

Introduction to CyberGIS

Dr. Junjun Yin

*CyberGIS Center for Advanced Digital and Spatial Studies
Department of Geography and Geographic Information Science
National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign, IL, 61801, USA*

January 19, 2016

GEOG 479: Advanced Topics in GIS – CyberGIS

- Course Goals
 - Be familiar with the high-performance computing components in CyberGIS
 - Leverage the CyberGIS architecture and resources to facilitate spatial thinking and knowledge discovery from geospatial Big Data.
 - Interact with geospatial Big Data in various forms (e.g., geo-located Twitter data, Taxi trajectory data and national DEM, etc.)
 - Solve geospatial problems with high-performance computing approaches.
- Course Syllabus
 - <https://bitbucket.org/geog479/course/downloads/Syllabus.pdf>
- Course materials
 - <https://bitbucket.org/geog479/>

GEOG 479: Advanced Topics in GIS – CyberGIS

- Course Outline
 - Introduction to CyberGIS, Geospatial Big Data and ROGER
 - High performance computing, cyberinfrastructure and GIS
 - Getting started with ROGER
 - Distributed Spatial Databases
 - Taming Big Geospatial Data with Hadoop
 - Advanced Geospatial Data Analytics
 - Mid-term project
 - Terrain analysis: Parallel MapAlgebra for Spatial Analysis
 - Interactive Geo-visualization of Large-scale Movement Data with Spark
 - Final project

What is CyberGIS?

Cyber + GIS ?

What is GIS?

- Systems
- Science
- Services
- Society
- Synthesis
- Spatial is special

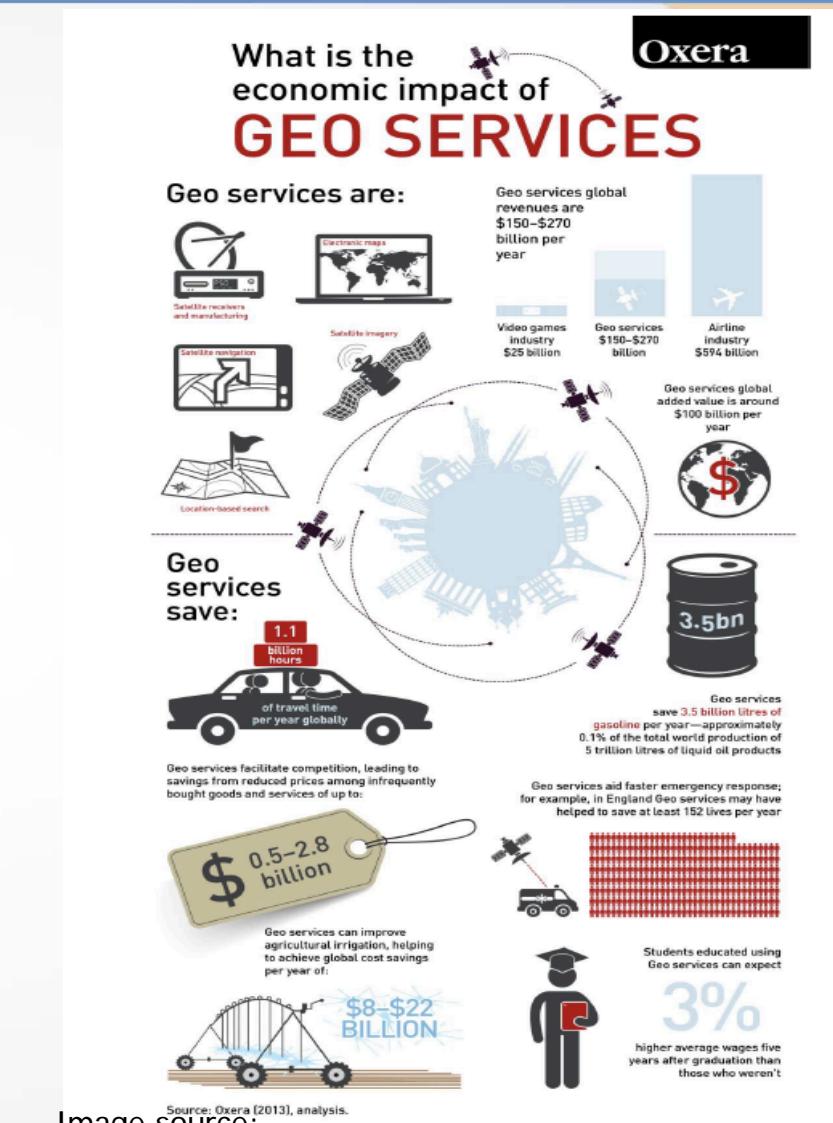


Image source:

http://www.oxera.com/Oxera/media/Oxera/download/s/reports/What-is-the-economic-impact-of-Geo-services_1.pdf

What is CyberGIS?

- CyberGIS is the synthesis of cyberinfrastructure, GIS and spatial analysis

Big data



Big compute

Big collaboration

Big problems

- Why CyberGIS?

