Data Challenges in Environmental Research

John McGee, Brian Blanton, Jeff Heard, Oleg Kapeljushnik

RENCI, University of North Carolina at Chapel Hill
Projects with common requirements

**Current**
- NWS WxEM, Big Board, NOAA Decision Support Prototype
- NCB Prepared
- CI-BER prototypes
- SPH grant tracking app
- SPH farmers’ market locator
- Coastal Risk Analysis

**Potential**
- Public health informatics
- Environmental science
- Decision support / Situational understanding
- Geographic collection browsing / Library Sciences
- NC First, version 2
Exploring a GeoAnalytics Framework

- Rapidly develop and deploy prototypes and solutions
- Reduce the barrier of entry to using geographic data for science
- Scale horizontally to Big Data, its update frequency, access patterns, and management requirements
- Vet, integrate, and federate open source geography tools

Led by Jeff Heard at RENCI
Data Cube

Store/query versioned, 4d regular-gridded data: X/Y, Elevation, Time

- Stores multiple elements as a set
- High level query routines support
  - re-projection
  - filtering
  - masking
- Server-side queries return Fortran-style arrays that can be processed by BLAS, native code, or Python’s Numpy/Scipy
- Client-side queries can be filtered through OWS or custom services
Distributed Task Queueing

Schedule tasks for regular execution
Defer long-running tasks for batch processing
Tasks are developed per-application and automatically discovered and scheduled by Geoanalytics, or can be used in server-side program as functions with deferrable execution

Example tasks include:
- Sunsetting ADCIRC and sensor data
- Running batch analytics on an offline dataset
Communicating Coastal Risk Analysis

science lead: Brian Blanton, RENCI
Coastal Risk Analysis

- Significant worldwide populations live along coastlines
- Risks include storm surge inundation, precipitation driven flooding, waves, coastal erosion
- Increased risk likely as impacts of a changing climate are felt through elevated sea levels and potentially increased storm intensity and frequencies
- Involves complex science and large amounts of diverse data
- New methods required for communicating to the public and government/business decision makers
Some questions of interest

- Do flood levels in the current climate (e.g., 100-year surge level) undergo a change in frequency such that they occur more often in a plausible future climate?
  - Under what conditions does the 100-year surge level, currently used to set flood insurance rates, become the 50-year surge level?

- Given anticipated coastal land use and development changes over the next century, under which climate scenarios are impacts expected to be greatest?
  - Which areas of the coast should expect the largest changes in risk?
ADCIRC Coastal Modeling

- Statistically meaningful 50 and 100 year flood plain mapping
  - Development of a high quality TIN

- Serial monte-carlo pre-processing step
- Hundreds of 256-way parallel ADCIRC runs
Exploring WWT for Viz and Communication

- Using the SDK, bring data into the environment from:
  - Tiled orthophotos
  - OGC Sensor Observation Service
  - OData Service
  - NetCDF
  - LIDAR data
NC Emergency Management’s Geospatial Technology Management Office

- statewide: 805 x 298 kilometers
- orthophotos 0.15 meters/pixel
- LIDAR at 3 to 6 meters/pixel
OGC service implementation
Implemented Sensor Observation Service with .Net

- .NET class library and web service template for implementation of the OGC
- http://ogc.codeplex.com
Unintended Consequences
Required programmatic access to Linux cluster for ADCIRC results

- Created an SSH Client Library for .Net
- Has become an active community, especially around providers of devices supporting SSH for command and control
- [http://sshnet.codeplex.com](http://sshnet.codeplex.com)
environmental data visualization with WWT

- model input/output
- sensed data
- image/LIDAR assets
© 2011 Microsoft Corporation. All rights reserved. Microsoft, Windows, Windows Vista and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries.

The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation.

MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.