We have PetaBytes of data you should really be looking at, and here's how

Ken Herner for the production group
19 October 2018

What is the production group doing?

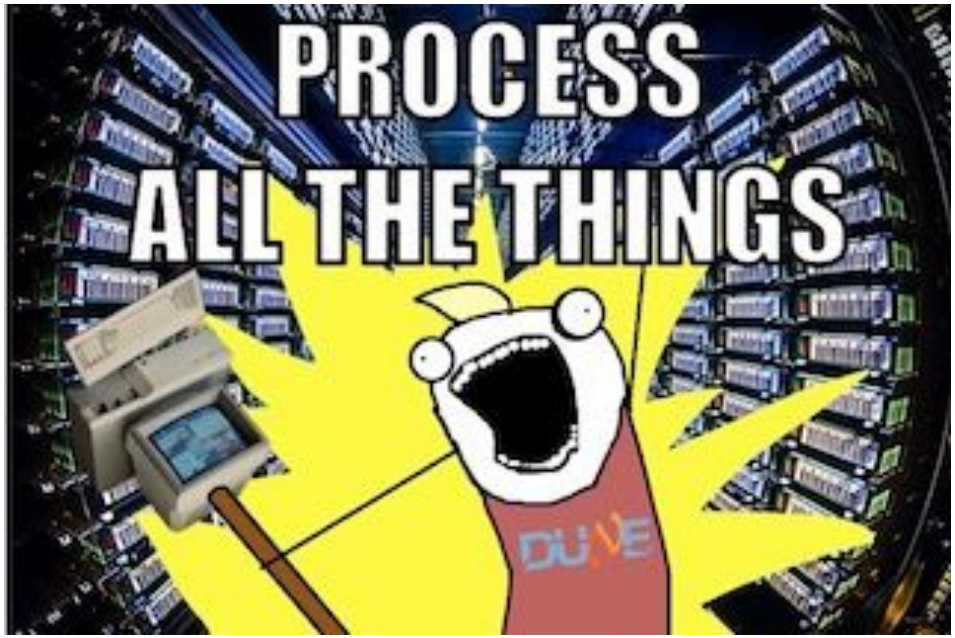
Two main efforts right now: protodDUNE data keepup processing (first priority) and MC production

Keepup: run raw data decoding and hit reconstruction (not tracking yet; will be added soon).

protoDUNE SP by the numbers

as of 18 Oct 2018

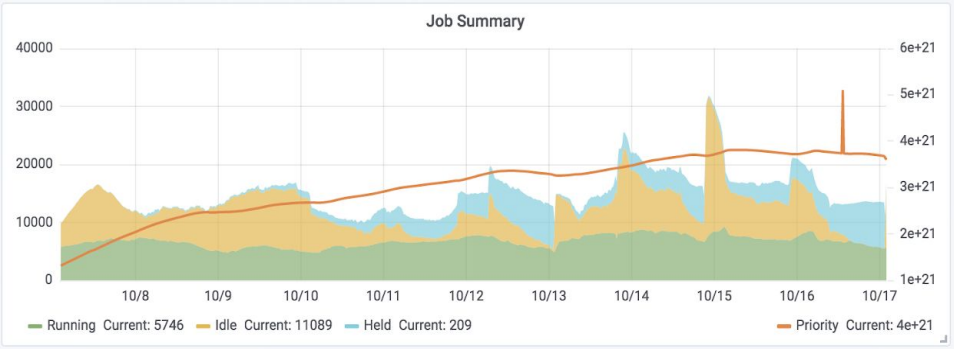
Raw data: 1319 TB (1.3 PB)
Raw “physics” data: 490 TB
Raw physics data fraction processed: 72.3% (now mostly processing Oct. 11-12)
Total hit reconstruction output size: 206 TB

