ProtoDUNE-DP: Computing, data readiness and organization

LBNC Meeting
CERN, 05/12/2019

Elisabetta Pennacchio, IPNL

Introduction

- ProtoDUNE-DP operations started on August 28th : 1.5M events have been collected so far.

- This presentation aims to explaining how these raw data are handled, processed, and more general how they are organized and how they can be accessed.

The following points will be discussed:

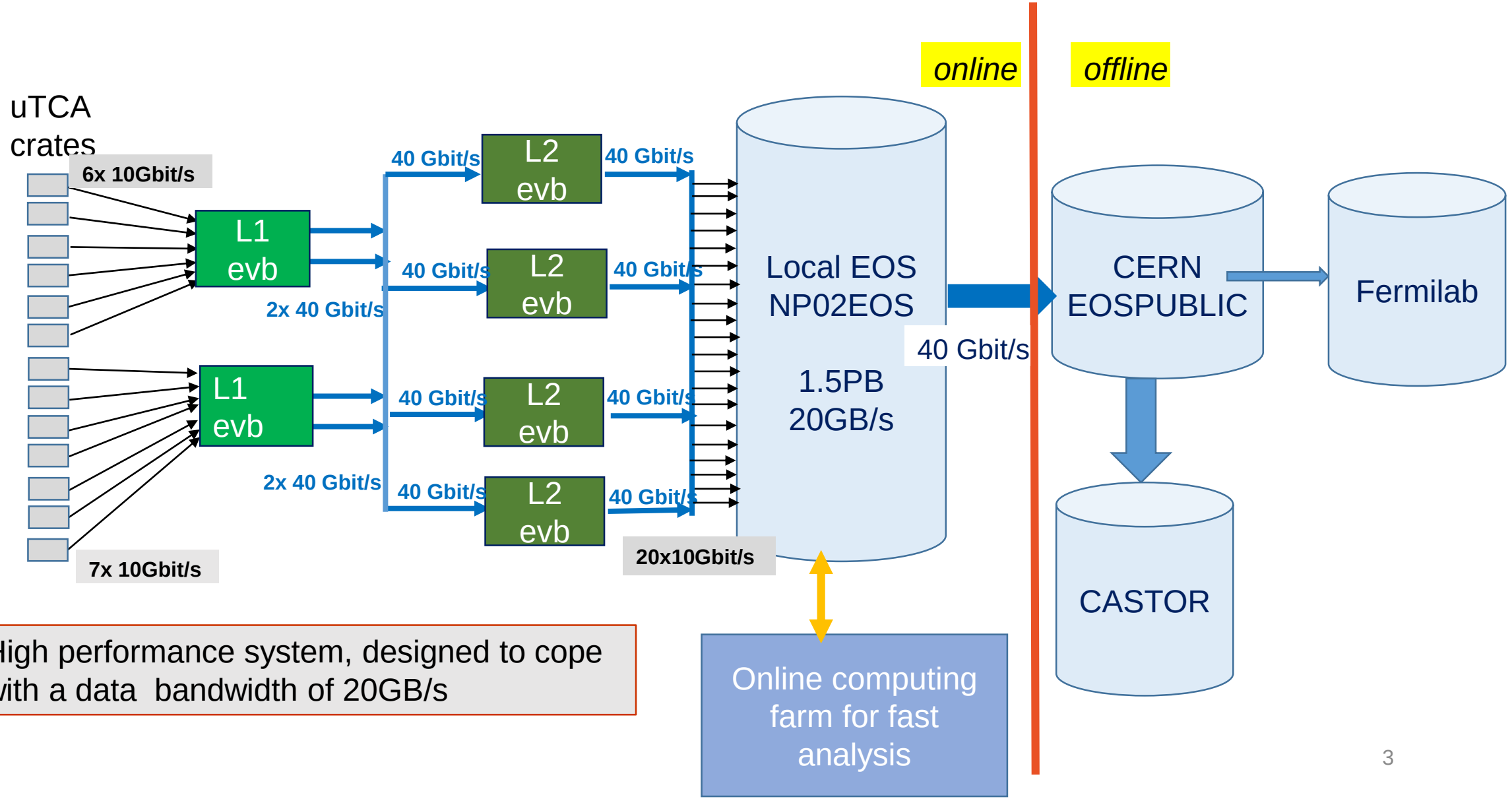
1. Online data organization : online storage and processing
2. Data transfer to CERN EOSPUBLIC: interface between online and offline
3. Offline data organization: data replication and offline processing
4. Data accessibility

I will discuss only the tools and the organization put in place.

Analysis activities are regularly going on in order to understand purity and LEM gain and performance, and have already been presented in previous talks

1. Online data organization : online storage and processing

- Reminder of ProtoDUNE-DP network architecture, back-end system and interface to offline computing



High performance system, designed to cope with a data bandwidth of 20GB/s

Raw Data description

- **A run corresponds to a well defined detector configuration (e.g. HV setting), and it is composed by several Raw Data files (sequences) of a fixed size of 3GB (optimized for storage and data handling)**

- **Raw Data files are produced by 2 levels of event building**:
  Level-1 event builders: 2 machines (**L1**) and Level-2 event builders: **4** machines (**L2**) **working in parallel**

The naming convention for Raw Data file is the following: ***runid_seqid_l2evb.datatype***, where

runid : run number,
seqid: sequence id, starting from 1
l2evb: can be equal to *a,b,c,d,* **to identify by which L2 event builder the file was assembled**
datatype can be *test*, *pedestal, cosmics,…*

- So, for the test run 1010 the Raw Data filenames will look like that:

*1010_**1**_a.test*
*1010_**1**_b.test*
*1010_**1**_c.test*
*1010_**1**_d.test*
*1010_**2**_a.test*
*1010_**2**_b.test*
*1010_**2**_c.test*
*1010_**2**_d.test*

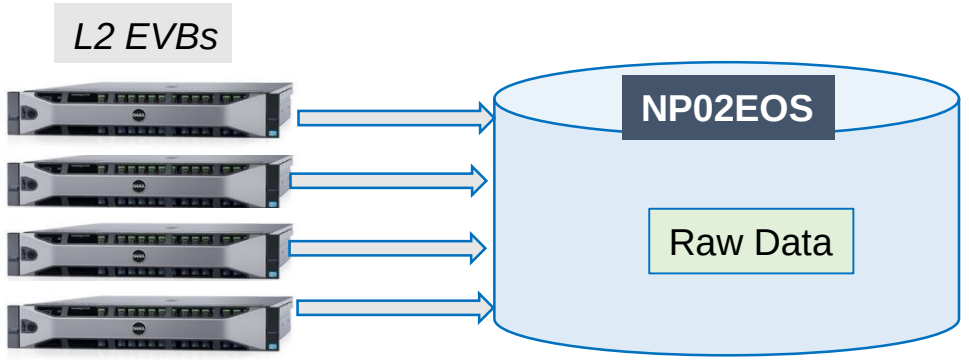
np02evbl2a	1,5,9,13,17,…
np02evbl2b	2,6,10,14,18,..
np02evbl2c	3,7,11,15,19,…
np02evbl2d	4,8,12,16,20,…

