Data Intensive Computing and I/O

Track 7   **August 8, 2024**

Rob Latham, **Phil Carns**, Shane Snyder,
Kevin Harms, Scot Breitenfeld, and Greg Nawrocki

extremecomputingtraining.anl.gov

Argonne
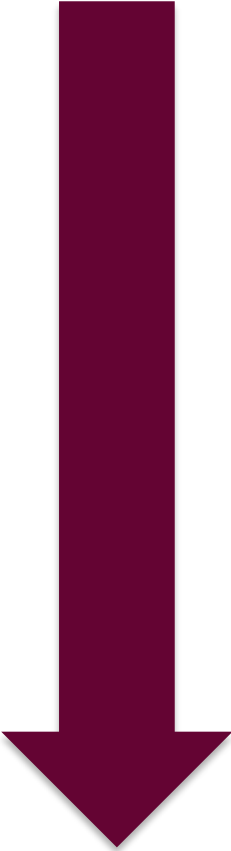NATIONAL LABORATORY

Welcome to Track 7 of ATPESC 2024

Data Intensive Computing and I/O

We want to help you answer the following questions:

1. What are the key things to know about HPC data storage?

2. What data management tools are available, and how do I use them?

3. How can I access data more efficiently in my application?

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

extremecomputingtraining.anl.gov

Today's topics

- Morning:
  - Introductory concepts
  - System walkthroughs
  - Instrumentation methods
  - Data movement
- Afternoon
  - I/O libraries
    - MPI-IO
    - PnetCDF
    - HDF5
  - Performance tuning
  - Discussion

Going into more detail as the day goes on

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

extremecomputingtraining.anl.gov

ATPESC2024
EXTREME-SCALE COMPUTING

Argonne
NATIONAL LABORATORY

Meet your 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.

Shane Snyder is a software engineer at Argonne National Laboratory. His research interests include the design of high-performance distributed storage systems and the characterization and analysis of I/O workloads on production HPC systems.

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

extremecomputingtraining.anl.gov

Argonne
NATIONAL LABORATORY

Meet your lecturers (expert guests)

Scot Breitenfeld specializes in HPC application use of HDF5 at The HDF Group. He has implemented, troubleshot, and tuned HDF5 for a broad spectrum of HPC applications and third-party HDF5 based libraries for a variety of platforms.

Greg Nawrocki is the Director of Customer Engagement of the Globus Department at the University of Chicago. He has also worked in high energy physics, the television and consumer products industry, and as the co-founder of a data analytics company.
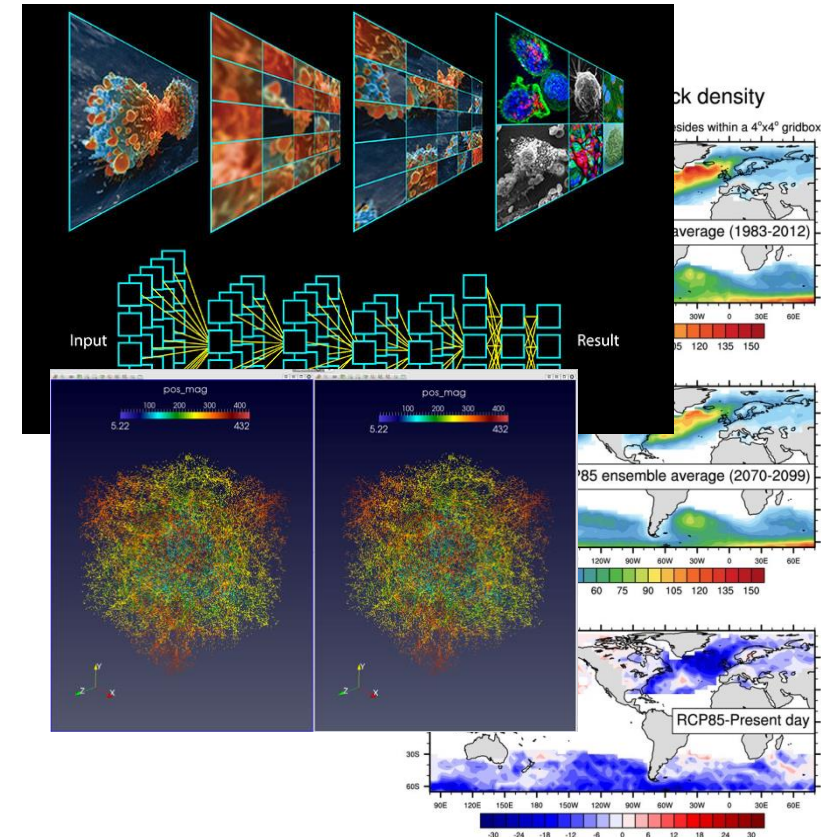

Kevin Harms is the technical director of the ALCF-4 project. His interests include high performance storage and I/O. He has helped to design, deploy, and operate storage systems across multiple generations of ALCF platforms.

