Overview of the Single-Phase protoDUNE online systems and data handling

M.Potekhin
B.Viren

Brookhaven National Lab

SBU-BNL Meeting
10/04/2016

Fermilab BROOKHAVEN NATIONAL LABORATORY DUNE

The protoDUNE program at CERN

- The protoDUNE program at CERN calls for the construction of large-scale engineering prototypes for both Single- and Dual-Phase Liquid Argon TPCs, and test-beam experiments with both.

- Scale: external dimensions of the cryostat for either detector are ~11×11×11m^3

- CERN designation (NP stands for “Neutrino Platform”):
  - NP02 Dual Phase
  - NP04 Single Phase

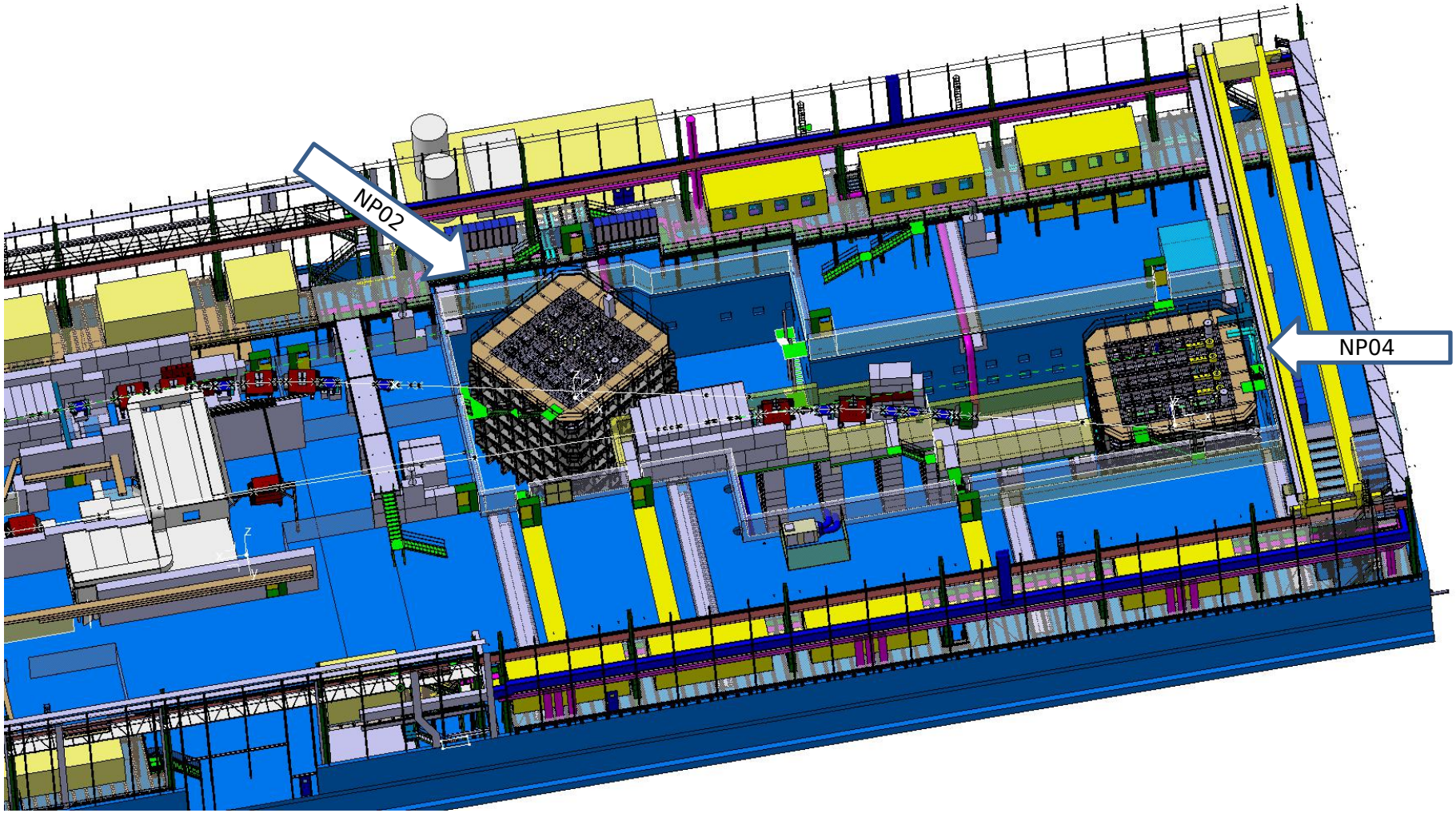
- An extension of the CERN North Area Experimental Hall is under construction.

