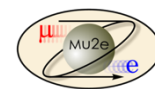




Mx2 Data Production Workflow

Tammy Walton for the Mx2 group
Production Meeting
October 31, 2024

In partnership with:



Objective

- This talk presents the workflow of the Mx2 data production
 - Preproduction
 - Calibration
 - Full Production

- We are requesting approval for the following:
 - The metadata for the offline production
 - Processing the offline data via production shifter
 - Archiving the Mx2 offline data to tape
 - Request data replicas at FNAL and NERSC

Overview of the data production

- Minerva-for-2x2 (Mx2) took NuMI data for the following periods
 - June 2023 (run 0)
 - [https://wiki.dunescience.org/wiki/Production_Run0_data_\(June_2023\)](https://wiki.dunescience.org/wiki/Production_Run0_data_(June_2023))
 - March 2024 – July 2024 (run 1)
 - https://wiki.dunescience.org/wiki/Production_Run1_data
- Mx2 data consist of the following data streams:
 - Pedestal data
 - PMT light injection data
 - NuMI data
 - Mixed mode pedestal and NuMI data
 - Mixed mode light injection and NuMI data
- The data are characterized by MetaCat datasets (see next slide)

Overview of the data production – Mx2 Run1 datasets

Pedestal Raw Data

Start Run	End Run	Dataset Name	Description	NFiles	Size	Location
1698	1795	twalton:mx2_pdstl_raw_run1_prod_a	files from dune:all where creator=dunepro and core.data_stream=pdstl and core.data_tier=binary-raw and core.group=minerva and core.event_count > 1 and core.runs[0] >= 1698 and core.runs[0] <= 1795	435	447.7 GB	/pnfs/dune/tape_backed/dunepro/neardet-2x2-minerva/binary-raw/2024/importedDetector/pdstl
1796	1805	twalton:mx2_pdstl_raw_run1_prod_b	files from dune:all where creator=dunepro and core.data_stream=numip and core.data_tier=binary-raw and core.group=minerva and core.event_count > 1 and core.runs[0] >= 1796 and core.runs[0] <= 1805	7	7.2 GB	/pnfs/dune/tape_backed/dunepro/neardet-2x2-minerva/binary-raw/2024/importedDetector/numip
50001	50018	twalton:mx2_pdstl_raw_run1_prod_c	files from dune:all where creator=dunepro and core.data_stream=numip and core.data_tier=binary-raw and core.group=minerva and core.event_count > 1 and core.runs[0] >= 50001 and core.runs[0] <= 50018	17	17.4 GB	/pnfs/dune/tape_backed/dunepro/neardet-2x2-minerva/binary-raw/2024/importedDetector/numip

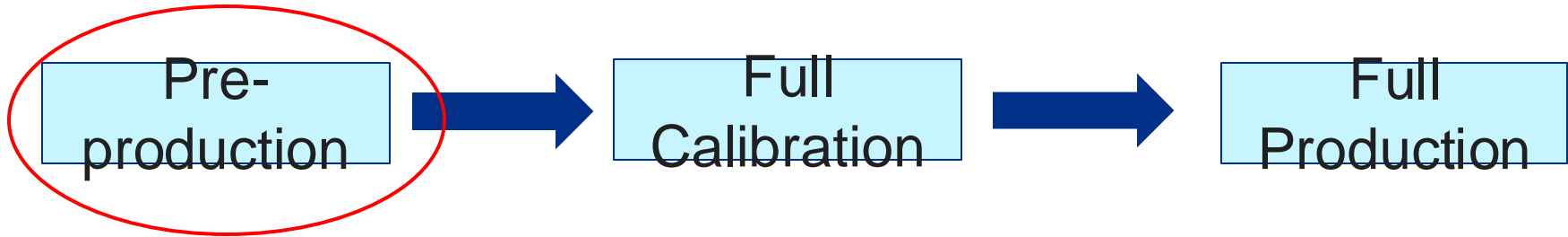
- “twalton” is the namespace, which can be changed
- The name of the dataset is designed to be short and simple
- The run periods for each dataset are based on simple

