Data Quality Monitoring

DAQ@LHC workshop

Introduction

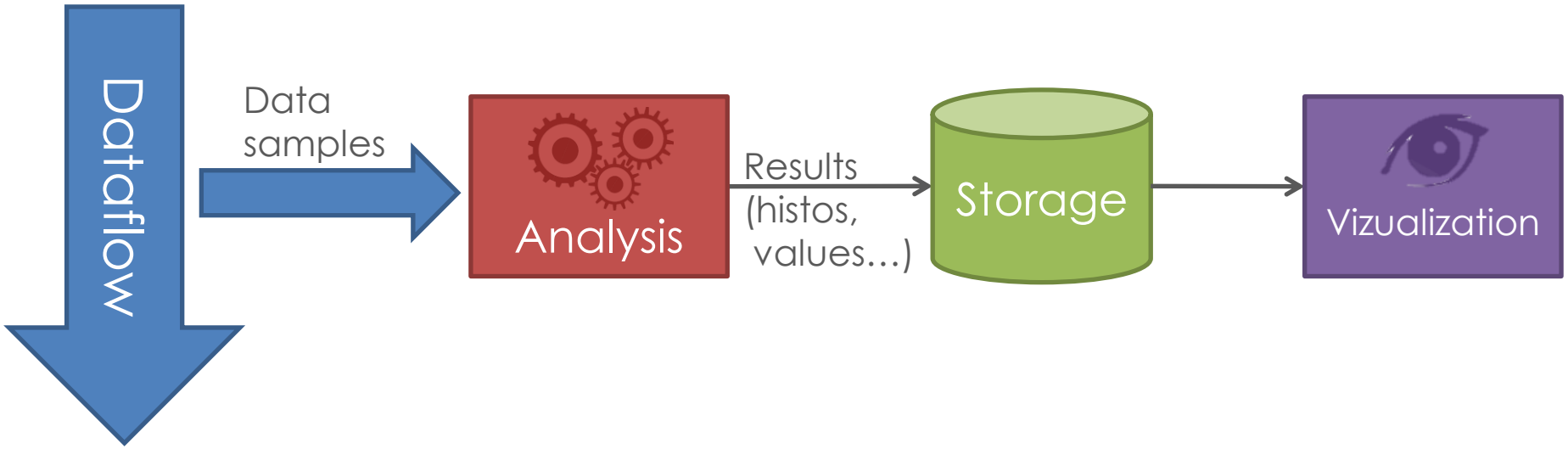
- What this presentation is not
- What it is and how it is organized
  - Definition of DQM
  - Overview of systems and frameworks
  - Specific chosen aspects
    - Data collection, Storage, Visualization, analyses
  - Operations
  - Qualitative assessment & discussion
  - Future

Data Quality Monitoring (1)

- Online feedback on the quality of data
- Make sure to take and record high quality data
- Use the data taking time and the precious bandwidth in an optimal way
- Identify and solve problem(s) early

Data Quality Monitoring (2)

- Data Quality Monitoring (DQM) involves
  - Online gathering of data
  - Analysis by user-defined algorithm(s)
  - Storage of monitoring results
  - Visualization



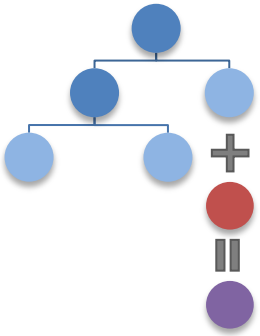
Data Quality Monitoring (3)

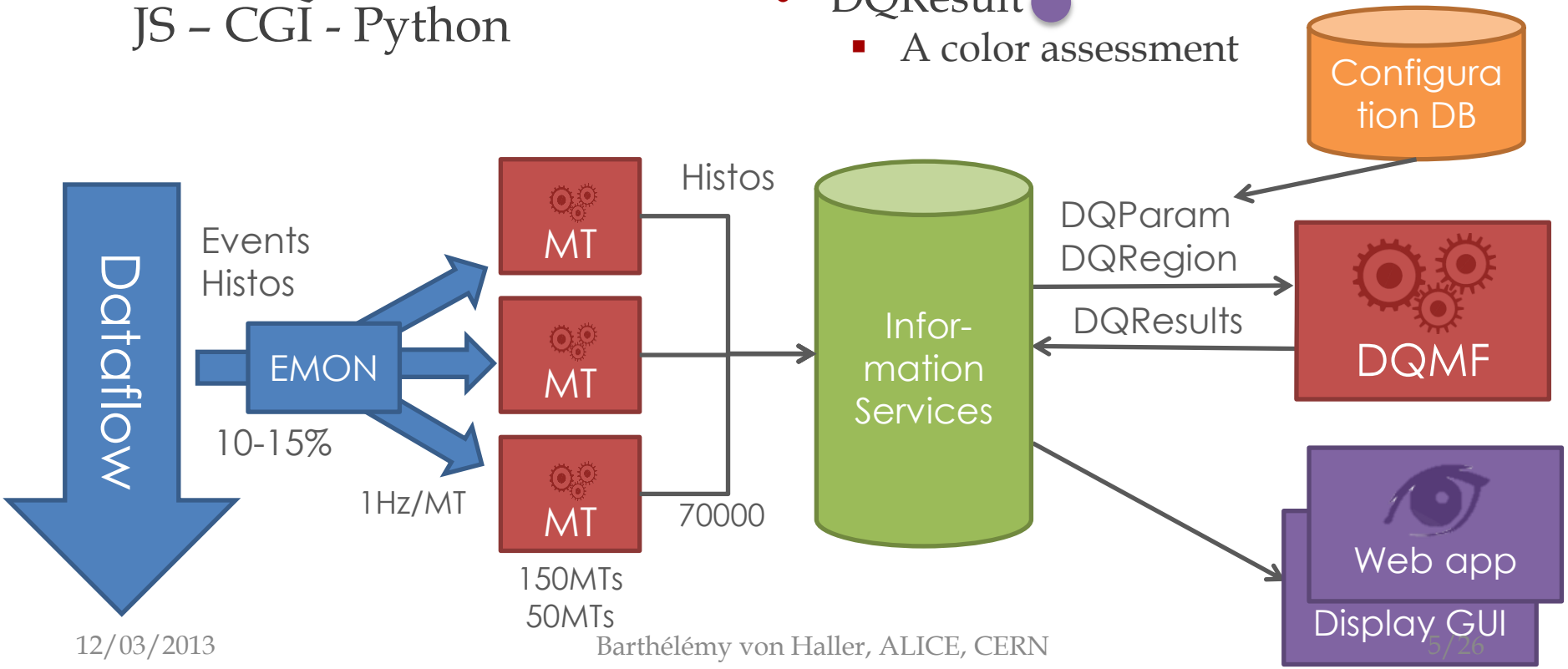
- Requirements
  - Range from formal documents to oral tradition
  - Changing → difficulties
  - No interference with data taking
  - Low latency update during runs
  - Centralized results
  - Modular, fast update of users code

