Establishing TauDEM as a Science Gateway Service on XSEDE for Scalable Hydrological Terrain Analysis

Yan Liu¹, Ye Fan¹, Shaowen Wang¹, David Tarboton², Ahmet Yildirim², Nancy Wilkins-Diehr³

¹ University of Illinois at Urbana-Champaign
² Utah State University
³ San Diego Supercomputer

XSEDE ECSS Symposium
September 16, 2014
• Introduction
  o TauDEM software
  o ECSS collaboration

• Performance Analysis and Improvement
  o Strategies
  o Findings & results

• CyberGIS-TauDEM Gateway Application Development
  o CyberGIS Gateway
  o GIResolve Open Service API
  o XSEDE-enabled execution
Scalable DEM-based Hydrological Information Analysis

• Digital Elevation Models (DEM)
  o Geospatial topographic data
  o Raster and vector representation

• DEM-based Hydrological Information Analysis
  o Use of topographic information in hydrological analysis and modeling
  o Examples
    • Derivation of flow directions, contributing area, stream network…

• Impact of High Resolution DEM Data
  o High resolution DEM data sources
    • National Elevation Dataset (NED) from the U.S. Geological Survey (USGS)
      o 10-meter resolution: 330GB raw data
      o 1-meter resolution: 4-5 PB raw data
    • OpenTopography Lidar-derived DEM data
  o Improved accuracy and reliability of analysis and modeling results
  o Revealing insights that were not possible to obtain before
Example: USGS NED 10m Resolution
Hydrological Information Extraction using DEM
TauDEM

• TauDEM - A Parallel Computing Solution to DEM-based Terrain Analysis
  o Open source software
  o A suite of DEM tools for the extraction and analysis of hydrologic information from topographic data
  o A growing user community

• Parallel Computing in TauDEM
  o Parallel programming model: Message Passing Interface (MPI)
  o Spatial data decomposition
    • Each process reads a sub-region for processing
    • MPI communication for exchanging runtime hydrological information
    • Each process writes a sub-region defined by output data decomposition
  o Parallel input/output (IO)
    • In-house GeoTIFF library (no support for big GeoTIFF)
    • MPI IO for DEM read and write
TauDEM Channel Network and Watershed Delineation Software (desktop)

- Stream and watershed delineation
- Multiple flow direction flow field
- Calculation of flow-based derivative surfaces

http://hydrology.usu.edu/taudem/
ECSS Goals

• Enhance TauDEM for large-scale terrain analysis on massive computing resources provided on national cyberinfrastructure through rigorous computational performance profiling and analysis

• Establishing TauDEM as a science gateway service empowered by cyberinfrastructure (e.g., XSEDE)
TauDEM-CyberGIS-OpenTopography Collaboration

TauDEM 5.0

Scalability Enhancement (XSEDE ECSS)

TauDEM 5.x

OpenTopography

Lidar-derived DEMs

OT TauDEM Services

CyberGIS

CyberGIS-TauDEM App

TauDEM-enabled Research

DEMs
USGS NED
OT
User DEMs
Collaboration Team

• National cyberinfrastructure
  o Extreme Science and Engineering Discovery Environment (XSEDE)
  o XSEDE Extended Collaborative Support Services (ECSS) provides computational science expertise
    • Ye Fan, Yan Liu, Shaowen Wang, National Center for Supercomputing Applications (NCSA)

• NSF OpenTopography LiDAR data facility
  o DEM generation services for LiDAR-derived TauDEM analysis
  o Integration of TauDEM in OpenTopography service environment
  o People
    • Chaitan Baru, Nancy Wilkins-Diehr, Choonhan Yeon, San Diego Supercomputer Center (SDSC)

• NSF CyberGIS project
  o Integration of TauDEM in CyberGIS Gateway
  o Integration of TauDEM in advanced CyberGIS analytical services (workflow)
  o People
    • University of Illinois at Urbana-Champaign (UIUC)
      o Yan Liu, Anand Padmanabhan, Shaowen Wang
    • San Diego Supercomputer Center (SDSC)
      o Nancy Wilkins-Diehr, Choonhan Yeon

