ProtoDUNE-Single Phase Software and Computing Project Plan

Tom Junk and Andrew Norman
DUNE Computing Coordinators, Fermilab
Ruth Pordes
ProtoDUNE-SP Computing Liaison

Version 10.4, June 16, 2017

This document outlines the full scope of the ProtoDUNE Single Phase Software and Computing deliverables. It points to separate documents for more details of the ProtoDUNE Single Phase organization's deliverables and gives more detail of the deliverables of the DUNE Software and Computing Organization to ProtoDUNE Single Phase.

History:

- Starting to add feedback from reviewers within ProtoDUNE-Single Phase.
- Starting with draft version 8.5 this document further classifies deliverables into those of the DUNE Core Computing and ProtoDUNE SP (protoDUNE Single-Phase) organizations.
- Starting with draft version 4.0, this document classifies deliverables outside of the Software and Computing domain as requirements, since components interfaced to Software and Computing must exist and interfaces must be defined and tested.
- Starting with draft version 9.0, DAQ deliverables are removed (they are kept in a backup document).

Table of Contents

Introduction **3**
Scope 3
Organization 4
Timeline 4
Effort 4
Global System Topology **4**
Data Acquisition Domain **5**

Prompt Processing Domain **5**

Data Management and Long Term Storage Domain **5**
Hardware Resources 6
File Transfer and Cataloging Services 7
Data Cataloging System 7
File Transfer System: EHN1 to EOS 8
File-Transfer System: EOS to Fermilab dCache 8
Analysis Data File Management 9
Metadata Schema Specification 9
Processed Metadata and Data File Management Specification 9

Offline Processing, Simulation and Analysis Domain **10**

Calibration and Configuration Domain **10**

Beam Instrumentation **10**

Databases **10**
Hardware Database 10
Calibration Database 11

Installation, Operations and Maintenance **11**

Other DUNE Software and Computing Services **12**
Security and Backup Services 12
Operational Security 12
Logfile Persistency Policy and Automated Management 12
Software Management 12
Production 14
Collaborative Tools 15
General Computing Services 16

External IT Components **19**
Service/Help Desk 19
E-mail 19
Phone and Video Conferencing 20
Calendar/Meeting Scheduling 20
Collaborative Websites/Document Sharing 20

Summary of Effort needed and Current Availability **20**

Introduction

The DUNE experiment requires a large, international effort in order to design, construct, and operate the beam, the near detector and the far detector in order to meet its physics goals. The largest portion of the investment is in the Far Detector (FD), a 40-kton (fiducial) liquid argon time projection chamber (LArTPC). The FD design must be validated with a comprehensive prototype. Two LArTPC technology choices are currently foreseen -- a single-phase design much like that used in ICARUS and MicroBooNE, and a new dual-phase design. Both of these technologies will have prototypes installed and tested in a test beam at CERN.

The ProtoDUNE Single-Phase prototype is called ProtoDUNE-SP (protoDUNE Single-Phase), or CERN Experiment NP04, and the dual-phase prototype is called ProtoDUNE-DP (PD-DP), or CERN Experiment NP02 and/or or WA105.

The test-beam runs will collect of order 2.5 PB of data each for NP02 and NP04, in a short (of order 120 days) run in 2018, the end of which is constrained by the start of CERN's Long Shutdown 2 (LS2). Software and computing is required to meet the needs of the prototypes: construction, assembly, testing, and commissioning, validating the quality of the data, storing it for future analysis, providing for common processing stages, and providing access to the raw and processed data to collaborators for analysis.

This document outlines the software and computing deliverables for protoDUNE Single-Phase, listing when the deliverable is needed, the system topology, boundaries and interfaces, needs for commissioning and data-taking, the required components for offline processing and analysis, and external components. If a deliverable is required for more than one phase of ProtoDUNE-SP operation, it will be listed the first time it is needed. Most deliverables require long-term maintenance and support. Synergies and components shared with ProtoDUNE-DP will be described.

Scope

The scope of this document is Software and Computing (S&C) deliverables, those delivered, coordinated and communicated under the DUNE Core Computing Group (DUNE S&C) coordinators and protoDUNE Single-Phase organization per se. The components, timelines and deliverables of data acquisition (including artdaq), prompt processing, beam instrumentation and physics (calibration, immediate processing, reconstruction, analysis, simulation etc.) algorithm software that are delivered by groups within the protoDUNE Single Phase organization are described in separate documents available, together with this document, in the DUNE Document Database #3450.