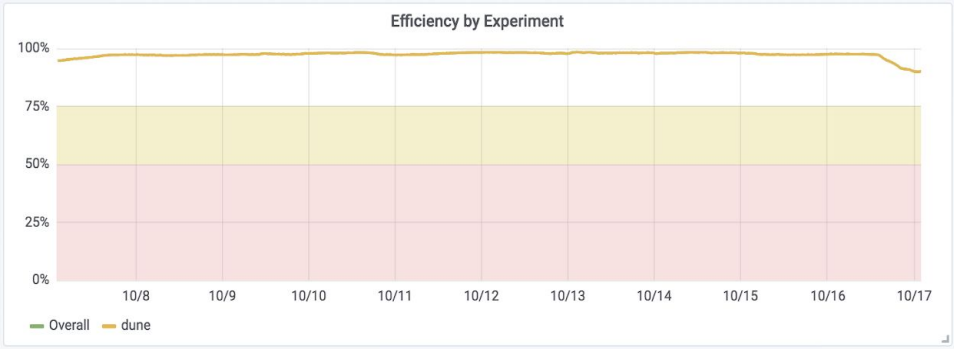


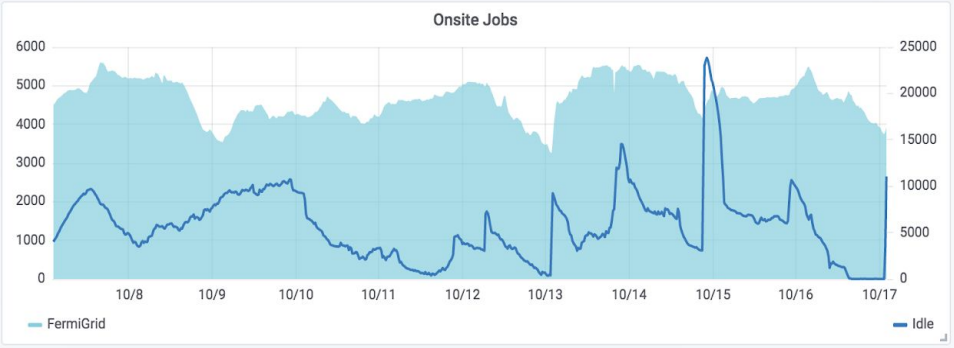
protoDUNE SP Keepup Processing

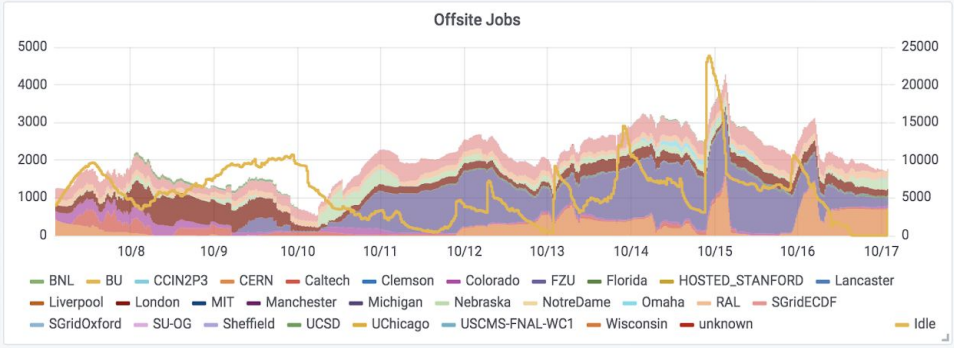
Processing all kinds of runs (test, noise, physics), so far in rough chronological order

Able to get lots of jobs running, but reconstruction takes a while for physics runs