- A test-beam facility (beamline and targets) will be placed in the extension, with a purpose-built tertiary beam from the SPS (H4) to provide various particle types.

- In addition to validating the design of the detectors with full-scale structural elements, protoDUNE will provide a unique opportunity for detector characterization and evaluation of reconstruction techniques in controlled test-beam conditions and with varying event types.

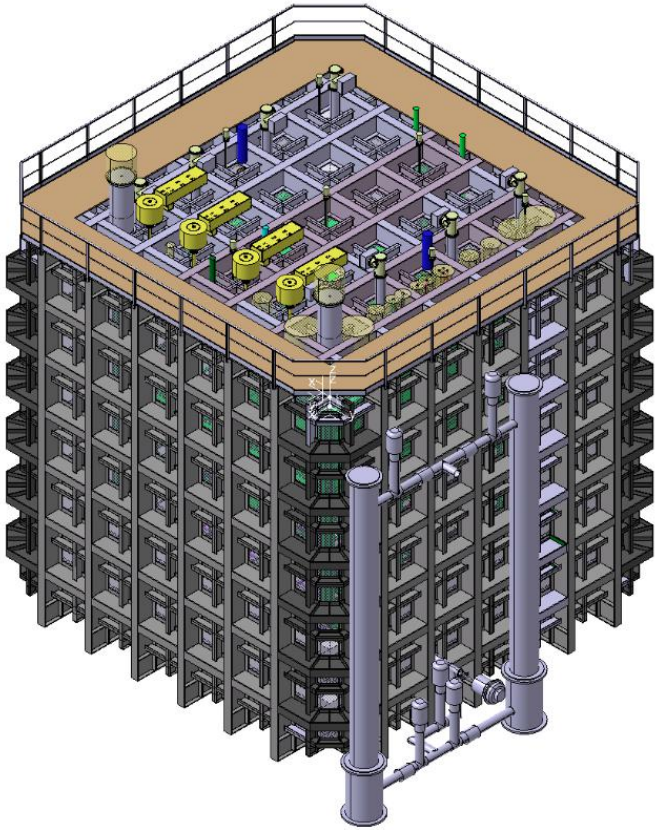
- There will be beam and cosmic ray triggers.