ProtoDUNE Single-Phase relies on local and external computing services from several internal organizations at Fermilab (including the Scientific Computing Division (SCD)), CERN (including CERN Neutrino Platform (CENF) and CERN IT). These dependencies and interfaces are included but the effort and organization for delivering them is not.

The overall scope and goals of the computing and software is guided by the ProtoDUNE-Single Phase TDR.

Organization

The organization (Management Plan Document # 2145 in the DUNE document database) of DUNE core computing and of protoDUNE Single-Phase is tailored to allow clear responsibility and interfaces between deliverables needed for the end-to-end protoDUNE Single-Phase computing solutions. In the areas of physics algorithm and other physics software deliverables DUNE Physics Working Groups and protoDUNE Single-Phase work closely together on common components. There is, in general but not completely, overlap between the names of these organizations and the components and domains of the system described in this document.

Timeline

The schedule for installing, commissioning, and running protoDUNE Single-Phase is highly compressed due to the projected date of the start of CERN Long Shutdown (LS2) in October 2018. The detector will continue to take cosmic-ray data for a period afterwards, of a length to be determined. The deliverables timeline is not included in this document and is being developed to follow the overall schedule available from the protoDUNE Single-Phase leaders (Flavio Cavanna, Gina Rameika, Christos Touramanis; as of 3/2017 it is available as Document #847 - page 20-24 in the DUNE Document Database). Currently the first APA is scheduled to arrive at CERN for testing in June or July of 2017 and several important S&C deliverables need to be in place for commissioning and testing when that is done.

Effort

An indication of the effort needed for each deliverable is included. Detailed effort profiles, planning and needs are kept by the relevant coordinators in DUNE and/or protoDUNE Single-Phase.

Global System Topology

The protoDUNE Single-Phase computing model is divided into discrete domains within the logical computing infrastructure. Each domain is separated from neighboring domains, but a

well-defined set of interface specifications is provided which allows for information/data to pass between the domains.

In this model the highest level domains defined in the ProtoDUNE model are the

- Data Acquisition (DAQ)
- Prompt Processing
- Data Management and Long Term Storage
- Offline Processing, Simulation, and Analysis
- Calibration and Configuration
- Beam Instrumentation
- Operations (including Commissioning)

Each of the computing deliverables discussed in this document are classified as belonging to one of these domains, and their impact are assessed based on the interaction points in the other domains. Note that “Domains” and “Organizational Responsibilities” can have overlapping names but are not always the same.

Data Acquisition Domain

The scope and deliverables in the DAQ domain are discussed in DUNE Document Database #3450, ProtoDUNE SP DAQ Overview.

Prompt Processing Domain

The scope, deliverables and timeline of the “prompt processing group” within the DP-SP organization are described by the prompt processing plan, DUNE Document Database #3219.

Data Management and Long Term Storage Domain

The following components are essential deliverables for the success of the long term/archival storage domain for protoDUNE Single-Phase. These deliverables provide the data pathway for data to migrate from the DAQ domain to durable and archival storage, for the data to be fully catalogued, and for the data to be retrieved by applications in the analysis domain. It includes the metadata services associated with the event data, as well as the management of the metadata and event data itself. Future deliverables will address the preservation of the data, which is to be directly accessible for 4 years after data taking, and available through a long term archive through the first data runs of the DUNE experiment itself.

The deliverables rely on services provided by the, external, Fermilab SCD and CERN IT organizations. Within DUNE S&C responsibility is part of the scope of the DUNE/PD/Fermilab/CERN Interface Group working with each of the protoDUNE Single-Phase

groups as needed. Meetings of this group are held bi-weekly. This domain depends on the DUNE Database group for some deliverables.

The Design of the Data Management system is available as a separate document at DUNE Document Database #1938.

Hardware Resources

As written in the Executive Summary of the software and computing plan, the size of the storage requested at Fermilab and CERN is listed below. Each of the resources has either been requested to match the FY17/FY18 need profiles for ProtoDUNE or has been requested and allocated for use by the DUNE collaboration.

Storage	**Type**	**Data Type**	**Status**	**Total(TB)**
Castor	Tape	Raw	Requested (FY18)	6000
EOS	Disk (Durable)	Raw	Allocated	700
EOS	Disk (Durable)	General	Allocated	3000
Enstore	Tape	Raw	Requested (FY18)	6000
Enstore	Tape	General	Allocated	1000
dCache	Write Cache	Raw	Allocated	40
dCache	Read Cache	General	Allocated	2000
dCache	Temp Shared	General	Allocated	1000
dCache	Disk (Durable)	Raw	Allocated	200

The following efforts are needed to support this hardware:

