BROOKHAVEN NATIONAL LABORATORY & Jefferson Lab

FUTURE TRENDS IN NUCLEAR PHYSICS COMPUTING

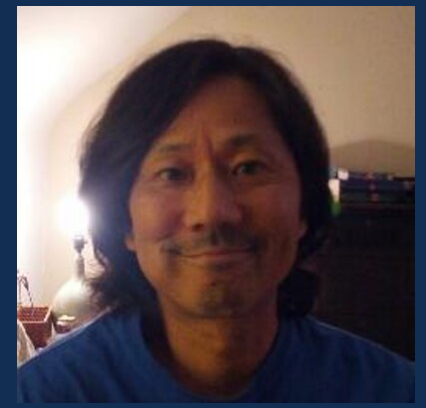
Workshop Summary

Organizers

Alexander Kiselev

Amber Boehnlein

Graham Heyes



Mark Ito

Markus Diefenthaler

Ofer Rind

Paul Laycock

Torre Wenaus

History

Future Trends in Nuclear Physics Computing in 2016

Goals

- Examined computing strategy at a time horizon of ten years
- Defined common vision for NP computing
- Recommended future directions for development

Website

- https://www.jlab.org/conferences/trends2016/

Future Trends in Nuclear Physics Computing in 2017

Goals

- Discussed trends in scientific computing
- Collected ideas on how to improve analysis
- Worked towards next-generation analysis techniques and tools

Website

- https://www.jlab.org/conferences/trends2017/

Goals

- Focus on the **Nuclear Physics Software & Computing community**
- Identify what is unique about our community
- Discuss how we could strengthen common efforts
- Chart a path for **Nuclear Physics Software & Computing** for the next ten years

THANK YOU FOR JOINING US!

207 participants

Organization

- We met for four hours each day in a time window chosen to be as inclusive as possible for participants around the world:

9:00 a.m. – 1:00 p.m. (EDT)

- Substantial discussion time was included in the agenda.

Common Scientific Software

- Tuesday, September 29
- **Moderated by** P. Laycock and T. Wenaus

The Role of Data Centers in Scientific Discovery

- Wednesday, September 30
- **Moderated by** G. Heyes and O. Rind

Unique Software Challenges for Nuclear Physics

- Thursday, October 1
- **Moderated by** M. Diefenthaler, M. Ito, and A. Kiselev

Discussion

Live notes

- https://docs.google.com/document/d/1mug_UB31WngFvvlLv8CFRBd4dSdKZR0iROL3m3lFY5Y/edit
- 26 (!) pages

Scope

- The **live notes** will be used to moderate and record the discussion and gather input from all participants, and after the workshop as the basis for summarizing and report writing.
- Questions and discussion points will be gathered in the **live notes** to guide a moderated common discussion following the talks.
- There will be time for short questions after each talk.

Common Scientific Software

Common Scientific Software: Password for watching the recording: sF0JQ^t#

Conveners: Paul Laycock (Brookhaven National Laboratory), Torre Wenaus (BNL)

Recording

9:00 AM	**Workshop Goals**	15m
	Speakers: Alexander Kiselev (BNL), Dr Amber Boehnlein (Jefferson Lab), Dr Graham Heyes (Jefferson Lab), Dr Mark Ito (Jefferson Lab), Markus Diefenthaler (Jefferson Lab), Ofer Rind (BNL), Paul Laycock (Brookhaven National Laboratory), Torre Wenaus (BNL)	
	TRENDS2020-Welc...	
9:15 AM	**Developing Common Software: ACTS**	30m
	Speaker: Andreas Salzburger (CERN)	
	2020-ACTS-Future-T...	
9:45 AM	**Questions on Talk**	10m
10:00 AM	**Software Sustainability**	30m
	Speaker: Daniel S. Katz (University of Illinois Urbana-Champaign)	
	nuclear-physics-co...	
10:30 AM	**Questions on Talk**	10m
10:45 AM	**Developing Common Software: Rucio**	30m
	Speaker: Mario Lassnig	
	Rucio @ Future Tren...	
11:15 AM	**Questions on Talk**	10m
11:30 AM	**Discussion on Common Software**	1h 30m

- No attempt to summarize the presentations, the slides stand on their own
- Our excellent speakers **on all days** made the session as interesting and useful as we could have hoped, thank you to them all
- And thanks to all participants for stimulating discussions, challenging questions, frank open answers

Common Scientific Software

- Engage other experiments early:
  - get early buy-in on cross-experiment interest if there's a thought to eventually making a software project common and cross-experiment
  - contribute to and collaborate on cross-experiment projects, cultivate reuse rather than building from scratch
- Be clear on your objectives and timeline: Short term goals are important, project should be grounded in real world deliverables from the beginning, but they need to be aligned with long term objectives
- Do not separate dev and ops. Developers are most effective when they are aware of their software in use, facing real operational issues
- Spread the effort; communities within a project working towards their own priorities, contributing code accordingly
- And data deserve the same attention to curation as the software (cf. the DAP talk on September 30)

- People are most important, not the software. Setting up an organization to create the right incentives to create and maintain the software.
- A strain repeated throughout the day: career support!
- Another strain supporting developers and their careers: software citations
- Common software projects create a pool of highly valuable, valued developers who can carry expertise on a key tool to other experiments and communities. cf. career path.
- Management support up the hierarchy is important for successful open source project
  - Acceptance of objectives wider than those of the home experiment
  - Recognition of the value of the wider investment
- Developers need the time and space to develop something new, not something just a little better

- **Ongoing discussion** Does NP need something like a NP Software Foundation? Or leverage/participate in HSF?

