Evolution of DUNE's Production System

Jake Calcutt, on Behalf of DUNE

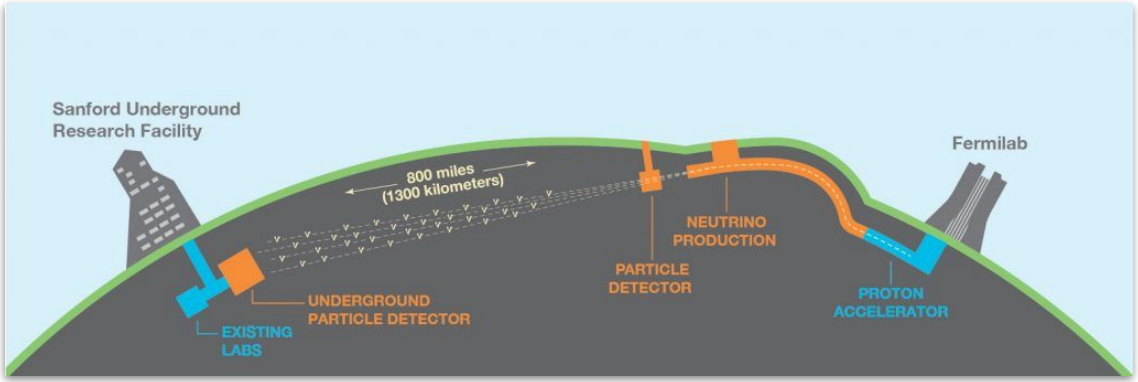
CHEP2024

October 24, 2024

Brookhaven National Laboratory

DUNE

Deep Underground Neutrino Experiment (DUNE)



Upcoming, next-generation long baseline neutrino experiment

First LArTPC Far Detector (FD) module expected in 2029

Near Detector (ND) complex in 2030

Physics Program

Long Baseline Physics

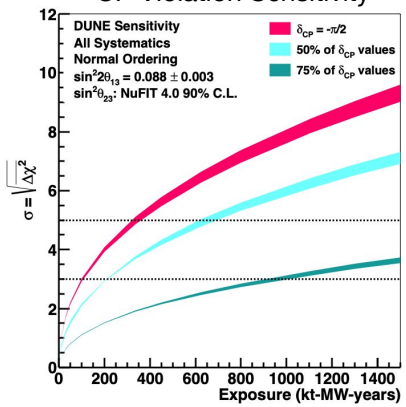
- Determination of CP Symmetry Violation
- Neutrino Mass Ordering

Solar Neutrinos

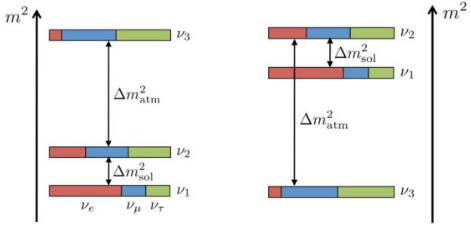
Supernova Burst (SNB) Detection

Beyond Standard Model Physics

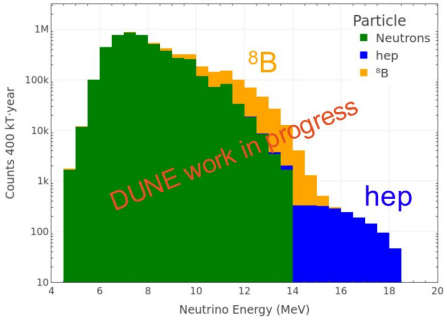
CP Violation Sensitivity



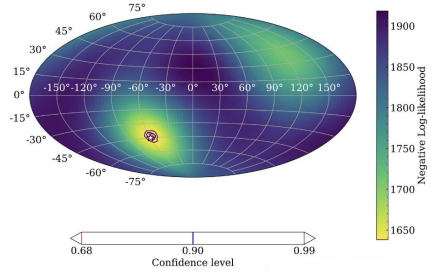
Neutrino Mass Ordering



Reco solar ν_e spectrum in DUNE



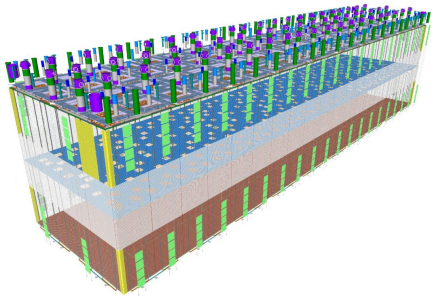
Supernova Pointing



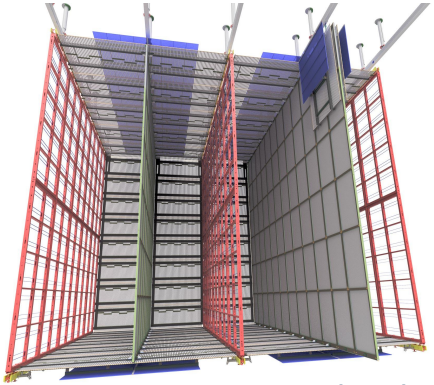
DUNE Detectors

LArTPC Far Detector Modules

- 10kt active LAr each
- HD: 384k readout channels, 2.6ms readout time
- VD: ~492k channels, 4.25ms readout time
- 100s extended readout for SNB



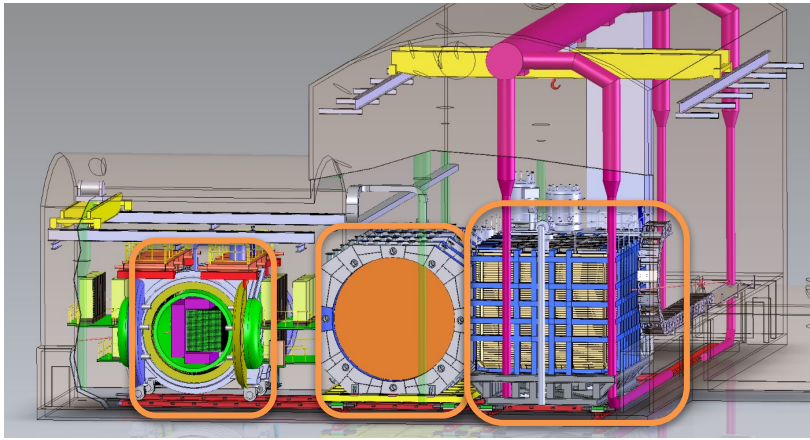
Vertical Drift (VD) FD Module



Horizontal Drift (HD) FD Module

Near Detector Complex

- ~12 million channels in ND-LAr
- Intense flux from beam → Large amounts pileup → Complicated reconstruction

