Accelerating TauDEM as a Scalable Hydrological Terrain Analysis Service on XSEDE

Ye Fan[1], **Yan Liu**[1], Shaowen Wang[1], David Tarboton[2], Ahmet Yildirim[2], Nancy Wilkins-Diehr[3]

[1] University of Illinois at Urbana-Champaign

[2] Utah State University

[3] San Diego Supercomputer

XSEDE'14

Atlanta, GA, July 15, 2014

Outline

- Introduction
  - TauDEM software
  - Parallelism
  - ECSS work plan
- Computational Intensity Analysis and Performance Improvement
  - Strategies
  - Findings & results
- TauDEM Gateway Application
  - Data integration
  - Workflow construction
  - XSEDE-enabled execution

Scalable DEM-based Hydrological Information Analysis

- Digital Elevation Models (DEM)
  - Geospatial topographic data
  - Raster and vector representation
- DEM-based Hydrological Information Analysis
  - Use of topographic information in hydrological analysis and modeling
  - Examples
    - Derivation of flow directions, contributing area, stream network…
- Impact of High Resolution DEM Data
  - High resolution DEM data sources
    - National Elevation Dataset (NED) from the U.S. Geological Survey (USGS)
      - 10-meter resolution: 330GB raw data
      - 1-meter resolution: 4-5 PB raw data
    - OpenTopography Lidar-derived DEM data
  - Improved accuracy and reliability of analysis and modeling results
  - Revealing insights that were not possible to obtain before

Example: USGS NED

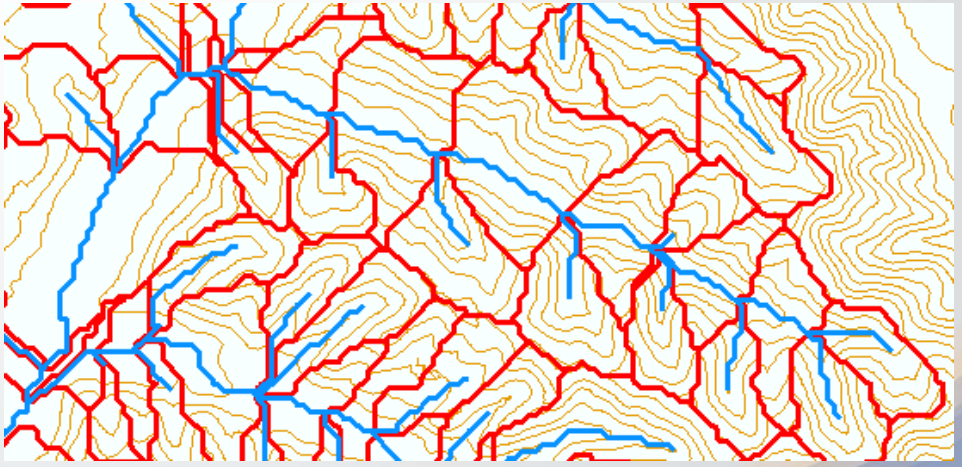
TauDEM

- TauDEM - A Parallel Computing Solution to DEM-based Terrain Analysis
  - Open source software
  - A suite of DEM tools for the extraction and analysis of hydrologic information from topographic data
  - A growing user community
- Parallel Computing in TauDEM
  - Parallel programming model: Message Passing Interface (MPI)
  - 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
  - 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

- TauDEM Tools
  - Basic Grid Analysis
    - D-Infinity Contributing Area
    - D-Infinity Flow Directions
    - D8 Contributing Area
    - D8 Flow Directions
    - Grid Network
    - Pit Remove
  - Specialized Grid Analysis
    - D-Infinity Avalanche Runout
    - D-Infinity Concentration Limited Accumulation
    - D-Infinity Decaying Accumulation
    - D-Infinity Distance Down
    - D-Infinity Distance Up
    - D-Infinity Reverse Accumulation
    - D-Infinity Transport Limited Accumulation
    - D-Infinity Upslope Dependence
    - D8 Distance to Streams
    - Slope Average Down
    - Slope Over Area Ratio
  - Stream Network Analysis
    - D8 Extreme Upslope Value
    - Length Area Stream Source
    - Move Outlets to Streams
    - Peuker Douglas
    - Peuker Douglas Stream Definition
    - Slope Area Combination
    - Slope Area Stream Definition
    - Stream Definition By Threshold
    - Stream Definition with Drop Analysis
    - Stream Drop Analysis
    - Stream Reach and Watershed
    - Watershed Grid to Shapefile

