Agenda Highlights

- Analytics and Data Scientist...in the Federation
- Other Activity Lifetimes: RDA, NIST
- Compiling use cases, analytics tools (internal and external to ESIP)
- Various guest speakers
- Cluster Information Sharing Website
- Describe/Demonstrate UV-CDAT and Climate/Network visualization analytics tools
- Use Case Information Needed/Template
- Defining, describing, and applying 5 Data Analytics Types
- Acquiring Use Case
- Planning Summer/2015 ESDA Sessions:
  - Yesterday, in case you missed it: Teaching Science Data Analytics Skills, and the Earth Science Data Scientist
  - Tomorrow, 10:30, don’t miss it

Events and Activities

- 2016-06-16: Fourteenth Telecon
- 2016-06-21: Thirteenth Telecon
- 2016-06-26: Fourteenth Telecon
- 2016-05-19: Eleventh Telecon
- 2016-05-26: Eighth Telecon
- 2016-05-02: Ninth Telecon
- 2016-04-07: January, 2016 ESDI Meeting notes (Washington), ESDI 201 Session
- 2016-01-07: January, 2016 ESDI Meeting notes (Washington), ESDI 201 Session
- 2014-11-20: Eighth Telecon
- 2014-09-20: Seventh Telecon
- 2014-08-26: Eighth Telecon
- 2014-07-20: Seventh Telecon
- 2014-06-20: Fifth Telecon
- 2014-06-22: Fourth Telecon
- 2014-05-17: Third Telecon
- 2014-03-30: Second Telecon
- 2014-02-20: First Telecon
- 2014-01-08: Initial ESDI Meeting notes

Use Case Template

- Use Case Title
- Author/Company/Email
- Actors/Workflows/Project URL and their roles and responsibilities
- Use Case Goal
  - Earth Science Data Analytics TYPES! (see below)
- Use Case Description
- Current technical considerations to take into account that may impact needed data analytics
- Data Analytics tools applied
- Data Analytics Challenges
- Type of User
- Research Areas
- Societal Benefit Areas
- Potential for and/or issues for generalizing this use case (e.g., for ref. architecture)
- More information and relevant URLs (e.g. who to contact or where to go for more information)

Analytics Tools/Techniques Examined (to mention a few)

- Dryad, MapReduce, Hadap, OpenMP, Powerset, Poreware, True Knowledge, WolframAlpha, myGrid, UV-CDAT, Climate/Network, MBIC II, CrazyEgg/Heat Maps

Presentations

- Wei Chen: NIST Big Data Public Working Group & Standardization Activities - 2/2/14
- Brand Nielsen: Surrounding Data Science and Data Analytics - 2/2/14
- Mike Schutze: MERDA Analytic Services Program at DoE Pacific Northwest National Laboratory - 1/7/15
- Paul McPherson: Data Science Masters Program at Duke University - 3/6/14
- Sarah Avison: Data Analytics Needs Scenario - 4/17/14
- Rashid Riazi: User-Directed Analytics and Tools using the Federated Data System DataFed - 4/17/14
- Tiffany Matthews: Atmospheric Science Data Center Sample Analytics Use Cases - 4/2/14
- Steve Kimplin: Analytics and Data Scientists, Earth Science Data Analytics 101 - 1/7/13
- Greg Heilman: From Maps, One (or create great precipitation data set from many good ones) - 7/17/13
- David Gaffney: Reconstructing Sea Ice Extent from Early Nimbus Satellites - 7/17/13
- Thomas Hearty: Simplifying Total Precipitable Water using AWS and MERRA - 1/7/15
- Radha Sekharay: Using Earth Observations to Understand and Predict Infectious Diseases - 7/17/13
- Tiffany Matthews: Promoting data analytics technologies - 7/17/13

Other References

- Education for Data Scientists
- Data Analytics Curriculum (example) - Data Analytics course
- Introduction to Data Science (example - online course)
- RDA Big Data Analytics Interest Group Charter
- NIST Big Data Program
- Schutze: MERDA Analytic Services paper
- Ralph Kahn, “Why we need huge datasets of Earth observations….”