Overview: ATLAS

- « Generic, yet flexible and nicely scalable »
- Shared online/offline and by ~10 independent communities
- C++ – Qt – ROOT – XML – JS – CGI - Python

- DQParameter
  - An histo
- DQRegion
  - A subset of detector
- DQAlgorithm
  - A specific analysis
- DQResult
  - A color assessment

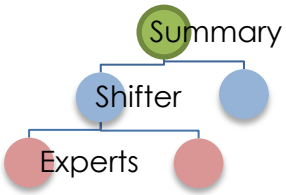


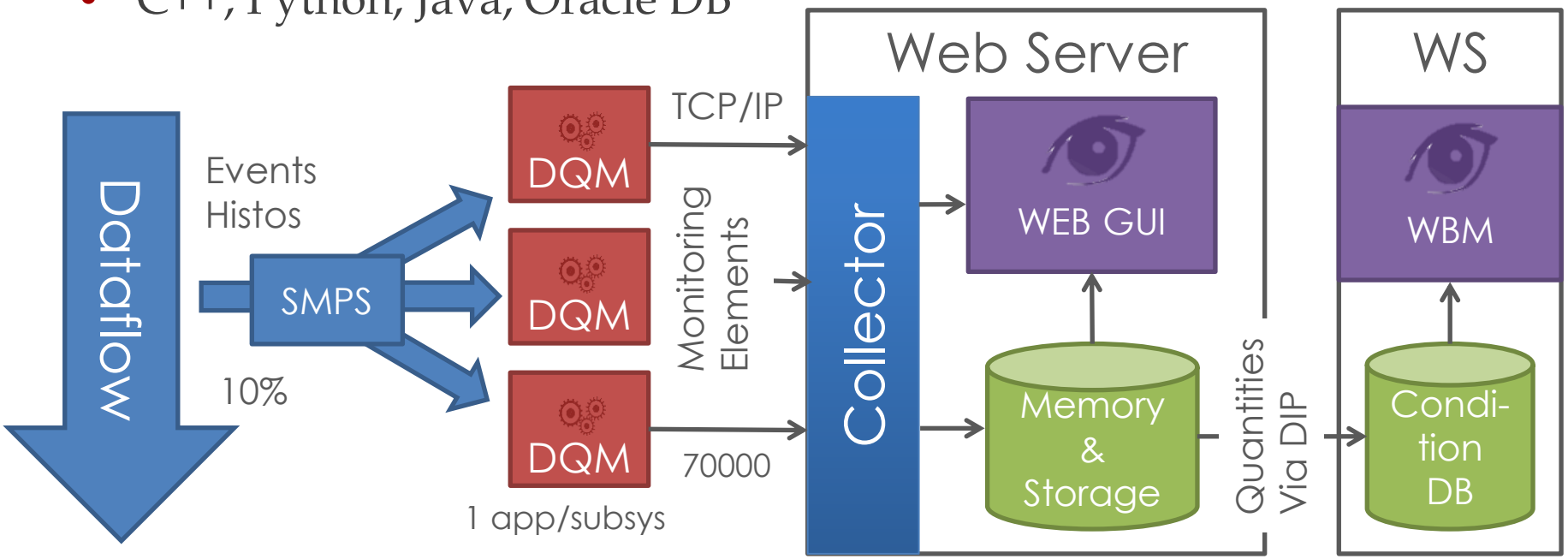


Overview: CMS

- DQM framework part of reconstruction framework
- SMPS allows for trigger path selection
- 1 DQM app per subsystem
  - Re-run reco according to needs
  - Run certain analysis
- C++, Python, Java, Oracle DB

- Monitoring element
  - Metadata
  - Reference histo
  - Quality
- Web based
  - HTTP for data exchange
  - Web UI only

