Accelerating Machine Learning with open standards and data

Roy Varghese[1], Tom Augspurger[1]
[1]Microsoft Corporation

Advances in weather forecasting and modelling, powered by high performance computing (HPC), have saved many lives and billions of dollars of property. To achieve this, complex workflows are being optimized with the latest scientific advancement as well as incorporation of AI/ML and deep learning techniques. As government organizations around the globe continue to collect and disseminate vast quantity of environmental datasets, their value for science is limited by the reality that those datasets are not often accessible to facilitate machine learning that can accelerate science. Microsoft's Planetary Computer offers a path forward to advancing science by combing a multi-petabyte catalog of global environmental data with intuitive APIs and a flexible scientific environment that allows users to answer global questions about that data.

The core of the Planetary Computer is a large collection of geospatial datasets, all available at scale through Azure Blob Storage. The data catalog includes many kinds of data, including remote sensing, climate model output, operational weather forecasts, biodiversity, and socio-economic. To facilitate easy access to the data, the Planetary Computer hosts a metadata API built using the open standard STAC, the SpatioTemporal Asset Catalog. With this API, users can query for specific records of interest by space, time, and more. Finally, users can compute on their data of interest, at scale, using the many compute options provided by Azure. The Planetary Computer includes a pangeo-style JupyterHub that's enabled with Dask for large-scale computing.

At this session, attendees will learn about Planetary Computer and it's capability as well as see a brief introduction to cloud-native geospatial data analysis at scale with the Planetary Computer. Presenters will be Roy Varghese, Director of Azure HPC and AI for Public Sector Americas and Tom Augspurger, Software Engineer for Microsoft Planetary Computer.