Contribution ID: **30**   Type: **Talk**

Implementing an Analysis Framework in the Marine Data Portal of the German Marine Research Alliance (DAM)

Friday 5 April 2024 09:25 (15 minutes)

The continuous growth of Earth System data, coupled with its inherent heterogeneity and the challenges associated with distributed data centers, necessitates a robust framework for efficient and secure data analysis. This abstract outlines the plans to implement an analysis framework within the marine data portal of the German Marine Research Alliance (Deutsche Allianz Meeresforschung, DAM) at https://marine-data.de. The proposed framework is based on the Data Analytics Software Framework (DASF), chosen for its decentralized, secure, and publisher-subscriber-based (pub-sub-based) architecture, which enables the execution of data analysis backends anywhere without exposing sensitive IT systems to the internet.

The challenges of analyzing Earth System data on the web are multifaceted. Data heterogeneity arises from the diverse sources, formats, and structures of earth system data, making seamless integration and analysis a complex task. The sheer volume of data compounds the challenge, demanding scalable solutions to handle vast amounts of information efficiently. Additionally, the computational power required for meaningful analysis is often expensive and can become a bottleneck in traditional data processing pipelines. Moreover, the distributed nature of data across multiple centers poses logistical challenges in terms of accessibility, security, and coordination.

To address these challenges, the integration of DASF into the marine data portal presents a comprehensive solution. DASF offers a secure and decentralized pub-sub-based remote procedure call framework, providing a flexible environment for executing data analysis backends. One of the key advantages of DASF is its ability to allow these backends to run anywhere without the need to expose sensitive IT systems to the internet, addressing the security concerns associated with data analysis.

The decentralized nature of DASF also mitigates data heterogeneity challenges by offering a unified platform for data integration and analysis. With DASF, disparate data sources can seamlessly communicate, facilitating interoperability and enabling comprehensive analysis across diverse datasets. The pub-sub mechanism ensures efficient communication between components, streamlining the flow of data through the analysis pipeline.

Security is a critical aspect of implementing a robust data analysis framework. DASF addresses this concern by incorporating an OAuth-based authentication mechanism at the message broker level. This ensures that only authorized users can access and interact with the data analysis functionalities. Additionally, the integration with the Helmholtz AAI empowers the sharing of analysis routines with users from other research centers or the general public.

The cost-effectiveness of DASF further enhances its appeal, as it optimizes the utilization of computational resources. By enabling the deployment of analysis components on diverse hardware environments, organizations can leverage existing infrastructure without significant additional investments.

In conclusion, the integration of DASF into the DAM portal marks a significant step toward overcoming the challenges inherent in analyzing Earth System data on the web. By addressing data heterogeneity, accommodating vast datasets, and providing a secure and decentralized architecture, DASF emerges as a key enabler for efficient and scalable data analysis. The adoption of DASF in the marine data portal promises to enhance the accessibility, security, and cost-effectiveness of data analytics, and finally facilitates open science in the research field Earth and Environment.

Primary authors: SOMMER, Philipp Sebastian (Helmholtz-Zentrum Hereon); HESS, Robin (Alfred-Wegener-Institut); SASS, Björn Lukas (Helmholtz-Zentrum Hereon); SCHAEFER, Angela (Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung)

Presenter: SOMMER, Philipp Sebastian (Helmholtz-Zentrum Hereon)

Session Classification: Session 3: Governmental Data and Transfer

Track Classification: Governmental Data and Transfer