DUNE Interface Document: DAQ/Offline Computing

Draft Version 0.5, 27-Dec-2017

Introduction: This document describes the interface between the DUNE far detector DAQ and Offline Computing. This document describes the necessary interfaces for both DAQ and Offline Computing to complete the design, fabrication and installation of their subsystems. This document describes the elements of the scope of each subsystem at the interface between them.
[Editor's note: The information presented here is intended to be a reasonable starting point for discussion. It should be understood that particular details may change based upon further discussions.]

Division of Responsibilities: This section describes the assignment of responsibilities among the groups involved.

1. The DAQ system will reduce the data volume to a level which is agreed upon by the DAQ, offline, and physics groups. The overall data rate is described in a later section of this document. By definition, the data that is written by the DAQ will be the "raw data".
    a. [The spirit of this item is to indicate that it is the goal of the DAQ group to reduce the data volume as much as technically possible before sending it downstream.]
    b. The contents of the raw data files will include the following:
        i. Digitized signals from the detector electronics
        ii. Derived quantities that are calculated as part of the analysis of the data for triggering purposes
    c. [to be discussed] The raw data files will not contain derived quantities that are not needed by the event selection in the DAQ. For example, the raw data will not contain wholesale conversion of the digitized electronics signals into physics quantities.
2. The raw data produced by the DAQ will be transferred from the SURF local area to Fermilab using a dedicated network connection. The capability of this network link is described in a later section.
    a. The design of this link will be the responsibility of networking experts drawn from SURF, Fermilab, CERN, and university collaborators. [Who will own the design process?]
        i. [It should be noted that the current wide-area network connections into SURF do not support the transfer rates that are anticipated for DUNE, and there may be significant work to upgrade these. Moreover, the security zone model that is used at SURF may need an exception to allow DUNE data to be transferred at the necessary rate.]
    b. Fermilab and SURF networking experts will be responsible for constructing the link. The costs of this construction will be paid by the DUNE project. The link will be fully operational by 2023, and procurement of the necessary hardware, etc.

will be executed with sufficient lead time to allow for installation and testing before 2023.

c. The operation of the link will be managed by Fermilab and SURF networking experts. The operations costs will be covered by who? [What fraction might be covered by Fermilab Detector Operations funds?]
d. A general-purpose WAN connection will be provided to the online environment at SURF. This networking is not part of the interface between DAQ and Offline Computing, per se; it is only mentioned here for completeness.

3. The DAQ system will provide a disk buffer that will be used to temporarily store the raw data if, or when, the network connection to Fermilab experiences a problem. It will also be used to store the data from supernova triggers, which produce a relatively large amount of data over the course of 10s of seconds. The size of this disk buffer is described below.
4. Near-line analysis, production reconstruction and analysis, and all other offline processing and analysis of the data will be the responsibility of Offline Computing.
5. Offline Computing will be responsible for the development and maintenance of the software that performs the transfer of the data from the DAQ disk buffer to Fermilab. The DAQ system will provide the CPU cycles, memory, and disk that are necessary to run this software within the DAQ cluster at SURF. The DAQ team will install all software libraries, etc. that are needed to support the data-transfer software, and they will develop and maintain the mechanism that alerts the data-transfer software when new files become available. It will be the responsibility of the data-transfer software to delete stale files once they are safely transferred to Fermilab, based on criteria that are mutually agreed upon.
   a. [The language of “CPU cycles, memory and disk” to run this software is intended to be somewhat vague. Whether or not dedicated servers are provided for this function, or if this software runs on servers that are also involved in other tasks, is an implementation detail that will be left to the DAQ group.]
6. The DAQ team will be responsible for the development, packaging, and documentation of a reference implementation of the software that is used to access and decode the raw electronics data. This software may be used directly by both online and offline, or it may be used as a reference for software developed by the offline.
7. The DAQ system will provide the framework for real-time data quality monitoring, and both groups will have responsibility for developing the algorithms that perform the analysis of the data and generation of the monitored quantities.

Physical Locations: This section describes the physical locations of various data-taking and analysis functions that help define the interface.

1. The data reduction that is performed by the DAQ will take place within the local area of the SURF lab (on site, or within a few kilometers on the surface).
   a. [As decisions are reached regarding the design of the DAQ system, we can add a reference to documents which describe that design here.]

2. The Offline Computing resources (servers, disks, software repositories, etc.) that will be used to perform offline analysis tasks will not be located in the vicinity of the SURF lab.
    a. [We can/should reference Computing documents that describe the plan.]

Rates, Sizes, and Latencies: This section describes the expected data rates, data sizes, and latencies in accessing the data. [These values may need to be updated as the results from various simulation activities reach their conclusion. Once those studies are available, we will include references to the relevant document(s) here.]

1. During stable running, the data volume that will be produced by the data acquisition systems of all four cryo modules will be no larger than 30 PB/yr.
    a. [We will provide a reference that motivates this rate as soon as one is available.]
    b. During commissioning, this rate may be substantially higher, and the value that has been discussed among members of the DAQ group is 30 PB per 'whatever timeframe encompasses the commissioning phase of each cryo module'.
    c. During the construction of the full far detector (four cryo modules), the full 30 PB/yr bandwidth may be produced by whatever subset of the full detector exists at each point in time.
    d. From the overall data volume, a typical instantaneous rate can be calculated:
        i. x (GB/sec) = y (PB/year) * 0.033 GB-year/PB-sec
        ii. 30 PB/year = 1 GB/s = 8 Gbps
        iii. If the duty factor for data-taking is noticeably different from 1.0, then that factor should be taken into account in this calculation.
    e. Based on discussions so far, it is expected that the deviations from the average rate will primarily be due to the readout of data from supernova triggers. (The fake rate of supernova triggers is still under study.) Each supernova trigger is expected to generate ~120 TB for a 10-20 second time window.
    f. [What information do we have about the data volume associated with calibrations?]
2. The networking link between SURF and Fermilab will provide bandwidth of 20 Gbps.
    a. This will provide a safety margin of approximately 2 above the typical instantaneous data rate, which is calculated above to be 8 Gbps.
    b. Along with providing a safety margin, this bandwidth will be useful in draining any backlog of data that builds up as a result of a temporary disruption in the link.
    c. The extra bandwidth will also allow the transfer of supernova trigger data over the course of ~1 day in parallel with the nominal transfer rate.
        i. 120 TB / 12 Gbps = 81920 sec = 22.8 day
3. The disk cache size will be sufficient to hold the data that is produced during 3 days of normal running along with the data that is produced by one supernova trigger. [Should it be 7 days or some other value?]
    a. [To help determine a reasonable duration, we'll ask people at SURF what the longest downtime in the WAN connection has been.]
    b. A 3-day buffer for data being written at 8 Gbps corresponds to ~250 TB.

    c. The size needed for a supernova trigger is ~120 TB.
    d. The sum of these two is 370 TB.
4. [What latency is desired for access to the raw data by near-line processing?]

Integration: TBD

Commissioning: TBD