ATLAS Data Quality Offline Monitoring

Peter Onyisi

THE UNIVERSITY OF CHICAGO

CHEP 2009, Prague

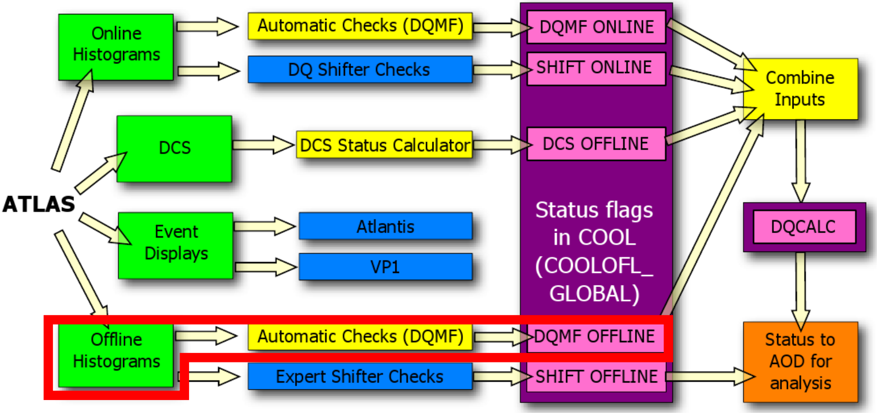
24 March 2009

Introduction

- The ATLAS experiment aims to reduce the time between data collection and the start of bulk data processing to 24 hours
  - Calibration, alignment, etc. loops must complete in this time
- Must determine quality of data as quickly as possible
- Also important to monitor time variation, both within a run and between runs

ATLAS has several data quality monitoring systems: focus here on checks performed on reconstructed data offline in near-real-time

Data Quality Monitoring Tools



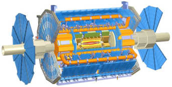
Covering one component of the full architecture here...

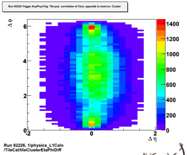
Histogram Production

- Monitoring tools run in the ATLAS software framework Athena
- Histograms can be produced with time granularity ranging from a few minutes to a LHC fill
- Tools are environment-aware
  - Same tool can produce different histograms depending on configuration
- “Trigger-aware”: different histograms for different types of triggers
- Monitor everything from readout errors to high level physics objects

Integration with Tier 0

- Initial prompt reconstruction of data occurs on the Tier 0 farm at CERN
  - Reconstruction as soon as partial data available — long before run ends
- Reconstruction jobs include monitoring tools
- Every 15 minutes, intermediate histogram files are merged and displayed on the web
- Full merge occurs at end of run, final histogram files are registered on the Grid




Run 92226 Trigger AnyPhysTrig: Tile pos. correlation of Clus. opposite to most en. Cluster
Δφ
6
4
2
0
-2
0
2
Δη
1400
1200
1000
800
600
400
200
0
Run 92226, 1/physics_L1Calo
/TileCal/tileClusterEtaPhiDiff

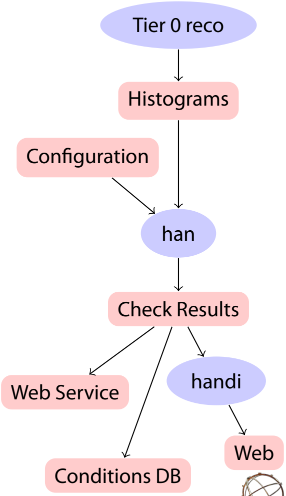
Data Quality Monitoring Core

- System for automatically checking histograms against expectations
- As histograms become available, they are run through “algorithms”
  - Algorithms can be simple (check number of filled bins) or complex (χ^2 comparisons, K-S tests)
  - Algorithms can use predetermined “reference” histograms for comparisons
  - Algorithms publish a “status” (red, yellow, green, or undetermined) and “results” (e.g. bins over threshold)
- Status checks are combined by summary algorithms
- Core software is shared with online monitoring

Offline Realization

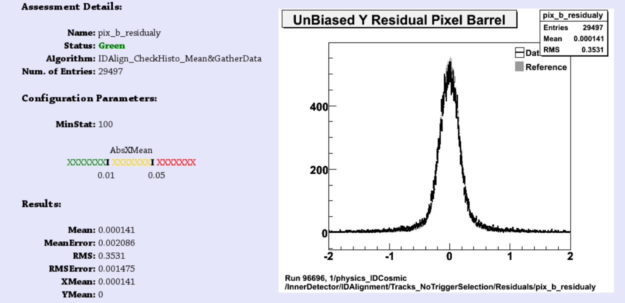
- Subdetector configurations merged and compiled into a binary configuration file
  - References included: binary configuration is self-contained
- "Han" (**H**istogram **an**alyzer) program reads configuration and an input ROOT file with histograms, runs the core monitoring code for automatic checks
- Han outputs a ROOT file containing
  - checked histograms,
  - check status and results,
  - configuration,
  - references

