

# Short term action items: protoDUNE computing infrastructure

Maxim Potekhin (BNL)

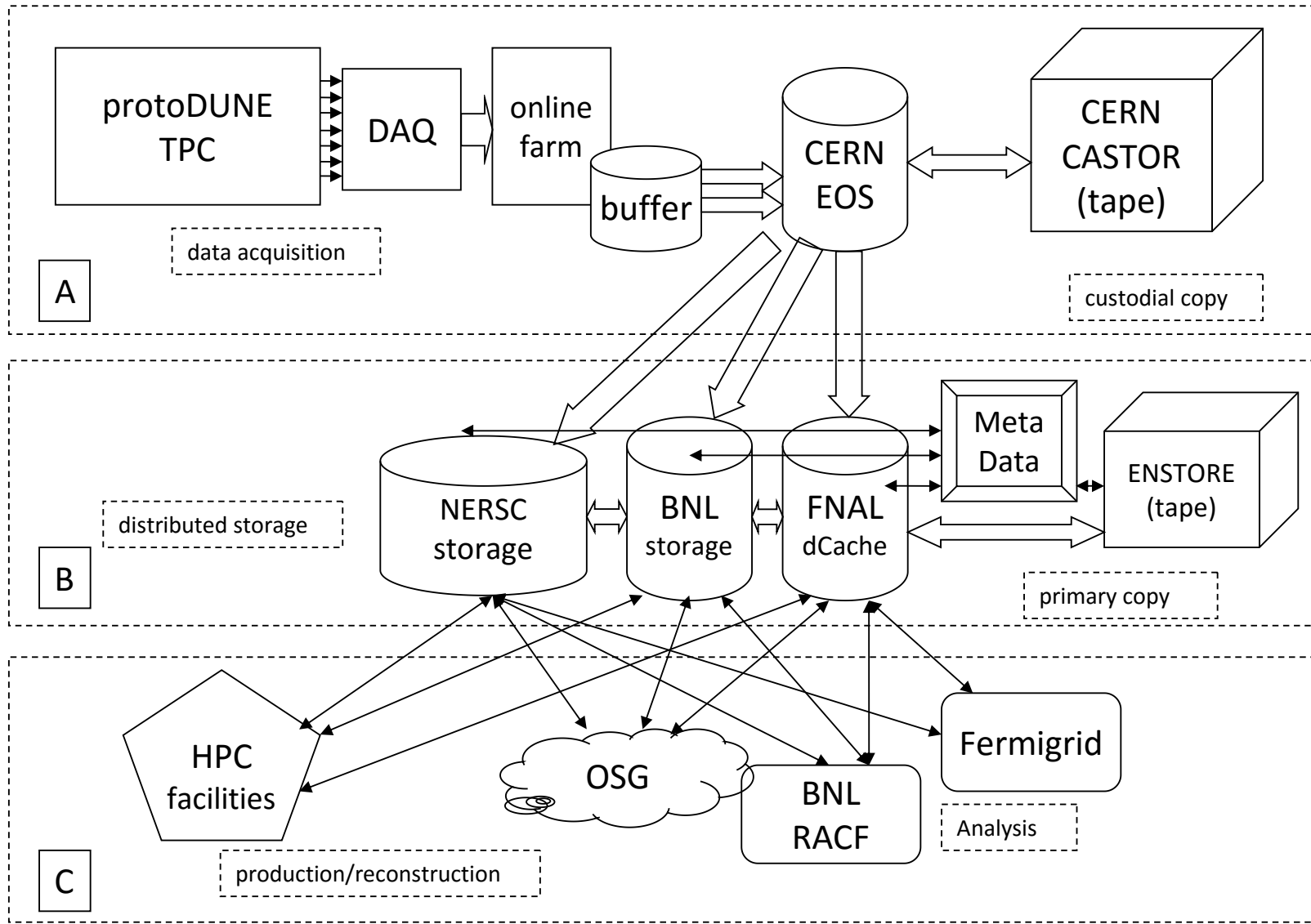
protoDUNE Meeting

12/18/2015

# Overview

- Scope of this presentation - not all of computing, but specifically data handling, after the data comes out of DAQ, and before it's ready for production at FNAL and other locations.
- This is the most time critical item (in addition to DAQ) to address.
- Recap of summary in October:
  - in general, data flow topology is quite similar to the LHC experiments
  - involves EOS as the staging area for outward transmission and express calibration streams, CASTOR for the custodial copy of the data
  - capability to run a subset of production at CERN for quick turnaround
  - main destination outside of CERN - FNAL (primary copy to be used in production), additional sites e.g. BNL, NERSC etc.
  - software reuse? ...not always feasible... (next slides)
- Nominal data rates and volume (perhaps TBD but the right ballpark)
  - 200MB/s sustained (e.g. to tape and out of CERN)
  - Assuming same bandwidth for protoDUNE as WA105, i.e. ~10gbps
  - CERN tape: O(100TB)

# Conceptual diagram of data flow in protoDUNE



# DAQ options

- According to Giles, due to the already tight schedule it may be too late to start from scratch, and the existing (a la 35t) solutions may have to be adopted.
- There is still a possibility that a new platform could be introduced for DAQ or a part of it.
- What does it mean for data handling? It means that an interface to DAQ needs to be designed, defined and agreed to with a view to decouple DAQ and data handling to the maximal degree possible.
- One of the difficult parts of doing so is the handshake among various components (including monitoring) for consensus on buffer flush.

# Notes

- To help planning, I created DocDB 883: notes on protoDUNE data handling infrastructure - attached to this meeting's Indico page.
- Looked at scripts used in ATLAS and CMS to transport data from DAQ to storage. Lots of dependencies on the respective experiments' computing infrastructure impedes reuse (express stream triggers, metadata and other DB interfaces etc).
- ***However, the basic protocols for EOS and CASTOR are always in place*** and we can script over these. This is also the approach of WA105.
- As laid out in DocDB 883, we'll need approximately 5FTE over two years to put proper data handling for protoDUNE in place. Approximate schedule is included in same document. Justification - time is already short and real effort is needed to meet deadlines.
- Staying in close contact with the DAQ group is crucial and we are doing this as best we can - that is a focus area. Handling of checksums, buffer flush handshake at general interfaces need to be agreed to and documented.

# Short term action items

- **ASAP:** get NP04 accounts and EOS to CASTOR sandboxes, and eventually allocation! May need involvement from protoDUNE spokes and deputy.
- Need a few VMs at CERN for development and testing.
- Do we need a single point of contact in CERN ITD/Central Services? Yes we do. WA105 do have a contact.
- Support - funding - for this effort (from DOE etc) is uncertain at this point so we need to start in interim by asking institutions to contribute on best effort basis, but at least with some commitment.
- Build a stable team.
- Spring 2016: meeting at CERN?