- Geospatial discovery and innovation with advanced digital technologies¹

1. Wang, S. (2013)

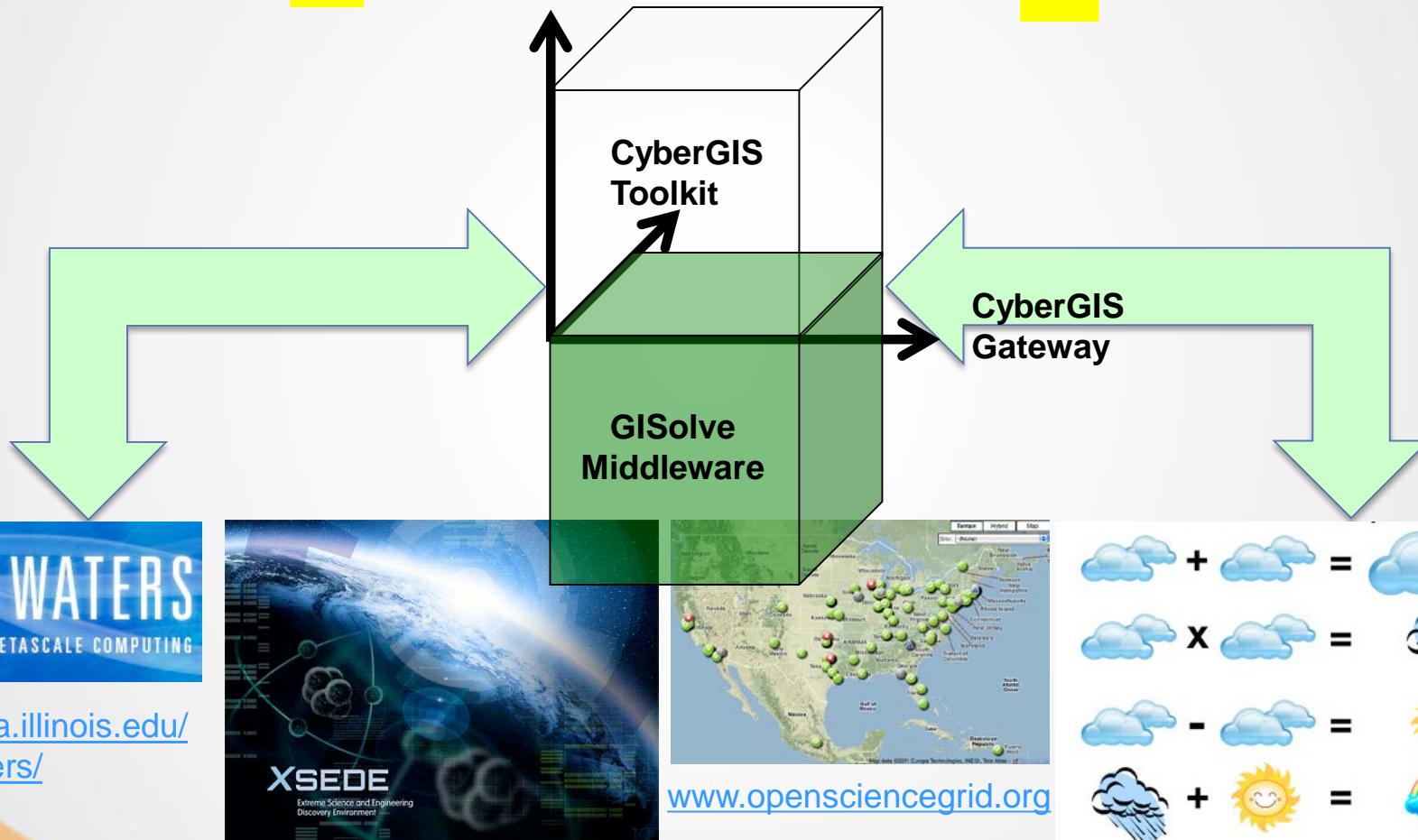
cyberinfrastructure

CyberGIS is Geographic Information Science and Systems based on advanced cyberinfrastructure.

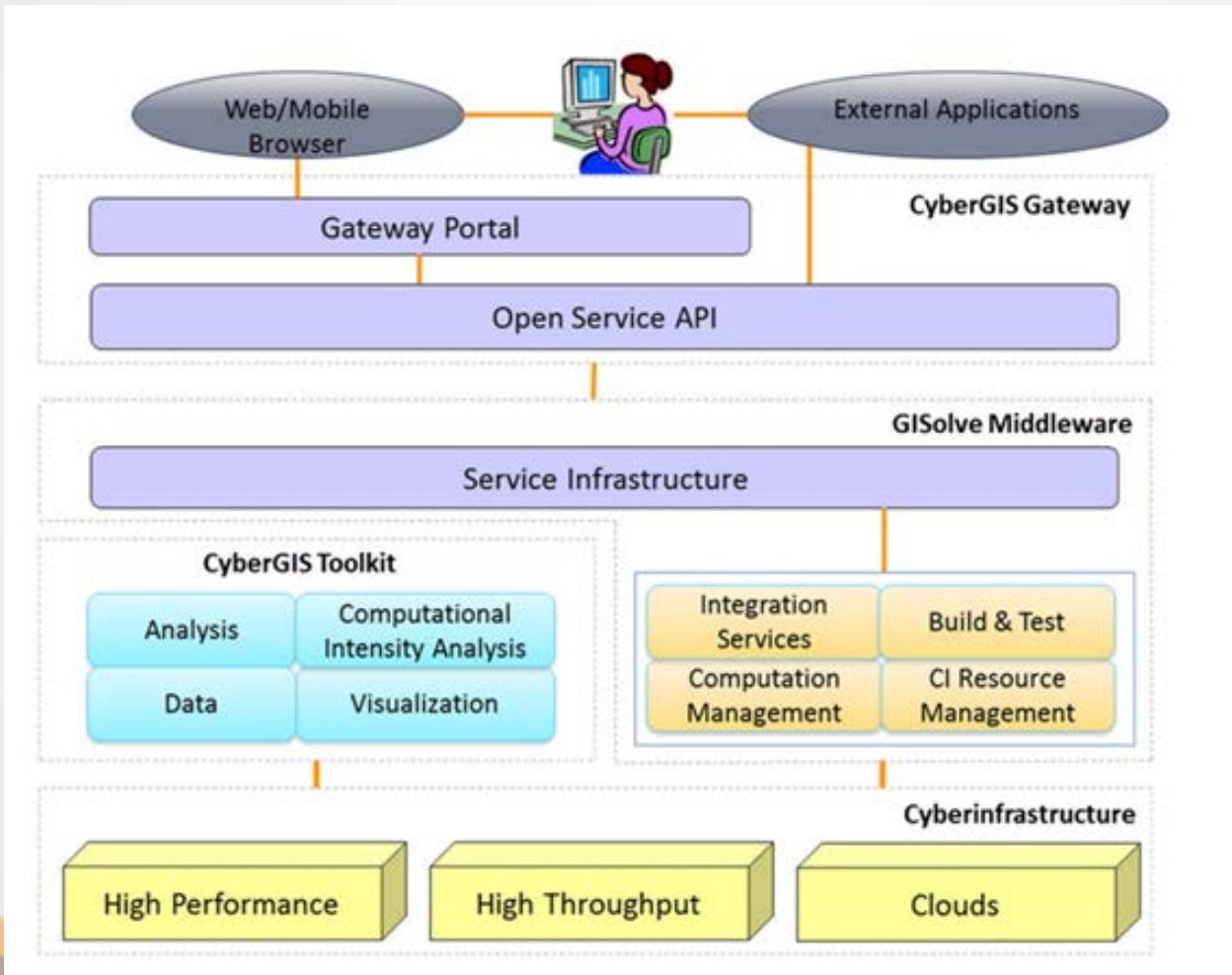
Cyberinfrastructure includes:

- (high performance) computing systems
- data storage systems
- advanced instruments
- data repositories
- visualization environments
- people
- linked by high speed networks

Advanced Digital and Spatial Studies



CyberGIS Software Environment



CyberGIS Gateway



The screenshot shows the CyberGIS Gateway homepage. At the top, there's a navigation bar with links for Home, Apps, Visualization, Community, and Help, along with Login and Register buttons. The main feature is the BIOSCOPE app, which is described as a "CyberGIS Gateway App for online biomass-to-biofuel supply chain system optimization." It features a map of the United States with a highlighted area in the Midwest. Below the app, there are sections for About, News, Status, and Recent Users.

About

CyberGIS Gateway is the leading online cyberGIS environment for a large number of users to perform computing- and data-intensive, and collaborative geospatial problem-solving enabled by advanced cyberinfrastructure. Supported in part by the National Science Foundation (NSF) [CyberGIS Project](#), the Gateway provides a suite of scalable spatial analysis and modeling applications built on [GISolve](#) middleware infrastructure. The computation of analytical and modeling

News

CyberGIS 2014
August 18-22, 2014: [The Second International Conference on CyberGIS and Geodesign](#) (CyberGIS 2014) will provide a forum for sharing cutting-edge research, education and training experiences ranging from new theories,

Status
CyberGIS Gateway v2.5b
Current build: r528.
11/16-17: BioScope resource maintenance.

