DUNE Software Training

Heidi Schellman, Oregon State University
David DeMuth, Valley City State University

October 19, 2024

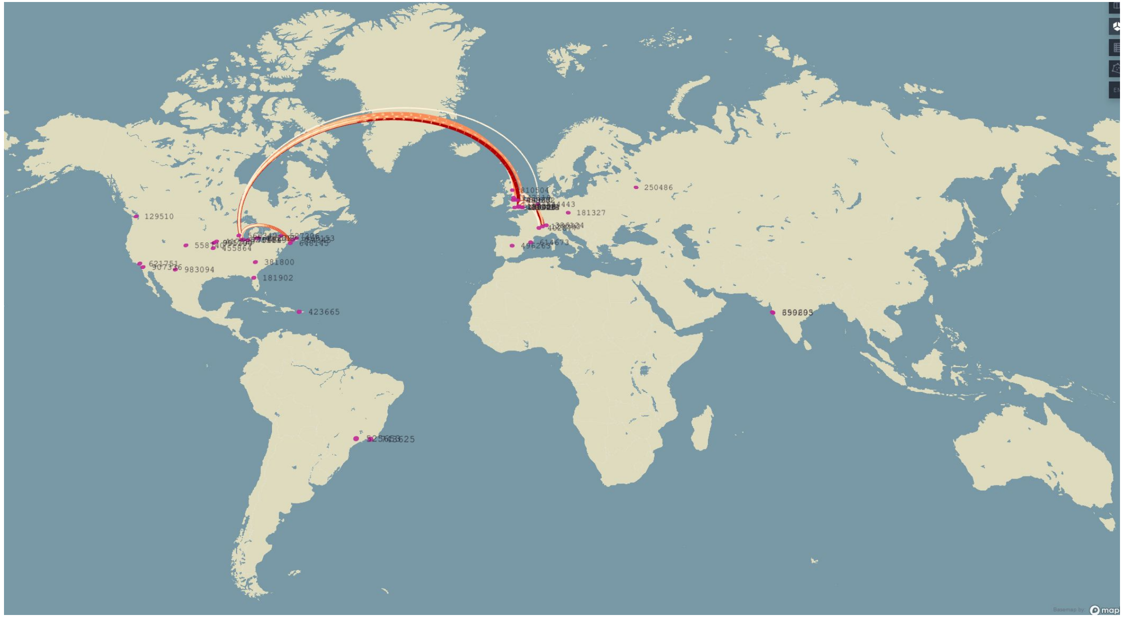
Outline

- Description of our basic training and development of a second module for the hardware database. Credit to Dave Demuth for these slides
- Some feedback from our user base
- Note: Neutrino physicists tend to come from smaller experiments with less formal structure.

DUNE

~1500 collaborators
~500 active FNAL accounts
~1281 in Slack
~37 countries
> 200 institutions
> 40 compute sites
~ 10 data sites

~ 10 PB/year → 30 PB/year



DUNE computing model

- Based on services not tiers
  - ~40 CPU sites, OSG/WLCG >= 2000 MB memory/core
  - HPC
  - ~10 Rucio controlled disk sites: FNAL, CERN, UK sites, NL … (need good network speeds)
  - Local caches
  - Tape store at FNAL, CERN, UK sites, IN2P3
  - Analysis facilities under construction
- CPU anywhere is useful for simulation, reconstruction over xrootd
- Our model tries to keep useful reconstructed samples on disk so dedicated disk resources near CPU are exceptionally valuable for fast analysis.

Communication channels

DUNE wiki (collab only)

DUNE docdb (collab only)

Slack (collab only)

Github pages (public)

- https://dune.github.io/computing-basics
- https://github.com/orgs/DUNE/projects/19

Having docs behind a firewall makes early training difficult

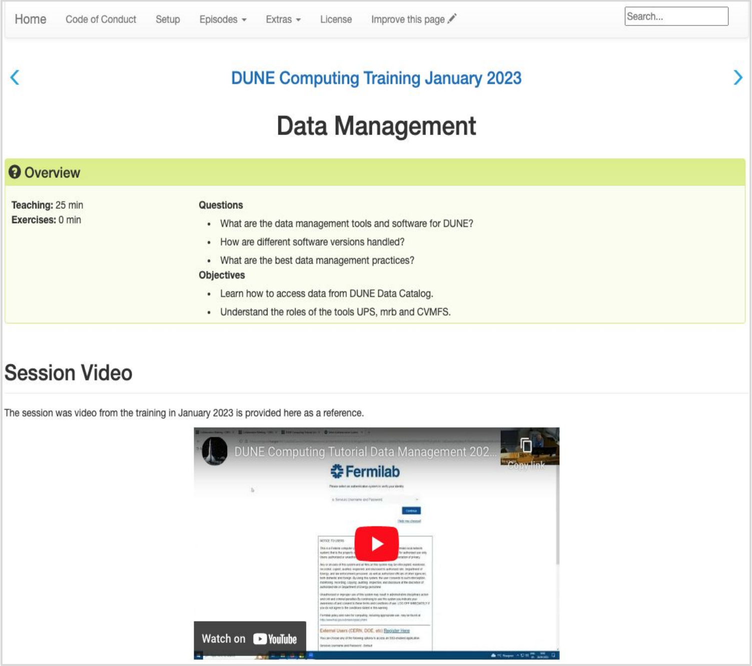
We do have some issues getting google to index the public information

Documentation DUNE Wiki

DUNE Computing hosts a one-stop for computing tutorials in several formats (pdf, wiki, html)

https://wiki.dunescience.org/wiki/Computing_tutorials

- Computing Basics
- FAQ
- The Justin Workflow System
- LArSoft training materials
- General HEP Training Courses
- Metacat documentation
- DUNE Hardware Database Training
- 2x2 and Minerva reconstruction tutorial



DUNE Computing Training January 2023

Data Management

Overview

Teaching: 25 min
Exercises: 0 min

Questions

- What are the data management tools and software for DUNE?
- How are different software versions handled?
- What are the best data management practices?

Objectives

- Learn how to access data from DUNE Data Catalog.
- Understand the roles of the tools UPS, mrb and CVMFS.

Session Video

The session was video from the training in January 2023 is provided here as a reference.

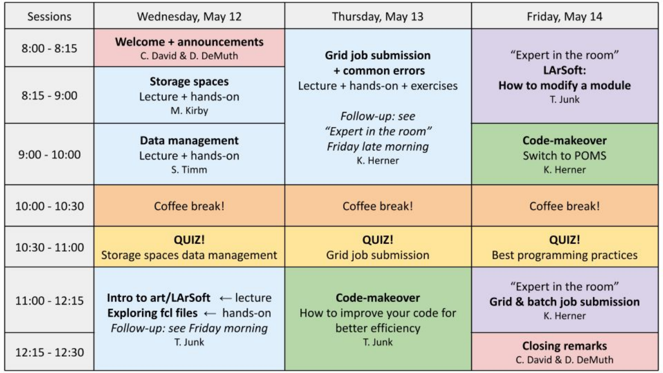
DUNE Computing Tutorial Data Management 202...

An overview of the Computing Basics training and the machinery to produce the lessons follows.