The Role of Data Centers in Scientific Discovery

09:00 → 13:00 **The Role of Data Centers in Scientific Discovery**

Conveners: Dr Graham Heyes (Jefferson Lab), Ofer Rind (BNL)

09:00	**Workshop Introduction and Recap** **Speakers**: Paul Laycock (Brookhaven National Laboratory), Torre Wenaus (BNL) Slides	15m
09:15	**Data Centers in a Decade** **Speaker**: Eric LANCON (BNL) Data Centers in a D...	30m
09:45	**Questions on Talk**	10m
10:00	**Data and Analysis Preservation and Open Data** ¶ **Speaker**: Maxim Potekhin (NPPS/Physics Department) Data and Analysis ...	30m
10:30	**Questions on Talk**	10m
10:45	**User Perspective and Requirements** **Speaker**: Dr Graham Heyes (Jefferson Lab) Users and Require...	30m
11:15	**Questions on Talk**	10m
11:30	**Discussion of the Role of Data Centers**	1h 30m

Our three speakers did a great job of examining the role of the data center from a number of perspectives:

- Looking at how this role may evolve over the next decade
- Looking at its particular role supporting long term data and analysis preservation for experiments
- Looking at the multifaceted interactions with its user base

Themes

- Local infrastructure decisions
- Distributed computing resources
- Long term viability of data
- Containerization - as a mechanism for aiding operation on distributed sites as well as preserving code for far future execution
- Interaction with users or stakeholders
- In all three talks communication and overall engagement was emphasized.

The Role of Data Centers in Scientific Discovery

- Common benchmarks that are consistently used by various sites participating in a distributed computing model.
  - Similarly common metrics that allow consistent evaluation of utilization and availability.
- The rapid growth of ML and AI in nuclear physics seen at both BNL and JLab.
  - Need to bring these groups together, at least within labs, to avoid duplication of effort and to share ideas.
- The balance between educating a user on the use of a complex system vs providing a simplified interface.
- The tension between HPC and HTC, where do supercomputers and large data centers fit.
  - Edge computing vs centralized and where the balance lies for NP.
- Collaboration between facility and users.
  - Embedding of people within experiments.
  - Physicists on Sci Comp staff as a bridge between physics and computer science.
  - Connecting ops oriented people with CS research.
  - Collaborating with CS researchers at universities.
- Promoting computing within the community
  - Career paths for joining science and computer science.
  - Pulling in a coherent direction.

Unique Software Challenges for Nuclear Physics

09:00 → 13:00 **Unique Software Challenges for Nuclear Physics Community: Password for watching the recording: D!thz2Vu**

Conveners: Alexander Kiselev (BNL), Dr Mark Ito (Jefferson Lab), Markus Diefenthaler (Jefferson Lab)

Recording

09:00	**Workshop Introduction and Recap** **Speakers**: Dr Graham Heyes (Jefferson Lab), Ofer Rind (BNL) Day 2 Summary · TRENDS2020-Day2...	15m
09:15	**Experience from 12 GeV Science Program** **Speaker**: David Lawrence (Jefferson Lab) 2020.10.01.12GeV... · Experience From th...	30m
09:45	**Questions on Talks**	10m
10:00	**Software Challenges in Streaming Readout** ¶ **Speaker**: Jan Bernauer (Stony Brook University and RBRC) futuretrends.pdf	30m
10:30	**Questions on Talks**	10m
10:45	**Survey from Nuclear Physics Students and Young Postdocs** TRENDS2020-Surv...	30m
11:15	**Questions on Survey**	10m
11:30	**Discussion of Unique Software Challenges for Nuclear Physics**	1h 15m
12:45	**Workshop Closing: Summary and Next Steps** **Speakers**: Alexander Kiselev (BNL), Dr Amber Boehnlein (Jefferson Lab), Dr Graham Heyes (Jefferson Lab), Dr Mark Ito (Jefferson Lab), Markus Diefenthaler (Jefferson Lab), Ofer Rind (BNL), Paul Laycock (Brookhaven National Laboratory), Torre Wenaus (BNL) TRENDS2020-Closi...	15m

- Importance of fast turnaround in data processing and having adequate monitoring and DQ tools
- DAP issue (again!): even big enough experiments do not have a working model in place by the time data taking starts
- Streaming readout: There is a substantial effort to unify the software in a way it can be used for small test setups and big experiments.
- Quality tutorials are extremely important for teaching young scientists, but the provided examples should work :-) at the very least; CI would not hurt either.

Unique Software Challenges for Nuclear Physics

- There is quite some connection between SRO and DAP, since the archived data is available for the analyses which were not anticipated at a time the experiment is taking data. A lot of data mining associated with the new physics topics can happen later in time.