Recent Users
Hao, UofI
Kelly, University of Illinois

GISolve Middleware



GISolve Open Service API

[Introduction](#)

[Status](#)

[Security API](#)

[Issue a token](#)

[Verify a token](#)

[Revoke a token](#)

[Integration API](#)

[Register an app](#)

[Get app information](#)

[Configure an app](#)

[Get app configuration](#)

[Update app integration status](#)

[Computation API](#)

[Launch a job](#)

[Monitor a job](#)

[Get job output](#)

[Version API](#)

[Glossary](#)

[Appendix 1: JSON Format for Application Configuration](#)

[Introduction](#)

[Configure Parameters](#)

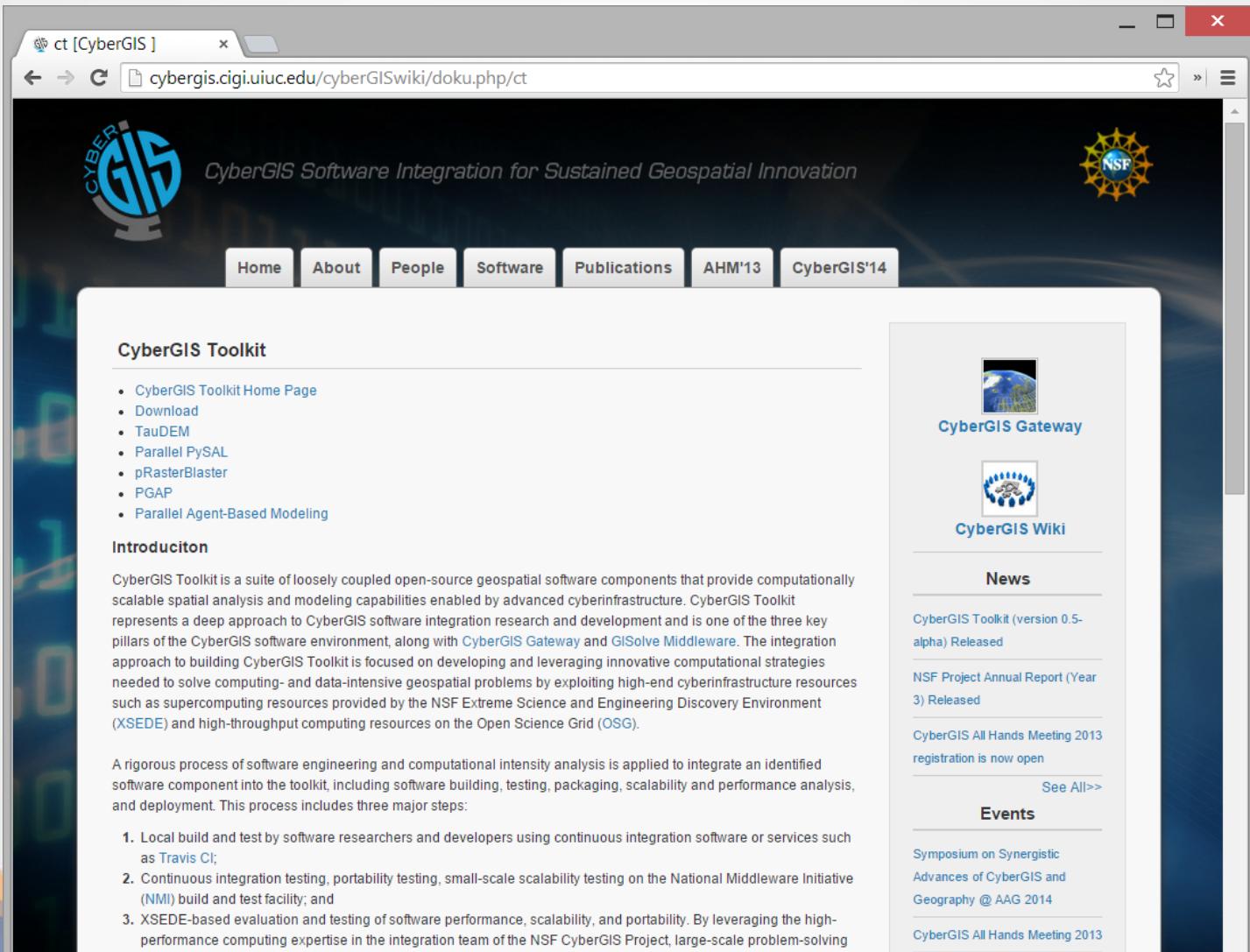
[Joints Parameters](#)

[Disjoints Parameters](#)

[Group Parameters](#)

[Appendix 2: JSON Format for Runtime Parameters](#)

CyberGIS Toolkit



The screenshot shows a web browser window displaying the CyberGIS Toolkit homepage. The URL in the address bar is cybergis.cigi.uiuc.edu/cyberGISwiki/doku.php/ct. The page features a dark blue header with the CyberGIS logo on the left and the NSF logo on the right. Below the header is a navigation menu with links to Home, About, People, Software, Publications, AHM'13, and CyberGIS'14. The main content area has a white background and contains sections for the CyberGIS Toolkit, Introduction, News, and Events.

CyberGIS Toolkit

- CyberGIS Toolkit Home Page
- Download
- TauDEM
- Parallel PySAL
- pRasterBlaster
- PGAP
- Parallel Agent-Based Modeling

Introduction

CyberGIS Toolkit is a suite of loosely coupled open-source geospatial software components that provide computationally scalable spatial analysis and modeling capabilities enabled by advanced cyberinfrastructure. CyberGIS Toolkit represents a deep approach to CyberGIS software integration research and development and is one of the three key pillars of the CyberGIS software environment, along with [CyberGIS Gateway](#) and [GISolve Middleware](#). The integration approach to building CyberGIS Toolkit is focused on developing and leveraging innovative computational strategies needed to solve computing- and data-intensive geospatial problems by exploiting high-end cyberinfrastructure resources such as supercomputing resources provided by the NSF Extreme Science and Engineering Discovery Environment ([XSEDE](#)) and high-throughput computing resources on the Open Science Grid ([OSG](#)).