DUNE Computing Tutorials

- DUNE Computing has coordinated 11 training events since 2016, with 400+ participants.
- Tutorials focused on three topics:
  - Data storage and management,
  - art & LArSoft,
  - Job submission and monitoring.
- The goal is to verify access to DUNE resources, understand the basics of logging in, storage areas, running applications, code modifications, submitting and monitoring batch jobs.
- Simultaneous face2face and Zoom modes, and post-production to allow asynchronous study.

https://dune.github.io/computing-basics/



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 *Follow-up: see* *"Expert in the room"* *Friday late morning* 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		
9:00 - 10:00	**Data management** Lecture + hands-on S. Timm		**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* 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

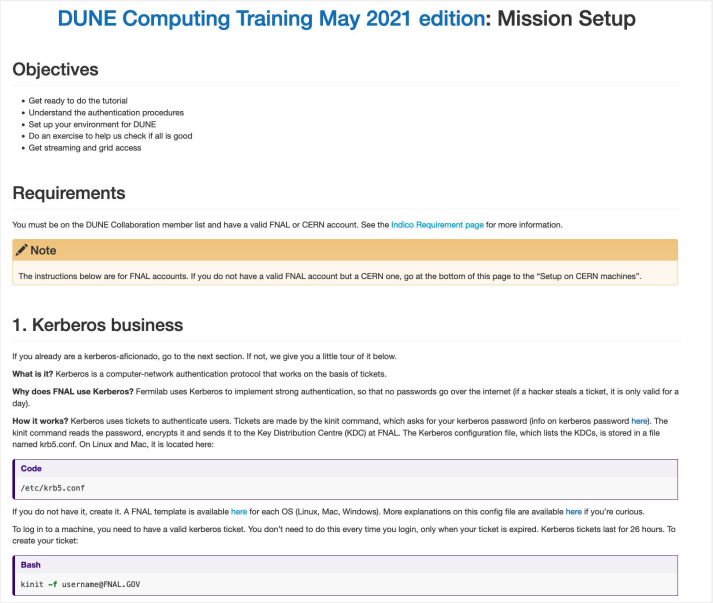
Amit Bashyal
ANL

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. Other training events were two half day, and one day events.

Training Logistics

- Event registration and communications are managed by Indico.
- Participants must verify their ability to use the Unix Shell.
- And verify access the DUNE general purpose virtual machines at FNAL or CERN
- Much of the tutorial can be done using CERN resources.
- Livecoding, quizzes, expert in the room sessions, and assigned mentors ensure the event as hands-on.
- Each session is delivered and captured via Zoom, then embedded into the Software Carpentries lesson framework (SWC) which is hosted at DUNE Computing's GitHub site for review.



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

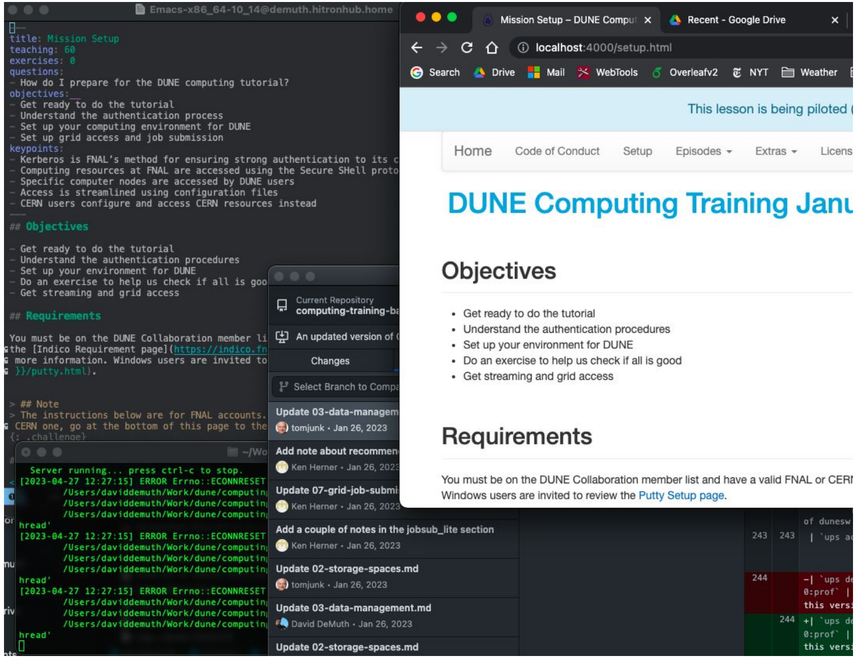
```
kinit -f username@FNAL.GOV
```

In advance, students 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** framework:

- A lesson template is imported as a new DUNE GitHub repo, configuration via a _config.yml file, and main lesson content as **markdown** files (.md) located in _episodes/.
- GitHub Desktop is used to manage the repository locally.
- Viewing edits in a localhost browser uses a Ruby/Jekyll rendering engine.
- Lessons are pushed to the site's main branch for access by multiple authors, and the public.
- Lessons are rendered elegantly on the web via GitHub Pages as a free service.



```
---
title: Mission Setup
teaching: 60
exercises: 0
questions:
- How do I prepare for the DUNE computing tutorial?
objectives:
- Get ready to do the tutorial
- Understand the authentication process
- Set up your computing environment for DUNE
- Set up grid access and job submission
keypoints:
- Kerberos is FNAL’s method for ensuring strong authentication to its c
- Computing resources at FNAL are accessed using the Secure SHell proto
- Specific computer nodes are accessed by DUNE users
- Access is streamlined using configuration files
- CERN users configure and access CERN resources instead
---
## 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 goo
- Get streaming and grid access

## Requirements

You must be on the DUNE Collaboration member li
the [Indico Requirement page](https://indico.fn
 more information. Windows users are invited to
 }}/putty.html).


> ## Note
> The instructions below are for FNAL accounts.
 CERN one, go at the bottom of this page to the
{: .challenge}
```

```
  Server running... press ctrl-c to stop.
[2023-04-27 12:27:15] ERROR Errno::ECONNRESET
        /Users/daviddemuth/Work/dune/computin
        /Users/daviddemuth/Work/dune/computin
        /Users/daviddemuth/Work/dune/computin
hread'
[2023-04-27 12:27:15] ERROR Errno::ECONNRESET
        /Users/daviddemuth/Work/dune/computin
        /Users/daviddemuth/Work/dune/computin
        /Users/daviddemuth/Work/dune/computin
hread'
[2023-04-27 12:27:15] ERROR Errno::ECONNRESET
        /Users/daviddemuth/Work/dune/computin
        /Users/daviddemuth/Work/dune/computin
        /Users/daviddemuth/Work/dune/computin
hread'
```

Update 03-data-managem
tomjunk • Jan 26, 2023

Add note about recommen
Ken Herner • Jan 26, 2023

Update 07-grid-job-submi
Ken Herner • Jan 26, 2023

Add a couple of notes in the jobsub_lite section
Ken Herner • Jan 26, 2023

Update 02-storage-spaces.md
tomjunk • Jan 26, 2023

Update 03-data-management.md
David DeMuth • Jan 26, 2023

Update 02-storage-spaces.md

```
         of dunesw
243  243  | `ups ac
244      -| `ups de
          0:prof` |
          this versi
     244 +| `ups de
          0:prof` |
          this versi
```

localhost:4000/setup.html

This lesson is being piloted (

DUNE Computing Training Janu

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

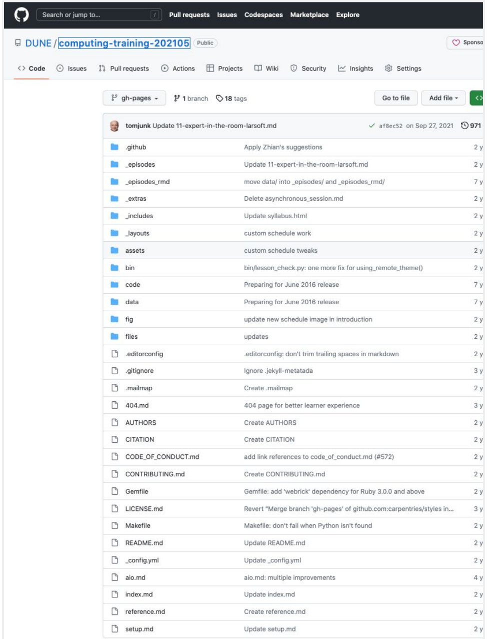
Requirements

You must be on the DUNE Collaboration member list and have a valid FNAL or CERI
Windows users are invited to review the Putty Setup page.