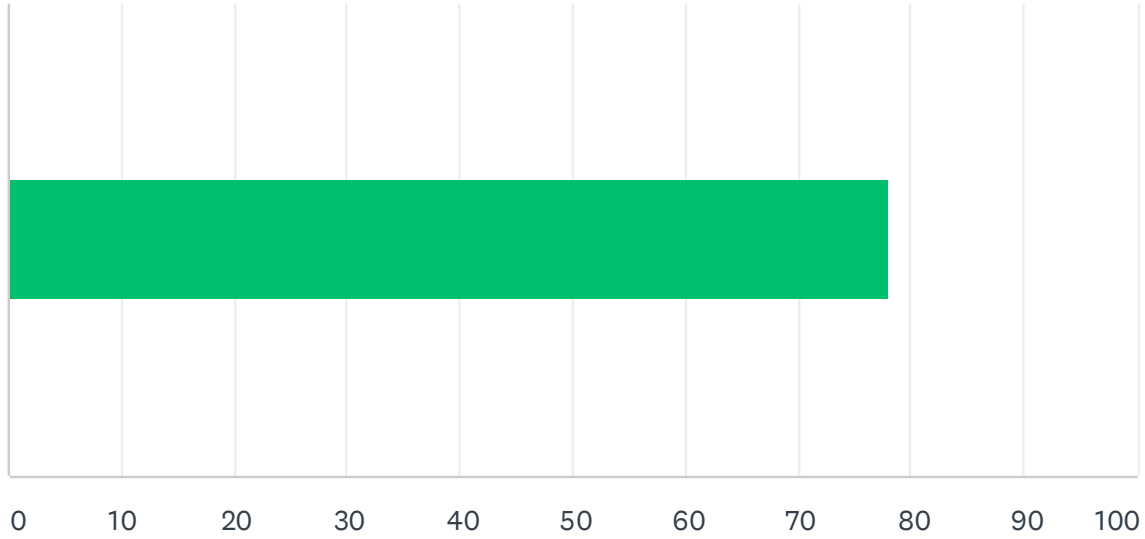
- What is unique about NP (software community)
  - New requirements for the MC generators
  - HEP MC generators are “more monolithic” in a sense (?)
    - NP: smaller experiments, faster turnaround times, also on the generator front
- Data analysis procedure is typically different: e.g. use the whole event sample to extract some angular modulation rather than to search for the specific event signatures.
- Harder requirements on the systematics.
- brute force approach to go through a data sample using custom analysis software works less and less, so: data size matters
- Hardware and software sustainability: should one rely on a big experiment hardware or software, having a concern that either of them disappear after a few years. Well, ROOT will stay forever :-)

FUTURE TRENDS IN NUCLEAR PHYSICS COMPUTING

Selected Survey results: Input from 44 Ph.D students and postdocs

Q1 What fraction of your time do you spend on the software and computing aspects of your research, such as programming, analysis jobs, etc.?

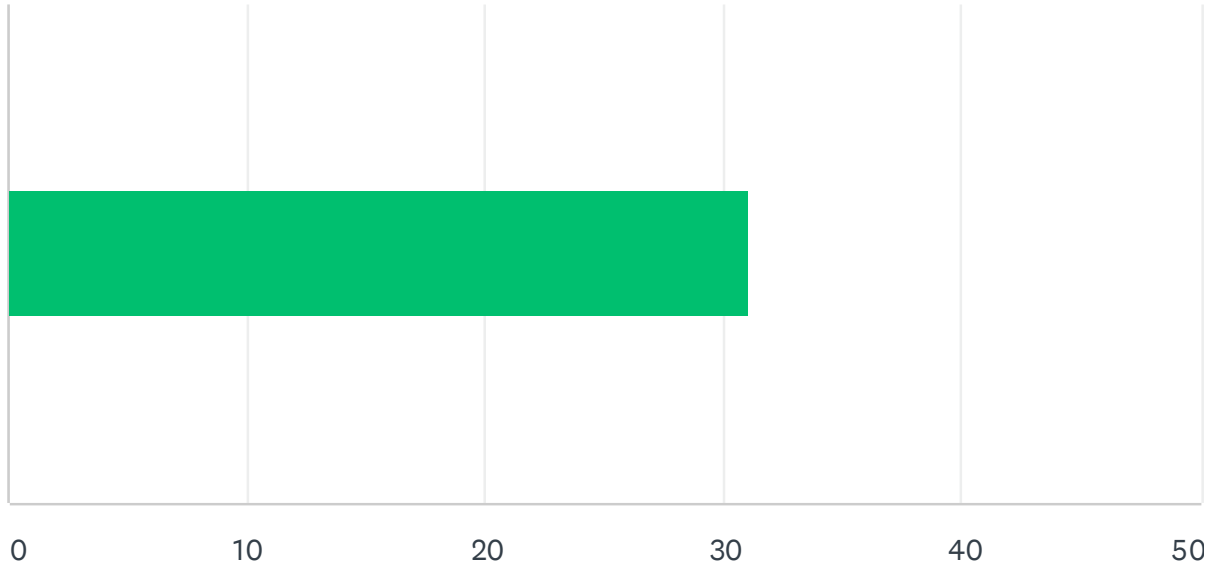
Answered: 44 Skipped: 0



ANSWER CHOICES	AVERAGE NUMBER	TOTAL NUMBER	RESPONSES
	78	3,435	44
Total Respondents: 44			

Q2 What fraction of your time do you spend contributing to software and computing tasks for other members of your research or analysis group?

