MICE COMPUTING FRAMEWORK

DURGA RAJARAM

MICE OPTICS REVIEW
JANUARY 14, 2016

Software & Computing Project

- Wide range of tasks
  - Hardware controls & monitoring
  - Read out detectors – DAQ
  - Reconstruction & Simulation
  - Online monitoring & reconstruction
  - Describe geometry, fields
  - Provide database tools to manage configurations
  - Data movement & curation
  - Manage and maintain Control Room servers
  - Web services

Configurations Management

- MICE: precision experiment with a range of complex elements
  - Muon beamline, absorbers, cooling channel magnets, tracking and particle identification detectors
- Data will be taken with a wide variety of configurations
  - Configurations could change on the order of hours/weeks/months
- Book-keeping of experiment's layout, configurations is essential
  - Historical record of experimental conditions
  - Accurate simulations
  - Reproducible analyses
- Configuration management framework to provide a central storage for the settings and tools to access them
- Primarily two types of configurations:
  - As-run (conditions during data collection, data processing..)
  - Expert-filled (electronics maps, calibrations, geometry..)

CONFIGURATIONS DATABASE

- Configurations stored in PostgreSQL database
- Temporal database
  - Transaction time
    - When was the setting recorded?
  - Validity time
    - What time period is the setting valid for?
- Possible to have multiple records with the same validity time
  - E.g. refined calibrations superseding an earlier one, geometry corrected with newer survey…
  - In such cases, the latest record is taken to be 'current'
- User access to the DB is through a set of Python/C APIs
- Source and various utilities shipped with the MICE Software (MAUS)

CONFIGURATIONS (RUN-TIME)

- Run-time configurations
  - To record conditions during data taking
  - Essential for simulation & analysis
    - Beamline magnets
    - Diffuser setting
    - Cooling channel magnets
      - Mode, polarity, current
    - Absorber setting
    - Scaler counts
    - …
  - Filled by Run Control during data-taking
  - Shifters verify that the information was recorded correctly

Run Number:	7591
Run Type:	Test
Start Date:	2015-12-17 20:00:05.0
End Date:	2015-12-17 21:59:57.0
Start Pulse:	1392183
End Pulse:	1397800
Target Depth (mm):	0.0
Target Delay:	0.0
Total Beam Loss (mV):	
Daq Version:	7.66
Daq Trigger:	TOF1
Daq Gate Width (ms):	3.0005
Start Notes:	task1 3+240 pion no Xe
End Notes:	end of shift looks good
Gdc Host Name:	miceraid5
Optics:	Individual Settings
Diffuser Thickness:	0
Beam Stop:	Open
Status:	false
Step:	4.0
Proton Absorber Thickness:	146

Magnets:

Name	Set Current	Polarity
D1	218.520004	1
D2	107.601997	1
DS	3.735294	1
Q1	62.029999	1
Q2	113.410004	1
Q3	69.300003	1
Q4	190.259995	1
Q5	254.947998	1
Q6	169.112	1
Q7	171.255997	1
Q8	259.115997	1
Q9	221.419998	1

Scalars:

Name	Value
GVa1 Triggers	5602
LMC-1234 Count	868120
LMC-12 Count	6389129
LMC-34 Count	6061263
Particle Triggers	45703
Requested Triggers	60601
ToF0 Triggers	304238
ToF1 Triggers	60602
ToF2 Triggers	20594

LDC Hosts:

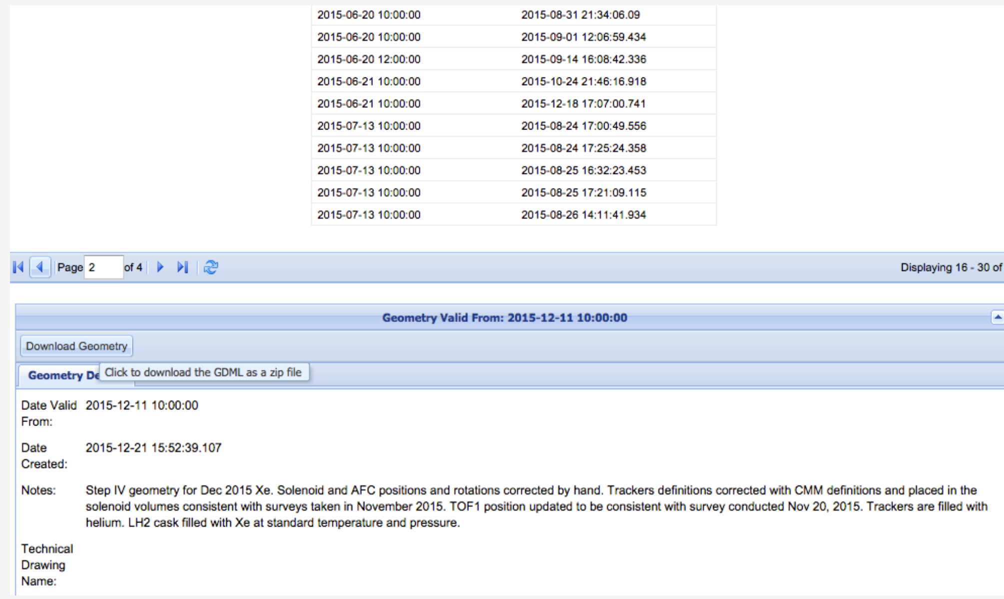
Host
miceacq20
miceacq17
miceacq16
miceacq15
miceacq14

