

# 2x2 DQM and data transfer

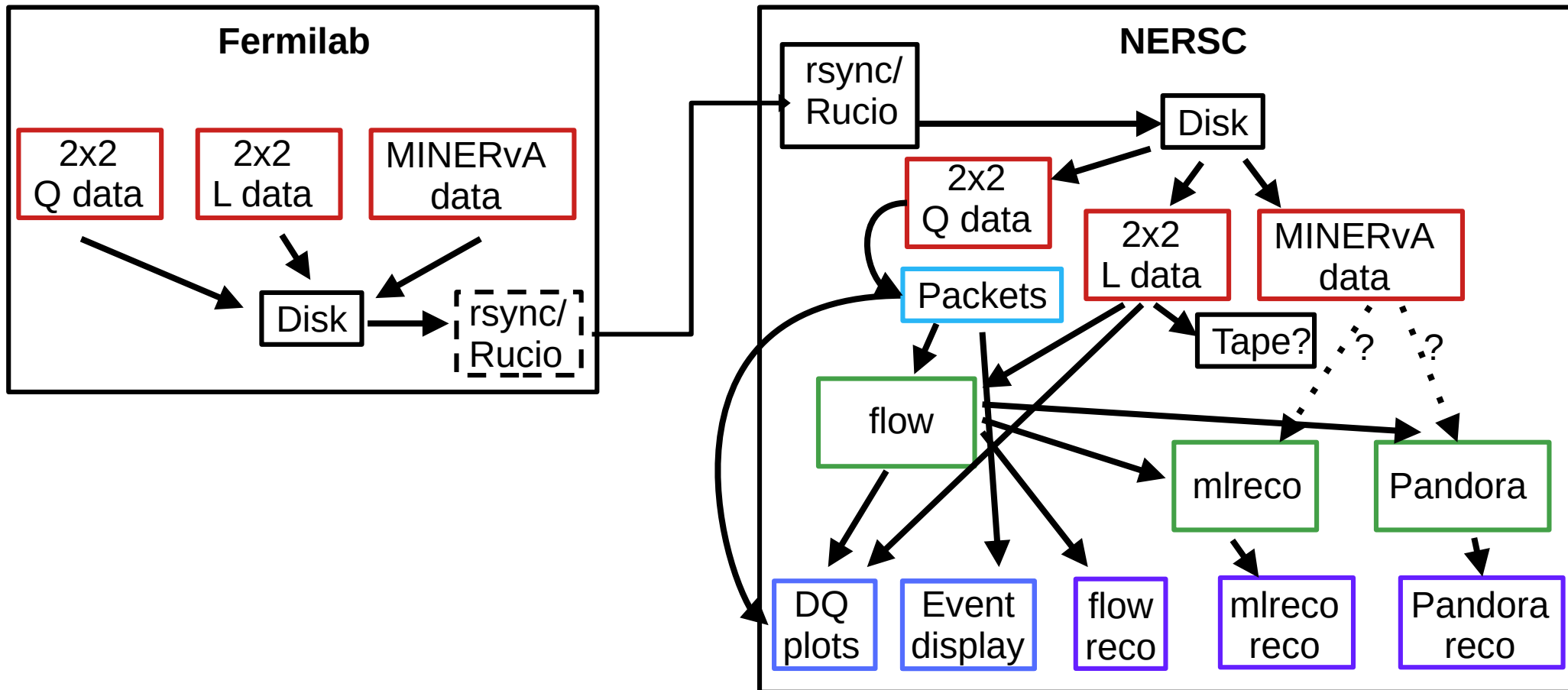
Matt Kramer, LBNL

2x2 analysis workshop, Bern  
Jan 20 2023

# Context

- **2x2 + MINERvA in NuMI beam soon**, expect  $\sim 5E20$  POT ( $\sim 1$ yr @ 50% uptime)
- Need realtime data quality monitoring (DQM)
- Need realtime data replication, cataloging
- Expect  $O(100$  TB) for  $5E20$  POT
  - Includes CRS and LRS, but not MINERvA

# 2x2 data processing diagram



# Data quality, event display

- Have Peter's larpix-monitor for monitoring filesystem, generating plots
  - Plots available through web portal
  - UI currently pretty basic (just browsing folders of images); “dashboard” would be nice
- Have Roberto's 3D event display (larnd-display)
  - Integration with DQ plots?
- Thanks to Dat Tran for stepping forward to implement new plots
- Need to incorporate flow in order to enable DQ plots based on calibrated data

# DQM at NERSC

- Currently, DQ plots (generated by larpix-monitor) can be browsed e.g. at <https://portal.nersc.gov/project/dune/data/Module1/TPC12/DQM>
- Going forward, hope is to create a web-based DQ “dashboard” to present these plots more formally. E.g.:
  - Use Python Dash framework (used in larnd-display)
  - Host at NERSC using Spin
  - Both a “one page per run” layout and a “global overview” layout (time-series etc.)
    - Add'l processing (beyond larpix-monitor) needed for the latter

# Adding DQ plots

- See <https://github.com/larpix/larpix-monitor>
- Adding a plot is straightforward
  - Create a class, implement `__call__` to take (packetized) HDF5 data and generate/update a matplotlib or Plotly plot
  - i.e. copy-paste and modify one of the existing plots
- Need volunteers for defining and implemented more plots (dQ/dx, electron livetime...)
  - Planning and progress tracking in [this Google doc](#)
- Thoughts
  - Do we also want plots that use higher-level info e.g. as produced by flow?
    - In that case would need to modify larpix-monitor core, an: run flow prior to larpix-monitor
  - Light readout system? Currently ignored in DQM

# Data management

- Metadata: To be stored in MetaCat.
  - Replicate metadata between NERSC and FNAL MetaCat instances? Need to define replication scheme.
  - Initial metadata filling at FNAL or at NERSC?
  - Need to solidify a metadata schema: Run configuration, conditions, calibrations, ...
  - Association between CRS, LRS, and MINERvA files
- Data replication
  - Quick and dirty: rsync to NERSC
  - Preferable: NERSC as a Rucio storage element
- Rucio for cataloging, tracking, accessing data
- At NERSC, web portal for providing “easy” access to data
- For recorded data, light waveforms to dominate space requirements
  - Full waveform data might be overkill for most analysis
  - Straight to tape? Use dCache to make access more transparent?
  - Process waveforms to higher-level quantities, store on disk?

# Module-2 experience

- Data transfer to NERSC
- Producing DQ plots on DAQ machine and transferring to NERSC
- Converting raw data to packet format
- Running `module0_flow` (thanks to Stephen; see “`add-module2-yamls`” branch)



# Summary

- A working DQM exists (largely thanks to Peter), has proven useful for single-module tests. Plans/needs:
  - More plots
  - Ability to consume higher-level calibrated/reconstructed data
  - Dashboard
- Need collaboration between FNAL and NERSC on data management:
  - Metadata: MetaCat hosting/replication, schema
  - Rucio: Integrating NERSC as a storage element

Backup

# Processing/calibration steps

- Convert raw data to packet format
- Make low-level DQ plots
- Run flow
- Make mid-level DQ plots
- Run mlreco, Pandora
  - + MINERvA?
- Make high-level DQ plots