Answered: 44 Skipped: 0

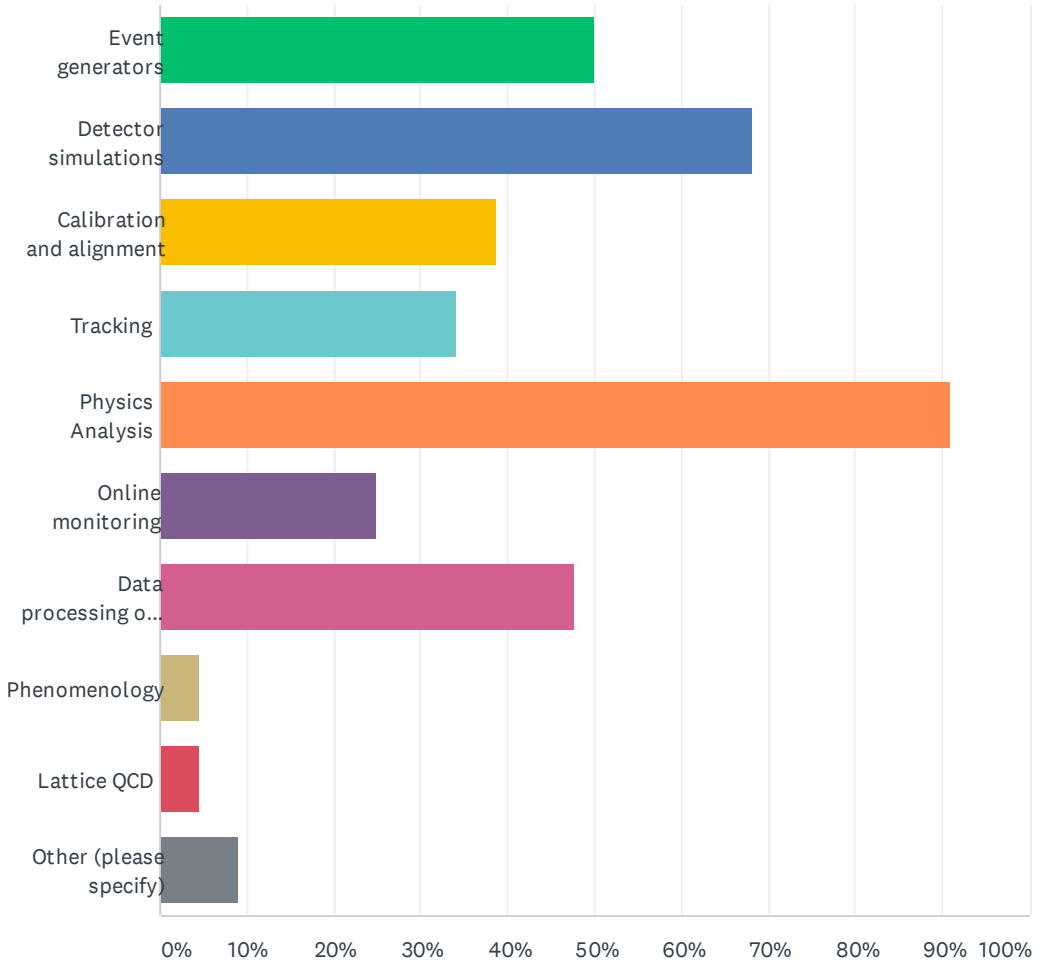


ANSWER CHOICES	AVERAGE NUMBER	TOTAL NUMBER	RESPONSES
	31	1,367	44
Total Respondents: 44			

Survey

Q5 What aspects of nuclear physics software and computing do you work on? Select all that apply.

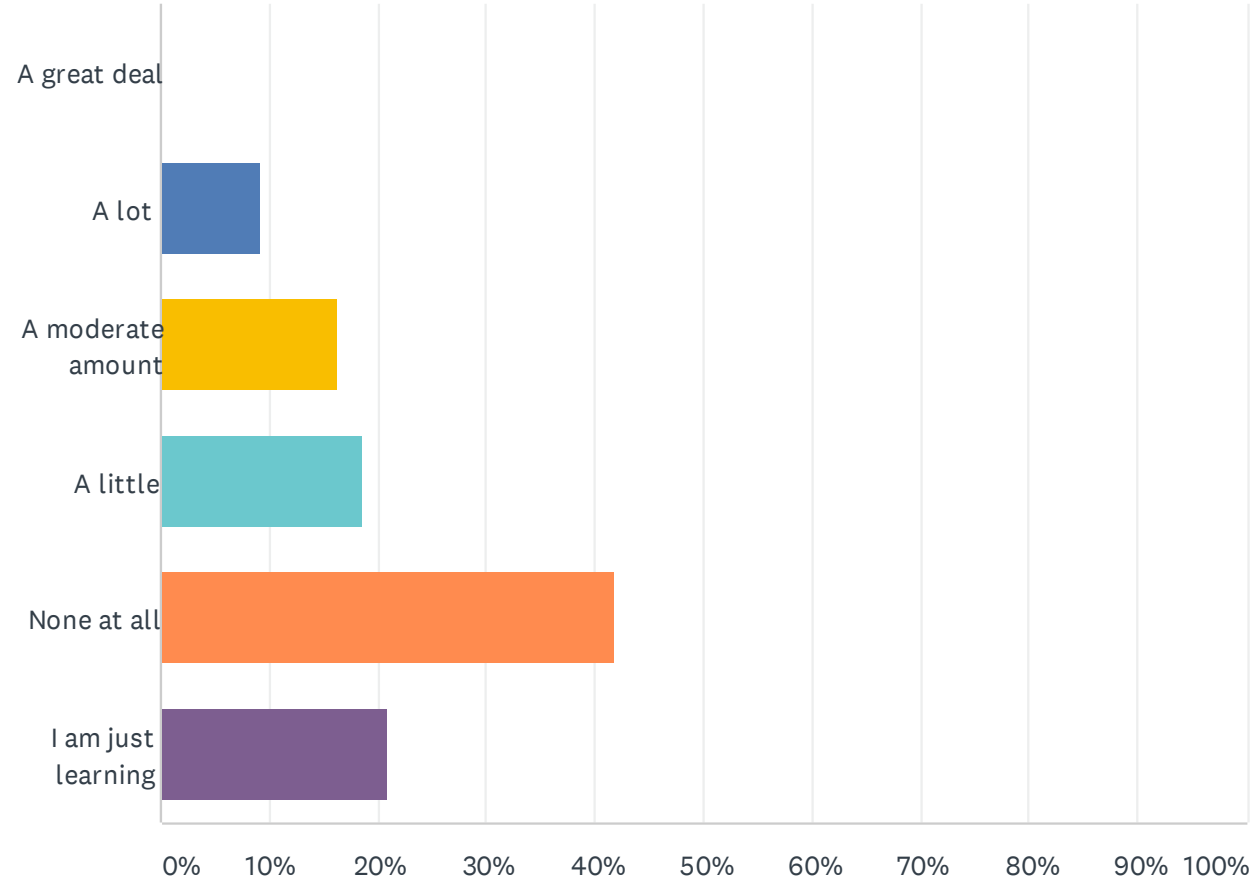
Answered: 44 Skipped: 0



#	OTHER (PLEASE SPECIFY)
1	DAQ and real-time data suppression
2	detector analysis with cosmic ray
3	Slow controls
4	Containerizing software

Q11 To what extent do you use ML or AI in your research?

