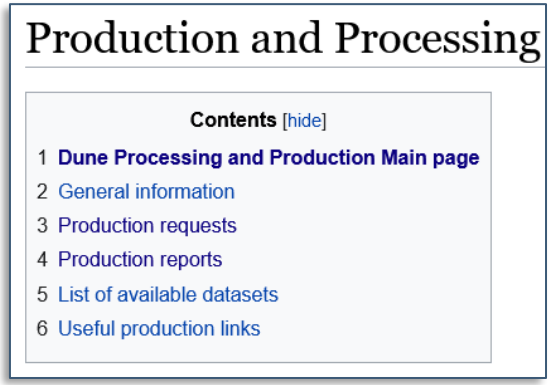
Production report

Aaron Higuera Pichardo, Rice University
Elisabetta Pennacchio, IP2I
Computing Resource Allocation Board Meeting
January 19, 2024

Outline:

1. Production group organization
2. Production campaigns in 2023
3. Plans for 2024

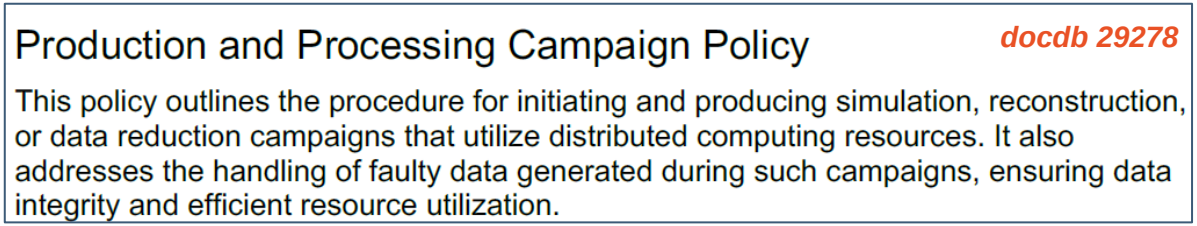
1. Production group organization (Production wiki pages)


Production and Processing

Contents [hide]

1 Dune Processing and Production Main page
2 General information
3 Production requests
4 Production reports
5 List of available datasets
6 Useful production links

- The production group manages production requests from the physics groups
- A set of policies has been developed to optimize the group working processes and to achieve results efficiently


Production and Processing Campaign Policy

docdb 29278

This policy outlines the procedure for initiating and producing simulation, reconstruction, or data reduction campaigns that utilize distributed computing resources. It also addresses the handling of faulty data generated during such campaigns, ensuring data integrity and efficient resource utilization.

- Best practices (based on the experience from past campaigns) are outlined, to use DUNE resources effectively, and obtain results in a timely matter
- Policies are available in docdb 29278. In particular the production request procedure is defined

Informations to be provided in the production request

- Working group leaders must formally request the initiation of a production campaign through the DUNE Computing Service Request/Production.
- Short description of the physics motivations
- Estimated deadline for the delivery of the output datasets
- Code version to be used (it has to be a tagged version and from a DUNE official repo)
- Description of the workflow: list of processing steps (generation, g4, detector simulation, reconstruction, analysis), and fcl file to be used for each step
- Validation sample, a test sample to be used as validation before full production, docdb 29278 "Production Policy"
- For each submitted job, the list of output files to be copied on tape
- Estimation of needed resources : CPU, memory and storage. If possible, it would be helpful to specify the statistic on which these values are based
- List of samples/runs to be processed, ordered by priority.
- Number of events/sample(run)
- Valid metadata
- If you have any questions, please do not hesitate to contact the Production team

Global organization, to optimize usage of DUNE computing resources

Technical aspects

- Once a request is received, an entry is created in the wiki

- A production shifter is appointed

Once a production campaign is finished:

- A dedicated production report summarizing resources usage (CPU and storage) and reporting issues (if any) is prepared and made available in docdb
- Output datasets are created

Output results

for each sample, output files can be listed by querying SAM *samweb list-definition-files (--summary)*:

Example for FD1-HD FD2-VD production

```
FD1 datasets:
1. higuera_fardet-hd__fd_mc_2023a__mc__hit-reconstructed__prodgenie_nu_dune10kt_1x2x6.fcl__v09_78_01d01__preliminary
2. higuera_fardet-hd__fd_mc_2023a__mc__hit-reconstructed__prodgenie_anu_dune10kt_1x2x6.fcl__v09_78_01d01__preliminary
3. higuera_fardet-hd__fd_mc_2023a__mc__hit-reconstructed__prodgenie_nue_dune10kt_1x2x6.fcl__v09_78_01d01__preliminary
4. higuera_fardet-hd__fd_mc_2023a__mc__hit-reconstructed__prodgenie_anue_dune10kt_1x2x6.fcl__v09_78_01d01__preliminary
5. higuera_fardet-hd__fd_mc_2023a__mc__hit-reconstructed__prodgenie_nutau_dune10kt_1x2x6.fcl__v09_78_01d01__preliminary
6. higuera_fardet-hd__fd_mc_2023a__mc__hit-reconstructed__prodgenie_anutau_dune10kt_1x2x6.fcl__v09_78_01d01__preliminary
FD2 datasets:
1. higuera_fardet-vd__fd_mc_2023a__mc__hit-reconstructed__prodgenie_nu_dunevd10kt_1x8x6_3view_30deg.fcl__v09_75_03d00__preliminary
2. higuera_fardet-vd__fd_mc_2023a__mc__hit-reconstructed__prodgenie_anu_dunevd10kt_1x8x6_3view_30deg.fcl__v09_75_03d00__preliminary
3. higuera_fardet-vd__fd_mc_2023a__mc__hit-reconstructed__prodgenie_nu_numu2nue_nue2nutau_dunevd10kt_1x8x6_3view_30deg.fcl__v09_75_03d00__preliminary
4. higuera_fardet-vd__fd_mc_2023a__mc__hit-reconstructed__prodgenie_nu_numu2nutau_nue2numu_dunevd10kt_1x8x6_3view_30deg.fcl__v09_75_03d00__preliminary
5. higuera_fardet-vd__fd_mc_2023a__mc__hit-reconstructed__prodgenie_anu_numu2nue_nue2nutau_dunevd10kt_1x8x6_3view_30deg.fcl__v09_75_03d00__preliminary
6. higuera_fardet-vd__fd_mc_2023a__mc__hit-reconstructed__prodgenie_anu_numu2nutau_nue2numu_dunevd10kt_1x8x6_3view_30deg.fcl__v09_75_03d00__preliminary
```

or MetaCat

Documentation:

Production shifter notes: google doc, DocDB

Production report DUNE-doc-29197

- Official output datasets for all campaigns are created and advertised in a dedicated wiki page

Inventory

Return to Data_Collections_Manager

This data inventory serves as a comprehensive repository that catalogues the latest datasets utilized by the various physics working groups within DUNE that were submitted through our official ticket system and follows our production policies docdb 29278. This structured data pool is maintained to ensure accuracy and consistency, facilitating a seamless flow of information. The inventory is also dynamically updated to reflect the latest request. This is maintained by DUNE's data collection data manager and includes only SAM datasets

DUNE data inventory

Working group	Dataset	Request	Summary	Comments
LBL	higuera_fardet-hd__fd_mc_2023a__mc__hit-reconstructed__prodgenie_nu_dune10kt_1x2x6.fcl__v09_78_01d01__preliminary	RITM1780305	File count:20784 Event count:2078400	Ready (preliminary dataset name), Horizontal drift
LBL	higuera_fardet-hd__fd_mc_2023a__mc__hit-reconstructed__prodgenie_anu_dune10kt_1x2x6.fcl__v09_78_01d01__preliminary	RITM1780305	File count: 19850 Event count:1985000	Ready (preliminary dataset name), Horizontal drift

- A document regarding the creation of official datasets can be found in docdb 29787

