

# Brief note on ProtoDUNE offline


Maxim Potekhin (BNL)

CERN DAQ workshop

02/26/2016

# Prioritization

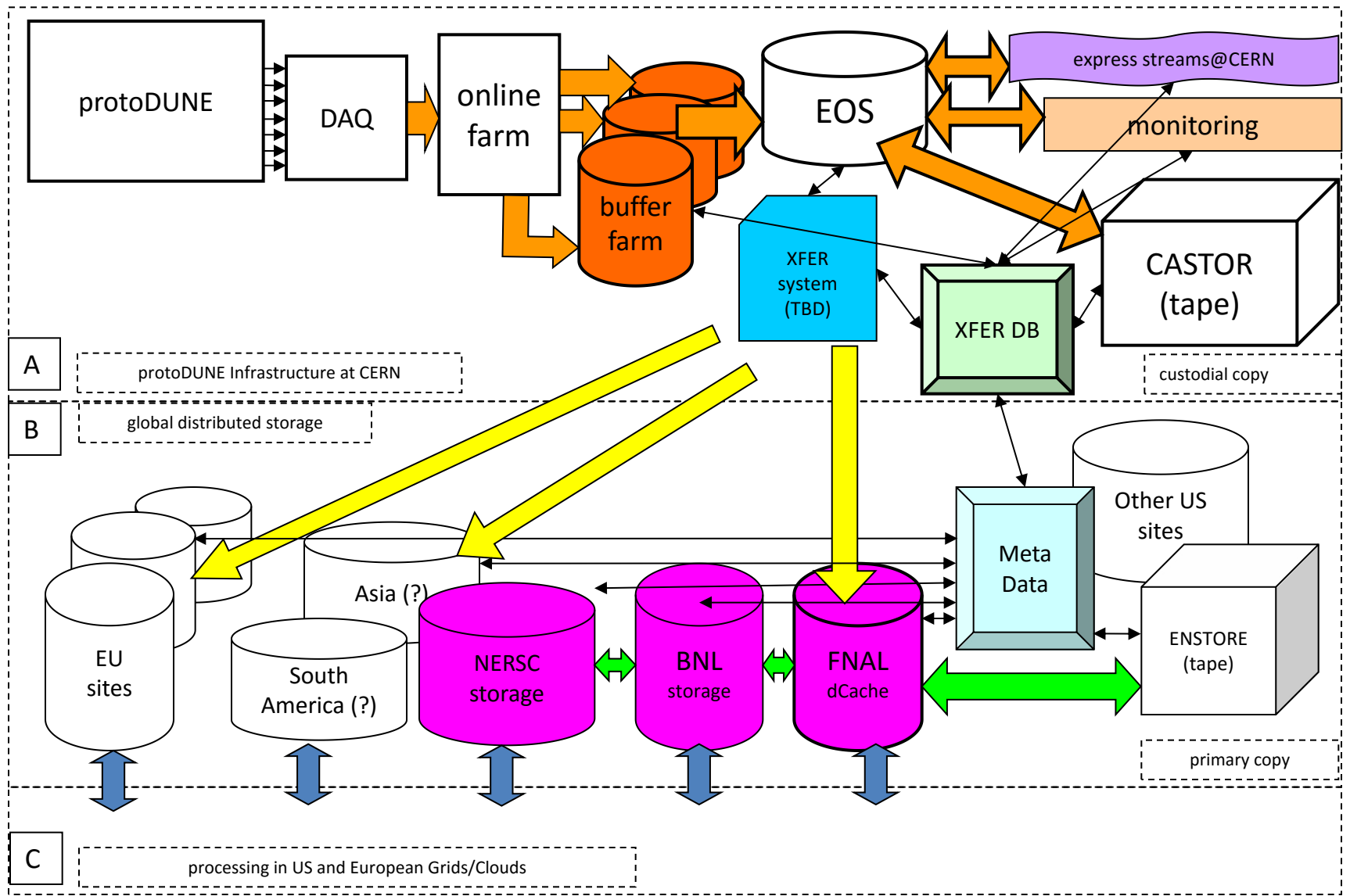
## Data Acquisition (DAQ)

