High-resolution National Elevation Dataset: Opportunities and Challenges for High-Performance Spatial Analytics

Yan Liu, Michael P. Finn, Babak Behzad, Eric Shook

University of Illinois
Urbana, IL 61801

yanliu@illinois.edu

The National Elevation Dataset (NED) produced by the U.S. Geological Survey (USGS) provides publicly accessible high-resolution elevation datasets for the U.S. coverage. As new technologies are adopted to produce higher resolution geospatial datasets, e.g., LiDAR-derived Digital Elevation Model (DEM), using these data for spatial analytics represents tremendous opportunities for researchers and GIS practitioners. However, even for high-performance spatial analysis and modeling methods that can efficiently leverage massive computing power, data access and associated performance issues become crucial challenges as high-resolution geospatial data of hundreds of gigabytes or even several terabytes are used. Two high-performance spatial analysis methods that use 1/3 arcsecond NED as input were investigated to identify the big data challenges. Specifically, data downloading and visualization issues and the impact of data transformation, transfer performance, and input/output (I/O) on the computational performance of spatial analysis methods were identified. A set of tools and services were thus developed and established to tackle aforementioned challenges using high-performance approaches.

Downloading services designed for very large datasets are more complex than direct file downloading. A downloading request has to be translated into a sequence of steps for data location, extraction, caching, and cleanup. The NED downloading service exposes each processing step as a service interface, making downloading a complex process for end users. We developed a NED downloading tool to hide the details of each step and present an easy-to-use interface for end users. A Web Mapping Service (WMS) that automates the pyramid operation of creating tiles at multiple zoom levels was also developed to enable the visualization of the half terabyte 1/3 arcsecond NED.

In integrating NED as a data source for high-performance spatial analytics, we established a parallel data transfer service between USGS and the Extreme Science and Engineering Discovery Environment (XSEDE), a leading cyberinfrastructure in the US, to greatly reduce the data transfer time by fully leveraging high network bandwidth. Through performance profiling, we found that necessary data processing functions such as clipping, transformation, and format translation could become serious bottleneck in computation process. For example, the high-performance viewshed analysis we developed can process 400MB DEM in 30 seconds, but data pre- and post-processing using the Geospatial Data Abstraction Library (GDAL) could take 5 minutes. A parallel I/O and data processing library is thus being developed to handle I/O-intensive operations efficiently and will be integrated in all of our spatial analysis methods.

keywords: big data, ned, dem, high performance, parallel computing, cyberinfrastructure

This abstract is for the Special Session on Big Data.