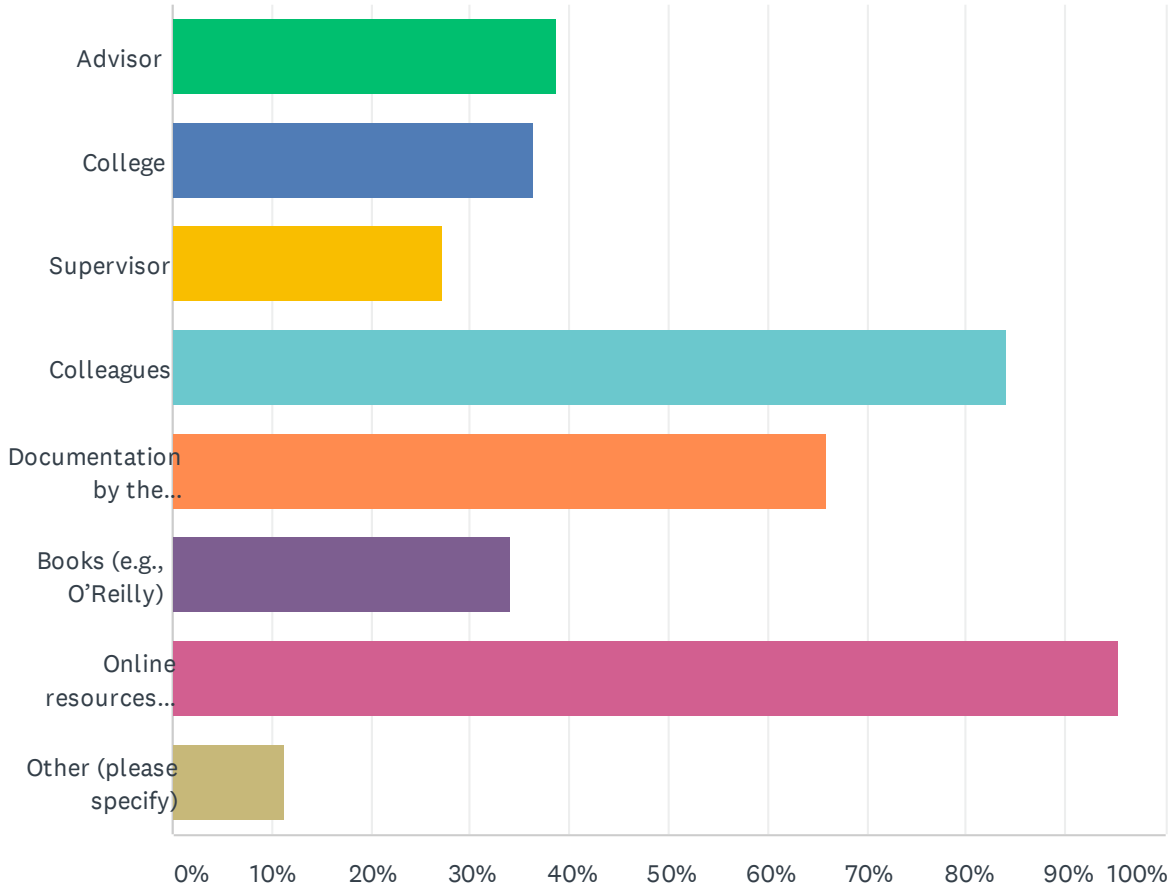
Answered: 43 Skipped: 1



Survey

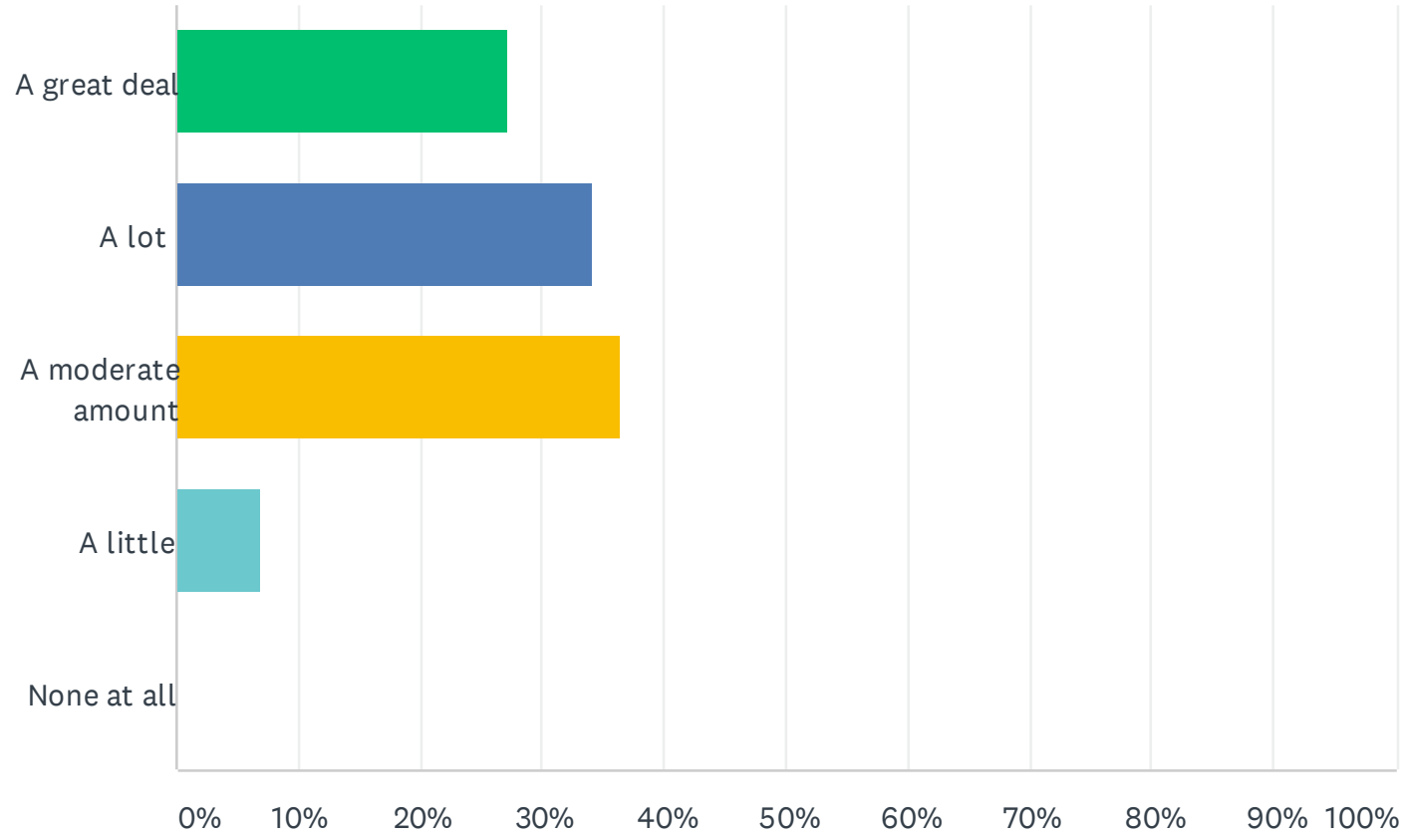
Q4 Which of the following resources have you used for your software and computing work