ARGONNE
ATPESC 2023
EXTREME-SCALE COMPUTING

Data Intensive Computing and I/O

August 10, 2023

Rob Latham, **Phil Carns**, Shane Snyder,
Scot Breitenfeld, Mike Brim, and Greg Nawrocki

extremecomputingtraining.anl.gov

ECP EXASCALE COMPUTING PROJECT

Argonne NATIONAL LABORATORY

Welcome to Track 7 of ATPESC 2023
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?

Today's topics at a high level

- Morning:
  - Introductory concepts
  - System walkthroughs
  - Instrumentation methods
  - Data movement
- Afternoon
  - I/O libraries
    - MPI-IO
    - PnetCDF
    - HDF5
  - Tuning performance
  - Discussion
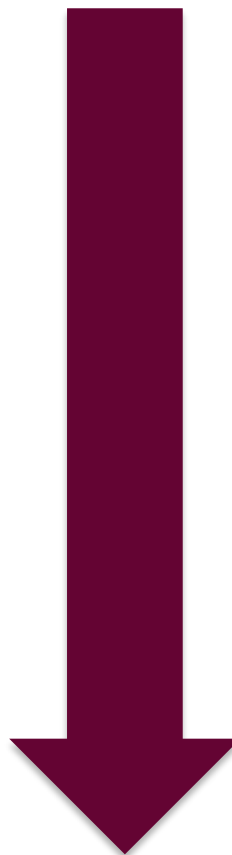
Going into more detail as the day goes on

Meet your lecturers (Argonne staff)

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

ATPESC2023
EXTREME-SCALE COMPUTING

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.

Mike Brim is a senior R&D staff member in the National Center for Computational Sciences at Oak Ridge National Laboratory. Michael has over 20 years of research experience in scalable tools, middleware, and systems software for HPC systems.

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

Why we do what we do: bridging the gap between science and storage systems
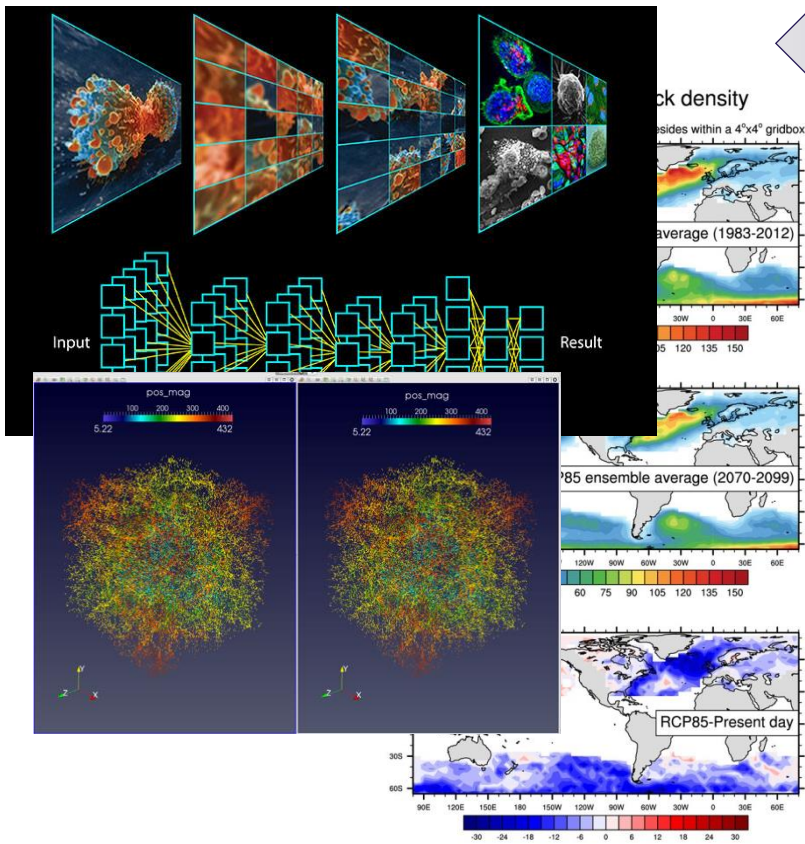

There are many different high performance storage technologies available. How can we use these technologies to meet the needs of scientists?

