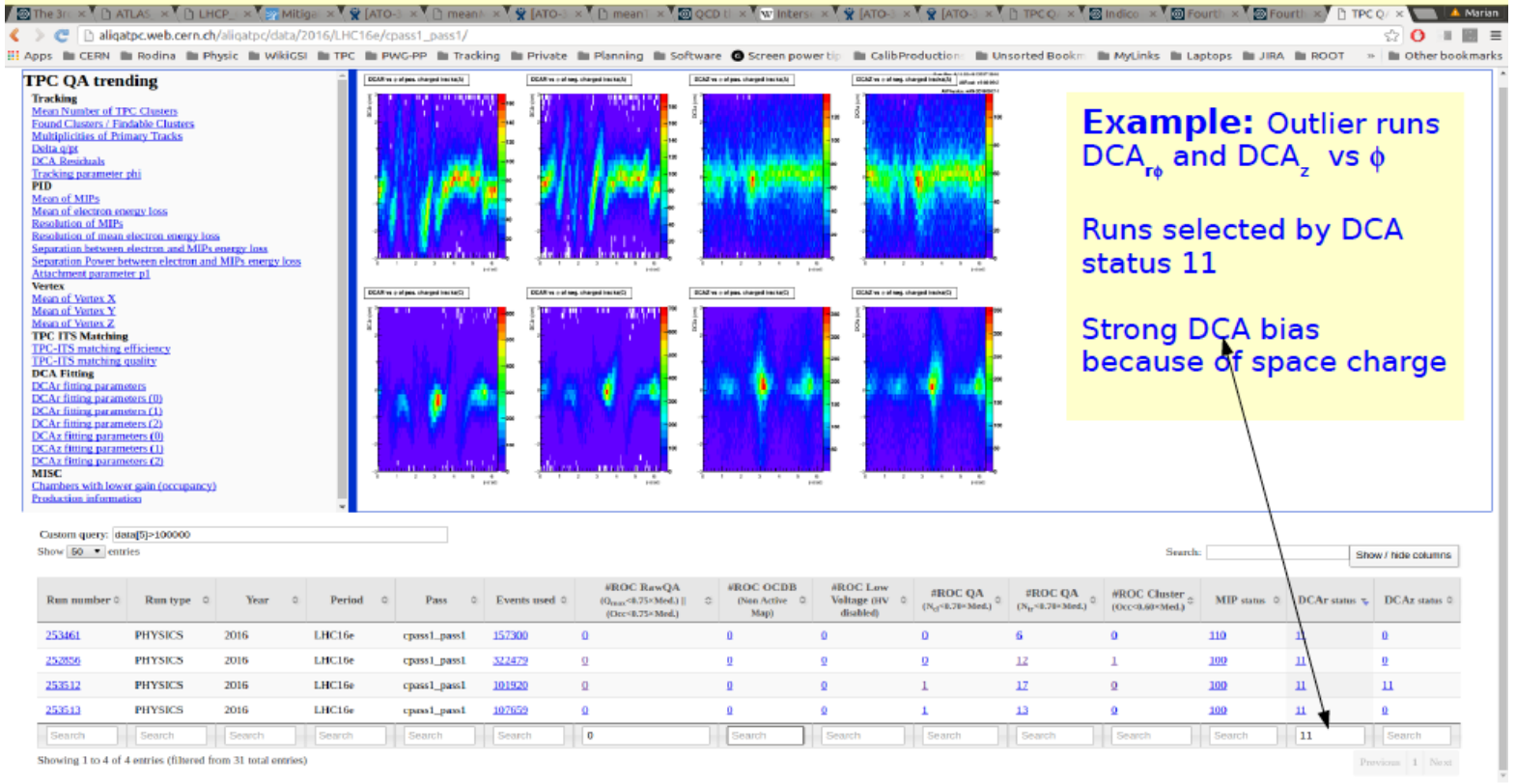
- Must store all correlated variables in one JSON file
- Good for the QA based on trending information



Example: TPC QA WEB interface

Outliers browsing example:

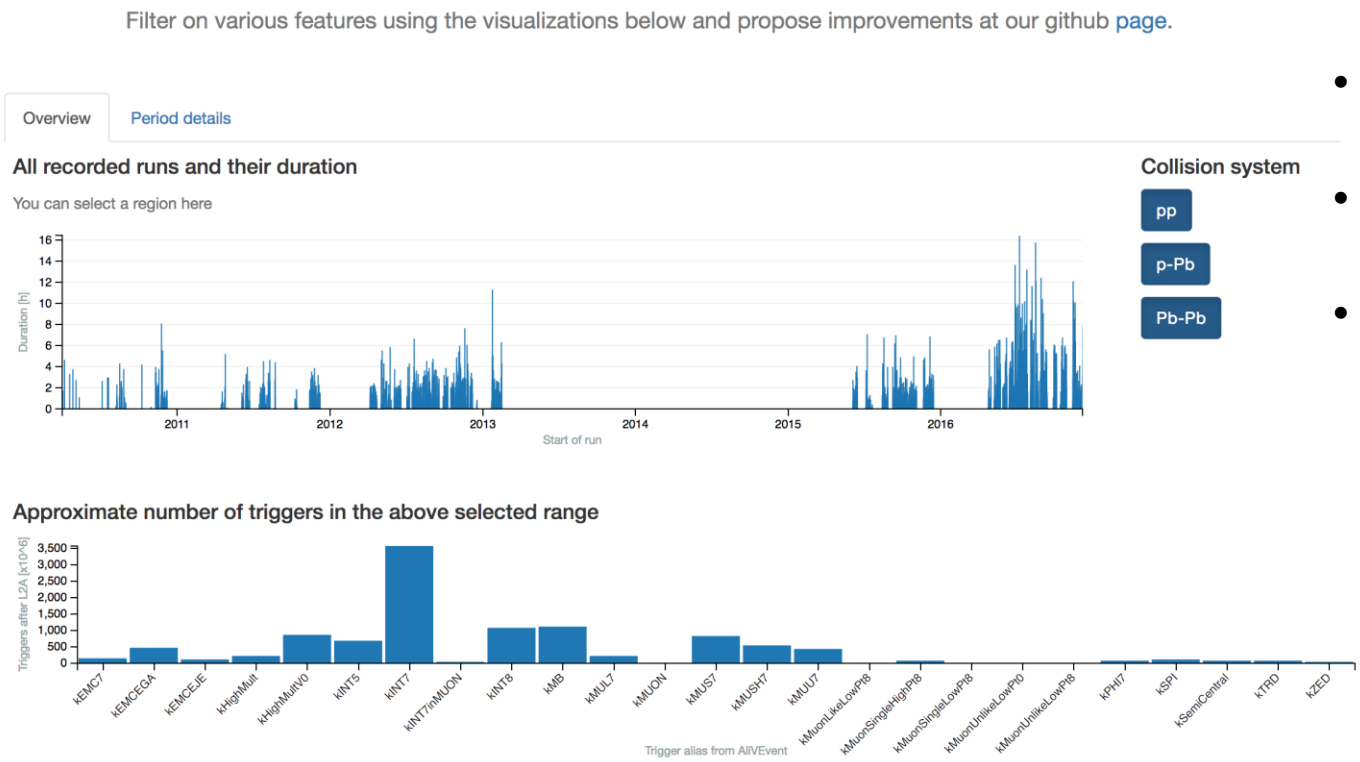
- Based on DataTables plug-in for the jQuery Javascript
  - Histograms organized into layouts to bundle related information
  - Status flag: defined by user defined logical expression (TTree query) using summary information (absolute bands, nσ bands, &&,||)



Marian Ivanov, Boris Rumantsev

Example: Dashboards

https://github.com/cbourjau/alice-dashboard
https://cbourjau.github.io/alice-dashboard/index.html



- JOIN sources on run number (SQLite DB)
- Convert root trees to csv files
- Vizualisation based on Dimensional Charting Javascript Library (dc.js)