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



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

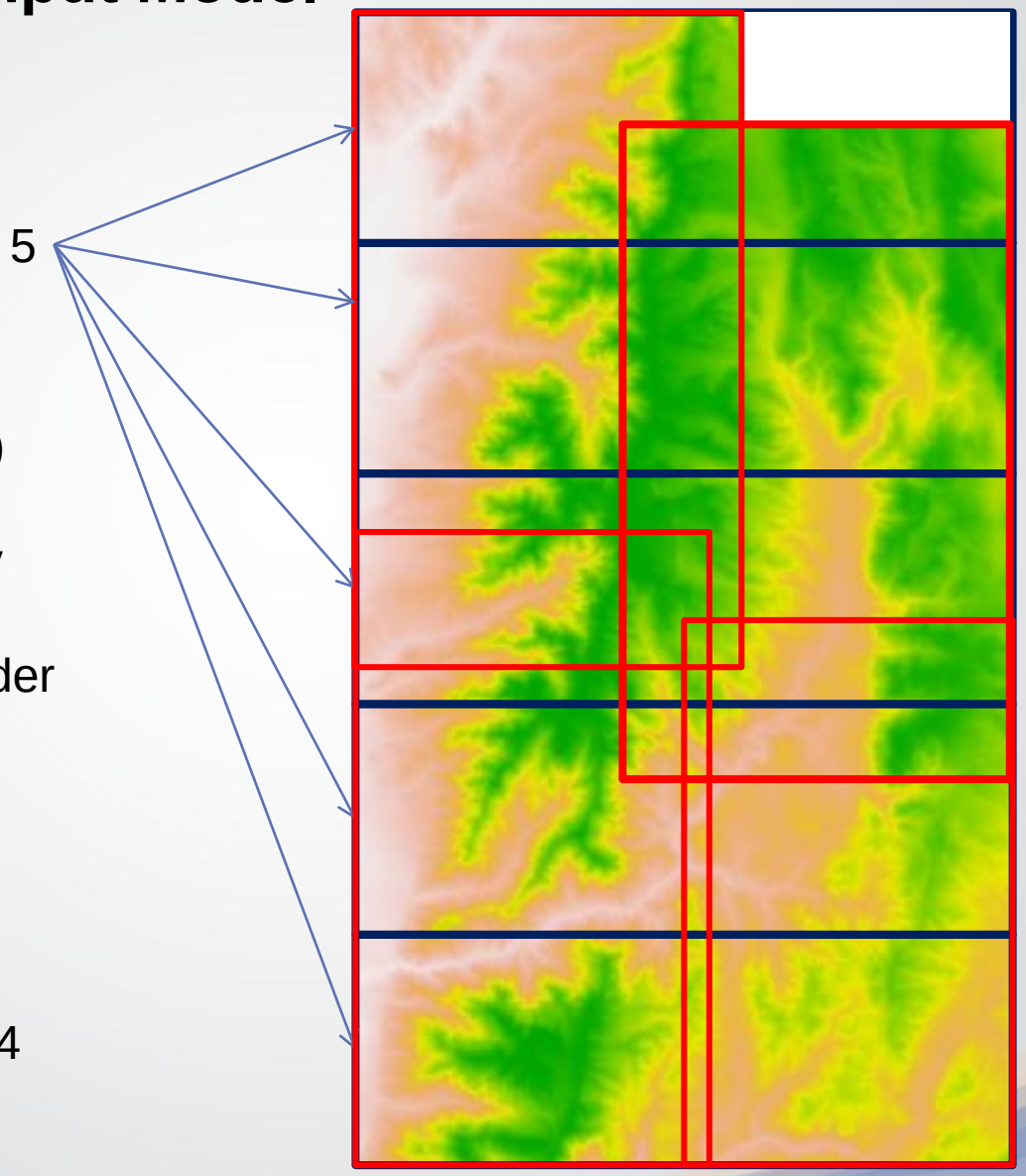
Multi-File Input Model

Number of processes **mpiexec –n 5 pitremove ...** results in the domain being partitioned into 5 horizontal stripes

On input files (red rectangles) data coverage may be arbitrarily positioned and may overlap or not fill domain completely. All files in the folder are taken to comprise the domain.

Only limit is that no one file is larger than 4 GB.