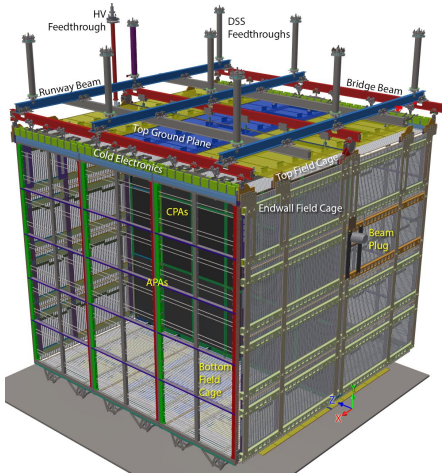


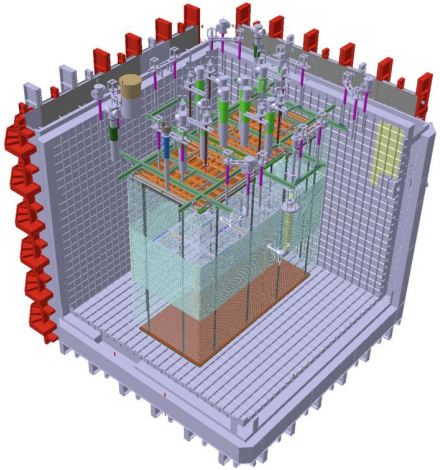
Near Detector complex with SAND, ND-GAr, ND-LAr

DUNE Prototypes

Several prototypes for FD at CERN

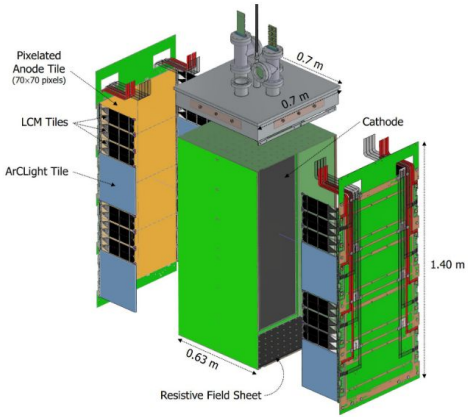
- Left: ProtoDUNE-SP
  - Now reconfigured and known as ProtoDUNE-HD (PDHD)
- Right: ProtoDUNE-VD (PDVD)

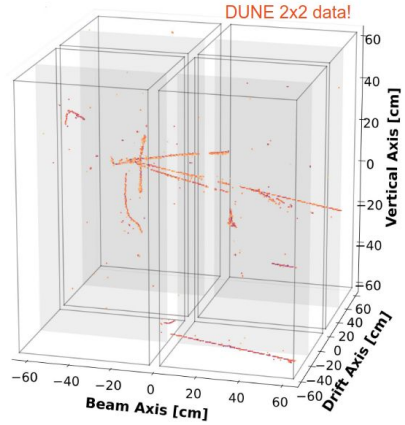




ND Prototype at FNAL

- 4 ND-LAr TPC modules (left) arranged 2x2 (right)
- Took data with FNAL NuMI beam this past summer





Processing & Storage Needs – Data

Detector	Processing Time (Event*)	Number of Events (Year)	Processing Time (Year*)	Raw Readout Size (Year)
FD HD	1.33 CPUh	2.2 million	2.93e6 CPUh	9.4 PB
FD VD	1.25 CPUh	2.2 million	2.75e6 CPUh	16 PB
ND	0.75 CPUh	15 million	11.3e6 CPUh	16 PB

~17 million CPUh / yr for processing new data

- Reprocessed every 3 years

*For normal, non-SNB readout

Intense Requirements

Evolution of DUNE's production system driven by:

- Scale of per-event data volume (compared to collider experiments)
- Worldwide-distributed data storage model
- Worldwide compute site usage
- Increased use of GPU
  - Machine Learning (i.e. pattern recognition, fast simulations)
  - Heterogenous workflows (transitioning between HTC/HPC sites)

Legacy Systems

For first two ProtoDUNEs and FD simulation employed legacy systems for distributed processing

- **Sequential Access via Metadata (SAM)** – All-in-one metadata server, replica manager, file deliverer developed for D0 & CDF
- **JobSub** – User interface/middleware to HTCondor & GlideinWMS

New Systems

Data Management:

- SAM replaced by
  - **Metadata Catalog (MetaCat)** of data file characteristics
  - **Rucio** for distributed storage and replica management

Workflow:

- **justIN*** – New workflow system developed at UK DUNE institutions
  - Job submission
  - Efficient allocation of resources considering
    - Job needs
    - Network connections between storage and compute sites

