Deriving Earth Science Data Analytics Requirements

AGU/15
IN23C
-1740

NASA/Goddard EARTH SCIENCES DATA and INFORMATION SERVICES CENTER (GES DISC)

Steve Kempler[1] , ESIP ESDA Cluster*

[1]NASA Goddard Earth Science Data and Information Services Center (GES DISC)
Steven.J.Kempler@nasa.gov

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

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

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.

Data Preparation | **Data Reduction** | **Data Analysis**

Earth Science Data Analytics: Goals

- To validate data
- To perform coarse data preparation
- To calibrate data
- To assess data quality
- To intercompare datasets
- To tease out information
- To glean knowledge
- To derive conclusions
- To forecast/ predict/model
- To derive new analytics tools

Earth Science Data Analytics: Initial Requirements

- Ingest from various sources; Homogenize data; Visualization; Sampling; Gridding
- Access large datasets; High speed processing; Subsetting, mining, machine learning
- Homogenize data; Intercomparison statistics; Pattern recognition
- Seek heterogeneous data relationships; Ingest from various sources; Image processing
- *Looking for Community input*
- Data exploration; Filter, mine, fuse, interpolate data; Manage custom code
- Ingest from various sources; High speed processing; Math functions
- Access large datasets; Assess erroneous data; Detect data anomalies
- Data exploration; Neural networks; Math/Stat modeling; Near Real Time data
- Access very large datasets; homogenize data; visualization

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

Types of Analytics	Tools	Techniques		Integrated Systems
• Data Preparation • Data Reduction • Data Analysis	• R, SAS, Python, Java, C++ • SPSS, MATLAB, Minitab • 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, GDAL	• 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	• EarthServer (http://www.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/)

Compiled from: http://practicalanalytics.co/predictive-analytics-101/ and http://cda.ornl.gov/research.shtml

Earth Science Data Analytics: Enabling Organizations

Research Data Sharing without barriers — RESEARCH DATA ALLIANCE

Federation of Earth Science Information Partners — Fostering connections to make data matter

National Institute of Standards and Technology — NIST National Institute of Standards and Technology U.S. Department of Commerce

OGC® Making location count.

The good news…

Earth Science Data Analytics: Preparing for the Future

Central England NERC Training Alliance — CENTA

Big data analysis to fuel environmental research at Reading University

2nd Annual Graduate Workshop on Environmental Data Analytics
July 27-31, 2015
Hosted by the National Center for Atmospheric Research in Boulder, CO

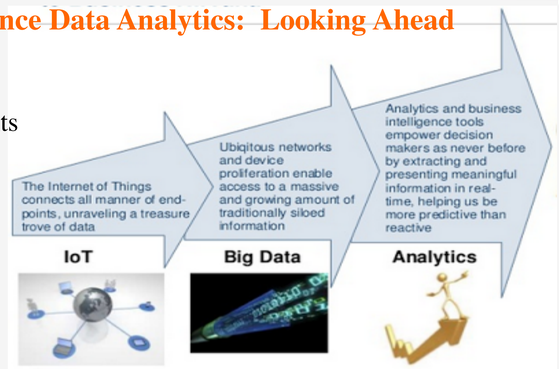
… offering degrees in Data Science

… summer school on Big Data Analytics

… online master's degree in data analytics

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'



http://hashtaggers586.blogspot.com (Source: http://iot.sys-con.com/)

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