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

Lesson Deployment

End users see:

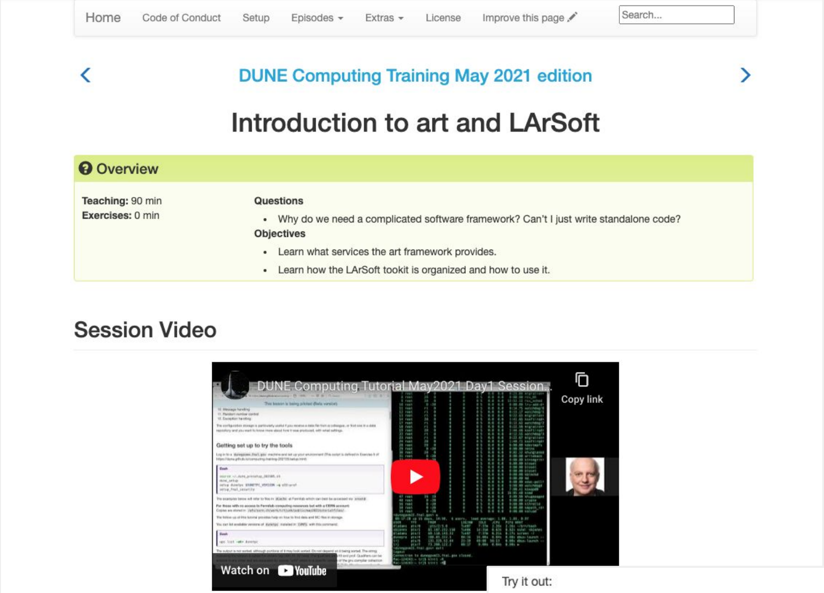


DUNE / computing-training-202105

gh-pages · 1 branch · 18 tags

tomjunk Update 11-expert-in-the-room-larsoft.md ✓ af8ec52 on Sep 27, 2021 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
_layouts	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
code	Preparing for June 2016 release	7 y
data	Preparing for June 2016 release	7 y
fig	update new schedule image in introduction	2 y
files	updates	2 y
.editorconfig	.editorconfig: don't trim trailing spaces in markdown	2 y
.gitignore	Ignore .jekyll-metatada	3 y
.mailmap	Create .mailmap	2 y
404.md	404 page for better learner experience	3 y
AUTHORS	Create AUTHORS	2 y
CITATION	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 y
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
reference.md	Create reference.md	2 y
setup.md	Update setup.md	2 y



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

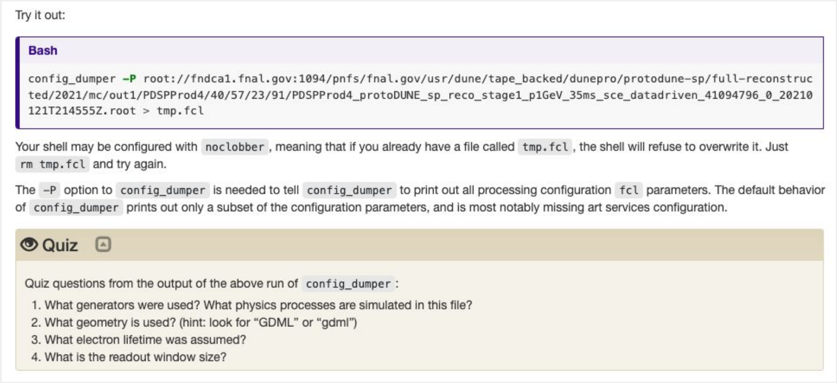
Introduction to *art*

Art is the framework used for the offline software used to process LArTPC data from the far because of the features it provides, but also because it allows DUNE to use and share algo ArgoNeuT, MicroBooNE and ICARUS. The section below describes LArSoft, a shared softw 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 impor configure an art job, how to define, read in and write out data products, how and when to u

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
6. Defines a format for associations between data products (for example, tracks have hits via art's association mechanism



Try it out:

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.

Quiz

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

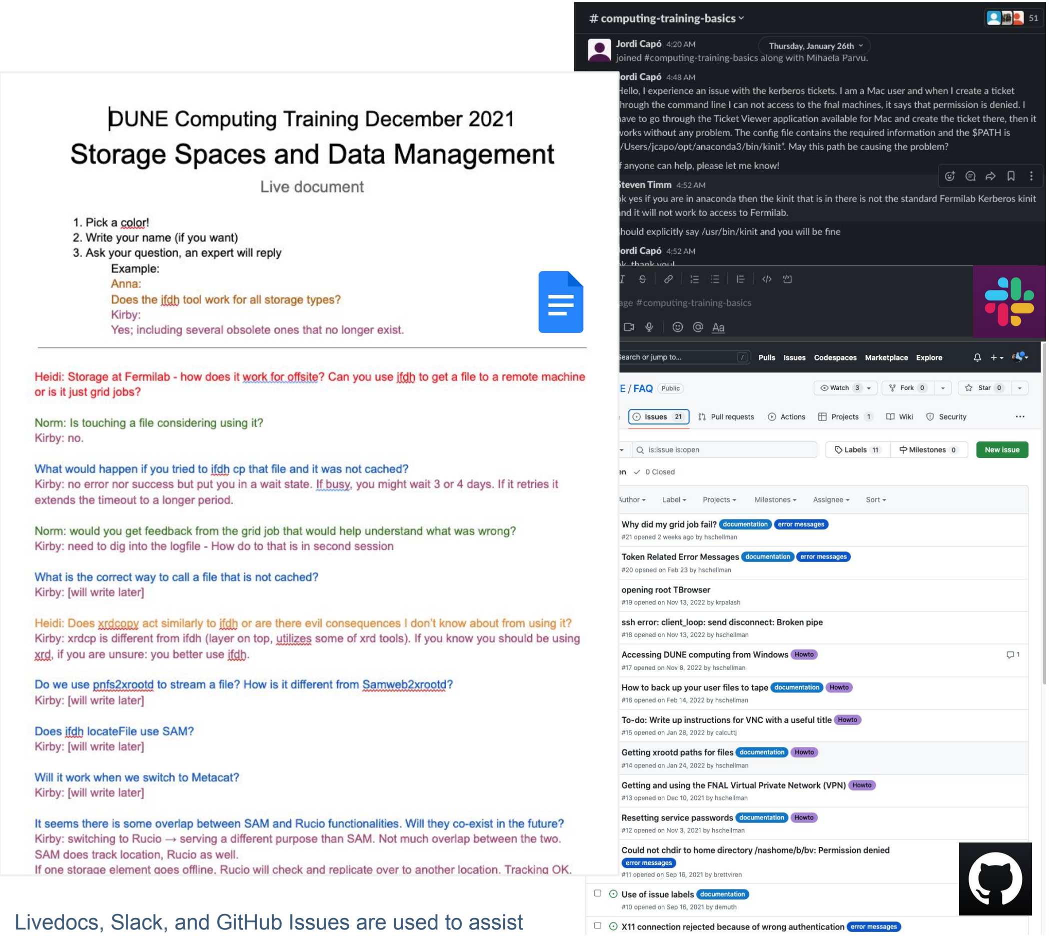
1. What generators were used? What physics processes are simulated in this file?
2. What geometry is used? (hint: look for "GDML" or "gdml")
3. What electron lifetime was assumed?
4. What is the readout window size?