*See Andrew McNab's Poster

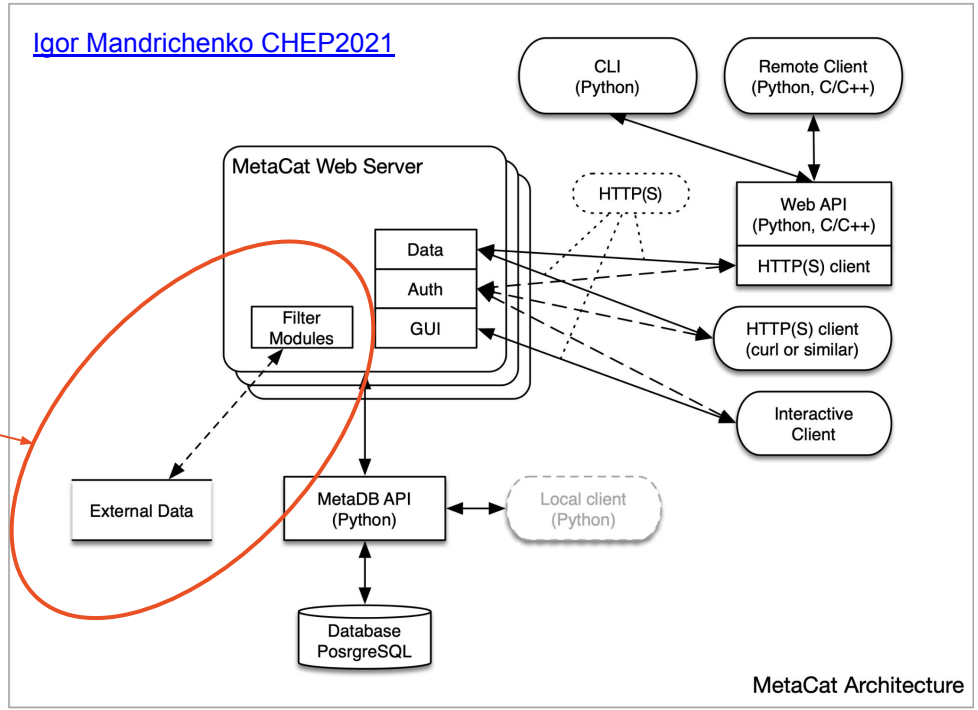
MetaCat

Database associating **files** and **datasets (collections of files)** with **metadata**

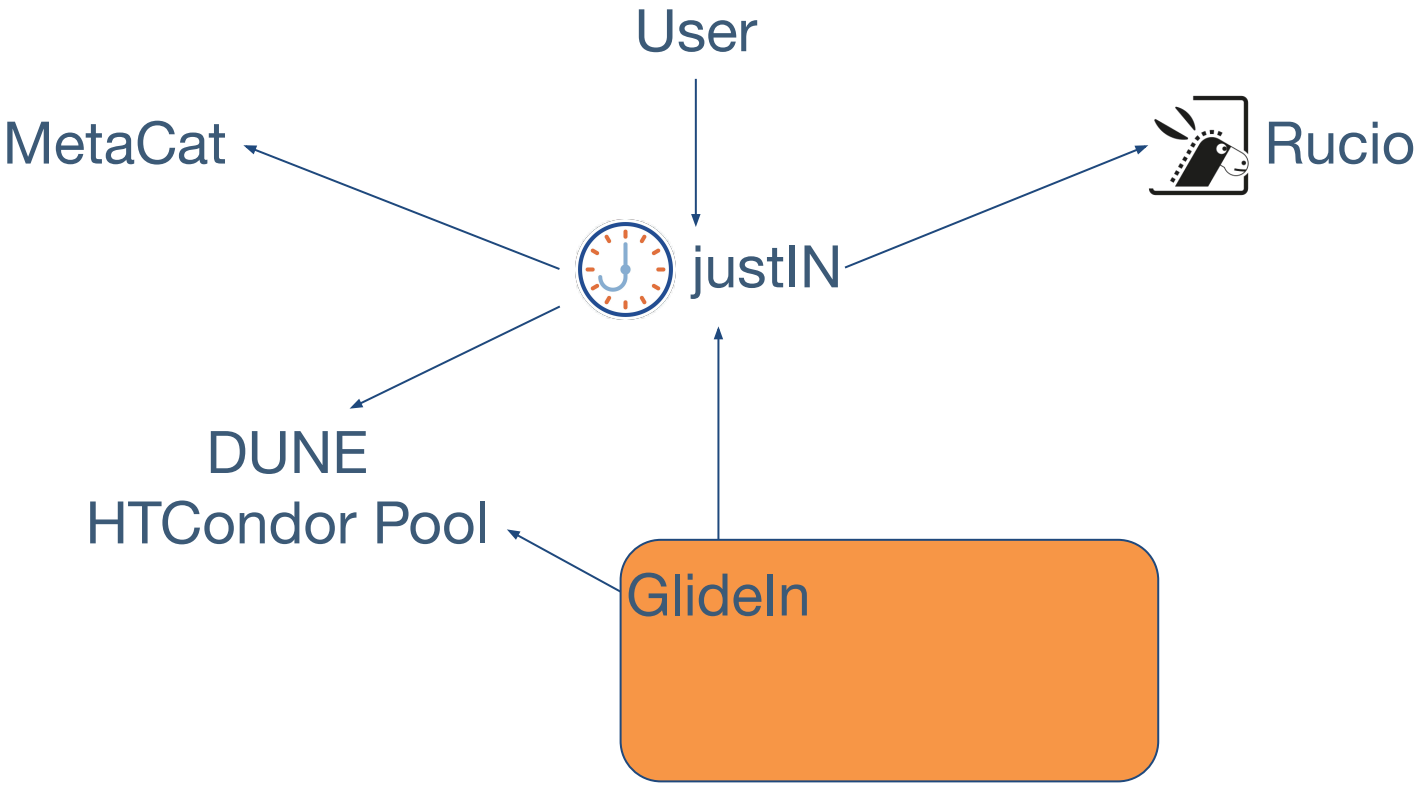
User submits query written in Metadata Query Language (MQL) to find interesting files