or research? Select all that apply.

Answered: 44 Skipped: 0



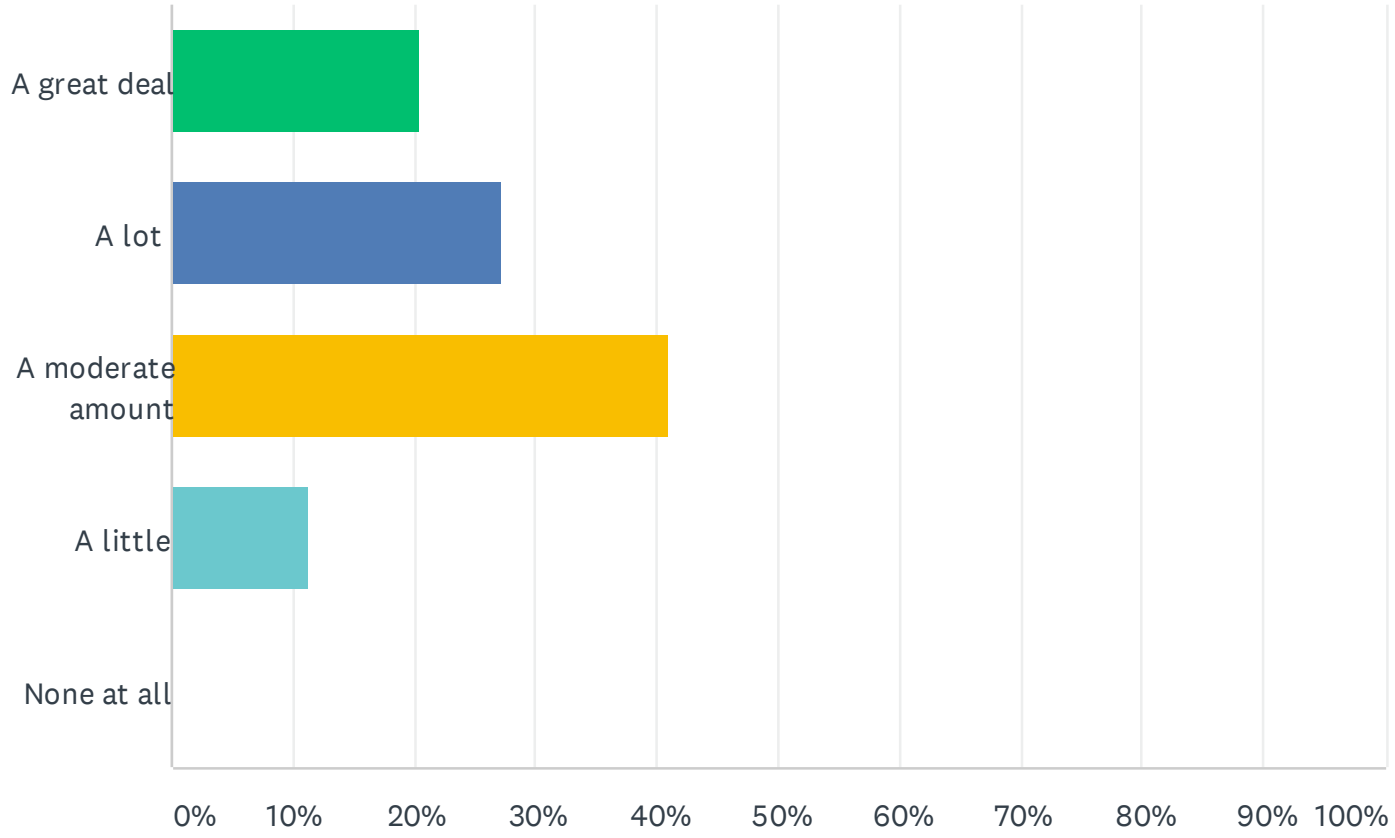
Q6 Are you currently able to perform the software and computing tasks needed for your research?