Storage	**Person Power**
EOS Disk Space at CERN	A small amount of time will be needed to make the request and coordinate with appropriate personnel, less than one week of a senior FTE. More time will be required if the request is denied or resource contributions from DUNE need to be identified.
Tape Storage on	A small amount of personpower is needed in order to

CASTOR at CERN	register the request with CERN computing and follow up with the needed project resources.
dCache Disk Buffer	At Fermilab: A request needs to be made to the SCPMT, and sufficient resources allocated within SCD to ensure that the disk space is available.
Tape Storage at Fermilab	DUNE/ProtoDUNE computing coordinators can make the request and keep the SCPMT up to date with the experiment requirements. SCD needs to be able to purchase tape in time for its use.
Offline Disk Storage	Several categories of disk space are required for offline analysis of the ProtoDUNE data. Users need space for building code, for storing temporary data sets, and for storing datasets with longer-term usage, such as ntuples. In order to access data that are stored on tape, a caching system is provided -- EOS at CERN, and dCache at Fermilab. Ongoing maintenance is required in order to use this space effectively -- cleanup of old files and requests for new space when it is clear that cleanup alone cannot meet the requirements of the Collaboration.

File Transfer and Cataloging Services

Data Cataloging System

Description: The data catalog is provided by the SAM data handling suite. It consists of applications which allow for the recording of information regarding individual files and the capability to queue on this information to locate and classify different collections of files. The data catalog is integrated with a storage aware replica catalog in this model and allows for the optimized retrieval and transport of files between storage and analysis applications.

Timeline: The data catalog system for protoDUNE Single-Phase is a subset of the catalog system that is to be used by DUNE. The catalog system for DUNE has already been deployed into production. Modifications are required to fully integrate the catalog with newly defined storage locations within the CERN computing campus (EOS and CASTOR based storage). Additional work is required to provide categorization information (i.e. protoDUNE Single-Phase metadata fields) that is specific to protoDUNE Single-Phase data and commissioning work.

Milestone Date: Q4 CY2016 for the initial version.

Personpower Required: Stu Fuess's project includes the necessary personpower needed to accomplish these goals.

File Transfer System: EHN1 to EOS

Description: This system constitutes an automated system which is responsible for transferring acquired data from the DAQ domain to one or more endpoints on the CERN EOS disk storage facility. This system also provides the automated registration of files with the data catalog system and perform the automated cleanup of files from the short term storage systems in the DAQ computing domain. This system will also perform the steps required for validation of the data being stored to the CASTOR archival tape system at CERN. This system requires setting up the proper software components with the appropriate privileges, as well as establishing the authentication needed to write data on remote systems. Regular monitoring and maintenance is required to resolve problems with the file transfer and authentication, such as expired certificates.

Timeline: A rudimentary system should be in place for the Vertical Slice Test in June 2017 -- performance is not at a premium for this. An optimized system needs to be in place by March 2018.

Personpower Required: Fermilab SCD effort. Stu Fuess's project will provision and test F-FTS for this purpose.

File-Transfer System: EOS to Fermilab dCache

Description: This system constitutes an automated system which is responsible for replicating acquired data from the endpoint that has been established on the CERN EOS disk storage facility to a similar (but distinct) endpoint on the Fermilab dCache disk storage system. This system also will provide the automated updating of replica information within the data catalog and provide any automated cleanup that is needed as part of the data flow between the systems. This system will also perform the steps required for validation of the data being stored to the Enstore archival tape system at Fermilab. This system requires setting up the proper software components with the appropriate privileges, as well as establishing the authentication needed to write data on remote systems. Regular monitoring and maintenance is required to resolve problems with the file transfer and authentication, such as expired certificates.

Timeline: A rudimentary system should be in place for the Vertical Slice Test in June 2017 -- performance is not at a premium for this. An optimized system needs to be in place by March 2018.

Personpower Required: Fermilab SCD.

Analysis Data File Management

Metadata Schema Specification

Description: The raw and processed data files are accessed using the data file catalog by queries on the metadata. It is anticipated that the schema will consist of a limited number of fields -- the run number(s), the begin and end dates and times, the names of contributing components, the number of events, the data tier, the file checksum, the version of the DAQ software and the software used for processing, run flags such as whether the run is beam physics, test, calibration, or other, and a simple indicator of the beam conditions are examples of metadata fields that are necessary to include in order to make efficient queries of the file catalog. Additional fields will be required to specify Monte Carlo-specific items, though these should be kept to a minimum, as the user-facing file catalog on the web should explain in detail what the Monte Carlo details are. The old LBNE metadata specification is clumsy in some respects and lacks support for ProtoDUNE-specific metadata.