Christian Bourjau

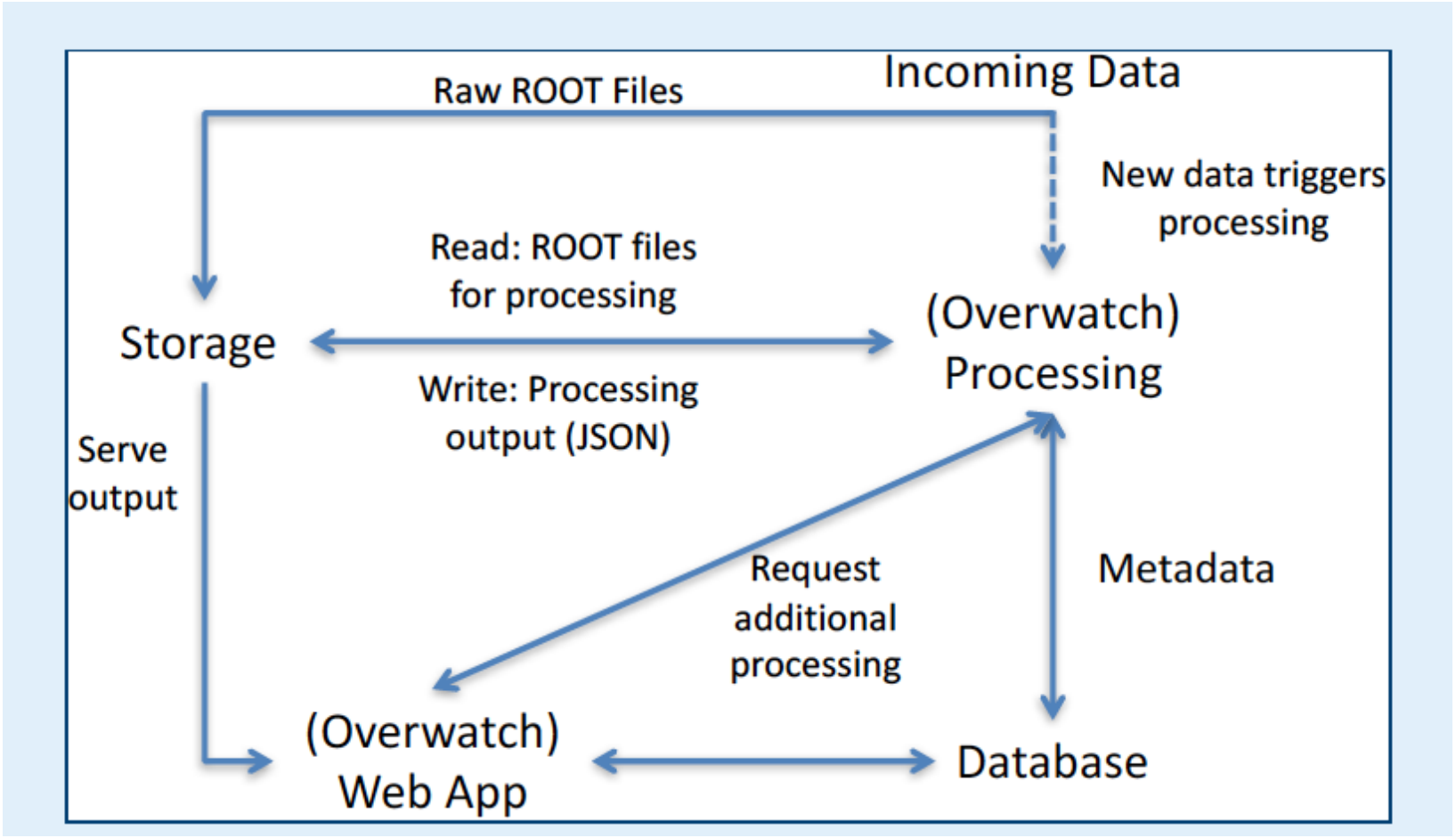
OVERWATCH - Online QC in Run2

Raymond Ehlers, Markus Fasel, Sarah LaPointe
Rafał Pachołek, Maciej Malawski (AGH, IT)