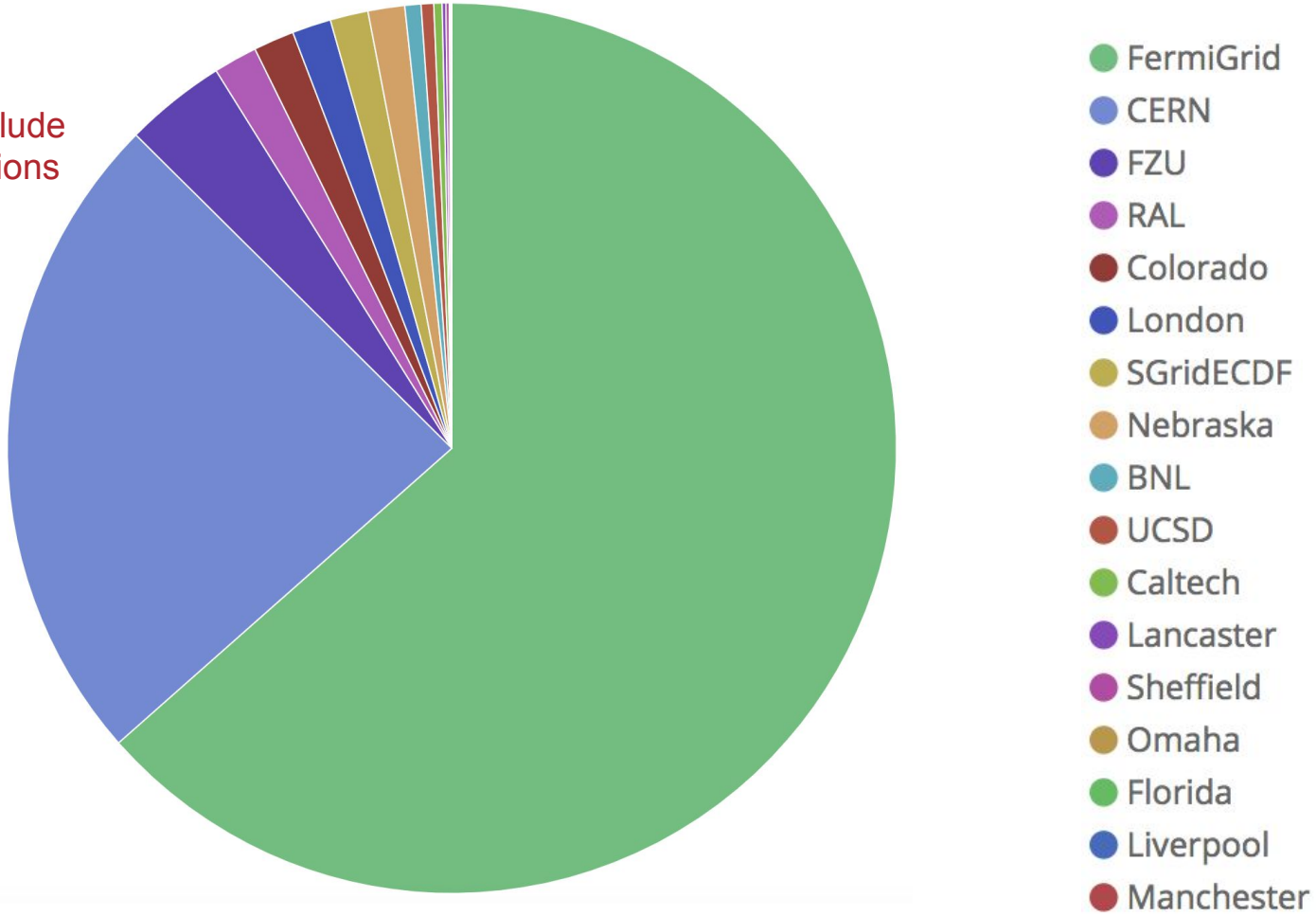


Running about 6500 1-core, 3500MB jobs continuously (~4500 FermiGrid, ~2000 offsite) with some higher peaks

Global contributions to keepup

Thanks to all institutions that have supplied resources!

N.B. does not include MCC11 contributions



Want to see your institution on this list? We do, too! Contact the computing, data management, or production coordinators.

What's in the keepup outputs

Keepup runs the raw data decoder and reconstruction through hit finding. Using dunetpc v07_06_00 so far

Saves outputs to separate data tiers (decoded-raw and hit-reconstructed) in SAM in separate files, accessible by specifying the data tier in SAM queries

Decoder outputs: ssp, tpc, timing raw decoder

Hit reco outputs: gaushit outputs, space point solver

Analyzers can start from these files (e.g. for tracking studies); no need to run expensive hit-finding (several minutes per event) many times

