Streaming, Storing, and Sharing Big Data for Light Source Science

Justin M. Wozniak,*† Kyle Chard,† Ben Blaiszik,† Michael Wilde,*† Ian Foster*†‡
*Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, USA †Computation Institute, University of Chicago and Argonne National Laboratory, Chicago, IL, USA ‡Dept. of Computer Science, University of Chicago, Chicago, IL, USA

I. Introduction

Synchrotron x-ray facilities, such as the Advanced Photon Source and Advanced Light Source, use a variety of scattering, imaging, and spectroscopic techniques to address a range of problems in materials science. The increasing brightness of such x-ray sources, coupled to recent developments in detector technologies, means that they must handle significantly greater data rates than in the past, both on individual beam lines and across facilities as a whole. Pulsed neutron sources, such as the Spallation Neutron Source, face similar big data challenges; their instruments include large multidetector arrays that require complex data reduction.

Access to such national scientific resources is granted via a competitive review process, with researchers often waiting months for a few hours or days of experimental time. Delays in data analysis can make all the difference between a successful and failed experimental session. Thus, techniques that can enable more rapid and effective (and, in particular, real-time) analysis can greatly enhance the productivity of both individual researchers and these large-scale scientific investments.

Single crystal x-ray scattering experiments provide a good illustration of the challenges associated with experimental data analysis. x-rays is scattered by a small crystalline sample into a large pixelated detector. A typical detector can collect images of size 2,048×2,048 pixels at around 10 Hz and thus can generate data at 160 MB/s. Given inevitable pauses for experiment set up, etc., a week-long experimental campaign may produce 10 TB raw data.

Such large data sets must be subjected to both automated and interactive analysis. For example, software tools are required to transform raw data from instrumental coordinates to the reciprocal lattice coordinates defined by the crystal structure. Sharp Bragg peaks that are periodic in the reciprocal lattice can be used to determine the average crystalline structure. In disordered materials, such as those used for batteries or electronics, the real structure deviates from this average, producing substantial *diffuse scattering* between the Bragg peaks.

Here we describe our approaches for for organizing both raw experimental data and subsequent derived data, and for managing and interacting with such large data over networks. We have applied these techniques to accelerate real-time analysis of x-ray scattering data and thus improve the effectiveness of large-scale scientific investments.

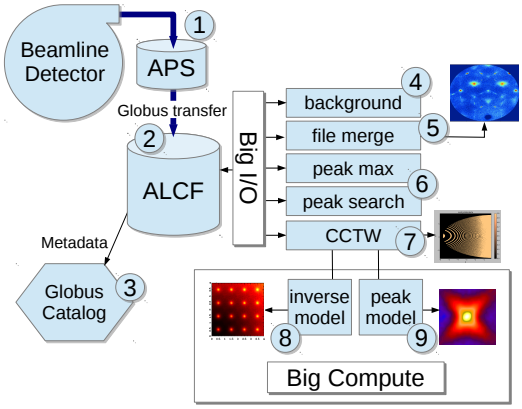

Figure 1: APS data analysis workflow, showing the data-intensive data ingest and compute-intensive phases.

II. Approach: Streaming APS x-ray data

We illustrate in Figure 1 a scientific workflow [1] that motivates this work. This workflow spans systems at the Advanced Photon Source (APS) at Argonne National Laboratory (ANL), the Argonne Leadership Computing Facility (ALCF) at ANL, and the cloud-hosted Globus service suite [2].

Streaming and storing: At the APS (1), detector data is collected at a beamline station, a lead-lined hutch containing the sample and a beamline computer [3], typically a PC with a specialized I/O interface to the detector hardware. This PC must not be delayed by any work other than transferring detector data to storage; if its buffers are filled, it will drop detector images. In our architecture, data is stored in a RAM-based filesystem, thus, storage is highly constrained. Data must be transferred off of this computer to a larger hard disk-based storage system. We do this by running a lightweight `rsync`-based script at a high `nice` level (that is, low priority) and deleting data once it has been safely transferred.

