

About The Project

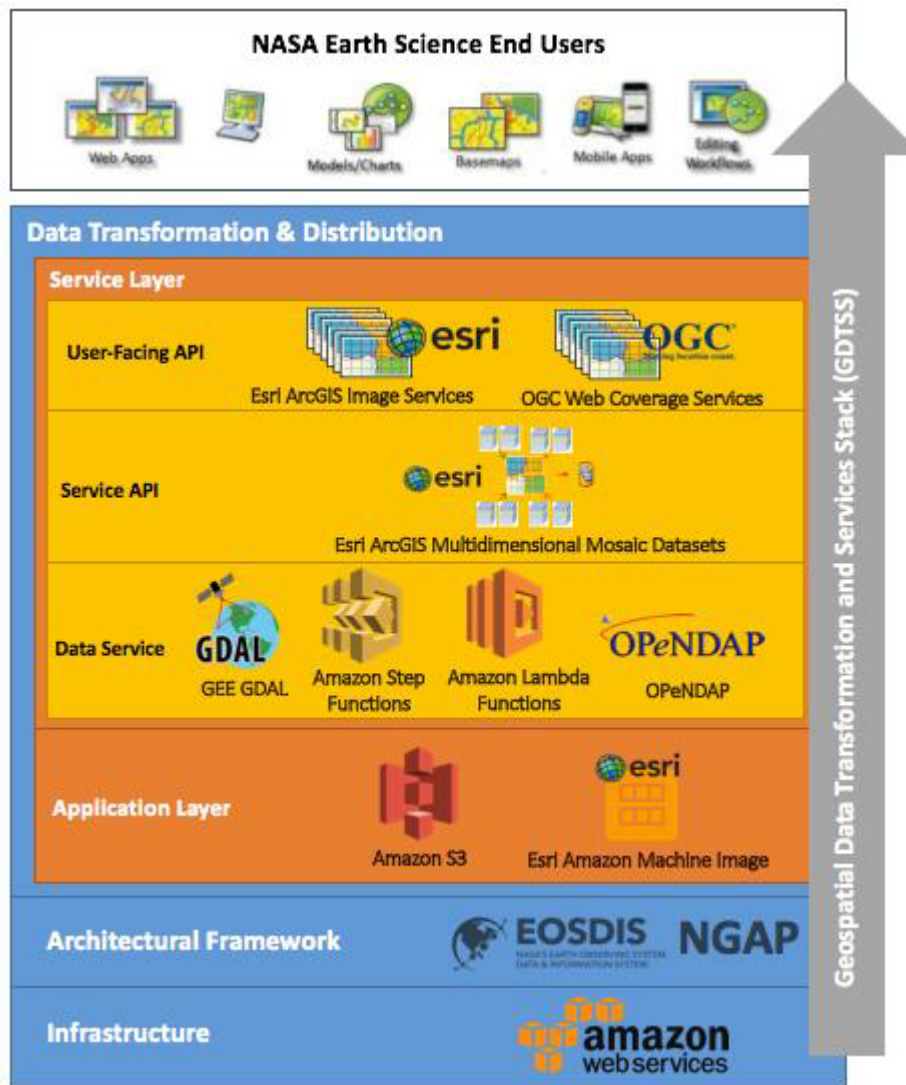
Systematic Data Transformation to Enable Web Coverage Services (WCS) and ArcGIS Image Services within ESDIS Cumulus Cloud

Objective

- Goal: Develop a systematic data transformation methodology to enable Web Coverage Services (WCS) and ArcGIS Image Services within ESDIS [Cumulus](#) Cloud.
- Project metrics
 - How will you measure the success of the project
 - Number of tools (e.g., QGIS, ArcGIS Desktop) and services (e.g., ArcGIS Server, Geoserver)
 - Number of data products (e.g., MOPITT, CERES, TES, SMAP, ICESat-2)
 - Number of users (e.g., EOSDIS metrics, web service logs analytics, [NPS survey](#))
 - Time saved (e.g., Cloud vs. Local data transformation performance, data access performance)
 - How will your project improve access to NASA data
 - Data centers can easily transform data instantly when they publish data on cloud. As soon as [Cumulus](#) publishes data on S3, we would like to transform HDF4/HDF5 data into GeoTIFF and cloud-optimized GeoTIFF so that users can access data easily with GIS tools including network data access (OPeNDAP/WMS/WCS) client and other NGAP services/tools.
 - Users can access data at scale from cloud in desired formats and services. Once GeoTIFF conversion is successful via Lambda, we can extend the framework to transform HDF5 into different formats (e.g., parquet / json) that native AWS solution (e.g., Elastic search / Athena / Rekognition / Alexa / QuickSight) can easily handle as well.