A rigorous process of software engineering and computational intensity analysis is applied to integrate an identified software component into the toolkit, including software building, testing, packaging, scalability and performance analysis, and deployment. This process includes three major steps:

1. Local build and test by software researchers and developers using continuous integration software or services such as [Travis CI](#);
2. Continuous integration testing, portability testing, small-scale scalability testing on the National Middleware Initiative ([NMI](#)) build and test facility; and
3. XSEDE-based evaluation and testing of software performance, scalability, and portability. By leveraging the high-performance computing expertise in the integration team of the NSF CyberGIS Project, large-scale problem-solving

CyberGIS Gateway

CyberGIS Wiki

News

CyberGIS Toolkit (version 0.5-alpha) Released

NSF Project Annual Report (Year 3) Released

CyberGIS All Hands Meeting 2013 registration is now open

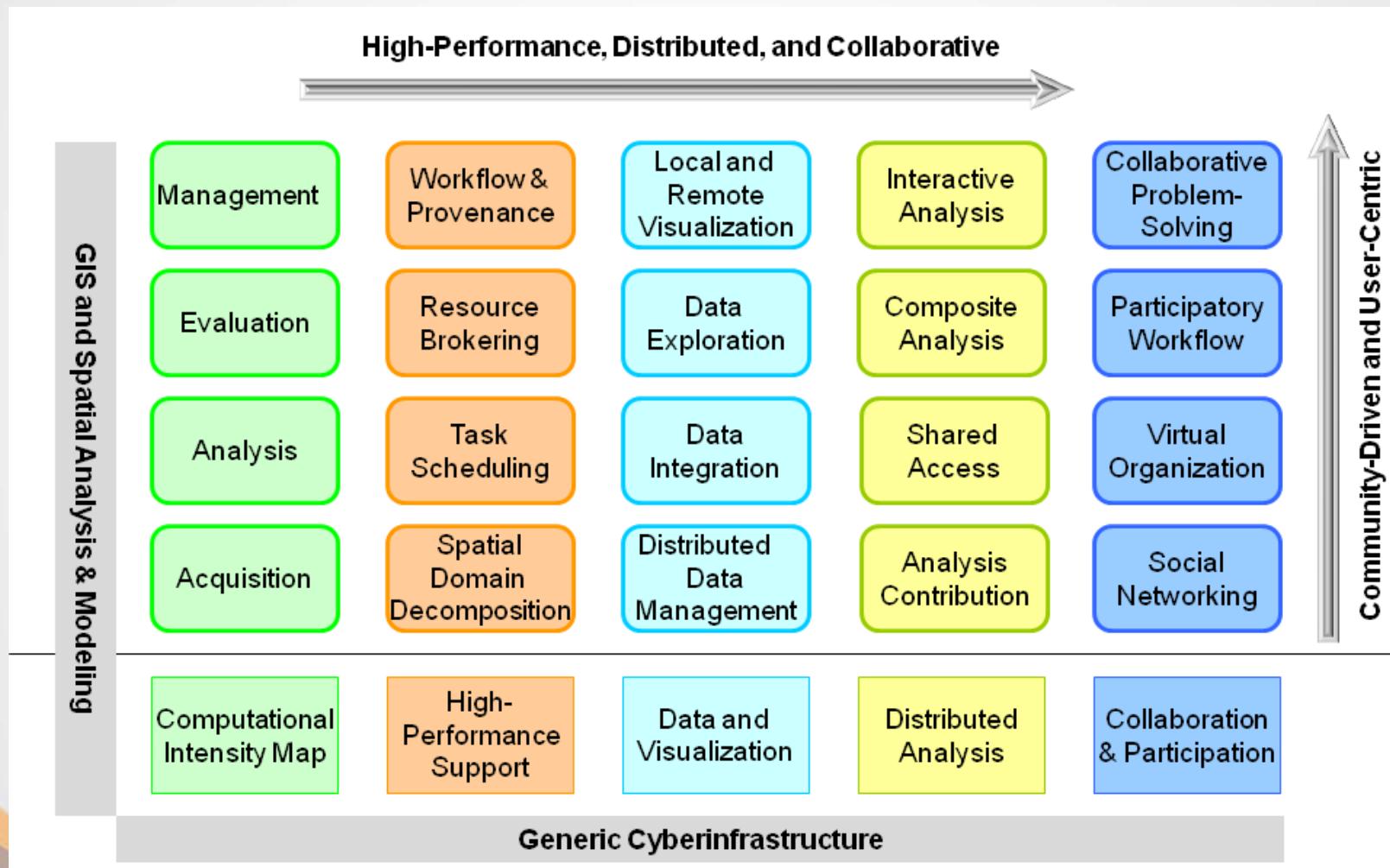
See All>>

Events

Symposium on Synergistic Advances of CyberGIS and Geography @ AAG 2014

CyberGIS All Hands Meeting 2013

CyberGIS Toolkit



CyberGIS architecture
(source: Wang et al., 2013)

Terminologies

- Cloud computing
- High performance computing
- High-throughput computing
- Distributed computing
 - Hadoop and Spark
 - Non-relational database (NoSQL)
 - Distributed database
- Parallel computing

- Computational geography
- WebGIS

CyberGIS Center for Advanced Digital and Spatial Studies

*Empower advanced digital and spatial studies
through innovation of
CyberGIS technologies and applications*

Research Application Focus Areas:

**Agriculture and Food
Disaster and Emergency
Earth and Environment
Energy and Water Resources
Health and Wellness**

CyberGIS Gateway and ArcGIS Online Integration

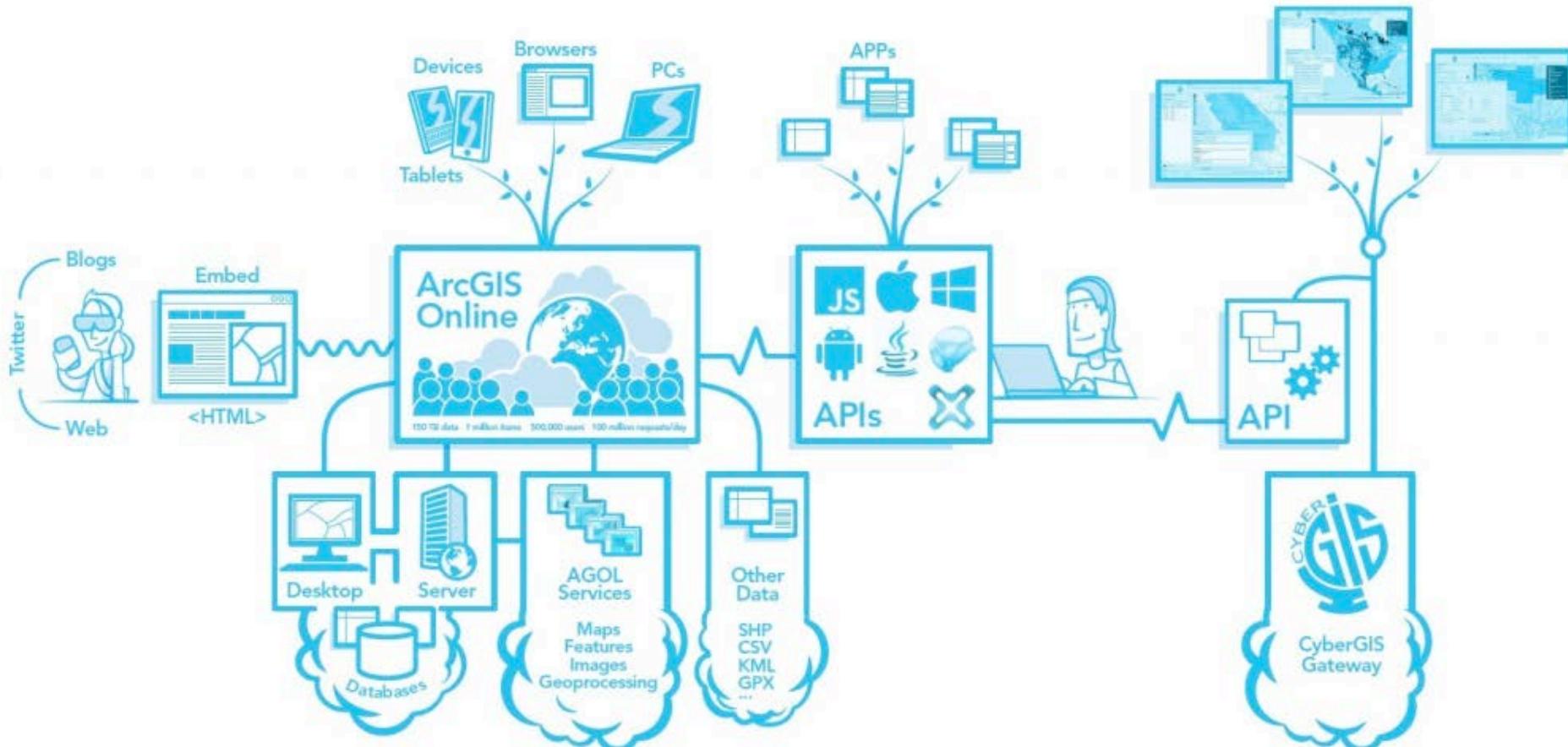


Image source: <http://blogs.esri.com/esri/arcgis/2013/10/01/what-is-cybergis/>

Domain Communities and Sciences

Multidisciplinary	Geosciences	Engineering	Social Sciences	Biosciences
<ul style="list-style-type: none">Computational and Data SciencesCyberGIS & GIScienceEmergency ManagementEnvironmental Sustainability...	<ul style="list-style-type: none">Atmospheric ScienceHydrologyGeologyRemote Sensing	<ul style="list-style-type: none">BioenergyEnvironmental EngineeringTransportation	<ul style="list-style-type: none">Geography and Spatial SciencesSocial Media & NetworkPolitical ScienceOrganizational Science	<ul style="list-style-type: none">AgriculturePlant biologyMedical Science

CyberGIS Capabilities & Services

Scalable Data & Analytics	Multi-scale Modeling & Simulation of Complex Spatiotemporal Systems	Interactive & Real-time Analytics	Spatial Decision-making	Collaborative Problem Solving
Data Integration <ul style="list-style-type: none">Data CloudInteroperable AccessProvenance	Scalable GIS <ul style="list-style-type: none">OptimizationSimulationStatistics	Visual Analytics <ul style="list-style-type: none">MappingOn-demand Visualization	GISolve <ul style="list-style-type: none">CI AccessOpen ServicesSpatial MiddlewareWorkflow	PSE <ul style="list-style-type: none">CollaborationSocial NetworkingUser Interfaces

Advanced CI Capabilities & Services

Computation	Data	Operation	Support	
<ul style="list-style-type: none">Resource ManagementPrivate CloudScalable ComputingSoftware Environment	<ul style="list-style-type: none">MetadataStorage & CurationDatabasesTransfer & Sharing	<ul style="list-style-type: none">AuditingScalability AnalysisSecuritySystem Monitoring	<ul style="list-style-type: none">Advanced ApplicationsEOTHelpdesk	
HPC (CPU, GPU, Data-Intensive)	Large Memory	Large Storage & High Bandwidth Input/Output	Fast & Low Latency Network	On-demand Virtualization

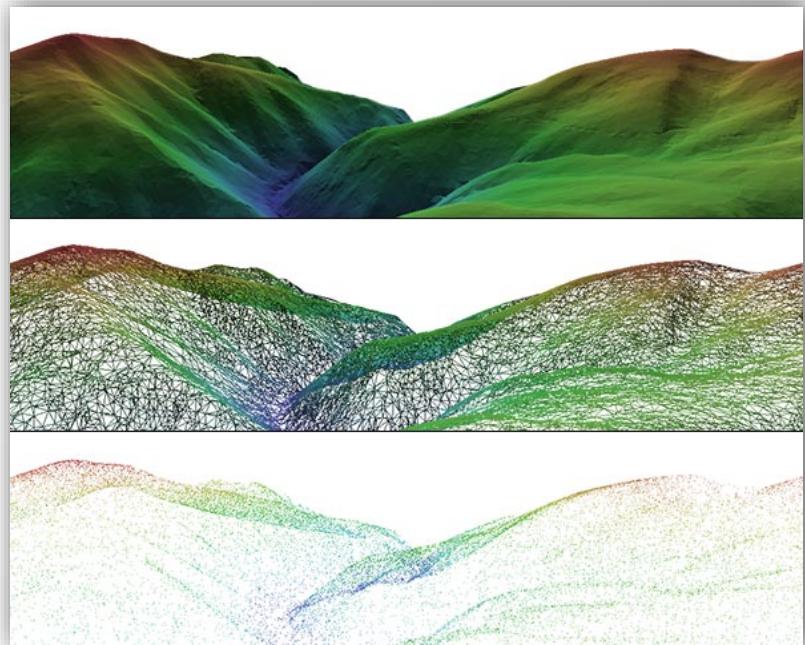
* EOT – Education, Outreach, and Training; HPC – High Performance Computing; PSE – Problem-Solving Environment

Education, Outreach, and Training Activities

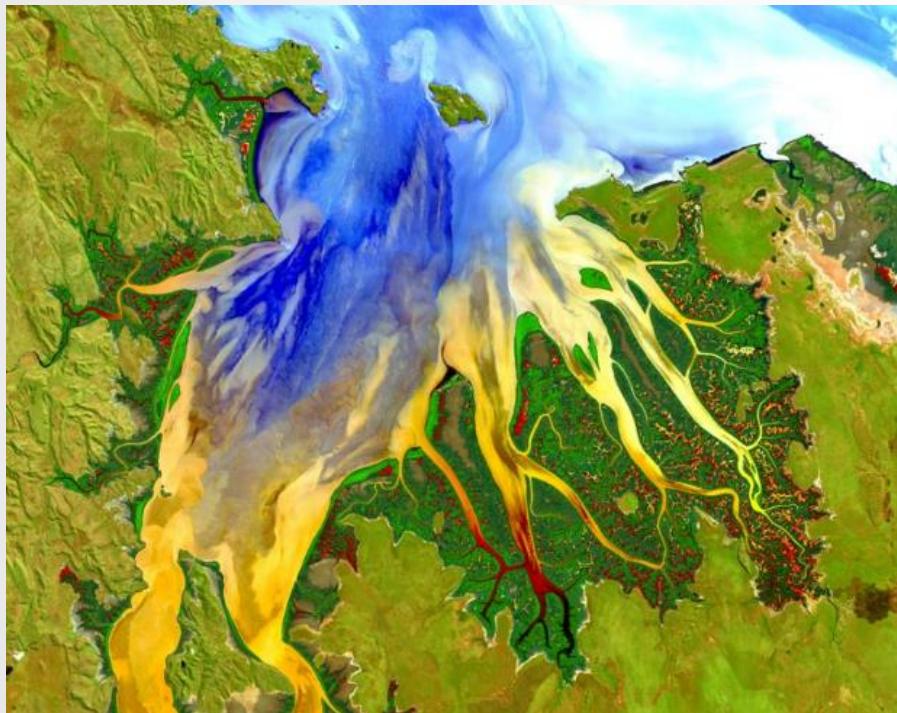
- CyberGIS Commons
 - For hands-on learning and training
- CyberGIS Collaboration Space
 - For groups to develop collaborative research
- CyberGIS Helpdesk
 - For in-depth technical consulting
- CyberGIS Brown Bag Seminars
 - Once every two to three weeks on average
- CyberGIS Distinguished Speaker Series
 - Once every two months on average
- CyberGIS MOOCs
- CyberGIS Webinars