- **Events in a given file are not strictly consecutive**: in order to fully parallelize processing each L2 event builder includes in its treated sequences only event whose number follows an arithmetic allocation rule (based on division module), as shown in the table here

Raw Data Storage

- Four L2 event builders first assembly and write Raw Data files in their RAM memory. **As soon as a data file is closed the process *L2EOS* running on each L2 event builder takes care of copying** it to the online storage facility (NP02EOS)

- NP02EOS **high performance EOS based distributed storage system (20 GB/s):** 20 storage servers (DELL R510, 72 TB per machine): up to 1.44 PB total disk space, 10 Gbit/s connectivity for each storage server.
  The version of eos running on NP02EOS instance is updated with the one running on EOSPUBLIC.

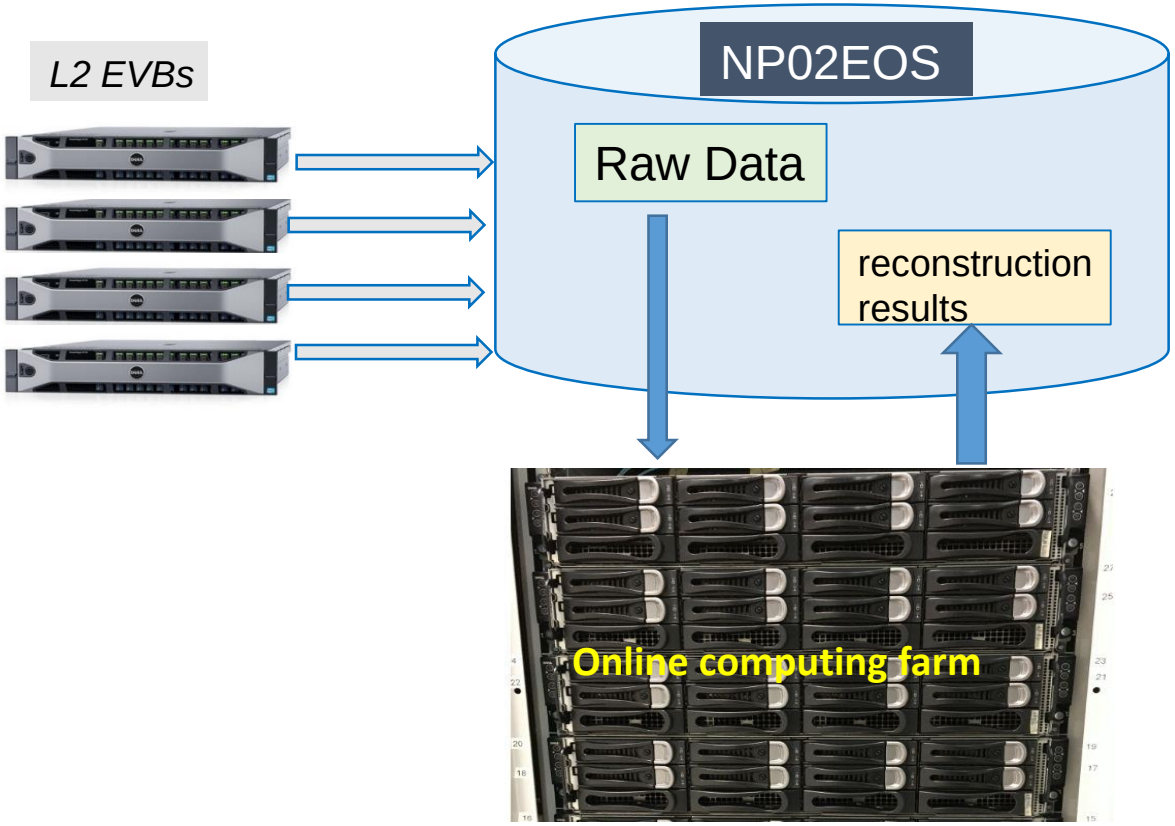


- The raw data files assembly by the event builders and their transfer to NP02EOS is done with a **dedicated software**, which has developed taking into account the network configuration and the characteristics of the EVBs. **This software has been intensively tested since 2018 with dedicated data challenges and it has been ensuring smooth data handling in 2019**

https://indico.fnal.gov/event/16526/session/10/contribution/164/material/slides/0.pdf
https://indico.fnal.gov/event/18681/session/7/contribution/151/material/slides/0.pdf

Online processing

- Once on NP02EOS, files are scheduled for automatic online reconstruction on the online processing farm.
  40 servers Poweredge C6200, corresponding to ~450 cores → **fast tracks reconstruction and data quality**

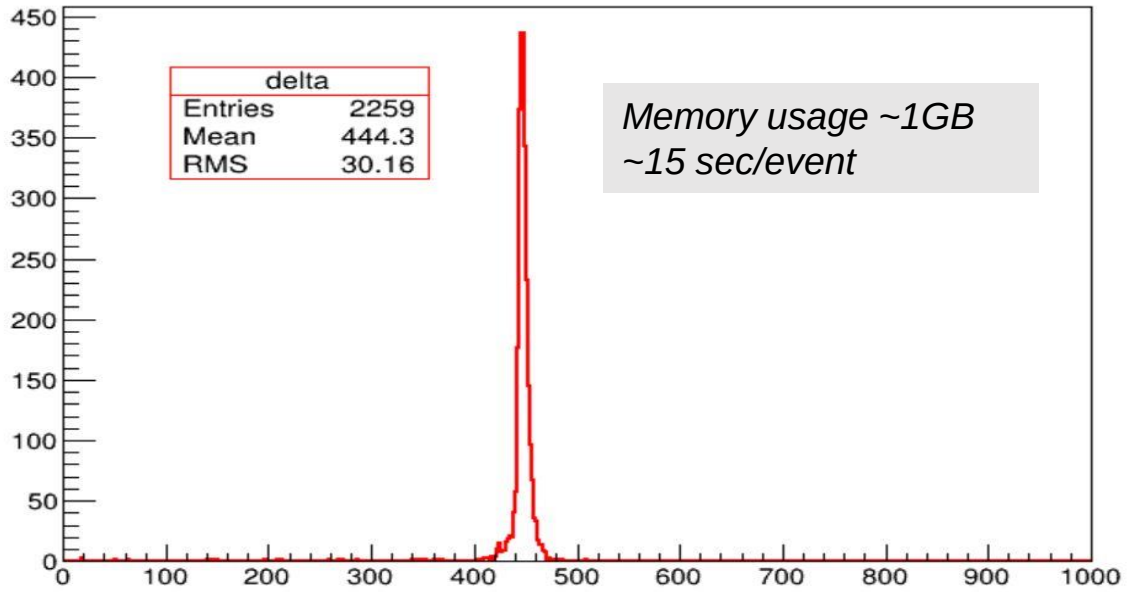


→ Short time interval in between the assembly of a file by one event builder and the availability of the reconstruction results is ~15 minutes

- <u>All events are systematically processed</u>.

