Great Lakes Tier-2 Computing and Muon Calibration Center Commissioning

Shawn McKee

DPF Computing Session

July 27th 2009

Overview

- ATLAS Great Lakes Tier-2 (Host for Calibration Center)
- ATLAS Muon Calibration Center
  - Physically a part of the AGLT2
  - Logically differentiated by policy and configuration
  - Requires additional, unique effort
    - Oracle installation, maintenance required
    - High priority configuration/resiliency
    - Special software (specific ATLAS releases)
    - Workflow management software (calibration splitter)
- Commissioning efforts and experiences

ATLAS Great Lakes Tier-2 (AGLT2)

- One of five ATLAS Tier-2 centers in the US
- Physically split between UM and MSU
- Logically one large center with redundant 10Gb network
  - Disk storage transparently split between sites
  - Compute nodes at sites appear identical to submitted jobs
- Currently comprised of 1800 job slots (cores), 500TB of dCache storage, 200TB of Lustre/test storage, 100TB other
- Condor is used for the job manager
- Extensive monitoring and management tools deployed

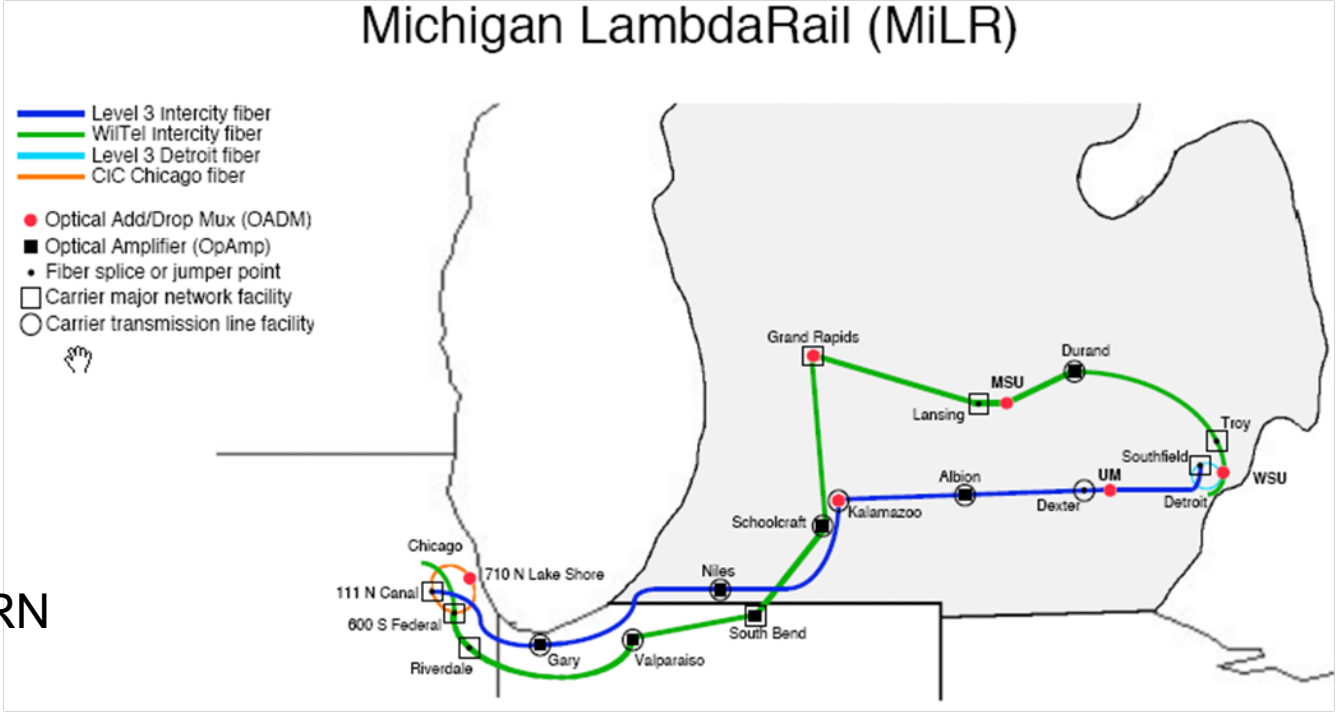
AGLT2 Networking

- AGLT2 has a resilient 10GE WAN connection with diverse paths to StarLight in Chicago (2 AGLT2 waves, 1 UltraLight)
- Participates in UltraLight and PLaNetS research programs

Has 19 10GE linked storage nodes

Is Dynamic Circuit Networked “enabled”

