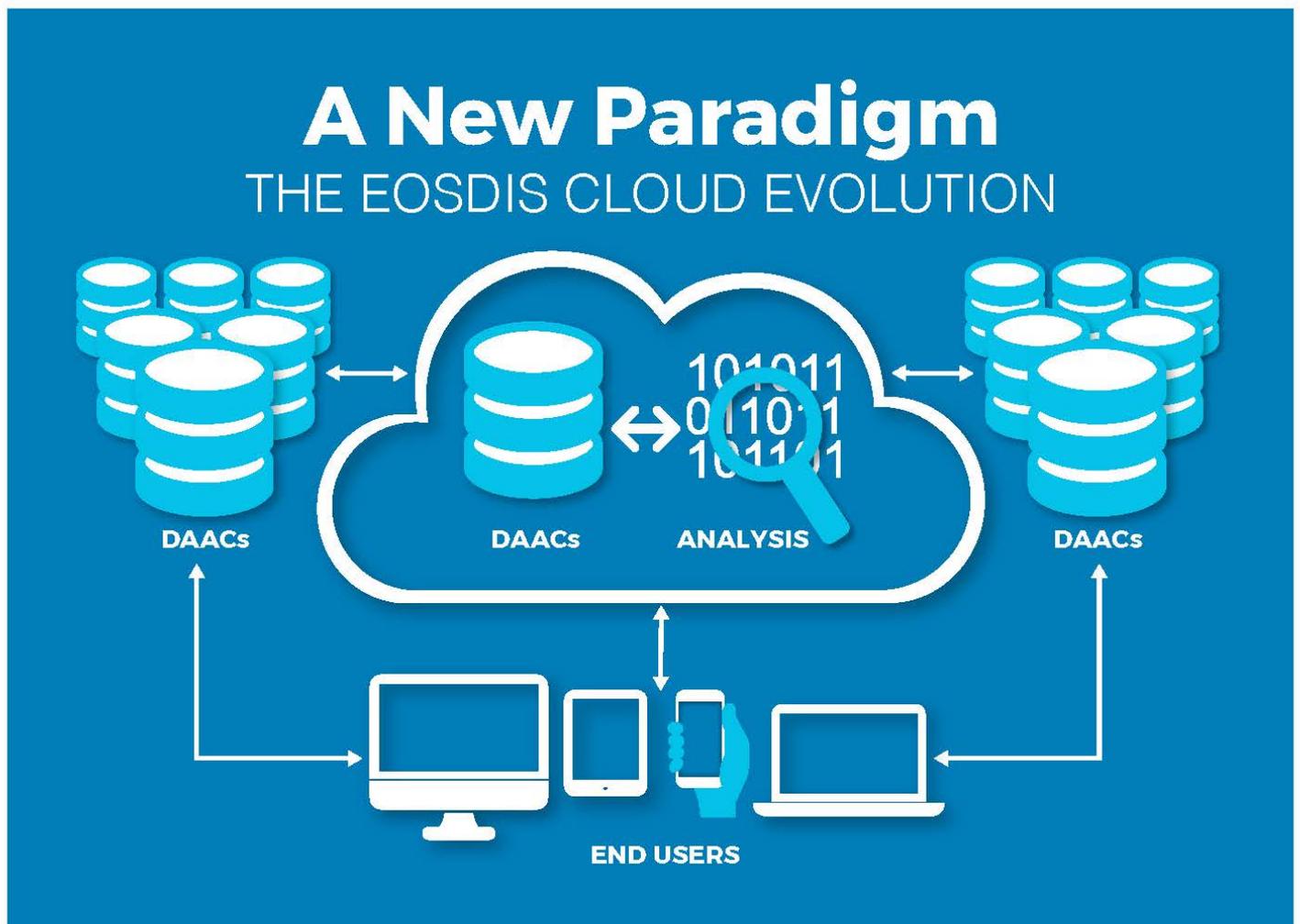


Getting Ready for NISAR

Getting Ready for NISAR — and for Managing Big Data using the Commercial Cloud

The upcoming NISAR satellite mission is expected to add as much as 85 TB of data each day to the EOSDIS archive. The commercial cloud is being explored as a way to archive and disseminate this extremely high volume of data efficiently.