Data sizes and other parameters vary from experiment to experiment. In our most recent use of the APS beamline facility, in April 2015, scattering data was collected as a sample was rotated 360 degrees in tenth of a degree increments, yielding a total of 3,600 images. Each image comprises 2048×2048 32-bit pixels, i.e., 16 MB, and thus a full dataset is 56 GB. (This high resolution is required to capture the subtle signals associated with the sample's structural defects.) Images are

produced at a rate of 10 per second, and thus at a peak rate of 160 MB/s. Following the collection of a complete 3,600 image dataset, which takes about 10 minutes, sample conditions may be changed (e.g., by changing temperature) and the process repeated. Or, a new sample may be substituted. The former process can be fully automated, allowing data collection to proceed 24/7 during the experimental window. Overall, data collection rates averaged 14.5 MB/s during our last one-week session (April 2015).

Once on ALCF resources (2), automated data-intensive operations (4)–(7) perform data bundling, metadata creation, and cataloging quickly and automatically. Background noise from the detector, determined via an earlier calibration process, is subtracted from the data (4). Each rotation of 3,600 image files is then assembled into a single NeXus-format file [4]. NeXus is a set of conventions for the use of HDF5 structured data format [5]. Each file is tagged with appropriate metadata such as temperature and other sample parameters.

The metadata and location information are also registered with a data catalog, allowing for near-real-time viewing with the NeXpy GUI [6] for NeXus data. This makes it easy for scientists to check for experimental errors while in the experiment hall, allowing for detection of calibration errors or other anomalies that could, after the fact, invalidate an entire campaign [7]. Automated data analysis operations then find Bragg peaks and perform a coordinate transform that requires significant computation and data movement.

The transformed data can be used for two purposes, both requiring large computation. An inverse modeling scheme (8) is used to compare against simulated crystal structures, iteratively converging on a good approximation to the true crystal structure, including defects. This scheme employs an evolutionary algorithm to optimize a large "population" of simulated crystal structures by running them through a forward model that produces a simulated scattering image [8].

Searching and viewing: We have developed a comprehensive system for remote access to datasets produced by light source experiments [9]. The system is highly modular, and consists of a metadatabase (Globus Catalog), bulk data movement system (Globus Transfer), and remote object interface for interactive operation (NXFS). It represents a complete, highly adaptable solution to indexing and accessing scientific big data as it is streamed.

Globus Catalog offers a highly collaborative, user-friendly metadata annotation system usable in multiple ways, from the web to Python interfaces to command line tools. It integrates well with the Globus Transfer system and supports data annotation at acquisition time, interactive queries, and long-term data management. We have developed a Catalog interface in NeXpy, which is allows users to locate data as though using a File→Open interface.

NXFS exposes NeXus/HDF datasets over Python remote object interfaces, offering a mix of filesystem and object service features, seamlessly enhancing NeXpy with remote data access techniques. It integrates well with Globus Catalog, and exceeds the performance of application-agnostic remote filesystem techniques.

III. Future work

Over the coming decade, the planned upgrade APS will present unprecedented scientific capabilities to study 4D material properties at length scales from angstroms to centimeters, time resolution from picoseconds to days, all while capturing chemical, magnetic and electronic state contrast. Along with these new capabilities come a host of challenging data, automation, and computation challenges that must be met in order to maximize instrument capabilities and scientific productivity. These challenges include 1) Processing data from higher resolution, faster detectors at rates orders of magnitude faster than currently produced at the APS [10]; 2) Developing advanced, scalable computational methods for interpretation of data; 3) Utilizing HPC resources to enable fitting and co-optimization of model and experiment and to provide real-time feedback to the user [11]; 4) Reducing barriers to integrate experiment and HPC, such as co-scheduling; and 5) Automating processes to embrace a wider demographic of intrument users.