Has persistent circuits setup to BNL and CERN

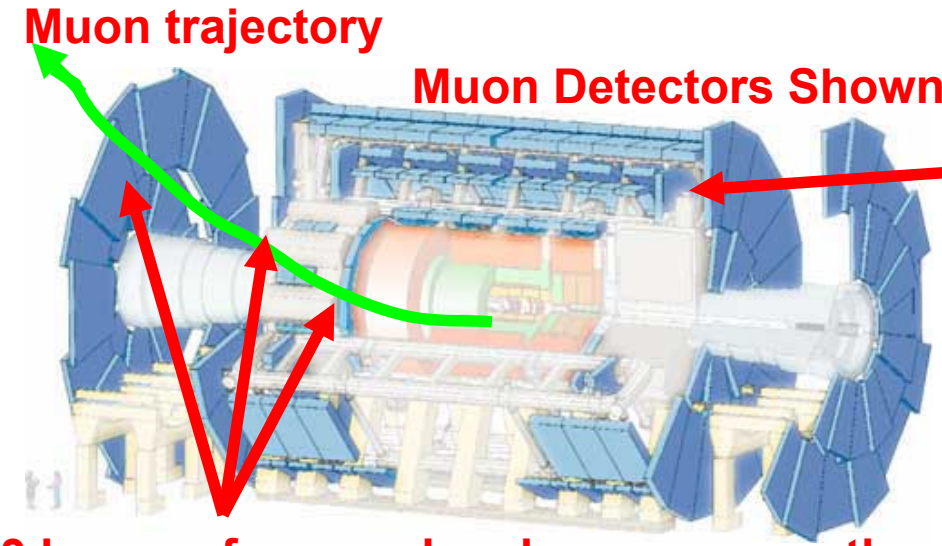


AGLT2 Status

- AGLT2 began operations in Fall of 2006 and has been performing very well, serving as a production site for ATLAS and supporting users and groups for analysis work.
- During STEP09 (June 2009) AGLT2 completed the second most number of analysis jobs worldwide by site.
- Regularly achieves 250-600% of WLCG MOU in CPU-hrs
- AGLT2 is unique among the US Tier-2 centers because it is also tasked with an additional role of being an ATLAS Muon Calibration center…

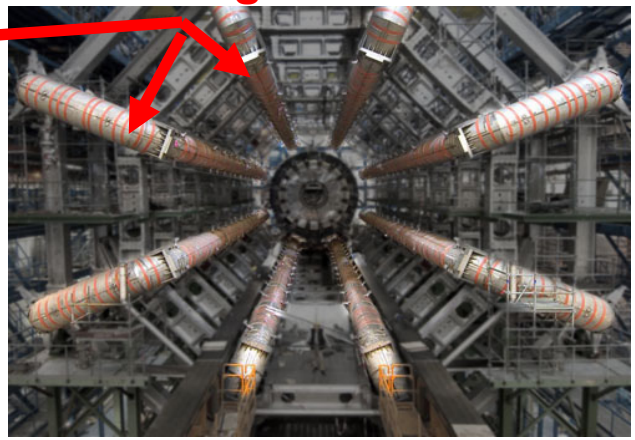
ATLAS Muon Spectrometer

- The ATLAS muon spectrometer measures the momentum of muons by measuring their trajectory as they pass through a magnetic field.
- Muons are very penetrating and in general pass completely through the ATLAS detector before decaying into an electron and 2 neutrinos.
- **To accurately measure muons requires a well calibrated system**



3 layers of muon chambers measure the path of the muon through the magnetic field

Toroidal magnets provide Magnetic field



End view of the ATLAS detector during construction

ATLAS Muon Calibration Centers

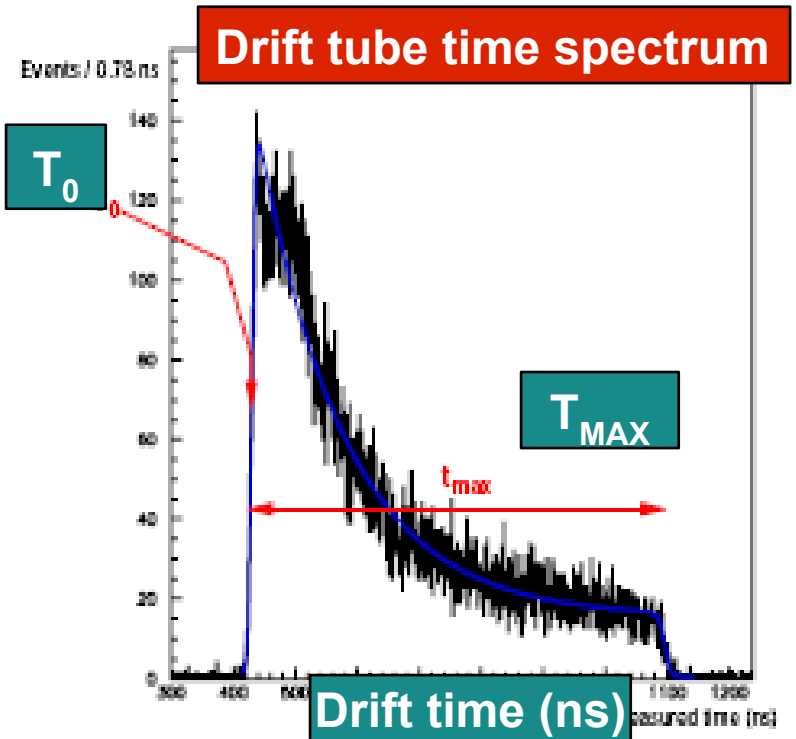
- In ATLAS, responsibility for calibrating the Muon system is split between 3 sites: University of Michigan, University of Rome, and Munich (Max Planck Institute+LMU)
- The calibration centers receive a **muon calibration data stream** via ATLAS distributed data management tools (the "**DQ2**" system).
- The data is reconstructed to find the muon tracks which are used as the input to calibrations. The calibrations are done on a daily basis and will occupy a ~100-cpu cluster at each calibration site.
- The calibrations must be provided to the Tier-0 within 36 hours so that the first-pass reconstruction on the data can be completed.
- A local ORACLE database is used to store/replicate all calibrations.

