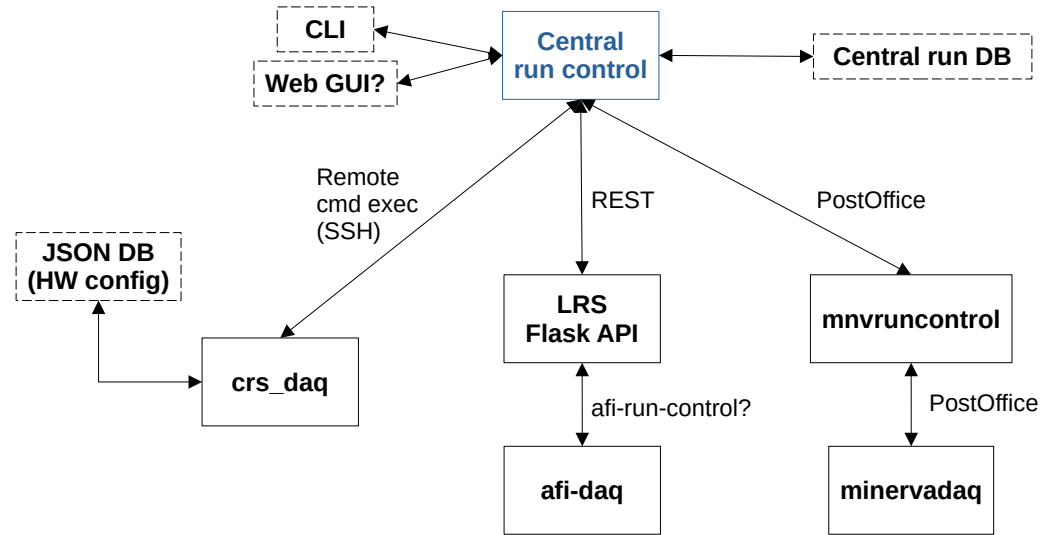# Run control and data flow

Matt Kramer

2x2 DAQ/computing meeting
Apr 4 2024

Run control

- No major progress since last update, but more questions:
  - CRS: Upload ASIC configs (tarred, compressed) to UConDB?
  - Start/stop subruns for 3 systems?
    - Already handled by minervadaq?

https://github.com/mjkramer/mx2x2_run_control/

Metadata, data management

- Script added for CRS
  - Should be called at end of run
- Output reviewed by DUNE data mgmt
- Should use .hdf5 extension, not .h5
- First/last "event" can be set to accelerator clock
  - Would assist matching with MINERvA
- Daily data volume:
  - CRS: O(1 GB)
  - LRS: O(100 GB) – ~dozen subruns/day
  - MINERvA: O(5 GB) (?) ~couple subruns/day

```
{
    "name": "binary-2024_03_14_16_44_59_CDT.hdf5",
    "namespace": "neardet-2x2-lar-charge",
    "checksums": {
        "adler32": "50967e92"
    },
    "size": 111407376,
    "metadata": {
        "core.application.family": "larpix",
        "core.application.name": "crs_daq",
        "core.application.version": "2x2",
        "core.data_stream": "commissioning",
        "core.data_tier": "raw",
        "core.file_type": "detector",
        "core.file_format": "hdf5",
        "core.file_content_status": "good",
        "core.start_time": 1710452700.0287526,
        "core.end_time": 1710452879.957409,
        "core.run_type": "neardet-2x2-lar-charge",
        "core.runs": [
            1234
        ],
        "core.runs_subruns": [
            12340001
        ],
        "core.first_event_number": 98765,
        "core.last_event_numer": 99887,
        "retention.class": "commissioning",
        "retention.status": "active"
    }
}
```

Transfers to NERSC

- Need low latency streaming transfers for nearline processing
  - Can't use Rucio for this (?), but still need Rucio for data mgmt beyond nearline ops

- Currently running "rsync in a loop":
  - while true; do rsync -avhP --chmod 777 /data/commission/April2024 dtn02.nersc.gov:/global/cfs/cdirs/dune/www/data/2x2/CRS/commission; sleep 60; done
  - "rsync2nersc" screen session on acd-daq02 (acdaq account)
  - Data accessible via NERSC web portal
    - https://portal.nersc.gov/project/dune/data/2x2/CRS/

- Need to provide shifters with instructions on monitoring

Nearline processing at NERSC

- "Watchdog" process at NERSC continuously monitors filesystem for incoming data

- Processing tasks are queued for new data
  - Currently just conversion from "binary" to "packet" format
  - Next: ndlar_flow processing, plot generation

- "Worker" script repeatedly pulls and executes tasks
  - Can run arbitrary number of workers

- Working on using filesystem notifications instead of scanning

- Need to add ability to continuously reprocess data for a run in progress
  - Expect long (day+) runs for CRS to avoid small files

Nearline underground

- Plan: Redundant nearline processing on acd-daq02 (CRS DAQ) or acd-daq04 (spare)
  - If acd-daq04, need to rsync (or use NFS, or fuse-sshfs)
- Required infrastructure: MongoDB
  - Can run in a Podman
- Run one watchdog and N workers

Summary

- Much work remains on central run control
  - For now, must continue manually starting/stopping the 3 systems
- Pipeline is running for low-latency transfers to NERSC, prompt processing
  - Underground nearline in process
- Able to generate metadata for charge files
  - Generate it, point ingestd to xrootd, watch the magic happen