Can interface with external databases to extend available information

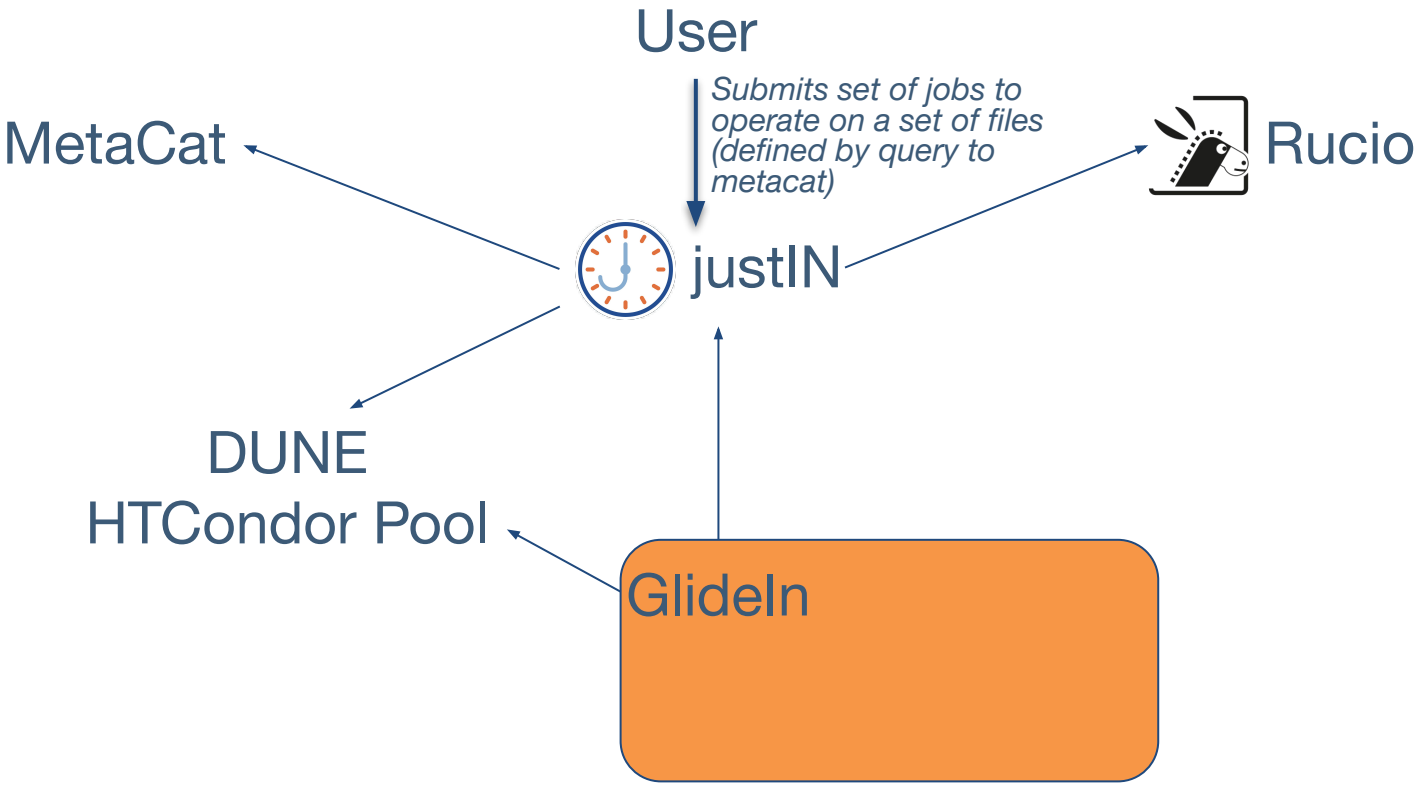
- I.e. Runs Conditions Database (see Nilay Bostan's poster)