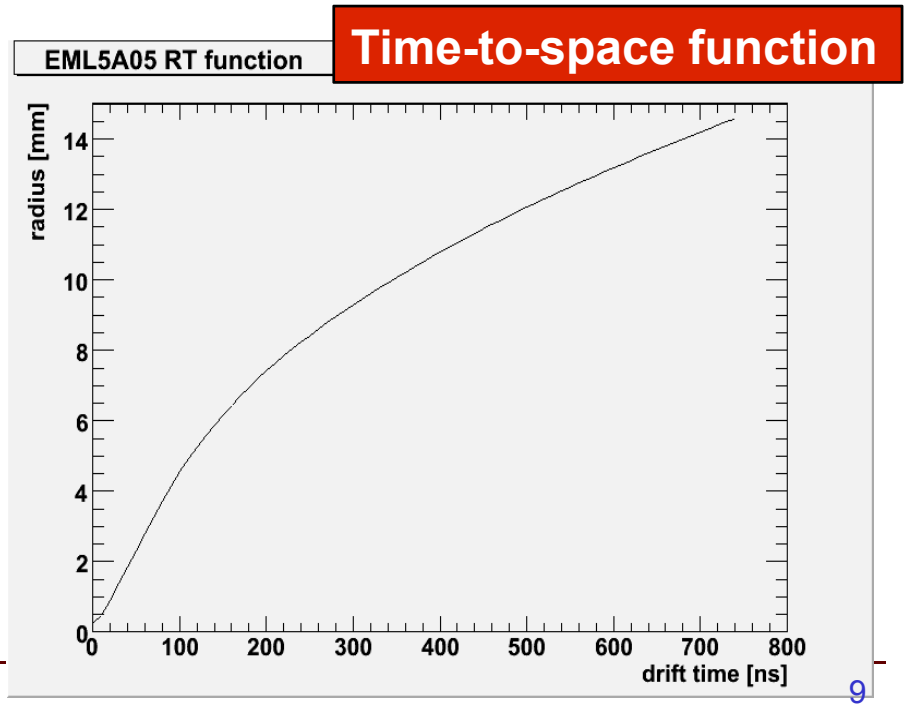
Muon Calibration Data Stream

- **Finding T_0 requires ~10^4 tracks to ensure a reliable fit.**
- Regular ATLAS trigger streams do not have enough muon tracks to calibrate the drift tubes in a timely basis. In order to get adequate statistics a special *calibration data stream* has been implemented.
- ATLAS data is recorded by the Trigger and Data Acquisition (TDAQ) system which has a 3-level structure to select which collision data to store. The system does a sophisticated pattern recognition to select only data which looks interesting , e.g., high-momentum muons (which could come from the Higgs decay, for example)
- This Muon calibration data stream takes muon tracks from 2-level of the TDAQ system which are then sent directly to 3 muon calibration centers for analysis. The calibration stream records about 10x as many muons as the ordinary data stream – about 10^8 per day, or 100 MB/day.

Muon Drift Tube Calibration

- To determine the drift radius two basic calibrations are needed:
  - Measurement of the time offset of the tube (T_0): The timing of the tube has an arbitrary offset due to electronic and cable delays. The T_0 represents the time when a particle passes next to the wire.
  - **Time-to-space function**: This function is the relationship between drift-time and drift radius





ATLAS Muon Calibration Workflow

- Data input
  - **Critical that calibration centers always are able to receive data**
  - DQ2 (ATLAS distributed data management) is used to manage muon calibration dataflow from the Tier-0->calibration centers
- Data processing
  - Sufficient on-demand (high-priority) storage and computing resources must be provided to achieve 36 hour turn-around
  - Access to needed conditions data must be quickly available
  - Operations and results must be checked and validated
- Data output
  - Calibration centers must run an Oracle instance which replicates calibration constants (via Oracle Streams) back to the Tier-0

Michigan ATLAS Muon Calibration Center

- AGLT2 also serves as an ATLAS Muon Calibration Center
- Tier-2 has 4 UM people: Bob Ball, Shawn McKee, Ben Meekhof and Wenjing Wu. **Calibration Center** is lead by Ed Diehl with help from many (See Ed's DPF talk on calibration at this conf.)
- The **unique additional** requirements:
  - High priority configuration/resiliency to meet calibration needs
  - Special software (specific ATLAS releases)
  - Workflow management software (calibration splitter)
  - Requires Oracle installation
- The next slides discuss how we address these requirements