- Online processing and interface for online detector monitoring and basic QA using data from the HLT (originally implemented for EMCAL)
  - https://aliceoverwatch.physics.yale.edu/monitoring
- Architecture similar to data processing in Run3 (parallel processing -> merging -> QA -> visualization)
- Allows slicing data in time windows
- Based around two main components
  - Processing based on PyRoot
  - WebApp based on Flask

https://indico.cern.ch/event/662424

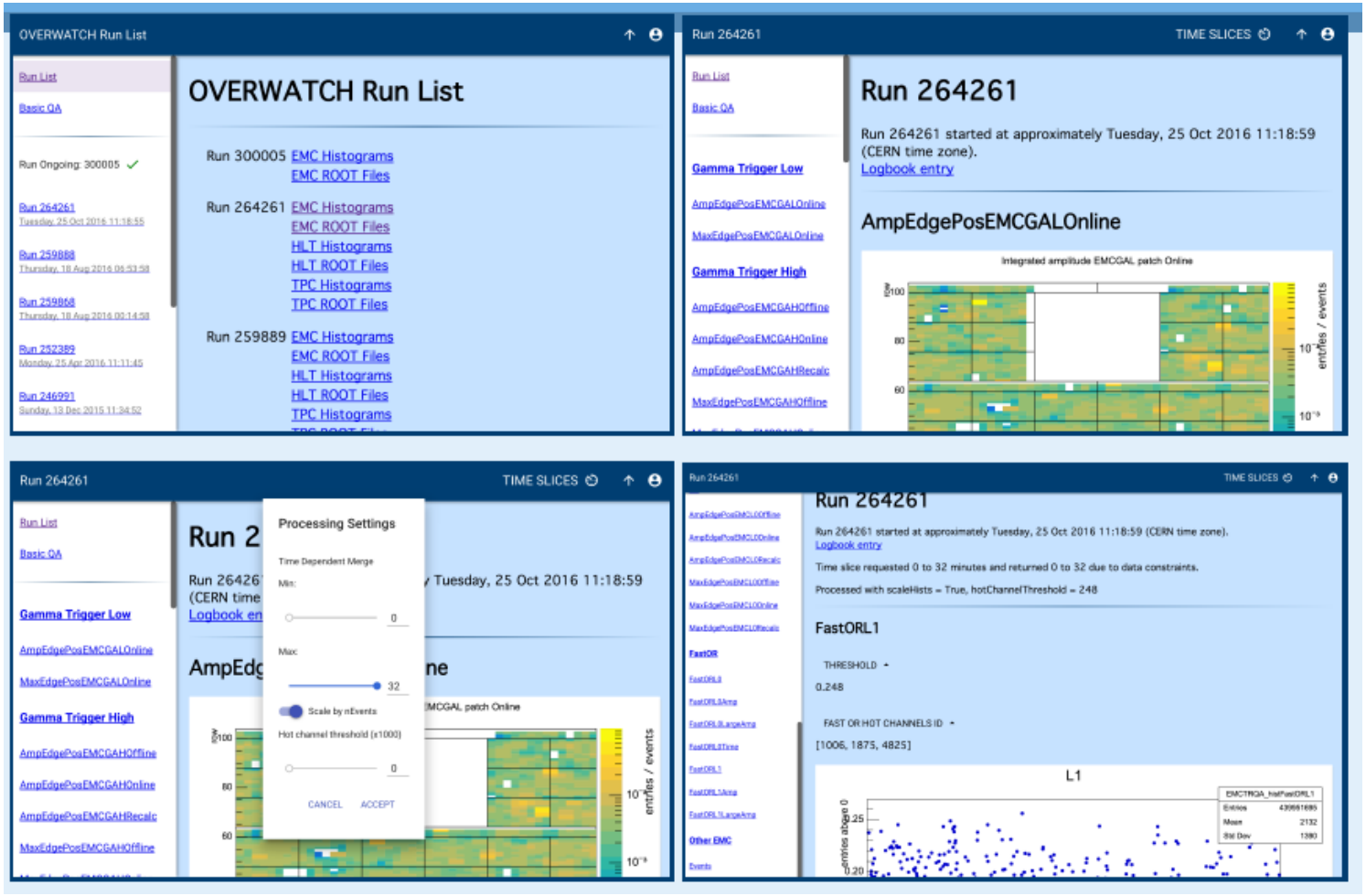
OVERWATCH - Current Architecture



Raymond Ehlers
https://indico.cern.ch/event/671097

- Raw QC objects on EOS
- Metadata in ZODB (a native object database for Python)

