Deriving Earth Science Data Analytics Requirements

Goal oriented Earth Science Data Analytics (ESDA) reveal requirements for needed data analytics tools/techniques

Earth Science Data Analytics: Definition
The process of examining, preparing, reducing, and analyzing large amounts of spatial (multi-dimensional), temporal, or spectral data using a variety of data types to uncover patterns, correlations and other information, to better understand our Earth.

Earth Science Data Analytics: Goals
- To glean knowledge
- To forecast/predict/model
- To derive new analytics tools
- To assess data
- To derive conclusions

Earth Science Data Analytics: Initial Requirements
To glean knowledge out of heterogeneous datasets to glean knowledge out of the data?

Types of Analytics
- Data Preparation
- Data Reduction
- Data Analysis

Tools
- R, SAS, Python, Java, C++
- SPSS, MATLAB, Mintab
- CPLEX, GAMS, Gauss
- Tableau, Spotfire
- VBA, Excel, MySQL
- Javascript, Perl, PHP
- Open Source Databases
- PIO, NCL, Parallel NetCDF
- AWS, Cloud Solutions, Hadoop
- MPI, GIS, ROI-PAC, GDAI

Techniques
- Statistics functions
- Machine Learning
- Data Mining
- Natural Language Processing
- Linear/Non-linear Regression
- Logical Regression
- Time Series Models
- Clustering
- Decision Tree
- Factor Analysis
- Principal Component Analysis
- Neural Networks
- Bayesian Techniques
- Text Analytics
- Graph Analytics
- Visual Analytics
- Map Reduce

Integrated Systems
- EarthServer (http://earthserver.eu)
- NASA Earth Exchange (https://nex.nasa.gov/nex/)
- EDEN (http://cda.ornl.gov/projects/eden/#)
- EARTHDATA (https://earthdata.nasa.gov)
- Giovanni (http://giovanni.gsfc.nasa.gov/giovanni/)


Earth Science Data Analytics: Exemplary Tools, Techniques, Integrated Systems

Methods to validate data
- To perform coarse data preparation
- To intercompare datasets
- To tease out knowledge
- To forecast/predict/model
- To derive new analytics tools

Motivation
How can we maximize the usability of large heterogeneous datasets to glean knowledge out of the data?

Methodology
Categorize/Analyze ESDA use cases; derive data analytics requirements; associate tools/techniques; perform gap analysis

Methodology
Categorize/Analyze ESDA use cases; derive data analytics requirements; associate tools/techniques; perform gap analysis

Earth Science Data Analytics: Looking Ahead
- Complete Gap Analysis between ESDA requirements and current tools/technologies
- Continue to evolve tools/techniques to address growing scope of the ‘Internet of Things’

The good news…
Earth Science Data Analytics: Preparing for the Future
Central England NERC Training Alliance

Big data analysis to fuel environmental research at Reading University

2nd Annual Graduate Workshop on Environmental Data Analytics
July 27-31, 2015

... offering degrees in Data Science
... summer school on Big Data Analytics
... online master’s degree in data analytics

Earth Science Data Analytics: Enabling Organizations

The Internet of Things connects all sources of sensing to improve decision making, presenting a new frontier of understanding our world, helping us to see more clearly than ever before.

* Thanks to the work of the Earth Science Information Partners (ESIP) Federation, Earth Science Data Analytics (ESDA) Cluster