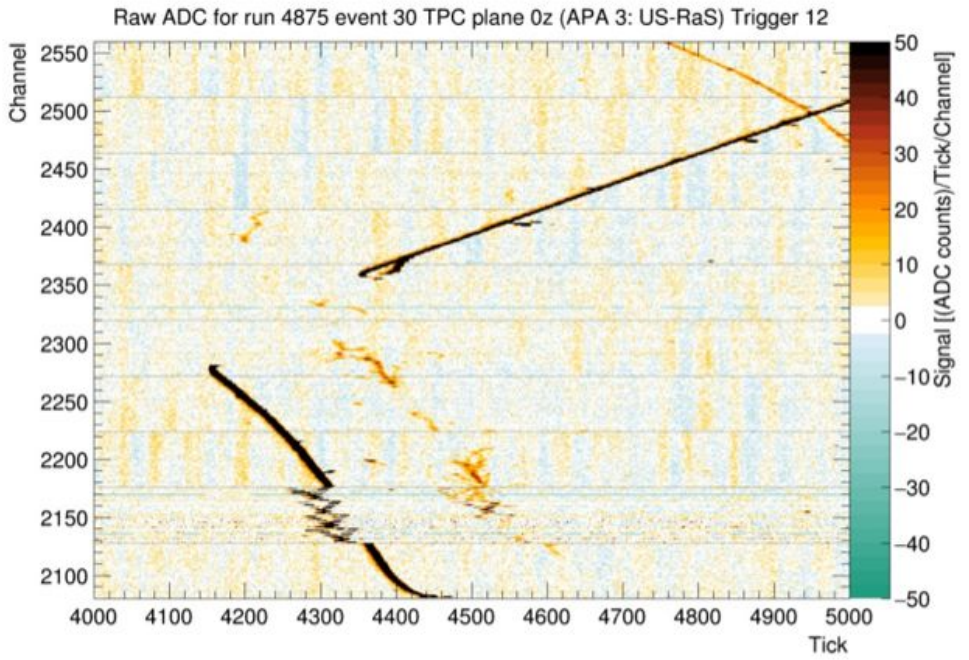
Not saving the artdaq fragments in outputs (saves a factor of ~2 in space)

Known issues and caveats

No tracking yet

dunetpc v07_06_00 does not include a fix for FEMB 302

Nevertheless, still a lot of information in these keepup outputs. The FEMB may not be an issue, depending on what you're doing.



How to access the data

dune-data.fnal.gov is the main portal

See what's available at http://dune-data.fnal.gov/data/protodune-sp/index.html

Protodune-SP (NP04) Data

Updated 18 Oct 2018

SAM Dataset Name	SAM Query Links	Comment
np04_hit-reconstructed_v07_06_00_physics	describe, summary, files	
np04_decoded-raw_v07_06_00_physics	describe, summary, files	
np04_hit-reconstructed_v07_06_00_allruns	describe, summary, files	
np04_decoded-raw_v07_06_00_allruns	describe, summary, files	
np04_raw_physics_online_good_runs	describe, summary, files	
np04_cryostat_commissioning	describe, summary, files	
np04_coldbox_test	describe, summary, files	
np04_coldbox_test_apa3	describe, summary, files	
np04_coldbox_test_apa4	describe, summary, files	
np04_coldbox_test_apa5	describe, summary, files	

Once you have chosen a dataset, follow the instructions at https://wiki.dunescience.org/wiki/Look_at_ProtoDUNE_SP_data

(The instructions work for accessing data from CERN, too!)

Future production plans for protoDUNE data

Next week: Update base release to >=v07_07_01 to pick up FEMB patch and a speedup to hit finding. Use no-beam week to begin running over data in current beam period (since Oct. 10; physics runs only). Data during final beam period (Oct. 31- Nov. 11) would use this new release only

Take a look at adding tracking if time budgets will allow (recent timing results encouraging)

Early December: start full reprocessing of protoDUNE SP data with up-to-date unpacking, hit finding, and tracking. **Need to freeze on this timescale to be complete by Feb. 1.**

Summary

Production group has been very busy the past few months with both protoDUNE data and MC processing.