OVERWATCH - Example



Raymond Ehlers

https://indico.cern.ch/event/671097

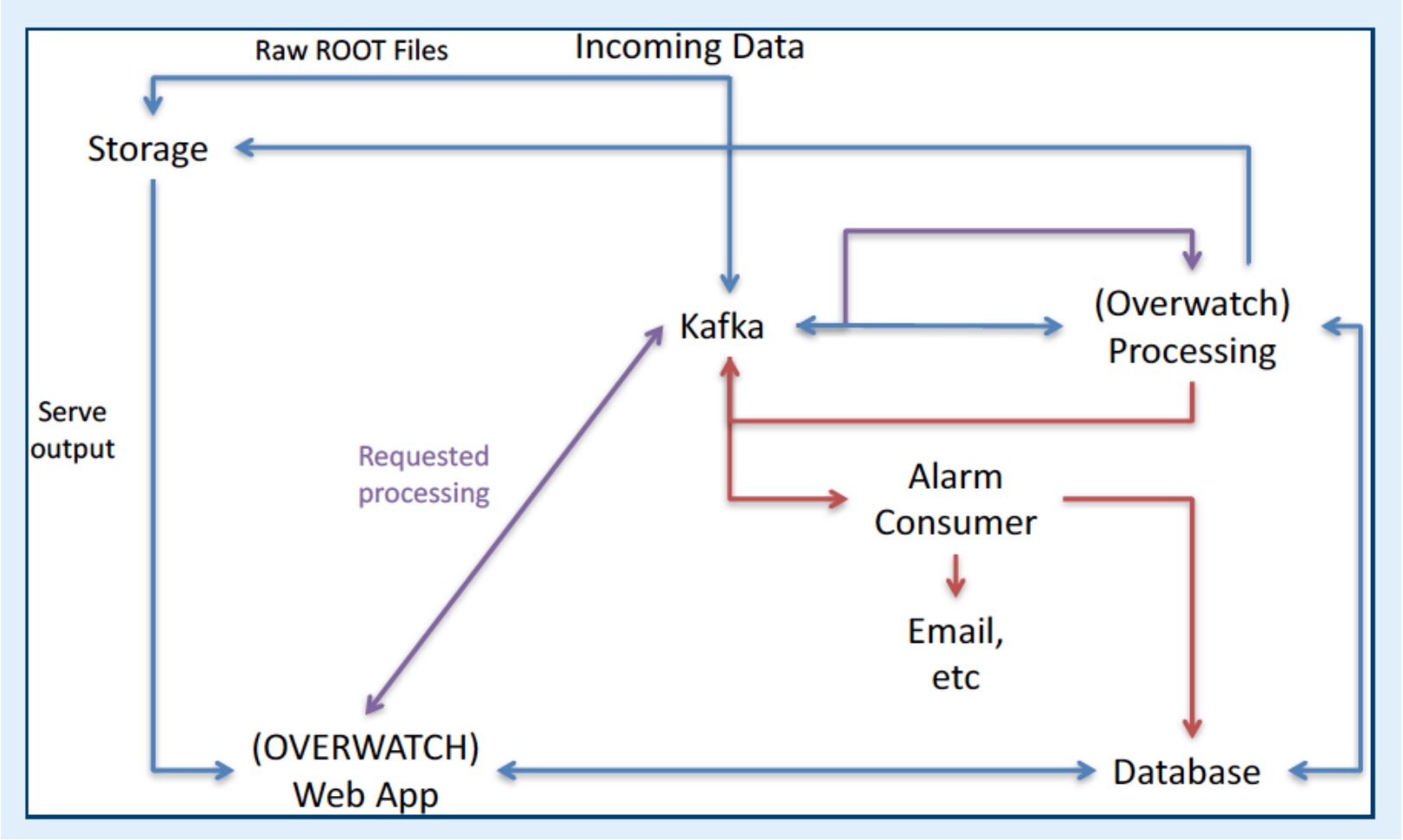
Very simple GUI at the moment

OVERWATCH - Implementation status

- Deployment of the Overwatch upgrade (partially completed)
  - https://github.com/raymondEhlers/OVERWATCH
- Triggering alarm system (implementation started)
  - Based on trending information
  - Following the TPC offline QA
- Consider to replace database ZODB with other DB (to be implemented)
  - ZODB stability concerns
  - Elasticsearch or Casandra considered
- Data aggregation and enrichment with Apache Kafka (to be implemented)

