ProtoDUNE-DP: Data organization and analysis

DUNE Collaboration Meeting
CERN, 27/1/2020

Elisabetta Pennacchio, IPNL

Introduction

- ProtoDUNE-DP operations started end of August 2019 : 1.9M events have been collected so far

- This presentation aims to explaining the access to ProtoDUNE-DP data, related software tools, the organization of data processing and analysis and some examples of preliminary results
  → Everybody is encouraged in looking at the ProtoDUNE-DP data

- Four main points will be discussed:

  a) *Online data organization: online storage and processing*

  b) *Offline data organization: data replication and offline processing*

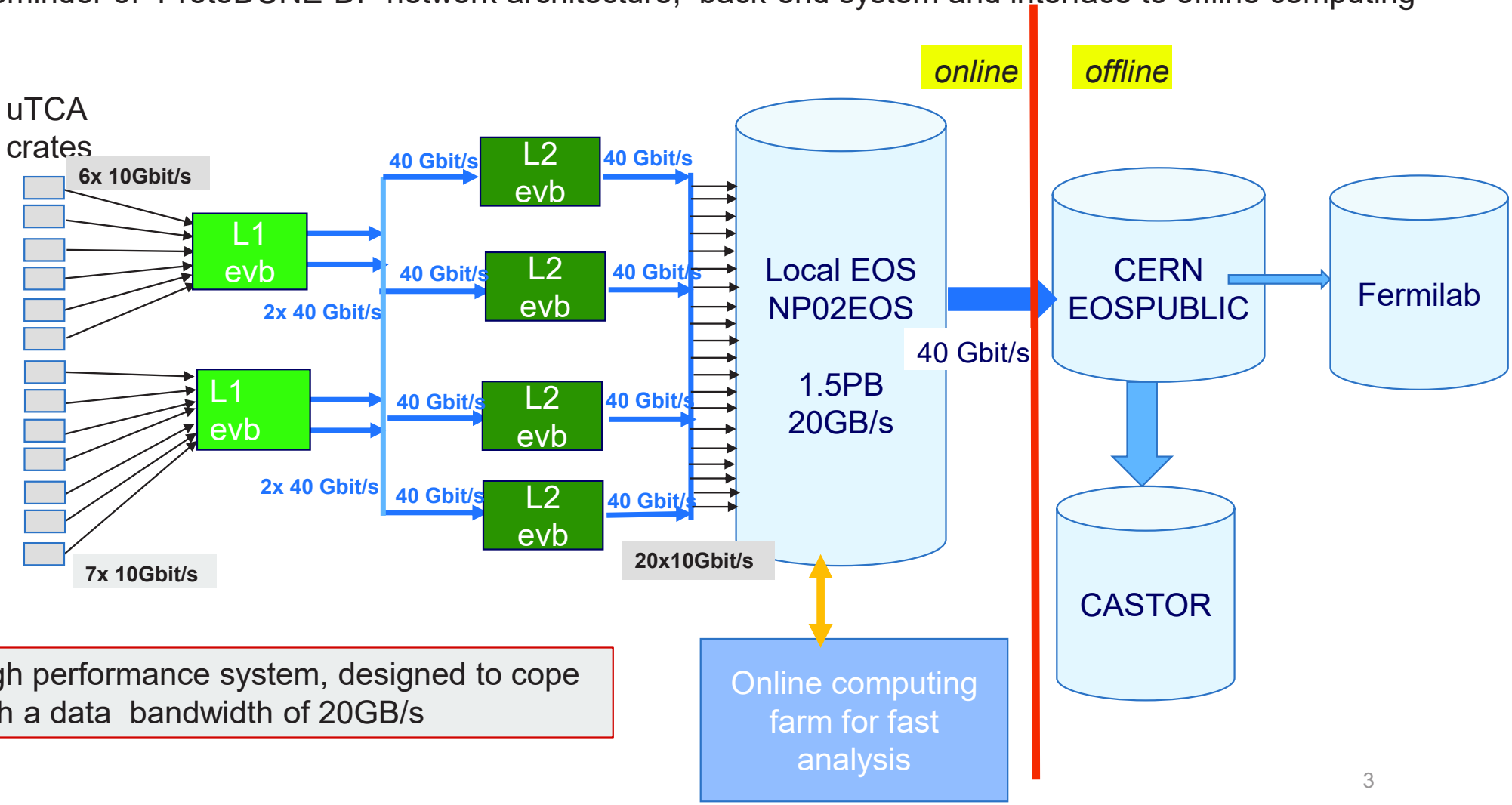
  c) *Data accessibility*

  d) *Data analysis and detector performance*

Analysis activities have been regularly going on during detector operation in order to understand purity and LEM/CRP gain and performance → some results have already been presented in previous talks