  (again, self-contained)



Visualization

- The default visualization of the results is HTML pages produced by “handi” (**han di**splay)
- Histogram, status, and algorithm results are shown
- Display options (e.g. axes ranges, linear/log scales) are specified in the han configuration
    - handi only needs the han output to run



Assessment Details:

Name: pix_b_residualy
Status: **Green**
Algorithm: IDAlign_CheckHisto_Mean&GatherData
Num. of Entries: 29497

Configuration Parameters:

MinStat: 100

AbsXMean
XXXXXXXI XXXXXXXI XXXXXXX
0.01 0.05

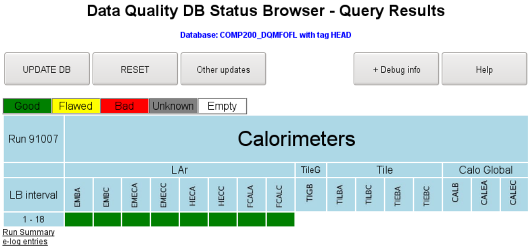
Results:

Mean: 0.000141
MeanError: 0.002086
RMS: 0.3531
RMSError: 0.001475
XMean: 0.000141
YMean: 0

Run 96696, 1/physics_IDCosmic
/InnerDetector/IDAlignment/Tracks_NoTriggerSelection/Residuals/pix_b_residualy

Database Storage

- Specific checks (or check summaries) are stored in the ATLAS conditions database
  - Results stored with the time granularity of the relevant check
- Results can be queried via C++, Python, web



Data Quality DB Status Browser - Query Results

Database: COMP200_DQMFOFL with tag HEAD

🟩 Good	🟨 Flawed	🟥 Bad	Unknown	Empty

Run 91007	Calorimeters															
	LAr								TileG	Tile				Calo Global		
LB interval	EMBA	EMBC	EMECA	EMECC	HECA	HECC	FCALA	FCALC	TIGB	TILBA	TILBC	TIEBA	TIEBC	CALB	CALEA	CALEC
1 - 18	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩								

Run Summary
e-log entries

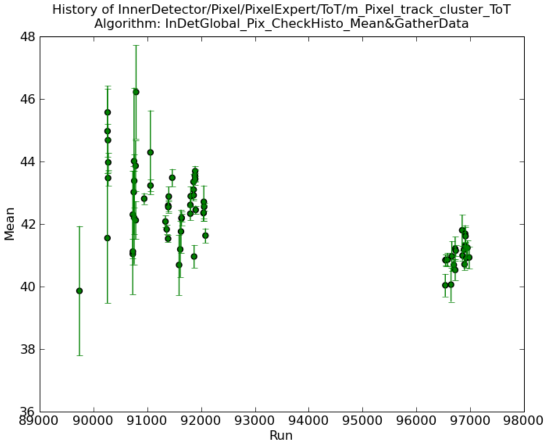
Exposing Results: Web Service

- An XML-RPC web service provides access to information in the han output files
  - Status flag
  - Algorithm results (mean, RMS, ...)
  - URL of PNG file of histogram from web display
  - Histograms themselves (transferred via ROOT’s XML serialization)
  - ...
- Allows distributed, language-neutral use of this information

Sample python code to retrieve information:

```
import xmlrpclib
s = xmlrpclib.ServerProxy(’http://atlasdqm.cern.ch’)
result = s.get_dqmf_results(
                {’stream’: ’physics_IDCosmic’,
                 ’source’: ’reproc’,
                 ’low_run’: 92226,
                 ’high_run’: 92226},
                 ’InnerDetector/Pixel/PixelExpert/ToT/m_Pixel_track_cluster_ToT:XMean’)
```

Web Service Example: History Plots



Plot dynamically generated across network using XML-RPC interface — no ROOT/database/... required

Other Uses

- The automatic check/web display code is relatively lightweight and is not tied to ATLAS Tier 0 reconstruction
- Has been deployed elsewhere:
  - Monitoring of later data reprocessing at Tier 1s
  - Software validation (does the output of today's nightly differ from yesterday's?)
  - Monte Carlo production validation

Production Experience

- Offline data quality monitoring was running at LHC startup
- Infrastructure fully exercised on Tier 0 during Sep-Dec 2008 cosmic ray data run with very few problems
- Check configurations rapidly developing as experience is gained with the detector
- Web service interface still very young, learning what we can do with it
- Still investigating usage in other contexts

Conclusion

- ATLAS offline data quality monitoring gives feedback on data within $\mathcal{O}(15 \text{ minutes})$ after collection
  - Automatic checks are performed
  - Results available on web
- System has been running smoothly in this role
- Other uses of infrastructure being explored and implemented