We need techniques, algorithms, and software to bridge the "last mile" between storage systems and scientific applications.

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

extremecomputingtraining.anl.gov

Why we do what we do: bridging the gap between science and storage systems
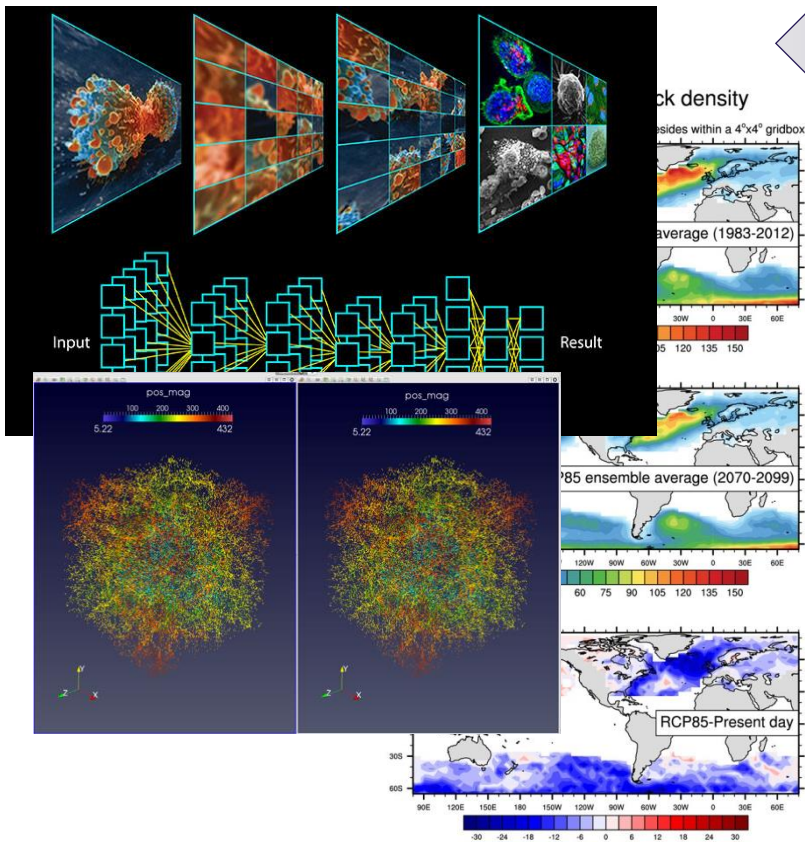
Examples of how we do this:

- Build and optimize data services
- Operate data centers
- Understand how storage is used
- Predict how storage will be used
- **Put new data storage technology into the hands of scientists**

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

extremecomputingtraining.anl.gov

Logistics for ATPESC-IO

- Agenda:
 - https://extremecomputingtraining.anl.gov/agenda-2023/#Track-7
- Discussion and questions:
  - Please ask questions as we go!
  - At least one of us will be monitoring the #atpesc-2023-track-7-io slack channel at all times.
  - We can provide one-on-one help and relay questions to lecturers if needed.
- Hands-on exercises and machine reservations:
  - See https://github.com/radix-io/hands-on
  - Please work on exercises at your own pace.
  - Continue to reach out to us through the remainder of the ATPESC program if you have questions.

ATPESC attendees have a dedicated reservation on Polaris (ALCF) today for experiments and exercises from noon to 9pm, but you are welcome to compile and run jobs on any of the ATPESC systems.

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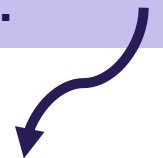
Thanks!

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