Maximum GeoTIFF file size: 4 GB = about 32000 x 32000 rows and columns

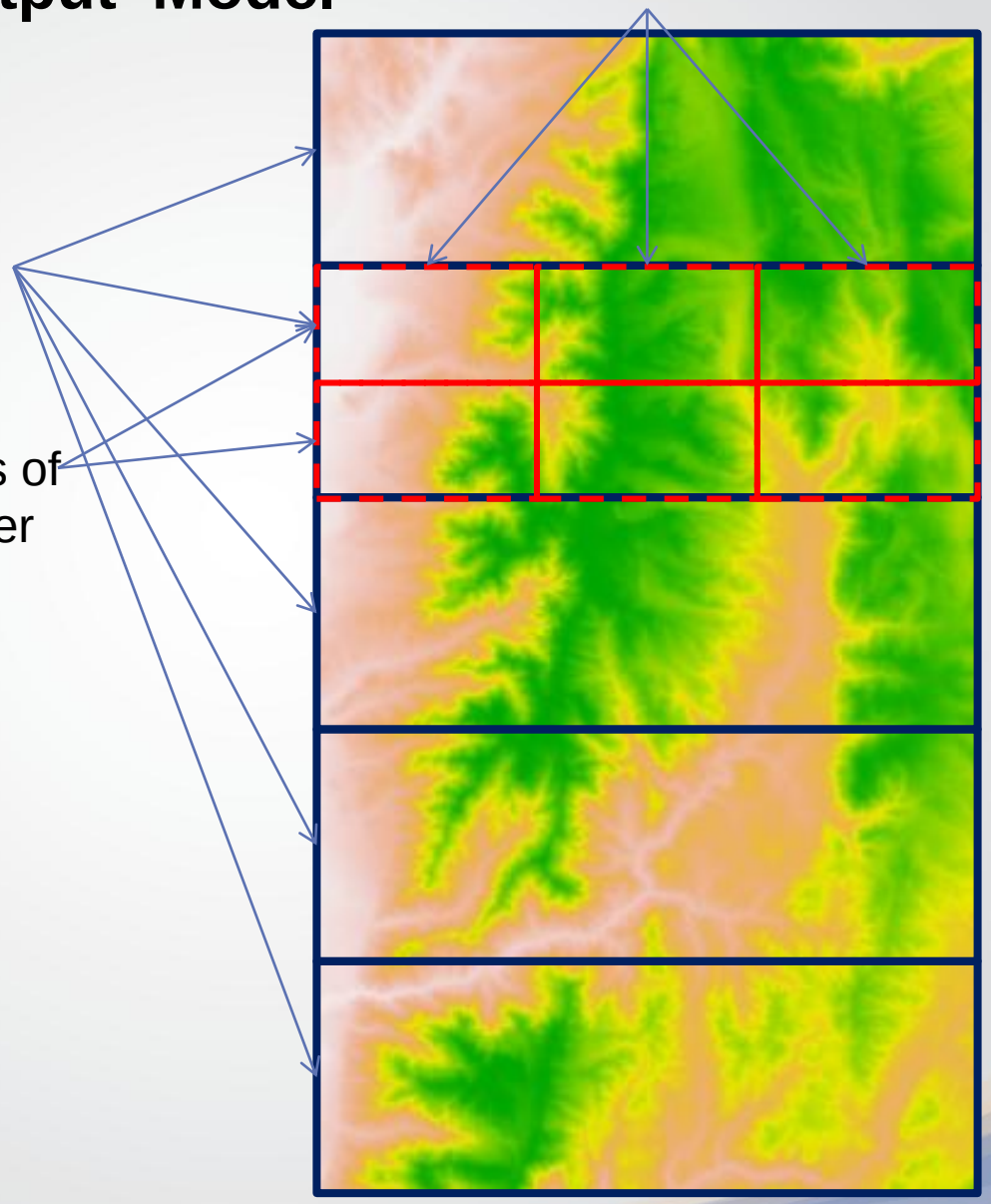


Multi-File Output Model

Number of processes
mpiexec –n 5 pitremove ...
results in the domain being partitioned into 5 horizontal stripes

Multifile option
-mf 3 2
results in each stripe being output as a tiling of 3 columns and 2 rows of files

Maximum GeoTIFF file size: 4 GB = about 32000 x 32000 rows and columns



Computational Challenges

- Scalability issues
  - PitRemove step on 2GB DEM
    - 681 seconds on an 8-core PC
    - 3,759 seconds on a 64-core cluster
    - Not acceptable on XSEDE resources
- Computational challenges
  - Scaling to large-scale analysis using massive computing resources is difficult
  - Cyberinfrastructure-based computational analysis needs in-depth knowledge and expertise on computational performance profiling and analysis