By Josh Blumenfeld, EOSDIS Science Writer



(https://cdn.earthdata.nasa.gov/conduit/upload/6695/EOSDIS_ExCEL_graphicV3a1.jpg)

The launch of the upcoming joint NASA/Indian Space Research Organisation (ISRO) Synthetic Aperture Radar (NISAR) (<https://nisar.jpl.nasa.gov/>) mission, currently scheduled for late 2021, will be a landmark undertaking. NISAR is not only the first joint mission between NASA(<https://www.nasa.gov/>) and ISRO(<http://www.isro.gov.in/>) scheduled to launch (the two organizations also have an agreement(<https://www.nasa.gov/press/2014/september/us-india-to->

collaborate-on-mars-exploration-earth-observing-mission/) to work on joint Mars exploration missions), but also the launch of the first dual-frequency synthetic aperture radar (SAR)(/asf-tutorials/sar-faq/). The data collected by the L-band(https://en.wikipedia.org/wiki/L_band) (produced by NASA) and S-band(https://en.wikipedia.org/wiki/S_band) (produced by ISRO) SAR systems aboard the NISAR satellite and processed into cloud-free, ultra-sharp imagery will facilitate cutting-edge research into some of the planet's most complex processes, including ecosystem disturbances, ice-sheet dynamics, earthquakes, tsunamis, volcanoes, and landslides.

“NISAR will produce copious amounts of data and these data will be in high demand, not only for the NISAR products themselves, but as ingredients in the generation of multiple higher-level informational products,” says NISAR Program Scientist Craig Dobson. “The global scope of NISAR science combined with NASA’s open data policy(<https://earthdata.nasa.gov/nasa-data-policy>) will stimulate and facilitate vast interest in these data.”



(https://cdn.earthdata.nasa.gov/conduit/upload/6968/NISAR_Mission_Logo.png)

As Dobson notes, NISAR is expected to generate a tremendous volume of data over its scheduled three-year mission — as much as 140 petabytes (PB). In comparison, the **total** volume of data in NASA’s Earth Observing System Data and Information System (EOSDIS)(<https://earthdata.nasa.gov/about>) archive at the beginning of 2017 was about 22 PB, according to metrics from NASA’s Earth Science Data Systems (ESDS) Program(<https://earthdata.nasa.gov/earth-science-data-systems-program>) . In fact, NISAR is expected to generate a data volume close to 85 terabytes (TB) each day. This is much greater than any currently operating NASA Earth observing mission. “This places considerable demands on the logistics of shipping data and on computational speed and efficiency,” Dobson says.

“NISAR has a two-part problem,” observes Chris Stoner, the Project Office Manager at NASA’s Alaska Satellite Facility (ASF) Distributed Active Archive Center (DAAC), which is one of EOSDIS’ discipline-specific(<https://earthdata.nasa.gov/discipline>) DAACs(<https://earthdata.nasa.gov/about/daacs>) and the future home for NASA’s NISAR data. “The file sizes are large and the overall volume of data will be huge. This means we have to do something different to ensure that NISAR data users have a good user experience and are able to do their research.”

The “different” approach ASF DAAC is exploring to efficiently store and distribute the tremendous amount of data expected from NISAR is to use the commercial cloud. While the ASF DAAC will archive and distribute NISAR data, these data will be processed at NASA’s Jet Propulsion Laboratory (JPL)(<https://www.jpl.nasa.gov/>) , located at the California Institute of Technology(<http://www.caltech.edu/>) in Pasadena, CA. ASF DAAC is working collaboratively with JPL to test and prototype ways of archiving and distributing NISAR data using the commercial cloud. This three-year project began in 2016 and is called Getting Ready for NISAR, or GRFN (pronounced *Griffin*).

Now in its second year, GRFN is part of the ongoing EOSDIS efforts to evolve NASA Earth observing data and EOSDIS services to the commercial cloud(<https://earthdata.nasa.gov/about/eosdis-cloud-evolution>) . The primary GRFN goals are to:

1. Obtain a better understanding of the costs and technical challenges associated with cloud-based and hybrid architectures for processing and storing NISAR data, and
2. Provide the opportunity for the science community to become comfortable interacting with and working with large SAR datasets in the cloud.

Since NISAR data will not be available until after launch, SAR data from the European Space Agency’s (ESA) (<http://www.esa.int/ESA>) Sentinel-1(http://ec.europa.eu/growth/sectors/space/copernicus_en) mission (which are archived and distributed by the ASF DAAC) are being used as a surrogate to prototype a cloud-based system using Amazon Web Services (AWS)(<https://aws.amazon.com/>) . AWS is currently the only commercial cloud provider approved by NASA for NASA data(<https://www.nasa.gov/press/2013/november/nasa-brings-earth-science-big-data-to-the-cloud-with-amazon-web-services/#.WdJNd9Frzcu>) .

