Linkages

Collaboration underway with two University of Utah CS colleagues (Feifei Li and Jeff Phillips) on two NSF funded projects (BIGDATA: Building the Mergeable and Interactive Distributed Data Layer for Big Data Summarization Systems and CIF21 DIBBs: STORM: Spatio-Temporal Online Reasoning and Management of Large Social and Science Data). Interdisciplinary geoscience collaboration with Scripps Institution of Oceanography colleague (Frank Vernon) using atmospheric pressure data collected on a NSF-sponsored seismic array (http://meso1.chpc.utah.edu/usarray/cgi-bin/usarray_home.cgi ). MesoWest (http://mesowest.utah.edu) is an ever growing repository of environmental information collected from sensors around the nation.

Computing Innovations Having Impacts in Field

1) Interest of CS researchers to work on problems of interest to geoscientists: how to deal with and analyze heterogeneous, dynamic, "messy" data

2) Hardware and software technology improvements: ever increasing hardware capabilities (CPU speed, easier data storage solutions); access to cloud resources; increasing availability of open source solutions, github-type code repositories; improved data compression methods; noSQL data access approaches and improvements to MySQL (e.g., TokuDB).

Science Challenges

1) Growing interest to utilize crowd sourced info (http://mping.nssl.noaa.gov/ ), inexpensive sensor technologies that often rely on sensors embedded within or attached to phones (wind http://weatherflow.com/windmeter/ , pressure http://www.pressurenet.io/, sensor suites http://www.valarm.net/ ), and sensors mounted on cars and trucks (http://www.weathertelematics.com/ ). A lot of issues about sensor accuracy and quality control of the information that requires improved handling of large volumes of low quality data.

2) Appropriate uses for autonomous and unmanned sensor platforms (UAVs, etc.) for measuring environmental conditions. Besides the privacy and policy issues, how to control UAVs and collect environmental information particularly in urban areas with large swaths of restricted air space?

3) Improved data analytics for complex heterogeneous data: make smart queries to large data repositories for dynamic pattern recognition.

4) Need for CS-trained data analysts as opposed to simply having programming skills. Harder to train people to utilize machine learning well than to teach language syntax.

5) Scientific computing still relies heavily on Fortran codes implemented on large compute clusters (e.g., for numerical weather prediction).