Warming climate is causing rapid and significant change to permafrost in the Arctic region. Fortunately a large quantity of satellite data is available for analysis. The goal of the Permafrost Discovery Gateway is to a) enable the creation of pan-Arctic geospatial products and b) make them accessible to both scientists and the public through visualization and analysis tools. To achieve the development of large geospatial data we are building a science gateway to manage hybrid machine learning pipelines using both Cloud and HPC Resources.

Part of this pipeline takes high resolution satellite imagery and maps permafrost thaw features across the Arctic region. This novel high performance image analysis framework, Mapping application for Arctic Permafrost Land Environment (MAPLE), detects ice wedge polygons from very high resolution optical imagery data archived at the Polar Geospatial Center, in three steps. The first step is image preprocessing, the second is DLCNN (Deep Learning Convolutional Neural Network) prediction, followed by a third post-processing step. The first and third steps have CPU implementations, but the DLCNN requires GPU resources.

Furthermore we create geospatial datasets of lake area change, fire scars and retrogressive thaw slumps, which occurred over the past 20 years across the Arctic permafrost region. These datasets are based on Landsat, which are pre-processed through Google Earth Engine and further analyzed using machine learning and geospatial data analysis in an automated processing pipeline.

For visualization, we incorporate Cesium as a 3D tile-based Imagery Viewer that allows exploration of pan-Arctic, sub-meter map products over time and can be exported as publication-quality map images. We also incorporate the Fluid Earth Viewer to enable global and regional visualization of Arctic data products over time. A third visualization tool will include the 2D-4D graph plotting of the big geospatial data.

We host data on an instance of Clowder using a Kubernetes cluster hosted on NCSA Radiant Openstack. We have adapted existing workflows (MAPLE and the analysis of data from Google Earth engine) as Clowder information extractors. Jobs that do not require GPU resources are executed in the local Clowder cluster, while those that require GPU resources are submitted to external clusters , such as XSEDE Bridges2, and the results uploaded back to the Clowder instance. We are in the process of automating the data ingestion and processing step.

Together, these components provide a starting environment to support permafrost science. Our long term goal is to apply lessons learned from implementing these solutions for specific use cases to other research questions around the study of the Arctic region.