Answered: 44 Skipped: 0



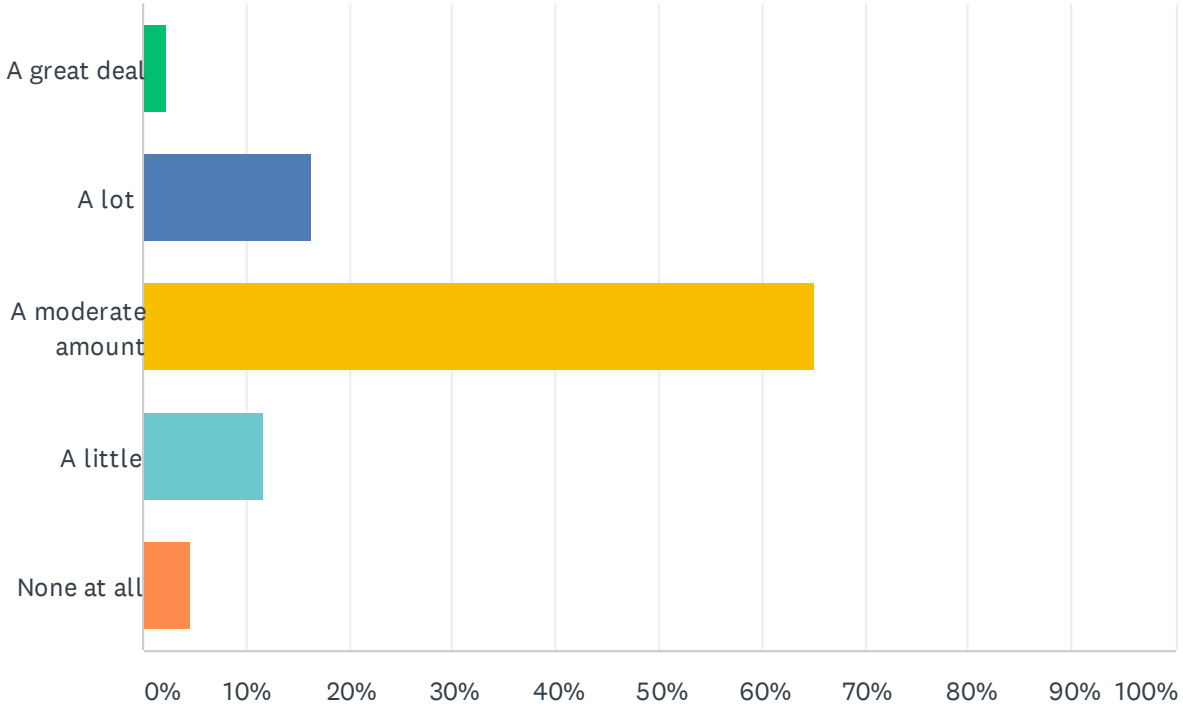
Q7 Are you confident performing software and computing tasks needed for your research?

Answered: 44 Skipped: 0



Q8 Are the documentation tools you use adequate?