Geospatial Big Data

- Volume
 - USGS NED 30m(175GB), 10m(500GB)
- Variety
 - Forms and formats (Raster and Vector)
 - Shapefile, GeoTiff, GeoJSON, WMS, WFS, NetCDF, HDF5
- Velocity
 - All Geocoded Tweets
 - 2.5kb/tweet
 - 6,000 average tweet per second
 - 3% of tweets are geo-located
 - 38GB per day
 - Landsat 8
 - 650GB per day
- Veracity
 - Uncertainty



Geospatial Big Data



Landsat 8

Twitter Firehose

All Geocoded Tweets

~1650 images per day

~1GB each (compressed)

~650gb per day

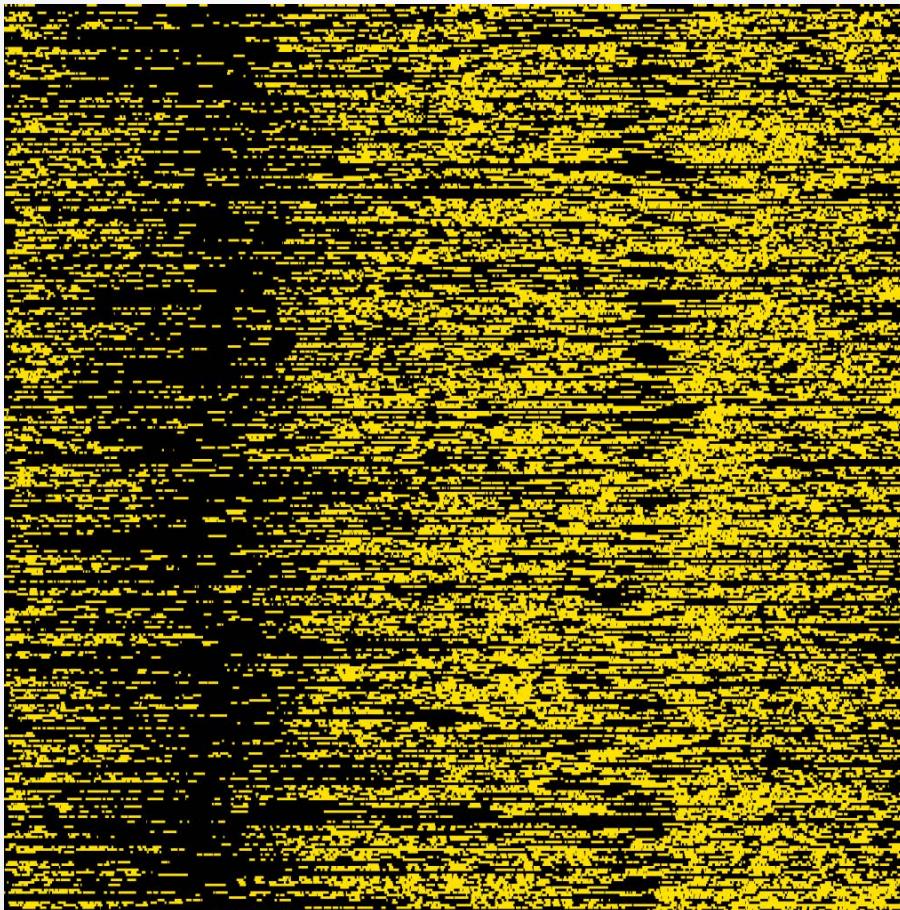
~2.5KB / tweet

~6,000 average tweet per second

~3% of tweets are geolocated

~38GB per day

Geospatial Big Data



Visualization of New York taxi data

The black pixel means the taxi did not have fare and yellows means it did
Source: <http://chriswhong.com/open-data/my-first-data-art-nyc-taxi-defrag/>

CyberInfrastructure Resources



NSF Blue Waters @ UofI
13,300 TFlop/s



NSF XSEDE: SDSC Comet
2,000 TFlop/s



UofI Campus Cluster

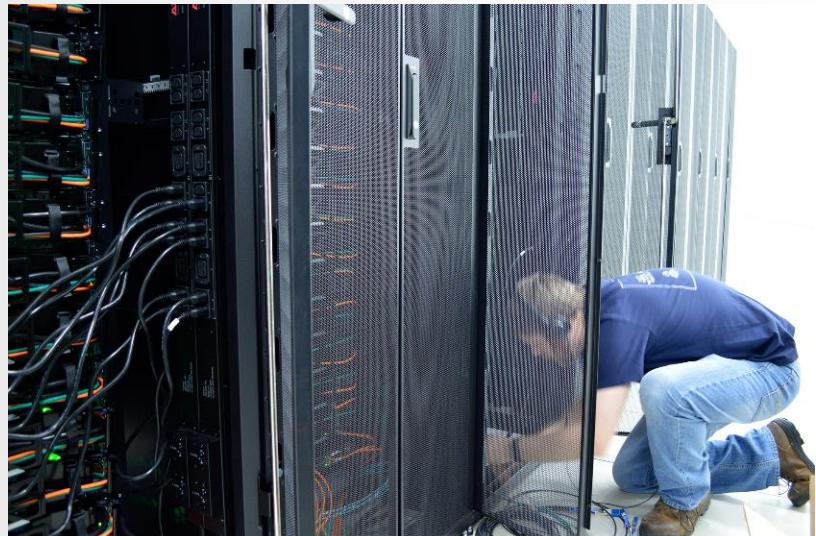
CyberGIS ROGER



CyberGIS ROGER
Resourcing Open Geospatial Education and Research
[After Roger Tomlinson](#)
goo.gl/8MpYC9
~60 TFlop/s

Why ROGER?

- Both a research project and a research resource
- Configured to best support geospatial data, with emphases on local memory and shared storage size and speed
- Supports multiple paradigms: traditional batch HPC, Hadoop, and Cloud (OpenStack).
- Integrate the three paradigms and leverage their strengths
- Inform the design of future geospatial supercomputers



ROGER: Overview

- Provides the CyberGIS Center with HPC, Hadoop and OpenStack functionality all on one system
- Managed by Systems Group
With support from our storage, network, security and services groups

ROGER: File system

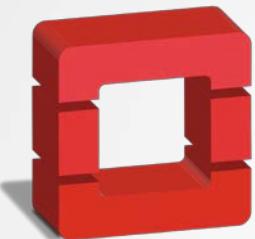
- IBM General Parallel File System (GPFS)
- Parallel file system means all of the nodes get fast access to hard drive storage.



