

DUNE Training Infrastructure

David DeMuth, Valley City State University

Database Group
August 22, 2023

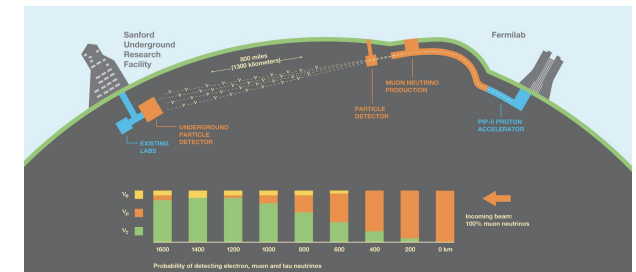


VALLEY CITY
STATE UNIVERSITY



Introduction

- Members of the DUNE Computing Consortium have developed tutorials for the graduate students, post-docs, young and old researchers joining the Collaboration each year.
- The focus has been on the essentials simulating and reconstructing events.
- Organizing training materials for asynchronous access with a common look and feel is a goal.
- Upcoming database training provides a use case to extend the basics training infrastructure to a broader curriculum.




Worldwide and sophisticated, DUNE Computing Tutorials are used regularly to train new colleagues.

DUNE Computing Tutorials


- DUNE Computing has mastered tutorials focused on:
 - data storage and management,
 - introduction to art and LArSoft, and
 - job submission and monitoring.
- A goal is to certify that new colleagues have access to DUNE computing resources, understand the basics of logging in, storage areas, running applications, code mods, and batch jobs.
- Another goal is to jumpstart physics analyses.
- Approximately 400 participants have joined for 10 training events offered during collaboration meetings since 2016.
- Organizers, lecturers, and mentors are few but committed.

Sessions	Wednesday, May 12	Thursday, May 13	Friday, May 14
8:00 - 8:15	Welcome + announcements C. David & D. DeMuth	Grid job submission + common errors Lecture + hands-on + exercises <i>Follow-up: see "Expert in the room" Friday late morning</i> K. Herner	"Expert in the room" LArSoft: How to modify a module T. Junk
8:15 - 9:00	Storage spaces Lecture + hands-on M. Kirby		Code-makeover Switch to POMS K. Herner
9:00 - 10:00	Data management Lecture + hands-on S. Timm	Coffee break!	
10:00 - 10:30	Coffee break!	Coffee break!	Coffee break!
10:30 - 11:00	QUIZ! Storage spaces data management	QUIZ! Grid job submission	QUIZ! Best programming practices
11:00 - 12:15	Intro to art/LArSoft ← lecture Exploring fcl files ← hands-on <i>Follow-up: see Friday morning</i> T. Junk	Code-makeover How to improve your code for better efficiency T. Junk	"Expert in the room" Grid & batch job submission K. Herner
12:15 - 12:30			Closing remarks C. David & D. DeMuth




Organizers

Claire David
York University / FNAL




David DeMuth
Valley City State University




DUNE Computing Consortium Lead

Heidi Schellman
Oregon State University


Lecturers



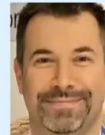
Mike Kirby
FNAL



Steven Timm
FNAL




Tom Junk
FNAL




Kenneth Herner
FNAL

Mentors



Amit Bashyal
ANL



Carlos Sarasty
U. of Cincinnati

The [May 2021 training](#) was offered as a three day event, each of four lecturers and two mentors doing the bulk of the work. The most recent training took place in [May 2023](#).

Training Logistics

- Event registration and communications use Indico.
- Participants verify ability to use the Unix Shell as homework.
- Hands-on activities include livecoding, quizzes, expert in the room sessions.
- Sessions are hybrid delivered, captured via Zoom for asynchronous review.
- Mentors are essential to ensure skill development and understanding.

DUNE Computing Training May 2021 edition: Mission Setup

Objectives

- Get ready to do the tutorial
- Understand the authentication procedures
- Set up your environment for DUNE
- Do an exercise to help us check if all is good
- Get streaming and grid access

Requirements

You must be on the DUNE Collaboration member list and have a valid FNAL or CERN account. See the [Indico Requirement page](#) for more information.

