Facilitating Big Data Sharing between Institutions

Dimitrios Katramatos and Dantong Yu
Computational Science Center, Brookhaven National Laboratory
{dkat, dtyu}@bnl.gov

Modern-day science programs are generating numerous extremely large complex datasets, expected to soon reach the scale of exabytes. Quantitative analysis of, and knowledge discovery from, these datasets can be accomplished most effectively through extensive collaborations between domain scientists and computational experts from within an institution and/or across multiple institutions. Currently, several key challenges impede scientists from efficiently and effectively sharing such datasets; For example, existing systems for moving scientific data are not adequately general nor scalable to the required degree; the costs of commercial solutions, such as cloud drives, are prohibitive above certain quota limits, and security might be inadequate. Furthermore, the solutions do not allow institutions to host their own data using their own resources, also their transfer speeds are limited by the public Internet. Institutional site networks also may have speed limitations due to their infrastructure and administrative policies.

We envisage a Big Data sharing environment that greatly can facilitate moving data from one place to another through a straightforward user interface, while taking full advantage of the capabilities of modern high-performance network to minimize transfer times. Key elements for realizing such an environment already are being put in place or soon will be. Examples include next-generation Wide Area Networks (WANs), such as ESnet [1] and Internet2 [2] with current speeds of 100 Gb/s, and production-deployed systems to reserve bandwidths for specific traffic, such as OSCARS [3]. Also, techniques are being installed to overcome bandwidth limitations within and through the perimeter of sites via special connectivity and/or by the establishment of Science Demilitarized Zones (SDMZs). There has been investment into developing experimental systems for reserving and scheduling bandwidth within institutional networks and across multiple network domains (including WAN); systems for co-scheduling network and storage resources; and communication protocols and tools for high-performance data transfers. We feel that a strategy towards developing and deploying a Big Data sharing environment will be highly beneficial to data users because it resolves a known need while taking advantage of existing investments. Such a framework will encompass existing technologies at the hardware and software level and focus future research towards developing more complex, enhanced capabilities that will further foster effective collaborations and scientific discoveries.

An example of such an environment would be a high-performance cloud drive for Big Data. Consider a Dropbox-style drag-and-drop service for moving and sharing files, but for files of such large sizes that would choke existing cloud drives, since the latter are meant for data volumes of a few tens to a few hundreds of megabytes. Much thought is needed on how to move a petabyte of experimental data for processing and analysis for each single dataset generated by multiple experimental cases. Such a cloud drive could be a federation of data stored at several institutions and made available through an integrating service with a user-friendly interface. Under the hood, and completely transparently to the user, this system would exploit a deployed network and storage co-scheduling infrastructure to guarantee the availability of the resources for any impending data transfer while utilizing a high-performance transfer tool with a suitable protocol to transfer the data reliably. The transfer would be scheduled to occur over special network connections created for the duration of that transfer; In this way, bandwidth bottlenecks would be avoided and the movement ultimately would conclude within a time window already known to, and agreed upon by the user.

In pursuit of the ideas presented above, we submit a few suggestions for consideration:

- There needs to be continued and increased investment in the research and development of software that exploits the principles of Software Defined Networking (SDN) for scheduling/reserving network resources. It also should take advantage of emerging technologies, such as the OpenFlow protocol, Remote Direct Memory Access (RDMA), and RDMA over Converged Ethernet (RoCE).
- Further investments should be made in the research on and development of software that ensures easy data sharing on an unprecedented scale. An example for this would be a federated cloud drive infrastructure.
- Finally, due to the highly distributed nature of the data sharing environment, there is an evident need for the formation of a body, such as an interest group and/or a steering committee that would coordinate the work of multiple institutions. Such a committee would develop best practices, rules, and regulations further to encourage and facilitate research and development as well as the deployment and testing of the necessary software infrastructure and tools needed to implement a Big Data sharing environment.

About the authors:
Drs. Katramatos and Yu have a combined 25 years of experience in big data handling as well as in research and development of high-performance network scheduling and data transfer software, first as the staff members of the RHIC-ATLAS Computing Facility (RACF) staff and subsequently as staff researchers with the Computational Science Center (CSC) at Brookhaven National Laboratory (BNL). Dr. Yu is the group leader of the computer science group at CSC and is the PI or Co-PI of several DOE ASCR-funded networking projects such as TeraPaths, StorNet, and FTP100. Dr. Katramatos is the lead developer of TeraPaths and the PI of the Virtual Network on Demand (VNoD) project. He has participated in several other related projects such as StorNet.

References:
1. Energy Sciences Network (ESnet): http://www.es.net
2. Internet2: http://www.internet2.edu
3. On Demand Secure Circuits and Advance Reservation System (OSCARS): http://www.es.net/services/oscars/