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

Support

As expert instructors work through a lesson plan, often live coding in an adjacent window, multiple methods are used to ensure near- and long-term support:

- Instructors encourage participants to pose questions via a shared and world editable Google Doc (**Livedoc**) which is monitored by mentors and other experts.
  - Each participant selects a unique color.
  - As questions are volleyed, experts interact with individuals to find solutions.
  - Livedocs can be studied by all, solutions can be contrasted, and becomes a resource.
- **Slack channels** are set up for each training event.
- **GitHub Issues** is also used to support learning.



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 a remote machine 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.

computing-training-basics

Thursday, January 26th

Jordi Capó 4:20 AM
joined #computing-training-basics along with Mihaela Parvu.

ordi Capó 4:48 AM
Hello, I experience an issue with the kerberos tickets. I am a Mac user and when I create a ticket hrough the command line I can not access to the fnal machines, it says that permission is denied. I have to go through the Ticket Viewer application available for Mac and create the ticket there, then it works without any problem. The config file contains the required information and the $PATH is /Users/jcapo/opt/anaconda3/bin/kinit". May this path be causing the problem?

f anyone can help, please let me know!

Steven Timm 4:52 AM
ok yes if you are in anaconda then the kinit that is in there is not the standard Fermilab Kerberos kinit and it will not work to access to Fermilab.

should explicitly say /usr/bin/kinit and you will be fine

ordi Capó 4:52 AM
ok, thank you!

age #computing-training-basics

E / FAQ Public

is:issue is:open

0 Closed

Why did my grid job fail? documentation error messages
#21 opened 2 weeks ago by hschellman

Token Related Error Messages documentation error messages
#20 opened on Feb 23 by hschellman

opening root TBrowser
#19 opened on Nov 13, 2022 by krpalash

ssh error: client_loop: send disconnect: Broken pipe
#18 opened on Nov 13, 2022 by hschellman

Accessing DUNE computing from Windows Howto
#17 opened on Nov 8, 2022 by hschellman

How to back up your user files to tape documentation Howto
#16 opened on Feb 14, 2022 by hschellman

To-do: Write up instructions for VNC with a useful title Howto
#15 opened on Jan 28, 2022 by calcuttj

Getting xrootd paths for files documentation Howto
#14 opened on Jan 24, 2022 by hschellman

Getting and using the FNAL Virtual Private Network (VPN) Howto
#13 opened on Dec 10, 2021 by hschellman

Resetting service passwords documentation Howto
#12 opened on Nov 3, 2021 by hschellman

Could not chdir to home directory /nashome/b/bv: Permission denied error messages
#11 opened on Sep 16, 2021 by brettviren

Use of issue labels documentation
#10 opened on Sep 16, 2021 by demuth

X11 connection rejected because of wrong authentication error messages

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

Lessons Learned

- SWC template is elegant and functional for delivering active learning materials.
- It is also very popular for asynchronous training!
- A common look and feel for training materials has value.
- Markdown lessons allows experts to provide content without worrying (too much) about format
- **Pre-event homework that checks that participants can access FNAL servers is a must.**
- It’s easy to be (over) ambitious with the material for events.
- Mentors ensure skill development and understanding.
- Coffee breaks allow students to catch-up on lesson activities, some assisted by mentors.
- Hybrid synchronous delivery of lessons works well.
- Zoom captures on YouTube allow asynchronous access, and a record of the event.
- Indico and Github sites require coordination.
- Providing build support streamlines all aspects of the training event.
- Undergraduates can contribute to lesson designs.

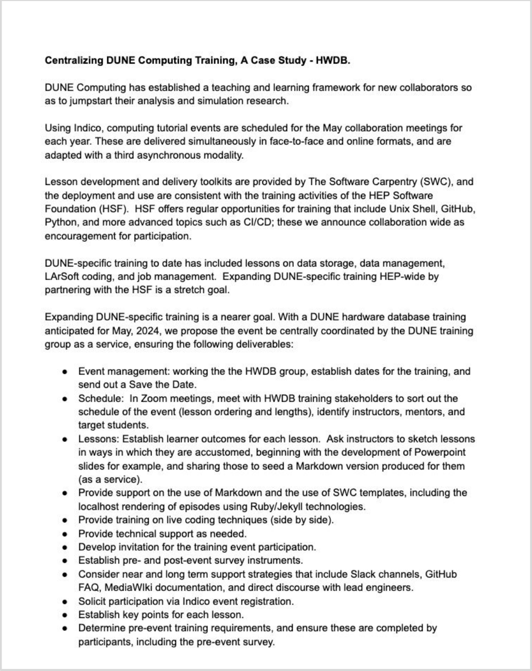
Case Study: HWDB

- Utilizing the same lesson templates and editing environment, a two half-day hardware database training was recently offered by the University of Minnesota (H. Muramatsu, A. Wagner, U. Ekka): Indico 65297

- Event management, lesson development environment setup, GitHub site management, were among the centrally coordinated tasks provided to the lead instructors: https://github.com/DUNE/computing-HWDB

- A four person team benefitted by weekly meetings over a two month development cycle; key was undergraduate U. Ekka's markdown and styling contributions, e.g. image shading and pop-ups.

- The HWDB training event evidences that other topics could be developed similarly, e.g., condoDB.



Centralizing DUNE Computing Training, A Case Study - HWDB.

DUNE Computing has established a teaching and learning framework for new collaborators so as to jumpstart their analysis and simulation research.

Using Indico, computing tutorial events are scheduled for the May collaboration meetings for each year. These are delivered simultaneously in face-to-face and online formats, and are adapted with a third asynchronous modality.

Lesson development and delivery toolkits are provided by The Software Carpentry (SWC), and the deployment and use are consistent with the training activities of the HEP Software Foundation (HSF). HSF offers regular opportunities for training that include Unix Shell, GitHub, Python, and more advanced topics such as CI/CD; these we announce collaboration wide as encouragement for participation.

DUNE-specific training to date has included lessons on data storage, data management, LArSoft coding, and job management. Expanding DUNE-specific training HEP-wide by partnering with the HSF is a stretch goal.

