Data Intensive Computing and I/O

ATPESC 2021

Rob Latham, **Phil Carns**, Shane Snyder, Scot Breitenfeld, Suren Byna, and Vas Vasiliadis

August 6, 2021

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

Welcome to Track 3 of ATPESC 2021:

Data Intensive Computing and I/O

The goal of this track is to help you answer the following questions:

How do HPC storage systems work?

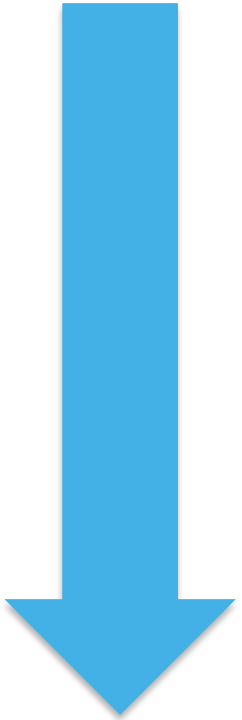
How can my application access data more efficiently?

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

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

Today's topics at a high level

- Morning:
  - Introductory concepts and tools
  - I/O libraries (MPI-IO and PnetCDF)
- Afternoon
  - I/O libraries (HDF5)
  - Tuning
  - Managing and transferring data
  - Discussion



Building up to more practical detail as the day goes on

Meet your lecturers (Argonne staff)

Phil Carns is a principal software development specialist at ANL who works on measurement, modeling, and development of data services. He has made key contributions to a variety of storage research projects, including Mochi, Darshan, CODES, and PVFS.

Rob Latham is a principal software development specialist 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 primarily 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

Meet your lecturers (Expert guests)

Scot Breitenfeld is with The HDF Group and specializes in HPC application use of HDF5. He has implemented, troubleshot, and tuned HDF5 for a broad spectrum of HPC applications and third-party HDF5 based libraries for various machine architectures and parallel file systems.

Suren Byna is a staff scientist in the Scientific Data Management Group at LBNL who works in optimizing parallel I/O and developing systems for managing scientific data formats. He leads the ECP ExaIO project, enhancing HDF5 for exascale, and the Proactive Data Containers project, managing data in deep storage hierarchies.

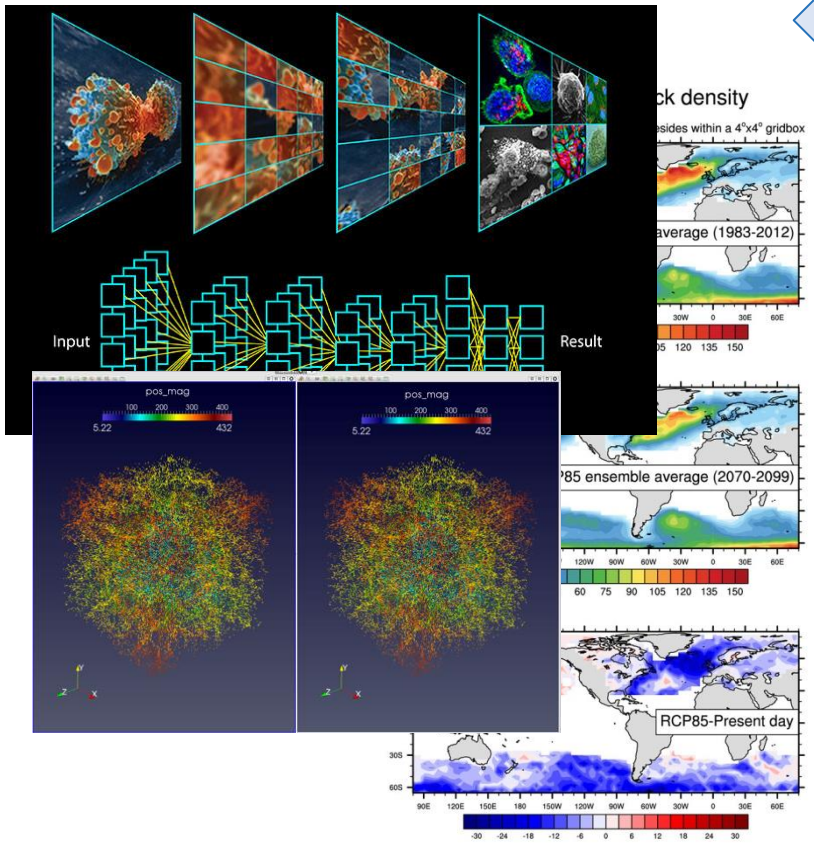
Vas Vasiliadis is the chief customer officer at Globus, where he works to help users get the most out of large scale data transfer technologies. He also teaches cloud computing and product management at the University of Chicago.