Timeline: An initial specification of the metadata should be incorporated in the DUNE SAM database as early as March 2017. The early test data may be more varied in format than the running data. Metadata fields can be added relatively easily.

Personpower Required: Small. These metadata can be decided upon in a meeting of interested people, written up, and circulated among the Collaboration. S&C coordinators, with a single postdoc or student, can draft the metadata and propose it for Collaboration review.

Processed Metadata and Data File Management Specification

Description: The processed metadata and data file naming will need specification and ongoing management in coordination with the Analysis groups.

Timeline: An initial specification of the metadata should be incorporated in the DUNE SAM database as early as March 2017. The early test data may be more varied in format than the running data. Metadata fields can be added relatively easily.

Personpower Required: Small. These metadata can be decided upon in a meeting of interested people, written up, and circulated among the Collaboration. S&C coordinators, with a single postdoc or student, can draft the metadata and propose it for Collaboration review.

Offline Processing, Simulation and Analysis Domain

The scope and deliverables are the responsibility of the protoDUNE Single-Phase Data Reconstruction and Analysis group. The Reconstruction and Analysis plan is available at DUNE Document Database #3450. The deliverables depend heavily on contributions from external groups including: DUNE reconstruction, existing experiments using LArTPC detectors, and the LArSoft project.

Calibration and Configuration Domain

The calibration and configuration plan and deliverables span multiple organizational units including DAQ/Online, prompt processing and Offline. The detailed plans are included in the documents at DUNE Document Database #3450

Beam Instrumentation

TBA

Databases

Several databases will be developed in support of the Domains above. A document giving a more complete plan for their design and implementation is in preparation.

Hardware Database

Description: A web-accessible hardware database is required in order to keep track of component locations, test results, current status, and history information. The database needs to be accessible from web browsers as well as from computer programs via an API. The database interface should provide tools for making summary reports such as installed fraction vs. time in order to satisfy the requests of review committees and funding agencies. Spreadsheet import is required. Currently use of this database is optional, though some sort of database will be needed to be maintained by each hardware group to track the progress of their components, in order to be able to communicate with the DUNE Collaboration and funding agencies regarding their progress. Calibrations that are performed on test stands that are needed for running and analysis of the data need to be communicated to the offline jobs either in the hardware database or the calibration database.

Timeline: Start as soon as possible, as protoDUNE Single-Phase components are already being fabricated and currently this information is stored in private spreadsheets.

Personpower Required: Fermilab's SCD will support instances of hardware databases that we request, based on the existing ones used by NOvA and other experiments. Users of the hardware database need to communicate the required schema to Fermilab staff and they will be implemented in the database.

Calibration Database

Description: The calibration databases refer to the need a system for storing and retrieving descriptive constants representing the states of the detector, that have a time indexed or based validity. Typical examples of tables that fall within calibration database systems are detector channel maps, electronics bad channel lists, pedestal measurements, gain calibrations on readouts, measured drift velocities within media, and other measurable or derived quantities that need to be know for proper configuration or interpretation of the data from the detectors.

Timeline: The work on this can start at any time, with default channel maps and detector properties entered. It must be ready by the time data are ready for analysis, in June 2018.

Personpower Required: Fermilab SCD can provide an instance of this database and hardware to run it on, likely shared with other experiments. The protoDUNE Single-Phase Calibration group will provide meta-data layout and payloads, with the database group providing libraries for data ingest and access.

Installation, Operations and Maintenance

Ongoing effort for operating the hardware and software systems as well as evolving them based on experiment needs. While each of the system domains will develop plans for this a summary of the needs is started here for reference:

- DAQ and Online component support
- Online monitoring
- Prompt Reconstruction support
- Monitoring of the data movement and storage
- Validation and correction of meta-data and analysis catalogs
- ...

Other DUNE Software and Computing Services

Security and Backup Services

Operational Security

Description: Operational security is provided by the Fermilab and CERN Security groups, consistent with that provided to other Neutrino and Intensity Frontier experiments, in collaboration with the DUNE software and computing organization.
Personpower Required: none from ProtoDUNE.

Logfile Persistency Policy and Automated Management

Description: The ProtoDUNE-SP experiment will establish a policy on the retention of logs and other diagnostic information within the DAQ environment. The experiment should also define an organizational structure for this information. A rules-based cleanup agent (currently provided as part of the SAM data handling suite) should be configured to implement the policy. Logfiles should be stored in DAQ server disk storage and all of the ones that accumulate need to be identified and backed up. New categories of logfiles may be created during the commissioning phase and so this task needs a low level of continuous attention. Logfiles can accumulate on a RAID6 disk array which gives a certain level of safety while a system is created to back them up. An index of archived logfiles should be made available on the web.

