

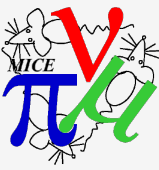


DATA PROCESSING & DATA FLOW

DURGA RAJARAM

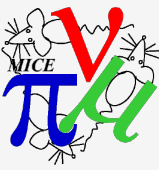
MICE PROJECT BOARD

APRIL 6, 2016



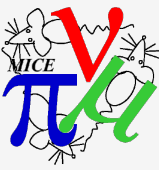
INTRODUCTION

- MICE: precision experiment with a variety of complex elements:
 - beamline, absorbers, cooling channel magnets, tracking and particle identification detectors
- Data taken with a spectrum of configurations
 - which can change over hours/weeks/months
- Thus, essential to have:
 - Accurate book-keeping of experimental configurations
 - Reliable data movement and curation
 - Real-time monitoring of data quality
 - Fast turnaround reconstruction for analysis



CONFIGURATIONS

- Book-keeping of configurations essential for
 - Historical record of experimental conditions
 - Accurate simulation
 - Reproducible reconstruction & analyses
- Primarily two types of configurations
 - As-run (conditions during data-taking, data-processing)
 - Expert-filled (electronics maps, calibrations, geometry...)



CONFIGURATIONS (RUN-TIME)

- Record conditions during data taking
- Needed during data-processing/simulation
 - Beamline magnets
 - Diffuser setting
 - Cooling channel magnets
 - Mode, polarity, current
 - Absorber settings
 - Scaler counts

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:

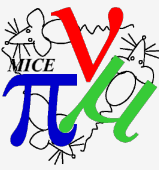
Name	Set Current	Polarity
D1	218.520004	1
D2	107.601997	1
DS	3.735294	1
Q1	62.029999	1
Q2	113.410004	1
Q3	69.300003	1
Q4	190.259995	1
Q5	254.947998	1
Q6	169.112	1
Q7	171.255997	1
Q8	259.115997	1
Q9	221.419998	1

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



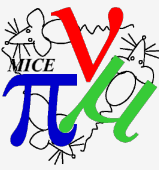
CONFIGURATIONS (DETECTORS)

- 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 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
 - Calibrations can be accessed by run#, or for a certain date



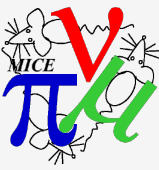
CONFIGURATIONS (PROCESSING)

- When processing raw data or simulating
- Essential to keep track of:
 - Software version used
 - Specific settings
 - Geometry, reconstruction flags, input beam (MC)
- Handled by “Iteration Number” DB tables
- e.g if a user requests (re)processing with non-default cards
 - They get uploaded to the DB and assigned a new iteration#
- Thus reconstructed dataset can be uniquely identified (and reproduced) by the software version & iteration #
- Thus, possible to re-process/re-simulate data with different software versions, or different cards and still maintain independent records



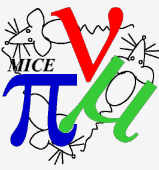
DATA CURATION

- Aim: to reliably store, & retrieve various types of data
- Raw data
 - Most critical, cannot be “recovered”
- Reconstructed data
 - Every iteration of reconstructed data is stored to tape
- MC data
 - Input beams, Simulated outputs
- Miscellaneous
 - Calibration data, geometry, surveys, field maps, etc
- CASTOR for long-term storage & “local” filestore
 - Data on CASTOR require GRID-certificate for access
 - For convenience, raw & reconstructed data are copied to Imperial College GRID store and made available via http



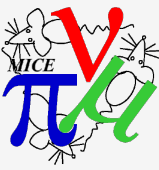
DATA FLOW (RAW)

- When a run ends, Run Control triggers a file compactor
 - Tar archive of raw binary data along with DAQ logs, histogram outputs from online reconstruction & online daq-level monitoring
 - Compacted data moved to temporary storage & any failures are raised to Run Control to alert shifters
 - Watcher on dedicated datamover machine watches for appearance of new compacted data
 - When data appears, copied to permanent storage and registered in a DB



