Mission:
To promote a common understanding of the usefulness of, and activities that pertain to, Data Analytics and more broadly, the Data Scientist; Facilitate collaborations to better understand the cross usage of heterogeneous datasets and to provide accommodating data analytics expertise, now and as the needs evolve into the future; Identify gaps that, once filled, will further collaborative activities.

Objectives:
• Provide a forum for ‘Academic’ discussions that provides ESIP members a better understanding of the various aspects of Earth Science Data Analytics
• Bring in guest speakers to describe external efforts, and further teach us about the broader use of Data Analytics.
• Perform activities that:
  - Compile use cases generated from specific community needs to cross analyze heterogeneous data
  - Compile sources of analytics tools, in particular, to satisfy the needs of the above data users
  - Examine gaps between needs and sources
  - Examine gaps between needs and community expertise
• Document specific data analytics expertise needed to perform Earth science data analytics
• Seek graduate data analytics/ Data Science student internship opportunities

Agenda Highlights
• Analytics and Data Scientist...in the Federation
• Other Activity Briefings: RDA, NIST
• NIST Big Data Program
• RDA Big Data Analytics Interest Group Charter
• ESIP Earth Sciences Data Analytics (ESDA) Cluster – Work in Progress
• Education for Data Scientists
• User-Oriented Data Analytics and Tools using the Federated Data System DataFed
• From Many, One (or creating one great precipitation data set from many good ones)
• Analytics. The Need for Earth Science Data Analytics to Facilitate
  - Deriving conclusions (i.e., that do not easily fall into another type)
  - To glean knowledge from data and information
  - To detect differences
  - To discern an event
  - To understand data
  - To ‘play’ together
  - To reduce inter-calibrations among datasets
  - To inter-calibrate datasets to produce a single source

Types of Earth Science Data Analytics

<table>
<thead>
<tr>
<th>Use Case</th>
<th>Data Description/Use Case</th>
<th>Use Case Goal</th>
<th>Data Analytics Challenges</th>
<th>Related Tools/Techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example 1:</td>
<td>Two events are occurring simultaneously in different regions</td>
<td>To calibrate data</td>
<td>- External dependencies</td>
<td>RDA, MapReduce, Hadoop, OpenCyc, Powerset, True Knowledge, ROI_PAC, GDAL, MeraXon, MapReduce, MPI, GIS, PDO, NCL, NCO, parallel NetCDF, scheduler, SQL</td>
</tr>
<tr>
<td>Example 2:</td>
<td>Data is coming in at a high rate of speed</td>
<td>To validate data</td>
<td>- High volume of data</td>
<td>RDA, MapReduce, Hadoop, OpenCyc, Powerset, True Knowledge, ROI_PAC, GDAL, MeraXon, MapReduce, MPI, GIS, PDO, NCL, NCO, parallel NetCDF, scheduler, SQL</td>
</tr>
<tr>
<td>Example 3:</td>
<td>Data is coming in from a variety of sources</td>
<td>To perform data integration</td>
<td>- Data heterogeneity</td>
<td>RDA, MapReduce, Hadoop, OpenCyc, Powerset, True Knowledge, ROI_PAC, GDAL, MeraXon, MapReduce, MPI, GIS, PDO, NCL, NCO, parallel NetCDF, scheduler, SQL</td>
</tr>
</tbody>
</table>

Use Case Template

- **Use Case Title:** Earth Science Data Analytics Types Example
- **Description:**
  - **Use Case Goal:** To create a comprehensive dataset for climate change analysis.
  - **Data Analytics Challenge:** Data heterogeneity and volume.
  - **Related Tools/Techniques:** MapReduce, Hadoop, OpenCyc, Powerset, True Knowledge, ROI_PAC, GDAL, MeraXon, MapReduce, MPI, GIS, PDO, NCL, NCO, parallel NetCDF, scheduler, SQL.
- **Data Description/Use Case:**
  - **Example:** Data is coming in from various sensors and remote sensing platforms.
  - **Related Tools/Techniques:** RDA, MapReduce, Hadoop, OpenCyc, Powerset, True Knowledge, ROI_PAC, GDAL, MeraXon, MapReduce, MPI, GIS, PDO, NCL, NCO, parallel NetCDF, scheduler, SQL.

Data Analytics Definition:
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

Case Study:
- **Use Case Title:** Earth Science Data Analytics Types Example
- **Description:**
  - **Use Case Goal:** To create a comprehensive dataset for climate change analysis.
  - **Data Analytics Challenge:** Data heterogeneity and volume.
  - **Related Tools/Techniques:** MapReduce, Hadoop, OpenCyc, Powerset, True Knowledge, ROI_PAC, GDAL, MeraXon, MapReduce, MPI, GIS, PDO, NCL, NCO, parallel NetCDF, scheduler, SQL.
- **Data Description/Use Case:**
  - **Example:** Data is coming in from various sensors and remote sensing platforms.
  - **Related Tools/Techniques:** RDA, MapReduce, Hadoop, OpenCyc, Powerset, True Knowledge, ROI_PAC, GDAL, MeraXon, MapReduce, MPI, GIS, PDO, NCL, NCO, parallel NetCDF, scheduler, SQL.