Mx2 running conditions

Beam Raw Data

Start Run	End Run	Dataset Name	Description	NFiles	Size	Location
1702	1795	twalton:mx2_numib_raw_run1_prod_a	files from dune:all where creator=dunepro and core.data_stream=numib and core.data_tier=binary-raw and core.group=minerva and core.event_count > 1 and core.runs[0] >= 1702 and core.runs[0] <= 1795	302	350.4 GB	/pnfs/dune/tape_backed/dunepro/neardet-2x2-minerva/binary-raw/2024/importedDetector/numib
1796	1805	twalton:mx2_numib_raw_run1_prod_b	files from dune:all where creator=dunepro and (core.data_stream=numip or core.data_stream=numil or core.data_stream=numib) and core.data_tier=binary-raw and core.group=minerva and core.event_count > 1 and core.runs[0] >= 1796 and core.runs[0] <= 1805	96	119.4 GB	/pnfs/dune/tape_backed/dunepro/neardet-2x2-minerva/binary-raw/2024/importedDetector/<numi data> where <numi data> := numip, numil, and numib
50001	50018	twalton:mx2_numib_raw_run1_prod_c	files from dune:all where creator=dunepro and (core.data_stream=numip or core.data_stream=numil or core.data_stream=numib) and core.data_tier=binary-raw and core.group=minerva and core.event_count > 1 and core.runs[0] >= 50001 and core.runs[0] <= 50018	117	141.1 GB	/pnfs/dune/tape_backed/dunepro/neardet-2x2-minerva/binary-raw/2024/importedDetector/<numi data> where <numi data> := numip, numil, and numib

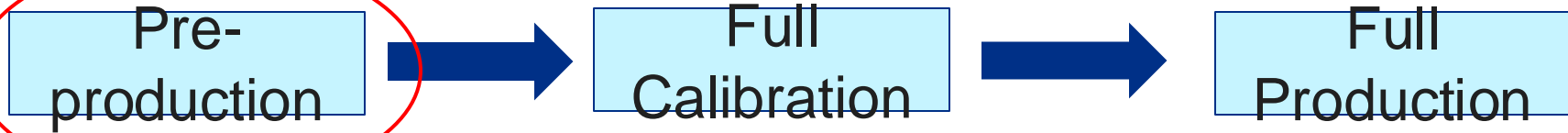
Overview of the data production



1. Pre-production

- Purpose:
 - Execute the minimal algorithms that are needed to perform the low-level calibrations
- Technical details
 - Performed by the calibration team
 - Jobs are submitted to the grid via JustIN, where the data are processed using an official release
 - Output files are transferred to the user scratch area
 - Calibration team analyzes the output data and uploads the constants to the development database
 - The analyses are presented at the Mx2 group meeting and then approved
 - Output data are transferred to the dune persistent area and the validation plots are posted on the wiki
 - The approved constants are migrated from the development database to the production database

Overview of the data production



Output of the Preproduction Stage

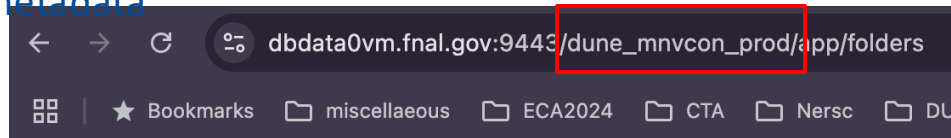
Pre-Calibration Data [\[edit\]](#)

				Pedestal Data			
Start Run	End Run	NFiles	Directory Size	Location	Validation Plots	Comments	
1698	1795	432	387 MB	/pnfs/dune/persistent/physicsgroups/dunendproto/twalton/mx2/constants/pedestals/run1a	mx2_pedestals_run1_prod_a.pdf	432 good tables (large pdf, might want to download)	
1796	1805	7	6.1 MB	/pnfs/dune/persistent/physicsgroups/dunendproto/twalton/mx2/constants/pedestals/run1b	mx2_pedestals_run1_prod_b.pdf	7 good tables	
50001	50018	16	14 MB	/pnfs/dune/persistent/physicsgroups/dunendproto/twalton/mx2/constants/pedestals/run1c	mx2_pedestals_run1_prod_c.pdf	16 good tables	

[How to read these plots \(at the bottom\)](#)

