Expressways from Scientific Instrument to Supercomputer: A Prototype Demonstration

Joaquin Chung, Zhengchun Liu, Tekin Bicer, Raj Kettimuthu, Linda Winkler,

Dimitrios Katramatosˆ, Nageswara S. V. Rao*, Ian Foster

Argonne National Laboratory, USA
ˆBrookhave National Laboratory, USA
*Oak Ridge National Laboratory, USA

Abstract — Moving data from modern scientific instruments to supercomputing facilities presents many challenges. For instance, extreme-scale scientific simulations and experiments can generate much more data than that can be stored and analyzed efficiently at a single site. Moreover, data movement must be finished within tight schedules. Another example is the growing number of scientific fields that require the ability to analyze data in near real-time, so that results from one experiment can guide selection of the next–or even influence the course of a single experiment. Expressways encompasses two demonstrations. First, we build upon our SC16 demonstration about moving a Petabyte in a day, and present novel techniques for moving a Petabyte in half a day. Second, we present a reference architecture and tools to establish network circuits from scientific instrument to supercomputer that will enable streaming data analysis.

1 Moving a Petabyte in half a day

Extreme-scale scientific simulations and experiments can generate much more data than can be stored and analyzed efficiently at a single site. For example, a single trillion-particle simulation with the Hardware/Hybrid Accelerated Cosmology Code (HACC) [1] generates 20 PiB of raw data (500 snapshots, each 40 TiB), which is more than petascale systems such as the Mira system at the Argonne Leadership Computing Facility (ALCF) and the Blue Waters system at the National Center for Supercomputing Applications (NCSA) can store in their file systems.

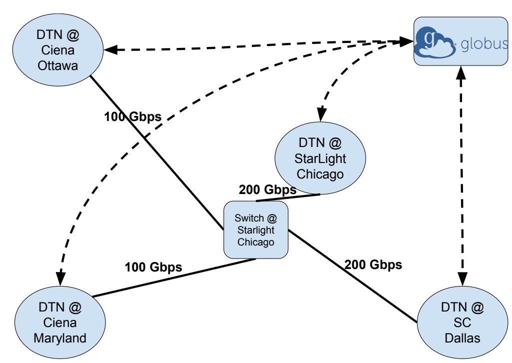

Figure 1: Moving a Petabyte in half a day. Sites involved in this demo

In 2016, at the NCSA booth of SC16 (the International Conference for High Performance Computing, Networking, Storage, and Analysis in 2016), a state-of-the-art, 29-billion-particle cosmology simulation combining high spatial and temporal resolution in a large cosmological volume was performed on Mira at ALCF. As this simulation ran, the Globus [2] transfer service was used to transmit simulation data to NCSA each of 500 temporal snapshots as it was produced. In total, this workflow moved one PiB in 24 hours from the ALCF to NCSA, achieved steady end-to-end rate of ∼93 Gb/s. Kettimuthu et al [3] presented experiences in transferring one petabyte of science data within one day.

In 2018, we will apply lessons we have learned from moving one PiB in 24 hour and developed new techniques to move one PiB data in half a day to meet the requirement of even more challenging simulation applications than the HACC we have supported in 2016. Specifically, we will show a case that we generate data (either from simulation or experiments like light source facility) at one place but the size of the data is too big to save locally, and transfer the data to another computing facility to analyze the data, then send back to reduced analyzing results. Meanwhile, the analyzed results will be sent to two other places for archival purpose. Figure 1 shows the sites involved in this demo.

2 End-to-end Network Orchestration for Streaming Analysis

A growing number of scientific fields require the ability to analyze data in near real-time, so that results from one experiment can guide selection of the next–or even influence the course of a single experiment. The experiments are often tightly scheduled, with timing driven by factors ranging from the physical processes involved in an experiment to the travel schedules of on-site researchers. Thus, the computing and network (as the computing resources may not be available locally) resources must often be available at a specific time, for a specific period. On-demand network bandwidth, though provided by backbone research and education networks such as ESnet and Internet2, is not easy to get end-to-end in an automated fashion. Even though compute resources can be obtained on-demand (at least in some institutions), those resources are not typically connected to the wide-area network. The typical model is that the data coming from the wide-area network goes into the parallel file system via the dedicated data transfer nodes, and compute nodes access the data from the parallel file system. This model does not work well for near real-time analysis of the data streams coming from an experiment or simulation.