ROGER for this class

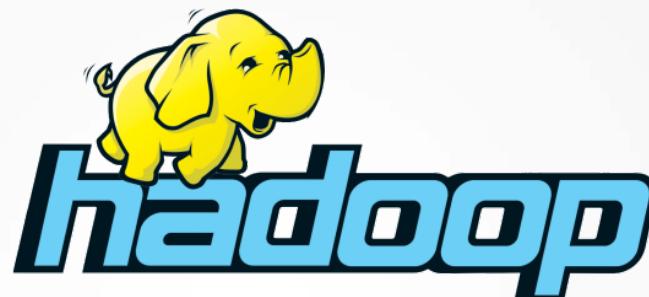
- Will create ROGER accounts for every student in this class
- Interact with various components in ROGER
 - Job submission
 - Handling and processing massive data with Hadoop, Spark, R, and Python, etc.
 - Getting to know OpenStack
 - Working with Geospatial libraries

ROGER for this class



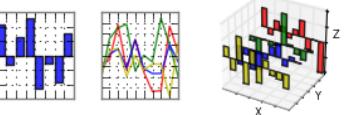
openstack.
CLOUD SOFTWARE

 mongoDB



spark

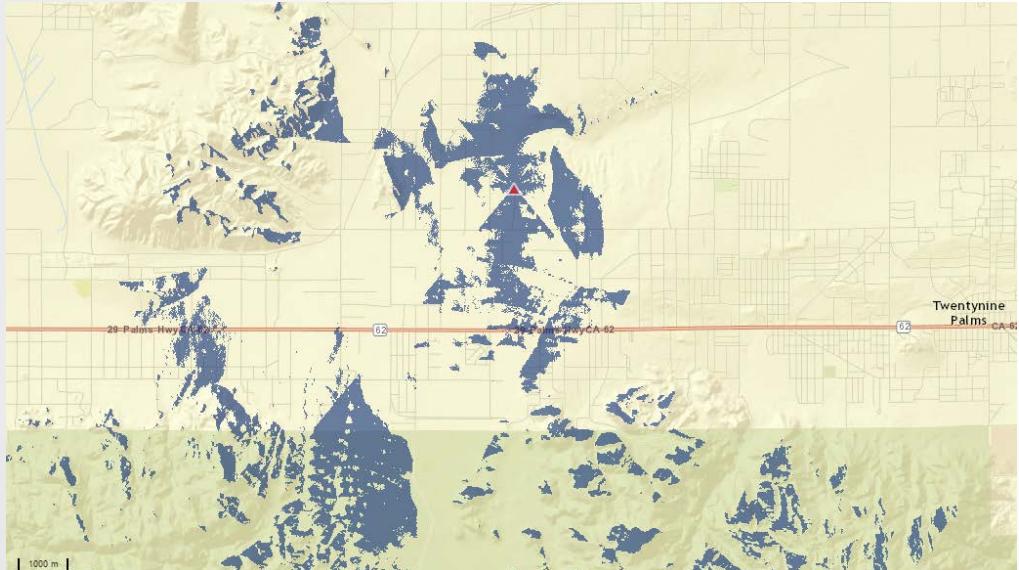
pandas
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



The pandas logo includes a mathematical equation for time series analysis and three small icons illustrating data manipulation: a heatmap, a line graph, and a 3D bar chart.



CyberGIS Viewshed App



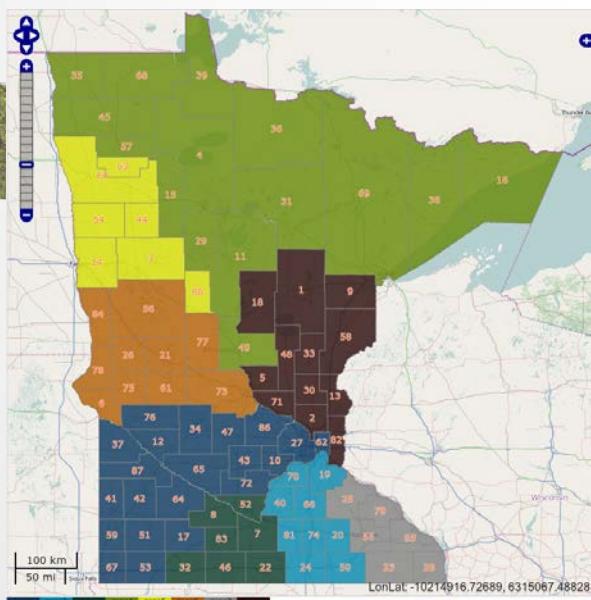
Viewshed



UofI Cluster



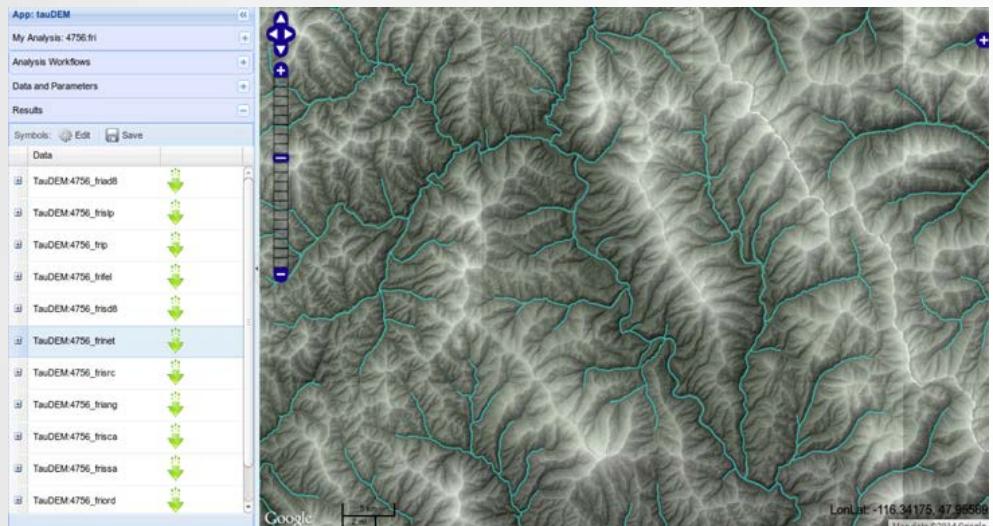
Parallel Genetic Algorithm for Redistricting Optimization on Blue Waters



