Leveraging XSEDE HPC resources to address computational challenges with high-resolution topography data

Choonhan Youn[1], Viswanath Nandigam[1], Minh Phan[1], David Tarboton[2], Nancy Wilkins-Diehr[1], Chaitan Baru[1], Christopher Crosby[3], Anand Padmanabhan[4], Shaowen Wang[4]

[1]San Diego Supercomputing Center (SDSC), University of California at San Diego

[2]Utah State University, [3]UNAVCO, [4]University of Illinois at Urbana-Champaign

{cyoun, viswanat, mnphan}@sdsc.edu, dtarb@usu.edu, {wilkinsn, baru}@sdsc.edu, crosby@unavco.org, {apadmana, shaowen}@illinois.edu

ABSTRACT

Leveraging service-oriented architectures and taking advantage of the high-performance compute resources provided by XSEDE, we have developed standards-based web services to address the challenges associated with processing large volumes of high resolution topography data. These web services make results from community software packages and other cyberinfrastructure-based applications available to the wider earth sciences community via the OpenTopography Facility and the CyberGIS Gateway.

Categories and Subject Descriptors

C.2.4 [**Distributed Systems**]: Distributed Applications

General Terms

Design, Management, Standardization

Keywords

OpenTopography, CyberGIS, web service, OGC WPS, TauDEM

1. EXTENDED ABSTRACT

The availability of high-resolution topography data, especially from LiDAR (Light Distance And Ranging) technology, has been revolutionary for earth and ecological sciences, environmental, and engineering applications. These data are one of the most powerful tools available for studying the earth's surface in a very high detail over large spatial extents. The NSF funded OpenTopography Facility (www.opentopography.org) provides community access to earth science oriented high-resolution topography data and related tools. OpenTopography has developed a cyberinfrastructure platform built using service oriented architecture (SOA) to enable efficient online access to these datasets along with web based processing tools to generate derived products including digital elevation models (DEMs). All data access functions and processing algorithms are packaged as web services deployed across multiple OpenTopography commodity clusters leveraging the open source OPAL toolkit (sourceforge.net/projects/opaltoolkit/). The service-oriented architecture of OpenTopography supports a distributed environment and service interoperability with its loose coupling using standards based communication protocols. The topography datasets are most useful to other communities when they can be easily integrated with processing capabilities for the derivation of higher-level analysis products. TauDEM (hydrology.usu.edu/taudem) is terrain analysis software for hydrologic analysis of DEMs that is used extensively by the hydrologic research community. TauDEM products [1] enrich the information content of gridded DEMs that are derived products available through OpenTopography. This paper describes the process in which the high performance XSEDE resources were leveraged to integrate the TauDEM software into the OpenTopography system for processing large amounts of topography data. The processing of these large amounts of high-resolution topography datasets would be extremely inefficient on commodity clusters. Integration of the TauDEM software into OpenTopography will enable wide spread utilization by the larger science community that are interested in hydrologic analysis and would like to leverage OpenTopography's large data collections.

CyberGIS represents a new generation of Geographic Information Systems (GIS) based on seamless synthesis of cyberinfrastructure, geographic information science, and spatial analysis and modeling. The NSF funded CyberGIS project (www.cybergis.org) advances the science of CyberGIS, with a particular focus on enabling the analysis of big spatial datasets, computationally intensive spatial analysis and modeling, and collaborative geospatial problem solving and decision-making. The CyberGIS Gateway provides an online, high-performance and collaborative geospatial problem-solving environment to allow for the contribution, sharing of and access to CyberGIS services and tools by a broad community of scientists to advance sciences enabled by CyberGIS. Service interoperability with OpenTopography making its data and processing algorithms available to the CyberGIS community would be mutually beneficial [2].

Current OpenTopography LiDAR point cloud data holdings alone exceed 18 terabytes (in compressed binary format) covering approximately 169,467 km^2. Users interact with the portal interface to perform sub-selections of the datasets and run processing algorithms on their area of interest to produce derivative products like DEMs, hillshade, visualizations, and morphometric products such as slope.

To support large jobs submitted both via OpenTopography and the CyberGIS Gateway, which integrates OpenTopography data

XSEDE '14, Jul 13-18 2014, Atlanta, GA, USA
ACM 978-1-4503-2893-7/14/07.
http://dx.doi.org/10.1145/2616498.2616564

and services, we applied for and were granted an XSEDE allocation on their Gordon HPC resource (www.sdsc.edu/supercomputing/gordon/). The XSEDE Gordon resource is a flash-based supercomputer at the San Diego Supercomputer Center designed for data-intensive workloads. We were able to deploy the complete OpenTopography processing services stack on a dedicated Gordon I/O node and compute nodes that were allocated to this project (See Figure 1). We worked jointly with the Gordon team to install the OpenTopography software stack including the Points2Grid (opentopography.org/index.php/Tools/otforge/points2grid) service that runs a local gridding algorithm for generating DEMs, TauDEM suite that process these DEMs, and a custom job scheduler based on PBS to manage the workloads across the Gordon cluster. All of the current OpenTopography datasets were replicated on Data Oasis, a Lustre based file system in a compressed LAS format, the standard LiDAR data format. Deegree (www.deegree.org), a geospatial web services toolkit, was also deployed to support the Open Geospatial Consortium (OGC) standard service platform widely used in CyberGIS community. In addition, we applied for and were granted an XSEDE Extended Collaborative Support Service (ECSS) allocation in order to optimize the TauDEM software suite for processing of large DEM data on XSEDE computing resources. This extends the recently developed parallel TauDEM capability [2] to the XSEDE environment. The highly optimized TauDEM tools can then be integrated into OpenTopography and CyberGIS for more efficient processing. The TauDEM XSEDE implementation enables the computation of TauDEM products from large DEMs such as those held in OpenTopography. The massive volumes of LiDAR data available via OpenTopography are a huge incentive for this project.