Hits, 2D tracks and 3D tracks are reconstructed by using the **fast reconstruction** based on QSCAN (WA105Soft) which was already used for the analysis of the 3x1x1 data → code simple and robust, (years of developments), suited to extract the **basic information at hits level and dQ/dx along single tracks** (not to reconstruct complicated topologies, showers etc … which is eventually the task of the offline analysis with LArSoft)

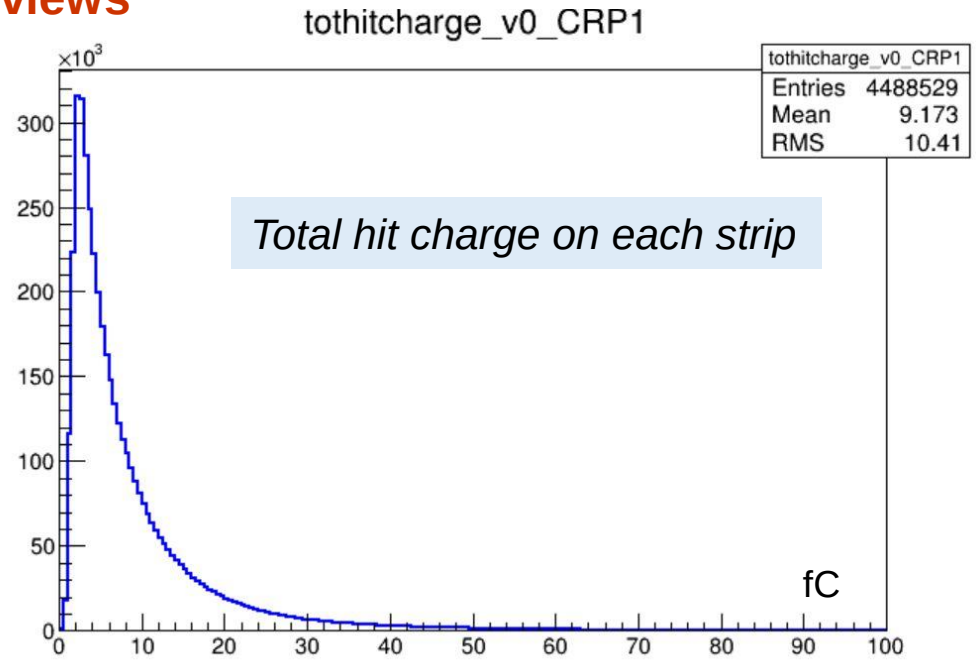
Processing time(no I/O) for Raw Data files of 30 events:



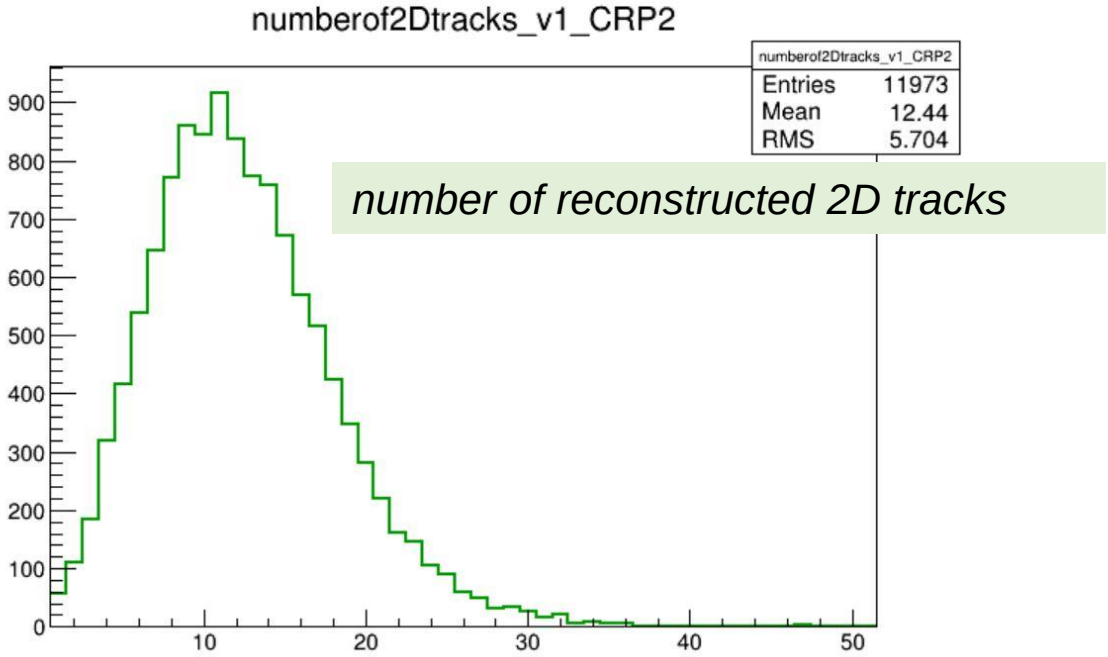
Memory usage ~1GB
~15 sec/event

Online reconstruction output used to produce a standard set of distributions **for Data Quality Monitoring**
→ see next slide

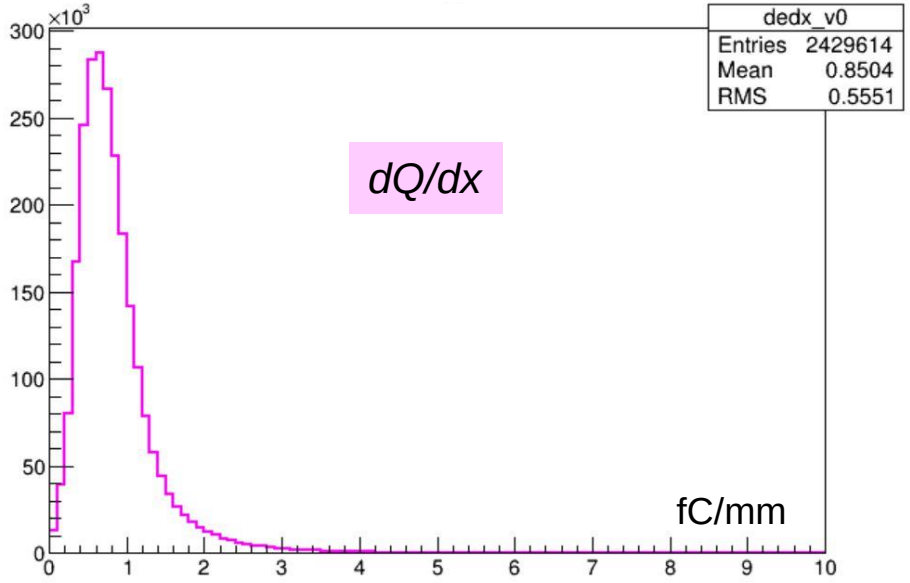
Some examples of distributions of online Data Quality Monitoring. Distributions are available for all CRPs, both views