Note

The instructions below are for FNAL accounts. If you do not have a valid FNAL account but a CERN one, go at the bottom of this page to the "Setup on CERN machines".

1. Kerberos business

If you already are a kerberos-aficionado, go to the next section. If not, we give you a little tour of it below.

What is it? Kerberos is a computer-network authentication protocol that works on the basis of tickets.

Why does FNAL use Kerberos? Fermilab uses Kerberos to implement strong authentication, so that no passwords go over the internet (if a hacker steals a ticket, it is only valid for a day).

How it works? Kerberos uses tickets to authenticate users. Tickets are made by the kinit command, which asks for your kerberos password (info on kerberos password [here](#)). The kinit command reads the password, encrypts it and sends it to the Key Distribution Centre (KDC) at FNAL. The Kerberos configuration file, which lists the KDCs, is stored in a file named krb5.conf. On Linux and Mac, it is located here:

```
Code
/etc/krb5.conf
```

If you do not have it, create it. A FNAL template is available [here](#) for each OS (Linux, Mac, Windows). More explanations on this config file are available [here](#) if you're curious.

To log in to a machine, you need to have a valid kerberos ticket. You don't need to do this every time you login, only when your ticket is expired. Kerberos tickets last for 26 hours. To create your ticket:

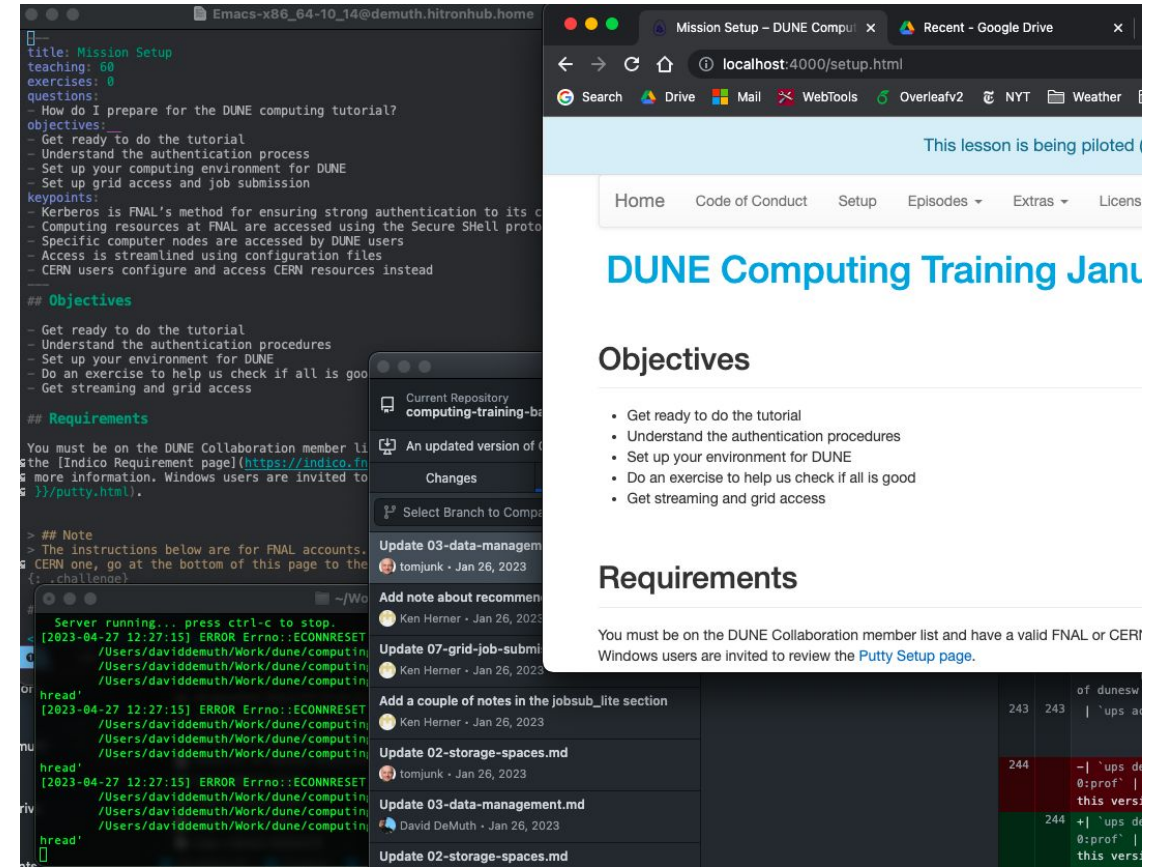
```
Bash
kinit -f username@FNAL.GOV
```

In advance of the opening salvo, students must demonstrate an understanding of using the Unix shell to access secure VMs.