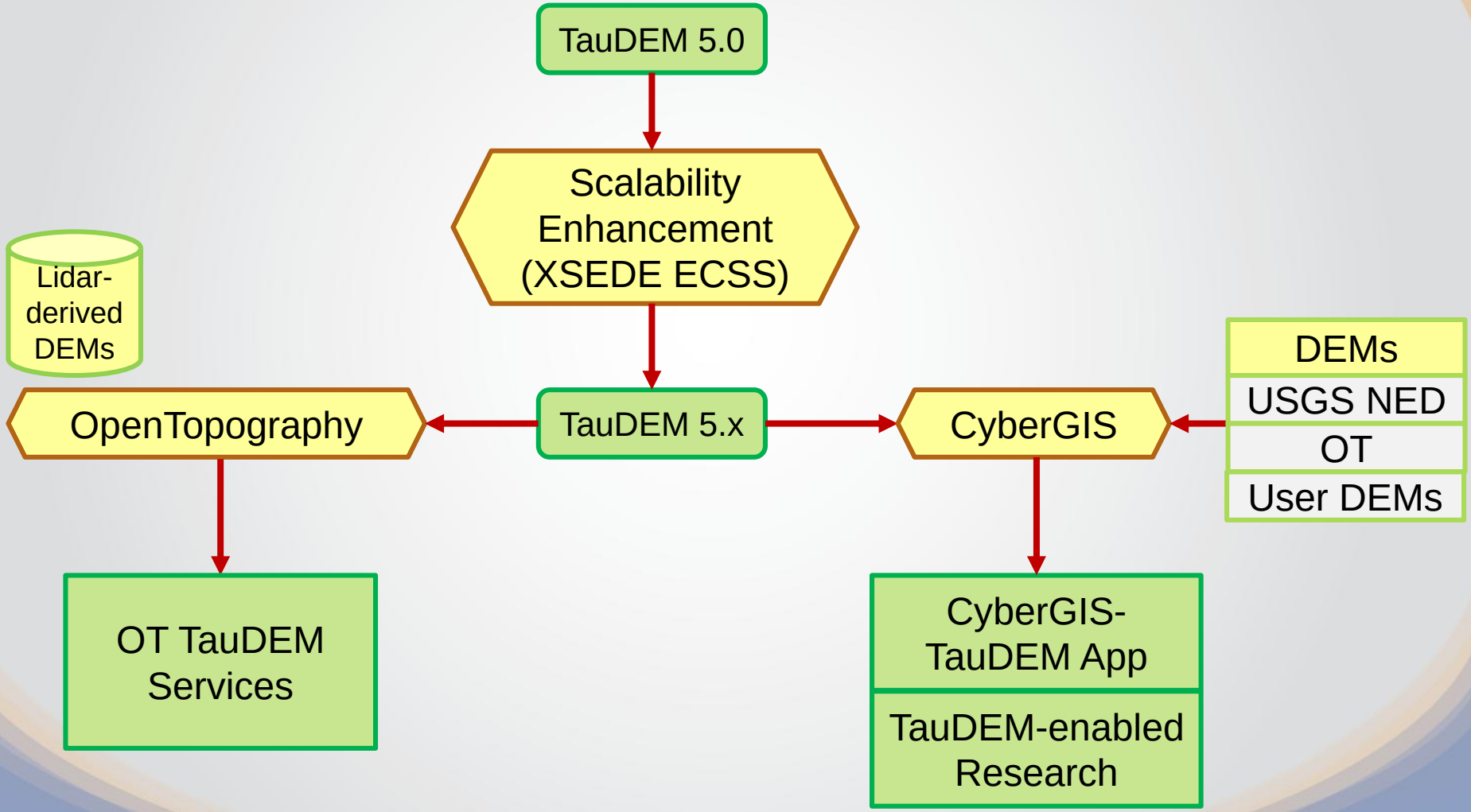
Computational Scaling Issues

Dataset	Size (GB)	Hardware	Number of Processors	PitRemove (run time seconds)		D8FlowDir (run time seconds)	
				Compute	Total	Compute	Total
GSL100	0.12	Owl (PC)	8	10	12	356	358
GSL100	0.12	Rex (Cluster)	8	28	360	1075	1323
GSL100	0.12	Rex (Cluster)	64	10	256	198	430
GSL100	0.12	Mac	8	20	20	803	806
YellowStone	2.14	Owl (PC)	8	529	681	4363	4571
YellowStone	2.14	Rex (Cluster)	64	140	3759	2855	11385
Boise River	4	Owl (PC)	8	4818	6225	10558	11599
Boise River	4	Virtual (PC)	4	1502	2120	10658	11191
Bear/Jordan/Weber	6	Virtual (PC)	4	4780	5695	36569	37098
Chesapeake	11.3	Rex (Cluster)	64	702	24045		

- Results collected on local cluster with Network File System (NFS) interconnet
- YellowStone dataset (27814x19320)
  - Using more processors reduced compute time, but suffered from longer execution time
- Chesapeake dataset (53248x53248)
  - Execution could not finish on D8FlowDir operation

http://hydrology.usu.edu/taudem/taudem5.0/TauDEM_4_12.pptx

CyberGIS-OT-TauDEM Collaboration



ECSS Goals

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

Collaboration Team

- National cyberinfrastructure
  - Extreme Science and Engineering Discovery Environment (XSEDE)
  - 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
  - DEM generation services for LiDAR-derived TauDEM analysis
  - Integration of TauDEM in OpenTopography service environment
  - People
    - Chaitan Baru, Nancy Wilkins-Diehr, Choonhan Yeon, San Diego Supercomputer Center (SDSC)
- NSF CyberGIS project
  - Integration of TauDEM in CyberGIS Gateway
  - Integration of TauDEM in advanced CyberGIS analytical services (workflow)
  - People
    - University of Illinois at Urbana-Champaign (UIUC)
      - Yan Liu, Anand Padmanabhan, Shaowen Wang
    - San Diego Supercomputer Center (SDSC)
      - Nancy Wilkins-Diehr, Choonhan Yeon