Defining the Calibration Center Resources

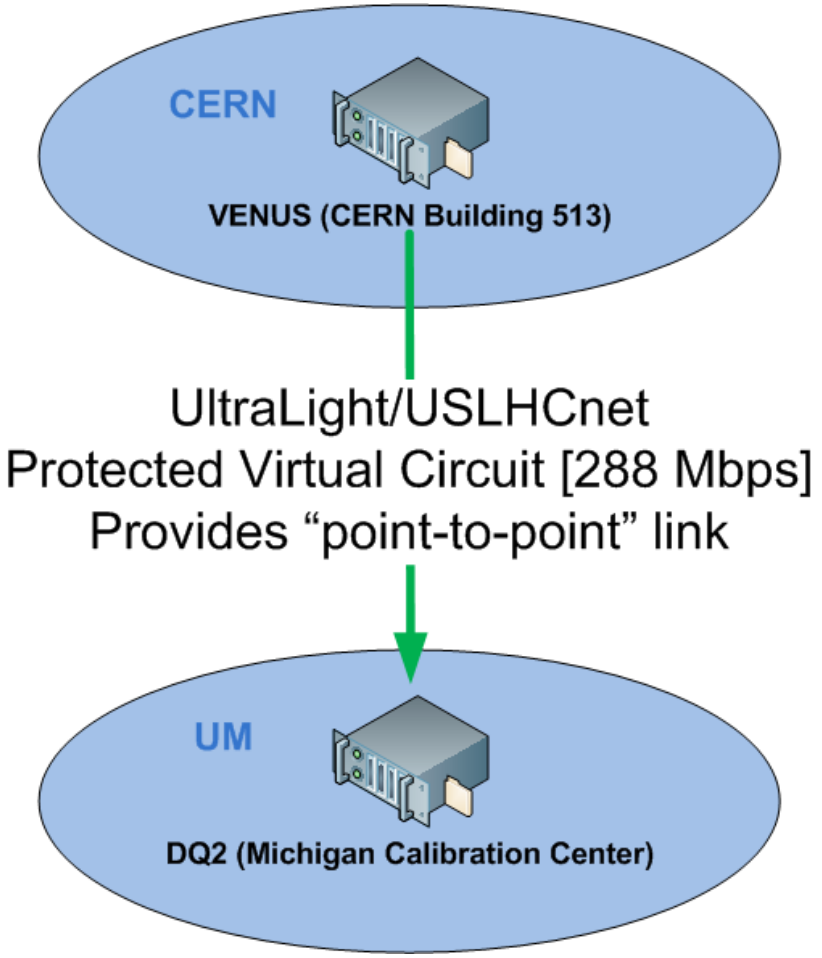
- The UM Calibration center is almost a subset of AGLT2 and is defined by policy and configuration rather than distinct hardware:
  - Calibration job slots are a subset of full available set
  - Separate Condor "Accounting Group" for calibration jobs
    - Quota is currently 100 job slots. Can easily adapt as required
    - Condor starts jobs first from "least filled quota" accounting groups
    - Calibration jobs start at high priority
  - Typical production job run times from 2-12hrs each
  - Can quickly go from 0 calibration jobs to full set of 100
    - Anywhere from 4 to 40 minutes to do this, depending on run time of production tasks (or Analysis tasks, if they are running)
- Some distinct additional physical resources: Oracle server, splitter server, network circuits, DDM storage (CALIBDISK)

ATLAS Muon Calibration Data Paths

- The Michigan calibration center is “further” in a network sense from the Tier-0 than the other calibration sites
- We need to ensure we can always get the muon calibration data **quickly** and **continuously**
- The **primary path** is to use the ATLAS distributed data management system (DDM/DQ2) to move the data to all the muon calibration sites. This has been working well…
- **However DDM/DQ2 doesn’t always work…what happens when there is a failure?**

Alt. ATLAS Muon Calibration Data Path

- A protected virtual circuit using UltraLight & USLHCnet is operational for moving ATLAS muon calibration data from CERN to Michigan
- Initial tests have moved data at **99%** of link capacity using FDT (98 GB of data) [March 5 2008]
- Future work will convert VENUS to a grid SE for CALIBDISK



While volume isn't large this is CRITICAL data. First pass – T0/TS reconstruction depends upon this chain