Expanding DUNE-specific training is a nearer goal. With a DUNE hardware database training anticipated for May, 2024, we propose the event be centrally coordinated by the DUNE training group as a service, ensuring the following deliverables:

- Event management: working the the HWDB group, establish dates for the training, and send out a Save the Date.
- Schedule: In Zoom meetings, meet with HWDB training stakeholders to sort out the schedule of the event (lesson ordering and lengths), identify instructors, mentors, and target students.
- Lessons: Establish learner outcomes for each lesson. Ask instructors to sketch lessons in ways in which they are accustomed, beginning with the development of Powerpoint slides for example, and sharing those to seed a Markdown version produced for them (as a service).
- Provide support on the use of Markdown and the use of SWC templates, including the localhost rendering of episodes using Ruby/Jekyll technologies.
- Provide training on live coding techniques (side by side).
- Provide technical support as needed.
- Develop invitation for the training event participation.
- Establish pre- and post-event survey instruments.
- Consider near and long term support strategies that include Slack channels, GitHub FAQ, MediaWIki documentation, and direct discourse with lead engineers.
- Solicit participation via Indico event registration.
- Establish key points for each lesson.
- Determine pre-event training requirements, and ensure these are completed by participants, including the pre-event survey.

Results from a survey

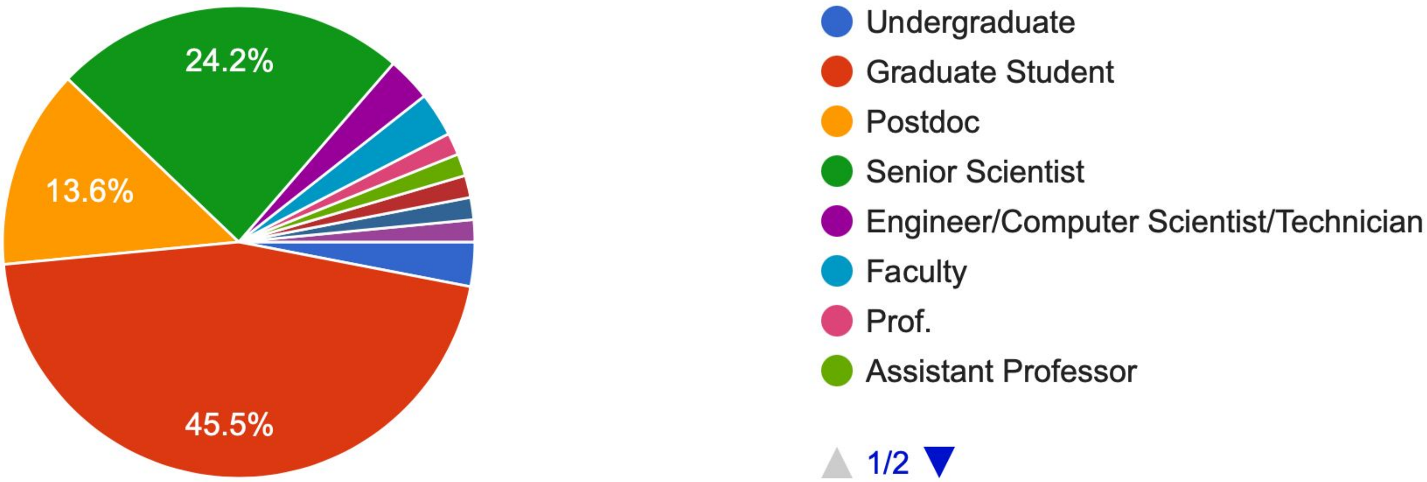
- Survey done before our last collaboration meeting in September 2024
- 66 responses from 20 countries

answers from

BR(2), CERN(3), CL(2), CO(2), CZ(1), DE(1), ES(1), FR(1), HU(1), IL(3), IN(1), IT(6), JP(1), MA(1), NL(2), PT(1), RO(1), TR(1), UK(3), US(30)

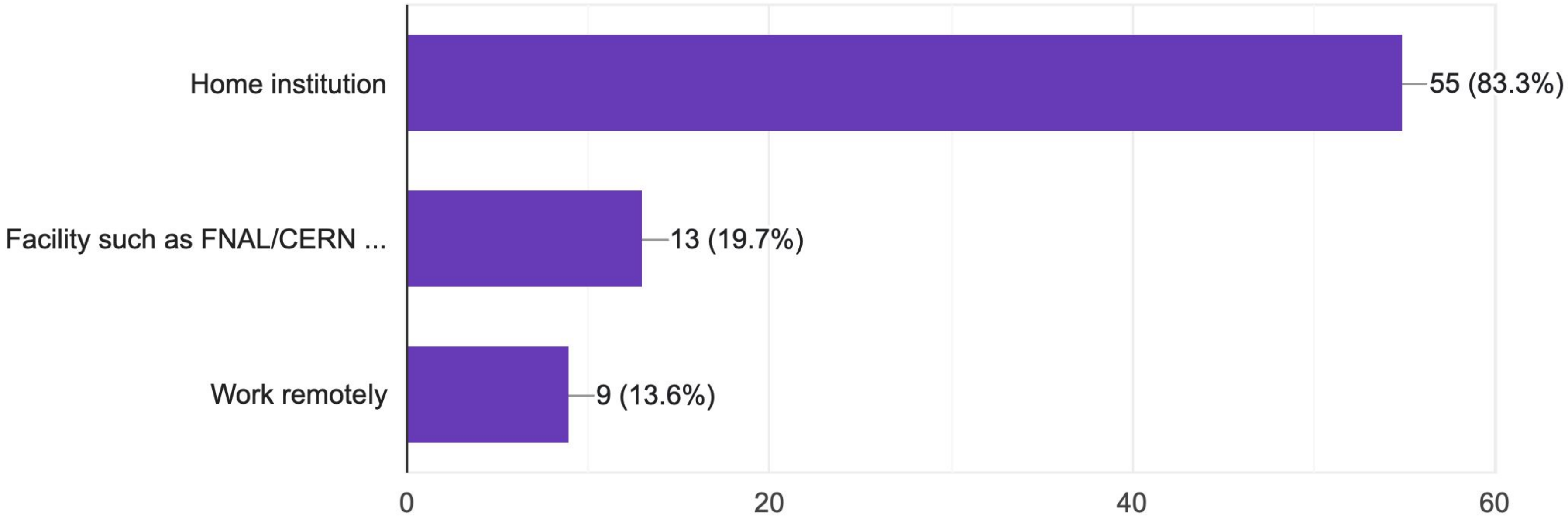
What career status are you?