Layout of the experimental area at CERN

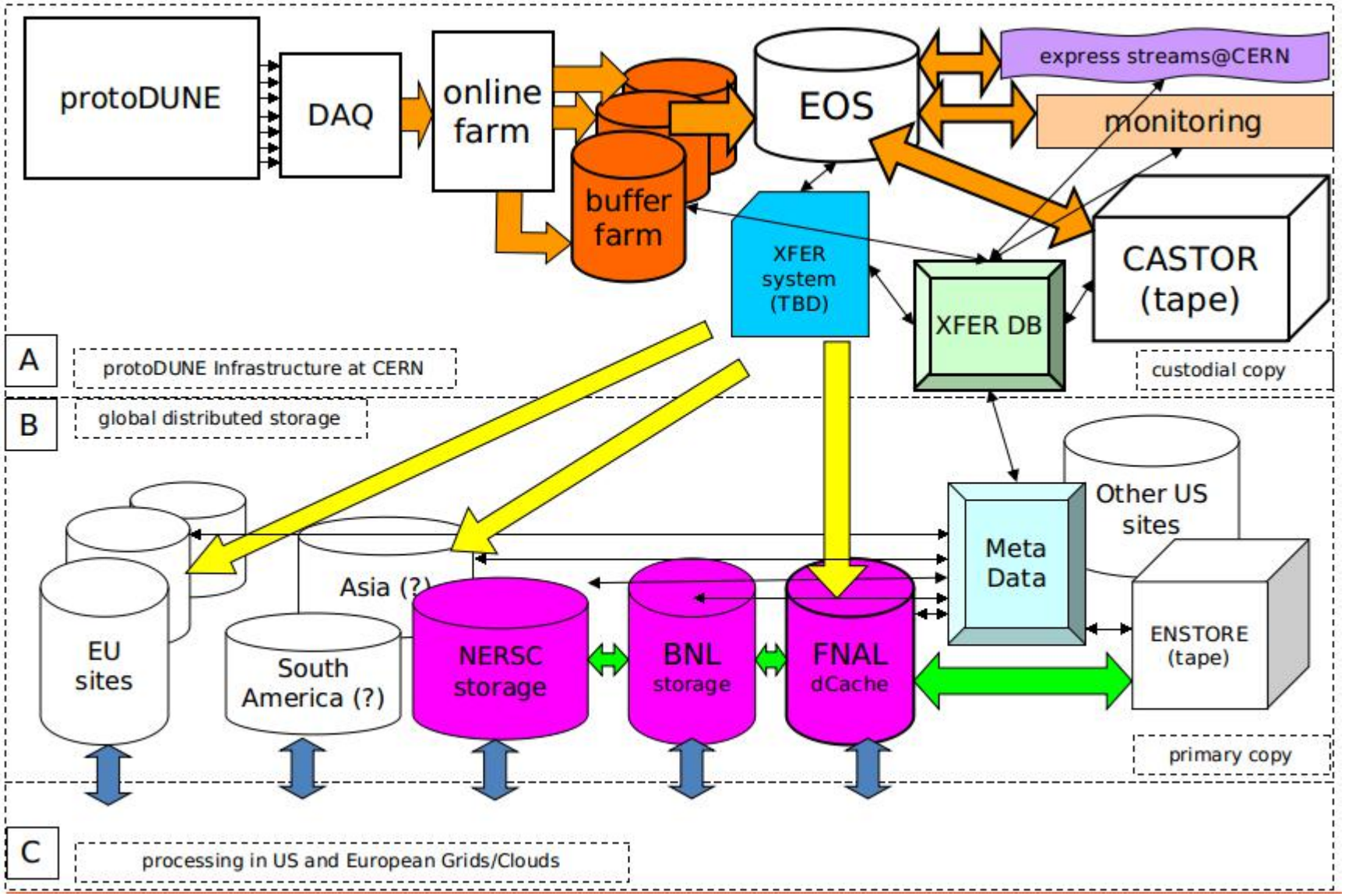


Single-Phase protoDUNE (experiment NP04)

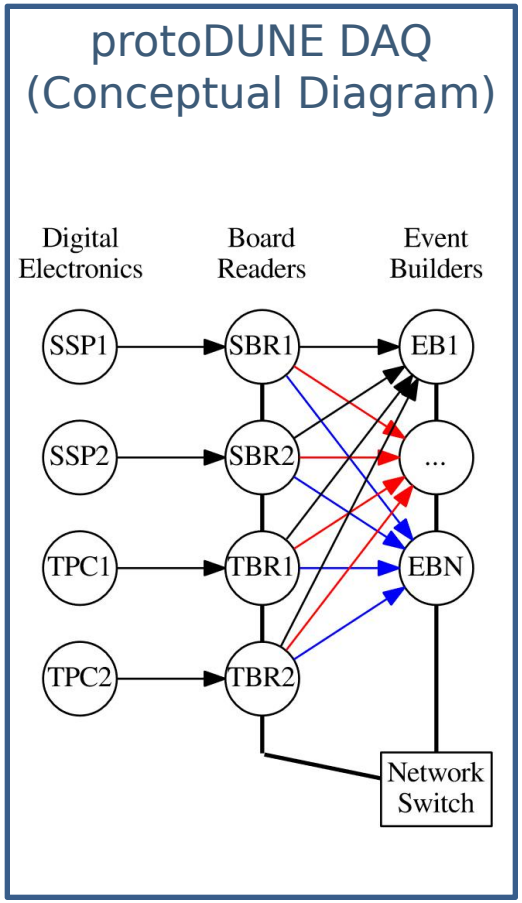
- Includes full-scale elements of the DUNE LArTPC: $2.3\times6.2m^2$ each
- TPC volume: $7.3\times7.4\times6.2m^3$
- External cryostat dimensions: $\sim11\times11\times11m^3$
- TPC channel count: 15,360
- Digitization frequency: 2MHz
- Nominal readout window: 5ms
- Nominal beam trigger rate: 25Hz
- Single readout size: 230MB
- Lossless compression factor: 4
- Post-compression peak data rate: 1.4GB/s
- Contingencies
  - higher trigger rate (50-100Hz)
  - lesser compression
  - large number of cosmic triggers
- To size up the data handling system, the benchmark of 3GB/s rate is assumed
- Nominal 20Gbps network bandwidth from the experiment to CERN central storage
- Up to 100TB daily data volume
- ~1.5PB of data to be collected during the run



