SeaDataCloud Virtual Research Environment (VRE): Implementation and Technical Aspects

IMDIS-remote, 12-14 April 2021

Merret Buurman (German Climate Computing Centre, DKRZ, Hamburg) for the SeaDataCloud VRE development team
SeaDataCloud Virtual Research Environment (VRE): Implementation and Technical Aspects

IMDIS-remote, 12-14 April 2021

Merret Buurman (German Climate Computing Centre, DKRZ, Hamburg) for the SeaDataCloud VRE development team
SeaDataCloud Virtual Research Environment (VRE): Implementation and Technical Aspects

IMDIS-remote, 12-14 April 2021

Merret Buurman (German Climate Computing Centre, DKRZ, Hamburg) for the SeaDataCloud VRE development team
SeaDataCloud Virtual Research Environment (VRE): Implementation and Technical Aspects

IMDIS-remote, 12-14 April 2021

Merret Buurman (German Climate Computing Centre, DKRZ, Hamburg) for the SeaDataCloud VRE development team
Large team!

- **Charles Troupin**, GHER/ULiège (Belgium), ctroupin@uliege.be
- **Alexander Barth**, GHER/ULiège (Belgium), a.barth@uliege.be
- **Leo Bruvry-Lagadec**, IFREMER (France), Leo.Bruvry.Lagadec@ifremer.fr
- **Sebastian Mieruch**, Alfred-Wegener-Institute (Germany), sebastian.mieruch@awi.de
- **Naranyanan Krishnan**, UKRI STFC (United Kingdom), naranyanan.krishnan@stfc.ac.uk
- **Giorgio Santinelli**, Deltares (The Netherlands), Giorgio.Santinelli@deltas.nl
- **Fedor Baart**, Deltares (The Netherlands), fedor.baart@deltas.nl
- **Peter Thijsse**, MARIS (The Netherlands), peter@maris.nl
- **Filip Waumans**, Vlaams Instituut voor de Zee (Belgium), filip.waumans@vliz.be
- **Themis Zamani**, Greek Research and Technology Network (Greece), themis@grnet.gr
- And more!

Thanks to all of you!
Focus of this presentation

This abstract focuses on the architecture and technical details!

For the user-facing functionalities, check out the presentation: "Working with the SeaDataCloud VRE: what can we do for you?“

Charles, Tuesday 13:56 – right after this!
What is a Virtual Research Environment (VRE)?

A VRE is an online environment for processing data - marine data, in this case. It provides a number of processing services that researchers use to perform their analyses. These services are mostly existing softwares (DIVA, ODV, etc.), mostly developed for Desktop usage, which we ported to a cloud-based environment.
Deployment basics

• „Cloud-based“ - load is distributed among various computing centres

• All services and tools run as Docker Containers, providing a standardized operating environment across all data centres (to facilitate deployment, load distribution, and the integration of heterogeneous services).
Two types of processing services

1 One instance serves many users - one Docker container is running continuously and waiting for users (BioQC, VIZ-ualisation, webODV).

2 Every user gets individual instance: One container per user (DIVA and ERDDAP).

- For more computationally intensive tools – easy to scale!
- For tools that are not optimized for handling multiple users sessions (this frequently occurs when desktop software is ported to the cloud).
(Mis?) Using JupyterHub for type II

JHub was developed to serve JupyterNotebooks but can be used for any tool that is "dockerized“ and interacts with the user via HTTP.

We use JHub to spawn a container when a user logs in.

Benefits:

- instance management
- authentication/authorization
- web security measures (reverse proxying, SSL termination, ...)

10
For later reference...

Shared central services
To sum up...

- Container-based deployment: convenient to deploy/maintain/distribute.
- One-container-for-all vs. one-container-per-user
- JupyterHub very useful for the latter! (Not just for Jupyter Notebooks)
Thank you!

Any questions?

buurman@dkrz.de
Subsetting service

The VRE's **subsetting service** embeds an instance of ERDDAP [https://coastwatch.pfeg.noaa.gov/erddap/index.html], configured dynamically and started on the fly according to a dataset selected by the user via the file selector. Its docker image includes the ERDDAP web application running in an Apache Tomcat web server and a visualization frontend using VueJS framework.
Summing up...

This was a short introduction to the technical side of our VRE. The container-based deployment has proved stable and convenient to maintain. In future, we hope to be able to extend and improve the VRE. Besides incorporating user feedback, possible areas of improvement concern the container orchestration (e.g. by introducing technologies such as *kubernetes* or *Apache Mesos & Marathon*) and the synchronization of data between the data centres.