CONFIGURATIONS (DETECTORS)

- Detectors for tracking and particle identification.
- Primarily two types of configurations for detectors
- Cabling maps to relate electronics readout channels to the physical detector
  - Feeds into the "unpacking" of raw data
- Further, the detectors have to be calibrated
  - E.g. convert ADC to #photoelectrons (tracker), TDC to time, timewalk corrections….
- Calibrations typically result from analysis of dedicated runs taken for the purpose
  - Performed by detector experts
  - Stored in the configurations database with a validity timestamp
  - Detector, calibration_type, valid_from
- Maps & Calibrations can be accessed by run#, or for a certain date

CONFIGURATIONS (GEOMETRY)

- Need accurate geometry to describe the layout of MICE
- Needed for accurate simulation & reconstruction
- CAD drawings folded with survey information to generate a GDML geometry description
- GDML stored in CDB with a validity timestamp
- Updated by geometry manager when corrections made or new survey performed
- Users download geometry by ID or date or simply for a certain Run#
  - Downloaded geometry description also contains magnet currents, diffuser, absorber settings



CONFIGURATIONS (PROCESSING)

- When processing raw data or simulating
- Essential to keep track of:
  - Software version used
  - Specific settings
  - Geometry, reconstruction flags, input beam (MC)
- Handled by "Batch Iteration" and "MC Iteration" tables
- If user requests (re)processing with non-default cards, they should be uploaded to the DB and are assigned a new iteration#
- Any reconstructed or simulated dataset can be uniquely identified (and reproduced) by the s/w version and an "iteration number" which identifies specific flags and inputs
- Thus possible to re-process/re-simulate data with different software versions, or different reconstruction flags and maintain independent records

Database Management

- PostgreSQL database
- Master server in MLCR rack room
- World-wide read-only access via web interface
  - http://cdb.mice.rl.ac.uk
- Write **only** from within MLCR network
- Read-only slave at RAL PPD
- Failover from master to slave ~5 minutes (not yet automated)
- Snapshot every 10 minutes, full backup every hour, backups kept for a month, monthly backups kept for a year

Data Curation

- Aim: to be able to reliably store and retrieve various types of data
- Raw data
  - Most critical, cannot be "recovered"
- Reconstructed data
  - Every iteration of reconstruction is stored to tape
- MC data
  - Input beams, Simulated outputs
- Miscellaneous
  - Calibration data, geometry, surveys, field maps, etc
- Data on CASTOR require a GRID certificate for access
- For convenience, data are also copied to Imperial College GRID store and made available via `http`

Data Movement

- Applies to raw data and MLCR-reconstructed data
- When a run ends, Run Control triggers a file-compactor
  - Tar archive of raw binary data along with data acquisition logs and histogram outputs from online reconstruction & online daq-level monitoring
  - Compacted data moved to temporary storage, and failure raised to Run Control for alerting shifters
  - Watcher on dedicated datamover machine watches for appearance of new compacted data
  - When new data appears, copied to permanent tape storage and file registered in a DB
- For reconstruction
  - An independent watcher on dedicated reconstruction machine launches reconstruction job upon appearance of new compacted data
  - A semaphore is written at end of reconstruction triggering the copy of the reconstructed output to permanent tape storage
  - Data stored in file-tree which references the *software version* & *iteration#*
  - E.g. `RECO/`**`MAUS-v1.3.0`**`/1/Step4/07500/`**`07591`**`_offline.tar`
  - The tarball contains reconstructed ROOT output, the processing log, and reconstruction cards

DATA PROCESSING

- Offline reconstruction has so far been on the GRID
  - During data-taking, use fast-response Tier-1 queue at RAL
  - Turns reconstructed data around within 24 hours
- With the speed improvements achieved with MAUS we decided to reconstruct data "live" in the MLCR.
  - Allows us to do the processing as we take data (or at worst just after a run ends)
  - Avoids queuing and submission issues (proxies, etc) on the GRID
  - Faster response reacting to patches & code fixes
- Tested in September: reconstruction happens ~parallel with data-taking
- A dedicated machine was then installed & was used for processing data during Nov-Dec 2015 runs
  - Reconstruction is automatically triggered at the end of a run
  - book-keeping improvements being developed (to register reconstruction status & reconstruction quality in DB)

Summary

- MICE is a precision experiment with a wide range of components run with a variety of configurations
- Configurations management is essential for ensuring accurate representation, simulation, and analysis of data
- A robust, flexible database has been developed to manage the configurations with user-friendly access
- All raw and processed data are curated to long term storage
- Data processing has been sped up ensuring quick turnaround for analysis