NSF Blue Waters @ UofI
13,300 TFlop/s

Collaborative Work with Wendy K. Tam Cho @ UIUC

CyberGIS TauDEM App



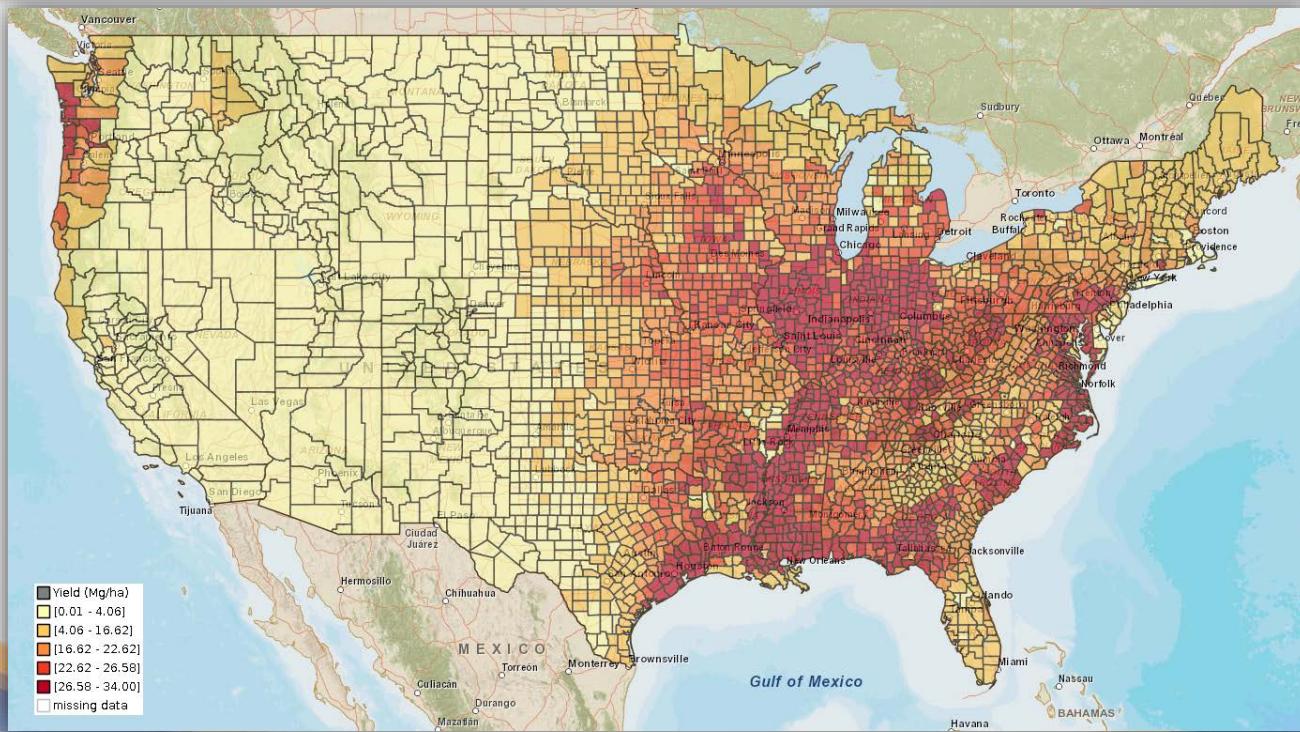
TauDEM



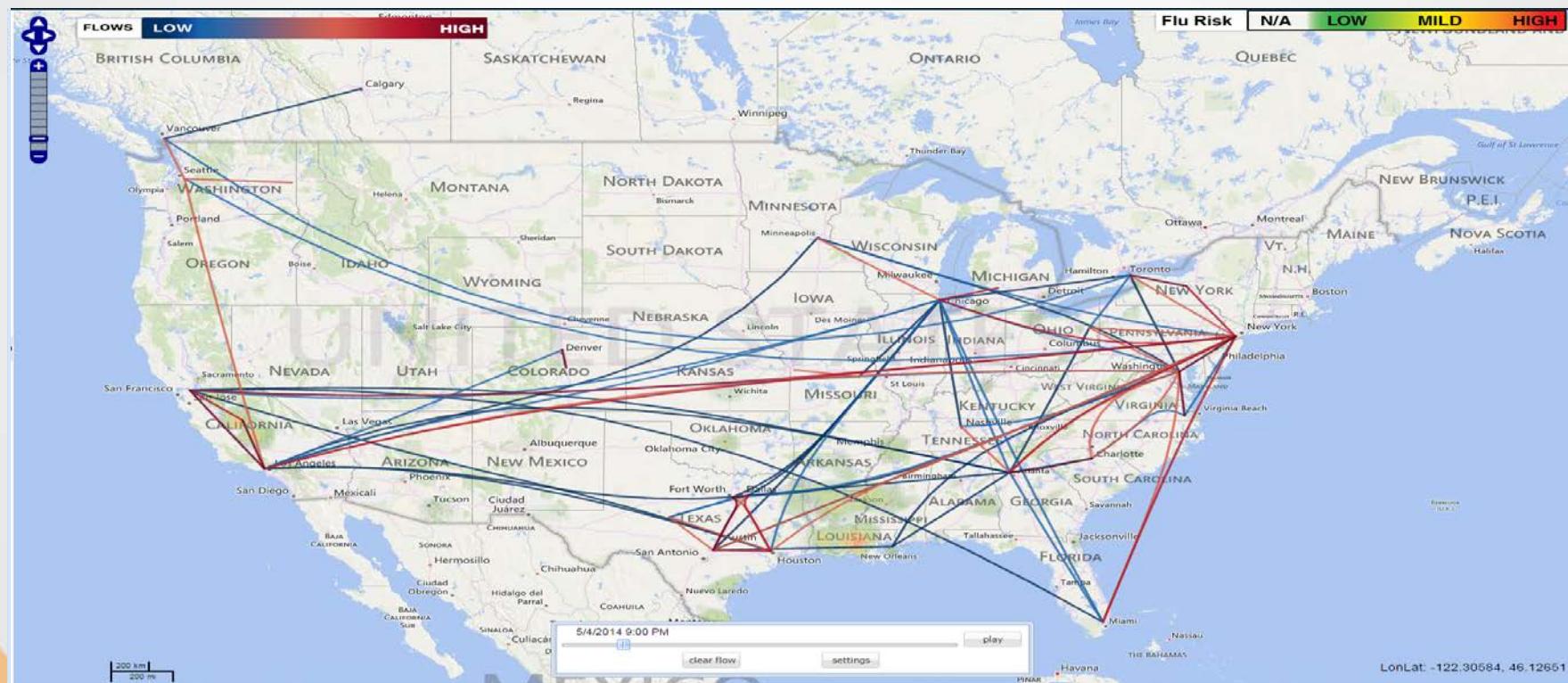
NSF XSEDE: SDSC Trestles
100 TFlop/s

Collaborative Work with David Tarboton @ Utah State University

CyberGIS BioScope App



CyberGIS FluMapper App



Lab

1. FluMapper

- Social media, point data, streaming data, Hadoop, routing, raster and vector visualization

2. BioScope

- Agriculture, supply chain optimization, chart generation, reference base maps

3. TauDEM

- Hydrologic data analysis, digital elevation models, code reuse

Reference

- Wang, S., Anselin, L., Bhaduri, B., Crosby, C., Goodchild, M. F., Liu, Y., and Nyerges, T. L. (2013). Cybergis software: a synthetic review and integration roadmap. *International Journal of Geographical Information Science*, 27(11), 2122–2145.
- Wang, S. (2013) “CyberGIS: Blueprint for Integrated and Scalable Geospatial Software Ecosystems.” *International Journal of Geographical Information Science*, 27 (11): 2119-2121