Total hit charge on each strip



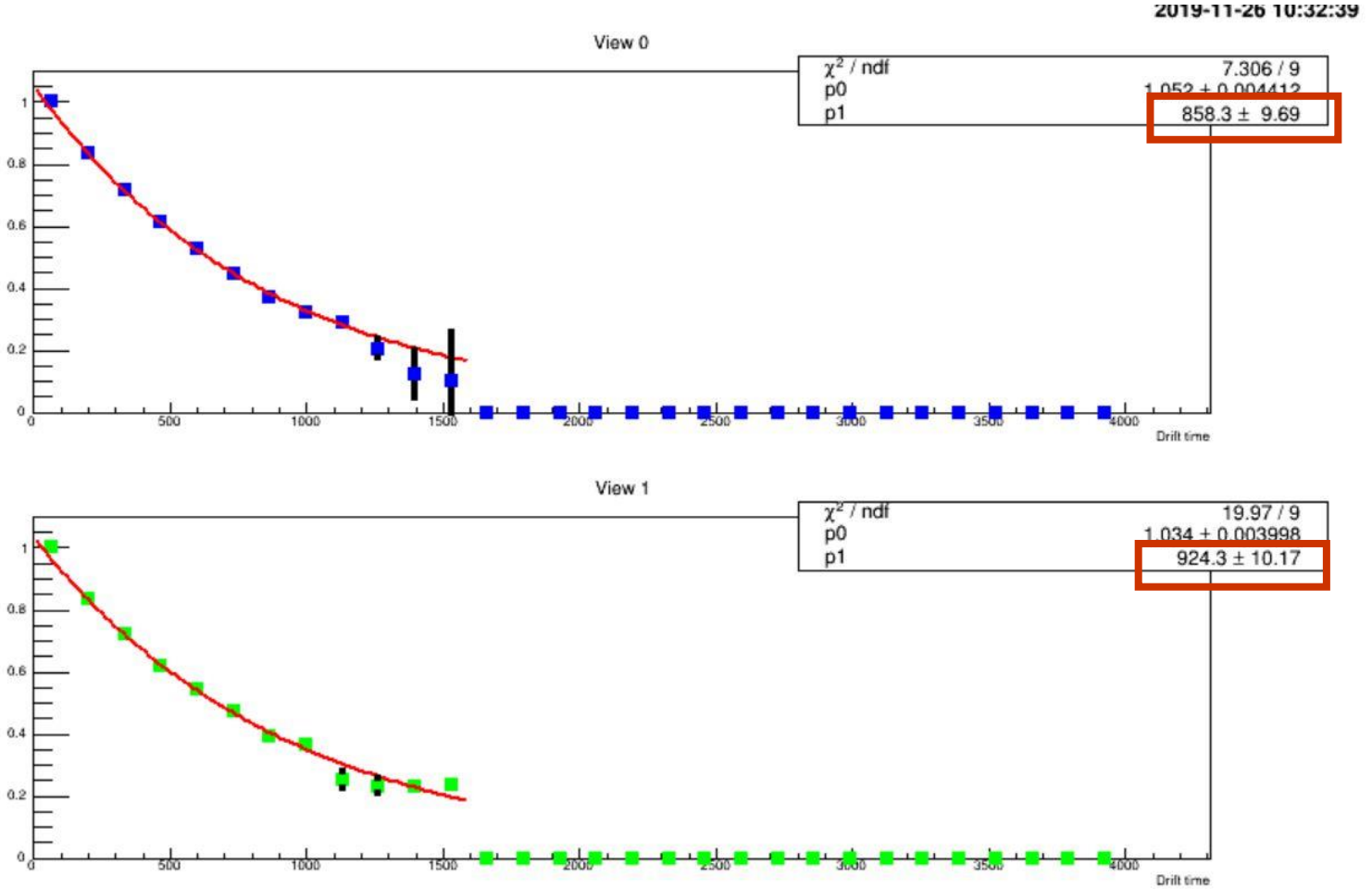
number of reconstructed 2D tracks



dQ/dx

These distributions can be used to check the behavior as a function of time of the detector and eventually detect unforeseen changes

Electron lifetime is also systematically measured for all cosmic runs .
The method used to evaluate the electron lifetime is based on 2D tracks reconstruction. For each run, two measurements of the charge attenuation along the track are obtained independently for view_0 and view_1



Lifetime ~1 msec

2. Data transfer to CERN EOSPUBLIC: interface between online and offline

- Raw data files (but also online reconstruction results, and purity measurement results) are copied from NP02EOS to CERN EOSPUBLIC, **to make them available to the DUNE collaboration.**

- Since the endpoint of the transfer is CERN storage, it has been decided to run the transfer by using FTS developed at CERN. This solution presents several advantages:

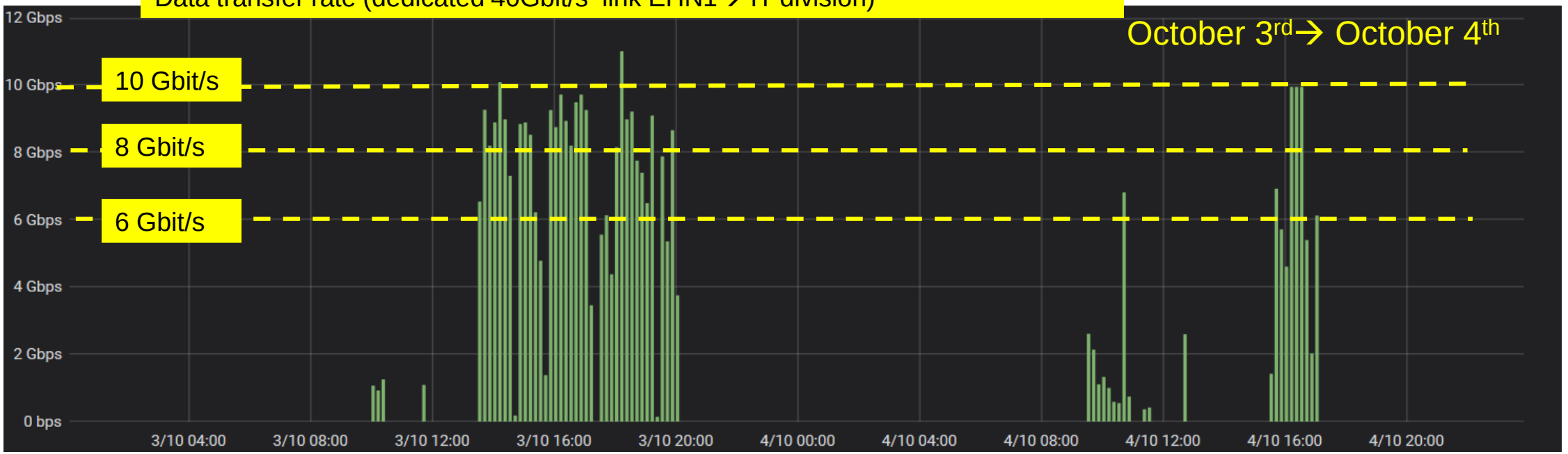
  - easy to put in place and to use
  - in case of transfer failure, retries are performed
  - optimization of the available bandwidth, to maximize data transfer rate
  - support and feedback from CERN IT division
  - dashboards to monitor files transfer status are available, detailed instructions on how to retrieve information about transfers (duration, problems…) from the FTS database are provided as well.