				Gain Data			
Start Run	End Run	NFiles	Directory Size	Location	Validation Plots	Comments	
1698	1795	7	45.1 MB	/pnfs/dune/persistent/physicsgroups/dunendproto/twalton/mx2/constants/gain/run1a		Not enough events for a gains table	
1796	1805	8	12.2 GB	/pnfs/dune/persistent/physicsgroups/dunendproto/twalton/mx2/constants/gain/run1b	mx2_gain_run1_prod_b.pdf	8 good tables	
50001	50018	15	20.3 GB	/pnfs/dune/persistent/physicsgroups/dunendproto/twalton/mx2/constants/gain/run1c	mx2_gain_run1_prod_c.pdf	14 good tables	

Pedestal and gain data do not have the accompanied metadata



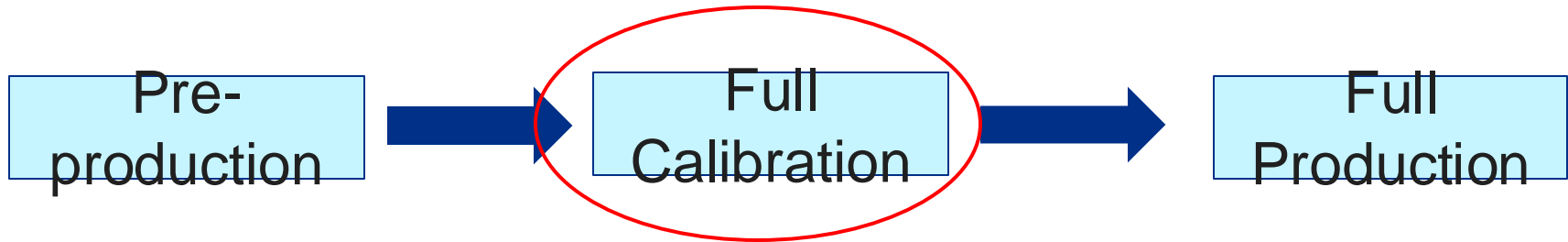
```

minerva2x2_feb_constants
minerva2x2_pedestal_constants
minerva2x2_atten_od_constants
minerva2x2_atten_id_constants
minerva2x2_pmtgain_constants
  
```

Preproduction constants are in the production database



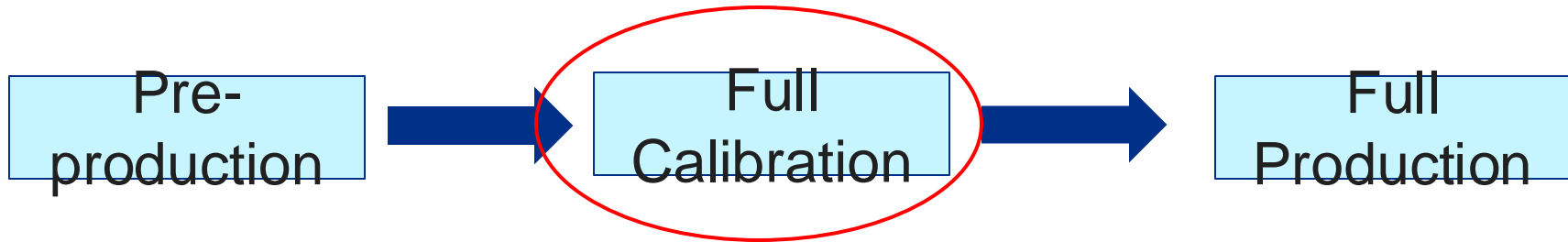
Overview of the data production



2. Full Calibration

- Purpose:
 - Using the approved pre-production constants, this stage reconstructs rock muons that are deployed to extract the high-level calibration constants.
- Technical details
 - Performed by the calibration team
 - Jobs are submitted to the grid via JustIN, where the data are processed using an official release
 - The full calibration stage involves multiple iterations between different analyses
 - All output files are transferred to the user scratch area
 - The final output files are transferred to the dune persistent area and the validation plots are posted on the wiki
 - The calibration team analyzes the output data and uploads the constants to the development database
 - The analyses are presented at the Mx2 group meeting and then approved

Overview of the data production



Rock Muon Calibration [\[edit\]](#)

