CDDIS Data Ingest Updates and Future Developments

Taylor Yates[1], Justine Woo[1], Nathan Pollack[1], Rebecca Limbacher[1], Marshall Finch[2], Patrick Michael[3]

Abstract NASA's Crustal Dynamics Data Information System (CDDIS) and the International VLBI (Very Long Baseline Interferometry) Service for Geodesy and Astrometry (IVS) have collaborated for several years to maintain completeness and availability of data and derived products for the geodesy community. To increase usability and better accommodate increasing storage needs introduced by the VGOS network and the SWIN dataset, the CDDIS is transitioning selected datasets to the AWS (Amazon Web Services) cloud environment. Additionally, the CDDIS strives to take advantage of its position serving multiple techniques by introducing best practices found in each technique to the rest of the data and product offerings. In 2023, several hardware and software upgrades were introduced that provide wide-reaching improvements to the CDDIS.

Keywords CDDIS, Geodesy, VLBI, SLR, GNSS, DORIS

1 Introduction

NASA's Crustal Dynamics Data Information System (CDDIS) continues to update and improve how space-geodetic data are ingested and stored in its repository, including moving some of its services into the cloud. In the following we describe a couple of examples from the GNSS world and cover our cloud migration efforts.

1. Science Systems and Applications
2. Adnet
3. Goddard Space Flight Center, NASA

2 GUARDIAN Products

The GUARDIAN (GNSS-based Upper Atmospheric Realtime Disaster Information and Alert Network) system is a near-real-time ionospheric monitoring software system for natural hazards warning. The CDDIS began hosting the GUARDIAN data in 2023. Hosting these data presented technical challenges due to processing a large quantity of small files with strict latency requirements. Dozens of GUARDIAN files arrive at the CDDIS every minute; therefore, a single threaded process would not be capable of sustaining service. Walking a directory tree on crontab would be too slow for incoming file detection, so this also required refactoring. These requirements were met by developing a new driver. It watches the kernel for uploads and dispatches worker threads to handle them. This is handled in a first-in-first-out (FIFO) manner, with the concurrent thread count limited by available processing cores. By implementing this new driver, we gained an order of magnitude of improvement to ingest latencies, with sub-second latencies from upload to public availability of the data. This new driver is running in production for GUARDIAN products, and the CDDIS plans to deploy it for other datasets as time permits

3 GNSS Real-Time Upgrade

The CDDIS provides a GNSS Real-Time Caster that streams data to customers. The CDDIS aimed to improve the availability and latency for customers.

Identified issues were a mean latency of two seconds and a standard deviation of 2.5 seconds. These latency spikes were difficult to trace, and a lack of di-

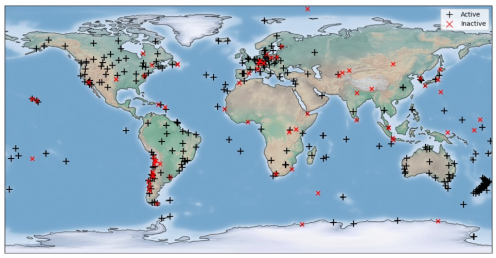

Fig. 1 CDDIS real-time caster station map.

agnostic tooling made root cause analysis difficult. In 2023 we opened a new caster to the public alongside the old caster while users migrated. This new caster was architected as a cluster of nodes, improving reliability and reducing the impacts of performing maintenance on the stream customers. We were able to reduce the mean latency to one second and the standard deviation to 0.75 seconds.

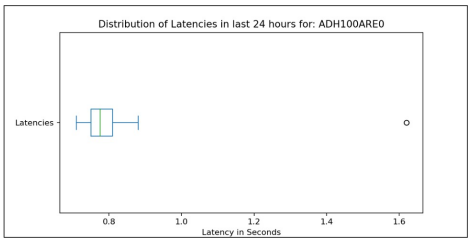

Fig. 2 Distribution of latencies.

The CDDIS also provides new visualizations for the new cluster. These include a stream status page, latencies in map and table formats, and box plots of latency distribution for each stream.

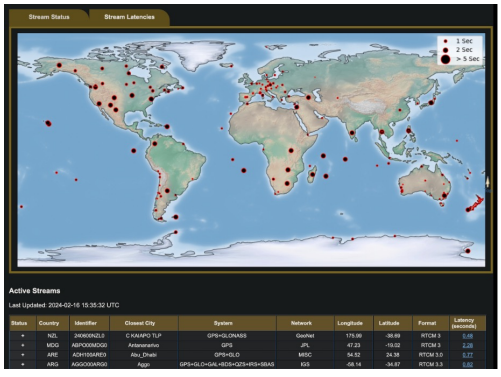

Fig. 3 Stream latencies map.

These latency maps and tables can be accessed on the CDDIS website (Figure 4) at the URL https://cddis.nasa.gov/Data_and_Derived_Products/GNSS/NTRIP_Caster_Streams_Status.html.

Fig. 4 https://cddis.nasa.gov.

4 Cloud Migration

The CDDIS is working towards deploying geodetic datasets to the cloud with Amazon Web Services (AWS). Motivations for this include better integration with data discovery tools that are provided by NASA such as the Central Metadata Repository (CMR) and EarthData Search and providing the ability to use data in place. It will be possible to continue to use the CDDIS as before. The directory structure and access via FTP-SSL and HTTPS will remain as before. In addition, the datasets hosted by CDDIS will be made available for cloud computing. To accomplish this, the CDDIS has joined other NASA teams in the Agile Scrum process in a team focused on the Ingest and Archive of data. The benefits of this are shared best practices and access to mentoring from teams that have already transitioned to AWS. There is a learning curve with any new technology, but the CDDIS is committed to helping users continue to access data as before. Currently, the CDDIS is building resources in a testing environment of AWS, and we will make an announcement before deploying a production environment to be used. The implementation order is listed in Figure 5.

Fig. 5 CDDIS cloud migration implementation order.