Performance Analysis: Challenges

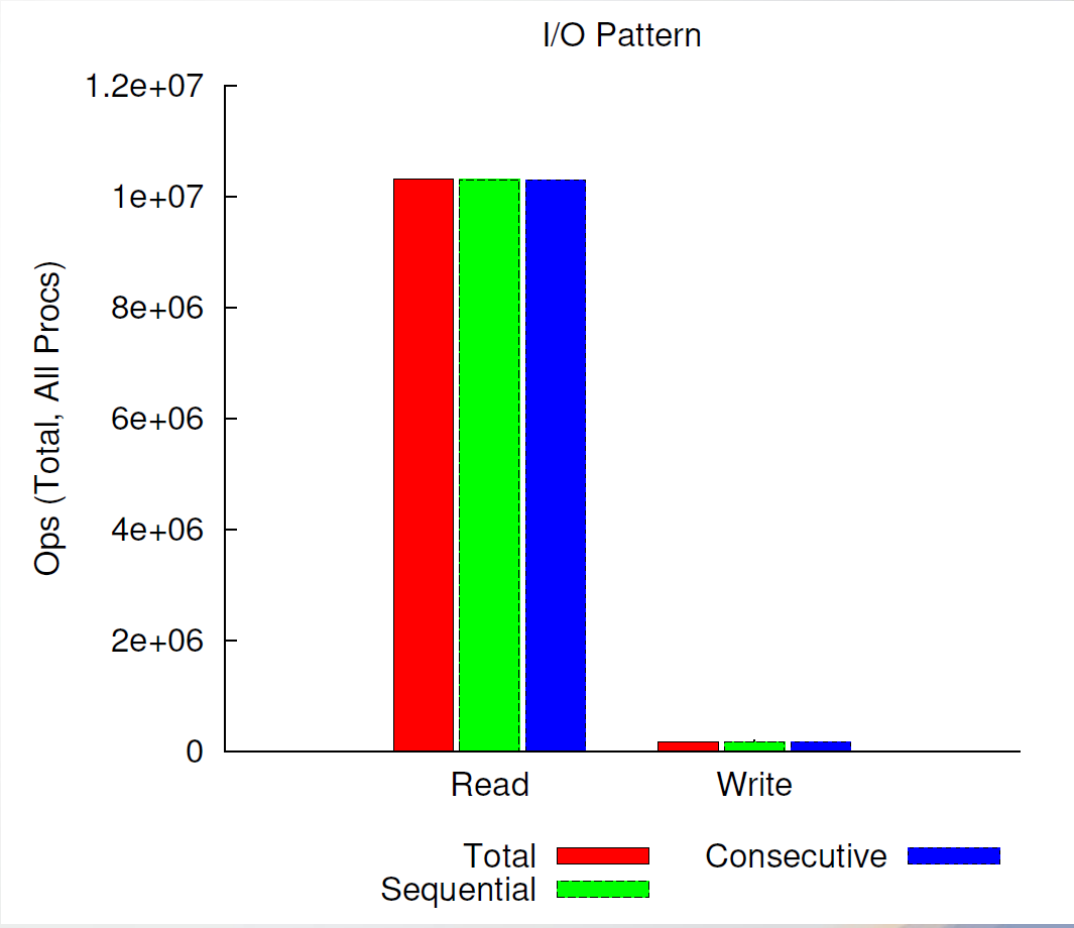
- System-level performance variation is very difficult to identify
  - Computing seemed not the reason for performance slowdown
  - Network issue or file system issue? NFS is difficult to debug
- Barrier for performance profiling
  - Performance profiling tools deployment need system administration skills
  - Using performance profiling libraries may need code change
  - Configuring profiling parameters and interpreting profiling results are not trivial

Strategies

- **Project management**
  - Code repository
    - TauDEM source code is moved to github to facilitate multi-party development and testing
    - http://github.com/dtarb/TauDEM
  - Documentation
    - Github wiki
    - Google Drive
  - Meetings
    - Bi-weekly teleconference
- **Build and test**
  - XSEDE resources:
    - Trestles@SDSC: for tests using up to 1,024 processors
    - Stampede@TACC: for tests using up to 16,384 processors
  - Profiling tools
    - Darshan: I/O profiling
- **Performance profiling and analysis**
  - Computational bottleneck analysis
    - Focus on I/O performance
  - Scalability to processors
  - Scalability to data size
  - Performance optimization

Generic I/O Profiling

- Darshan profiling found anomaly on file read operations
- The finding is confirmed in TauDEM timing data



IO Bottlenecks - Input

- Inefficient File Reading
  - *n* processes, *m* files
  - Original version: *n x m* file reads for getting geo-metadata
  - Fix: *1 x m* file reads + MPI_Bcast
- Coding Issue
  - File read deadlock situation caused by too many opened file descriptors
  - File not closed in time
  - Fix: close a file as soon as read operation is done

IO Bottleneck - Output

- Inefficient MPI IO
  - Original spatial domain decomposition did not consider IO performance
  - Improvement: domain decomposition strategy is changed to reduce the number of processes needed by an output file
- No Collective IO
- Parallel File System
  - Use as many OSTs on Lustre file system

Scalability Results

- Scalability Tests
  - Processors: up to 1,024
  - Data sizes: 2GB, 12GB, 36GB DEMs
- IO No Longer a Bottleneck