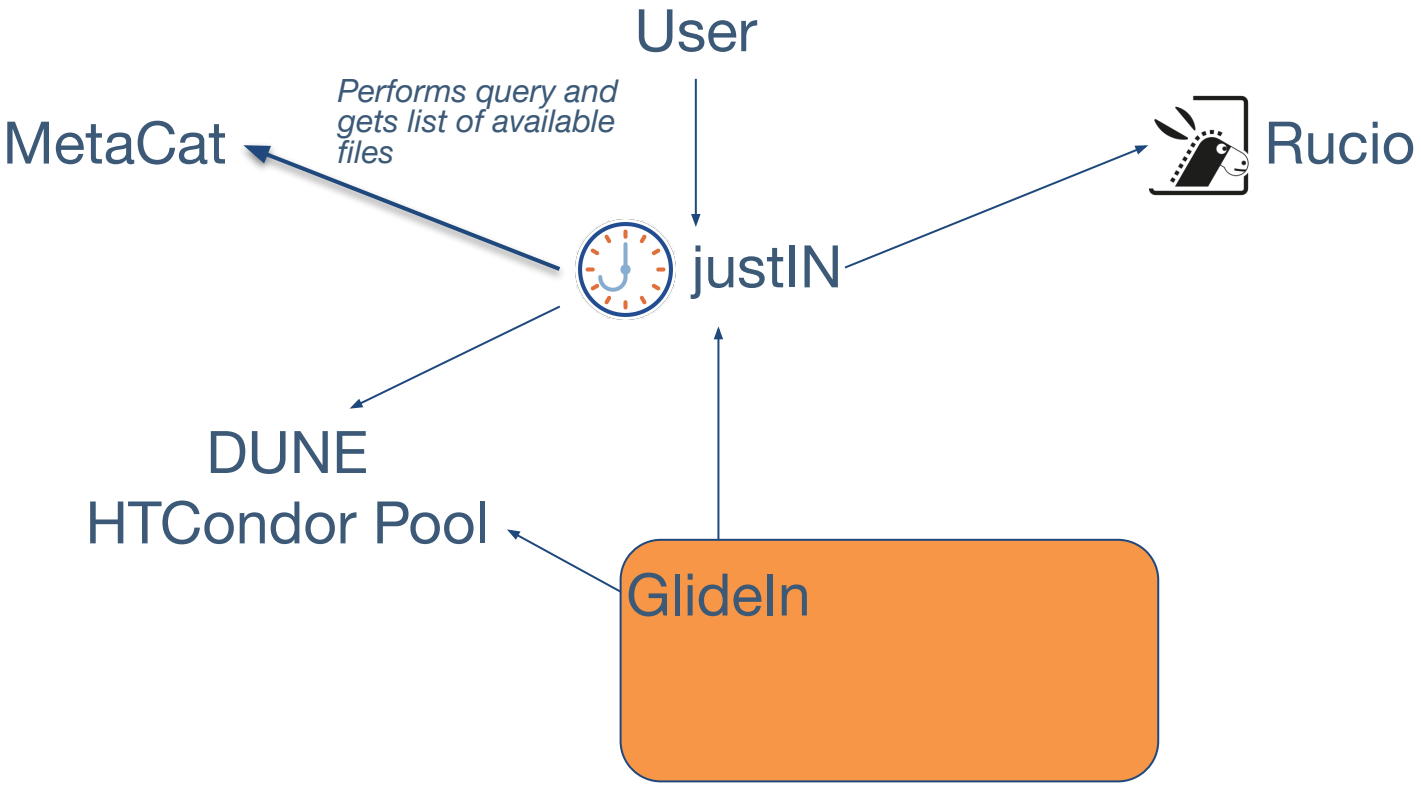
New Systems – Integration



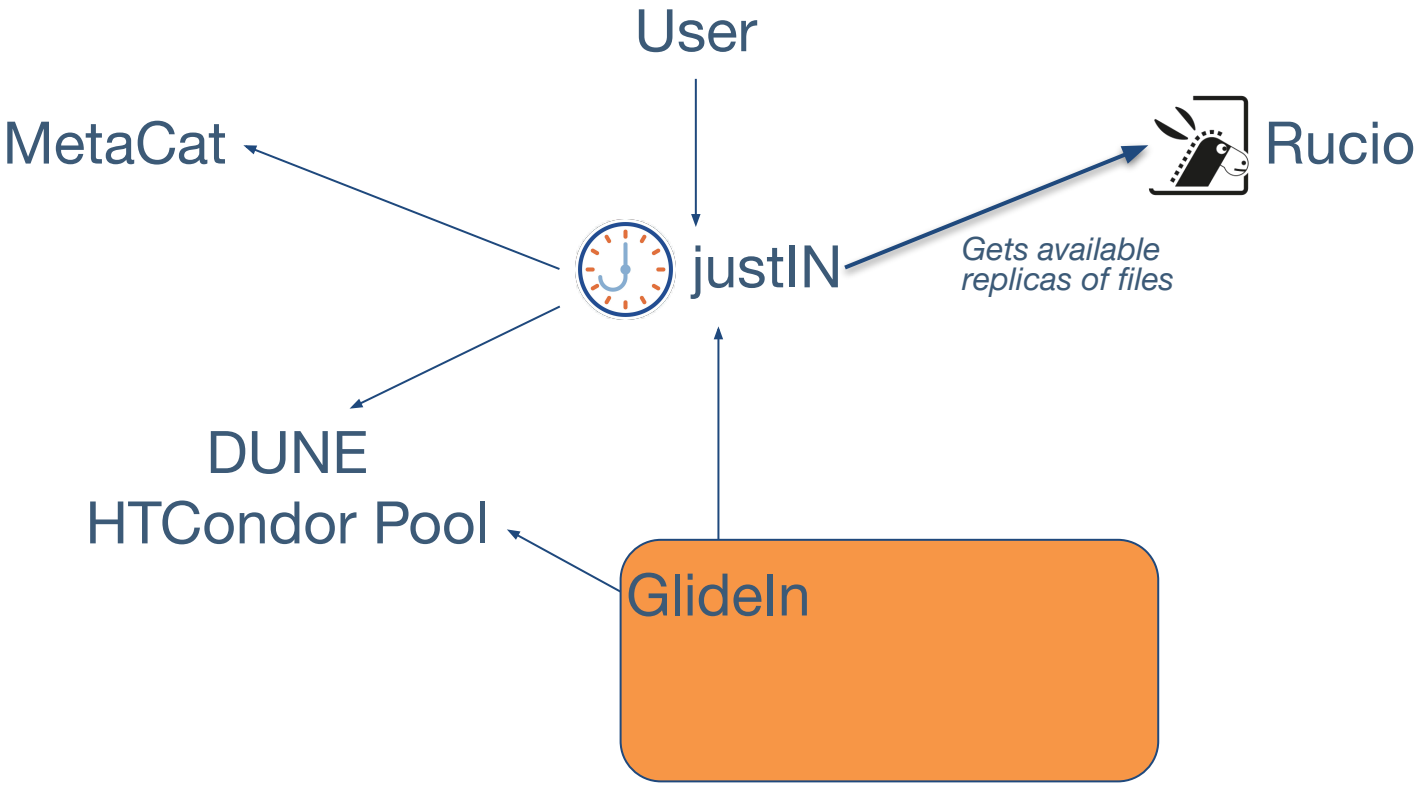
New Systems – Integration



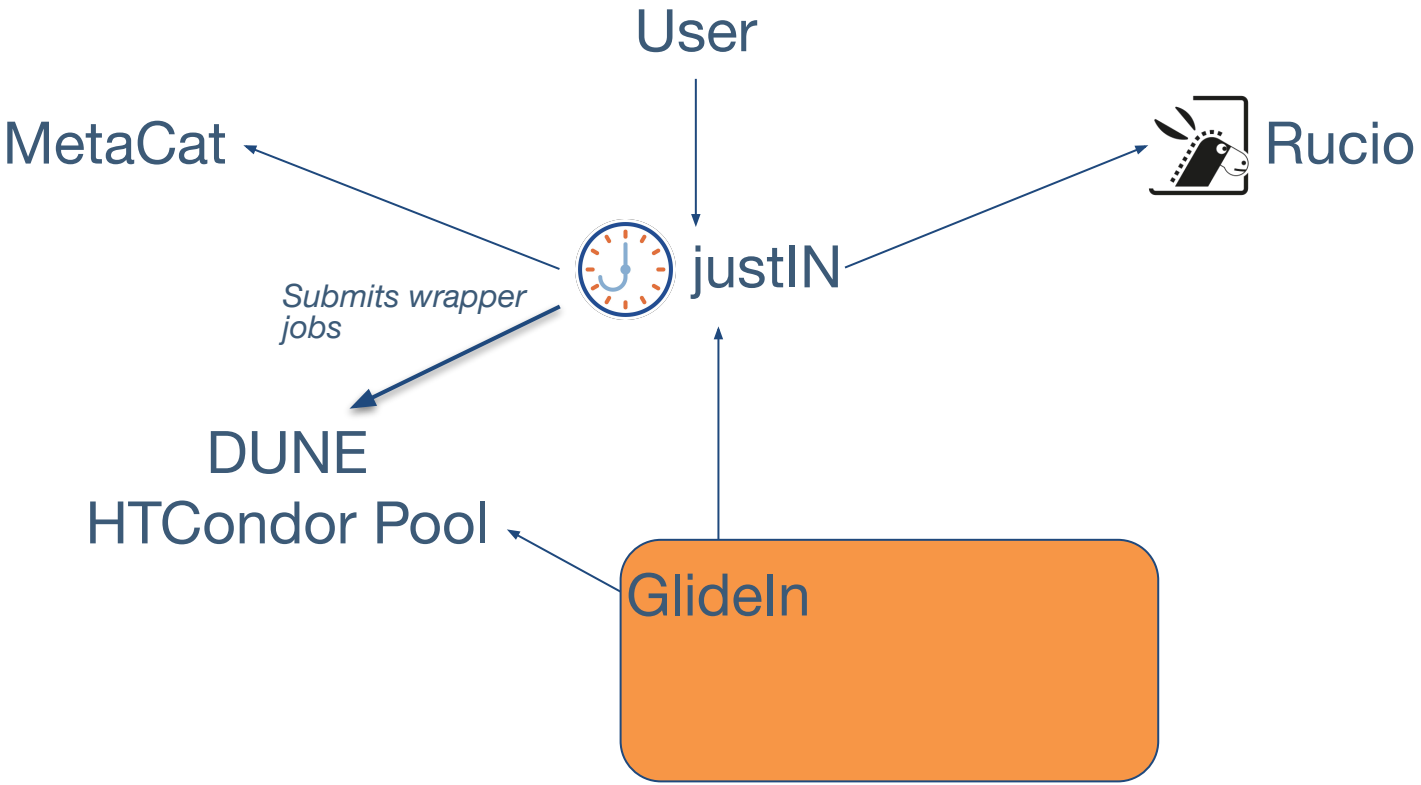
New Systems – Integration



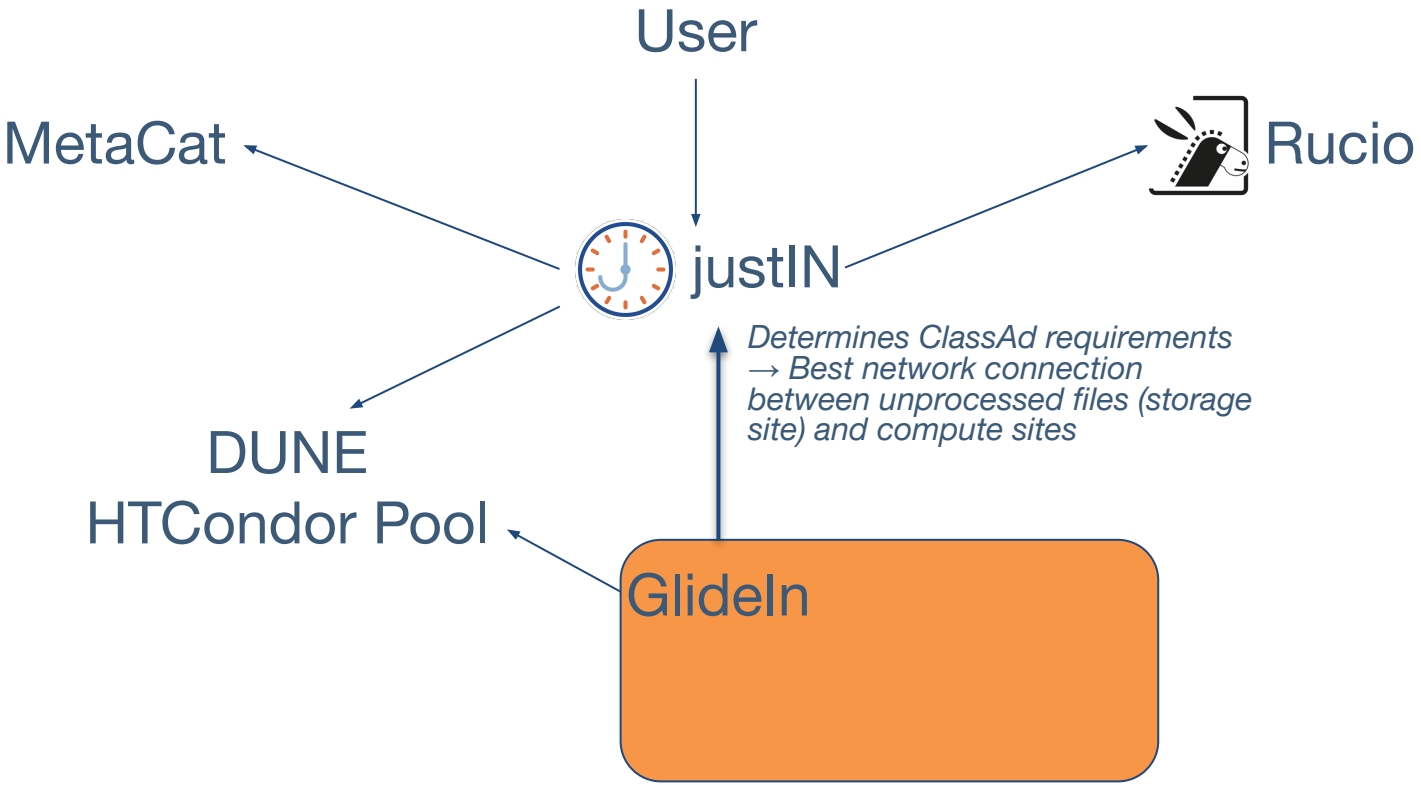
New Systems – Integration



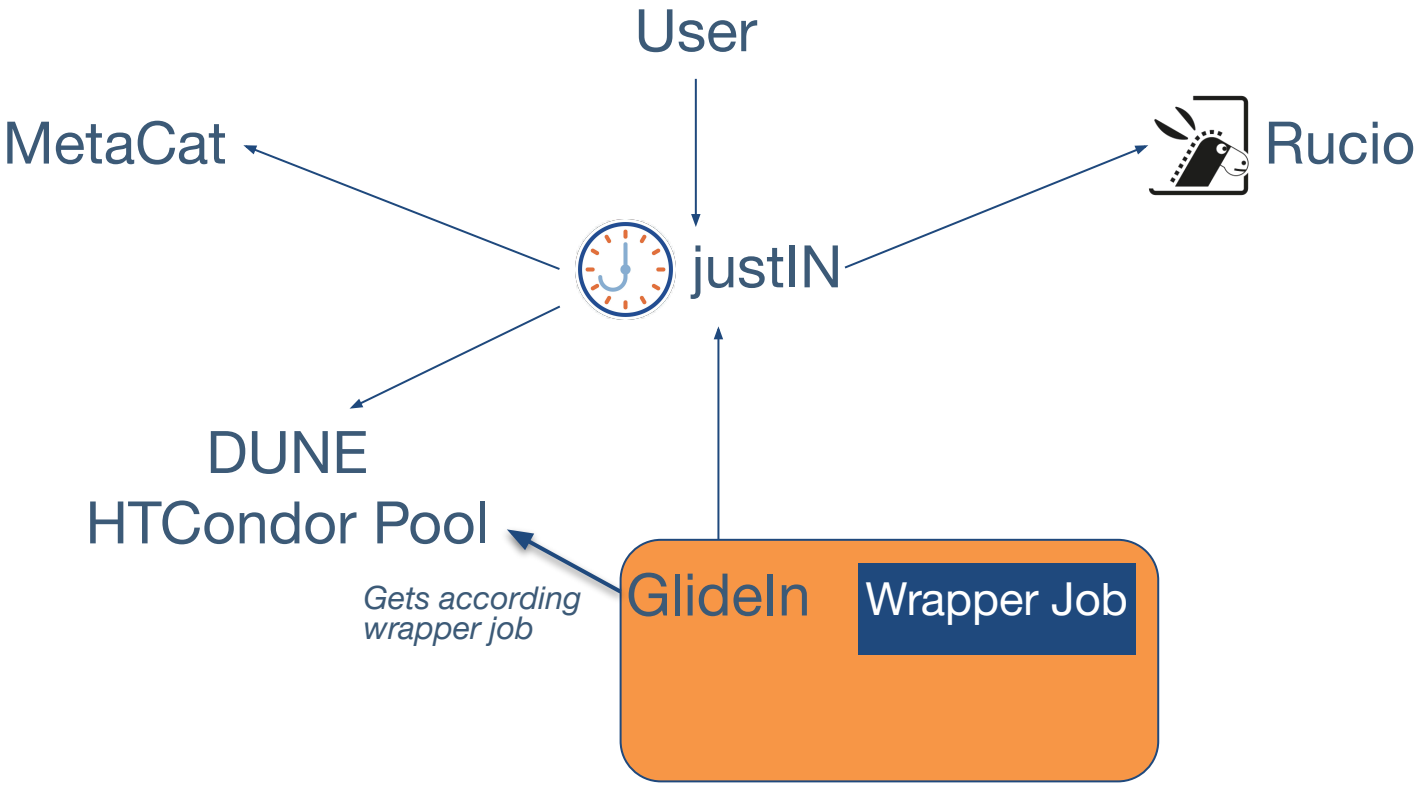
New Systems – Integration



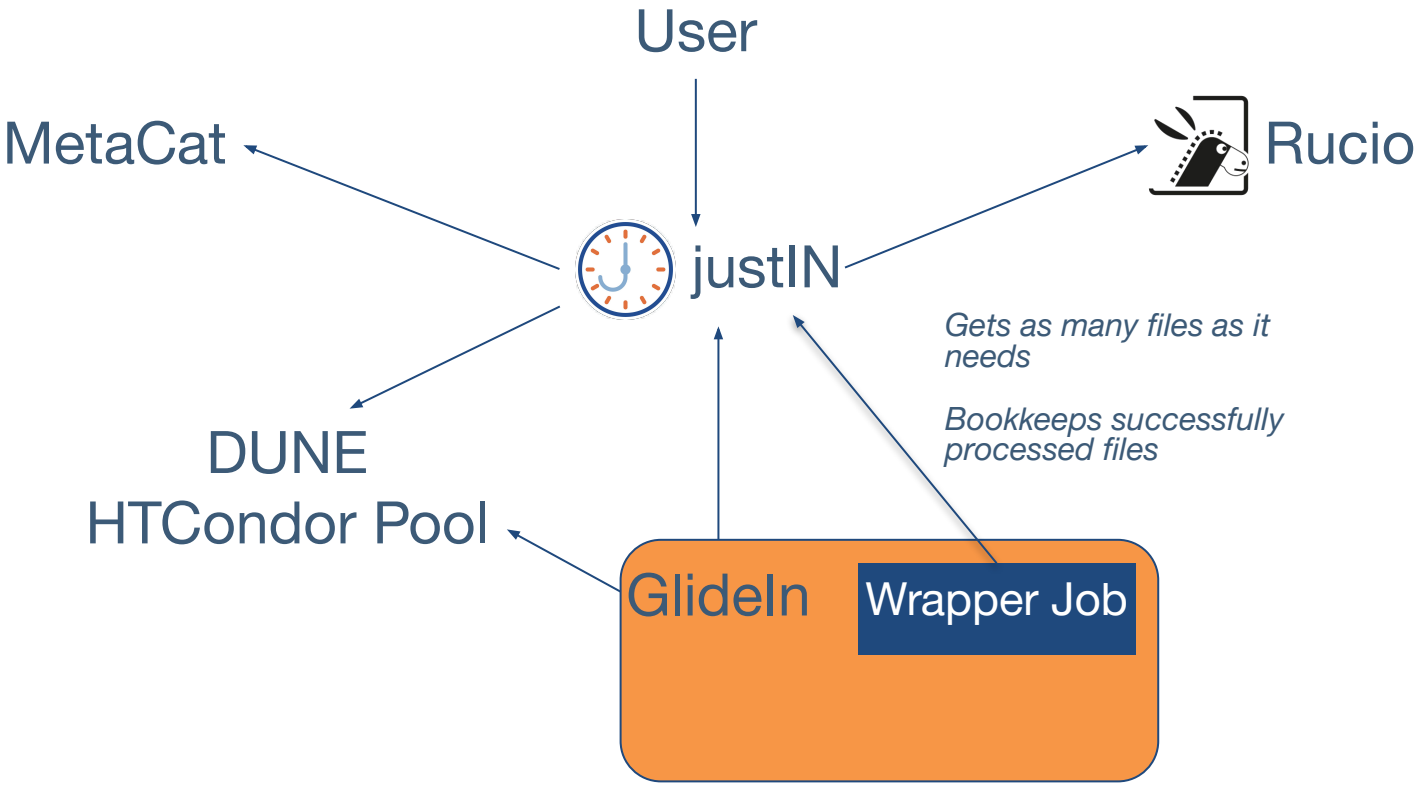
New Systems – Integration



New Systems – Integration



New Systems – Integration



Recent & Ongoing Productions

ProtoDUNE-HD data processing

- First real data production using justIN
- Twice-daily cron jobs submit requests to justIN
  - Query MetaCat for newly-produced files coming from detector
    - Periodic searches of new files matching query
- Bookkeeping & logging done by hand
  - Also by-hand resubmissions if site issues prevent jobs from finishing

Several simulation campaigns

- PDHD, PDVD, second-pass FD reco., low-energy FD sample

2024 WLCG Data Challenge

- First validation of justIN operation

Future Developments

justIN Campaigns

- Multi-stage workflows
- Periodic submissions to process newly produced data
  - As opposed to user-written cron jobs
- Automated logging of requests & submissions