Stoner notes that the huge size of NISAR files is a primary reason for using the commercial cloud for processing, archiving, and distributing mission data. Modern SAR sensors produce extremely high-resolution images, day or night, without the need for outside illumination (such as from the sun). By comparing SAR imagery from different days, subtle changes easily can be observed and measured, such as uplift from earthquakes or subsidence from excessive groundwater pumping.

The trade-off for this detailed imagery, though, is extremely large data files necessitating high data transfer rates. According to the ASF DAAC, Sentinel-1 data products average 5 gigabytes (GB) per frame. NISAR data products will be much larger than Sentinel-1 products, and average 25 GB per frame. According to Stoner, users of Sentinel-1 data typically download hundreds or sometimes thousands of scenes for their research. The time researchers need to download data is time away from conducting research using the data.

“The difference between a 5 gigabyte Sentinel file and a 25 gigabyte NISAR file is significant,” says Stoner. “Researchers on Sentinel-1 now are asking us if there are alternatives to downloading a bunch of files because it takes a while. Well, ‘it takes a while’ is going to evolve for NISAR to ‘I cannot work this way.’ It will simply take too long to process NISAR data using conventional processing technology.”

The commercial cloud provides an elegant alternative to downloading huge data files. In a cloud-based system, researchers have the ability to work with data directly in the cloud and only have to download the finished, derived product. “Data users can bring their algorithms to the cloud and process [data] next to data storage; no downloads,” Stoner says, noting that while the derived product still could be very large in size, a researcher will not have to wait to download all the individual files necessary to produce the finished product.

During the first year of GRFN, JPL built a prototype science data system (SDS) environment in AWS to generate Sentinel-1 interferograms(</asf-tutorials/data-recipes/insar-using-sentinel-1-toolbox/>) and deliver these to the ASF DAAC. Meanwhile, the ASF DAAC built a prototype system to manage all aspects of the data lifecycle for these products in AWS, including data ingest, storage, discovery, distribution, and on-demand product generation. This not

only allows data users to become familiar with using SAR products in the cloud, but, by analyzing how these products are being used, also provides a better understanding of the associated costs of cloud-based storage and dissemination of these data.

GRFN is simulating several processing scenarios applicable to SAR data and the NISAR mission:

- Forward stream processing, or “keeping up” with normal data flow from the satellite
- Bulk reprocessing
- On-demand processing
- Urgent response (such as after a tsunami, an earthquake, or other natural event)

The GRFN team already has demonstrated sustained processing speeds of 10 gigabits per second (Gbps), which is the speed required for NISAR forward stream processing. The next objective will be in achieving 40 gigabits per second, which is the NISAR bulk reprocessing rate. The goal, says Stoner, is 50 gigabits per second, which is the rate required for receiving both forward and bulk processing. Processing rates this fast will occur when the satellite is sending data to the SDS for processing while the SDS is coincidentally reprocessing a year’s worth of data and sending these data to the ASF DAAC as rapidly as possible.

“Right now, we’ve achieved very fast processing, and it’s going to get faster,” Stoner says. “Our challenge is to keep up with the data processing needs while keeping the whole process cost-effective. Overall, the cloud shouldn’t be excessively expensive [compared with more traditional data processing and storage methods]. The cloud should be cost effective, especially at scale.”

Although only in its second year, GRFN already is successfully demonstrating the viability of using the commercial cloud for SAR data missions, with tools and techniques that easily can be applied to similar high-volume data missions. “There is certainly a time constraint to research, and we’re trying to react to that. Cloud computing really answers a lot of the problems [in dealing with big data] and it offers probably a better and a broader opportunity,” says Stoner.

By laying the groundwork for processing, archiving, and disseminating mission data using the commercial cloud now, well in advance of the scheduled NISAR launch, JPL, the ASF DAAC, NASA’s EOSDIS, and data users will be better prepared to maximize the tremendous amount of data expected from this groundbreaking mission.



About Us

The Alaska Satellite Facility downlinks, processes, archives, and distributes remote-sensing data to scientific users around the world. ASF's mission is to make remote-sensing data accessible.

Alaska Satellite Facility

 2156 Koyukuk Drive
Fairbanks, AK 99775

 (907) 474-5041

 [Contact Us](#)

*UA is an AA/EEO employer and educational institution and prohibits illegal discrimination against any individual:
www.alaska.edu/nondiscrimination(<http://www.alaska.edu/nondiscrimination>) .*