2. Production campaigns in 2023 (wiki pages)

- **FD1-HD FD2-VD** production report

  step 1: up to hit reconstruction

  6 neutrino samples for FD1-HD, (3 FHC, 3 RHC), 2M events/sample

  6 neutrino samples for FD2-VD, (3 FHC, 3 RHC), 2M events/sample

  24 M events, output size ~350 TB, total CPU: 2.2 Mh

- **FD1-HD (High energy)** production report

  step 1: up to hit reconstruction

  15 M events, output size ~150 TB, campaign run at NERSC, total CPU: 10879 NERSC core hours

- **ProtoDUNE-SP cosmics reconstruction** production report

  450K events, output size ~40 TB, total CPU: 91Kh

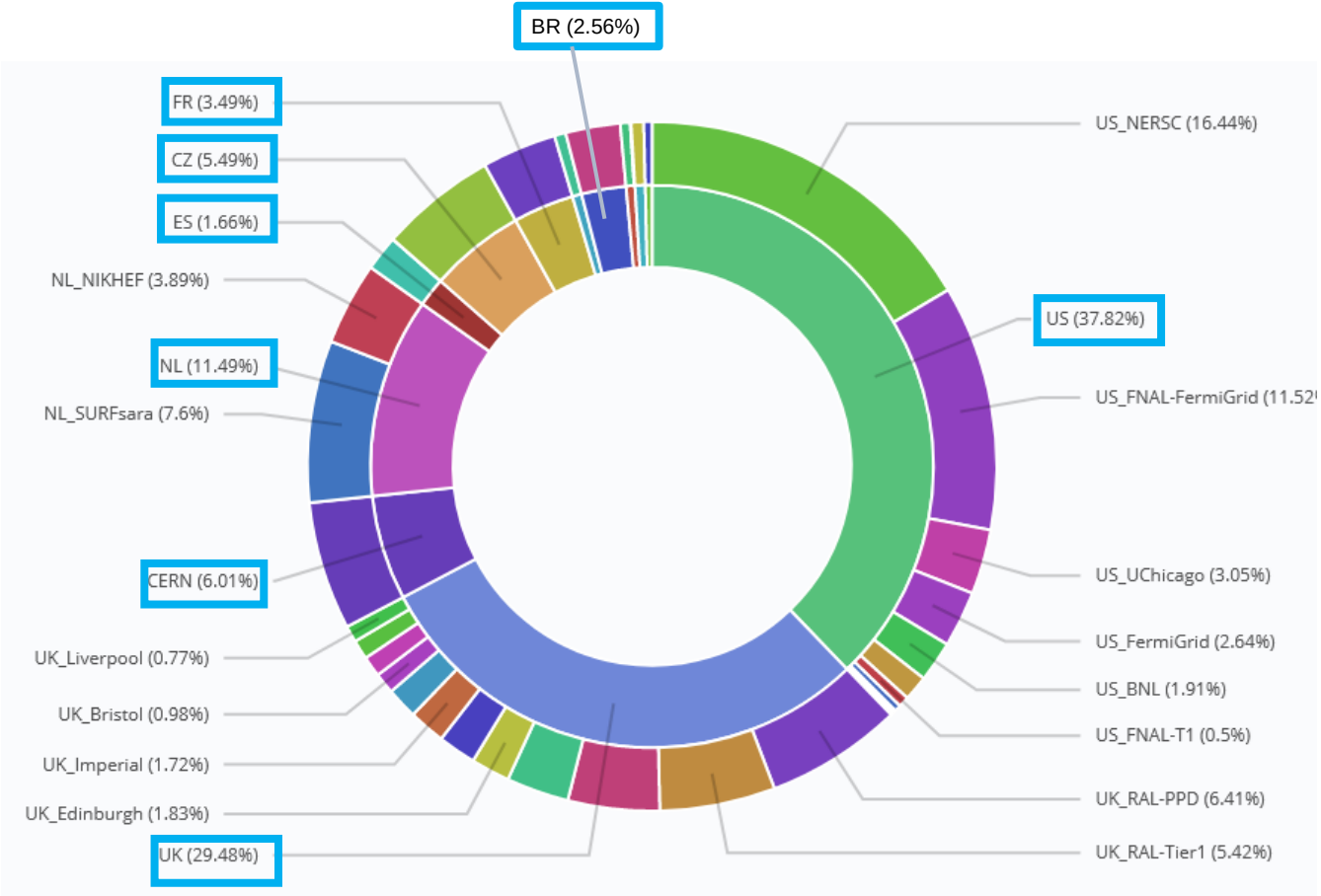
- **Coldbox CRP3 raw data reconstruction** production report

  continuation of the campaign started in 2022

  2023: ~200k events, output size < 10 TB, CPU ~400 CPU h

Production group worked in collaboration with data management group

Total CPU for production jobs in 2023: 11 kHS23-year



3. Plans for 2024

Active campaigns:

- FD1-HD FD2-VD step 2 full reconstruction: validation sample in preparation
- SP 2GeV MC: validation sample prepared

Production requests, resources footprint and shifter notes are avaibale [in the wiki]

What's next (*not a complete list*):

- protoDUNE-VD MC
- FD LE MC production

campaigns for which a preliminary discussion with physics groups started

production group will ask to this committee how to prioritize/coordinate requests

- FD1-HD (High energy) step 2 full reconstruction, ….
- **protoDUNE keeup processing**

➢ **Continue the integration of ND workflow**