Managing Calibration Workflow

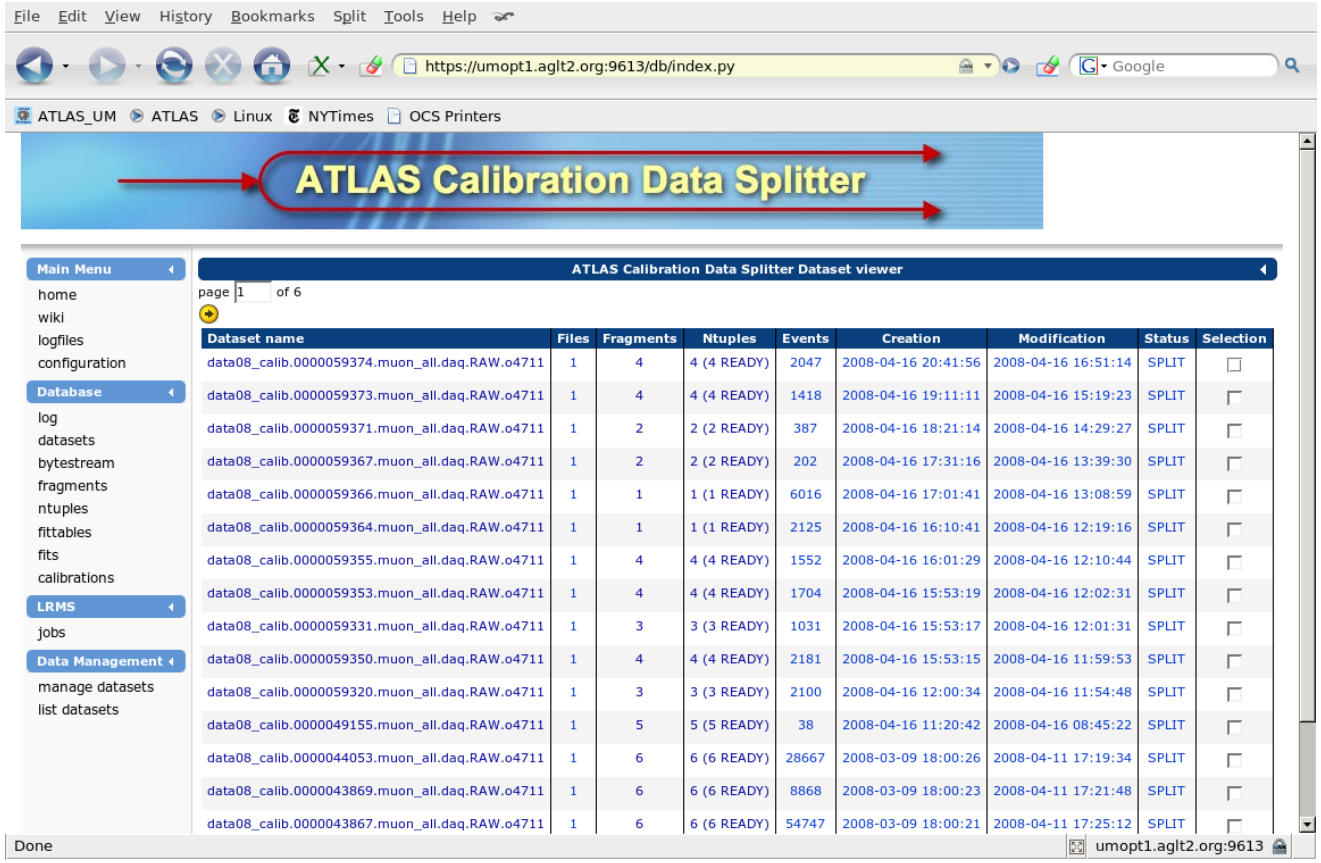
- All 3 calibration centers are utilizing a common “ATLAS Calibration Data Splitter” from Alessandro de Salvo (Rome)

Integrates with local site DDM and job management systems

Provides secure web interface to calibration work. Accessible anywhere

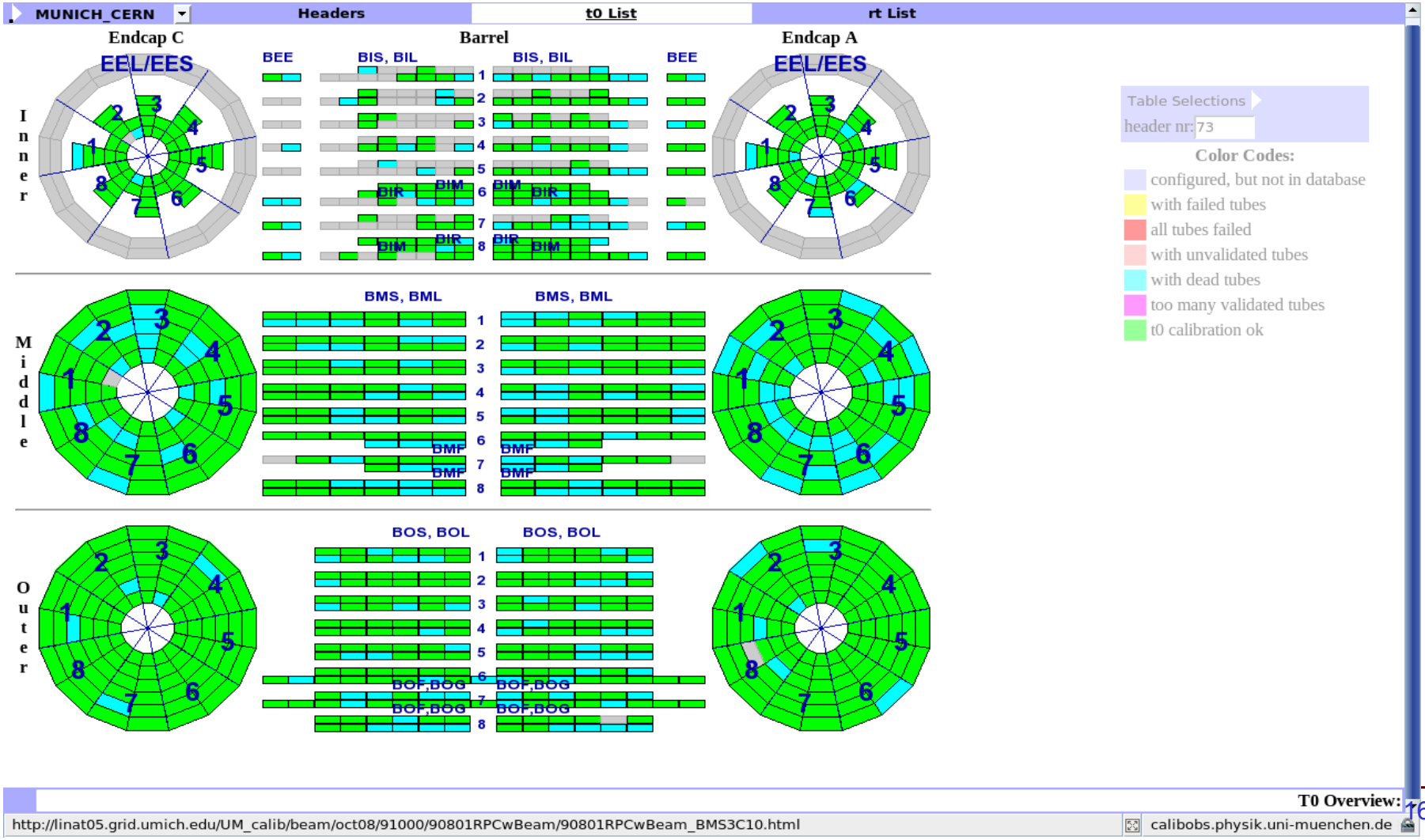
Organizes and tracks calibration workflow

