Designing a Cloud and HPC Based M&S platform to Investigate the IVD Diseases Mechanisms

Maria Paola Ferri
Barcelona Supercomputing Center
Spain
maria.ferri@bsc.es

Laia Codó Tarraubella
Barcelona Supercomputing Center
Spain
laia.codo@bsc.es

Josep Lluis Gelpi Buchaca
Barcelona Supercomputing Center
Spain
josep.gelpi@bsc.es

1. Introduction

The main issue with the investigation for multi-factorial pathologies is the high variety and contributions from cross related biological, computational and medical fields. Their integration into a holistic and rational map is the underlying purpose of creating a centralized secure and easy-to-access platform, powered by a tailor-made Cloud and HPC environments. The platform concerned is developed in the HORIZON EU MSCA Disc4All, for data sharing and tools analysis related to the Intervertabral Disc Degeneration (IVD), as on of the main causes related to LBP (Low Back Pain).

2. Methods

The Cloud dedicated resources are allocated in the **ELIXIR-ES** Cloud infrastructure, on-premise cloud located at the Barcelona Supercomputing Center (BSC).

1. The tools provided within the project, that would be used to generate an integrated workflows for simulations and investigation, first they would be adapted into **BioBBs (BioExcel Building Blocks).** The BioBB-based workflow would be deployed in NextFlow, an emergent workflow manager that supports already Docker and Singularity.

2. A hybrid computing infrastructure is built, depending on the tools requirements, through the use of Docker/Singularity containers and HPC job submission service combined. The computational infrastructure was set up on the **OpenVRE** as a **PaaS** as the foundation. To include the simulation tools, it is necessary to wrap them following the OpenVRE protocol. Subsequently, the tools and workflows would be exposed for users to a user-friendly front-end, to grant access to a virtual environment for the analysis tools and workflow.

2.1 Infrastructure Architecture

Its components would be designed and implemented as independent Docker containers:

- A **Web-base front-end** to provide a unique access point;

- A **Data Catalog** built on the metadata provided by the data repositories;

- An **Access Manager,** supporting authentication for OpenID Connect access providers;

- A **Virtual Research Environment,** as a central workspace for the users, built on top of the **OpenVRE** framework, and relying on a OCCI Cloud provider- such as OpenStack and/or Microsoft Azure;

- A **Data Storage** to provide for data synchronization among data providers, based on MongoDB database manager;

- **Software Schedulers**, to manage analysis workflows in an adapt following computational demands.
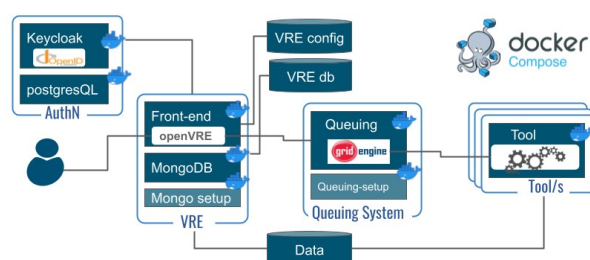
3. Results

The most complete BioBB in production at the moment is based on the prioritization algorithm of genes variants attributed to the IVD. This was used as a first attempt for a BioBB-based workflow in Nextflow. To test the efficiency of the mentioned workflow manager, also a use case for a simulation for high performance computational mechanics was implemented in Nextflow, launched on the *MareNostrum* supercomputer (BSC) through **BioBB Remote,** ensuring the access through Ssh credentials locally. Instantiated on Docker containers, the platform is up and available at the url *www.disc4all.eu*, and the aforementioned installation for the BioBB-based version for the tools are in progress to be inserted in the platform.

ACKNOWLEDGMENTS

REFERENCES

[1] *BioExcel Building Blocks, a software library for interoperable biomolecular simulation workflows*. **Pau Andrio et Al. Nature Scientific Data, 09/2019,Volume 6, Issue 1, p.169, (2019).**

[2] *Nextflow enables reproducible computational workflows*. **Di Tommaso et AL. (2017). Nature Biotechnology, 35(4), 316–319. doi:10.1038/nbt.3820**

[3] **Open Virtual Research Environment (openVRE) [https://github.com/inab/openVRE]**