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

extremecomputingtraining.anl.gov

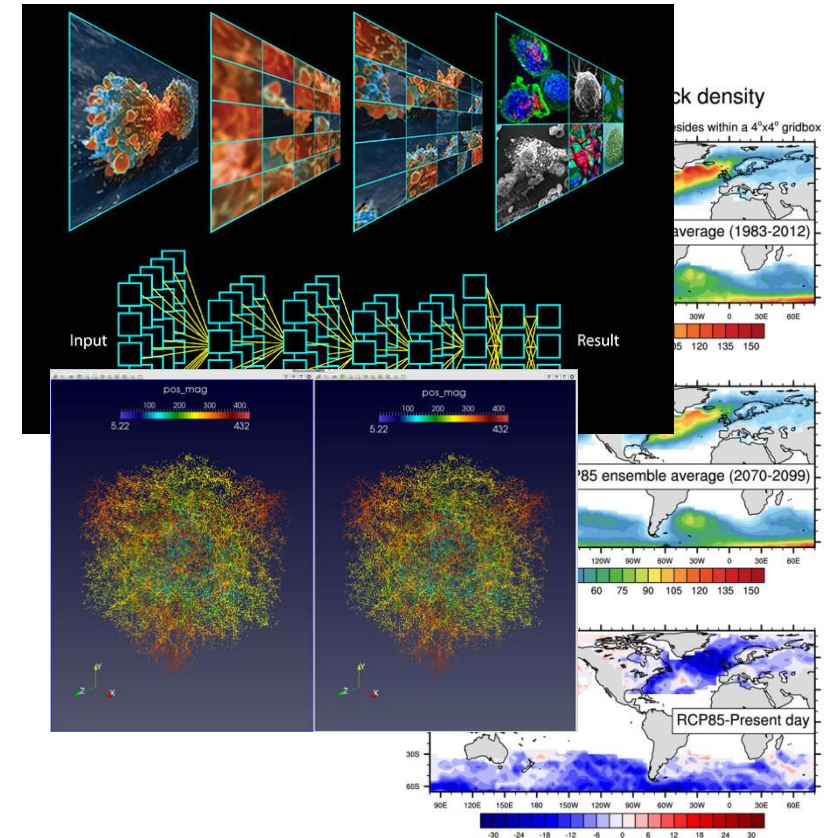# The overall theme of our work: bridging the gap between science and storage systems

What are the most effective and productive ways to use high performance storage devices to solve scientific problems?

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

extremecomputingtraining.anl.gov

Argonne
NATIONAL LABORATORY

The overall theme of our work:
bridging the gap between science and storage systems

- Building and operating effective systems
- Developing and applying efficient interfaces
- Measuring and improving performance
- **Putting new data storage technology into the hands of scientists**

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

extremecomputingtraining.anl.gov

Argonne
NATIONAL LABORATORY

Logistics for ATPESC-IO

- Agenda:
  - https://extremecomputingtraining.anl.gov/agenda-2024/#Track-7

- Discussion and questions:
  - Please ask questions as we go!
  - At least one of us will monitor the #atpesc-2024-track-7-io slack channel at all times.
  - We can provide one-on-one help and relay questions to lecturers if needed.

- Note that tomorrow's Machine Learning track will delve into some data management topics as well.

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

extremecomputingtraining.anl.gov

Argonne
NATIONAL LABORATORY

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.**

- We will be available on Slack for the remainder of the ATPESC program.

ATPESC attendees have a dedicated reservation on Polaris (ALCF) today for experiments and exercises, but you are welcome to compile and run jobs on any of the ATPESC systems.

ARGONNE
ATPESC 2024
EXTREME-SCALE COMPUTING

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

extremecomputingtraining.anl.gov

Argonne
NATIONAL LABORATORY

Thanks!

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

ATPESC2024
EXTREME-SCALE COMPUTING
ARGONNE

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

extremecomputingtraining.anl.gov

Argonne
NATIONAL LABORATORY