ARGONNE
ATPESC2025
EXTREME-SCALE COMPUTING

Data Intensive Computing and I/O

Rob Latham, **Phil Carns**, Kevin Harms,
Scot Breitenfeld, Jean Luca Bez, and Lev Gorenstein

August 7, 2025

Welcome to Track 7 of ATPESC 2025

Data Intensive Computing and I/O

- How many of you rely on large amounts of data in your work?
- *Would you like to generate / move / analyze that data 2x faster?*

We want to equip you with the knowledge and techniques to get there!

Hands on exercises:
https://github.com/radix-io/hands-on

An overview of today's topics

Conceptual Overview

"Why do we do things this way?"

Systems and Tools

"How do I interact with storage systems and manage my data?"

Application I/O Libraries

"What's the most effective way to access data from my application?"

Understanding and Optimizing

"I'm using the right tools and APIs; why isn't my performance better?

Hands on exercises:
https://github.com/radix-io/hands-on

Meet your Argonne lecturers

Phil Carns is a computer scientist at ANL focused on measurement, modeling, and development of data services. He has contributed to several influential storage research projects including Mochi, Darshan, CODES, and PVFS.

Rob Latham is a research software developer at ANL who strives to make applications use I/O more efficiently. He has played a prominent role in the ROMIO MPI-IO implementation, the PVFS file system, and the PnetCDF high level library.

Kevin Harms is the technical director of the ALCF-4 project. His interests include high performance storage and I/O. He has helped to design, deploy, and operate storage systems across multiple generations of ALCF platforms.

Meet your expert guest lecturers

Scot Breitenfeld is the director of engineering at the HDF Group. His technical specialty is HPC application use of HDF5. He has implemented and tuned HDF5 for HPC applications and third-party libraries across a wide range of platforms.

Jean Luca Bez is a research scientist at LBNL. His research focuses on optimizing the I/O performance of scientific applications guided by I/O performance analysis, I/O scheduling, and ML-driven automatic tuning and reconfiguration.

Lev Gorenstein is a Solutions Architect at the Globus Department of the University of Chicago. After running long MD simulations in a structural biology lab, he switched to the dark side and began supporting HPC clusters and users. He now facilitates Globus adoption and applications worldwide.

Hands on exercises:
https://github.com/radix-io/hands-on

Logistics for ATPESC-IO

Agenda:

https://extremecomputingtraining.anl.gov/agenda-2025/#Track-7

Discussion and questions:

- Please ask questions as we go!
- At least one of us will monitor the #atpesc-2025-track-7-io slack channel at all times.
- We can provide one-on-one help and relay questions to lecturers if needed.
- We will be available on Slack for the remainder of the ATPESC program.

Hands-on Exercises

Hands-on exercises and machine reservations:

See https://github.com/radix-io/hands-on

How to use the hands-on exercises:

- Some exercises will be covered in real time during today's lectures.
- Additional exercises are available for you to explore at your own pace according to your interests.
- Please work on the ones that are the most important to you and ask questions as you go!

We are also happy for you to experiment with I/O strategies in your own codes and "ask the experts" for help.

ATPESC attendees have dedicated reservations on Aurora and Polaris (ALCF) today for experiments and exercises, but you are welcome to use and ask questions about any of the ATPESC systems.

Thanks!

Any questions about logistics before we roll up our sleeves and get to work?

Hands on exercises:
https://github.com/radix-io/hands-on

ARGONNE
ATPESC2025
EXTREME-SCALE COMPUTING

ARGONNE TRAINING PROGRAM ON EXTREME-SCALE COMPUTING

Produced by Argonne National Laboratory, a U.S. Department of Energy Laboratory managed by UChicagoArgonne, LLC under contract DE-AC02-06CH11357.

Special thanks to the National Energy Research Scientific Computing Center (NERSC) and Oak Ridge Leadership Computing Facility (OLCF) for the use of their resources during the training event.