DUNE Training Infrastructure

David DeMuth, Valley City State University

Database Group August 22, 2023





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 "Expert in the room	
8:15 - 9:00	Storage spaces Lecture + hands-on М. Кirby	+ common errors Lecture + hands-on + exercises Follow-up: see	LArSoft: How to modify a module T. Junk
9:00 - 10:00	Data management Lecture + hands-on S. Timm	"Expert in the room" Friday late morning K. Herner	Code-makeover Switch to POMS K. Herner
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 Follow-up: see Friday morning	Code-makeover How to improve your code for better efficiency	"Expert in the room" Grid & batch job submission K. Herner
12:15 - 12:30	T. Junk	T. Junk	Closing remarks C. David & D. DeMuth



The <u>May 2021 training</u> 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 <u>May 2023</u>.



Training Logistics

- Event registration and communications use Indico.
- Participants verify ability to use the <u>Unix Shell</u> 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
- 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, nerrypts 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.comf. On Linux and Mac, it is located here:

Code /etc/krb5.con1

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 <u>lesson template</u> 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 <u>Ruby/Jekyll</u>.
- Lessons are rendered on the web via <u>GitHub Pages</u> as a free service.



David DeMuth - Jan 26, 2023
Update 02-storage-spaces.md

/Users/daviddemuth/Work/dune/computing

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



this ver

Lesson Deployment

End users experience:

Search or jump to	p 7 Pull requests issues	Codespaces Marketplace Explore	
DUNE / comput	ing-training-202105 (Public)	Ø	Sponso
Code 💿 Issues	13 Pull requests 💿 Actions 🖽 Project	s 🖽 Wiki 🛈 Security 🗠 Insights 🕸 Settings	
	₽ gh-pages - ₽ 1 branch © 18 ta	gs Go to file Add file ~	
	tomjunk Update 11-expert-in-the-room	n-larsoft.md 🗸 af8ec52 on Sep 27, 2021	③ 971
	github	Apply Zhian's suggestions	2 y
	_episodes	Update 11-expert-in-the-room-larsoft.md	2 y
	episodes_rmd	move data/ into _episodes/ and _episodes_rmd/	7 y
	extras	Delete asynchronous_session.md	2 y
	includes	Update syllabus.html	2 y
	ayouts	custom schedule work	2 y
	assets	custom schedule tweaks	2 y
	bin	bin/lesson_check.py: one more fix for using_remote_theme()	2 y
	i code	Preparing for June 2016 release	7 y
	ata data	Preparing for June 2016 release	7 y
	🖿 fig	update new schedule image in introduction	2 y
	iles files	updates	2 y
	.editorconfig	editorconfig: don't trim trailing spaces in markdown.	2 y
	🗅 .gitignore	Ignore .jekyll-metatada	3 у
	🗋 .mailmap	Create .mailmap	2 y
	🗅 404.md	404 page for better learner experience	3 y
	AUTHORS	Create AUTHORS	2 y
		Create CITATION	2 y
	CODE_OF_CONDUCT.md	add link references to code_of_conduct.md (#572)	2 y
	CONTRIBUTING.md	Create CONTRIBUTING.md	2 y
	🗋 Gemfile	Gemfile: add 'webrick' dependency for Ruby 3.0.0 and above	2 y
	LICENSE.md	Revert "Merge branch 'gh-pages' of github.com:carpentries/styles in	3 у
	D Makefile	Makefile: don't fail when Python isn't found	2 y
	README.md	Update README.md	2 y
	🗋 _config.yml	Update _config.yml	2 y
	🗋 aio.md	aio.md: multiple improvements	4 y
	index.md	Update index.md	2 y
	C reference.md	Create reference.md	2 y
	🗅 setup.md	Update setup.md	2 v
	Contraction and Contraction		

Home Code of Conduct Setup Episodes - Extras - License Improve this page Search. < **DUNE Computing Training May 2021 edition** Introduction to art and LArSoft

Overview

Teaching: 90 min Exercises: 0 min	Questions Why do we need a complicated software framework? Can't I just write standalone code? Objectives		
	 Learn what services the art framework provides. 		
	Learn how the LArSoft tookit is organized and how to use it.		

Session Video



Introduction to art

Art is the framework used for the offline software used to process LArTPC data from the far detector and the ProtoC because of the features it provides, but also because it allows DUNE to use and share algorithms developed for oth ArgoNeuT. MicroBooNE and ICARUS. The section below describes LArSoft, a shared software toolkit, Art is also us experiments. The primary language for art and experiment-specific plug-ins is C++.

The art wiki page is here: https://cdcvs.fnal.gov/redmine/projects/art/wiki. It contains important information on com configure an art job, how to define, read in and write out data products, how and when to use art modules, services Art features

1. Defines the event loop

- 2. Manages event data storage memory and prevents unintended overwrites 3. Input file interface - allows ganging together input files
- 4. Schedules module execution
- 5. Defines a standard way to store data products in art-formatted ROOT files

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

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.

Bash

config_dumper -P root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/tape_backed/dunepro/protodune-sp/full-reconstruc ted/2021/mc/out1/PDSPProd4/40/57/23/91/PDSPProd4_protoDUNE_sp_reco_stage1_p1GeV_35ms_sce_datadriven_41094796_0_20210 121T214555Z.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.

Ouiz ⊡

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?



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.



Kirby: no.

O X11 connection rejected because of wrong authentication error me



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.



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 $\fbox{1}$. 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 $[\bar{1}]$. 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 II. The unique identifiers (Parts ID) will be transferred to the hardware database automatically once they have been entered into the PartsID databased. 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.