A) 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 and files naming

- **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:

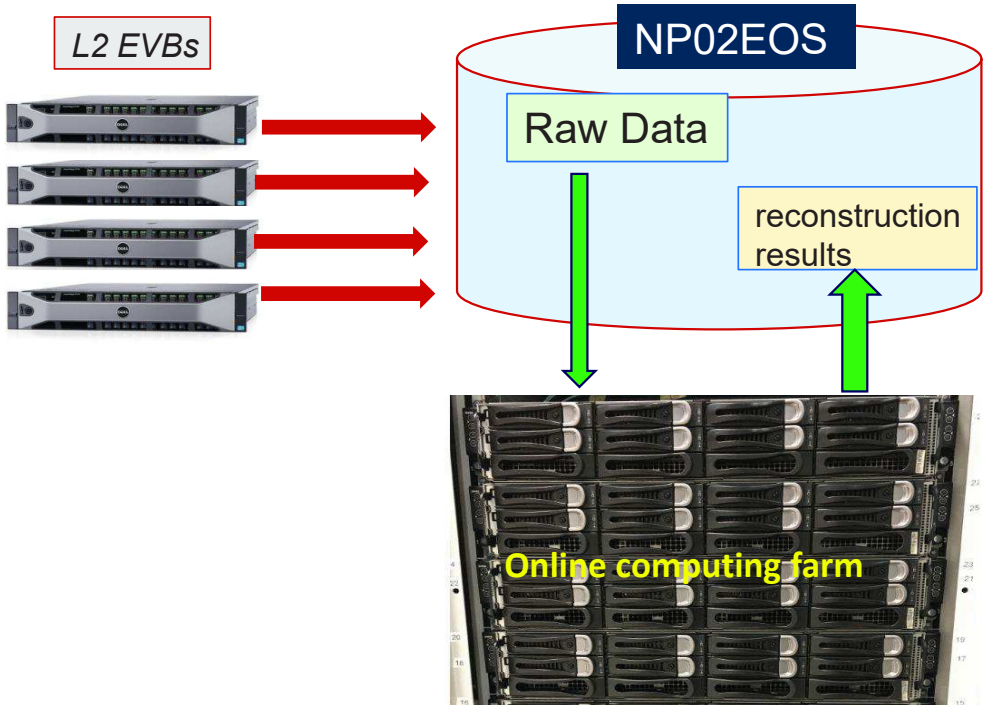
1010_1_a.test
1010_1_b.test
1010_1_c.test
1010_1_d.test
1010_2_a.test
1010_2_b.test
1010_2_c.test
1010_2_d.test

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 online storage and fast processing scheme

1) Four L2 event builders continuously assemble the events in raw data files in their RAM memories
→ As **soon as a raw data file is completed it is transferred to the online storage** facility NP02EOS (1.5PB, 20GB/s) based on the EOS distributed file system



2) Once stored on NP02EOS, files are scheduled for **automatic online reconstruction** on the DAQ 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** (~15 minutes). → All events are systematically processed.