• TauDEM Developers
  o Utah State University Civil and Environmental Engineering
    • David Tarboton
  o Utah State University Computer Science
    • Ahmet Yildirim, Dan Watson
Performance Analysis and Improvement
 TauDEM Build and Run (Trestles) 

- Github link: [https://github.com/dtarb/TauDEM](https://github.com/dtarb/TauDEM)
- Code written in C++
- Prepared makefile template
- mvapich2_ib/1.5.1p1

```bash
mpirun_rsh -np 64 -hostfile $PBS_NODEFILE /home/yefan/TauDEMgit/TauDEM/pitremove -z YellowMF -fel YellowMFFel
mpirun_rsh -np 64 -hostfile $PBS_NODEFILE /home/yefan/TauDEMgit/TauDEM/d8flowdir -fel YellowMFFel -p YellowMFp -sd8 YellowMFSd8
mpirun_rsh -np 64 -hostfile $PBS_NODEFILE /home/yefan/TauDEMgit/TauDEM/dinflowdir -fel YellowMFFel -ang YellowMFFang -slp YellowMFSlp
mpirun_rsh -np 64 -hostfile $PBS_NODEFILE /home/yefan/TauDEMgit/TauDEM/areadinfl -ang YellowMFFang -sca YellowMFFsca
mpirun_rsh -np 64 -hostfile $PBS_NODEFILE /home/yefan/TauDEMgit/TauDEM/peukerdouglas -fel YellowMFFel -ss YellowMFSs
mpirun_rsh -np 64 -hostfile $PBS_NODEFILE /home/yefan/TauDEMgit/TauDEM/aread8 -p YellowMFp -ad8 YellowMFFss -w YellowMFss
mpirun_rsh -np 64 -hostfile $PBS_NODEFILE /home/yefan/TauDEMgit/TauDEM/aread8 -p YellowMFp -ad8 YellowMFAad8
mpirun_rsh -np 64 -hostfile $PBS_NODEFILE /home/yefan/TauDEMgit/TauDEM/threshold -ssa YellowMFFssa -src YellowMFSrc -thresh 300
mpirun_rsh -np 64 -hostfile $PBS_NODEFILE /home/yefan/TauDEMgit/TauDEM/streamnet -fel YellowMFFel -p YellowMFp -ad8 YellowMFAad8 -src YellowMFSrc -ord YellowMFFord -tree YellowMFTree.dat -coord YellowMFcoord.dat -net YellowMFnet.shp -w YellowMFW
```
TauDEM Profiling Summary

- I/O profiling
- MPI communication profiling
- Code tuning
- Algorithmic analysis
Test DEM Datasets

- Logan (7M)
- Yellowstone (2G)
- Chesapeake (9G)
- USGS NED (36G)
Multi-File Input Model

- Support multiple input files which form the entire domain
  - Red rectangles in the graph
- Spatial data decomposition
  - Horizontally partitioned into stripes (blue rectangles in the graph)
- Constraint on input
  - The size of single file must be less than 4GB
  - Maximum single GeoTIFF file 32000 X 32000
Multi-File Output Model

• For illustration a total of $n=5$ processors are employed to run the code (in general $n>>5$)

• The entire domain is partitioned into $n$ horizontal stripes indicated.

• Multi-output option “-mf”
  
  o With “-mf 3 2” option running TauDEM, each stripe being output as a tiling of 3 columns and 2 rows of files
Previous TauDEM Status

• Scalability issues
  o PitRemove on 2GB DEM
    • 681 seconds on an 8-core desktop
    • 3,758 seconds on a 64-core cluster

• Computation challenges
  o Scaling to large-scale analysis using massive computing resources is difficult
  o Cyberinfrastructure-based computational analysis needs in-depth knowledge and expertise on computational performance profiling and analysis
I/O Profiling with Darshan

• Large amount of small read operation
• All I/O operation in non-collective fashion
PitRemove on Trestles Using the Yellowstone Dataset (Lustre)
PitRemove on Trestles using the Yellowstone Dataset (SSD)
TauDEM Code Review

• TiffIO costing a lot of time
  o Reason: Every process opens every tiff file to retrieve header information in order to identify the working spatial domain
  o Solution: Let one process gather this information and broadcast to other processes
Improved PitRemove on Trestles Using the Yellowstone Dataset (Lustre)
Improved D8FlowDir on Trestles Using the Yellowstone Dataset (Lustre)
Large DEM Data Run on Stampede