Timeline: This project should start after the data transfer and metadata creation systems are up and running, so that similar data handling techniques can be used to get logfiles to tape.
Personpower Required: 1 week time for a skilled shell-script programmer is needed for this task.
Interface to S&C: The archived logfiles need to be kept in persistent storage, accessible to users. S&C can provide to this storage and cataloging for easy retrieval.

Software Management

The following are the responsibility of DUNE software and computing software management group.

Software Repository

Description: The code needed to analyze the protoDUNE Single-Phase data is to be stored and versioned in a repository at Fermilab in order to meet the Data Management Plan requirements. Because the software depends on LArSoft and *art*, it is natural to group the repositories needed to analyze the protoDUNE Single-Phase data with the other LArSoft repositories. This repository will contain the prompt processing code components that depend on LArSoft. A separate repository containing data reformatting and unpacking code will be shared with the online and DAQ efforts.
Timeline: These repositories exist now and have been used for the 35-ton prototype. New ones with appropriate names can be created easily. Software needs to be archived and made retrievable for the duration of the DUNE experiment. This may involve migration to newer systems over time.
Personpower Required: No additional personpower beyond software authors and libraries/release managers.

Documentation Wikis
Description: Instructions for new users, code developers, shift takers and run coordinators needs to be drawn up. Wiki pages provide the easiest ways to keep such documentation up to date, as they do not require specialized document preparation systems. The wikis for shift taking at CERN should be hosted at CERN to minimize the risk due to downtime.
Timeline: Each software repository in Fermilab's Redmine system comes with a wiki for documentation and tools. Wikis exist for the 35-ton prototype and the offline repository. New ones can be created as needed. At CERN,solutions such as twiki may be easier to implement and maintain.
Personpower Required: No additional personpower is needed beyond the DAQ/online and software authors. Group leaders need to encourage members to write documentation and keep it up to date.

Build System
Description: All protoDUNE Single-Phase software needs to be compiled and installed. A common build system is supplied and supported by DUNE S&C software management group
Some domains may have domain specific build systems to meet specific needs. For example:

- The DAQ group may use its own build tools for its software, though *artdaq*-related components will be built most easily using mrb.
- The prompt processing group may use its own build tools for fast turnaround of software versions that include new payload algorithms.

Timeline: mrb and cetbuildtools already exist. SPACK tools should be ready in 2017, but are not needed for ProtoDUNE-SP operations or analysis. They may require some effort to migrate to as *art* and LArSoft move to them.

Personpower Required: Each organization will supply its own effort for domain specific build software. For the common DUNE software management build system only maintenance is required.

Code Distribution System
Description: Pre-built, released code must be distributed to interactive and batch workers, as well as online and DAQ nodes. Released code must be packaged for easy download for installation on user computers for code development away from networks. CVMFS fulfills the role of distributing source and pre-built binaries to computers on the network, and scisoft.fnal.gov provides pre-built tarballs and installation scripts that can also compile code from source on appropriately configured computers. For the online and DAQ system, it may suffice to distribute pre-built code only to the online cluster.
Timeline: CVMFS and scisoft.fnal.gov distributions already exist. Online and DAQ code repositories need these capabilities set up. Timeline for online and DAQ CVMFS repositories -- March-April 2017.

Production

The following are the responsibility of the DUNE software and computing production group:

Processing Campaign Management Tool
Description: The large datasets produced by ProtoDUNE-SP will require significant automated processing steps. The production of the data for analysis may require more than one job per input file, to execute serially, requiring a DAG. Input and output cataloging, failed job handling, and resource management functionality is needed in a way that minimizes the labor burden on the data production team.
Timeline: The production system needs to be in place for the July 2018 data run, and test versions of it can start with the Vertical Slice Test data in the Summer of 2017.
Personpower Required: Fermilab's SCD provides POMS development, documentation, and support.

ProtoDUNE-SP Processing Team
Description: Keeping up with the data, calibration, and reprocessing requires personnel to take responsibility for running jobs, checking for failures, updating documentation, and solving problems. This team can also shepherd the Monte Carlo production.
Timeline: Monte Carlo Challenge 8 took place in early 2017, and used project.py. Two persons were needed to run Monte Carlo Challenge 7 using project.py.
Personpower Required: About 0.3 FTE averaged, but with at least two people, until July 2018, at which point 2 FTEs will be needed until the end of the data run and the first processing step is over, at which point the reprocessing and MC steps will require 0.3 FTEs on average again. About 0.2 FTE of a senior person's time will be needed for supervision.