Types of Earth Science Data Analytics

- To calibrate data
- To validate data (quality; note it does not have to be via data intercomparison)
- To perform course data reduction (e.g., subsetting, data viewing)
- To intercompare data (i.e., any data intercomparison; could be used to better define validation/quality)
- To derive new data product
- To derive output information from data
- To glean knowledge from data and information
- To forecast/predict phenomena (i.e., special kind of conclusion)
- To derive conclusion (i.e., that do not easily fall into another type)
- To derive analytics tools
- To recover/reduce data

Data Analytics Definitions

- The process of examining large amounts of data of a variety of types to uncover hidden patterns, unknown correlations and other useful information, involving one or more of the following:
  - Data Preparation – Preparing heterogeneous data so that they can “play” together
  - Data Reduction – Smartly removing data that do not fit research criteria
  - Data Analysis – Applying techniques/methods to derive results

Use Case Template

- Use Case Title
- Author/Company/Email
- Actors/Workflows/Project URL and their roles and responsibilities
- Use Case Goal
  - Earth Science Data Analytics TYPES! (see below)
- Use Case Description
- Current technical considerations to take into account that may impact needed data analytics
- Data Analytics tools applied
- Data Analytics Challenges
- Type of User
- Research Areas
- Societal Benefit Areas
- Potential for and/or issues for generalizing this use case (e.g., for ref. architecture)
- More information and relevant URLs (e.g. who to contact or where to go for more information)

Analytics Tools/Techniques Examined (to mention a few)

- Dryad, MapReduce, Hadap, OpenMP, Powerset, Poreware, True Knowledge, WolframAlpha, myGrid, UV-CDAT, Climate/Network, MBIC II, CrazyEgg/Heat Maps

Presentations

- Wei Chen: NIST Big Data Public Working Group & Standardization Activities - 2/2/14
- Brand Nielsen: Surrounding Data Science and Data Analytics - 2/2/14
- Mike Schutze: MERDA Analytic Services Program at DoE Pacific Northwest National Laboratory - 1/7/15
- Paul McPherson: Data Science Masters Program at Duke University - 3/6/14
- Sarah Avison: Data Analytics Needs Scenario - 4/17/14
- Rashid Riazi: User-Directed Analytics and Tools using the Federated Data System DataFed - 4/17/14
- Tiffany Matthews: Atmospheric Science Data Center Sample Analytics Use Cases - 4/2/14
- Steve Kimplin: Analytics and Data Scientists, Earth Science Data Analytics 101 - 1/7/13
- Greg Heilman: From Maps, One (or create great precipitation data set from many good ones) - 7/17/13
- David Gaffney: Reconstructing Sea Ice Extent from Early Nimbus Satellites - 7/17/13
- Thomas Hearty: Simplifying Total Precipitable Water using AWS and MERRA - 1/7/15
- Radha Sekharay: Using Earth Observations to Understand and Predict Infectious Diseases - 7/17/13
- Tiffany Matthews: Promoting data analytics technologies - 7/17/13

Other References

- Education for Data Scientists
- Data Analytics Curriculum (example) - Data Analytics course
- Introduction to Data Science (example - online course)
- RDA Big Data Analytics Interest Group Charter
- NIST Big Data Program
- Schutze: MERDA Analytic Services paper
- Ralph Kahn, “Why we need huge datasets of Earth observations….”

Types of Earth Science Data Analytics

- To calibrate data
- To validate data (quality; note it does not have to be via data intercomparison)
- To perform course data reduction (e.g., subsetting, data viewing)
- To intercompare data (i.e., any data intercomparison; could be used to better define validation/quality)
- To derive new data product
- To derive output information from data
- To glean knowledge from data and information
- To forecast/predict phenomena (i.e., special kind of conclusion)
- To derive conclusion (i.e., that do not easily fall into another type)
- To derive analytics tools
- To recover/reduce data

Data Analytics Definitions

- The process of examining large amounts of data of a variety of types to uncover hidden patterns, unknown correlations and other useful information, involving one or more of the following:
  - Data Preparation – Preparing heterogeneous data so that they can “play” together
  - Data Reduction – Smartly removing data that do not fit research criteria
  - Data Analysis – Applying techniques/methods to derive results

Use Case Template

- Use Case Title
- Author/Company/Email
- Actors/Workflows/Project URL and their roles and responsibilities
- Use Case Goal
  - Earth Science Data Analytics TYPES! (see below)
- Use Case Description
- Current technical considerations to take into account that may impact needed data analytics
- Data Analytics tools applied
- Data Analytics Challenges
- Type of User
- Research Areas
- Societal Benefit Areas
- Potential for and/or issues for generalizing this use case (e.g., for ref. architecture)
- More information and relevant URLs (e.g. who to contact or where to go for more information)

Analytics Tools/Techniques Examined (to mention a few)

- Dryad, MapReduce, Hadap, OpenMP, Powerset, Poreware, True Knowledge, WolframAlpha, myGrid, UV-CDAT, Climate/Network, MBIC II, CrazyEgg/Heat Maps