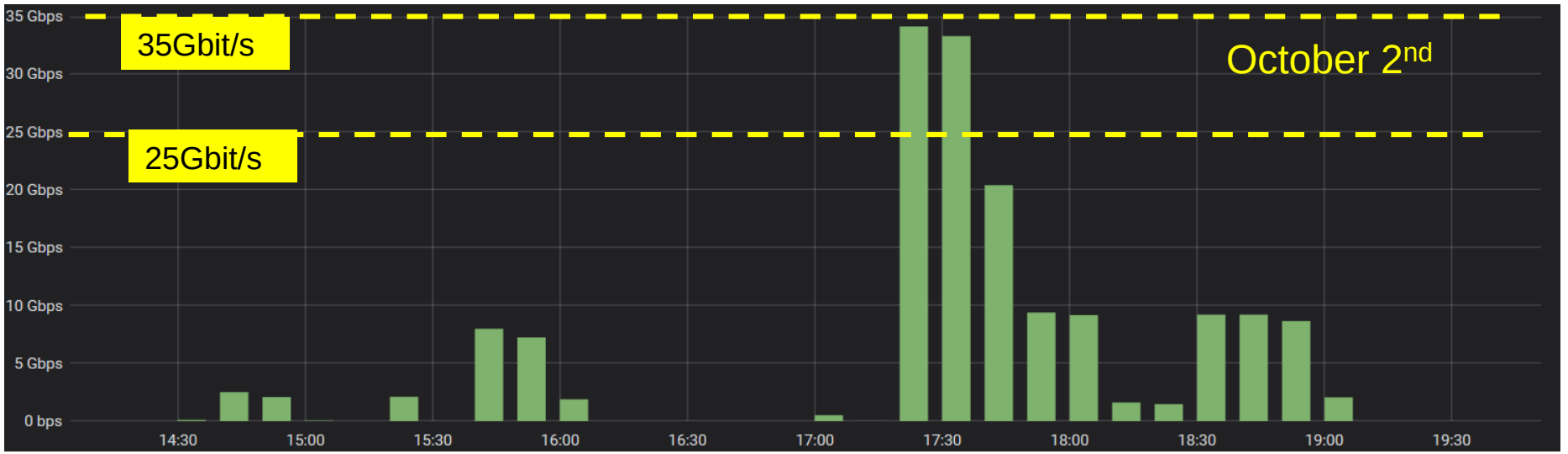
- **The FTS transfer** is run from some DAQ service machines connected to the online storage system; for each Raw Data file a *metadata file* is generated as well in order to allow the logging of the data file in the overall **DUNE data management scheme**.

- The delay Δt between the creation of a Raw Data file and its availability on EOSPUBLIC is ~10 minutes

Raw Data Flow Monitoring (NP02EOS→EOSPUBLIC) some examples:

Data transfer rate (dedicated 40Gbit/s link EHN1→ IT division)

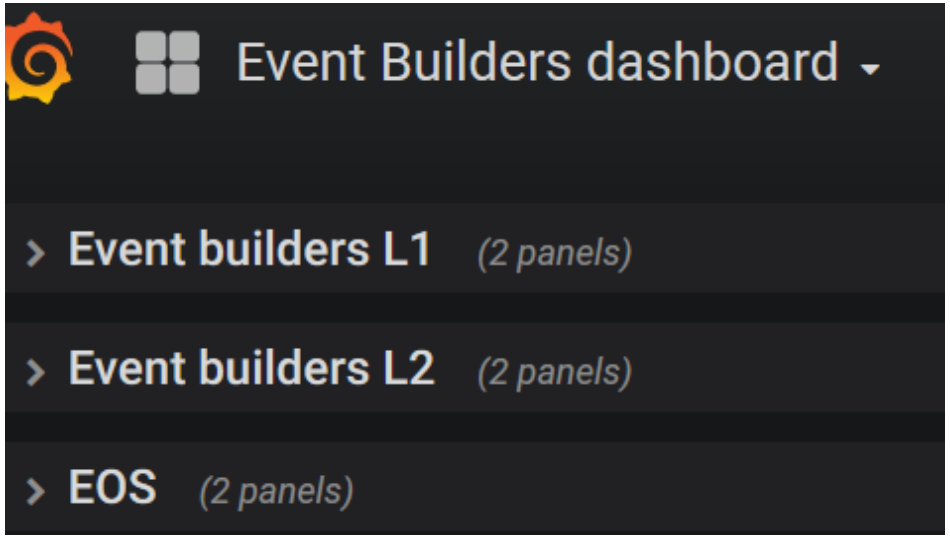


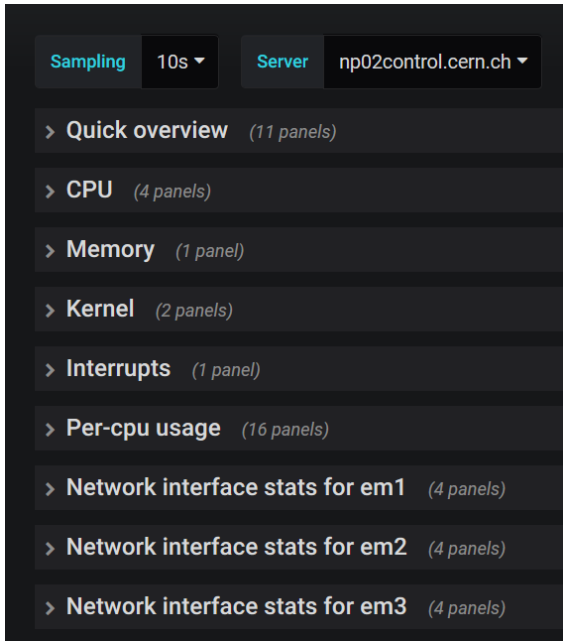


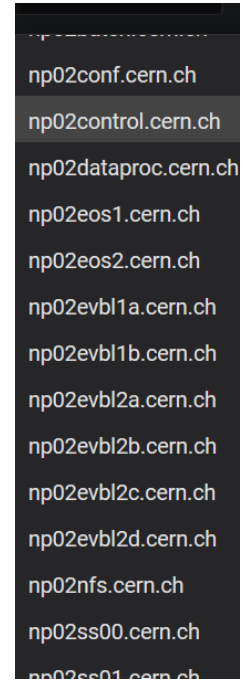
Back-end activity logging and monitoring:

1) All steps of **Raw Data handling** are stored in a dedicated **online database**

2) The monitoring of the **activity of the DAQ machines, storage and processing farm** is performed with 2 dedicated Grafana dashboards.