- 1. Detector Monitoring and data QA**
  - 2. Transport of data to mass storage**
  3. Distributed data management
  4. Production (managed processing of the data which includes reconstruction)
  5. Analysis
- 

# ZS vs ROI

DAQ Strategy	Advantages	Disadvantages	Nominal Data Volume
ZS (DAQ1)	Compact data	Data quality dependence on a particular choice of ZS algorithm, lack of sensitivity to low-amplitude signals	0.1PB
ROI (DAQ2)	Ability to tune ZS offline and to more precisely study signal shapes, even at low amplitudes	Substantially larger rate and volume of data, lesser ability to characterize the cosmic ray background due to effective cuts on fiducial	1.0PB

# Conceptual diagram of raw data flow in protoDUNE



# Main design elements

Component	Function	Motivation
Buffer Farm (in round robin or similar configuration)	Absorb instantaneous data rate over of DAQ	Provide scalable bandwidth to disk, fault tolerance and modularity (DAQ decoupled from mass storage)
XFER DB	Consistent keeping of the state of units of data in the system as it is transmitted and read locally	Allows to manage and monitor all aspects of data transmission, facilitates express and monitoring streams
Express and monitoring streams	Prompt QA of the data and detector monitoring	Maximum efficiency of detector operation

# Items

- Two components we didn't discuss so far:
  - xfer to external sites (FNAL+others), tech downselect?
  - provisioning of DB - very much hope by CERN - having the xfer DB is the tactic used by the LHC experiments
- Interface to SAM at FNAL - do we have enough coverage?
- CPU request stands at 1M in 2017 and 8M in 2018, will probably need to scale out
  - additional sites?
- If we cross 2PB cap on raw data processing will become complicated due to logistics (disk and network in addition to CPU).

# Notes for WBS development and headcount (before data taking)

Item	FTE	Duration (months)
DAQ emulator to enable buffer integration	1	2
Prototyping of the buffer farm ("subfarm")	2	3
EOS interface and "consensus buffer flush"	2	2
EOS-CASTOR interface (bidirectional)	1	3
Prototyping express-streams to run at CERN, out of EOS	2	2
Provisioning/installing DUNE software at CERN	2	2
Mock data challenge (simulated express-streams at CERN)	1	2
Tech. evaluation and downselect for CERN-US data link	1	2
Metadata generation and SAM interface	1	3
Procurement, configuration and testing of bandwidth out of CERN	1	2
Scalability test of data transport CERN-US	1	1
Data flow monitoring system	2	2
Dress rehearsal of complete chain of data transmission (including metadata) using emulated components	3	1
Installation and configuration of the HW buffer (subfarm)	2	3
protoDUNE production system	3	4
Integration with actual DAQ	2	3
Data challenge with actual components in place (DAQ, subfarm, CERN streams, data link, production at FNAL and other sites)	3	4
XRootD data storage federation	1	3
Commissioning	3	3



$\geq 4\text{FTE} \times 2\text{yrs}$

Table 1: WBS items for protoDUNE computing infrastructure

# Development platform at CERN: action items

- To make progress, we need:
  - To define teams and responsibilities
  - To get all "standard" CERN accounts for all team members with proper NP04 group attribution, auth/auth etc. (top priority).
  - "Sandbox" allocation of space in EOS and CASTOR
  - Machines: can start with VM but eventually will have to graduate to hardware, getting a few VMs soon would be very helpful
  - Development DB server (TBD)
  - Potentially, open port(s) for CERN-to-FNAL data link
- Some of these items require a degree of stewardship by CERN personnel...
- ...especially liaisons in ITD, Central Services and general support (accounts etc).
- If available, help from CERN experts (e.g. LHC experiments) acting as consultants will be appreciated as time is short.
- Perhaps a seminar with experts from LHC experiments could be helpful.
- Need to understand which fraction of resources such as VM etc will be provided by CERN vs DUNE.