66 responses



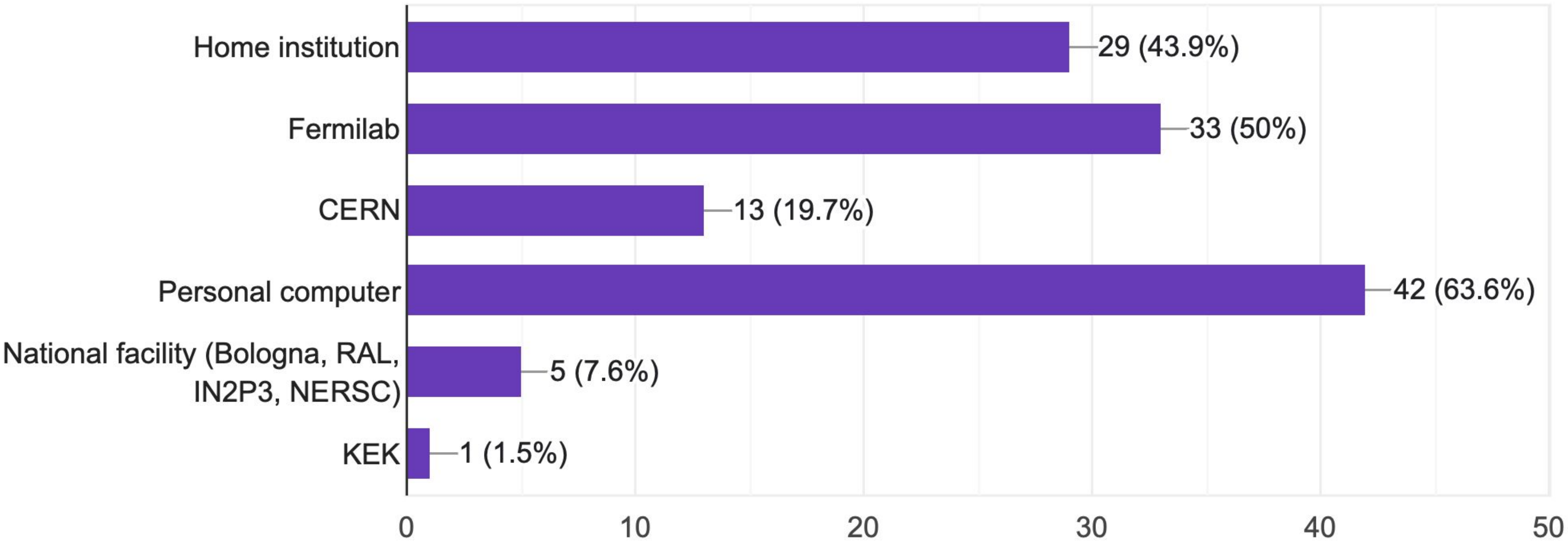
Where are you mainly located

66 responses



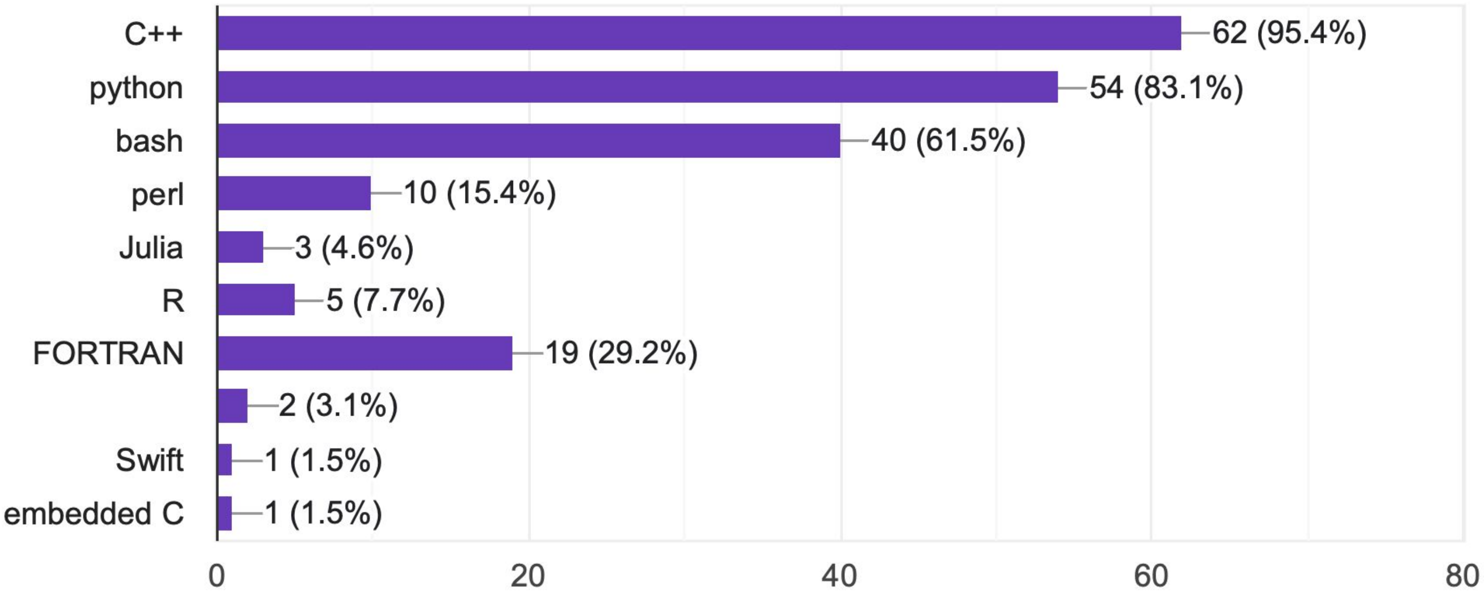
Where do you do your computing

66 responses



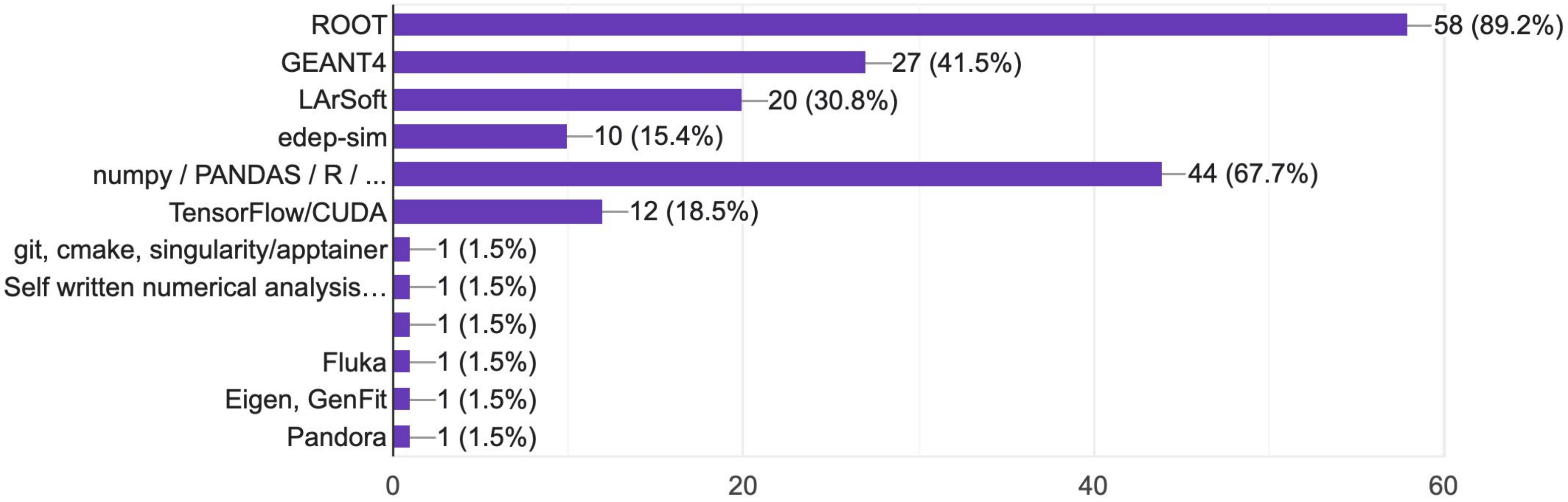
What Languages are you familiar with

65 responses