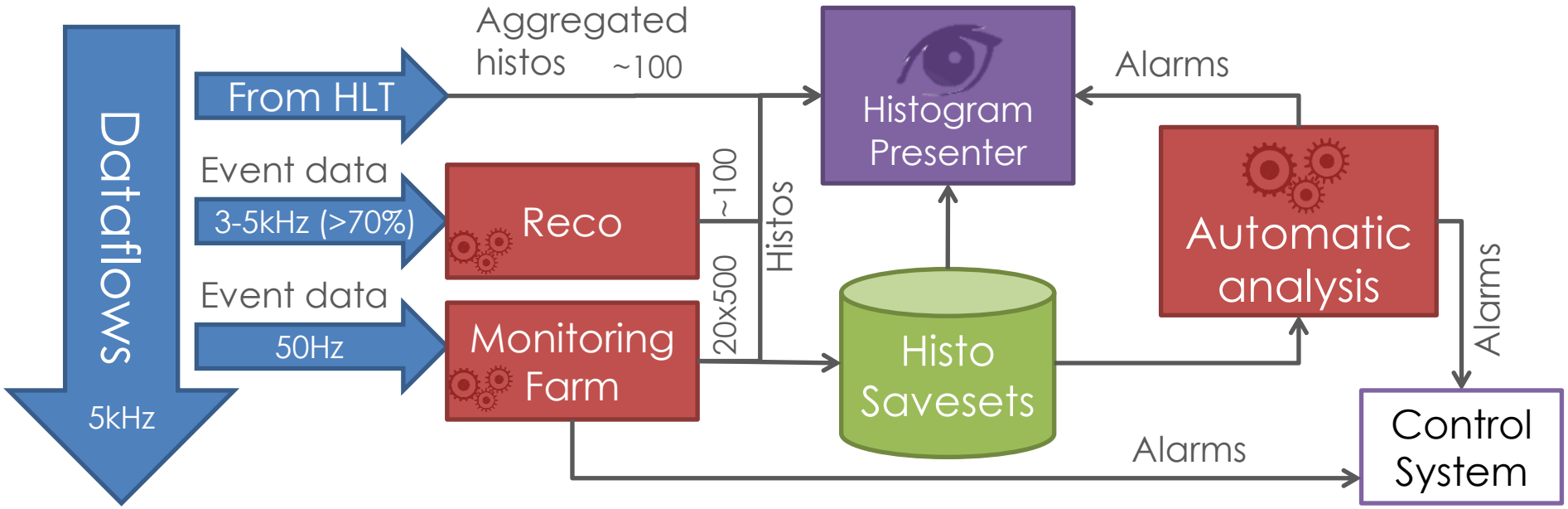




Overview: LHCb

- DIM
  - Network
  - RPC
- 3 distinct histo-gramming places

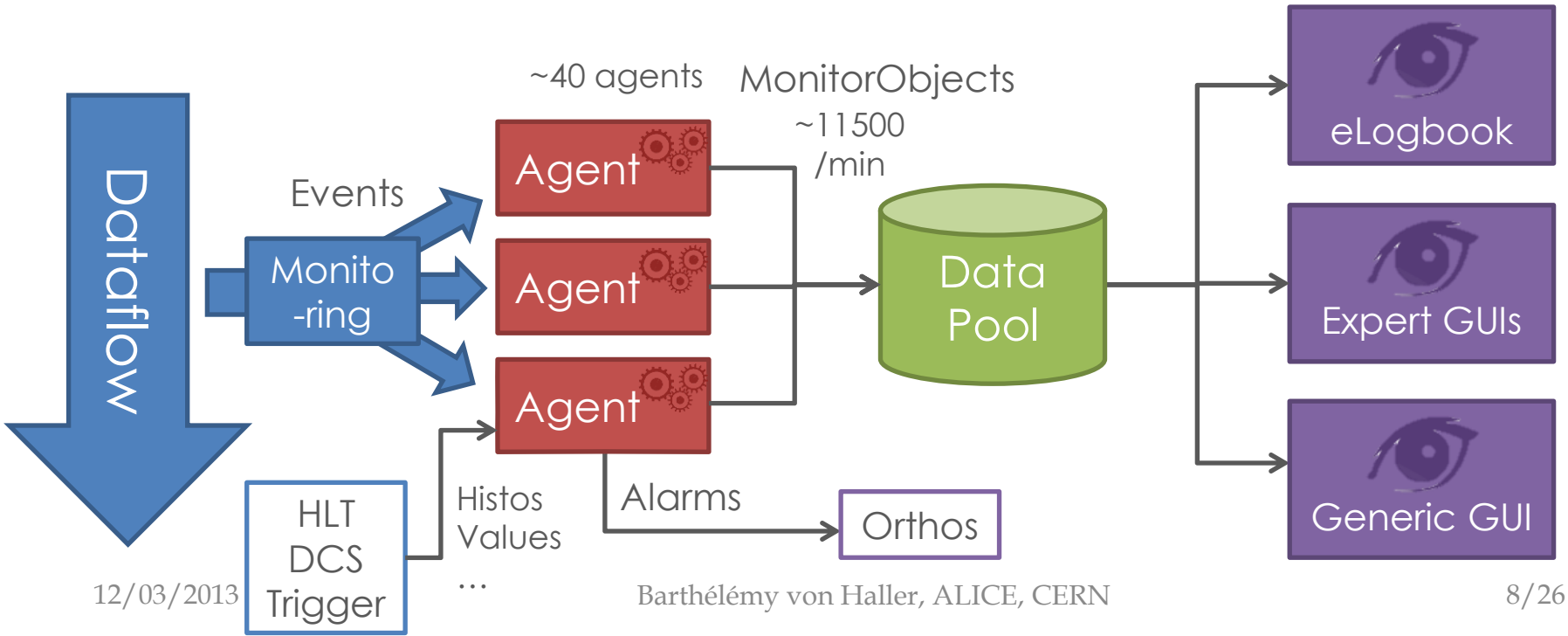
- C++ - ROOT
- Monitoring on 70-100% of data
- Auto. analysis every 15 minutes
- Automatic actions !



Overview: ALICE

- AMORE framework
  - C++ - ROOT
  - Plugin architecture
  - Notifications via DIM
  - MySQL database
- Plugins developed in 20 institutes worldwide

- Histograms or values also retrieved in other systems
- MonitorObject
  - Metadata (run, quality, expert-shifter flag)
  - ROOT base type encapsulated (TH1, TObject, scalar, …)



Data samples collection

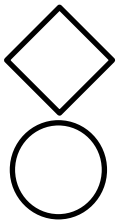
- Main input to DQM
  - Events (all) and sub-events (ALICE, ATLAS)
  - Histograms or values coming from other systems, esp. HLT.
- Challenges
  - Bandwidth issues: monitoring processes do events selection themselves instead of using a proxy (ATLAS, ALICE)
  - Merge of all histograms from HLT (LHCb, CMS)

Data analysis (1)