What we learned after these months of activities :

1. Several activities related to the setting up and commissioning of the back-end were performed in strict collaboration with CERN/IT (network deployment, setting up of NP02EOS, usage of FTS and EOS) and Fermilab Data Management group (integration of Raw Data files in the overall DUNE scheme) .
In particular *IT-ProtoDUNE coordination (Single Phase and Double Phase) meetings* are regularly organized by CERN/IT.

It is fundamental to keep strong links, since they allow to anticipate any possible problem in data flow management that would delay the availability of Raw Data on EOSPUBLIC (and to the DUNE collaboration).

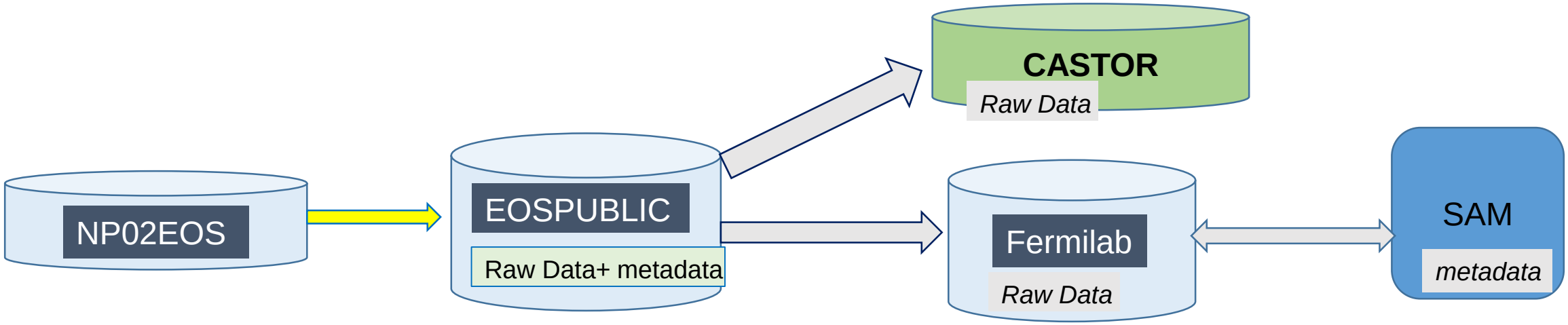
2. Every time a new component (hardware or software) of ProtoDUNE-DP back-end has been put in place, a data challenge was run, to test this new part. More generally, data challenges to stress the system have also been regularly organized. This allowed to find out and fix weak points and problems well before the start of the operations.

→ Indeed by carefully preparing all the mechanism the data taking and data handling went ahead quite smoothly

3. Offline data organization: data replication and offline processing

- As mentioned before, Raw Data and online reconstructed data are copied by the DAQ system from NP02EOS to EOSPUBLIC.

- The integration of NP02 Raw Data in the general DUNE data management scheme is done by metadata files. On the online machine a metadata file is generated for each Raw Data file and copied as well to EOSPUBLIC. These **metadata files trigger the data transfer to CASTOR (storage on tapes), and Fermilab (data replication) .**
- These transfers are run by the **Fermilab data management group, as it is done for ProtoDUNE-SP.**

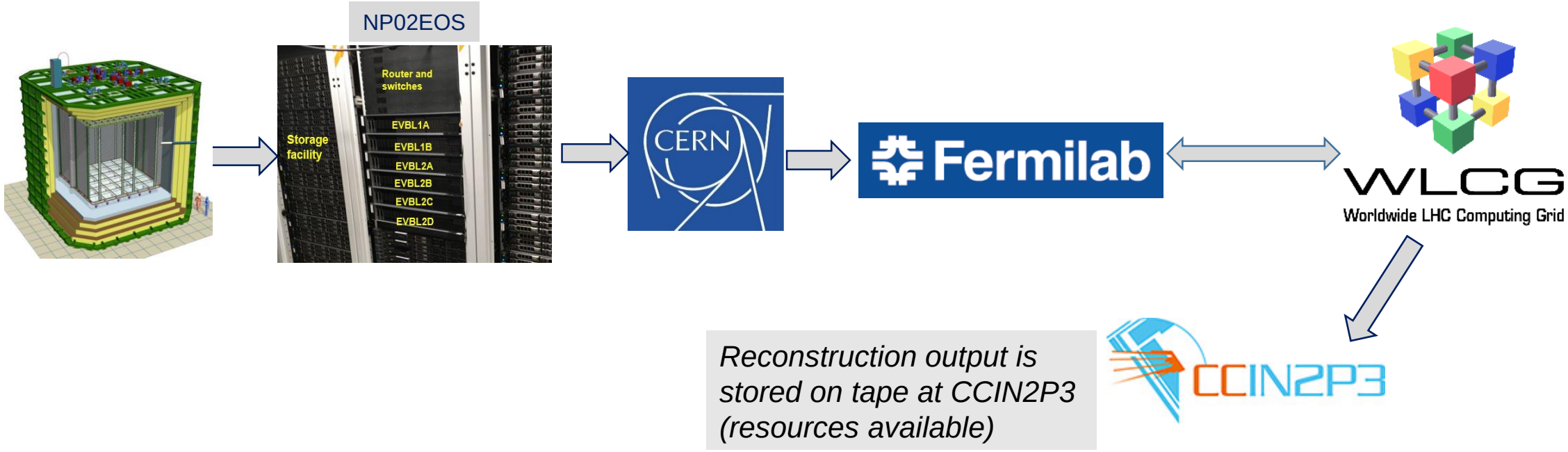


- Once Raw Data are transferred to Fermilab → they become available for LArSoft reconstruction

- **The offline reconstruction and analysis of ProtoDUNE-DP data is performed with LArSoft**, similarly as for ProtoDUNE-SP, and benefit of the same environment and tools.