- USGS NED Dataset (36G)
- Stampede does not allow consecutive “ibrun” in a single job
- Run workflow with job dependency
  - Under development
Perfsuite Profiling on Computation and Communication

- Build code with “-g” option
- Monitor “PAPI_TOT_CYC” event and check program stack every 10000000 cycles
- Find out where the hot spot is and the function taking most of the time
MPI Communication Profiling Result on Stampede with USGS NED Dataset
Intel VTune Profiling on Local Desktop (D8FlowDir)

• `resolveflats()` function taking most of the time
  o Data synchronization on the boundary taking too much time
  • Not every iteration needs data synchronization
  o Processes are waiting on each other to get into the next iteration
  o Load imbalance make the performance even worse
Algorithm Design Analysis

- Inevitable data synchronization on the boundary
- Too many iterations → costly synchronization
- Algorithmic improvement
  - A research topic
  - Develop new parallel algorithm
    - Reduce the number of iterations
TauDEM Code Tuning Effort

• Make frequent called routines inline functions
• Remove un-necessary inheritance to enable compile time static dispatch on critical functions
• Combine MPI_Send() immediately followed by MPI_Recv() to MPI_Sendrecv()
• Fix potential memory leak causing by unmatched deallocations
CyberGIS-TauDEM Gateway Application Development

- Streamlined TauDEM Analysis in CyberGIS Gateway
  - Web environment for sharing
  - Transparent integration of DEM data sources
  - Customized TauDEM analysis workflow
  - Online visualization

- Status
  - Beta release in August 2014
  - Developing an enhanced version for online course teaching
    - To be released end of Sept. 2014
CyberGIS Application Integration Framework

GISolve Middleware

CyberGIS Gateway

- Job Panel: Data Selection
- Analysis Input Panels: Geo-Input Editing
- Workflow: Mapping
- Visualization: Sharing

CyberGIS Toolkit

- Data Storage
- Computing Environment
- Job Wrappers
- Data Retrieval
- Geo Data Processing
- Execution Setup
- Parallel Computing
- Post-processing
- Geo-visualization
CyberGIS Gateway Objectives

• Empower online high-performance and collaborative geospatial problem solving environment
• Enable easy access to cyberGIS analytics and data sources
• Seamless integration of advanced cyberinfrastructure, GIS, and spatial analysis and modeling
  o Provide Web-based user interactions for handling huge volumes of data, complex analysis and visualization without exposing complexities in cyberinfrastructure
• Gain fundamental understanding of scalable and sustainable geospatial software ecosystems through an online integration platform
Important Characteristics

• Usability
  o Highly interactive online user interface
  o Transparent access to backend service infrastructure and cyberinfrastructure

• Scalability
  o Number of users
  o Amount of CI resources
  o Number of gateway applications

• Interoperability
  o User environment
  o Services
  o Cyberinfrastructure

• Sustainability
  o Consistent interfaces
  o Based on open service APIs
Design Approach

• Rich-client Web portal
  o Geographic Information Science (GIS) applications require highly interactive user interfaces
  o HTML5-based Web technology advances make Web interface usability comparable to desktop GUI

• Modular user interface component development
  o Customizable portal components
  o Modular gateway application packages and reusable components

• Open API approach for service environment interactions
  o Open service framework
  o Standard-based implementation for service interoperability

• Scalable application integration
  o Streamlined service integration
  o Agile user interface development

• Transparent access to cyberinfrastructure
  o Hide cyberinfrastructure details from users
Gateway Portal

http://gateway.cybergis.org/
Technologies

• Portal
  o Rich client: Ext JS (Web) or Sencha Touch (Mobile)
    • A rich set of ready-to-use interface components
  o Server: PHP + Open Service API
  o Client-server communication
    • AJAX + JSON

• Application development
  o Client-side model-view-controller (MVC)

• Mapping
  o Rich client: OpenLayers
    • Customized to work with Ext JS 4+
  o Backend services
    • Base maps
      o Google, Open Street Map, Bing, ArcGIS
    • GeoServer for large data and raster visualization

• Service infrastructure
  o GISolve middleware
Application User Interface Development