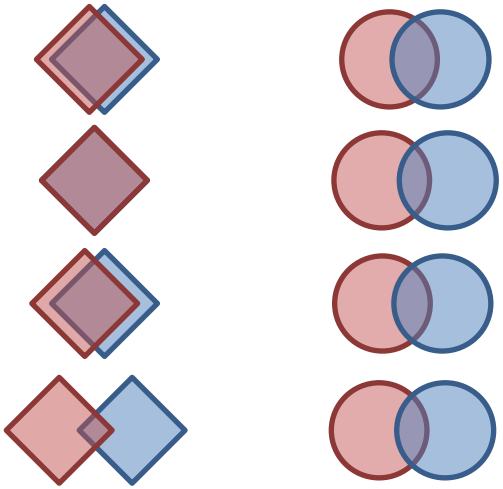
- Most typical analysis include
  - Check for dead channels (holes in distribution)
  - Check for noisy channels (peaks in distribution)
  - Check for differences with a reference
  - Check for non-empty error histograms
- Also
  - Check for infrastructure issues
  - Check for reconstruction issues
  - Check for trigger/dead time issues

Data analysis (2)

- Offline - Online
  - Same framework ?
  - Same analysis ?



- ATLAS
- CMS
- LHCb
- ALICE

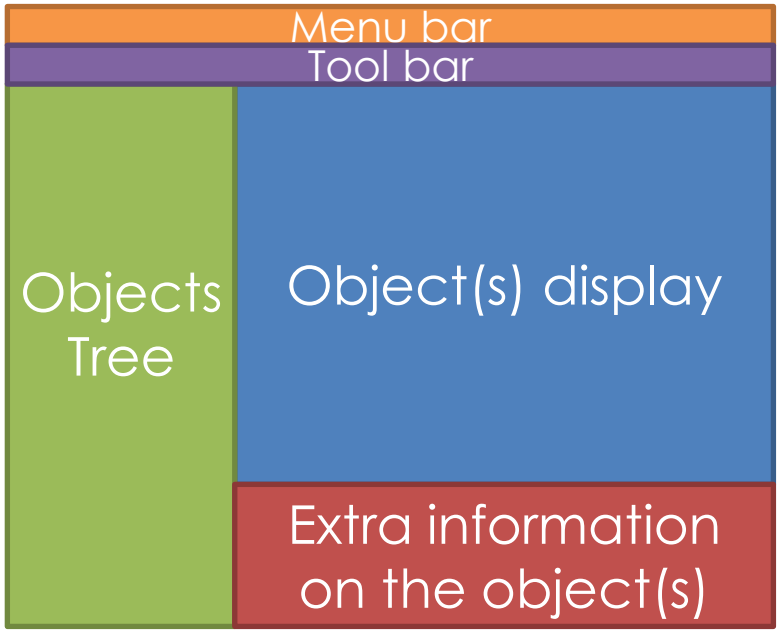


Storage

- Online Storage
  - In-memory (ATLAS, CMS, LHCb)
  - Database (ALICE)
- Online recent history
  - Short-term, detailed (all)
- Archiving
  - Long-term per run/per time interval (all)
- Does long-term archiving make sense ?
  Is it still online DQM ? Rather offline QA ?

Visualization (Desktop)

- Desktop GUI (all but CMS)
  - C++, Qt or ROOT
- Very similar display showing online
- Also past objects (LHCb)
- Summary panel (ATLAS)
- Easy customization of the layout
- Quality and alarms



Amore Visualization Main Window

Amore Agents Objects Help

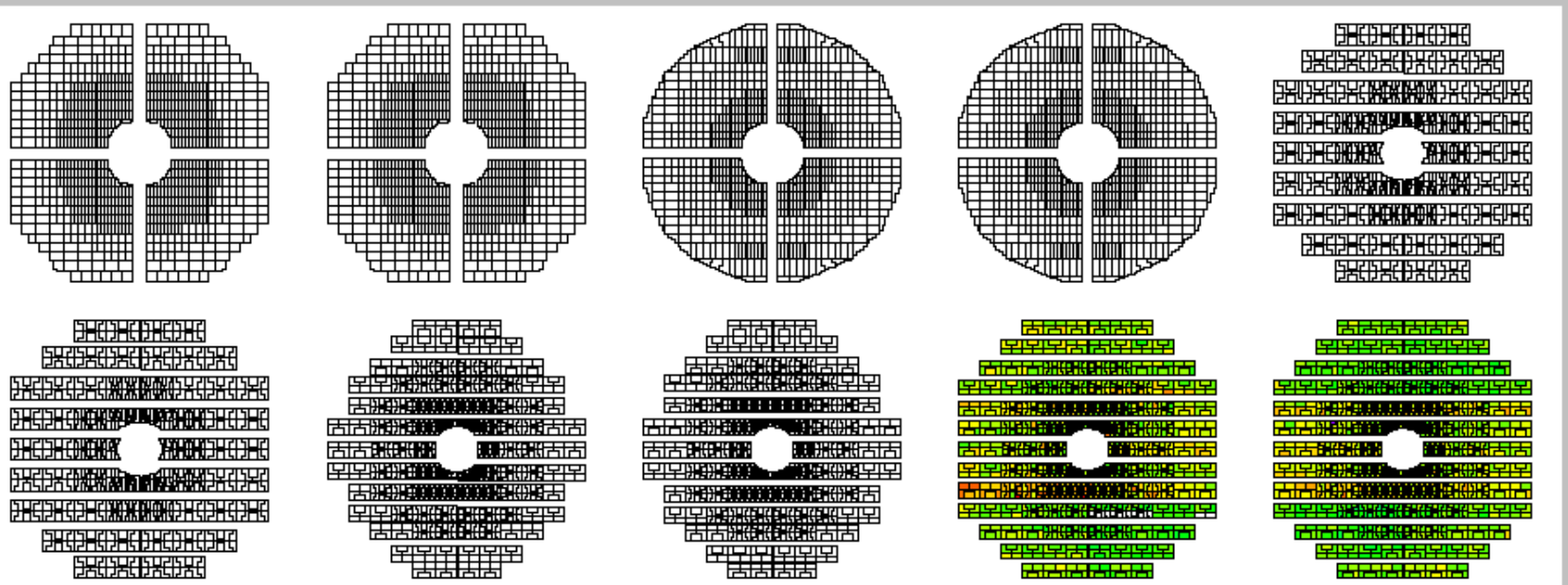
ALICE's AMORE is batigne@nanpca9.in2p3.fr