Collaborative Tools

The following are the responsibility of the DUNE software and computing Collaborative Tools organization.

Document Database
Description: A repository for documents that can be written to and read by DUNE collaborators is essential to the proper functioning of PD--SP. Documents need to be versioned and indexed via metadata, and a tiered protection scheme is required. It must be accessible via the web.
Timeline: DocDB already exists and meets the requirements, though routine maintenance is required for the length of ProtoDUNE-SP.
Personpower Required: Maintenance only. No additional personpower is required beyond that needed to maintain the DUNE DocDB. Migration to future platforms may be required in the future, during the DUNE run, in order to accommodate evolving technology.

Data Catalog Web Site
Description: The data are to be cataloged with metadata using SAM. More information needs to be specified for users to select the right datasets, and to communicate to users what the available datasets are, as well as to provide documentation about how to access the data.
Timeline: dune-data.fnal.gov already exists. It needs to be updated and maintained as new data and Monte Carlo samples arrive and as access patterns evolve.
Personpower Required: None beyond the already-accounted personpower requirements will be needed for maintenance of this site.

Analysis Group Web Pages - DRA responsibility to discuss
Description: Analysis groups need a place to coordinate activities and communicate this coordination to group members and DUNE collaborators who would like to become members. These pages should contain documentation about how to join, what analysis topics are covered and what topics need additional coverage, and instructions for accessing resources such as Monte Carlo and data sets relevant to the analyses conducted by the group. Results that are in preparation or review may be posted to the analysis group web page, and drafts of public results pages shown here, or on another internal DUNE analysis page.
Timeline: These can be created at any time, as needed by the analysis groups.
Personpower Required: A small amount of time for setup. If Redmine is used, then the wiki infrastructure already exists.

Public Results Web Pages
Description: A web site where public DUNE (and ProtoDUNE) results are available without the need of a password or other authentication is needed to provide physicists seeking results, plots, documentation, and illustrations for conference talks and other fair use. It should be hand-editable and also allow for automated scripts to perform maintenance on it by extracting plots and html components from DocDB. Tools exist on NOvA to do this, and can be ported to DUNE in order to reduce the effort needed.
Timeline: A web site, http://internal.dunescience.org/plots/ has already been set up and the scripts to collect graphs and images from appropriately specified DocDB entries has been written and tested. A public version should be set up to display publicly approved plots, likely with less automation and more hand-customization to make the web display appealing.
Personpower Required: This topic has already been addressed as of Feb. 27, 2017. Only the public web page needs to be set up.

General Computing Services

The following are the responsibility of the DUNE software and computing organization, relying in many cases on External IT deliverables and services. The effort/person power indicated is that from the DUNE and protoDUNE Single-Phase collaborations/programs themselves. If the deliverables are completely external no DUNE/protoDUNE Single-Phase effort is needed except for interfacing, using and tracking.

User Authentication Support
Description: In addition to supporting interactive logins, support is needed for batch job submission and running, as well as access to data on disk and tape.
Timeline: This is an ongoing task, starting now and continuing as long as ProtoDUNE-SP data are analyzed.
Personpower Required: Fermilab SCD and the Service Desks of both CERN and Fermilab should respond to authentication requests. A small amount of time, approximately 1 FTE week per year, will be needed from the DUNE S&C coordinators to update certificates, including reading and understanding documentation.

Grid CPU Allocation
Description: Simulation, reconstruction, and analysis jobs need to run on grid compute nodes. The allocations are reviewed yearly at Fermilab, which requires estimation of required CPU needs as well as special requests such as large-memory nodes and short-term spikes in usage. Recommendations to use OSG resources creates a bit of effort on DUNE's part to adapt code and scripts to be able to run on computers with fewer services available than Fermilab grid nodes.

CERN Tier-0 nodes provide an important computing resource -- Approximately 1500 batch slots have been formally allocated to DUNE. These batch slots can be used for detector simulation, reconstruction, analysis, and testing for production until operations start, and will play an important role in the production processing of experiment data.

Timeline: Every year in January and February, resource requests are collected and presented to Fermilab's Scientific Computing Portfolio Management Team. Moving forwards, requests will need to be made to the equivalent team at CERN, and negotiations for OSG sites will be required.

The DUNE/Fermilab/CERN interface group coordinates bi-weekly with CERN-IT on resource needs and plans. An operations bi-weekly technical meeting bewteen ProtoDUNE and CERN-IT members provides communication of how to use the CERN resources, and a forum for more detailed technical discussions.

As of February 2017, 500 Tier-0 batch slots have been allocated, and 1500 will be available starting in August 2017. This allocation has been made for a period of four years.