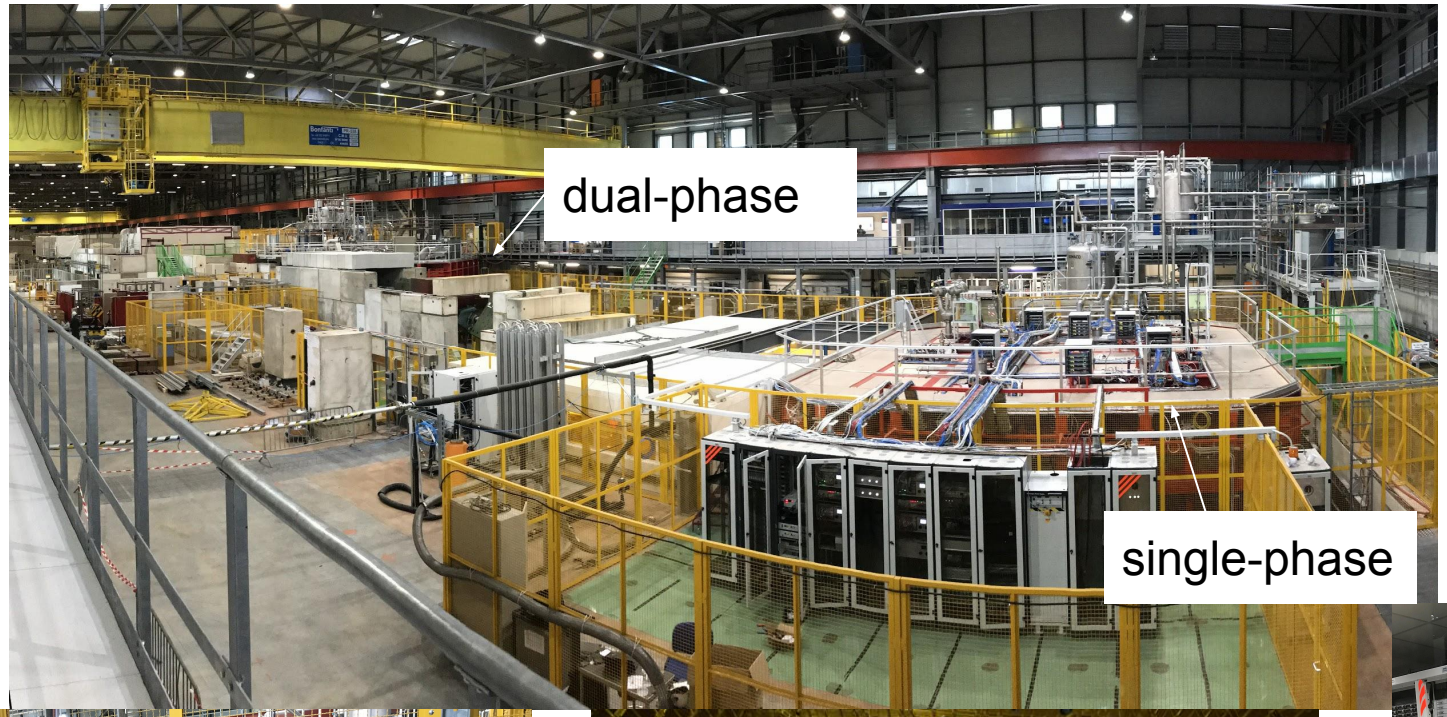
Keepup outputs (raw decoder and hit reconstruction) available to the collaboration; avoids re-running computationally expensive pieces many times

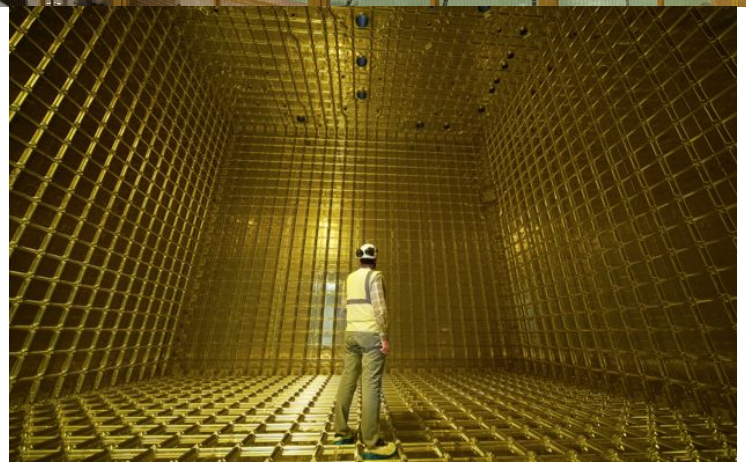
Plan to update release in ~one week and continue running keepup throughout beam run