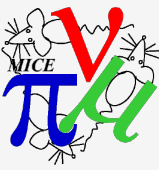
DATA FLOW (RECONSTRUCTION)

- Similar to movement of raw data
- An independent watcher on dedicated reco machine watches for newly compacted data
 - Upon new data, reconstruction of that run is triggered
 - A semaphore is created at the end of reconstruction job
 - A “reco-mover” watches for appearance of reconstruction semaphores & copies the output to tape
 - Data stored in filetree with software version & iteration # embedded in the file-path



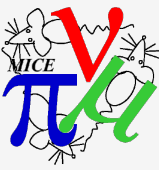
DATA PROCESSING

- Offline reconstruction used to be done on the GRID
 - Were able to reconstruct data around within ~6 hours
- With the speed improvements to MAUS we decided to reconstruct data 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
- A dedicated machine was then installed & has been used for doing the official reconstruction since Nov 2015



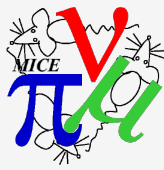
DATA PROCESSING FRAMEWORK

- Reconstruction is performed using:
 - Official major release of MAUS
 - or a tagged patch release
 - Geometries from the Configurations Database (CDB)
 - Geometry is downloaded by-run-number
 - DAQ and detector electronics maps from the CDB
 - Maps are downloaded by run-number
 - Detector calibrations from the CDB
 - Calibrations are downloaded by run-number
 - Data cards from the default configuration in MAUS



DATA PROCESSING OUTPUT

- All Step4 data since June 2015 have been reprocessed
 - *~June is when downstream detectors moved into their final Step4 positions & tracker commissioning was stable*
- Output is a tarball which contains:
 - Reconstructed ROOT file
 - Reconstruction log
 - Data cards used
 - Geometry files retrieved for this run
 - Tracker map & calib. & dead channel list for the run
- Outputs available from:
 - <http://reco.mice.rl.ac.uk>
- Outputs also copied to permanent tape store



DATA PROCESSING OUTPUT

- Outputs available within ~hr after run ends

reco.mice.rl.ac.uk

Apps faw mice.iit micemine lp:maus jenkins elog Other Bookmarks

MICE Reconstructed Data

Show 25 entries Search:

Run Number	Run Date	Optics	Triggers	Reco Version	Download	Reco Date
7866	2016-03-24 21:32:31	3-172+M0_WDS	101638	MAUS-v2.1.0	07866 offline.tar (1 GB) (md5)	2016-03-25 00:03:36
7865	2016-03-24 19:26:42	3-172+M0_WDS	96126	MAUS-v2.1.0	07865 offline.tar (583.7 MB)	2016-03-24 21:46:55
7864	2016-03-24 17:22:50	3-172+M0_WDS	95279	MAUS-v2.1.0	07864 offline.tar (988.7 MB) (md5)	2016-03-24 19:40:35
7863	2016-03-24 16:40:03	3-172+M0_WDS	29187	MAUS-v2.1.0	07863 offline.tar (302.2 MB) (md5)	2016-03-24 17:24:48
7861	2016-03-24 14:19:20	3-172+M0_WDS	96507	MAUS-v2.1.0	07861 offline.tar (1006.2 MB) (md5)	2016-03-24 16:51:32
7860	2016-03-24 12:31:28	3-240+M0	63099	MAUS-v2.1.0	07860 offline.tar (824.9 MB) (md5)	2016-03-24 14:25:58



DATA PROCESSING (IMPROVEMENTS)

- Want to improve and add a few things to the processing:
 - Reconstruction plots
 - Currently, Online Reconstruction (live) & a “physics devil” (post-process) produce plots of the data
 - Integrating the physics devil into the processing framework so that the plots are available as part of output
 - Reconstruction quality flags
 - Have created CDB infrastructure to hold flags for each run, for each detector, for each software version used
 - Will be a 4-bit word with a flag from each detector
 - Up to each detector to determine if quality was OK/not



SUMMARY

- MICE is a precision experiment with a wide range of components run with a variety of configurations
- 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 quality continually monitored
- Data processing has been sped up ensuring rapid turnaround for analysis