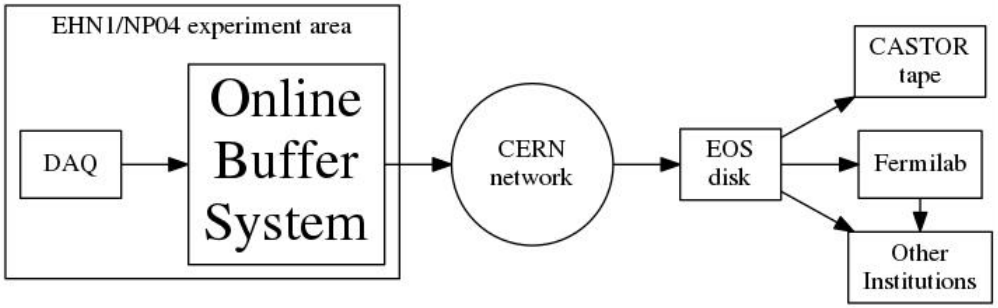
Raw Data Management: the “big picture”



Elements of the protoDUNE data flow



The online buffer is necessary to satisfy the CERN requirement for experiment-local storage and to absorb high instantaneous data rate coming out of DAQ



The online buffer characteristics:

- Nominal capacity: 300TB
- Data ingestion rate: up to 3GB/s

High bandwidth implications:

- requires high-speed storage media
- SSD prohibitively expensive at this volume
- multiple HDDs required; cf. at nominal 50MB/s write access 60 spindles will be needed

The protoDUNE online buffer: design options

- Two design options have been considered
- Option A is currently considered as preferred and option B is a back-up solution:
  A. a high-bandwidth storage appliance (such as from DELL PowerVault MD3 series) combined with a storage array, and few servers (e.g. DELL R630) to perform computational tasks
  B. a XRootD cluster with approximately 60 nodes

- Different kind of fault tolerance in the two options:
  A. fault tolerance built into storage array itself
  B. fault tolerance achieved by having a large number of nodes

- Different kind of scalability:
  A. vertical scalability (scale-up): large units of hardware need to be added to provide the required capacity and bandwidth, cf. there have to be at least two systems based on PowerVault hardware in protoDUNE
  B. horizontal scalability (scale-out): small, low-cost units can be added elastically as needed

- Different hardware:
  A. High-end hardware to be purchased form DELL
  B. Older commodity servers which can be re-purposed from other applications through a HDD upgrade

Data transport – component reuse

- Requirements have been formulated for data transmission in protoDUNE which include fault tolerance, recovery, creation of metadata, accounting, monitoring etc.
- There requirements are met by a combination of systems developed and currently in use at FNAL:
  - F-FTS: the "Fermilab File Transfer System"
  - SAM: the file metadata and catalog system
- Both system successfully utilized in Intensity Frontier and HEP experiments
- F-FTS supports automation, recovery, monitoring, is compatible with a variety of protocols and can orchestrate third-party transfers
- protoDUNE online data handling infrastructure mimics the systems used in major LHC experiments, i.e.
  - The CERN EOS distributed disk storage which receives data from the online buffer
  - CASTOR - the principal central tape storage facility at CERN
- F-FTS will be used to manage data transfer between the buffer, EOS, CASTOR, FNAL and other sites (see next slide)
- Principal storage at FNAL will consist of dCache and "Enstore" - the FNAL tape mass storage
- The dCache instance at FNAL is accessible to outside clients via XRootD

