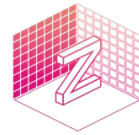




GES DISC Collaborations with Open-Source Communities to Migrate Data Collections to Cloud-friendly



Zarr Stores

Dieu My Nguyen, Hailiang Zhang, Brianna Pagán,
Christine Smit, Mahabal Hegde

*Software for the NASA Science Mission Directorate
Workshop – NASA HQ – May 7-9, 2024*

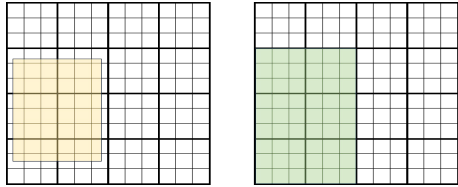


tinyurl.com/gesdisc-slides

dieumy.t.nguyen@nasa.gov



A cloud-friendly file storage format for chunked, compressed, N-dimensional arrays based on an open-source specification.



Desired selection Corresponding chunks

Our collaborations with open source software development and communities:

1. Leading the community development process for GeoZarr, an Open Geospatial Consortium (OGC) specification to extend Zarr for storing geospatial observations.
2. Exploring solutions for managing and keeping up-to-date dynamic data. A potential solution is the generic lakeFS platform for git-like data version control.
3. Developing a Zarr extension for fast and cost-efficient data analysis by pre-computing and storing cumulative sums at the chunk level.