https://indico.cern.ch/event/662424

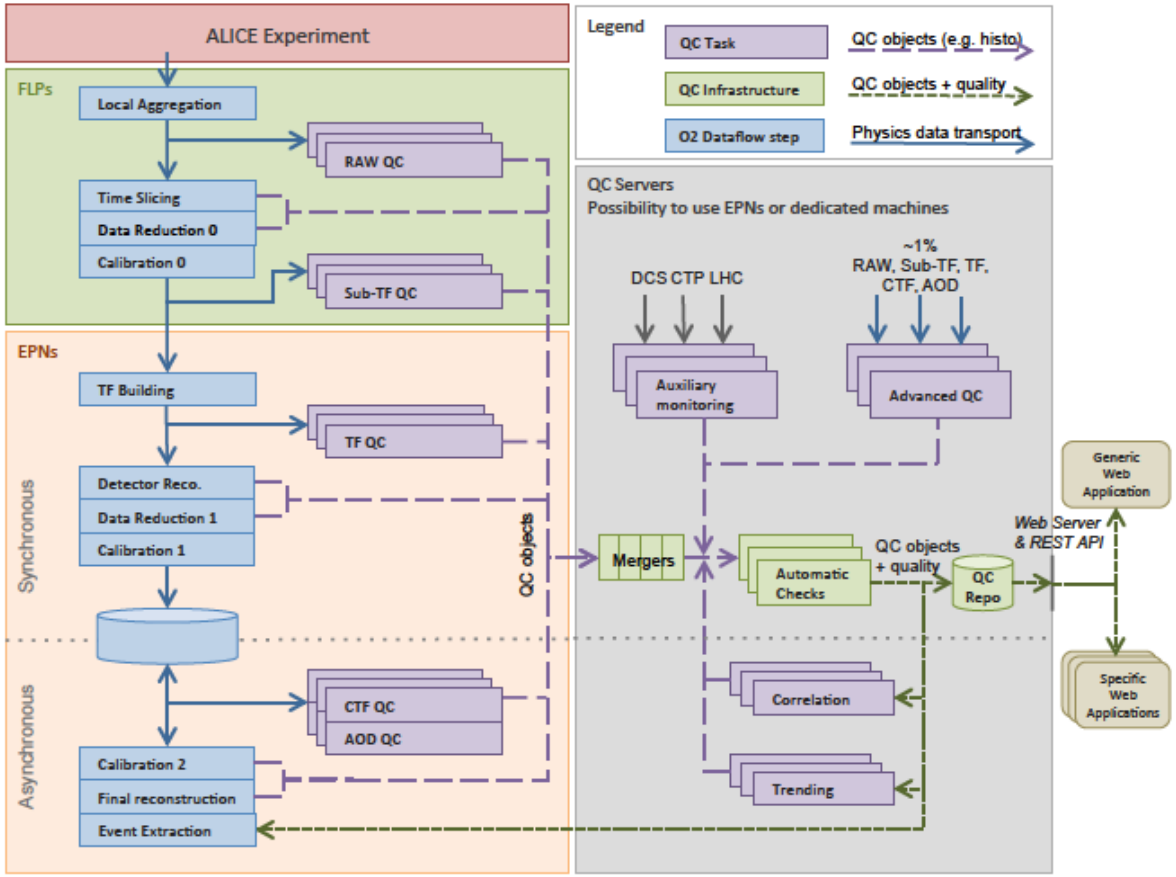
OVERWATCH and Apache Kafka



Raymond Ehlers
https://indico.cern.ch/event/671097

https://kafka.apache.org

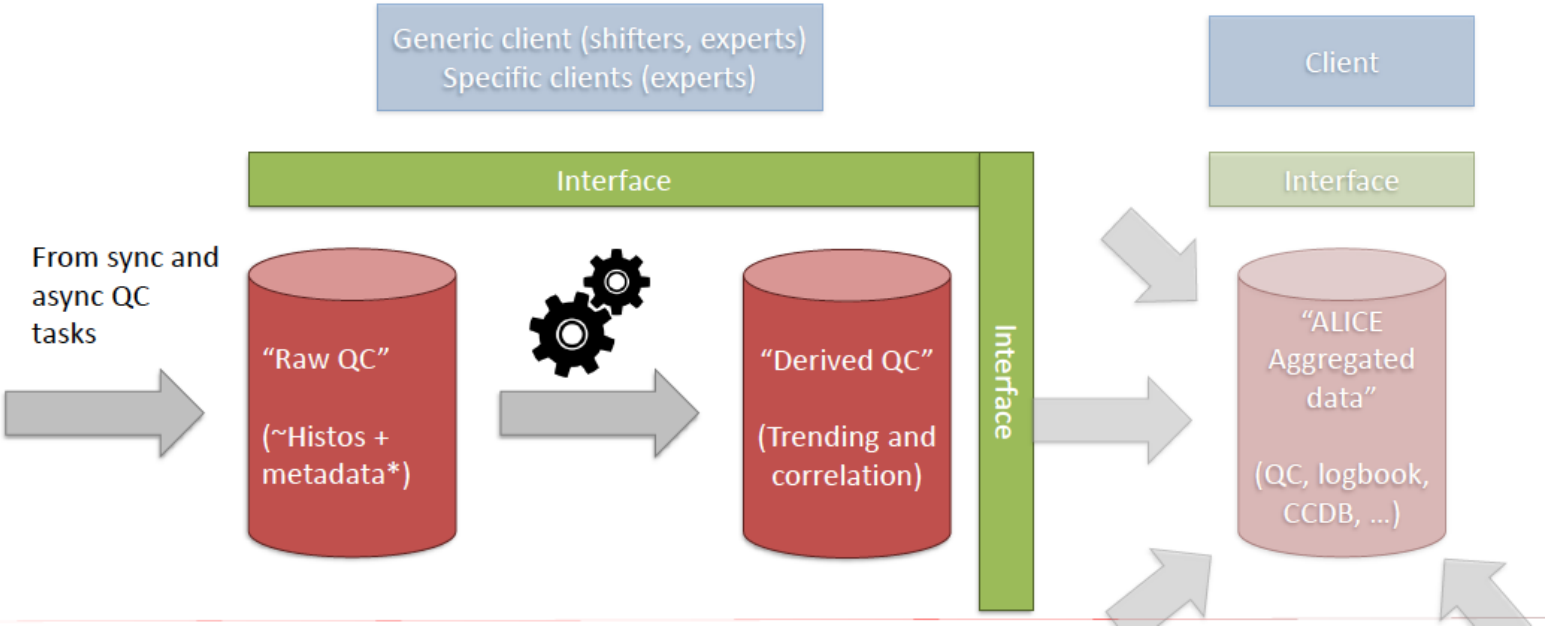
QC in Run3



- Data aggregation and enrichment: via databases or Kafka like system?
- Data repositories: CCDB and/or noSQL/SQL databases?
- GUI / Visualization / Dashboards
  - Any general solution?