Results – Resolving I/O Bottlenecks

#cores	Compute	Header Read	Data Read	Data Write
32	42.7 / 42.8	193.5 / 3.8	0.4 / 0.4	**153.5 / 3.5**
64	35.3 / 34.8	605.5 / 3.9	1.5 / 1.1	**160.2 / 2.3**
128	33.7 / 33.0	615.2 / 2.6	0.9 / 1.0	**173.2 / 2.3**
256	37.5 / 38.0	831.7 / 2.3	0.5 / 0.9	**391.3 / 1.6**

Table 1. I/O Time Comparison (before / after; in seconds)
(Fan et al. 2014)

Results – Execution Time

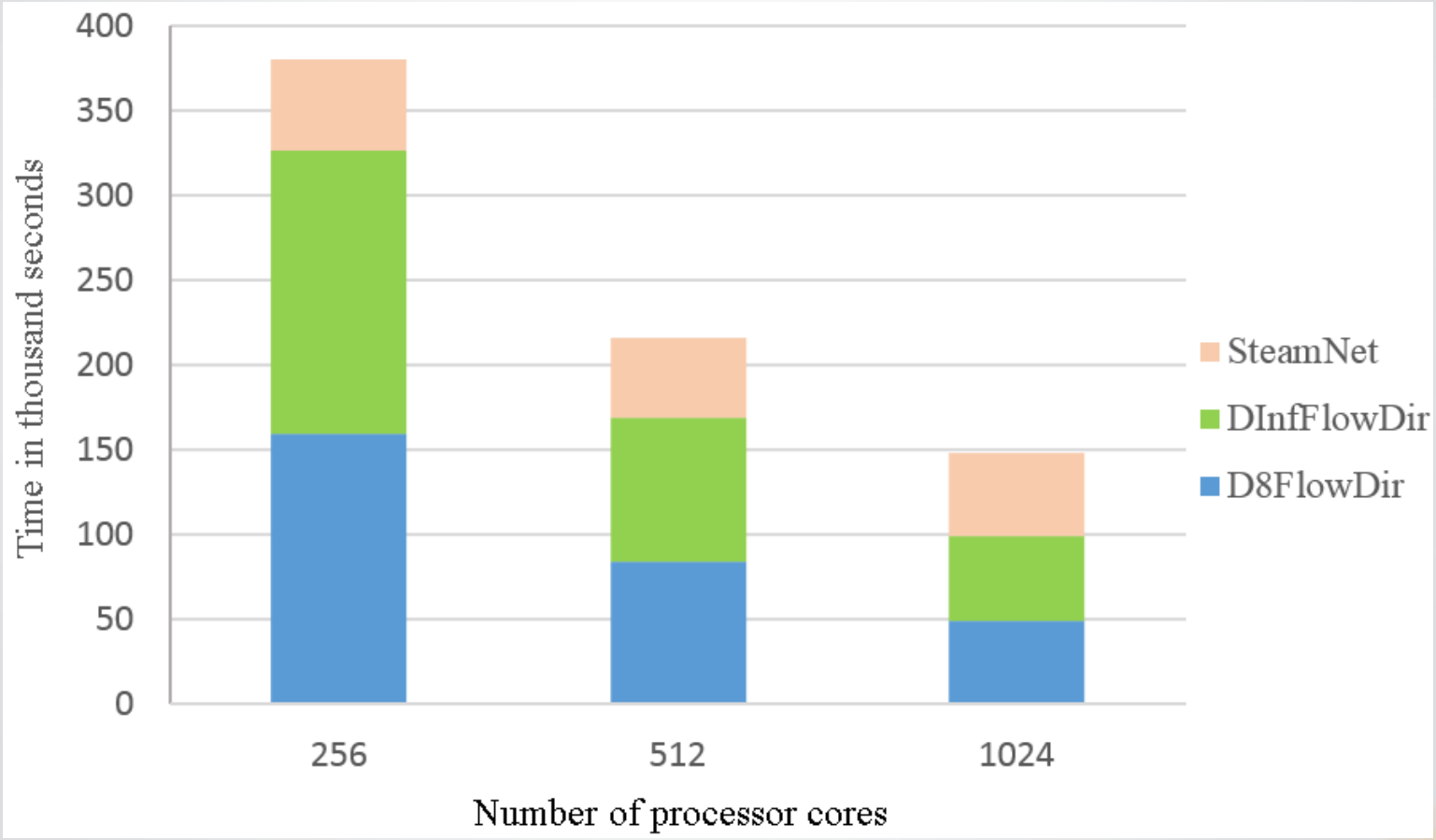


Figure 2. Execution time of the three most costly TauDEM functions on a 36GB DEM dataset. (*Fan et al. 2014*)