Rock Muon Data						
Start Run	End Run	NFiles	Directory Size	Location	Validation Plots	Comments
1702	1795	302	37G	/pnfs/dune/persistent/physicsgroups/dunendproto/twalton/mx2/rockmuons/run1a/data/03613/1/		
1796	1805	96	12G	/pnfs/dune/persistent/physicsgroups/dunendproto/twalton/mx2/rockmuons/run1b/data/03612/1/		
50001	50018	117	181M	/pnfs/dune/persistent/physicsgroups/dunendproto/twalton/mx2/rockmuons/run1c/data/03597/1/		

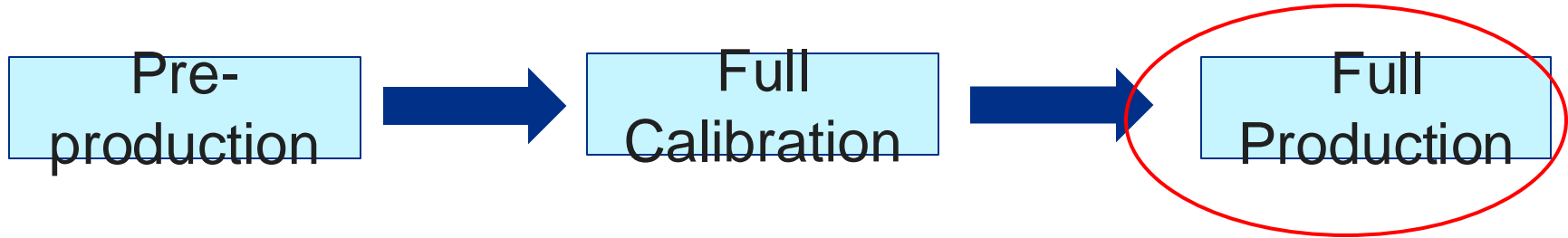
Calibration Data [\[edit\]](#)

Validation Plots						
Start Run	End Run	Timing	Strip-to-Strip	Muon Energy Scale	Comments	
1698	1795					
1796	1805					
50001	50018					

In Progress

The rock muon files do have the accompanied metadata
 The rock muon files are small
 Request these data to be catalogued in MetaCat (not archived to tape)

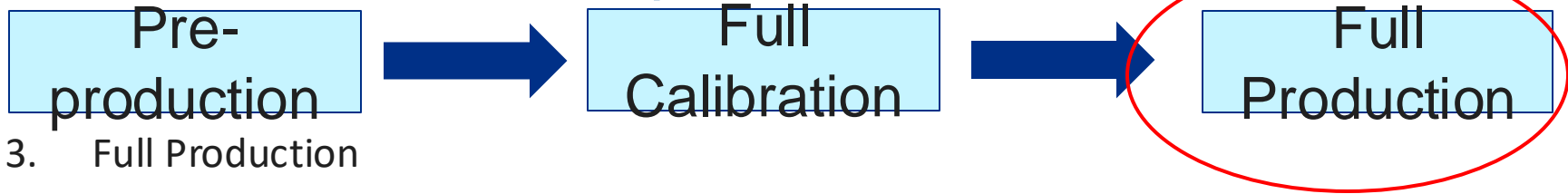
Overview of the data production



3. Full Production

- Purpose:
 - Using the approved calibrations, this stage reconstructs data into physics objects
- Technical details
 - Need production to be submitted by dunepro
 - Noe Roy has volunteered to serve as the dune pro shifter on behalf of the 2x2 group
 - Jobs are submitted to the grid via JustIN, where the data are processed using an official release
 - Need details on where to transfer the reconstructed data and there are two types of data
 - Gaudi root files (nested root trees)
 - Flat root files (input to the 2x2 Common analysis framework)
 - Request approval for the metadata