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



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.

Texture of the P II and Textur	All Thing Claire David and Dav			
Abstract	ORGANIZATION & PARTNERS	OPERATIONS & MONITORING	GETTING STARTED	Redmine
ADJUSTICAL Decomposition of a will be an envolved collaboration of sciencits, in the science and science to apply collaboration of sciencits, in the science and science to apply collaboration of a science and science and science and science science and according to a science and sci	About DUNE Compliant Satisfament 2. Constant, & Sapasizational chart Meetinga 2. Computing wateky meetings 3. Computing midd coategory) 3. Compaining and Software workstops Mesciated colloborations 4. Failot for Frontier Experiments (FFE) 5. Emergy Sonce and extern ESint 1. High Energy Physics Software Fundation (HSF) 1. Urdoff Collaborations 3. Workforded LHC Computing Grid (MLCG)	Online Data Taking PhotoDURH Frendman 2014 PhotoDURH Frendman 2014 PhotoDURH Service and Print Service Service 2014 PhotoDure PhotoDure 2014 PhotoDure 2014	New to DUNE? + Hork to jan ONE + Hork to gait a Fermibia account Computing Basics Tutorials - The basics: Unix Shell, Python, not + 18/87 Training Center: Giblub, Docker, CUCD DUNE-Specific Tutorials - Geting Startiet Tutorial - Geting Startiet Tutorial - Geting Startiet Tutorial - Geting United - Decision Tutorial - Ruscio Tutorial	Interacting Andreim publicly host DMC IT and Practicular Hosting with any allow brains and summarized to consult and consultation of the second second second second second second second second second second second discontration of the algorithm and official second second discontration of the algorithm and second second second discontration of the algorithm and the algorithm and discontration and second second second second second and the algorithm and the al
	COMPUTING TOOLBOX Data • Data access: disks, dCache • Data management tods: rucio, SAM, lidho Code • Ode access: cvmS, GitHub, docker images • Detector software: protoUNIE, ND, FD Frameworks & Toolbits	WORKING GROUPS Permanent Working Groups software management Database Data Management Vorkload and Workflow management Monitoring and Information system management	Help Desk UNK Slack channel UNK Slack channel UNK Computing Service Request UNK Computing FAQ (GitHub) RESOURCES Administration of information Information protection Unta preservation	community of users, theogo DUM does not yet use (Shuhi's wike notenively: that and build adgrad Software (McKoff) teams have transitioned from the Relmine system to using distribu's tickeling yetsen. A peak is for doe to be publicly readable; but protected from nonflication. Code Standard: for LA/Soft are well evelopeet, while DUM-specific coder gundards are in self-evelopeet.
Documentation	CAFAna CArSoft	Networking Training and Documentation	Computing documents	User Support
The documentation related to the DUNE's computing apperts will be accessible on a variety of platform, each with specific goals and access policies. As a rule, access policies. As a rule, access and requires sight sign-on ISS0 access to materials hosted at both Fermibia and CEN. A challenge has been in maritaining consistency across non-public and public documentation.	UARDOTT Instruction UARDOTT Instruction UARDOTT Instruction Instruction UARDOTT Instruction UARDOTT UARDO	Temporary Taskforce Data Challenge Framework Task Force Calibration Task Force DUNE Liaisons	Unconcentry of Winkith and Computing Documents Conference talks List of talks and proceedings List of talks	For issues associated to DUNE Computing, CBHub issues is sering as an IAD (3), where separts regularly monitor an provide feedback. Commercial grade Static provides a patient for all DUNE collaborators to communicate sugget us spattions, with declarade DUNE experiment monitoring activity ServiceNow (6) is the official Fermida issue reporting system Problems, including operational issues that can not be resolved with peer support through Stack (7) are directed to the ServiceNow Services, and reso DUNE Service Siles.
Wikis	With a validated login, DUNE users are empowered to use, peruse, search, and study relevant content.	Hardware Groups Physics Groups	DUNE Computing Wiki "Where To Put	References
The DUNE wiki provides a starting point for DUNE Computing. Access is limited to those with SSD credentials as sensitive information such as node names are prevalent.	Selected DUNE Computing editors have been selected to take on an enhanced role to develop content, often from pre-Wiki documentation streams.	ProtoDUNE Far Detector Near Detector	The"	C. David, CHEP 2023, Poster 130 [2] D. DaMuth, CHEP 2023, Talk 625 [3] DUNE Offline Computing Conceptual Design Report, Mann (Kenner (der D310) 1999)
Six main blocks form the top level (entry) of the DUNE computing wilk: Organization & Partners, Computing Toolbox, Operations & Monitoring, Working Groups, Resources, and Getting Started. These are Illustrated in the middle part of this poster, much of which is self explanatory.	Abouts, calendar meetings, and collaborations make up th collects links to data access, analysis tools, and algorithm accordingly. Conveners of Working Groups manage their ass colleagues with how-too, and the Resources block collate access. Managing a well designed information portails an im More on DUNE Computing [3]. Contacts: claire.david@rem.cc	e Organization block. The Computing Toolbox frameworks. Operations activities are linked octated space. Getting started helps new DUNE is information such as policies and document portnat strategy for DUNE. h, david.demuth@vcsu.edu		https://annx.ofg/abs/2210.15665 [4] Sphine Hilps://www.sphine.cdoc.org/en/master/ [5] FAQ. https://github.com/DUNE/FAQ/issues [6] ServiceNow. https://anvicedesk.fnal.gov [7] Stack https://dunescience.akk.com [8] DUNE science: www.dunescience.org

CHEP 2023 All Things Computing. Poster 130