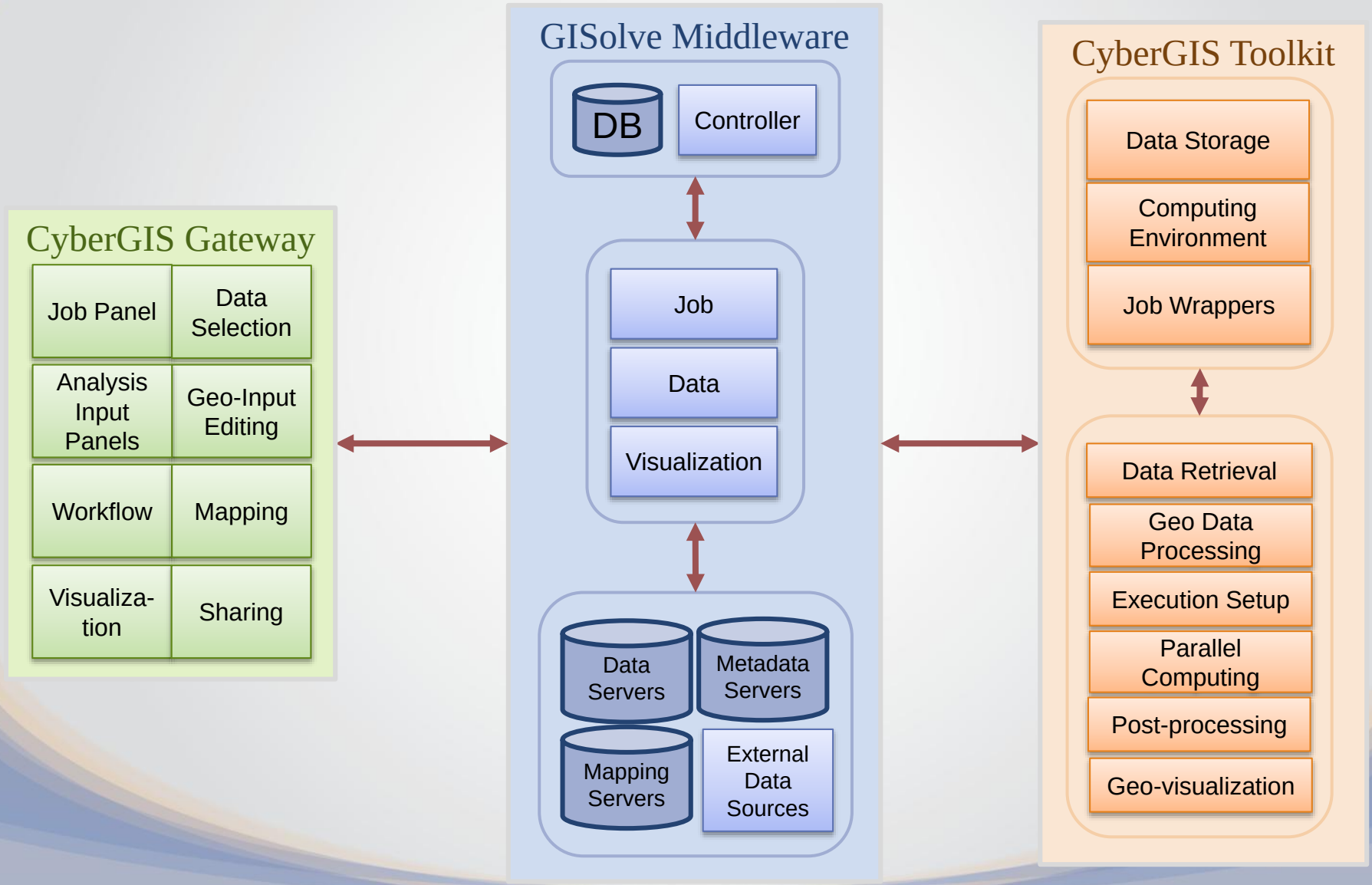
Next Steps

- More Room to Improve
  - 41.6 hours using 1024 cores on 36GB DEM
- Communication Pattern Analysis
- Methodological Investigation on Algorithm Design

CyberGIS-TauDEM Gateway Application

- Streamlined TauDEM Analysis in CyberGIS Gateway
  - Web environment
  - Transparent integration of DEM data sources
  - Customized TauDEM analysis workflow
  - Online visualization
- Status
  - 2 prototypes in April and May, respectively
  - Alpha release in early July
  - Beta release in August

CyberGIS Application Integration Framework



Data Integration

- Multiple High Resolution DEM Sources
  - USGS NED (10-meter)
    - Hosted at UIUC
    - Map preview
  - OpenTopography LiDAR-derived DEMs
    - Web service API
- Data Retrieval
  - USGS NED: wget
  - OT: Dynamic DEM generation and downloading
  - Data caching
    - XWFS?
- Data Processing
  - Study area clipping
  - Multi-file generation
  - Reprojection
  - GDAL library (http://gdal.org)
  - High-performance map reprojection
    - Collaborative work with USGS

