New Era, New Opportunity, Is GES DISC Ready for Big Data Challenge?

A. Li1, C. Shie1,2, J. Wei1, L. Pham1 and D. Meyer1

Abstract

The new era of Big Data has opened doors for many new opportunities, as well as new challenges, for both Earth science research/application and data communities. As one of the twelve NASA data centers - Goddard Earth Sciences Data and Information Services Center (GES DISC), one of our great challenges has been how to help research/application community efficiently (quickly and properly) accessing, visualizing and analyzing the massive and diverse data in natural hazard research, management, or even prediction. GES DISC has archived over 200TB data on premises and distributed over 23,000 TB of data since 2010. Our data has been widely used in every phase of natural hazard management and research, I.e., long term risk assessment and reduction, forecasting and predicting, monitoring and detection, early warning, damage assessment and response.

The big data challenge is not just about data storage, but also about data discoverability and accessibility, and even more, about data migration/mirroring in the cloud. This paper is going to demonstrate GES DISC’s efforts and approaches of evolving our overall accessibility, and even more, about data migration/mirroring in the cloud. The new era of Big Data has opened doors for many new opportunities, as well as new challenges, for both Earth science research/application and data communities. As one of the ESSDIC project for “native” cloud-based data ingest, archive, distribution and management system. Benefit: Scalable performance - Cloud based data ingest, archive, distribution and management system, fast data access - 20% reduction in access times for some services.

Cloud Data Migration - Cumulus

• Cumulus – ESSDIC project for “native” cloud based data ingest, archive, distribution and management system.
• Benefit: Scalable performance
• Cloud based data ingest, archive, distribution and management system, fast data access - 20% reduction in access times for some services.

Cloud Data Services - Harmony

• Increase usage and ease of use of EOSDIS data.
• Focus on opportunities when multiple DAAC’s data all exist in the Earthdata Cloud.
• Users can work seamlessly with data from different DAACs in ways previously not possible.
• GES DISC is evaluating services to migrate to Harmony and expect to have some services in operation later in 2020.

Conclusions

• GES DISC is moving high-value data into AWS to test core archival functions and cloud analytics.
• For more information on GES DISC cloud efforts, please contact the author.

Acronyms

ESDIS - Earth Science Data and Information System
MERRA2 - Modern-Era Retrospective Analysis for Research and Applications, Version 2
OMI – Orbiting Carbon Observatory
OCO – Orbiting Carbon Observatory

DAAC - Data Active Archive Center
AIRS – Atmospheric Infrared Sounder
TROPOMI – Tropospheric Monitoring Instrument
EOSDIS – Earth Observing System Data and Information System

Related Poster

• IN12B-06 - Metrics Learning at NASA GES DISC
• IN12B-14 - Integrated Analysis of Multiple User Metrics - A “Sequel”, and Introducing the Google Analytics
• IN13B-0708 - Cloud Giovanni: Raining in Costs and Improving Performance with Analytical Data Stores using Scalable Serverless Architecture

Authors

1 NASA Goddard Space Flight Center
2 University of Maryland, Baltimore County

http://disc.gsfc.nasa.gov