F-FTS

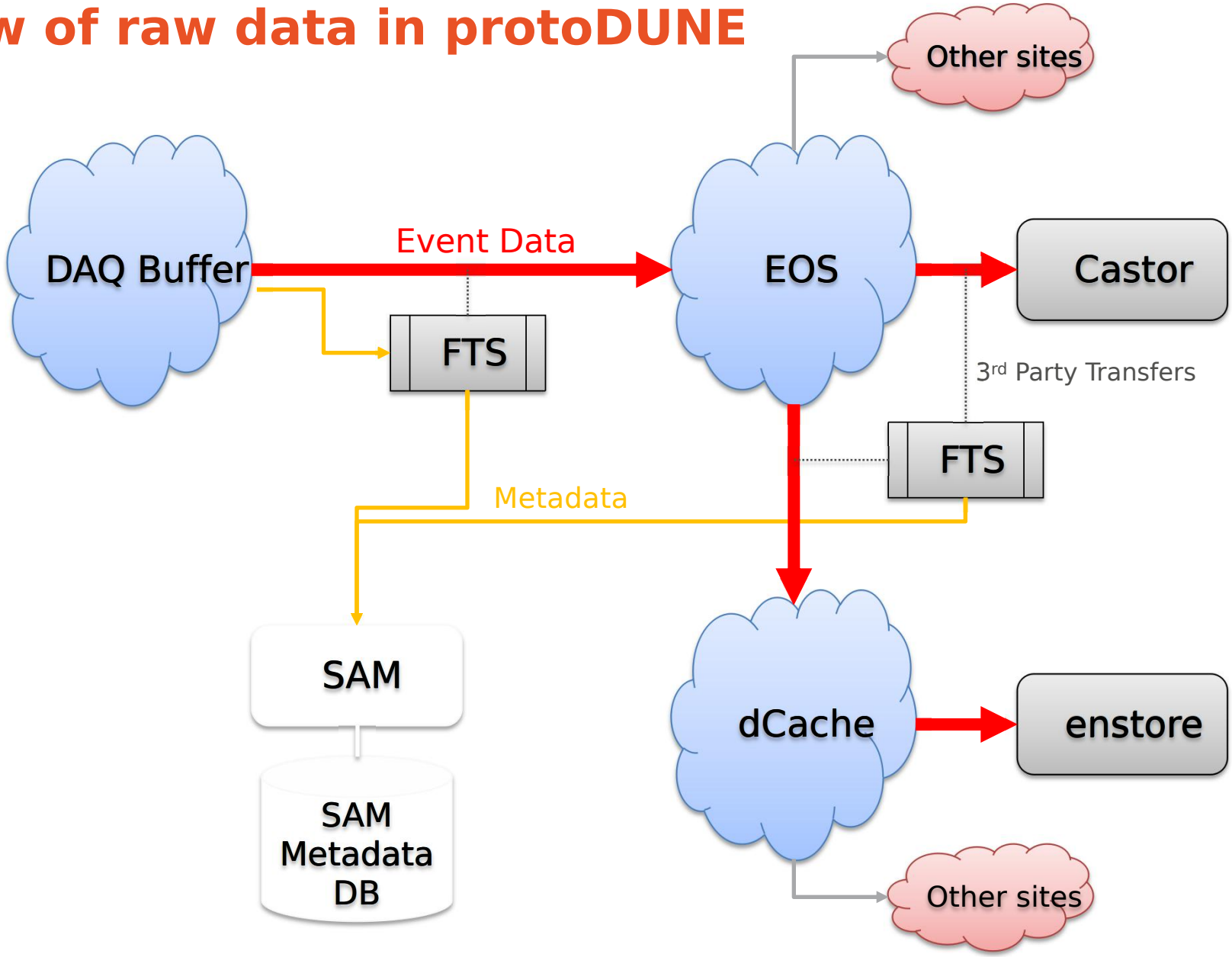
- F-FTS fulfills all of the basic requirements of a typical HEP and/or IF experiment
- It includes automation, error recovery, monitoring, interaction with the metadata system
- Supports a variety of 3rd party transfer protocols
- Most commonly, data transfer is triggered via a "dropbox" logic, whereby newly arrived files are identified and scheduled for transmission

Challenges

- For the metadata to be useful, it may be necessary to extract some of its elements by reading a large fraction or all of the data file written to the buffer
  - in the very high I/O bandwidth scenario of protoDUNE this put additional stress on storage
  - checksum is often considered a crucial part of the file metadata
  - if checksums are required from the very beginning of the data transfer chain, reading of the complete files is necessary
  - checksums require a non-negligible amount of CPU

- In adddition to the basic monitoring there are plans for "prompt processing" - express streams for data QA
  - it is defined as processing of a portion of the raw data which takes place on the time scale of minutes rather than hours after data is taken
  - the scope of the prompt processing is now work in progress
  - depending on the eventual decision and available resrouces, it can be limited to calculating the channel noise spectra (FFT), or in the high-end scenario include complete reconstruction of a subset of events
  - will require basic workflow management with high-velocity data

Flow of raw data in protoDUNE



Status and plans

- Characteristics of the protoDUNE data have been estimated and documented, and will drive the design and scale of the systems to handle raw data
- protoDUNE will leverage the CERN storage infrastructure in a way similar to the LHC experiments (cf. the data rate and volume comparable to ATLAS in Run 1)
- Data transport and metadata systems (F-FTS, SAM) developed and maintained by FNAL will be used for data transfer and distribution.
- The online buffer is a critically important component which must operate at a very high data rate (up to 3GB/s) and provide substantial storage volume
- Various design options for the buffer have been investigated and DELL PowerVault platform has been chosed as the primary option and will be tested at CERN in 2017
- Prompt processing is in the initial phase of development
- Deployment of the data handling system at CERN will start in May 2017