Allows other sites to easily take-over for “down” sites



DBObserver Interface Example

DB Observer (Munich) view of the T0 results from one calibration run. Green or light-blue indicate the calibration succeeded. Clicking can "drill down" to more details.



AGLT2 Oracle Streams Setup

- The design of the ATLAS Muon Calibration centers utilizes Oracle with STREAMS replication at each site.
- The Michigan Oracle server was the initial example installation for the ATLAS muon calibration centers and has been using streams replication to CERN since September 2006 (Dell PE1950, RAID10)
- Data entered in our server is continually replicated to a central Oracle server at CERN for use in the first-pass reconstruction.
- Currently “standard” IP routing is used between sites is used.
- Future possibilities include special routing across protected network circuits to guarantee a level of service for this traffic.

Commissioning UM Calibration Center

- During the last year we have worked on ‘commissioning’ our calibration center
- The primary issues we encountered:
  - Meshing the splitter with our local site configuration (often OSG vs LCG differences or security related)
  - Oracle maintenance and upgrades in the presence of STREAMS
  - Data management: insuring timely access to input data
  - Calibration slow job initialization
- **All solved**…the next few slides provide some details on this last issue and our solution.

Slow Oracle DB Access at AGLT2

All studies are done with ATHENA release kit 14.5.0.x and 14.5.1 by including *include("RecJobTransforms/UseOracle.py")* inside top job options *MuonDataRec_myTopOptions.py* from MuonSpectrometer/MuonReconstruction/MuonRecExample/ to use nearby TIER1 Oracle DB for detector calibration constants and conditions, where BNL TIER1 Oracle DB is used for Muon Calibration Center at AGLT2.

Using ATHENA AtlasProduction 14.5.0.1 to do muon reconstruction for 100 events with normal access to BNL's Oracle DB, where raw data were used at CERN, and stripped MUON raw data were used at BNL and AGLT2.

Site	Node	Processing Time
CERN	pcatum11.cern.ch, my desktop	~24 minutes
BNL	acas002.bnl.gov, user login node	~14 minutes
AGLT2	user login nodes and batch farm	~64 minutes

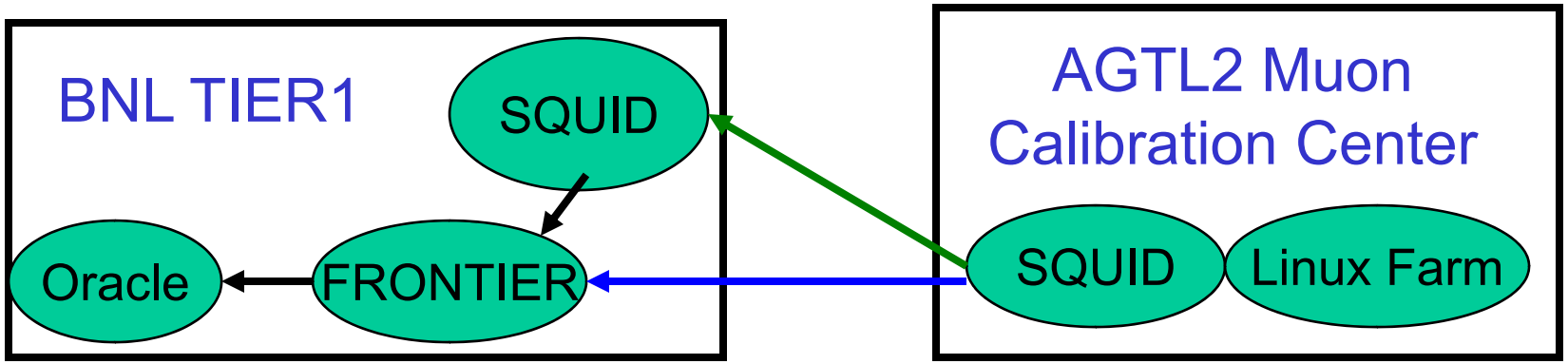
Fix DB Access via Squid/Frontier

To fix the slow DB access problems, we discussed:

1. Using SQLite replica, which only contains limited DB tables of certain period for particular task, such as Muon Calibration Task;
2. Replicate Oracle DB from TIER1 to TIER2: issue would be **license and support** of Oracle at Tier-2s
3. **Using FRONTIER combined with SQUID caches**.