Personpower Required: Two weeks of time from the DUNE S&C coordinators are needed each year to request resources, compile the requests, prepare a presentation, give a practice presentation, and give the presentation to the SCPMT. Additional effort is required throughout the year to monitor grid usage and encourage DUNE collaborators to use the OSG.

Data Movement Software - GridFTP and XROOTD

Description: GridFTP is a standard tool for high-throughput file transfer to centralized storage locations from grid nodes or DAQ systems, and vice versa. protoDUNE Single-Phase uses this tool or similar ones to transfer data. XROOTD - TBA

Timeline: GridFTP is already installed at Fermilab but it may need installation on the CERN DAQ and CENF nodes. XROOTD - TBA

Personpower Required: Fermilab SCD and CERN system managers provide this service and assist with authentication and debugging.

Job Management Services

Description: This includes the Fermilab FIFE job management service (jobsub), CERN batch system job management service, and potentially other job management services for other resources use and needed by protoDUNE Single-Phase. These are tools for submitting batch jobs to the grid and other distributed resources. It must scale to accommodate many users and many jobs, as well as providing command-line status information about jobs and logfile retrieval.

Timeline: Jobsub already exists and provides the required functionality. TBA
Personpower Required: Fermilab Scientific Computing Division (SCD) FIFE project provides jobsub service and maintains it.

Batch Monitoring and Accounting Services
Descriptions: 1) A web-accessible monitoring tool that allows a user to query the status of waiting, running and completed jobs is required. CPU time used, memory usage, i/o, and job efficiency metrics are to be displayed. 2) Tools are needed to sum over the total wall-clock hours, CPU hours, and data transmission volume and rates on the timescales of days, weeks, months, and years, in order to make reports and identify bottlenecks. These should be broken down by user and by site.
Timeline: For Fermilab and OSG jobs: FIFEMON and Grid Accounting already exists and meets the requirements for Fermilab's computers. For CERN jobs: For other distributed computing jobs:
Personpower Required: None.

Interactive Computing Servers
Description: Fermilab: Ten four-core interactive servers, each with 12 GB of memory and access to the NAS and dCache disk space is sufficient for offline work at Fermilab. Additional nodes are required in order to evaluate upcoming operating system versions or to test the batch environment interactively. CERN: A collection of interactive nodes is required at CERN for software and DAQ development there. These nodes must be administered, maintained, and upgraded as needed.

Timeline: The Fermilab nodes already exist and are maintained by SCD. At CERN CENF provides computers and support for the CERN software effort.

Personpower Required: Fermilab SCD maintains the DUNE interactive servers. DUNE S&C staff spend 2-3 days per year testing new releases and evaluating the system software on these nodes, and submit tickets when components are missing or a problem arises. A system administrator is needed at CERN to take care of the CENF cluster.

Build Nodes
Description: The software build process is CPU intensive, and the build tools in use allow parallelization to many cores. Fermilab: The current limit on the number of cores on an interactive general purpose virtual machine at Fermilab is 4, and so these nodes are not optimal for building code in large repositories. A build machine with at least 16 cores is desirable. The number of cores that can be profitably used by a build machine depends on the I/O capabilities to the disks storing applications. Since the build node should not be used to run programs, even to test built code, then the code must reside on shared disks. CERN: A request is being made to CERN IT for an equivalent machine at CERN.

Timeline: A build node at Fermilab, dunebuild01, already exists, with 16 cores.

Personpower Required: At Fermilab, SCD provides and maintains the build node.

Data Release Management
Description: In accordance with the DUNE Data Management Plan, data must be made publicly available after a time allotted to the Collaboration to extract scientific results. This is a plan for data archiving, management and release that is compliant with the DoE directives on data management. The plan is executed through the data management group.

Timeline: Data release may happen one or more years after the data are collected. Small amounts of data may be shared with colleagues or the public at any time on an *ad hoc* basis.

External IT Components

The following are all ongoing services. They are considered to be generic external resources that are used by DUNE/protoDUNE-SP that are provided outside of any collaboration effort either by CERN IT or Fermilab's Core Computing [Infrastructure] or Scientific Computing divisions:

Service/Help Desk

Description: The Service/Help desks at the collaborating labs provide portals for the creation and tracking of incidents and requests made by users. The labs use the Service Now (SNOW) platform for these interactions. All DUNE/ProtoDUNE users should have access to the appropriate service interfaced based on their laboratory credentials (i.e. access to the CERN SNOW system requires CERN identification, FNAL requires similar credentials to access the system) and should be able to submit tickets to the CERN and Fermilab service desks and receive responses through these systems.