Lesson Development

The infrastructure to develop lessons is provided in the Software Carpentries (SWC) framework:

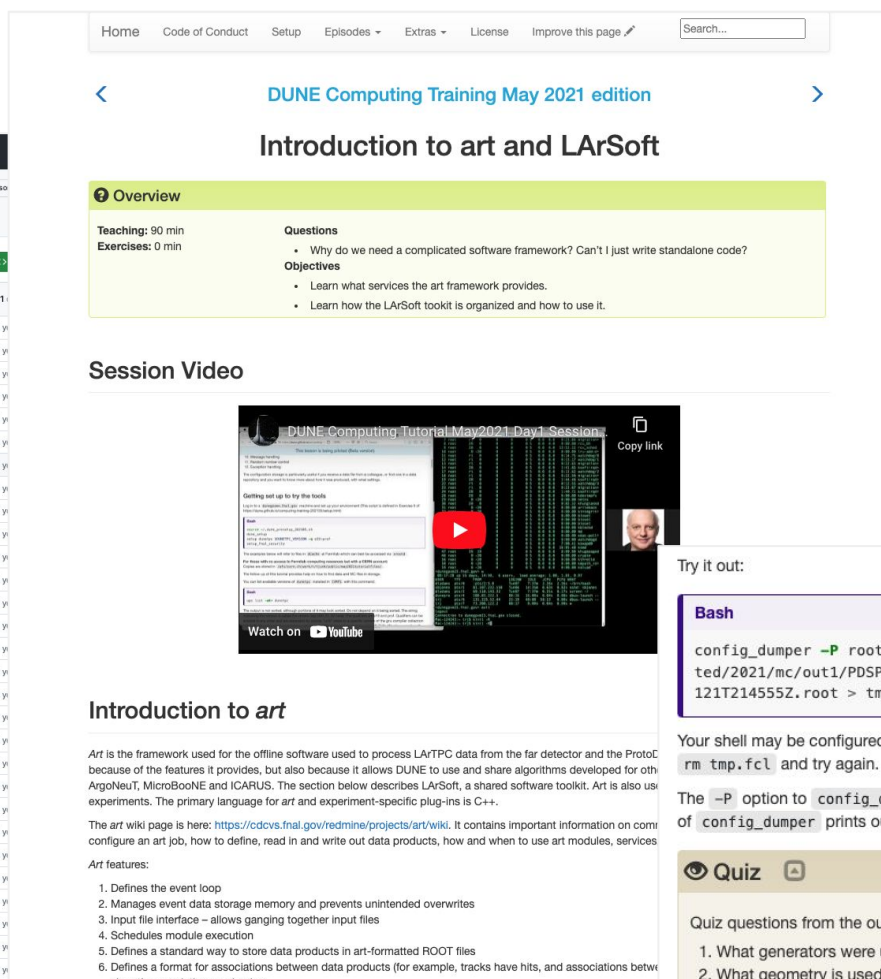
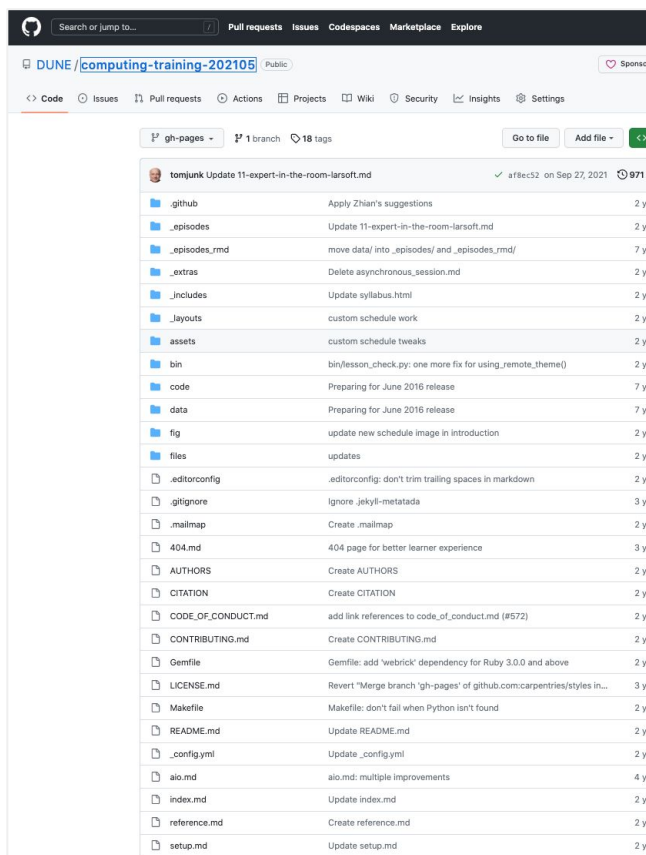
- The SWC [lesson template](#) is imported as a new DUNE GitHub repo, configurable, and main lesson content as hosted as easily edited markdown files.
- GitHub Desktop is used to manage the repository locally.
- Viewing edits in a localhost browser uses a [Ruby/Jekyll](#).
- Lessons are rendered on the web via [GitHub Pages](#) as a free service.