We focused on option 3: using SQUID/FRONTIER

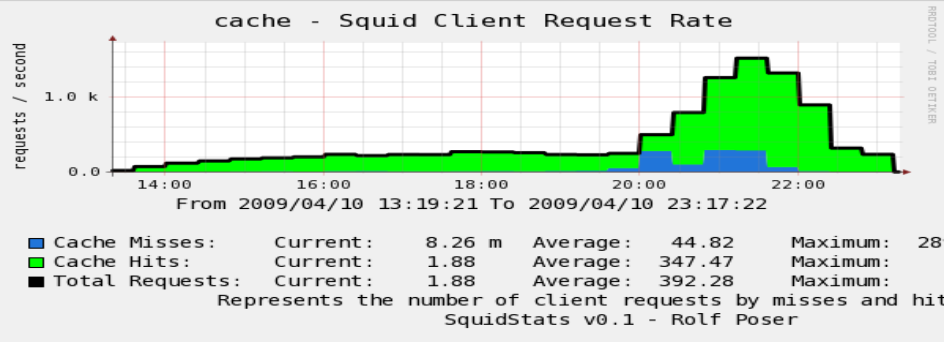
This was a collaborative effort with BNL (John DeStefano, Carlos Gamboa)

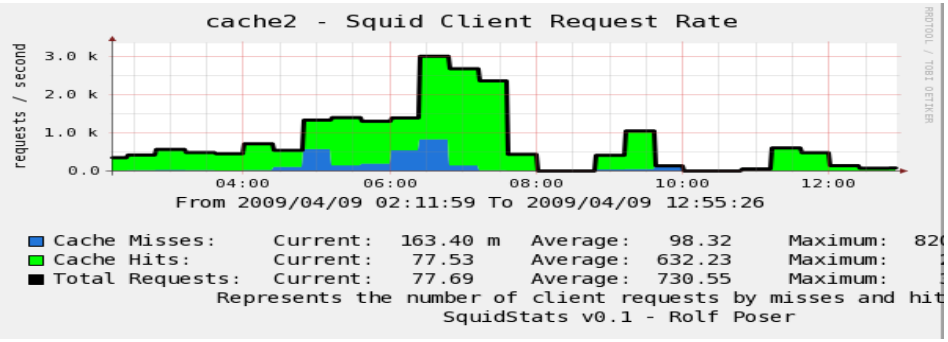


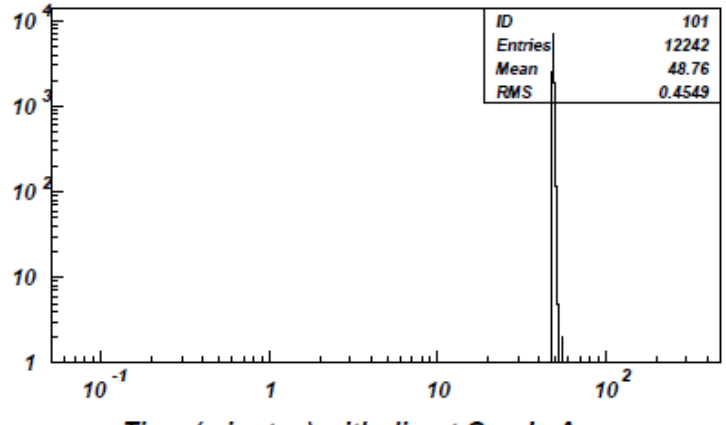
Finished intensive testing with/without local SQUID caches to find limits. SQUID at “local” site seems necessary for best performance

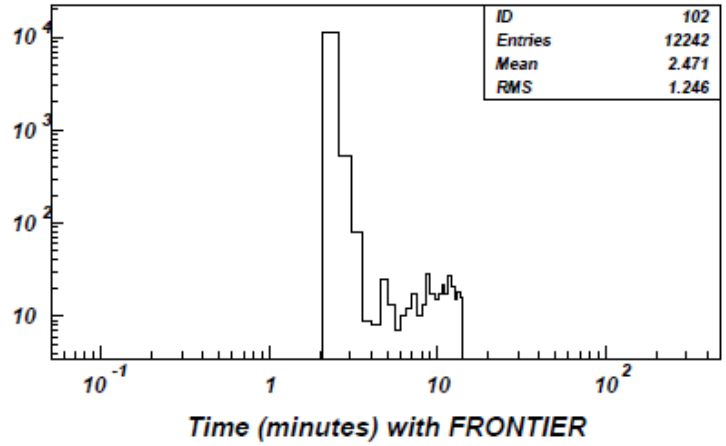
Further Squid/Frontier Production Tests

Job crashes with DB connection issues were observed. After several days tests and debug, it is concluded that remote SQUID server is not sufficient for massive production, LOCAL SQUID servers are needed, one SQUID per roughly 1000 CPUs **(IMPLEMENTED at AGLT2 with 2 SQUIDs)**