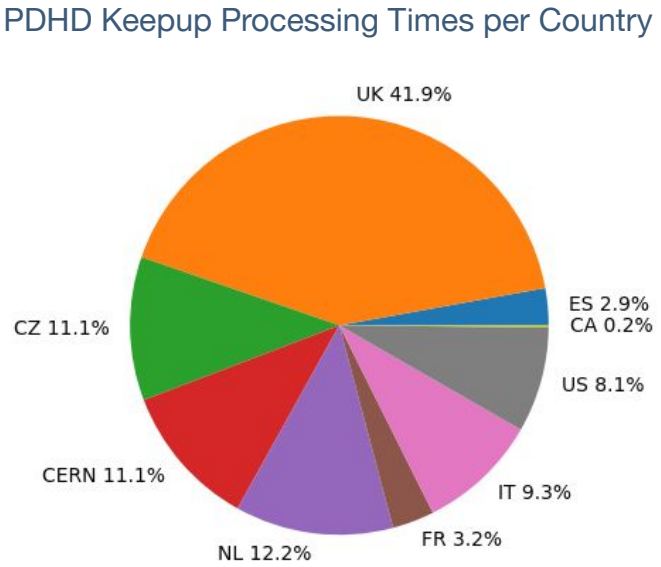
Improved monitoring & reporting of production

Final Thoughts

DUNE's computing model relies on global resources & operation

We welcome new partnerships for infrastructure

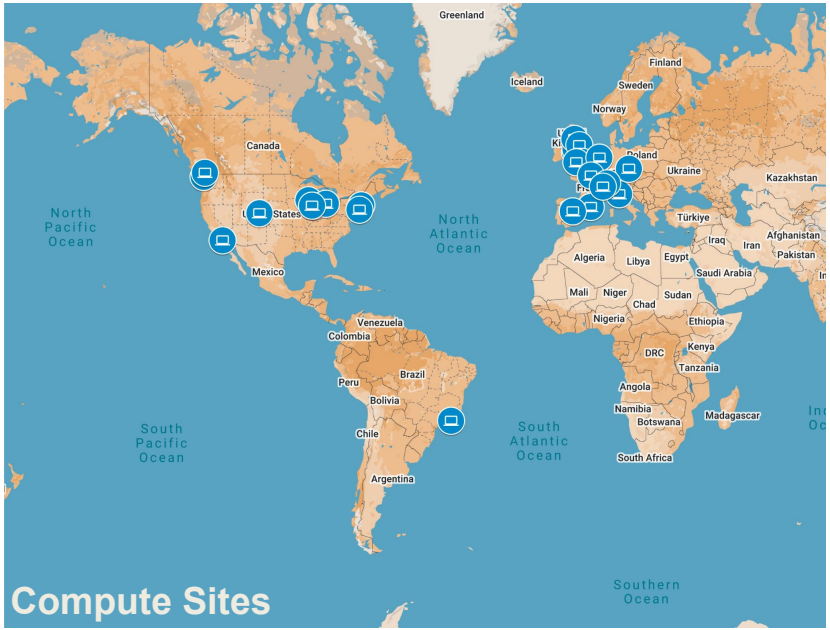
- Weekly operations meeting to coordinate between computing/storage maintainers and DUNE computing/physics

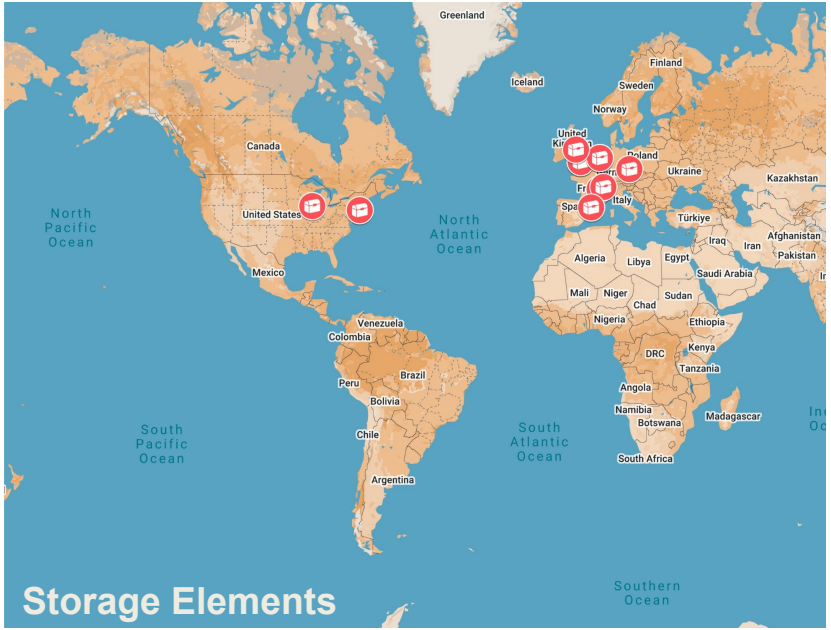


Thank you!

Backup Slides

Global Distribution – Sites & Storage





Near Detector Prod. System

ND production developed in parallel to FD system

- Larger direct use of GPU for det. sim. and ML-based reconstruction
- Multi-step heterogenous (CPU/GPU) workflow
- Direct submission of multi-node/multi-pilot jobs using Slurm & Fireworks for workflow management

Processing & Storage Needs – Data

Data Rate (expected)

- Far Detector: 30 PB/yr
- Near Detector: 250 TB/yr

Processing Time (CPU only)

- Horizontal Drift FD: 1.33 hr/event
  - Expect 2.2 million events/yr → **2.93 million CPUh/yr**
- Vertical Drift FD: 1.25 hr/event
  - Expect 2.2 million events/yr → **2.75 million CPUh/yr**
- Near Detector: 0.75 hr/event
  - Expect 15 million events/yr → **11.3 million CPUh/yr**
- **Total: ~17 million CPUh/yr**
  - Just for data reconstruction

Rucio

Distributed Data Management Service

- Originally developed for ATLAS
  - Now used by many scientific collaborations
- Unified frontend to heterogeneous storage infrastructures
- Rules-based management to automatically distribute data

Now employed by DUNE. See Wenlong Yuan's talk on DUNE's Rucio monitoring