File History Help

MCH Status | Error Report | Data Display

Cathode: 0 1 | Plane: B NB | ViewPoint: Front Back

Tracker-B-Front



Responder: Chamber, DE, BUSPATCH, MANU, PAD, PCB

Outline: Chamber, DE, BUSPATCH, MANU, PAD, PCB

Plot: Chamber, DE, BUSPATCH, MANU, PAD, PCB

Sources: none, MCHPed, MCHGain, MCHRun

MCHRun: mean of Calibrated charge, sigma of Calibrated charge, occ, N, n

Init | Start | Stop | Update View Period (sec) 30 | Update

amoreAgentSPDHits

amoreAgentSPDQAshifter

The threshold is set to 100%.
The vertical lines change color according to the quality flag.

Remote data access

- How can experts access DQM results ?
- Remote ssh, interactive access to control room
  - LHCb: open to collaboration
  - ATLAS: provides tokens for a few hours access + replica
  - CMS: closed but has a collaboration replica
  - ALICE: closed
- Web GUI
  - Images and ROOT objects download (ALICE)
  - Generated static set of pages (ATLAS)
  - Full-fledged web application (CMS)

Visualization (Web, ALICE)

Run Details - 104864

104864 | Quick Access | Print tab | Print all

Run Conditions | Run Statistics | Trigger Clusters Info | Run Quality | LDCs Statistics | GDCs Statistics | Shuttle Info | File Info | Log Entries | InfoLogger Messages | DQM

Data Quality Monitoring Info - agent 'amoreAgentTRD03'

Overview | Permanently Archived MOs (0) | Temporarily Archived MOs (147) | Online MOs (147)

AMORE Agent 'amoreAgentTRD03' Overview for run 104864

General

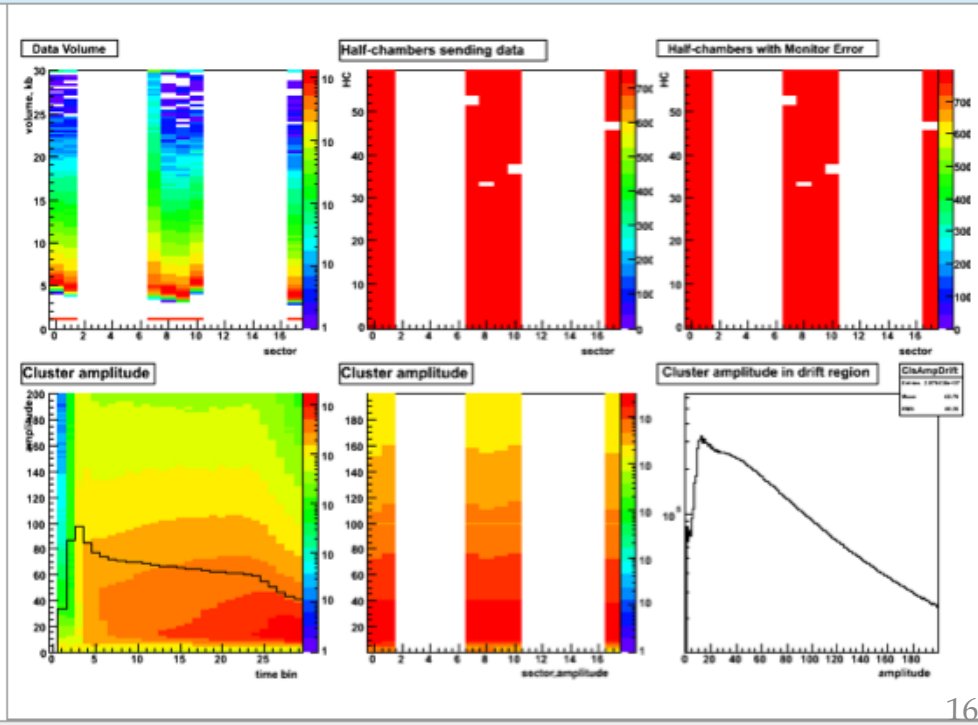
Detector: TRD
Version: 1.6
 # Monitor Objects: 147
 # Versions: 21462
Total Size: 419.3 MB
Last Updated: 12/12/2009 15:19:12