Once installed on a curriculum designers local machine, the lesson production environment is slick.

Lesson Deployment

End users experience:



Markdown language features:

- Simplified editing structure for web browsers,
- Bash script blocks easily copied and pasted in terminal windows,
- Drop down Quiz blocks that can be opened on demand.

Try it out:

Bash

```
config_dumper -P root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/tape_backed/dunepro/protodune-sp/full-reconstructed/2021/mc/out1/PDSPProd4/40/57/23/91/PDSPProd4_protoDUNE_sp_reco_stage1_p1GeV_35ms_sce_datadriven_41094796_0_20210121T214555Z.root > tmp.fcl
```

Your shell may be configured with `noclobber`, meaning that if you already have a file called `tmp.fcl`, the shell will refuse to overwrite it. Just `rm tmp.fcl` and try again.

The `-P` option to `config_dumper` is needed to tell `config_dumper` to print out all processing configuration `fcl` parameters. The default behavior of `config_dumper` prints out only a subset of the configuration parameters, and is most notably missing art services configuration.

Quiz

Quiz questions from the output of the above run of `config_dumper`:

1. What generators were used? What physics processes are simulated in this file?
2. What geometry is used? (hint: look for "GDML" or "gdml")
3. What electron lifetime was assumed?
4. What is the readout window size?

GitHub's gh-pages are rendered seamlessly, and for free.

DUNE Computing Support

- As expert instructors work through lessons, live coding in an adjacent window is helpful, particularly for complex commands.
- Livedocs are monitored by mentors and other experts provide a support interface that afterwards become a permanent educational resource.
- Slack channels are set up for each training event, where questions can be posed asynchronously, questions receive attention from a cadre of experts, ideally triggering relationships.
- GitHub Issues (FAQ) is another method to field questions, and ensure the novice become proficient in DUNE computing.
- Fermilab's ServiceNow provides user support.
- Documentation, training, and support are interconnected activities.

DUNE Computing Training December 2021

Storage Spaces and Data Management

Live document

1. Pick a color!
2. Write your name (if you want)
3. Ask your question, an expert will reply

Example:
Anna:
Does the `ifdh` tool work for all storage types?
Kirby:
Yes; including several obsolete ones that no longer exist.

Heidi: Storage at Fermilab - how does it work for offsite? Can you use `ifdh` to get a file to or is it just grid jobs?
Norm: Is touching a file considering using it?
Kirby: no.

What would happen if you tried to `ifdh cp` that file and it was not cached?
Kirby: no error nor success but put you in a wait state. If busy, you might wait 3 or 4 days. If it retries it extends the timeout to a longer period.

