New Features

- A Python library (CDSlib) that supports full temporal, spatial, and grid-based resolution services
- A new reanalysis collections reference model to enable operator design and implementation
- An enhanced library of sample queries to demonstrate and use case scenarios
- Extended operators that enable single- and multiple-reanalysis area average, vertical average, re-gridding, and trend, climatology, and anomaly computations
- Full support for the MERRA-2 reanalysis and the initial integration of additional reanalyses (CFSR, ECMWF, 20CR, JRS-55)
- A Jupyter notebook-based distribution mechanism that combines CDSlib documentation with interactive use case scenarios and personalized project management
- Prototyped uncertainty quantification services that combine ensemble products with comparative observational products
- Convenient, one-stop shopping for commonly used data products from multiple reanalyses, including basic subsetting and arithmetic operations over the data and extractions of trends, climatologies, and anomalies

Virtual Real-Time Analytics Testbed

- Powerful computing resources are necessary for on-demand analytic processing across 34+ years of reanalysis data.
- The MapReduce operations currently leverage a cluster consisting of 36 Dell R720 servers each populated with six 2-terabyte hard drives in addition to their operating system disks ( CentOS 6.8).
- For interconnectivity, there is a 36-port 56 Gbps Ethernet switch plus a 48-port Gigabit switch. Initial cluster metrics show peaks of 314 gigaflops/second and an overall capability of approximately 11 teraflops.
- This testbed, which will ultimately replace the current cluster, will incorporate new technologies (e.g., alternate software stacks, parallel HDFS file systems).

ABSTRACT

We are extending our Climate Analytics-as-a-Service (CAaaS) capabilities to include the following: (1) A Reanalysis Ensemble Service (RES) offering a basic set of commonly used operations over multiple reanalysis collections that are accessible through NASA’s climate data analytics web services and our client-side Climate Data Services Python library, CDSlib; (2) a high-performance Virtual Real-Time Analytics Testbed supporting multiple major datasets using advanced technologies like Apache Spark and Hadoop-based MapReduce analytics over native NetCDF files; and (3) an Open Geospatial Consortium (OGC) WPS-compliant web service interface to CDSlib to accommodate the Earth System Grid Federation (ESGF) web service endpoints.

Example RES Workflow: Uncertainty Quantification

Uncertainty analysis pervades all levels of climate impact assessments and climate data analysis. Preparing data for uncertainty quantification is a particularly time-consuming step. The Reanalysis Ensemble Service (RES) is an effort to make data quickly and easily ready for assessment. RES provides an Uncertainty Quantification Package (UQP) that contains the requested RES products along with context-sensitive peer products that can be used to characterize uncertainty. The following example demonstrates a possible use of a UQP for evaluating global precipitation among multiple reanalysis datasets. The workflow consists of the following tasks:
1) Calculate global average precipitation from five reanalysis collections (MERRA2, CFSR, JRA-55, NOAA 20CRv2c, and ERA-Interim).
2) Calculate global average precipitation from two observed climate records (CMAP and GPCP).
3) Calculate global ensemble average for the reanalysis collections.
4) Generate line plots to show long-term variability of global precipitation in each dataset.
5) Generate a Taylor diagram to summarize the similarity of global precipitation pattern between CMAP and other datasets.

Reanalysis Ensemble Service - Notional Architecture

- External Data Services
- MapReduce
- Data Nodes
- No. 1
- No. 2
- ... No. 10
- Data Nodes
- Metadata
- Data Analysis
- Processing
- Data
- Providers
- Other
- Geo spatial
- Service
- Library
- Persistence
- Transformation
- Utility Services
- Analytics Services
- REanalysis Ensembl

For Additional Information

- glenn.s.tamkin@nasa.gov
- john.l.schnase@nasa.gov
- daniel.q.duffy@nasa.gov