- The processing of ProtoDUNE-DP data is organized in a centralized way by the **Fermilab production and data management groups.**

- The workflow is the following:



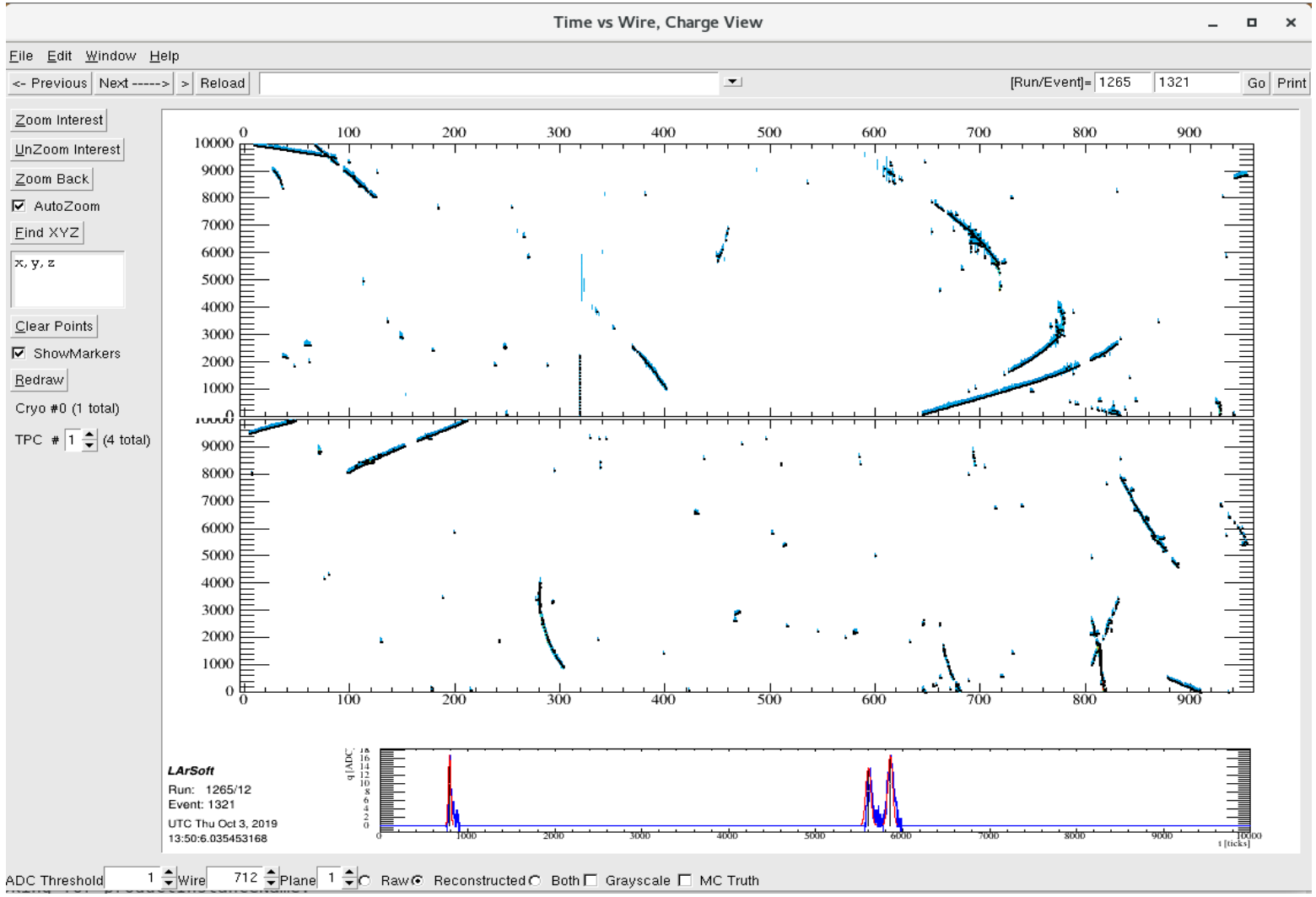
- The **grid processing scheme** is the result of several months of development; in particular it was **validated and commissioned** by running 2 dedicated **data challenges** (April 2018, July 2019) organized with CERN IT division and Fermilab production and data management groups.

LArSoft is a shared framework with other Fermilab LAr TPC experiments, based on Art framework. To include ProtoDUNE-DP in the framework, a LArSoft interface to the ProtoDUNE-DP Raw Data was developed and provided. Reconstruction (hits and 2D tracks) has been tested and validated.

https://indico.fnal.gov/event/21266/contribution/1/material/slides/0.pdf
https://indico.fnal.gov/event/22125/contribution/1/material/slides/0.pdf
https://indico.fnal.gov/event/22190/contribution/3/material/slides/0.pdf

Example

Results of reconstruction performed with LArSoft on a cosmic event in ProtoDUNE-DP



- A first subsample of data has been reconstructed with LArSoft: validation of physical results is ongoing. Once finished, the plan is to move to the **massive production of all data taken in 2019. Resources (CPU and space) are already defined and available**

  → **LArSoft reconstruction of ProtoDUNE-DP data takes ~10 min/events, 1.5GB memory footprint**

- As already discussed before, a strict collaboration with DUNE software and computing and data management groups has been crucial, together with the organization of data challenges