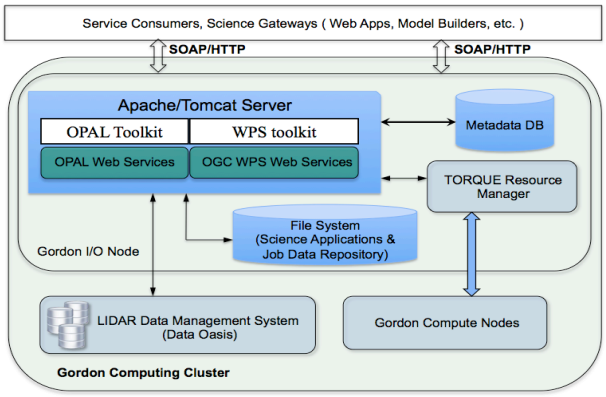

Figure 1 System configuration on Gordon Cluster

Our XSEDE allocation includes 16 Gordon compute nodes as well as the dedicated Gordon I/O node. Since Gordon's primary PBS job scheduler does not manage these compute nodes, when a job request comes through, its submission and management are performed through the web services deployed on a Tomcat server. But some of these web services run a job script that needs a scheduler. We have configured and installed our own local TORQUE resource manager for managing the allocated compute nodes. To speed up the computation, we also configured the compute nodes to access the local SSD scratch directories on I/O node.

We have integrated key TauDEM functions into the standard OpenTopography point cloud processing workflow to generate hydrologic data products (See Figure 2). TauDEM products are derived for any OpenTopography dataset and any subset of the products and made available for download in the OpenTopography portal.

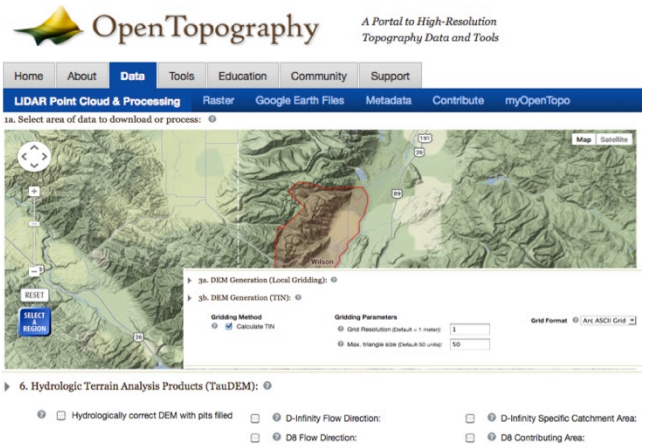

Figure 2 Illustration of TauDEM services within the OpenTopography web interface.

As a result of this work OpenTopography now provides users with the capability to obtain TauDEM hydrologic terrain analysis products for any DEM produced from OpenTopography LiDAR data holdings. This enhances the value and use of these data in hydrologic research by providing access to these computational products without users having to learn TauDEM or run HPC. The OpenTopography and CyberGIS communities are benefiting with the availability of XSEDE Gordon resources by going beyond the resource restrictions of commodity clusters, allowing larger jobs, and by utilizing the larger compute capability to deploy and use TauDEM like services.

ACKNOWLEDGMENTS

This work is supported in part by the National Science Foundation under Grant Number 1047916 for CyberGIS. Computational experiments used the Extreme Science and Engineering Discovery Environment (XSEDE) (resource allocation Award Number TG-EAR130018). OpenTopography is supported by the National Science Foundation under NSF awards 1226353 & 1225810. TauDEM development was supported by the US Army Research and Development Center under contract number W912HZ-11-P-0338 and the U.S. National Science Foundation under grant EPS 1135482

REFERENCES

[1] Tesfa, T.K., Tarboton, D.G., Watson, D.W., Schreuders, K.A.T., Baker, M.E., Wallace, R.M., 2011. Extraction of hydrological proximity measures from DEMs using parallel processing. Environmental Modelling & Software 26(12) 1696-1709, http://dx.doi.org/10.1016/j.envsoft.2011.07.018.

[2] Padmanabhan, A., Youn, C., Hwang, M., Liu, Y., Wang, S., Wilkins-Diehr, N., Crosby, C. 2013. Integration of Science Gateways: A Case Study with CyberGIS and OpenTopography, In Proceeding of *XSEDE '13 Proceedings of the Conference on Extreme Science and Engineering Discovery Environment: Gateway to Discovery*, ACM, 28p. DOI=10.1145/2484762.2484808