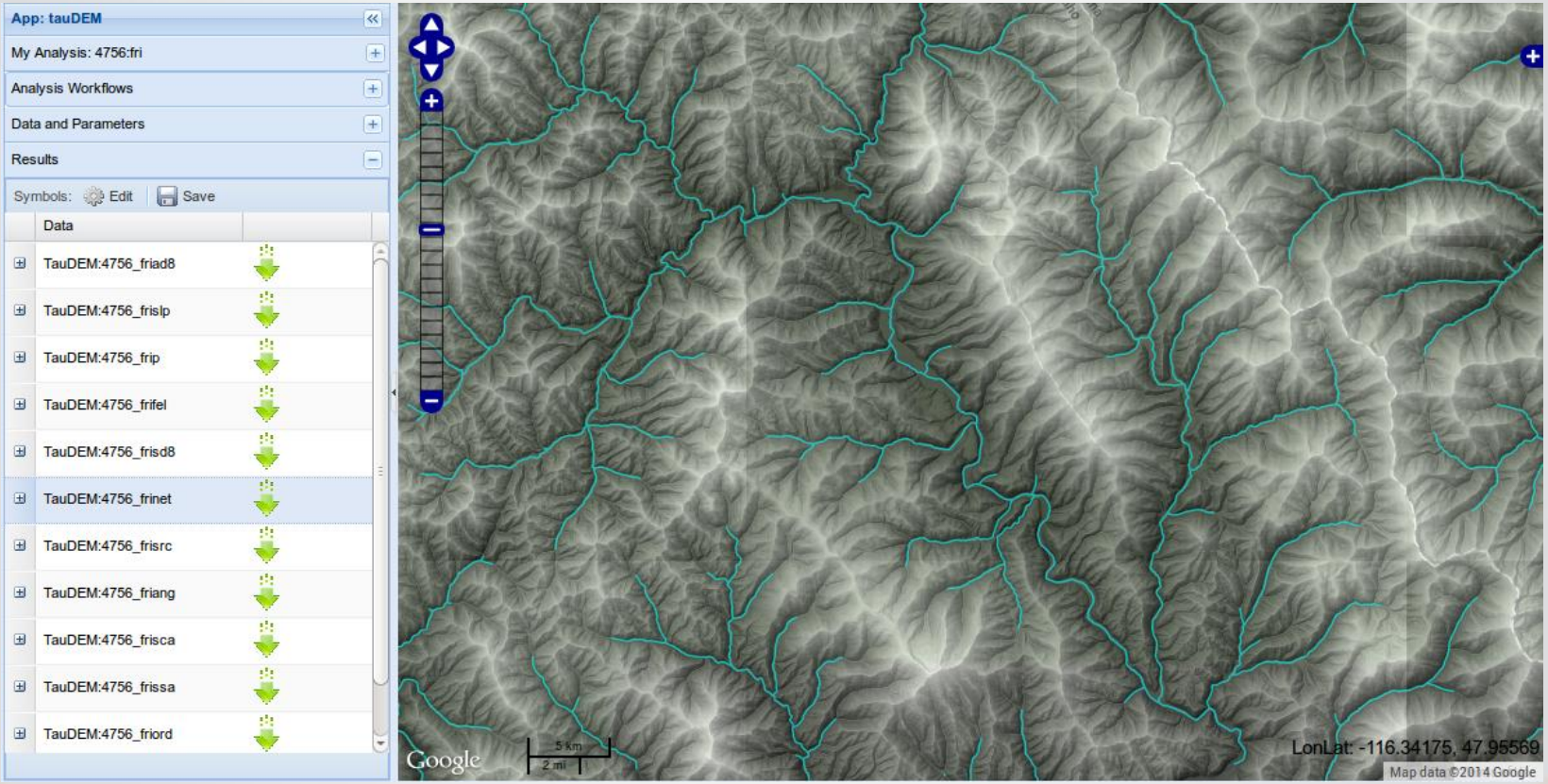
Analysis Workflow

- Approach
  - 26 TauDEM functions
  - Template-based customization of TauDEM functions
    - Pre-defined dependency
    - Dynamic workflow construction in Gateway
    - Data format: JSON
- Implementation
  - Interactive workflow configuration
    - Ext JS + SigmaJS
- Execution
  - Runtime command sequence generation
    - On Trestles: command sequence
    - On Stampede: a set of jobs linked based on job dependency

Visualization

- Visualization Computation
  - Reprojection
  - Pyramid generation for multiple zoom levels
  - Coloring (symbology)
- Online Visualization
  - Each product is a map layer accessible through the OGC-standard Web Mapping Service (WMS)

DEMO


App: tauDEM
My Analysis: 4756:fri
Analysis Workflows
Data and Parameters
Results
Symbols: Edit Save
Data
TauDEM:4756_friad8
TauDEM:4756_frislp
TauDEM:4756_frip
TauDEM:4756_frifel
TauDEM:4756_frisd8
TauDEM:4756_frinet
TauDEM:4756_frisrc
TauDEM:4756_friang
TauDEM:4756_frisca
TauDEM:4756_frissa
TauDEM:4756_friord
Google
5 km
2 mi
LonLat: -116.34175, 47.95569
Map data ©2014 Google

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
  - Explicit computational thinking in software development lifecycle (design, coding, testing)
  - Performance analysis remains challenging.
  - Collaboration with computational scientists and conducting performance profiling on cyberinfrastructure are important
  - 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

Acknowledgements

- XSEDE (NSF Grant No. 1053575)
- 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.
- Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation

Thanks!