Norm: would you get feedback from the grid job that would help understand what was wrong?
Kirby: need to dig into the logfile - How do to that is in second session

What is the correct way to call a file that is not cached?
Kirby: [will write later]

Heidi: Does `xrdcopy` act similarly to `ifdh` or are there evil consequences I don't know about from using it?
Kirby: `xrdcp` is different from `ifdh` (layer on top, utilizes some of `xrd` tools). If you know you should be using `xrd`, if you are unsure: you better use `ifdh`.

Do we use `pnfs2xrootd` to stream a file? How is it different from `Samweb2xrootd`?
Kirby: [will write later]

Does `ifdh` `locateFile` use SAM?
Kirby: [will write later]

Will it work when we switch to Metacat?
Kirby: [will write later]

It seems there is some overlap between SAM and Rucio functionalities. Will they co-exist in the future?
Kirby: switching to Rucio → serving a different purpose than SAM. Not much overlap between the two. SAM does track location, Rucio as well.
If one storage element goes offline, Rucio will check and replicate over to another location. Tracking OK.

Livedocs, Slack, and GitHub Issues are used to assist DUNE colleagues with computing questions.

The image shows two overlapping screenshots. The top one is a Slack channel titled '# computing-training-basics' with a dark theme. It shows a conversation where a user named Jordi Capó asks a question about Kerberos tickets on a Mac, and another user, Steven Timm, responds. The bottom screenshot is a GitHub Issues page for the repository 'E / FAQ'. It shows a list of issues with titles like 'Why did my grid job fail?', 'Token Related Error Messages', and 'How to back up your user files to tape'. The page includes search filters, labels, and a 'New Issue' button.

DUNE DB Training Infrastructure

- Build GitHub instance of training arranged as SWC episodes:
- Each episode is “owned” by an instructor who takes responsibility for updating the lesson to match their instruction.
- Episodes are markdown files created from teaching resources that might be extracted by hand from LaTeX and slides.
- Instructional materials become a step by step learning map for both the instructor and learner.
- Video is captured during the training, and afterwards embedded in the episodes for asynchronous study/review.
- A first training might be more easily delivered traditionally from slides, then formulated into markdown for use in subsequent training events.

PDS (HD) Hardware Database Schema and API Description

Norm Buchanan
August 25, 2021

1 Introduction

This document describes the hardware database schema for the XArapuca modules that will populate the ProtoDUNE detector and the DUNE vertical drift far detector. The database schema is used to define the various tables and relationships between them in the DUNE hardware database. The overall schema layout will be described, followed by detailed descriptions of each table. Lastly, the Application Programming Interface (API) used to collate the various table contents, or data, and insert them into the central database will be described.

2 Schema Overview

The overview of schema tables is shown in figure 1. Lines between boxes, or tables, denote relationships between the data stored in them.

Figure 1: Overview diagram of DUNE PDS (HD).

The PDS schema will adhere to the HWDB requirements as outlined in reference [1]. The critical requirement is that each component stored in the hardware database must be mapped to a unique identifying Part identifier.

The unique PartsID will be generated and entered into the PartsID database prior to the component being entered into the HWDB. The part numbers will be designated as shown in Table 1. The unique identifiers (Parts ID) will be transferred to the hardware database automatically once they have been entered into the PartsID database. Once the PartID has been transferred into the hardware database the part record can be completed. The PartID will be the unique identifier, and key, associated with each component in the hardware database.

¹Once the PartIDs have been defined they can be entered in a batch operation into the PartsID database

1

DUNE Computing Documentation

- Newly on-boarded and longstanding researchers access DUNE computing resources which are inventoried as a protected one-stop MediaWorks wiki with main blocks:
 - Organization & Partners,
 - Computing Toolbox,
 - Operations & Monitoring,
 - Working Groups,
 - Resources,
 - Getting Started.
- Redmine and GitHub are used for code documentation.
 - Sphinx & Read-the-Docs are systems being explored.
- For unrecognized topics: the ABC DUNE glossary.
- Self-studies of documentation is an important training step.

[CHEP 2023 All Things Computing, Poster 130](#)

<https://abc.dunescience.org/>