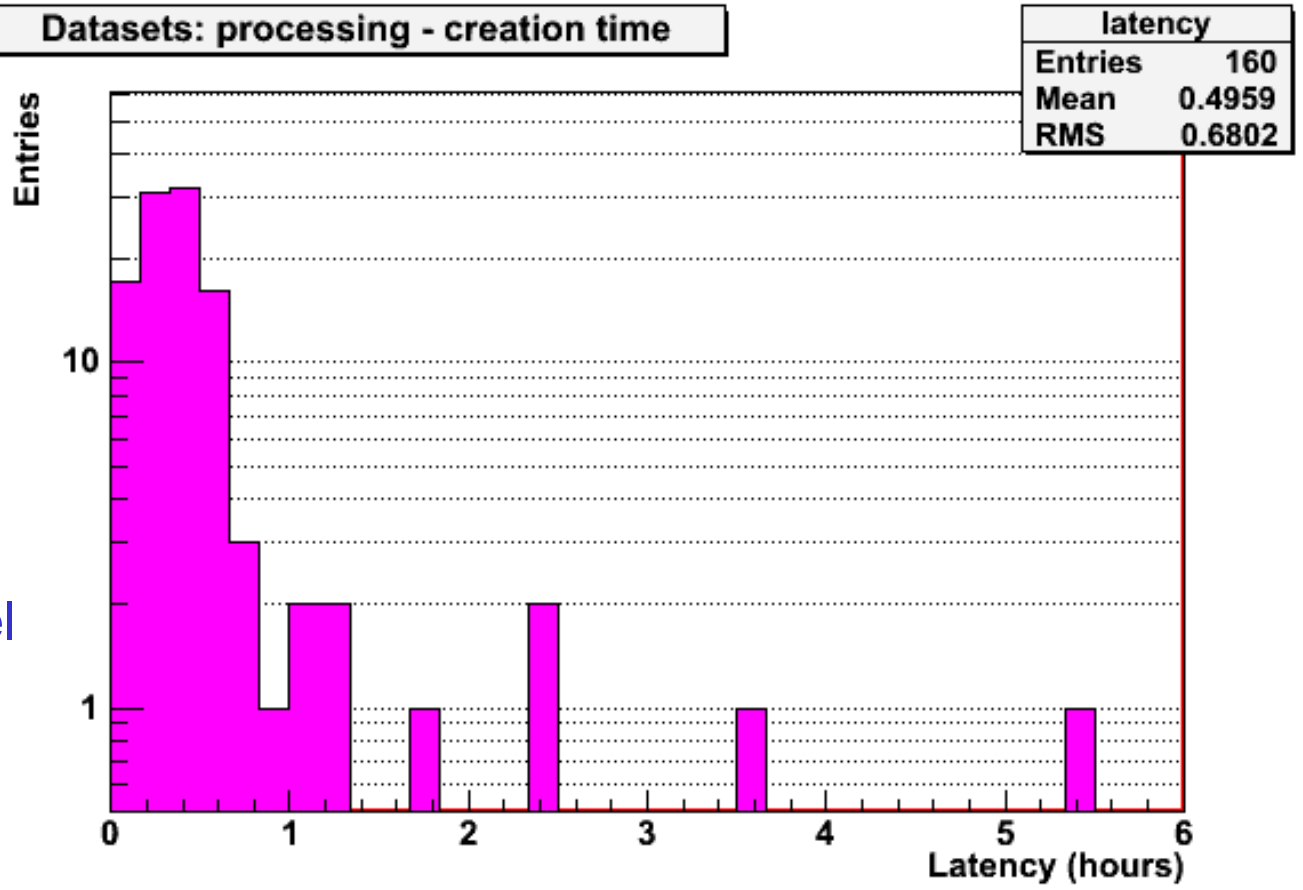


2008 Cosmic-Runs: Data Latency

Shown below is the latency in getting the calibration stream files during the October 2008 cosmic-data-runs.

NOTE that even the worst-case time is less than 5.5 hours and most are retrieved within 30 minutes

We should have no problems with this level of latency for 36 hour turn-around



Calibration Center Near-Term Plans

- We need to better integrate the Muon Splitter job-submission with the ATLAS PANDA system to properly **account for** and **prioritize** these jobs within the ATLAS job-mgmt framework
- Work on the backup data path to ensure robust operations:
  - SE at CERN to be used as a backup data source
  - Oracle STREAMS replication via the protected circuit path
- Continued Splitter testing and integration with local site config.
- Calibration tools are in good shape but can benefit from further testing. **Algorithms are well developed and tested already**.
- Add more MDT data quality assurance components to the system

Conclusion

- The Michigan ATLAS Muon Calibration Center has been successfully deployed and operated as a prioritized subset of the AGLT2
- Issues related to the required services and their robust operation have been identified and addressed during the last year: **splitter integration, DB access, DDM issues**
- Further testing and improvements are underway/planned
- **The tools and infrastructure are in place already to provide muon calibration results as required**