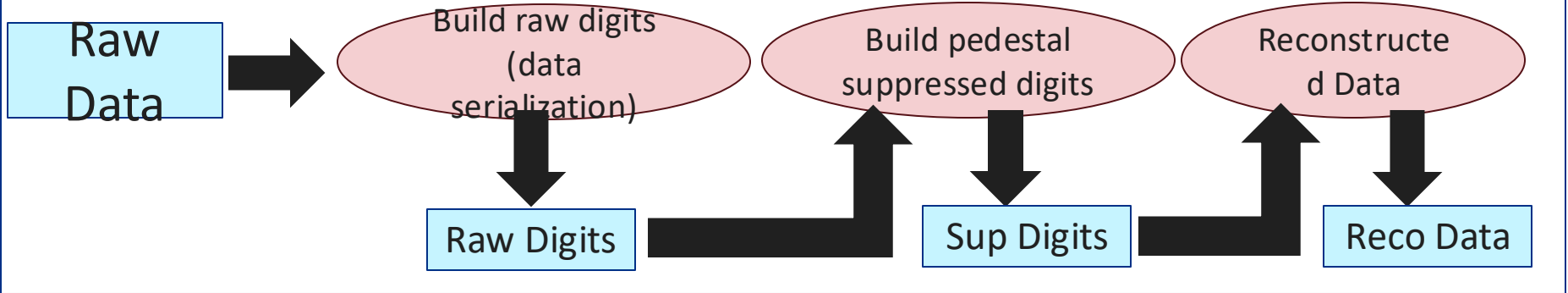
Overview of the data production



3. Full Production

– Overview of the submission workflow

CPU node – Execute a sequential three-staged workflow



twalton:mx2_numib_raw_run1_prod_c				
	Common File Size	Smallest File Size	Largest File Size	Archived to Tape
Raw Digits	~1.0 GB	6.9 MB	1.2 GB	No
Sup Digits	~800 MB	6.8 MB	805 MB	No
Reco Gaudi Data	~500 MB	3.1 MB	545 MB	Yes
Reco DST Data	~300 MB	1.8 MB	325 MB	Yes

Overview of the data production : Metadata

New Metadata Fields:

- dune.dqc_quality – used to create “good” datasets (may change it to a list)
- dune.mx2x2_global_subrun_numbers – used for matching the files from the 2x2 charge and light readout systems

```
{
  "name": "TS1_00050017_0017_numib_v09_2410292332_RecoData.root",
  "namespace": "neardet-2x2-minerva",
  "creator": "twalton",
  "size": 545918104,
  "metadata": {
    "core.data_tier": "reco",
    "core.start_time": "1730244766",
    "core.end_time": "1730245882",
    "core.file_format": "root",
    "core.events": [],
    "core.last_event_number": 7200,
    "core.first_event_number": 1,
    "core.file_content_status": "good",
    "dune.dqc_quality": "unknown",
    "dune.campaign": "run1",
    "dune.requestid": "",
    "dune.config_file": "TS1_00050017_0017_numib_v09_2410292332_RecoData.opts",
    "dune.workflow": "justin",
    "dune.output_status": "true",
    "core.application.family": "gaudi",
    "core.application.name": "reco",
    "core.application.version": "v102r1p1",
    "retention.status": "active",
    "retention.class": "physics",
    "core.runs": [
      50017
    ],
    "core.group": "minerva",
    "core.run_type": "neardet-2x2-minerva",
    "core.file_type": "importedDetector",
    "core.data_stream": "numib",
    "core.event_count": 7200,
    "core.runs_subruns": [
      706732721
    ],
    "core.lum_block_ranges": [
      [
        26511430,
        26518629
      ]
    ]
  },
}
```

We need to fix the core.runs_subruns

```
,
  "dune.mx2x2_global_subrun_numbers": [
    500170000542,
    500170000543,
    500170000544,
    500170000545,
    500170000546,
    500170000547,
    500170000548,
    500170000549,
    500170000550,
    500170000551,
    500170000552,
    500170000553,
    500170000554,
    500170000555,
    500170000556,
    500170000557,
    500170000558,
    500170000559,
    500170000560,
    500170000561,
    500170000562,
    500170000563,
    500170000564,
    500170000565,
    500170000566,
    500170000567,
    500170000568,
    500170000569,
    500170000570,
    500170000571,
    500170000572,
    500170000573,
    500170000574,
    500170000575,
    500170000576,
    500170000577,
    500170000578
  ],
  "parents": [
    {
      "did": "neardet-2x2-minerva:TS1_00050017_0017_numib_v09_2407100856_RawData.dat"
    }
  ],
  "checksums": {
    "adler32": "efd92f00"
  }
}
```