December: begin full reprocessing and continue over break; finish some time in January

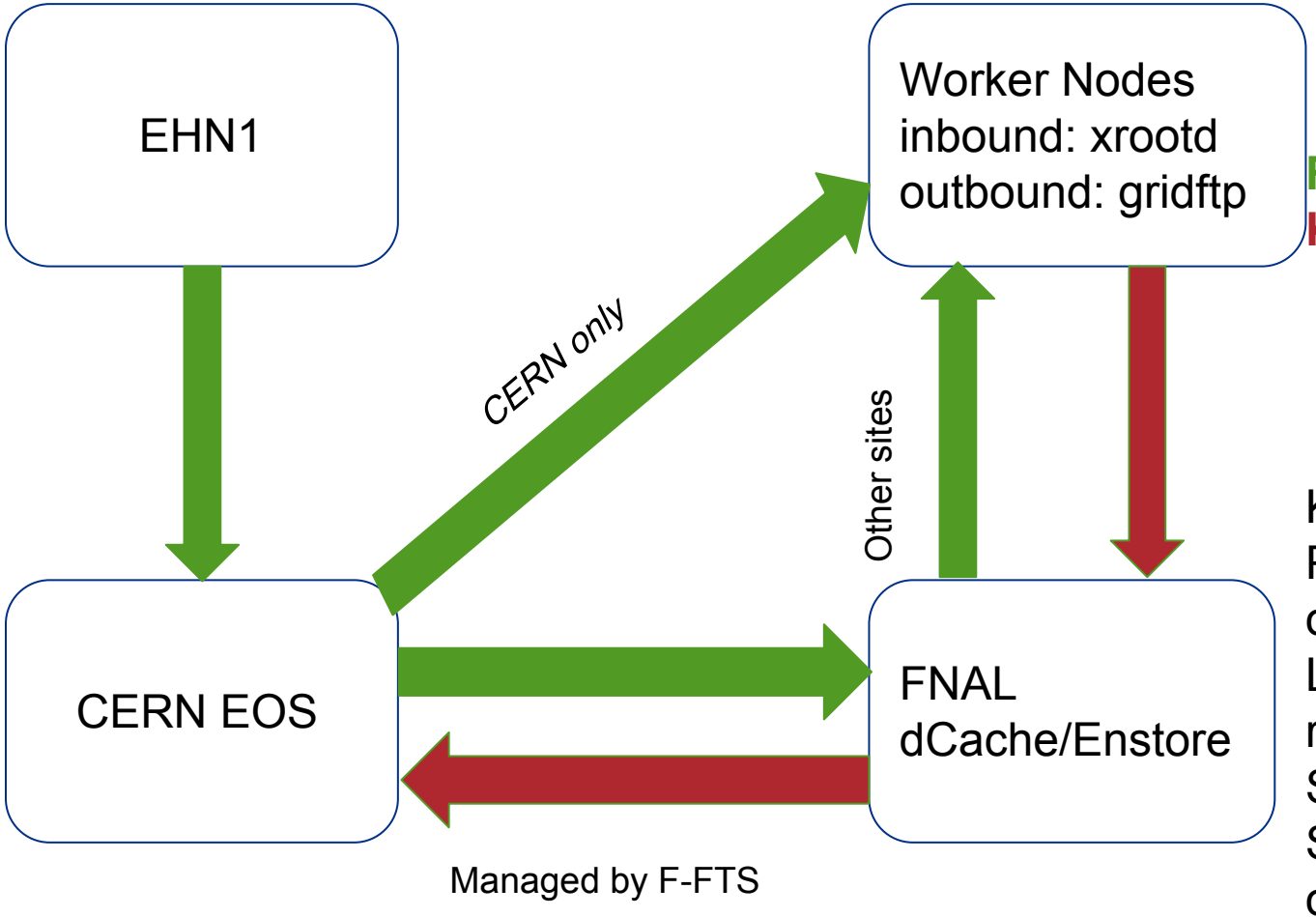
BACKUP

protoDUNE: Inside EHN1