Approach

- **Explain your approach to the project.**
 - **Include major project elements.**
1. Develop geospatial data transformation plugins using the Geospatial Data Abstraction Library (GDAL) Enhancements for ESDIS (GEE) software within the existing NASA ESDIS Cumulus cloud environment.
 2. Correct sample MOPITT and CERES data products in the ESDIS Cumulus cloud environment using the newly created plugins and serve them out as OGC Web Coverage Services (WCS) and Esri ArcGIS Image Services.
 3. Demonstrate the potential performance gains by leveraging a cloud optimized data format and the reproducible process of transforming data to be geospatially enabled and accessible in commercial off-the-shelf and open source GIS software and online data catalogs.



Geospatial Data Transformation and Services Stack (GDTSS)

Co-Investigators:

- A. Jason Barnett, Matthew Tisdale, Booz Allen Hamilton (BAH), NASA LaRC Atmospheric Science Data Center (ASDC)
- Hyo-Kyung (Joe) Lee, The HDF Group

Collaborators

- Dr. Daniel Ziskin, MOPITT Data Manager, NCAR – Atmospheric Chemistry Observations & Modeling Laboratory
- Dr. Kathleen Moore, CERES Deputy Data Management Team Lead, NASA LaRC, Climate Science Branch
- Tom Maersperger, NASA LP DAAC Scientist, NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS
- John Kusterer, ASDC Head, NASA LaRC Atmospheric Science Data Center (ASDC)
- Joseph Koch, ASDC GIS User Services, NASA LaRC Atmospheric Science Data Center (ASDC), BAH
- Dr. Paul W. Stackhouse, Jr., NASA Senior Research Scientist, Prediction Of Worldwide Energy Resources (POWER)
- Tyler Bristow, Bradley Macpherson, Geospatial Analysis Developers, Prediction Of Worldwide Energy Resources (POWER), BAH
- Brian Tisdale, Geospatial Web Services Working Group (GWSWG) Co-Chair, BAH
- Leah Schwizer, NASA Disasters GIS Strategy and Governance, BAH

Proposal Summary

- The NASA Earth Science Data and Information System (ESDIS) Project has recently launched the Cumulus prototype which provides a scalable cloud-based platform to ingest, archive, distribute, and manage Earth Science data within the Amazon cloud environment ("Earth Science Data," 2017). One of the goals for the Cumulus project is to increase the flexibility and effectiveness of NASA Earth Science data access and distribution to end users. Earth science data is geospatial in nature. However, according to the Geospatial Data Abstraction Library (GDAL) Enhancements

for ESDIS (GEE) Assessment ("GDAL Enhancements," 2016), many Earth science data products (e.g. MOPITT, CERES, to name a few) are difficult to access and use within commercial off-the-shelf (COTS) and open source Geographic Information Systems (GIS) software, such as Esri's ArcGIS and the open source QGIS software.

- According to the 2017 American Customer Satisfaction Index (ACSI) survey, ArcGIS is the most used software tool/package at 64%, followed by QGIS (37%), ENVI (32%) and Excel (27%) to work with NASA Earth science data ("ACSI Reports," 2018). We propose to develop geospatial data transformation plugins that could be used within the ESDIS Cumulus environment to serve out transformed MOPITT and CERES data product(s) as OGC Web Coverage Services (WCS) and Esri ArcGIS Image Services. These services will then be easily consumed into COTS GIS software such as ArcGIS and QGIS. These plugins will perform the transformation to fix the data issues (e.g. incorrect image sizes, orientation, multidimensional variable interpretation, georeferenced metadata recognition, etc.) and this extensible framework can be further expanded to add new plugins as additional issues are identified and addressed. In addition, ArcGIS Image Services and Web Coverage Services provide easy access to actual data values, not just static images, and alleviate the need to navigate complex and storage-intensive processes of raw data downloads. The importance of delivering geospatially enabled web services is emphasized through initiatives like the establishment of the NASA Earth Science Data System (ESDS) Geospatial Web Services Working Group (GWSWG), the NASA Big Earth Data Initiative (BEDI), and the ESDIS Cumulus Seamless 360 Degrees of Services that includes the implementation of a common user-facing API such as WCS. By developing these geospatial data transformation plugins and subsequently providing the corrected data as OGC and Esri web services, it will remove the barrier of transforming each NASA data product after download since it will be correctly served, leading to more easily accessible NASA data products by the Earth Science community.
- This proposal will utilize Amazon Web Services (AWS) Step Functions and Lambda Functions as also utilized by the ESDIS Cumulus prototype, though if the ESDIS Cumulus system is not operationalized, our project can be decoupled and operate independently within a commercial cloud environment. We will approach this problem in three phases. Phase one, planning, will focus on the AWS workflow structure. Phase two, development and implementation, will deploy the AWS Step Functions and Lambda Functions used to orchestrate a workflow of customized micro-services executing GDAL transformations in order to geospatially enable and serve a new cloud-optimized MetaRaster Format (MRF) product as an OGC Web Coverage Service and ArcGIS Image Service. Phase three, testing, will allow us to measure the discoverability of NASA Earth Science data web services within major data catalogs (e.g. Earthdata Search, GeoPlatform, ASDC ArcGIS Portal, etc.), and their use, speed, accessibility and analytical capabilities within QGIS, ArcGIS, custom web mapping applications, and Data Cubes.