Presentation Title:

Petascale data architectures for portals and computing centers

Keywords:

Performance, Cyberinfrastructure, Science DMZ, Data Portals, Science Gateways

Presentation Abstract/Description:

Many science collaborations use data portals (also called Science Gateways) to provide discovery, search, and access/download services for the datasets used by the collaboration, or for the entire field. Astronomy, genomics, and climate science are just a few of the fields that make heavy use of data portals. Often, the datasets published in portals are large or complex enough that they require the capabilities of a supercomputer center for scientifically productive data analysis.

This presentation will describe advances made in data portal architecture and in the production deployment of high-performance data services at supercomputer centers, and the ways in which these services synergistically interact. The Science DMZ model provides an architectural framework, and the Modern Research Data Portal design pattern builds on the Science DMZ to enable high-performance data portals which are well-matched to the high performance data capabilities of today's supercomputer centers. A demonstration of these capabilities on production infrastructure will be included in the presentation.

Additional information for reviewers:

This work builds on the Science DMZ design pattern and describes functional best practices for deploying and using high-performance data services for large scale data analysis. The architectures and deployment models described in this presentation require high-performance networks in order to function correctly, and demonstrate the value of high performance networks to science collaborations. References include:
- The Science DMZ:
    - http://www.es.net/assets/pubs_presos/sc13sciDMZ-final.pdf
    - http://fasterdata.es.net/science-dmz/
- The Modern Research Data Portal:
    - https://peerj.com/preprints/3194/

Author:

Eli Dart, ESnet, Lawrence Berkeley National Laboratory, USA

Presenter:

Eli Dart, ESnet, Lawrence Berkeley National Laboratory, USA

Biography:

Eli Dart is a network engineer in the ESnet Science Engagement Group, which seeks to use advanced networking to improve scientific productivity and science outcomes for the DOE science facilities, their users, and their collaborators. Eli is a primary advocate for the Science DMZ design pattern and works with facilities, laboratories, universities, science collaborations, and science programs to deploy data-intensive science infrastructure based on the Science DMZ model. Eli also runs the ESnet network requirements program, which collects, synthesizes, and aggregates the networking needs of the science programs ESnet serves.

Eli has over 15 years of experience in network architecture, design, engineering, performance, and security in scientific and research environments. His primary professional interests are high-performance architectures and effective operational models for networks that support scientific missions, and building collaborations to bring about the effective use of high-performance networks by science projects.