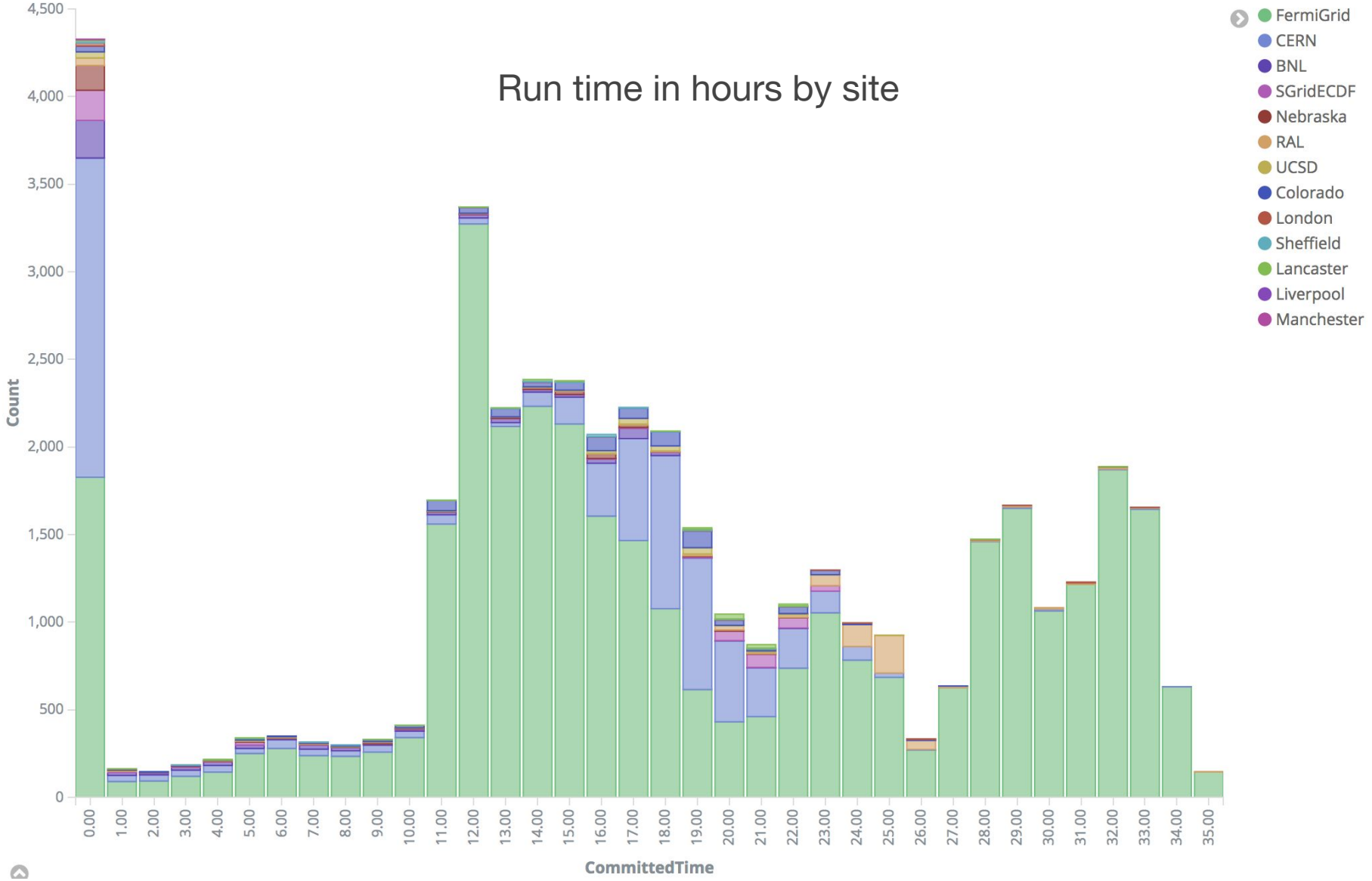


Keepup dataflow

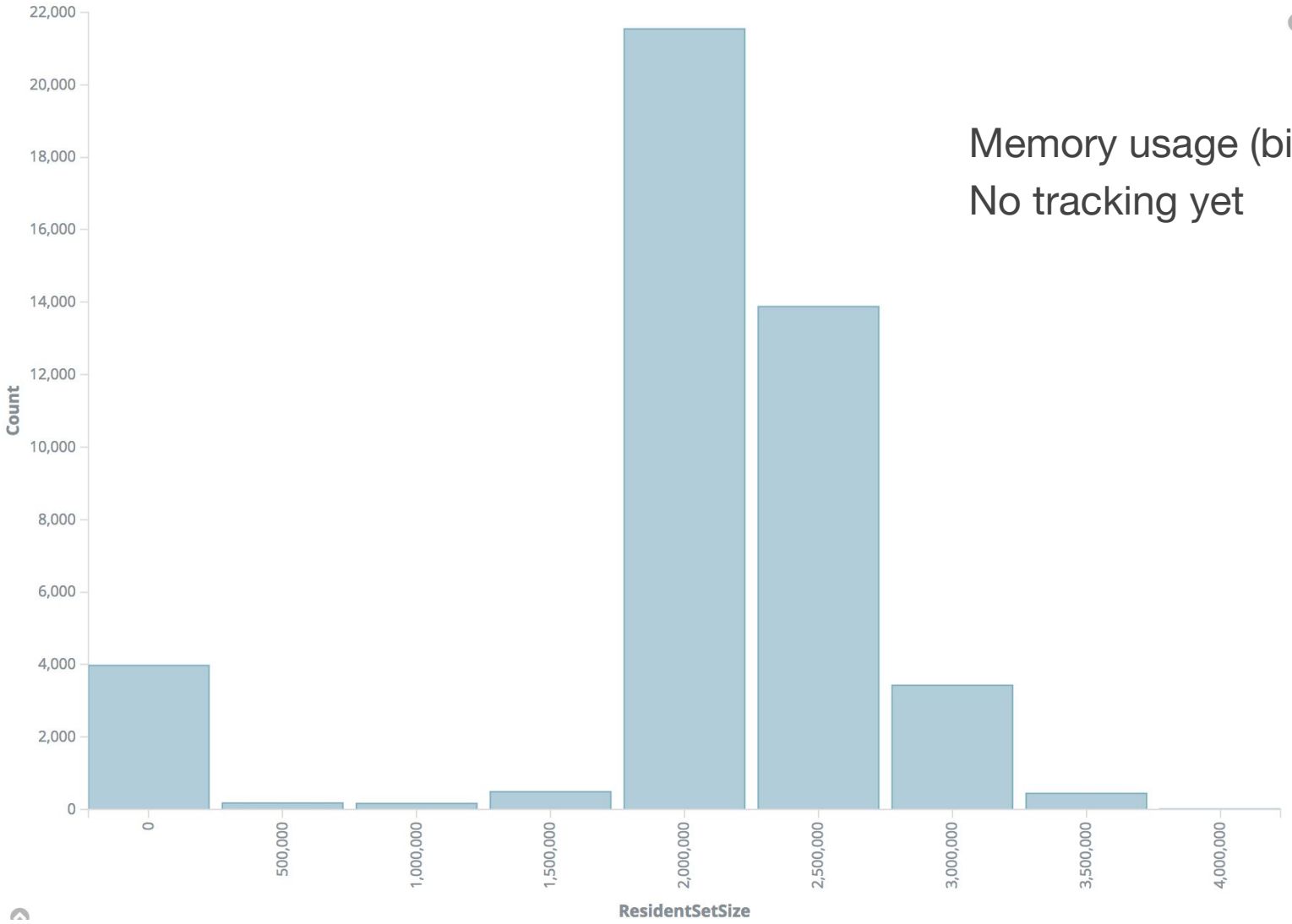


Keepup launches via POMS
on a crontab
Launches occur after raw data show up in SAM
SAM metadata for outputs declared within the job

Keepup Processing (2)



Keepup Processing (3)



Memory usage (bins of 500MB)

No tracking yet