To address the aforementioned challenges, in this demo we will present a reference architecture and tools to establish network circuits from scientific instrument to supercomputer. We will use a real-world, high-speed testbed for end-to-end software-defined networking (SDN) [4] orchestration between scientific instrument (e.g., a lightsource) and supercomputing facility (see Figure 2). We will deploy a data acquisition server generating streaming data at the instrument site, and a compute node receiving and processing this data at supercomputing site. The two nodes will be connected by a high-speed WAN connection (10 Gbps). Our end-to-end network orchestration framework, which is based on the Software-Defined Network Science Flows (SDN-SF) stack [5], will provision dedicated QoS circuits between the two endpoints programmatically. The objectives of this demo are the following:

1. To enable elasticity in network resources in a non-disruptive fashion.
2. To provision network resources for streaming science workflows end-to-end.
3. To develop network resource management strategies (best practices) to extend wide area flows to compute nodes in supercomputing facilities.

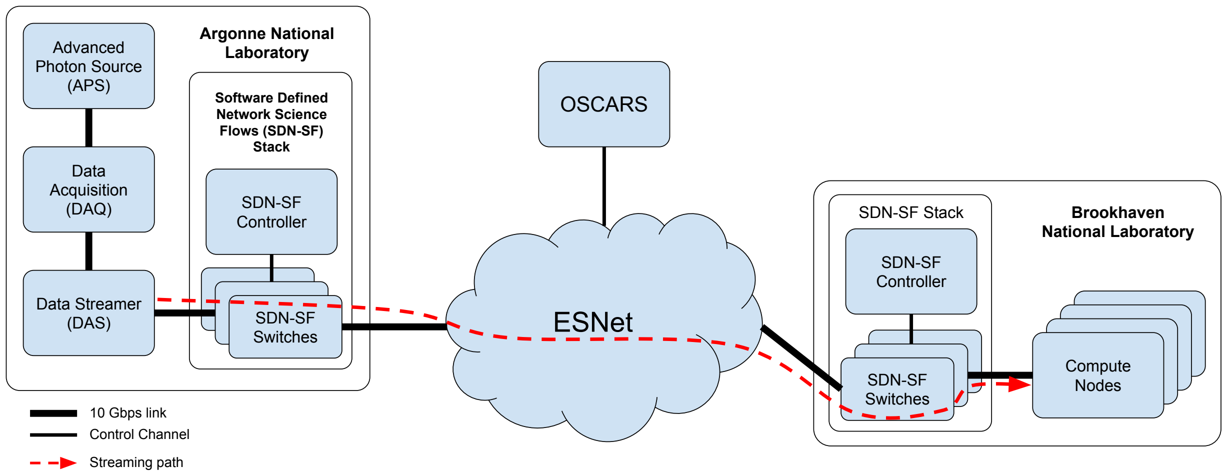

Figure 2: Expressways architecture for streaming data analysis using the SDN-SF Stack for network orchestration

3 Future Work

For our future work, we will continue developing techniques for reducing the transfer time of extreme-scale datasets. Regarding our streaming analysis use case, we will collaborate with science network operator to deploy our network orchestration framework in production. Our goal is to produce the best practices for enabling end-to-end orchestration in a non-disruptive fashion.

References

[1] S. Habib, A. Pope, H. Finkel, N. Frontiere, K. Heitmann, D. Daniel, P. Fasel, V. Morozov, G. Zagaris, T. Peterka, V. Vishwanath, Z. Lukić, S. Sehrish, and W. Liao, "HACC: Simulating sky surveys on state-of-the-art supercomputing architectures," *New Astronomy*, vol. 42, pp. 49–65, 2016.

[2] www.globus.org, *globus*, 2018 (accessed October 3, 2018). `https://www.globus.org`.

[3] R. Kettimuthu, Z. Liu, D. Wheeler, I. Foster, K. Heitmann, and F. Cappello, "Transferring a petabyte in a day," *Future Generation Computer Systems*, vol. 88, pp. 191 – 198, 2018.

[4] D. Kreutz, F. M. V. Ramos, P. E. Verssimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, "Software-defined networking: A comprehensive survey," *Proceedings of the IEEE*, vol. 103, pp. 14–76, Jan 2015.

[5] N. S. V. Rao, Q. Liu, S. Sen, R. Kettimuthu, J. Boley, B. W. Settlemyer, H. B. Chen, D. Katramatos, and D. Yu, "Software-defined network solutions for science scenarios: Performance testing framework and measurements," in *Proceedings of the 19th International Conference on Distributed Computing and Networking*, ICDCN '18, (New York, NY, USA), pp. 53:1–53:10, ACM, 2018.