QC repositories in Run3

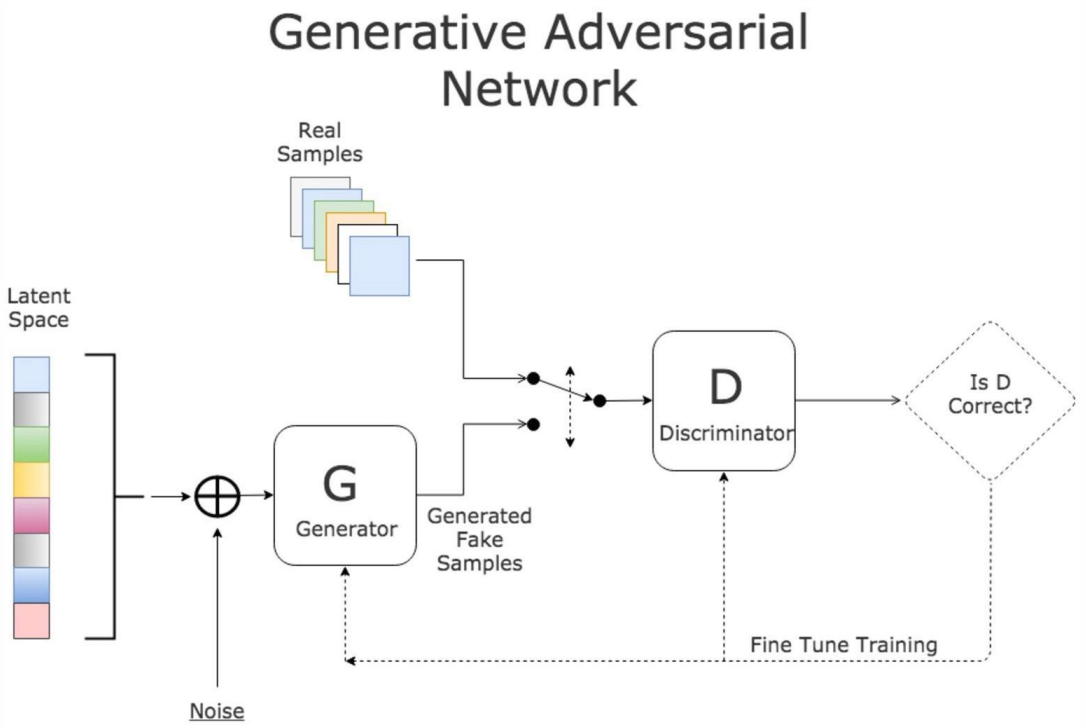
https://indico.cern.ch/event/649759



Barth von Haller

- RAW QC objects in CCDB (or MySQL or others …)
- Derived QC objects in DB (Casandra or others…)
- One database e.g. CCDB or Cassandra for both RAW and derived QC is also considered
- Prototype backend class to use the CCDB in the QC with curl – Barth von Haller
  - https://alice.its.cern.ch/jira/browse/QC-53

Machine Learning and QC



Tomasz Trzciński, Piotr Deja
(WUT, IT)

https://indico.cern.ch/event/676027

- Search for anomalies: real QC data comparison with GAN-generated data points
  - Work started on the HLT QA data (DQM and TPC QA)
- Other applications:
  - Fast MC detector simulations
  - Physics analyses

Outlook

- Several tools implemented and tested for offline QA
  - ROOT files + Elasticsearch + Swan
  - Data sets comparison based on trending information (TPC QA)
- Data aggregation and enrichment
  - Based on ROOT files (TPC QA)
  - Other solutions e.g. Kafka or via databases…?
  - Good solution needed for triggering online alarms
- Data repositories
  - Raw QC stored in CCDB or MySQL …?
  - Derived QC in Elasticsearch or Casandra…?
    - Do we need differrent DB for RAW and derived QC?
- GUI / Visualisation / Dashboards
  - Several prototypes: TPC QA WEB interface, SWAN/Jupyter, Dashboards based on dc.js
  - Several use cases
    - Do we need a general solution?
- ML/GANs for the QC started