4. Data accessibility

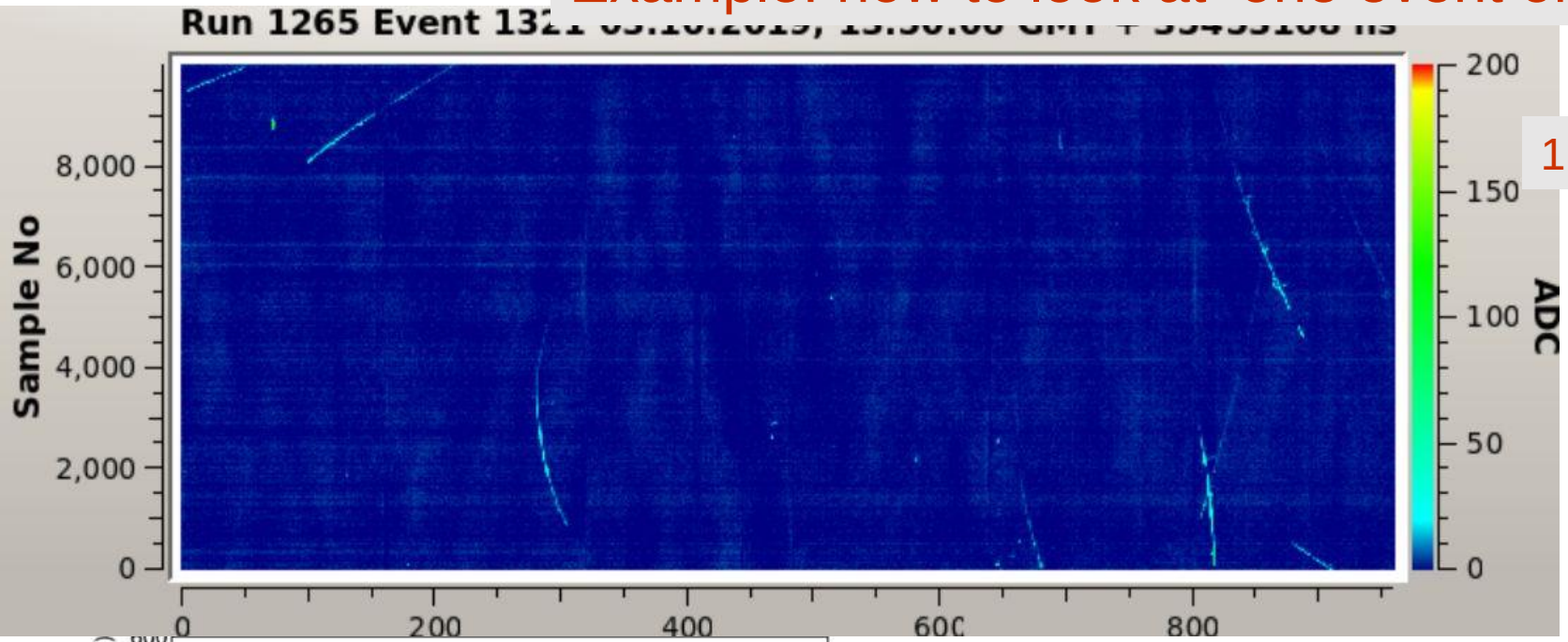
- **Raw data** are available at CERN , on EOSPUBLIC and CASTOR, and at Fermilab

- **Offline reconstructed data**: availability of reconstructed data with LArSoft is document in the general DUNE data catalog https://dune-data.fnal.gov
  The documentation about LArSoft is available from the DUNE WIKI pages here https://wiki.dunescience.org/wiki/Main_Page

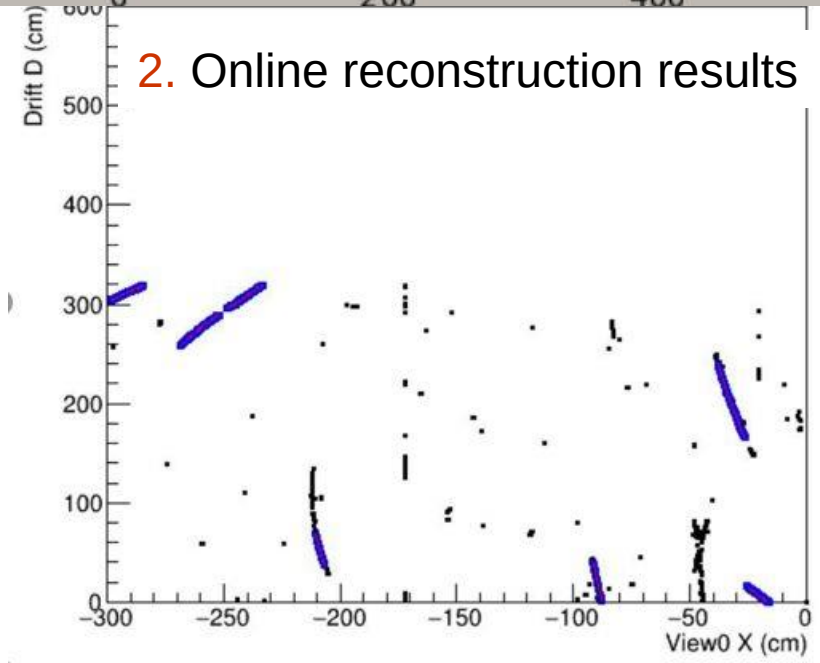
- **Online reconstructed data: results from the fast online reconstruction are also available on LXPLUS**.
  In addition to LArSoft on LXPLUS it has been set up as well a simpler environment to **access Raw Data (and online reconstruction results) based on the online reconstruction software → these data sets can be used also for some simple offline analyses**
  → A copy of the same software installed on the online machines to run the online processing and perform the electron lifetime measurement is available
  → The online event display which is part of the DAQ system is also available. It can be used offline to look at the Raw Data files

User documentation is being written and updated continuously. It is available on NP02 twiki operation pages.
A "Software and data access tutorial session" has been organized (October 31st)

Example: how to look at one event on LXPLUS

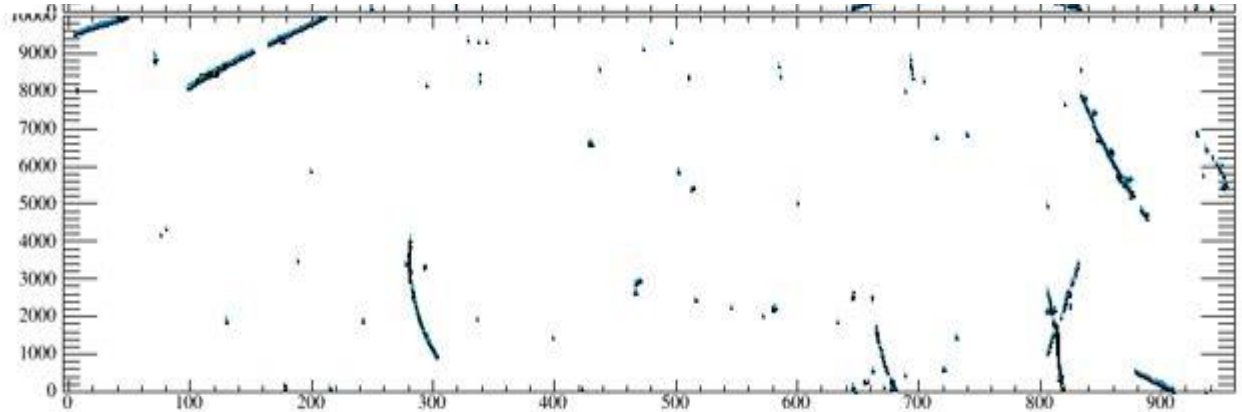


1. Raw event: NP02 event display



2. Online reconstruction results

3. LArSoft reconstruction results



Conclusions

- DP-ProtoDUNE data organization is well in place.
- Both online and offline treatment of the data have been defined in previous months, and tested with several dedicated data challenges. Since the start of operations, online handling and processing are working automatically without manual interventions. As soon as raw data files are written, they are immediately made available to the DUNE collaboration. Interactions with CERN/IT and Fermilab are fundamental.
- Raw data are integrated in the overall DUNE scheme, and the processing with LArSoft has started.
- In addition to LArSoft, also online reconstructions results are available on LXPLUS for fast analysis checks