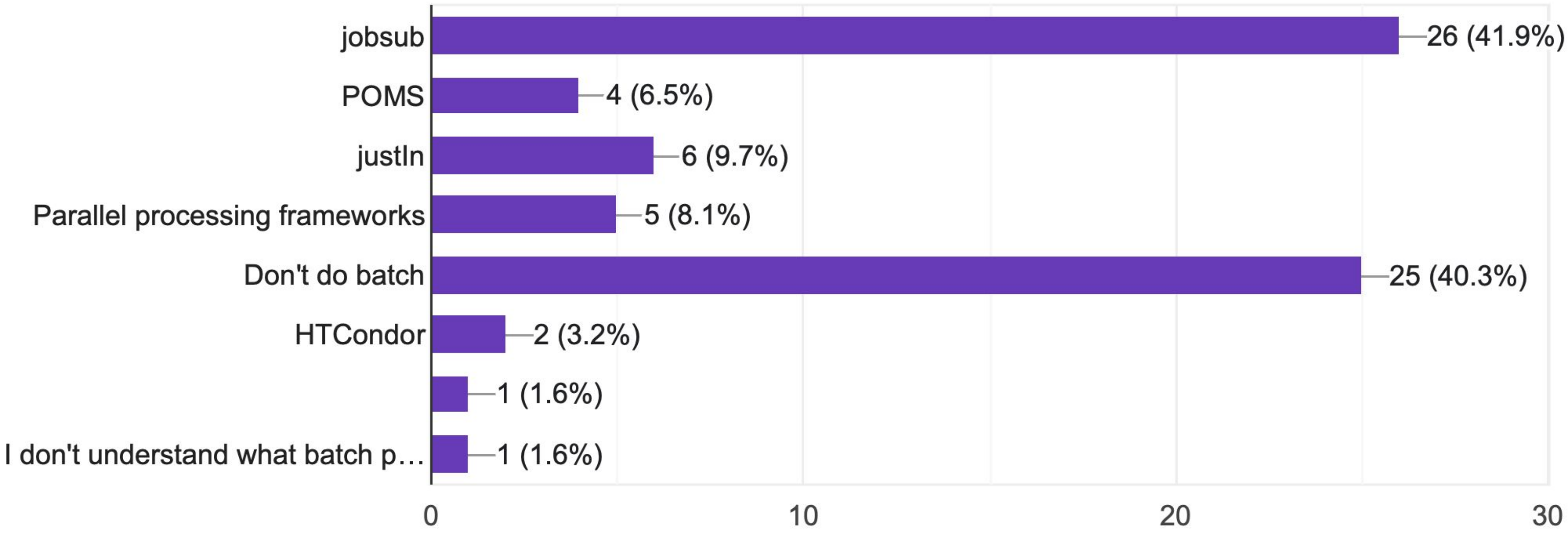
What software packages are you familiar with?

65 responses



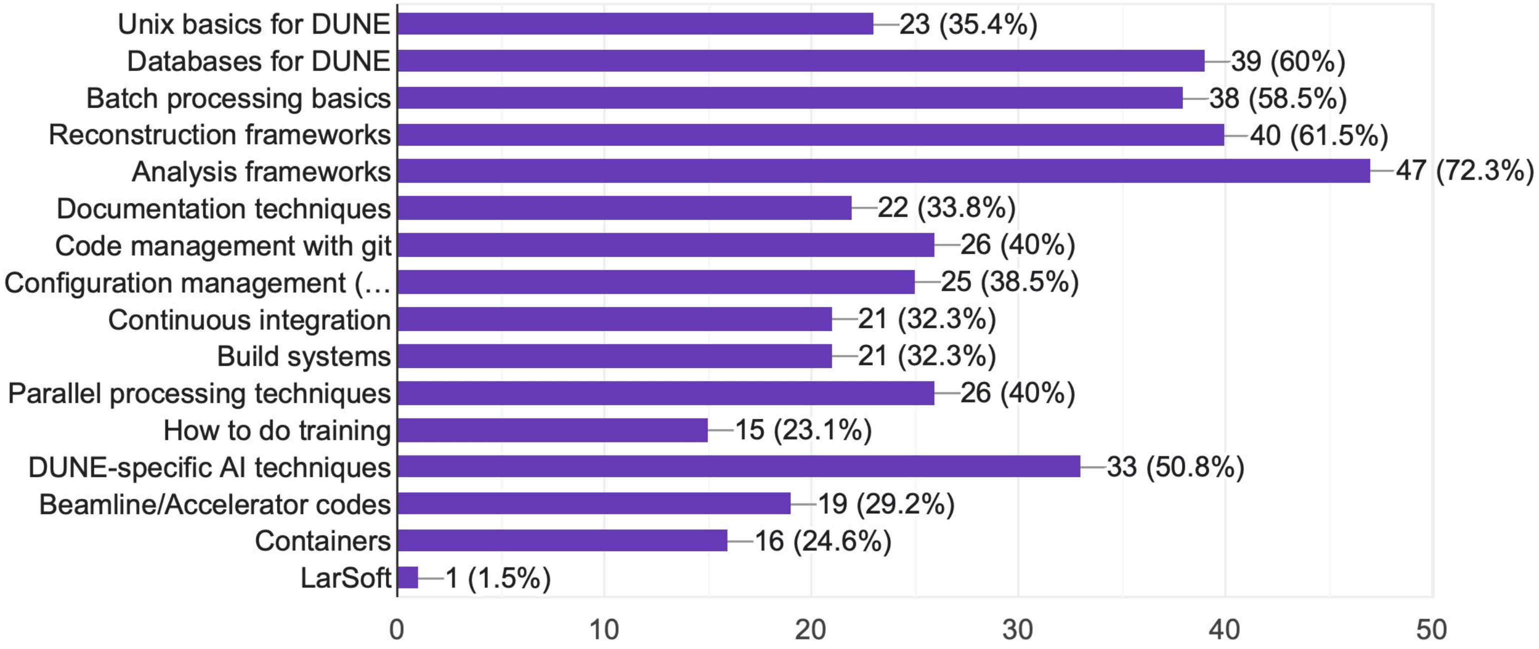
How do you do batch processing?

62 responses



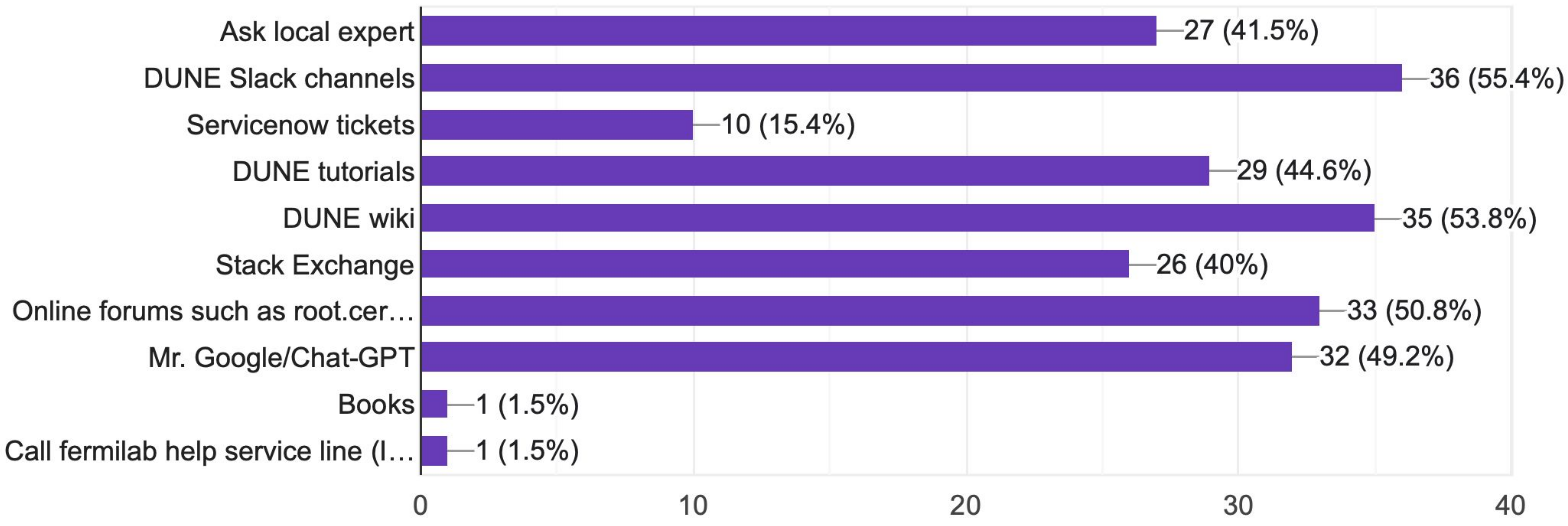
Would you be interested in learning more about:

65 responses



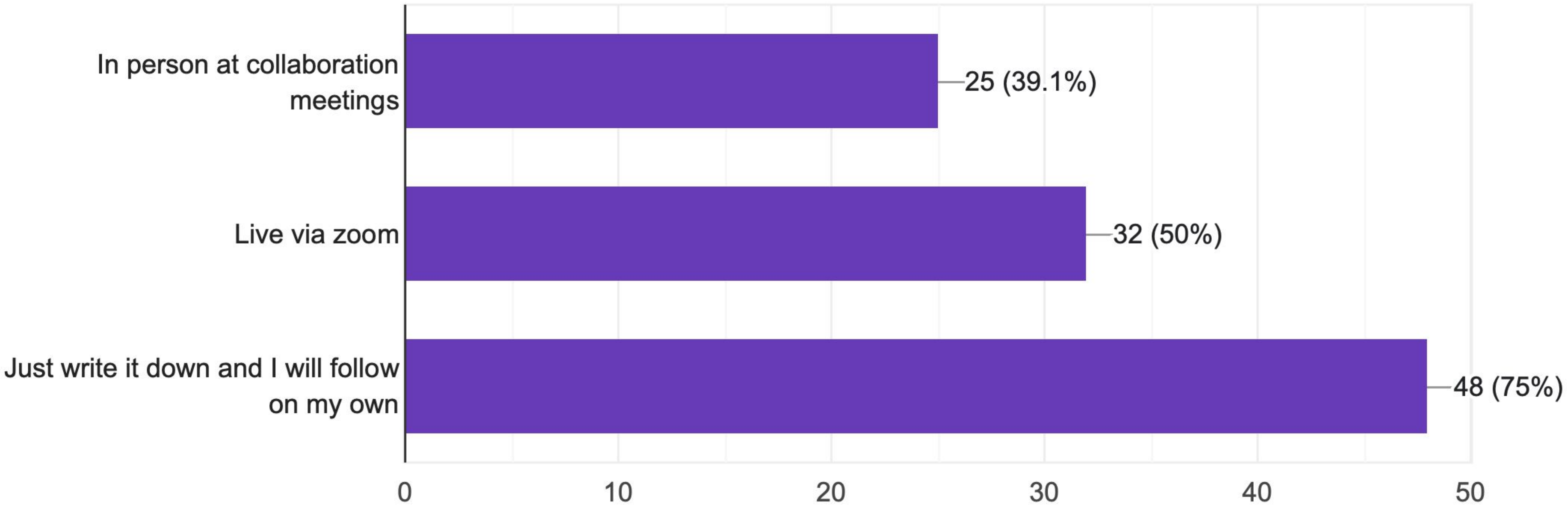
How do you get computing help?

65 responses



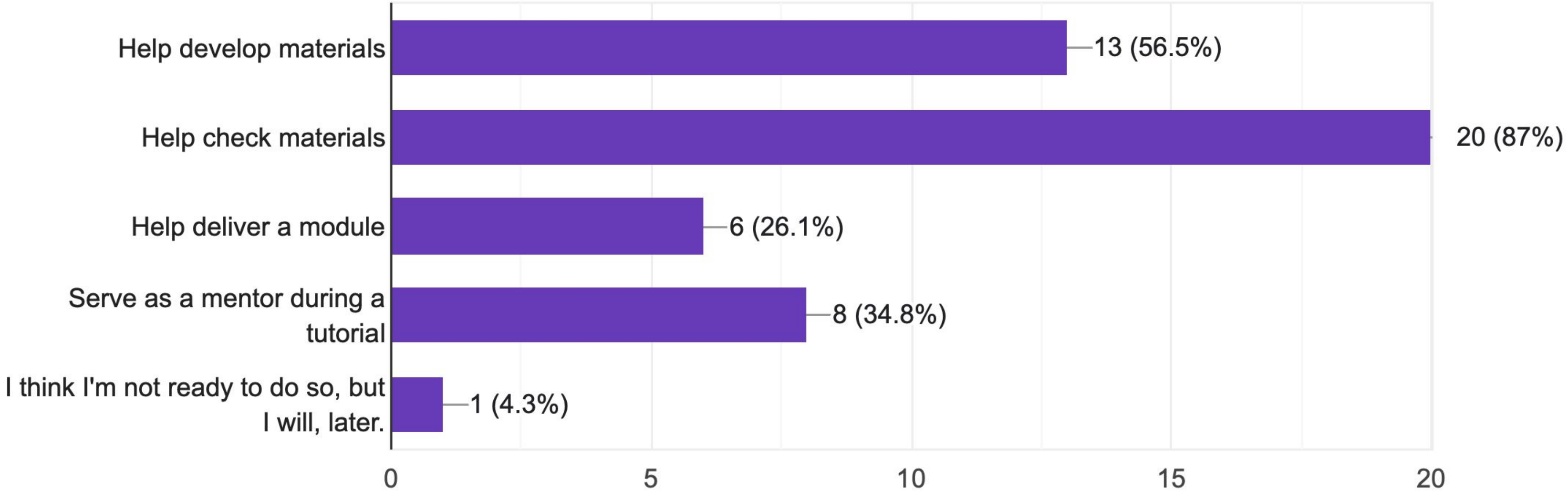
What format do you like for tutorials?

64 responses



Are you interested in helping with a tutorial?

23 responses



what would you like to help develop?

- When DUNE software is ready and available with Spack I would be happy to help prepare or give some tutorials
- Python
- Databases, anything else I know well enoug to be useful.
- LArSoft Basics
- Software development
- Reco/Sim framework, Analysis framework
- job submission
- Data management
- I'm interested in helping data analysis and simulation/reconstruction.
- GEANT4 and ROOT related topics mainly
- HWDB, but willing to help other DB-related topics
- python, analysis tools, git
- Unix, code management, configuration though mostly wherever my skill set best serves the groups needs.
- C, Python

When you joined DUNE what resources were you pointed to?

- DUNE wiki
- DUNE docb
- LArSoft workshop
- dune.github.io/computing-basics (the online tutorial)
- Google (which doesn't work well as most items are now hidden)
- Nothing….
- No mention of the DUNE slack among respondents.

Lessons learned

- Onboarding needs to be more uniform
- People like the tutorial format (Carpentries) but we need to extend to more topics
- DUNE software environment is changing rapidly
  - SAM -> Metacat/Rucio for data ✓
  - POMS□ Justin for batch 🔋
  - ups □ spack for configuration management ☄
- Any documentation we write now will be pretty up-to-date as so much is new.
- Basic training on batch systems (what they even are?) would be useful
- But many things are not yet documented