Our ongoing work aims to generalize our remote data access and operation model via a standard and secure Globus HTTP endpoint model. This approach will provide a standard interface to our services that can be used by third-party applications to access data remotely for real-time visualization and steering of ongoing experiments and simulations. Methods are required for remotely subsetting data and directly accessing file contents, especially in the case of structured file formats (e.g., HDF).

Globus Catalog supports the management of distributed and heterogeneous data, however it must be enhanced to meet new challenges as the number and complexity of data sources and elements (files, directories, and datasets) increases. Specifically, hybrid storage models (e.g. entity schema for dense data and decomposed storage for sparse data) are required to efficiently represent and query data; enhanced command line interfaces and APIs are needed to support myriad scripting and application integrations; enhanced provenance tracking is required to support the exploration of a datasets lineage; and automated methods are needed to extract and catalog crucial metadata embedded within files.

We are also working to integrate new features into NXFS, including data modification operations and more advanced remote computation. This will allow users to visualize the results of data transformations, such as background subtraction or data projections, while moving only minimal amounts of data over the network. An additional critical need is the ability to apply versioning over small changes to big datasets. Application-specific optimizations could greatly exceed the storage efficiency of general purpose version control systems.

A large part of our future work will focus on the ability to generalize our approach and apply it to a variety of different domains. Specifically, we see many opportunities to integrate these capabilities in external applications via standard interfaces. We will start by integrating these capabilities with a variety of workflows within the x-ray community at the APS, before broadening these workflows to include other facilities and applications.

REFERENCES

[1] Ian Foster, Tekin Bicer, Raj Kettimuthu, Michael Wilde, Justin Wozniak, Francesco de Carlo, Ben Blaiszik, Kyle Chard, Francesco de Carlo, and Ray Osborn. Workflows at experimental facilities: Use cases from the Advanced Photon Source. In *DOE Workshop on the Future of Scientific Workflows*, 2015.

[2] Ian Foster, Rachana Ananthakrishnan, Ben Blaiszik, Kyle Chard, Ray Osborn, Steven Tuecke, Mike Wilde, and Justin M. Wozniak. Networking materials data: Accelerating discovery at an experimental facility. In Gerhard Joubert and Lucio Grandinetti, editors, *Big Data and High Performance Computing*. In press, 2015.

[3] Takeshi Egami and Simon J. L. Billinge. *Underneath the Bragg Peaks: Structural Analysis of Complex Materials*, page 118. Elsevier, 2003.

[4] Przemek Klosowski, Mark Koennecke, Jonathan Tischler, and Raymond Osborn. NeXus: A common format for the exchange of neutron and synchrotron data. *Physica B: Condensed Matter*, 241243:151 – 153, 1997.

[5] Mike Folk, Robert E. McGrath, and Nancy Yeager. HDF: An update and future directions. In *Proc. Geoscience and Remote Sensing Symposium*, 1999.

[6] NeXpy: A Python GUI to analyze NeXus data. http://nexpy.github.io/nexpy.

[7] Laura Wolf. Boosting beamline performance, 2014. https://www.alcf.anl.gov/articles/boosting-beamline-performance.

[8] Justin M. Wozniak, Timothy G. Armstrong, Daniel S. Katz, Michael Wilde, and Ian T. Foster. Toward computational experiment management via multi-language applications. In *DOE Workshop on Software Productivity for eXtreme scale Science (SWP4XS)*, 2014.

[9] Justin M. Wozniak, Kyle Chard, Ben Blaiszik, Ray Osborn, Michael Wilde, and Ian Foster. Big data remote access interfaces for light source science. In *Proc. Big Data Computing*, 2015.

[10] APS upgrade web site. https://www1.aps.anl.gov/APS-Upgrade.

[11] Justin M. Wozniak, Hemant Sharma, Timothy G. Armstrong, Michael Wilde, Jonathan D. Almer, and Ian Foster. Big data staging with MPI-IO for interactive X-ray science. In *Proc. Big Data Computing*, 2014.