Argonne NATIONAL LABORATORY

ECP EXASCALE COMPUTING PROJECT

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

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



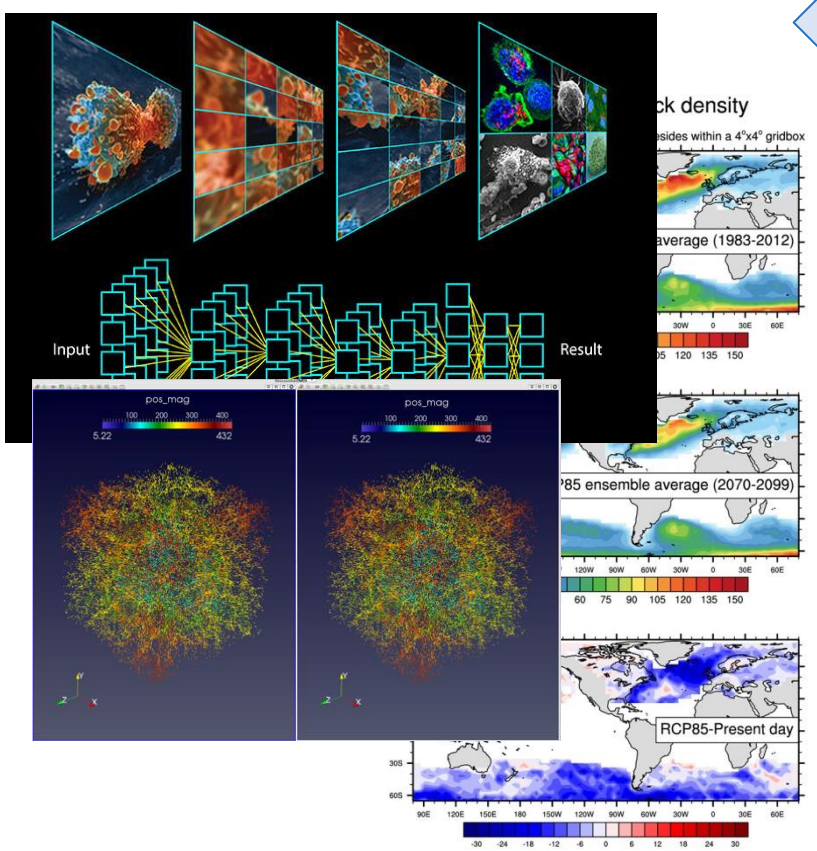
There are many different high performance storage technologies available. How can we use these technologies to meet the needs of scientists?

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

Argonne NATIONAL LABORATORY

ECP EXASCALE COMPUTING PROJECT

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



Examples of how we do this:

- Building/optimizing data services
- Operating data centers
- Understanding how storage is used
- Predicting how storage will be used
- **Putting new data storage technology into the hands of scientists**

Logistics for ATPESC-IO

ATPESC attendees have a dedicated reservation on Ascent (OLCF) and Theta (ALCF) today for experiments and exercises. See the link at the top of each slide for details.

- Agenda:
  - https://extremecomputingtraining.anl.gov/agenda-2021/#Track-3
- Discussion and questions:
  - Please ask questions as we go!
  - At least one of us will be monitoring the #track-3-data-and-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
  - We don't have much time blocked specifically for hands-on exercises.
  - 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.

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?