Runtime Parameters

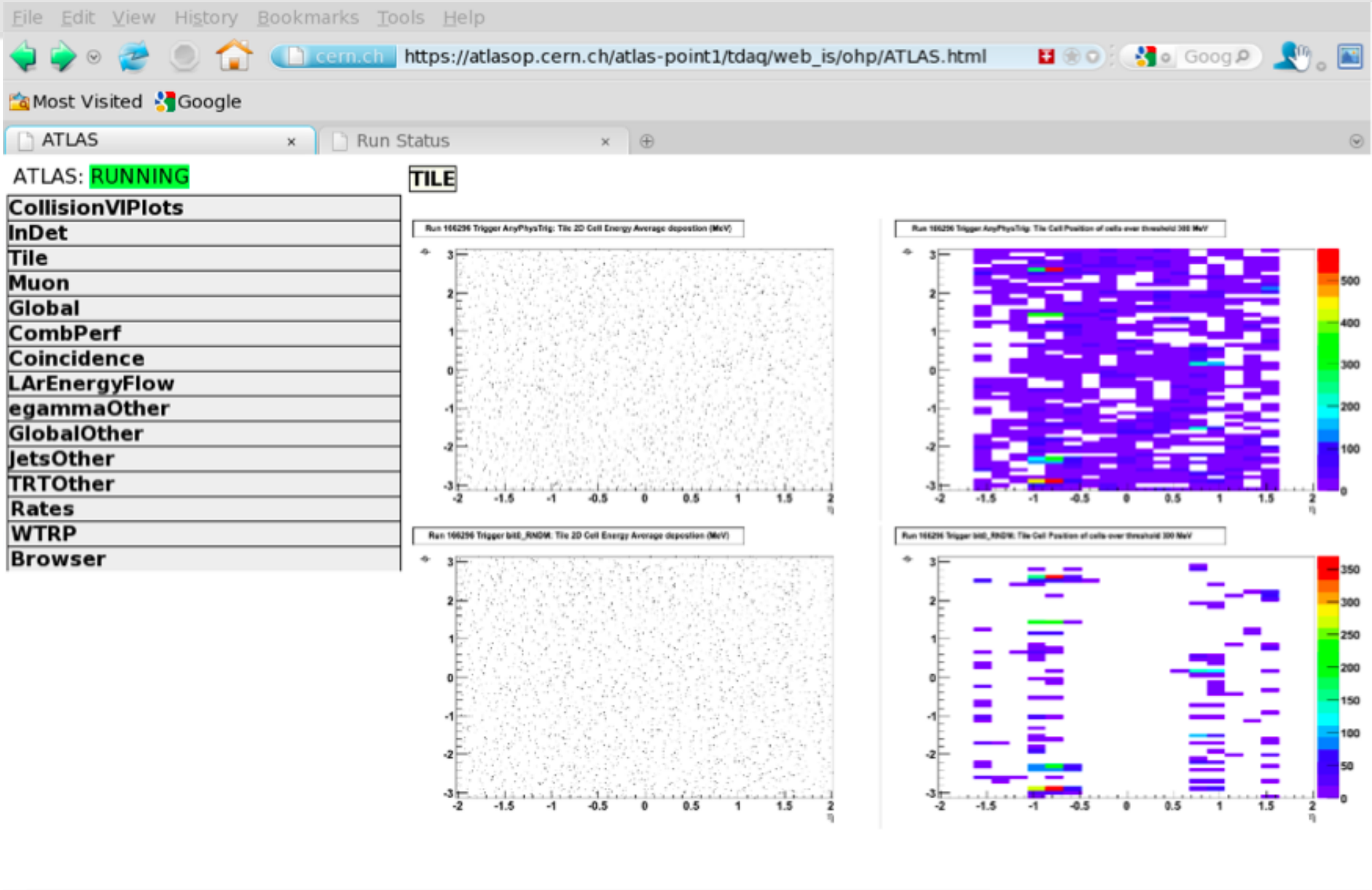
```
amoreAgent -u -a amoreAgentTRD03 -e50 -s =PHYSICS_1
Configuration file :
```

Monitoring Objects

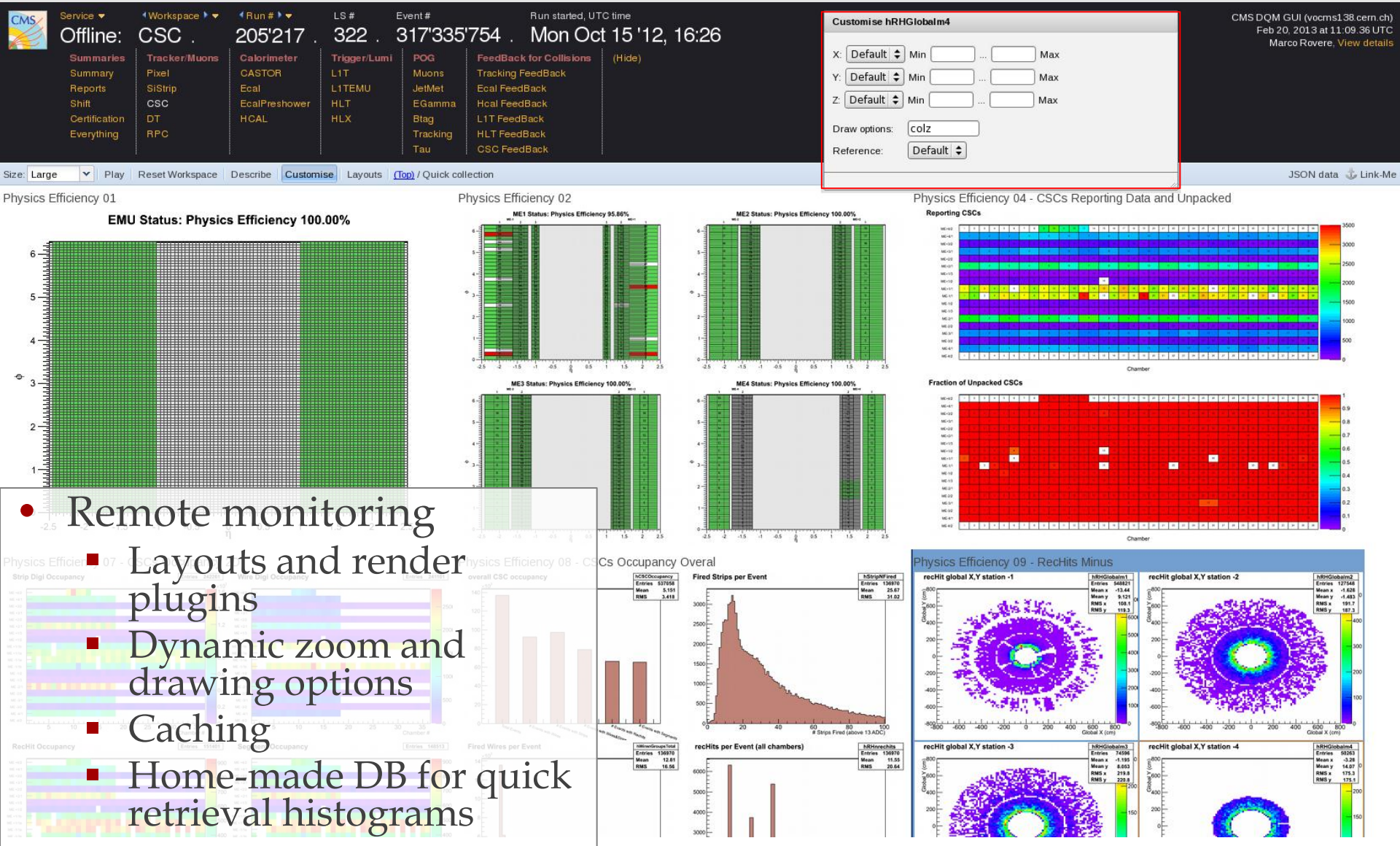
MOs Permanently Archived: 0 (0.0 KB)
MOs Temporarily Archived: 0 (0.0 KB)
MOs Online: 0 (0.0 KB)
First Object From: 30/01/2010 11:10:23
Last Object From: 30/01/2010 11:10:23



