

GEOANALYTICS.CA – A DEMONSTRATION CLOUD-NATIVE OPEN PLATFORM FOR BIG DATA GEOSCIENCE

Jason Suwala, Andy Dean, Matthew Fung, Hussain Hassam, Barry Pierce

Hatfield Consultants, 200-850 Harbourside Drive, North Vancouver, BC V7P 0A3

ABSTRACT

Current data-driven and computing intensive research on climate-change, ecosystem modeling, and environmental and natural resources monitoring is based on the collection, management, analysis, and dissemination of geospatial data. Currently, the acquisition and analysis of satellite Earth Observation (EO) data is most often executed through traditional, computational, and data analysis approaches, which require users to download data to their desktops to perform any analysis. As the volume of satellite EO data continues to grow, new analytical possibilities arise requiring new approaches to data management and processing. The traditional approach to acquire and process satellite EO data sets cannot address this new potential and quickly becomes unsustainable due to the volume of EO data to be analyzed.

To demonstrate how cloud computing systems can overcome issues with traditional approaches to satellite EO data analytics, Hatfield created the GEO Analytics Canada platform. Our approach to designing and implementing the GEO Analytics Canada Demonstration Platform integrates the following:

- Bring the user to the data – to achieve high performance geospatial data analytics, it's critical to bring the user to the data and avoid downloading wherever possible.
- Cloud native - Cloud geospatial involves more than simply migrating desktop apps to the cloud. The GEO Analytics Canada platform is built from the ground-up to leverage the power of cloud computing.
- Infrastructure vendor agnostic – the GEO Analytics Canada platform can be installed on a wide variety of cloud computing providers. We can pursue hybrid and multi-cloud architectures that exploit pre-existing distributed data stores, such as Landsat and Sentinel data.
- Part of an ecosystem of open architected systems - GEO Analytics Canada platform is a starting point towards an open architected, distributed ecosystem approach to satellite EO data analytics. We believe that platforms should not require all data and tools to be centralized in one place. Instead, data and processing resources should be distributed to exploit pre-existing distributed data stores.
- Supporting open science – all GEO Analytics Canada platform tools and systems support the key tenants of open science: “openness, transparency, scrutiny and traceability of results, access to large volume of complex data, and the availability of community open tools”.
- Canadian focused – the platform stores its data completely in Canada and uses Canadian hosted compute resources. This supports Canadian organizations that are required to fulfill Canadian privacy laws which require data to be kept in Canada.

The Demonstration Platform comprises custom built, fully integrated systems that are built on top of cloud-based storage and computational systems. These systems remove the need to download EO data in order to conduct visualization and analysis of large satellite EO datasets in a scalable, performant manner. Tools provided in the platform include: authentication, security and user management systems; EO data query and discovery systems; massively scalable EO data ingestion and pre-processing systems; a Jupyter-Lab based scalable data analysis environment; on-demand personal Ubuntu desktops in a browser; a file browser system; and a ground truth data management system. For machine-to-machine integrations, SpatioTemporal Asset Catalogue (STAC) and OGC API-Features API endpoints have been included.

Keywords— cloud native geospatial, big geospatial data analytics, ground truth data management, STAC, OGC API-Features