E-mail

Description: Electronic mail service is provide through individual's host labs or universities.

Phone and Video Conferencing

Description: Phone and video conferencing infrastructure is provided by the hosts labs or collaborating universities as appropriate.

Calendar/Meeting Scheduling

Description: Calendar and meeting scheduling is provided by the hosts labs via the Indico platform. Additional scheduling services are provided by the host lab's portals for booking meeting rooms and other associated resources

Collaborative Web Sites/Document Sharing

Description: The DUNE collaboration will use a combination of collaborative web tools for generating and disseminating documentation relating to the experiment. These tools in the forms of WIKIs, SharePoint sites, Wordpress pages/sites, Document Databases and other tools are provided by the host laboratory's computing division or IT departments as appropriate

Summary of Effort needed and Current Availability

Area of Work **DUNE/ pDUNESP**	**Effort Estimate**	**Responsible**	**Effort Available**
Tape and Disk Storage	0	DUNE Data Management Manager	Fermilab SCD & CERN IT
Data Cataloging System	0	DUNE Data Management Manager	Fermilab SCD
File Transfer System: EHN1 to EOS	0	DUNE Data Management Manager	Fermilab SCD & CERN IT
File-Transfer System: EOS to Fermilab dCache	0	DUNE Data Management Manager	Fermilab SCD & CERN IT
Tape storage and Retrieval	0	DUNE Data Management Manager	Fermilab SCD & CERN IT
Reconstruction and Analysis Algorithms	6 FTE	Data Reconstruction and Analysis Leads	?3 FTE?
Hardware Database		DAQ	
Metadata Schema Specification	~2 FTE Weeks/Year	DUNE Database Manager	?

Logfile Persistency Policy and Automated Management	~2 FTE Weeks/Year		Effort from sub-areas - DAQ and Production
Calibration Database	0.1 FTE	DUNE Databases Manager	DRA, Fermilab ND + SCD & CERN IT +
Software Repository	0	DUNE Software and Release Manager	
Documentation Wikis	2 FTE Weeks/Year	DUNE/ProtoDUNE-SP subareas	DAQ, DRA, Prompt Processing, DUNE S&C
DUNE/ProtoDUNE-SP Specific Build Software/System			
Code Distribution System	4-6 FTE Weeks/year	DUNE Software and Release Manager	?
Resource Requests (compute, disk, network) Allocation and Management	4 FTE Weeks/Year	DUNE S&C Coordinators	Interface with Fermilab SCD, CERN IT, Collaboration Resource providers
Offline Disk Caching/Storage	0	DUNE Data Management Manager	Fermilab SCD & CERN IT
Processed Data Metadata Schema	2 FTE Weeks	DUNE Data Management Manager	Fermilab SCD.
Processing Campaign Management Tool	0	DUNE Production Manager	Fermilab SCD
Data Production Processing Till start of data taking Operations Post data taking	 0.3 FTE 2.0 FTE. 0.3 FTE	DUNE Production Manager	0.3 FTE available (Anna) ? FTE additional help
Data Catalog Web Site	0	DUNE Data Management Coordinator	Fermilab SCD contributed effort for operations
Document Database	0	DUNE Collaborative Tools Manager	LBNF/DUNE joint effort
Analysis Group Web Pages	?	DUNE Collaborative Tools Manager	
Public Results Web Pages	1 FTE Week/Year	DUNE Collaborative Tools Manager	LBNF/DUNE joint effort
General Computing Services: Data Release Management	0	DUNE Data Management Manager	Fermilab SCD
General Computing Services: Build Nodes	0	DUNE Release Manager	Fermilab SCD & CERN IT

General Computing Services: Interactive Computing Servers	2-4 FTE Weeks/Year	DUNE S&C Release Manager for testing	Fermilab SCD & CERN IT
General Computing Services: Batch Monitoring and Accounting Services	0	DUNE Production Coordinator Lead(?)	Fermilab SCD & CERN IT.
General Computing Services: Job Management Services	0	DUNE Production Coordinator Lead	Fermilab SCD & CERN IT.
General Computing Services: Data Movement Software (GridFTP and XROOTD)	0	DUNE Data Management Lead	Fermilab SCD & CERN IT.
General Computing Services: Grid CPU Allocation	2 FTE Weeks/Year	DUNE S&C coordinators	
General Computing Services: User Authentication Support	1FTE Week/Year	DUNE S&C coordinators	
External IT components	<1 FTE Week/Year	DUNE S&C coordinators interfacing to SCD and CCD.	Available per Fermilab SCD and CCD