Visualization (Web, ATLAS)



Visualization (Web, CMS)

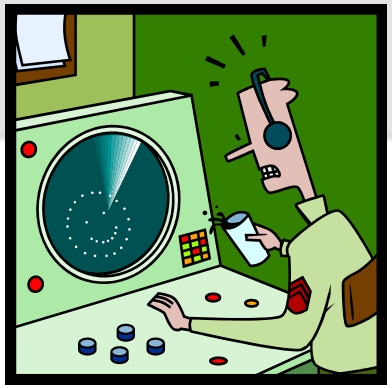


- Remote monitoring
  - Layouts and render plugins
  - Dynamic zoom and drawing options
  - Caching
  - Home-made DB for quick retrieval histograms

Operations (1)

- A typical DQM shifter
  - Check alarms
  - Go through plots layouts (LHCb, ALICE, ATLAS)
  - Find information in a Wiki or directly the GUI
  - Contact experts / shifters (if present)
  - Inform the shift leader

Operations (2)



- Is it worth it ? Yes !
  - ATLAS
    - Transition Radiation Tracker timing problem spotted
  - CMS
    - Wrong DAQ configuration of SST detector used during beam "ramp" → audio alarm triggered by DQM → fix before stable beam was declared.
  - LHCb
    - Bad mass plot -> anomaly -> inverse value of the magnetic field used by the HLT
  - ALICE
    - HLT started compressing the TPC laser events → recons-truction problems → DQM shifter spotted anomalies →fix

Qualitative review (1)

- **Personal** hit-list
  - ATLAS: scalability, service-oriented
  - CMS: interactive and dynamic web interface
  - LHCb: automatic actions
  - ALICE: versatility
  - General:
    - Flexibility and reusability

Qualitative review (2)

- Biggest challenges
  - ATLAS: initial display was in Java and couldn't cope with the number of histograms (70k)
  - CMS: handling huge quantities of histograms in the web gui and yet be responsive
  - LHCb: nothing dramatic, summing up HLT histograms with data rates of the order of a LEP experiment
  - ALICE: ensure stability despite of the multiple contributions

Qualitative review (3)

- Could we have used the same system ?
  - 4 efficient and scalable DQM
  - Similar architecture and technologies
  - I believe the answer is: Yes…
  - … up to a certain extent

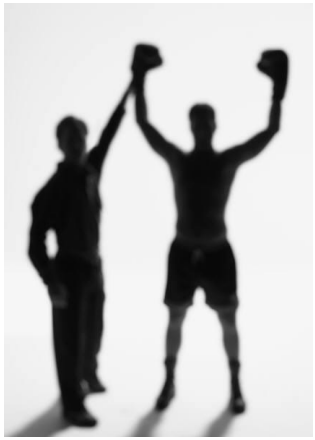
- If not the framework, maybe have the same (Web) UI ?
- Could we at least share more ?
  - (Regular) common meetings ?

Future

- LS1
  - Algorithm and thresholds adjusting on run conditions (*ALICE, ATLAS*)
  - Adapt SW to dependencies changes (*LHCb, CMS*)
  - Adapt framework for multi-core multi-thread environment (*CMS, ALICE*)
  - Web GUI and tools (*ALICE, [LHCb]*)
  - Full archiving (*ATLAS*)
  - Proxy monitoring (*ALICE*)
- LS2
  - Complete rewrite along with online-offline framework (*ALICE*)

Conclusion

- Much similarities
  - General LHC DQM framework ?
  - Trans-experiments DQM meetings ?
- DQM in all 4 experiments
  - Work well and up to the demand
  - Proved to be of great importance
- Future
  - DQM is crucial, still often considered late
  - Today is an opportunity to attack it early !



A big thanks to:

