Supporting Multi-Cloud Model Execution with VLab

M. Santoro (1), P. Mazzetti (1), N. Spadaro (2), S. Nativi (2)

(1) CNR-IIA
(2) EC-JRC
The VLab

- Developed in the context of the European projects ECOPOTENTIAL and ERA-PLANET
- VLab is a cloud-based platform to support the activity of environmental scientists in sharing their models
- The main challenges addressed by VLab are:
  - Minimization of interoperability requirements in the process of model porting (i.e. to simplify as much as possible the process of publishing and sharing a model for model developers)
  - Support multiple programming languages and environments (it must be possible porting models developed in different programming languages and which use an arbitrary set of libraries)
VLab Concept

The ECOPOTENTIAL VLAB Concept
Code Sharing and Containerization Technologies

Git

Git is a distributed version control system designed to handle everything from small to very large projects with speed and efficiency.

Docker

Docker is a tool designed to make it easy to run applications by using containers. Containers allow a developer to package up an application with all of the parts it needs.
Usage of Git

**Source code of the model**
- Python files
- Java files
- Etc.

**Scripts to launch the model execution**
- sh files
- exe files

**Machine-readable information about:**
- Data input ingestion
- Model execution
Usage of Docker

- **Model Environment**
  - Vlab uses the provided Docker image to instantiate and run Docker container, which is where the actual model execution takes place.
  - If the model is coded in python, a Docker image with the required python installation libraries must be defined by model provider.

- **Resources**
  - When configuring docker image, required resources such as memory (RAM) can be defined.
How VLab works

1. User requests model execution
2. VLab fetches model source code and instantiates Docker container
3. VLab downloads model inputs and ingests them in the execution environment
4. VLab launches the Docker container on one of the computing platforms
5. Output is saved and displayed to user
Cloud Services for Model Execution

The design of VLab allows the execution of models on different cloud environments. To this aim, VLab utilizes a set cloud services. Following table lists which cloud specific technology was utilized in the different environments.

<table>
<thead>
<tr>
<th>Cloud Environment</th>
<th>Amazon Web Services</th>
<th>OpenStack/Kubernetes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud Hosting</td>
<td>Elastic Compute Cloud</td>
<td>OpenStack Compute (nova)</td>
</tr>
<tr>
<td>Auto Scaling</td>
<td>Auto Scaling Group</td>
<td>OpenStack Heat</td>
</tr>
<tr>
<td>Container Management</td>
<td>Elastic Container Service</td>
<td>Kubernetes</td>
</tr>
<tr>
<td>Shared File System</td>
<td>Elastic File System</td>
<td>Network file System (Shepler, et al., 2010) + Kubernetes Persistent Volume</td>
</tr>
<tr>
<td>Web Storage</td>
<td>Simple Storage Service</td>
<td>MinIO (under experimentation)</td>
</tr>
<tr>
<td>Hosted Queue</td>
<td>Simple Queue Service</td>
<td>KubeMQ (under experimentation)</td>
</tr>
</tbody>
</table>
Advantages of Supporting Multi-Cloud Execution

- Allows user (or client application) to choose which platform to use for a specific model run, e.g. depending on the required data which might be available on a specific platform only (move code to data).

- Another possible scenario empowered by this multi-platform support feature is the possibility to let the user choose the computational platform and utilize her/his credentials to request the needed computational resources.

- Finally, it is also possible to exploit this feature for benchmarking different cloud platforms with respect to their performances.
VLab was experimented on several processing platforms.
Demonstration

- In the context of the EuroGEOSS Sprint to Ministerial activity, promoted by the European Commission, a web application was developed by the Joint Research Centre of the European Commission (EC JRC) to show the use of Copernicus Sentinel data to calculate Land Cover and Land Cover change in a set of Protected Areas belonging to different ecosystems.
- To this aim, the Earth Observation Data for Ecosystem Monitoring (EODESM) model (Lucas & Mitchell, 2017) was utilized to classify land covers and change.
- The web application utilizes VLab APIs to select the area and dates of interest and request the execution of EODESM model on one of the supported cloud environments.

Figure 10 - Protected Areas Analysis Web Application (https://pademo.geodab.org/)
Future Work

- Enhance multi cloud support in Vlab implementing:
  - automatic selection of cloud according to data availability
  - use of users' credentials to request the needed computational resources
- Enhance VLab to support model chaining
- Integrate VLab with Knowledge repositories to increase the interoperability level to semantic and pragmatic
Acknowledgement

The research leading to these results has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreements N. 641762 (ECOPotential), 689443 (ERA-PLANET), 777536 (EOSC-hub), 776136 (EDGE)