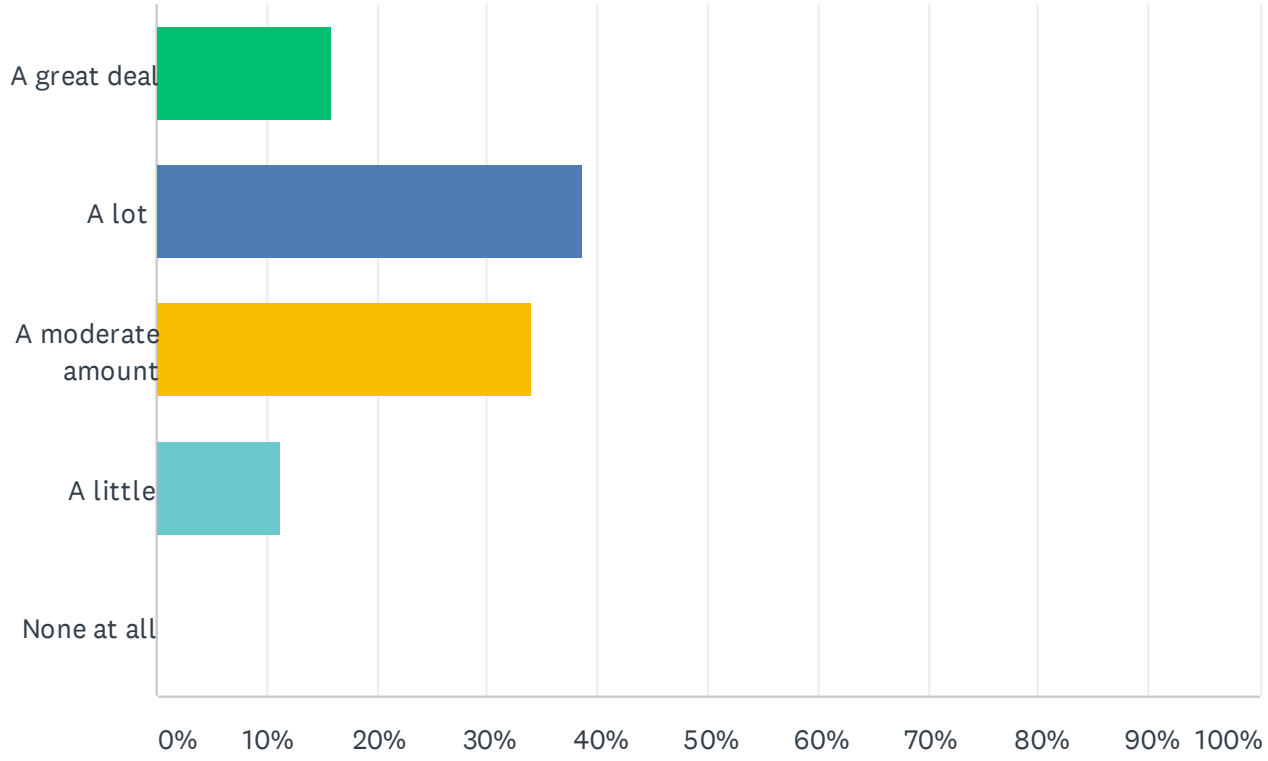
Answered: 43 Skipped: 1



#	PLEASE COMMENT
1	For Python, I can find easy to use resources and/or ask questions online and get ready answers. For ROOT, it feels like the documentation is outdated and esoteric, and trying to get answers online is merely an exercise in frustration.
2	I don't understand the question
3	Not always, no. But this may be a symptom of the code base I use not being supported (just bad timing).
4	Documentation in the two collaborations that I have worked with are awful, outdated, and only experts know what still works and what changes have been done.

Q9 Are the computing resources you need readily available?

Answered: 44 Skipped: 0



4	It seems that funds are spent mostly on computing resources, but very little on computing expertise (if we had half the computers and twice the number of software experts we would be in a much better place)

Q15 What barriers do you encounter in your research that are software and computing related?

Selected comments

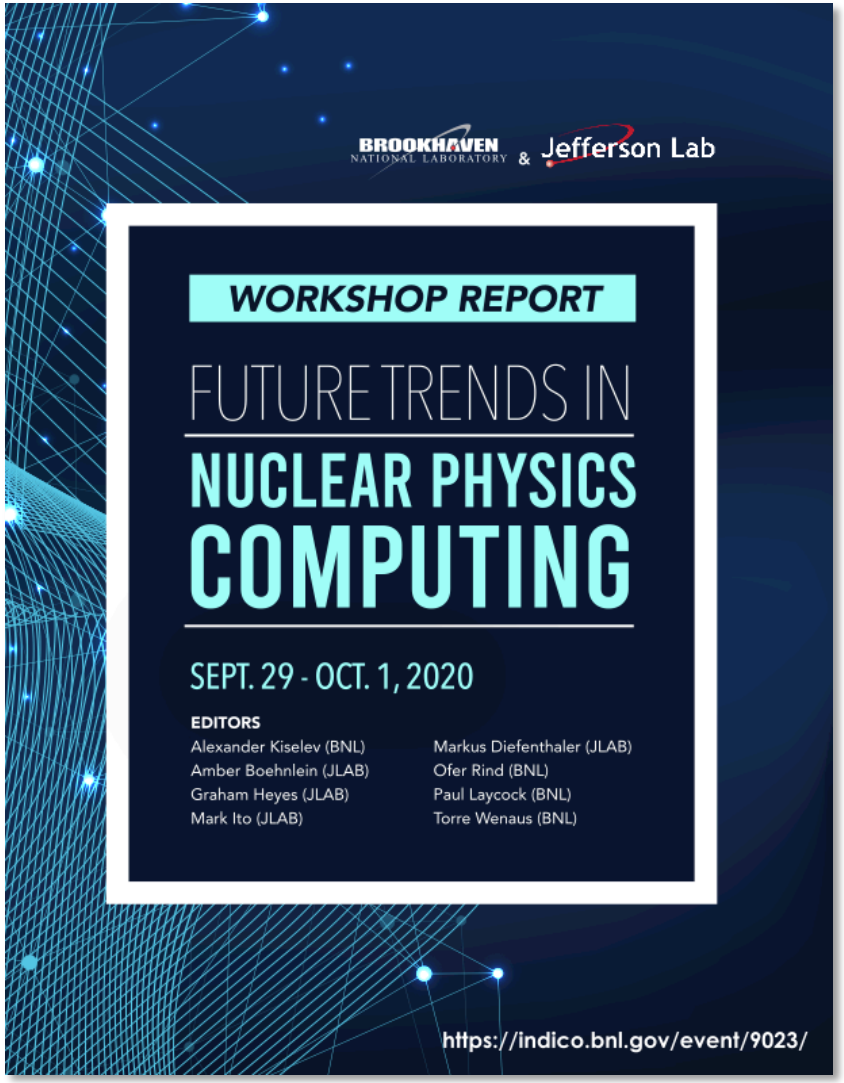
1. Training: Need the fundamentals as well as advanced topics
2. Out of date or otherwise inadequate documentation
3. Resources are difficult to find or to use or quite time consuming
4. Computing resources, especially disk space
5. Lack of support for software and computing systems.
6. Some really specific comments that may have merit, but cannot be addressed.

Q16 Anything we missed? Anything you would like to add? Please contribute here.

Selected comments

1. Career issues…
2. Changing the culture to respect S&C as a contribution and that one reason the training & document issues are so acute is a general lack of knowledge in the research group coupled with a lack of standardization of tool suites so the students have no place to turn.
3. Having “Research Software Engineers” available.

Workshop Summary and Report



- The workshop concluded with a short summary.

Workshop Summary

- Strong community participation also in the workshop summary with a lot of comments.