

DARE as a platform to support Climate Data Analytics using Cloud Infrastructures

Thursday, 11 October 2018 15:00 (15 minutes)

Supporting data analytics in climate research with respect to data access is a challenge due to increasing data volumes, especially for end users, as the whole climate data archive is expected to reach a volume of 30 Pb in 2018 and up to 2000 Pb in 2022. Several international and European initiatives have emerged and provide standalone solutions that offer potential for interoperability. The DARE e-science platform (<http://project-dare.eu>) is designed for efficient and traceable development of complex experiments and domain-specific services on the Cloud.

In Europe, the IS-ENES (<https://is.enes.org>) consortium has developed a platform, that is a component of the ENES CDI (Climate Data Infrastructure), to ease access to climate data for the climate impact community (C4I: <https://climate4impact.eu>). One of the important aspect of the C4I platform is that it enables users to perform on-demand data analysis calculations through its backbone based on a collection of OGC WPS (Web Processing Service). These, coupled with authorization mechanisms based on access tokens, enable the delegation of the calculations onto distributed infrastructures and the controlled management of the results.

These characteristics have been further extended with provenance integration, especially to obtain the traceable calculation of climate impact indicators, in the context of the FP7-CLIPC project. A solution based on a standard representation (W3C-PROV) and a set of lineage management and workflows tools that will scale to other computational use cases, and that will be interoperable with ongoing European initiatives. In the DARE project, the provenance system will be also built on top of W3C-PROV, ensuring interoperability.

DARE will also integrate services from the EUDAT CDI, enabling generic access and cross-domain interoperability, as well as providing compliance and integration with the future EOSC platform. As DARE will use containerization technologies, it will be easily deployed on heterogeneous architectures.

A scientific pilot has been designed within the DARE project for the ENES community (climate domain). The objectives are to enable delegation of on-demand computational-intensive calculations to the DARE platform. In the presented Use Case, on-demand data analytics will be initiated on the IS-ENES C4I platform by end users of climate data, in a seamless fashion. A schematics of the architecture and Use Case will be presented, along with initial development status.

Type of abstract

Presentation

Summary

Supporting data analytics in climate research with respect to data access is a challenge due to increasing data volumes, especially for end users, as the whole climate data archive is expected to reach a volume of 30 Pb in 2018 and up to 2000 Pb in 2022. Several international and European initiatives have emerged and provide standalone solutions that offer potential for interoperability. The DARE e-science platform (<http://project-dare.eu>) is designed for efficient and traceable development of complex experiments and domain-specific services on the Cloud.

Primary author: PAGE, Christian (CERFACS)

Co-authors: SPINUSO, Alessandro (KNMI); KLAMPANOS, Iraklis (National Centre for Scientific Research "Demokritos"); PLIEGER, Maarten (KNMI); ATKINSON, Malcolm (UE); KARKALETSIS, Vangelis (National Centre for Scientific Research "Demokritos"); CERFF, W.J. (KNMI)

Presenter: PAGE, Christian (CERFACS)

Session Classification: Thematic Services

Track Classification: Area 3. Computing and Virtual Research Environments