- MVC code structure
- Mashup library creation
  - Dependent libraries
    - Web vs. Mobile
  - Code compilation (compression)
  - Packaging
- Mashup API documentation
  - JSDuck-based Ext JS/Sencha Touch documentation
- Integration with portal
  - Template-based configuration
    - Entry points (menu, app panel)
    - Portal environment setting
      - user information, token, data servers
    - Coherent layout and Web element placement

- app.js
- app/
  - view/
    - map.js
  - model/
    - mapConfig.js
  - store/
    - maps.js
  - controller/
    - mapControl.js
- css/
- resources/
- server-scripts/
  - getMap.php
Reusable Components

• Portal
  o Layout
  o Menu
  o Header
  o Footer

• User interface components
  o Map panel
  o Map layer loading utilities
  o Legend configuration
  o Visualization panel
  o Analysis list
  o Data selection
  o Upload panel
  o Etc.
GI Solve Open Service API
GISolve Open Service API Capabilities

• **Usability**
  o Easy to use: Web-based
  o Multiple programming languages support
    • PHP, Java, Python, Perl, JavaScript
  o Documentation and tutorial resources

• **Standard programming interface**
  o Standard REST Web service interface
  o Provide a programming way to integrate gateway applications
  o Consistent interface to computation management

• **Scalable software integration**
  o Easy-to-use for cyberGIS application/software contributors within or outside of the project

• **Interoperability with Gateway user environment**
  o Coupled with Gateway identify management service
  o API for Gateway application development for accessing underlying service infrastructure
TauDEM App: Data Integration

• Multiple High Resolution DEM Sources
  o USGS NED (10-meter)
    • Hosted at UIUC
    • Map preview
  o OpenTopography LiDAR-derived DEMs
    • Web service API

• Data Retrieval
  o USGS NED: wget
  o OT: Dynamic DEM generation and downloading
  o Data caching
    • XWFS?

• Data Processing
  o Study area clipping
  o Multi-file generation
  o Reprojection
  o GDAL library (http://gdal.org)
  o High-performance map reprojection
    • Collaborative work with USGS
TauDEM App: Analysis Workflow

• Approach
  o 26 TauDEM functions
  o Template-based customization of TauDEM functions
    • Pre-defined dependency
    • Dynamic workflow construction in Gateway
    • Data format: JSON

• Implementation
  o Interactive workflow configuration
    • Ext JS + SigmaJS

• Execution
  o Runtime command sequence generation
    • On Trestles: command sequence
    • On Stampede: a set of jobs linked based on job dependency
TauDEM App: Visualization

• Visualization Computation
  o Reprojection
  o Pyramid generation for multiple zoom levels
  o Coloring (symbology)

• Online Visualization
  o Each product is a map layer accessible through the OGC-standard Web Mapping Service (WMS)
DEMO
Concluding Discussions

• Multidisciplinary collaboration is a key to the success so far
• Great potential for further performance improvement
• Performance profiling and analysis at large scale is critical
• Guidelines for future software research and development
  o Explicit computational thinking in software development lifecycle (design, coding, testing)
  o Performance analysis remains challenging.
  o Collaboration with computational scientists and conducting performance profiling on cyberinfrastructure are important
  o Cyberinfrastructure provides a set of abundant and diverse computational platforms for identifying computational bottlenecks and scaling code performance
• CyberGIS-TauDEM Gateway application significantly lowers the barrier of conducting large-scale TauDEM analyses by community users
  o Large-scale TauDEM analysis online
  o Software as a service
  o EOT
Acknowledgements

• This material is based in part upon work supported by NSF under Grant Numbers 0846655 and 1047916.

• TauDEM team work is supported by the US Army Research and Development Center contract No. W912HZ-11-P-0338 and the NSF Grant Number 1135482.

• Computational experiments used the Extreme Science and Engineering Discovery Environment (XSEDE) (resource allocation Award Number SES090019), which is supported by NSF under Grant Number OCI-1053575.

• The authors acknowledge the San Diego Supercomputer Center for providing the Trestles supercomputer for our research and development work in this research.

• The authors acknowledge the Texas Advanced Computing Center (TACC) at the University of Texas at Austin for providing HPC resources that have contributed to the research results reported.
Thanks!