Markus Frank

Clara Gaspar

Serguei Kolos

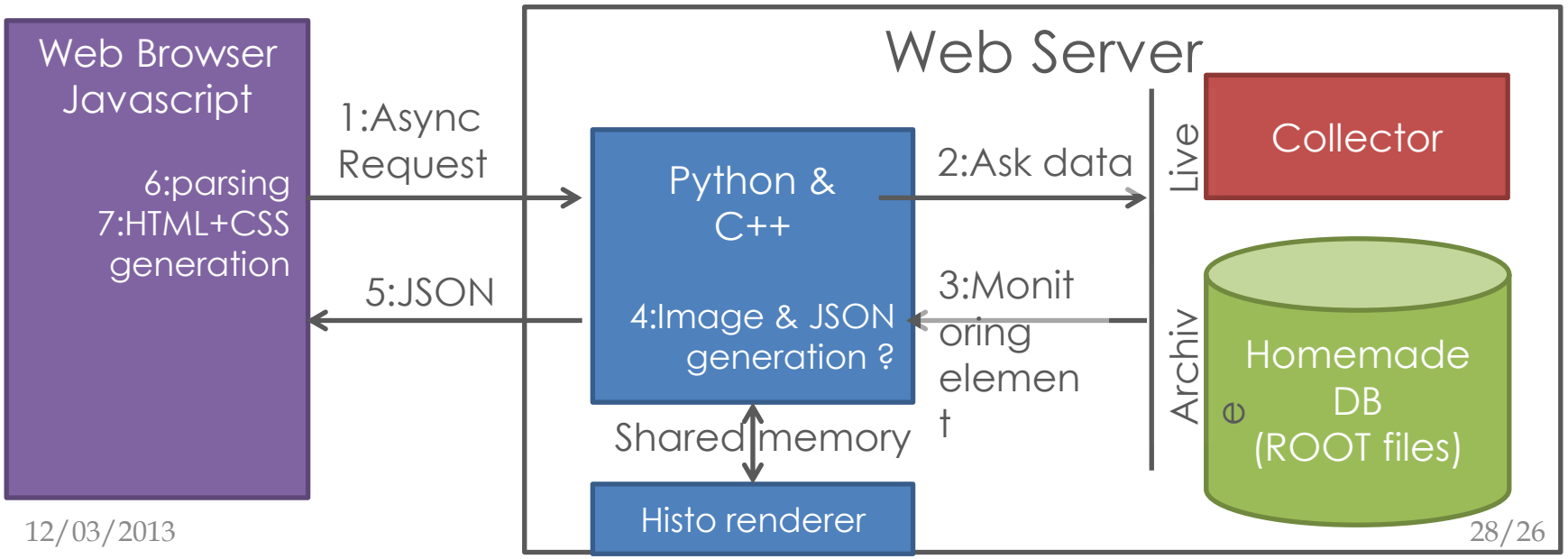
Marco Rovere

Questions and discussion

Backup

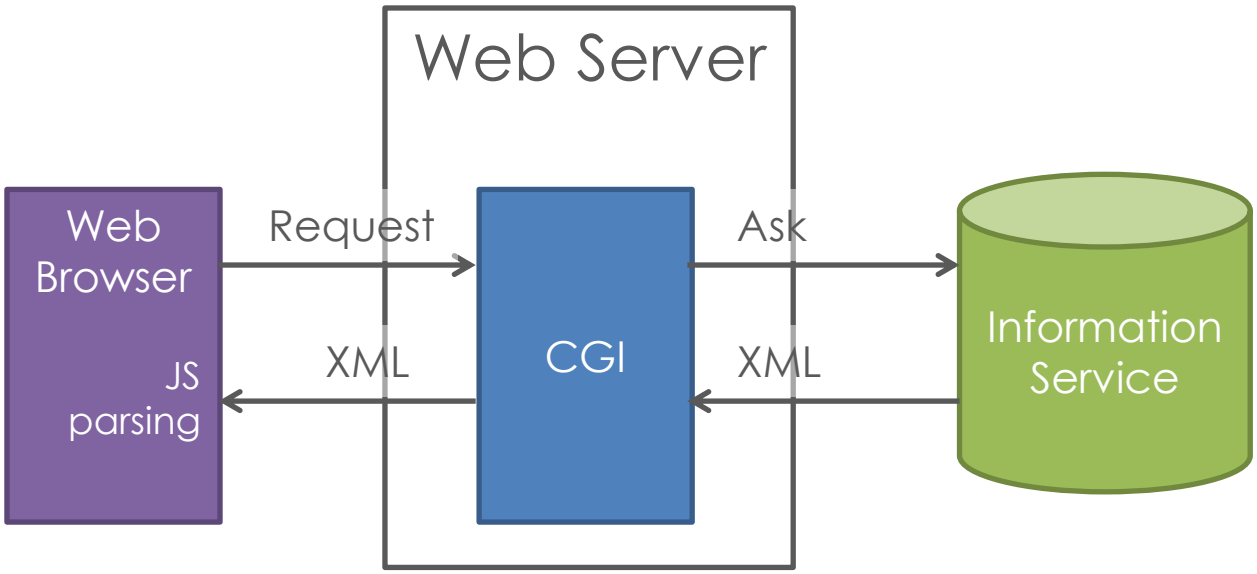
Visualization (Web, CMS)

- Remote monitoring
  - Layouts and render plugins
  - Dynamic zoom and drawing options
  - Caching



Visualization (Web, ATLAS)

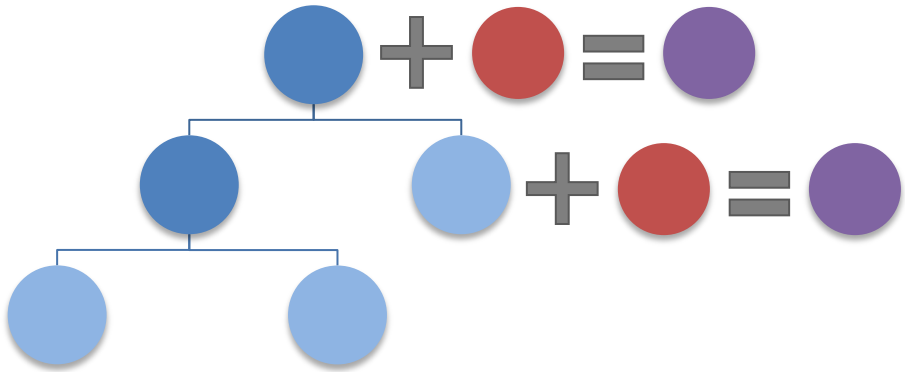
ATLAS



Overview ATLAS

- Services approach
- C++ – Qt – ROOT – XML – JS
- Framework DQMF based on DQMCore

- DQParameter
  - An histo
- DQRegion
  - A subset of detector
- DQAlgorithm
  - A specific analysis
- DQResult
  - A color assessment
  - [New histo]



Overview: ATLAS