

GNSS RADIO OCCULTATION DATA IN THE AWS CLOUD: AWS ARCHITECTURE

A. McVey, S. Leroy, S. Leidner, and J. Martin

Verisk Atmospheric and Environmental Research

Multiple independent centers and retrieval systems have processed GNSS radio occultation data beginning with GPS/MET in 1995. Each center has processed different subsets of those data, but the retrieval algorithms have been nonhomogeneous in time as have the formats of the output data themselves. Moreover, there has been no universally agreed upon definition for the identification of a unique RO sounding, making matchups of RO retrievals difficult. Finally, access to RO data has proven limited because of the volume of data, constraints on bandwidth, and inadequate metadata on RO soundings. In order to enable easier and more efficient use of RO data for scientific studies, we at Verisk Atmospheric and Environmental Research (AER) have new standardized formats for RO data, translated repositories of RO data from UCAR COSMIC Project Office, the NASA Jet Propulsion Laboratory (JPL), and the Radio Occultation Meteorology Satellite Application Facility (ROM SAF) into the new formats, assigned unique identifiers to each RO sounding, and developed a database that enables RO sounding matchups between different contributing processing centers and allows for highly efficient subsetting of RO data according to geolocation, time, mission, transmitter, receiver, geometry (rising v. setting), and contributing processing center. Using Amazon Web Services (AWS) we were able to process and catalogue over 17 million RO soundings in a matter of a few days at minimal cost. We also developed an automated system that routinely checks for new RO data and then retrieves and processes them to keep the dataset current with a latency of approximately 2 months. All processes require a complex system that simultaneously translates incoming data and interacts with the database. To meet these objectives, we developed an AWS architecture to enable easy updates and efficient processing. AWS services used include Batch, DynamoDB, S3, ECR, Lambda, and IAM. This presentation will dive into detail on the AWS architecture and discuss the hurdles we had to overcome.