



Building a cloud based distributed active archive data center

Rahul Ramachandran (1), Katie Baynes (2), and Kevin Murphy (3)

(1) NASA/MSFC, Huntsville, United States (Rahul.Ramachandran@nasa.gov), (2) NASA/GSFC, Greenbelt, MD, United States (kathleen.baynes@nasa.gov), (3) NASA/HQ, Washington, DC, United States (kevin.j.murphy@nasa.gov)

NASA's Earth Science Data System (ESDS) Program serves as a central cog in facilitating the implementation of NASA's Earth Science strategic plan. Since 1994, the ESDS Program has committed to the full and open sharing of Earth science data obtained from NASA instruments to all users. One of the key responsibilities of the ESDS Program is to continuously evolve the entire data and information system to maximize returns on the collected NASA data. An independent review was conducted in 2015 to holistically review the EOSDIS in order to identify gaps. The review recommendations were to investigate two areas: one, whether commercial cloud providers offer potential for storage, processing, and operational efficiencies, and two, the potential development of new data access and analysis paradigms. In response, ESDS has initiated several prototypes investigating the advantages and risks of leveraging cloud computing.

This poster will provide an overview of one such prototyping activity, "Cumulus". Cumulus is being designed and developed as a "native" cloud-based data ingest, archive and management system that can be used for all future NASA Earth science data streams. The long term vision for Cumulus, its requirements, overall architecture, and implementation details, as well as lessons learned from the completion of the first phase of this prototype will be covered. We envision Cumulus will foster design of new analysis/visualization tools to leverage collocated data from all of the distributed DAACs as well as elastic cloud computing resources to open new research opportunities.